Scale factor dependent equation of state for curvature inspired dark energy, phantom barrier and late cosmic acceleration

S.K. Srivastava

Department of Mathematics, North Eastern Hill University, Shillong-793022, India

Received 25 August 2006; accepted 11 October 2006
Available online 30 October 2006
Editor: T. Yanagida

Abstract

Here, it is found that dark energy and dark matter emerge from the gravitational sector, if non-linear term of scalar curvature is added to Einstein–Hilbert Lagrangian. An equation of state for dark energy, having the form \(p_{\text{de}} = -\rho_{\text{de}} + f(a)\) (with \(p_{\text{de}}\) (\(\rho_{\text{de}}\)) being the pressure (density) for dark energy, \(f(a)\) being a function of scale factor \(a(t)\) and \(t\) being the cosmic time) is explored. Interestingly, this equation of state leads to a phantom barrier \(w_{\text{de}} = -1\) at \(a = a_w\). It is found that when \(a < a_w, w_{\text{de}} > -1\) and \(w_{\text{de}} < -1\) for \(a > a_w\), showing a transition from non-phantom to phantom phase at \(a = a_w < a_0\) (\(a_0\) being the current scale factor of the universe).

© 2006 Published by Elsevier B.V.

PACS: 98.80.Cq

Late cosmic acceleration is the most remarkable astronomical observation in the recent past [1,2]. Theoretically, it is possible when the universe is driven by a fluid having negative pressure and equation of state parameter (EOSP) \(w_{\text{de}} < -1/3\). Observational data also indicate that \(w_{\text{de}}\) should be close to \(-1\) (the phantom barrier) and most probably less than \(-1\) for the current universe. The source of this mysterious fluid is still in dark. So, many phenomenological models, suggesting possible sources of dark energy (DE), have been proposed in the past few years [3,4]. In spite of non-gravitational sources of DE, non-linear terms of curvature are also proposed as gravitational alternative for DE [5]. A recent comprehensive review on DE is available in [6].

In [5], non-linear terms of curvature are taken as DE Lagrangian a priori and its various consequences are discussed. In the present model, the story is different. Here also, investigations begin from the modified gravity stemming from addition of non-linear term to Einstein–Hilbert Lagrangian, but unlike [5], here, non-linear term of curvature is not taken as a DE source. Rather, DE emerges spontaneously as a combined effect of linear as well as non-linear terms of Ricci scalar curvature [7,8], whereas, in [5], only non-linear term of curvature contributes to DE. Interestingly, here, it is found that curvature can be a possible source of dark matter too. In [9], various equations of state (EOS) for DE, dependent on Hubble’s expansion rate \(H\) and its derivatives, are proposed. Contrary to this, in what follows, EOS for DE is not proposed but derived. EOS for DE, obtained here, has the form \(p_{\text{de}} = -\rho_{\text{de}} + f(a)\) (with \(\rho_{\text{de}}, p_{\text{de}}, a(t)\) and \(t\) being dark energy density, pressure, scale factor and cosmic time respectively). It is interesting to see that \(a = a_w < a_0\) (\(a_0\) is the current scale factor) gives the phantom barrier \(w_{\text{de}} = -1\). It is found that when \(a < a_w, -1 < w_{\text{de}} < -1/3\) and \(w_{\text{de}} < -1\) for \(a > a_w\). It shows a transition from a quintessence to phantom phase at \(a = a_w\). In [10], a possibility of this type of transition is discussed taking EOS in Jordan frame. The present work is different from [10] as no EOS is taken here a priori, rather it is derived yielding phantom barrier and quintessence to phantom transition spontaneously.

Natural units (\(h = c = 1\)) are used here with GeV as the fundamental unit, where \(h\) and \(c\) have their usual meaning. In this unit, it is found that \(1\,\text{GeV}^{-1} = 6.58 \times 10^{-25}\) s.
The action for higher-derivative gravity is taken as

\[
S_g = \int d^4x \sqrt{-g} \left[ \frac{R}{16\pi G} - \alpha R^{(2+r)} \right],
\]

(1)

where \( R \) is the Ricci scalar curvature, \( G = M_P^{-2} \) (\( M_P = 10^{19} \text{ GeV} \) is the Planck mass). Moreover, \( \alpha \) is a coupling constant having dimension (mass)\(^{-2}\) with \( r \) being a positive real number. As \( (2+r) > 0 \), instability problem does not arise like the model containing the \( R^{-1} \) term [11]. But, as mentioned above, here approach is different from the papers [10,11].

The action (1) yields gravitational field equations

\[
\frac{1}{16\pi G} \left( R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R \right) - \alpha \left\{ (2+r) \left\{ \nabla_\mu \nabla_\nu R^{(1+r)} - g_{\mu\nu} \Box R^{(1+r)} + R^{(1+r)} R_{\mu\nu} \right\} - \frac{1}{2} g_{\mu\nu} R^{(2+r)} \right\} = 0
\]

(2)

using the condition \( \delta S_g / \delta g^{\mu\nu} = 0 \). Here, \( \nabla_\mu \) denotes covariant derivative and the operator \( \Box \) is given as

\[
\Box = \frac{1}{\sqrt{-g}} \frac{\partial}{\partial x^\mu} \left( \sqrt{-g} g^{\mu\nu} \frac{\partial}{\partial x^\nu} \right)
\]

(3)

with \( \mu, \nu = 0, 1, 2, 3 \) and \( g_{\mu\nu} \) as metric tensor components.

Taking trace of (2) and doing some manipulations, it is obtained that

\[
\Box R + \frac{r}{R} \nabla^\nu R \nabla_\nu R = \frac{1}{3(2+r)(1+r)} \left[ \frac{R^{(1-r)}}{16\pi G \alpha} + r R^2 \right] 
\]

(4)

with \( \alpha \neq 0 \) to avoid the ghost problem.

Experimental evidences support spatially homogeneous flat model of the universe [12]. So, the line-element, giving geometry of the universe, is taken as

\[
dS^2 = dt^2 - a^2(t) \left[ dx^2 + dy^2 + dz^2 \right]
\]

(5)

with \( a(t) \) as the scale factor. It gives expansion rate \( H = \dot{a}/a \).

In the homogeneous space–time, given by (5), (4) is obtained as

\[
\ddot{R} + \frac{3}{2} \frac{\dot{R}}{a} + \frac{r \dot{R}^2}{R} = \frac{1}{3(2+r)(1+r)} \left[ \frac{R^{(1-r)}}{16\pi G \alpha} + r R^2 \right].
\]

(6)

In most of the situations, for example, radiation model, matter-dominated model, and accelerated models, we have \( a(t) \) as a power-law solution yielding \( R \) as the power-law function of \( a(t) \). So, it is reasonable to take \( R \) as

\[
R = \frac{A}{a^n}
\]

(7)

with \( n > 0 \) being a real number and \( A \) being a constant with mass dimension 2.

\( R \), given by (7), satisfies (6), if

\[
\frac{\ddot{a}}{a} + (2 - rn - n) \left( \frac{\dot{a}}{a} \right)^2 = \frac{1}{3(2+r)(1+r)} \left[ - \frac{a^{3r}}{8\pi G \alpha} - \frac{r A}{n} a^{-n} \right].
\]

(8)

Eq. (8) integrates to

\[
\left( \frac{\dot{a}}{a} \right)^2 = - \frac{C}{a^{2[3-n(1+r)]}} - \frac{1}{3(2+r)(1+r)} \left[ \frac{a^{3r}}{8\pi G \alpha} - \frac{r A}{3n(-n + 2[3 - n(1+r)])} a^{-n} \right]
\]

with \( C \) being an integration constant. Setting \( n = 3 \), (9) is rewritten as

\[
\left( \frac{\dot{a}}{a} \right)^2 = - Ca^{6r} + \frac{a^{3r} A^3}{72\pi Gar(2+r)(1+r)} a^{3r} + \frac{r A}{27(1+2r)(2+r)(1+r)a^{2r}}.
\]

(10)

This is the modified Friedmann equation giving dynamics of the universe. The third term on r.h.s. (the right-hand side) of this equation has the form of density for pressureless matter. This term emerges due to non-linear term of curvature in the action (1), hence it is termed as dark matter density. So, \( 8\pi G \rho_{dm} = \frac{r A}{27(1+2r)(2+r)(1+r)a^{2r}} \).

(11)

The first term, on r.h.s. of (10), emerges spontaneously and second term is the combined effect of linear as well as non-linear term of curvature in the action (1). A very interesting cosmic scenario is obtained on using these two terms (first and second) and taking energy density \( \rho_{de} \) as

\[
\rho_{de} = B - \frac{A^r}{72\pi Gar(2+r)(1+r)} a^{3r} \left[ 1 - \frac{a^{3r}}{2\lambda} \right].
\]

(12a)

with

\[
\lambda = \frac{A^r}{144\pi Gar(2+r)(1+r)C}.
\]

(12b)

If \( \rho_{de} = \rho_{de(s)} \) at \( a = a_s \) such that \( a_s^{3r} = 2\lambda \), (12a) looks like

\[
\rho_{de} = \rho_{de(s)} - \frac{A^r}{72\pi Gar(2+r)(1+r)} a^{3r} \left[ 1 - \frac{a^{3r}}{2\lambda} \right].
\]

(13)

According to WMAP [13], current values of dark matter density and dark energy density are \( \rho_{dm}^0 = 0.23\rho_c^0 \) and \( \rho_{de}^0 = 0.73\rho_c^0 \) with

\[
\rho_c^0 = \frac{3H_0^2}{8\pi G}.
\]

(14)

\( H_0 = 100h \text{ km/Mpc s} = 2.32 \times 10^{-42}h \text{ GeV} \) and \( h = 0.68 \). Using these observational values in (11) and (13), we obtain

\[
\frac{8\pi G}{3} \rho_{dm} = 0.23H_0^2 \left( \frac{a_0}{a} \right)^3 \frac{[1 - a^{3r}/2\lambda]}{[1 - a_0^{3r}/2\lambda]}.
\]

(15)

and

\[
\rho_{de} = \rho_{de(s)} - (\rho_{de(s)} - \rho_{de}^0) \left( \frac{a}{a_0} \right)^{3r} \frac{[1 - a^{3r}/2\lambda]}{[1 - a_0^{3r}/2\lambda]}.
\]

(16)
where
\[ \frac{A'}{72\pi Ga r (2+r)(1+r)} = (\rho_{de(s)} - \rho_{de}) \left( a_0^3 \left[ 1 - a_0^{3r}/2\lambda \right] \right)^{-1}. \] (17)

Using (11), (15) and (17), Friedmann equation (10) looks like
\[ \left( \frac{\dot{a}}{a} \right)^2 = (\rho_{de(s)} - \rho_{de}^0) \left( \frac{a}{a_0} \right)^3 \left[ 1 - a^{3r}/2\lambda \right] \]
\[ + 0.23H_0^2 \left( \frac{a_0}{a} \right)^3. \] (18)

If the term, proportional to \( a^{-3} \), dominates other terms on the r.h.s. of (18), it reduces to
\[ \left( \frac{\dot{a}}{a} \right)^2 \simeq 0.23H_0^2 \left( \frac{a_0}{a} \right)^3 \]
yielding
\[ a(t) = a_d \left[ 1 + 0.72H_0a_d^{-3/2}(t - t_d) \right]^{2/3} \] (19)
which shows decelerated expansion as \( \ddot{a} < 0 \). Here \( a_d \) and \( t_d \) are constants.

When other terms dominate the term proportional to \( a^{-3} \) on the r.h.s. of (18) and \( \rho_{de(s)} > \rho_{de}^0 \) (18) reduces to
\[ \left( \frac{\dot{a}}{a} \right)^2 = (\rho_{de(s)} - \rho_{de}^0) \left( \frac{a}{a_0} \right)^3 \left[ 1 - a^{3r}/2\lambda \right] \]
\[ \frac{1}{1 - a_0^{3r}/2\lambda}, \] (20)
which is integrated to
\[ a(t) = a_s \left[ \frac{a_0^{3r}}{2\lambda} + \left\{ \sqrt{1 - \frac{a_0^{3r}}{2\lambda} - \frac{3}{2} r B a_0^{3r/2}(t - t_0)} \right\}^2 \right]^{-1/3r} \] (21)
giving acceleration as \( \ddot{a} > 0 \). Here, \( t_s \) is the time for transition from deceleration to acceleration and \( a_s \) is the corresponding scale factor. Moreover, the constant \( B \) in (21) is given by
\[ B^2 = (\rho_{de(s)} - \rho_{de}^0) (a_0^{3r}/2\lambda)^{-1}. \] (22)

It is obvious that for \( t > t_s \), (18) reduces to (20). So, if \( \rho_{de(s)} < \rho_{de}^0 \), (20) shows that \( (\dot{a}/a)^2 < 0 \) for \( t > t_s \). It leads to an unphysical situation. So, this possibility is rejected and it is concluded that
\[ \rho_{de(s)} > \rho_{de}^0. \] (23)

Further, DE conservation equation is given as
\[ \dot{\rho}_{de} + 3H(\rho_{de} + \rho_{de}) = 0. \] (24)

Connecting (16) and (24), it is obtained that
\[ \rho_{de} = -\rho_{de} + \left( \rho_{de(s)} - \rho_{de}^0 \right) \left( \frac{a}{a_0} \right)^3 \left[ 1 - a^{3r}/\lambda \right] \]
\[ \left[ 1 - a_0^{3r}/2\lambda \right], \] (25)
This equation shows that at \( a = a_t^{1/3r} = a_w \), \( \rho_{de} = -\rho_{de} \). So,
\[ \rho_{de} = -\rho_{de} + \left( \rho_{de(s)} - \rho_{de}^0 \right) \left( \frac{a}{a_0} \right)^3 \left[ 1 - (a/a_w)^{3r} \right] \]
\[ \left[ 1 - \frac{3}{2}(a_0/a_w)^{3r} \right]; \] (26)
This is the EOS for DE, obtained in this model, which is scale factor dependent. \( a_s \), given above, and \( a_w \) are related as \( a_s = 2^{1/3r}a_w > a_0 \) and \( a_s < a_w < a_0 \).

Now, (16) and (26) yield
\[ \rho_{de} + 3\rho_{de} \]
\[ = -2\rho_{de(s)} + \left( \rho_{de(s)} - \rho_{de}^0 \right) \left( \frac{a}{a_0} \right)^3 \left[ 1 - 2(a/a_w)^{3r} \right] \]
\[ \left[ 1 - \frac{3}{2}(a_0/a_w)^{3r} \right]. \] (27)

Due to inequality (23), (26) and (27) yield
\[ \rho_{de} + \rho_{de} > 0, \] (28a)
\[ \rho_{de} + 3\rho_{de} < 0, \] (28b)

for \( a < a_w \) and
\[ \rho_{de} + \rho_{de} > 0, \] (29a)
\[ \rho_{de} + 3\rho_{de} < 0, \] (29b)

for \( a > a_w \).

(28a), (28b) give quintessence phase of DE, whereas (29a), (29b) give phantom phase of DE. These two phases are divided by
\[ \rho_{de} = -\rho_{de} \] (30)
at \( a = a_w \) as given by (27). Thus, we have transition from non-phantom to phantom at \( a = a_w > a_0 \).

It is given above that at \( a_s^{3r} = 2\lambda = 2a_w^{3r} \). So, (18) is obtained as
\[ \left( \frac{\dot{a}}{a} \right)^2 = \left( \rho_{de(s)} - \rho_{de}^0 \right) \left( \frac{a}{a_0} \right)^3 \left[ 1 - 2(a/a_s)^{3r} \right] \]
\[ \left[ 1 - (a_0/a_s)^{3r} \right], \] (31)
(31) shows that when \( a(t) \) reaches \( a_s \), acceleration given by (20) stops and deceleration driven by matter resumes. It shows transient acceleration as obtained in [8]. Sahni [14] had obtained this type of result in the context of brane-gravity cosmology earlier.

(31) also shows that \( \dot{a} = 0 \) at \( a = a_m > a_s \) if
\[ \left( \rho_{de(s)} - \rho_{de}^0 \right) \left( \frac{a}{a_0} \right)^3 \left[ 1 - 2(a_m/a_s)^{3r} \right] \]
\[ \left[ 1 - (a_0/a_s)^{3r} \right] \]
\[ + 0.23H_0^2 \left( \frac{a_0}{a_m} \right)^3 = 0. \] (32)
It shows that expansion will reach its maximum at \( a = a_m \) and it will begin to contract taking a turn around.

Thus, a transition from deceleration to acceleration, at some time \( t_s \) in the recent past, is obtained giving a possible explanation for late cosmic acceleration [2]. EOS for DE, derived here, depends on the scale factor \( a(t) \). Interestingly, it is found that another transition, from quintessence phase of DE to phantom phase, takes place at \( a = a_w \) with \( a_s < a_w < a_0 \). Characteristics of quintessence DE are different from phantom one. Kinetic energy for the latter is negative, whereas it is positive for the former. Moreover, \( w_{de} > -1 \) for the former, \( w_{de} < -1 \).
for the latter. So, it indicates that possibly, DE has two components (i) quintessence and (ii) phantom. Former dominates when $a_0 < a < a_w$ and latter dominates when $a_w < a < a_s$. It is found that, after acceleration for some time, universe decelerates. The decelerated expansion continues till scale factor acquires its maximum at time $t_m$ and will begin to contract for $t > t_m$.

References

First observation of the decay $\tau^- \to \phi K^- \nu_\tau$

Belle Collaboration

K. Inami $^{t,*}$, K. Abe $^{b}$, K. Abe $^{an}$, I. Adachi $^{g}$, H. Aihara $^{ao}$, D. Anipko $^{a}$, K. Arinstein $^{a}$, V. Aulchenko $^{a}$, T. Aushev $^{P,k}$, S. Bahinipati $^{c}$, A.M. Bakich $^{aj}$, E. Barberio $^{s}$, M. Barbero $^{f}$, I. Bedny $^{a}$, I. Bizjak $^{l}$, A. Bondar $^{a}$, A. Bozek $^{x}$, M. Bračko $^{g,r,l}$, T.E. Browder $^{f}$, A. Chen $^{v}$, K.-F. Chen $^{w}$, W.T. Chen $^{v}$, R. Chistov $^{k}$, Y. Choi $^{ai}$, Y.K. Choi $^{ai}$, J. Dalseno $^{s}$, M. Danilov $^{k}$, A. Drutskoy $^{c}$, S. Eidelman $^{a}$, D. Epifanov $^{a}$, S. Fratina $^{l}$, N. Gabysev $^{a}$, T. Gershon $^{g}$, G. Gokhroo $^{ak}$, H. Ha $^{a}$, J. Haba $^{g}$, K. Hara $^{a}$, H. Hayashii $^{u}$, M. Hazumi $^{a}$, D. Heffernan $^{ac}$, T. Hokuue $^{t}$, Y. Hoshi $^{am}$, S. Hou $^{v}$, W.-S. Hou $^{w}$, Y.B. Hsiung $^{w}$, T. Iijima $^{t}$, A. Ishikawa $^{ao}$, R. Itoh $^{g}$, Y. Iwasaki $^{g}$, J.H. Kang $^{at}$, P. Kapusta $^{x}$, S.U. Kataoka $^{u}$, H. Kawai $^{b}$, T. Kawasaki $^{z}$, H.R. Khan $^{aj}$, H. Kichimi $^{g}$, Y.J. Kim $^{d}$, S. Korpá $^{r,l}$, P. Križan $^{q,l}$, P. Krokovny $^{g}$, R. Kulasiri $^{c}$, R. Kumar $^{ad}$, A. Kuzmin $^{a}$, Y.-J. Kwon $^{at}$, G. Leder $^{l}$, T. Lesiak $^{x}$, S.-W. Lin $^{w}$, D. Liventsev $^{k}$, G. Majumder $^{ak}$, F. Mandl $^{j}$, T. Matsumoto $^{aq}$, A. Matyja $^{x}$, S. McOnie $^{aj}$, W. Mitaroff $^{f}$, H. Miyake $^{ac}$, H. Miyata $^{z}$, Y. Miyazaki $^{l}$, R. Mizuk $^{k}$, J. Mueller $^{af}$, Y. Nagasaka $^{h}$, I. Nakamura $^{g}$, E. Nakano $^{ab}$, M. Nakao $^{a}$, S. Nishida $^{g}$, O. Nitoh $^{ar}$, S. Ogawa $^{al}$, T. Ohshima $^{s}$, S. Okuno $^{m}$, Y. Onuki $^{ag}$, H. Ozaki $^{g}$, H. Palka $^{x}$, C.W. Park $^{ai}$, H. Park $^{o}$, L.S. Peak $^{aj}$, L.E. Pilonen $^{as}$, A. Poluektov $^{a}$, Y. Sakai $^{g}$, T. Schietinger $^{p}$, O. Schneider $^{p}$, A.J. Schwartz $^{c}$, R. Seidl $^{l,ag}$, K. Senyo $^{t}$, M.E. Sevior $^{s}$, H. Shibuya $^{al}$, B. Shwartz $^{a}$, V. Sidorov $^{a}$, J.B. Singh $^{ad}$, A. Somov $^{c}$, N. Soni $^{ad}$, S. Stanič $^{aa}$, M. Starič $^{l}$, H. Stoeck $^{aj}$, S.Y. Suzuki $^{g}$, O. Tajima $^{g}$, N. Tamura $^{z}$, M. Tanaka $^{g}$, G.N. Taylor $^{s}$, Y. Teramoto $^{ab}$, X.C. Tian $^{ae}$, T. Tsukamoto $^{s}$, S. Uehara $^{g}$, K. Ueno $^{w}$, T. Uglov $^{k}$, S. Uno $^{g}$, P. Urquijo $^{s}$, Y. Usov $^{a}$, G. Varner $^{f}$, S. Villa $^{p}$, E. Won $^{n}$, C.-H. Wu $^{v}$, B.D. Yabsley $^{aj}$, A. Yamaguchi $^{an}$, Y. Yamashita $^{y}$, L.M. Zhang $^{ab}$, V. Zhilich $^{a}$, A. Zupanc $^{l}$

$^a$ Budker Institute of Nuclear Physics, Novosibirsk, Russia  
$^b$ Chiba University, Chiba, Japan  
$^c$ University of Cincinnati, Cincinnati, OH, USA  
$^d$ The Graduate University for Advanced Studies, Hayama, Japan  
$^e$ Gyeongsang National University, Chinju, South Korea  
$^f$ University of Hawaii, Honolulu, HI, USA  
$^g$ High Energy Accelerator Research Organization (KEK), Tsukuba, Japan  
$^h$ Hiroshima Institute of Technology, Hiroshima, Japan  
$^i$ University of Illinois at Urbana-Champaign, Urbana, IL, USA  
$^j$ Institute of High Energy Physics, Vienna, Austria  
$^k$ Institute for Theoretical and Experimental Physics, Moscow, Russia  
$^l$ J. Stefan Institute, Ljubljana, Slovenia  
$^m$ Kanagawa University, Yokohama, Japan  
$^{n}$ Korea University, Seoul, South Korea  
$^o$ Kyungpook National University, Taegu, South Korea  
$^p$ Swiss Federal Institute of Technology of Lausanne, EPFL, Lausanne, Switzerland  
$^q$ University of Ljubljana, Slovenia  
$^r$ University of Maribor, Maribor, Slovenia  
$^s$ University of Melbourne, Victoria, Australia  
$^t$ Nagoya University, Nagoya, Japan  
$^u$ Nara Women's University, Nara, Japan  
$^v$ National Central University, Chung-li, Taiwan
Abstract

We present the first observation of τ lepton decays to hadronic final states with a φ-meson. This analysis is based on 401 fb$^{-1}$ of data accumulated at the Belle experiment. The branching fraction obtained is $\mathcal{B}(\tau^- \to \phi K^- \nu_\tau) = (4.05 \pm 0.25 \pm 0.26) \times 10^{-5}$.

© 2006 Elsevier B.V. All rights reserved.

PACS: 13.35.Dx; 14.40.Cs

Keywords: Tau; Phi

1. Introduction

Hadronic τ decays with a φ-meson in the final state are valuable to investigate QCD at a low mass scale. However, they have never been observed due to their small branching fractions. The decay $\tau^- \to \phi K^- \nu_\tau$ is Cabibbo-suppressed and further restricted by its small phase space, while the decay $\tau^- \to \phi \pi^- \nu_\tau$ is suppressed by the OZI rule although it is Cabibbo-allowed (Fig. 1). The branching fraction of the former can roughly be estimated by scaling the analogous Cabibbo-allowed decay $\tau^- \to K^+ K^- \nu_\tau$ [1] by $\tan^2 \beta_t$ and the ratio of the phase space of the two decays, resulting in $\mathcal{B}(\tau^- \to \phi K^- \nu_\tau) \sim 2 \times 10^{-5}$. Similarly, the vector dominance model predicts $\mathcal{B}(\tau^- \to \phi \pi^- \nu_\tau) = (1.20 \pm 0.48) \times 10^{-5}$ [2], whereas the CVC upper limit following from the cross section for $e^+ e^- \to \phi \pi^0$ is $\mathcal{B}(\tau^- \to \phi \pi^- \nu_\tau) < 3 \times 10^{-4}$ at the 90% confidence level [3].

Previously, the CLEO Collaboration searched for these decays using 3.1 fb$^{-1}$ of data taken on the $\Upsilon(4S)$ resonance. They set upper limits of $\mathcal{B}(\tau^- \to \phi K^- \nu_\tau) < (5.4-6.7) \times 10^{-5}$ and $\mathcal{B}(\tau^- \to \phi \pi^- \nu_\tau) < (1.2-2.0) \times 10^{-4}$ at the 90% confidence level, depending on the mechanism assumed for the decay [4]. Here we report the first measurement of the $\tau^- \to \phi K^- \nu_\tau$ decay. (Throughout this Letter charge-conjugate states are implied.) We also observe for the first time the decay $\tau^- \to \phi \pi^- \nu_\tau$, but it is treated here as a background process, together with the kinematically allowed but phase-space suppressed decays $\tau^- \to \phi \pi^- (n\pi) \nu_\tau$ ($1 \leq n \leq 4$). The result is based on a data sample of 401 fb$^{-1}$ corresponding to $3.6 \times 10^8 \tau^+ \tau^-$ pairs collected near the $\Upsilon(4S)$ resonance with the Belle de-
tector at the KEKB asymmetric-energy $e^+e^-$ (3.5 on 8 GeV) collider [5].

The Belle detector is a large-solid-angle magnetic spectrometer that consists of a silicon vertex detector, a 50-layer central drift chamber, an array of aerogel threshold Čerenkov counters, a barrel-like arrangement of time-of-flight scintillation counters, and an electromagnetic calorimeter comprised of CsI(Tl) crystals located inside a superconducting solenoid coil that provides a 1.5 T magnetic field. An iron flux-return located outside of the coil is instrumented to detect $K^0_L$ mesons and identify muons. The detector is described in detail elsewhere [6].

Two inner detector configurations were used. A 2.0 cm radius beampipe and a 3-layer silicon vertex detector were used for the first sample of 158 fb$^{-1}$, while a 1.5 cm radius beampipe, a 4-layer silicon detector and a small-cell inner drift chamber were used to record the remaining 244 fb$^{-1}$ [7].

2. Event selection

We look for $\tau^- \rightarrow \phi K^- \nu_\tau$ candidates in the reaction $e^+e^- \rightarrow \tau^+\tau^-$ with the following signature:

\[
\begin{align*}
\tau_{\text{signal}} & \rightarrow \phi + K^- + \text{(missing)} \\
\rightarrow & \ K^+ K^-, \\
\tau_{\text{tag}} & \rightarrow (\mu/e)^+ + n (\leq 1) \gamma + \text{(missing)},
\end{align*}
\]

where ‘missing’ denotes other possible daughters not reconstructed. The detection of $\phi$ mesons relies on the $\phi \rightarrow K^+ K^-$ decay ($B = (49.2 \pm 0.6\%$) [1]); the final evaluation of the signal yield is carried out using the $K^+ K^-$ invariant mass distribution.

The selection criteria described below are determined from studies of Monte Carlo (MC) simulated events. The background samples consist of $\tau^- \tau^+$ (1570 fb$^{-1}$, which does not include any decay mode with a $\phi$ meson) and $q\bar{q}$ continuum, $B^0 \bar{B}^0$, $B^+ B^-$ and two-photon processes. For signal, we generate samples with $2 \times 10^6 \tau^- \rightarrow \phi K^- \nu_\tau, \phi K^- \pi^0 \nu_\tau, \phi \pi^- \nu_\tau$ and $\phi \pi^- \pi^0 \nu_\tau$ events.

The transverse momentum for a charged track is required to be larger than 0.06 GeV/c in the barrel region ($-0.6235 < \cos \theta < 0.8332$, where $\theta$ is the polar angle relative to the direction opposite to that of the incident $e^+e^-$ beam in the laboratory frame) and 0.1 GeV/c in the endcap region ($-0.8660 < \cos \theta < -0.6235$, and $0.8332 < \cos \theta < 0.9563$). The energy of photon candidates is required to be larger than 0.1 GeV in both regions.

To select a $\tau$-pair sample, we require four charged tracks in an event with zero net charge, and a total energy of charged tracks and photons in the center-of-mass (CM) frame less than 11 GeV. We also require that the missing momentum in the laboratory frame be greater than 0.1 GeV/c, and that its direction be within the detector acceptance, where the missing momentum is defined as the difference between the momentum of the initial $e^+e^-$ system, and the sum of the observed momentum vectors. The event is subdivided into 3-prong and 1-prong hemispheres according to the thrust axis in the CM frame. These are referred to as the signal and tag side, respectively. We allow at most one photon on the tag side to account for initial state radiation, while requiring no extra photons on the signal side to reduce the $q\bar{q}$ backgrounds.

We require $\cos \theta_{\text{thrust-miss}}^\text{CM} < -0.6$ to reduce backgrounds from other $\tau$ decays and $q\bar{q}$ processes, where $\theta_{\text{thrust-miss}}^\text{CM}$ is the opening angle between the thrust axis (on the signal side) and the missing momentum in the CM frame. In order to remove the $q\bar{q}$ background, we require that the invariant mass of the particles on the tag side (if a $\gamma$ is present) be less than 1.8 GeV/$c^2$ ($\sim m_\tau$). Similarly, the effective mass of the signal side must be less than 1.8 GeV/$c^2$. Moreover, we require that the lepton likelihood ratio $P_{\ell/\gamma}$ be greater than 0.1 for the charged track on the tag side. Here $P_{\ell/\gamma}$ is the likelihood ratio for a charged particle of type $x$ ($x = \mu, e, K$ or $\pi$), defined as $P_x = L_x/(\sum_x L_x)$, where $L_x$ is the likelihood for particle type hypothesis $x$, determined from responses of the relevant detectors [8]. The efficiencies for muon and electron identification are 92% for momenta larger than 1.0 GeV/c and 94% for momenta larger than 0.5 GeV/c, respectively.

We require that both kaon daughters of the $\phi$ candidate have kaon likelihood ratios $P_\phi > 0.8$ and $\cos \theta > -0.6$. The kaon identification efficiency is 82%. To suppress combinatorial backgrounds from other $\tau$ decays and $q\bar{q}$ processes, we require that the $\phi$ momentum be greater than 1.5 GeV/c in the CM frame. After these requirements, the remaining contributions from $B^0 \bar{B}^0$, $B^+ B^-$, Bhabha, $\mu$ pair and two-photon backgrounds are negligible.

To separate $\phi K^- \nu_\tau$ from $\phi \pi^- \nu_\tau$, the remaining charged track is required to satisfy the same kaon identification criteria as the $\phi$ daughters. The $\tau^+ \tau^-$ and $q\bar{q}$ backgrounds are reduced by requiring that the opening angle ($\theta_{\phi K}$) be between the $\phi$ and $K^-$ in the CM frame satisfy $\cos \theta_{\phi K}^\text{CM} > 0.92$, and that the CM momentum of the $\phi K^-$ system be greater than 3.5 GeV/c. For $\phi \pi^- \nu_\tau$, we require that the charged track be identified as a pion, $P_\pi > 0.8$, and that the opening angle between the $\phi$ and $\pi^-$ in the CM frame satisfy $\cos \theta_{\phi \pi}^\text{CM} < 0.98$. This last requirement suppresses the background from $\tau^- \rightarrow \phi K^- \nu_\tau$ and $\tau^- \rightarrow \phi K^- \pi^0 \nu_\tau$.

Fig. 2(a) shows the $K^+ K^-$ invariant mass distribution after all $\tau^- \rightarrow \phi K^- \nu_\tau$ selection requirements. As there are two possible $K^+ K^-$ combinations from the $K^- K^+$ tracks on the signal side, this distribution has two entries per event. Therefore, the signal MC shape includes a long tail due to the wrong $K^+ K^-$ combination. Non-resonant backgrounds arise mainly from $\tau^- \rightarrow K^+ K^- \pi^- \nu_\tau$, which has a branching fraction of $B = (1.53 \pm 0.10) \times 10^{-3}$ [1]. Small contributions are expected from $q\bar{q}$ processes as described below.

3. Signal and background evaluation

The detection efficiencies $\epsilon$ for $\tau^- \rightarrow \phi K^- \nu_\tau$ and the cross-feed rates from $K^- \pi^0 \nu_\tau$ and $\phi \pi^- \nu_\tau$ are evaluated, as listed in Table 1, from MC simulation using KKMC [9], where the $V - A$ interaction is assumed at the vertices and the final hadrons decay according to non-resonant phase space. The efficiencies include the branching fraction for $\phi \rightarrow K^+ K^-$. 
The signal yields are extracted by a fit to the $K^+K^-$ invariant mass distribution. For signal, we use a $p$-wave Breit–Wigner (BW) distribution convolved with a Gaussian function (of width $\sigma$) to account for the detector resolution. The $\phi$ width is fixed to be $\Gamma_\phi = 4.26$ MeV$/c^2$ [1] but $\sigma$ is allowed to float. First- and second-order polynomial background functions are used for $\tau^- \rightarrow \phi K^-\nu_\tau$ and $\phi\pi^-\nu_\tau$ decays, respectively. The fit results are also shown in Fig. 2. The obtained signal yields are $N_{\phi K^+} = 573 \pm 32$ and $N_{\phi\pi^+} = 753 \pm 84$. The $\sigma$'s from the fits are $1.2 \pm 0.3$ MeV$/c^2$ and $1.2 \pm 0.7$ MeV$/c^2$ for $\phi K^-\nu_\tau$ and $\phi\pi^-\nu_\tau$, respectively, which are consistent with MC simulation.

MC studies show that only the $\tau^- \rightarrow \phi\pi^-\nu_\tau$, $\tau^- \rightarrow \phi K^-\pi^0\nu_\tau$, and $q\bar{q}$ samples yield significant contributions peaking at the $\phi$ mass. The contributions of other backgrounds are less than 0.01% and can be neglected. The contribution of $\tau^- \rightarrow \phi\pi^-\nu_\tau$ events to the $\phi K^-\nu_\tau$ sample is estimated using $N_{\phi\pi^-\nu_\tau}$ and the misidentification rate, as discussed below. Other contributions are estimated as follows.

To evaluate the branching fraction and background contribution from $\tau^- \rightarrow \phi K^-\pi^0\nu_\tau$, we select $\pi^0 \rightarrow \gamma\gamma$ candidates and combine them with $\phi K^-\nu_\tau$ combinations that satisfy the requirements listed above. The signal yield is estimated by fitting the resulting $K^+K^-$ invariant mass distribution with a $p$-wave BW distribution plus a linear background function, as shown in Fig. 3. The resulting yield is $8.2 \pm 3.8 \phi K^-\pi^0\nu_\tau$ events. Using a detection efficiency $\epsilon_{\phi K^+\pi^0\nu_\tau} = (0.396 \pm 0.007\%)$ obtained from MC simulation, and an $e^-\bar{e}^+ \rightarrow \tau^+\tau^-$ sample normalization $N_{\tau^+\tau^-} = 401 \text{ fb}^{-1} \times 0.892 \text{ nb} = 3.58 \times 10^8$, we obtain a branching fraction $B(\tau^- \rightarrow \phi K^-\pi^0\nu_\tau) = (2.9 \pm 1.3) \times 10^{-6}$. However, this must be corrected for the unknown contamination of $\tau^- \rightarrow \phi\pi^-\pi^0(\pi\pi)\nu_\tau$ $(0 \leq n \leq 3)$ decays. Using this value, we estimate the $\tau^- \rightarrow \phi K^-\pi^0\nu_\tau$ background in the $\tau^- \rightarrow \phi K^-\nu_\tau$ sample to be $N_{\phi K^-\pi^0\nu_\tau} = (6.8 \pm 3.1)$ events, given a cross-feed rate for $\tau^- \rightarrow \phi K^-\pi^0\nu_\tau$ to the $\tau^- \rightarrow \phi K^-\nu_\tau$ sample of $(0.328 \pm 0.006\%)$ (see Table 1).

From a MC study, we find a $q\bar{q}$ contamination of $N_{q\bar{q}} = 6.6 \pm 2.5$. To take into account the uncertainty in $\phi$ production in the $q\bar{q}$ MC, we compare MC results with enriched $q\bar{q}$ data by demanding that the effective mass of the tag side be larger than $1.8$ GeV$/c^2$. With this selection, the background is $q\bar{q}$ dominated and the other backgrounds are negligible. The yield in data is $262 \pm 21$ events, and the yield in the $q\bar{q}$ MC is $117 \pm 10$ events. We subsequently scale the above $q\bar{q}$ background estimate by the ratio $f = 2.23 \pm 0.26$; the result is $N_{q\bar{q}} = 14.8 \pm 5.8$ events.

Table 1

Detection efficiencies $\epsilon$ and cross-feed rates (%) from MC simulation. The errors are from the MC statistics.

<table>
<thead>
<tr>
<th>Candidates</th>
<th>Decay modes</th>
<th>$\phi K^\nu$</th>
<th>$\phi\pi^\nu$</th>
<th>$\phi K^-\pi^0\nu_\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau^- \rightarrow \phi K^\nu$</td>
<td>1.826 ± 0.009</td>
<td>0.049 ± 0.002</td>
<td>0.328 ± 0.006</td>
<td></td>
</tr>
<tr>
<td>$\tau^- \rightarrow \phi\pi^-\nu_\tau$</td>
<td>0.110 ± 0.002</td>
<td>1.663 ± 0.014</td>
<td>0.009 ± 0.001</td>
<td></td>
</tr>
</tbody>
</table>

4. Results

The peaking backgrounds described above, $\tau^- \rightarrow \phi K^-\pi^0\nu_\tau$ and $q\bar{q}$, are subtracted from the signal yield, leaving $N_{\phi K^-\nu_\tau} = (573 \pm 32) - (6.8 \pm 3.1) - (14.8 \pm 5.8) = 551 \pm 33$ events.

To take into account cross-feed between $\tau^- \rightarrow \phi K^-\nu_\tau$ and $\tau^- \rightarrow \phi\pi^-\nu_\tau$ due to particle misidentification ($K \leftrightarrow \pi$), we solve the following simultaneous equations:

$$\begin{align*}
N_{\phi K^+} &= 2N_{\tau^+} (\epsilon_{\phi K^+} B_{\phi K^+} + \epsilon_{\phi\pi^-} B_{\phi\pi^-}) , \\
N_{\phi\pi^-} &= 2N_{\tau^+} (\epsilon_{\phi\pi^-} B_{\phi K^+} + \epsilon_{\phi\pi^-} B_{\phi\pi^-}) .
\end{align*}$$

Fig. 2. $K^+K^-$ invariant mass distributions for (a) $\tau^- \rightarrow \phi K^-\nu_\tau$ and (b) $\tau^- \rightarrow \phi\pi^-\nu_\tau$. Points with error bars indicate the data. The shaded histograms show the expectations from $\tau^-\tau^-$ and $q\bar{q}$ background MC simulations. The open histogram is the signal MC with $B(\tau^- \rightarrow \phi K^-\nu_\tau) = 4 \times 10^{-5}$ in (a) and $B(\tau^- \rightarrow \phi\pi^-\nu_\tau) = 6 \times 10^{-5}$ in (b). The curves show the best fit results, and the dashed curves indicate the non-resonant background contributions. See the text for details.

Fig. 3. $K^+K^-$ invariant mass distributions for $\tau^- \rightarrow \phi K^-\pi^0\nu_\tau$. Points with error bars indicate the data. Histograms show the MC expectations of $\tau$-pairs (shaded) and signal (open) with a branching fraction of $3 \times 10^{-6}$. The solid curve shows the best fit result and the dashed curve shows the non-resonant background contribution.
where $B_{\phi K}^{}$ and $B_{\phi \pi}^{}$ are the branching fractions for $\tau^- \rightarrow \phi K^-\nu_\tau$ and $\tau^- \rightarrow \phi \pi^-\nu_\tau$, respectively. The detection efficiencies, $\epsilon$'s, are listed in Table 1. The factor $\epsilon_{\phi K}^{}$ is the efficiency for reconstructing $\tau^- \rightarrow \phi K^-\nu_\tau$ as $\tau^- \rightarrow \phi \pi^-\nu_\tau$ while $\epsilon_{\phi \pi}^{}$ is the efficiency for reconstructing $\tau^- \rightarrow \phi \pi^-\nu_\tau$ as $\tau^- \rightarrow \phi K^-\nu_\tau$. The resulting branching fraction for $\tau^- \rightarrow \phi K^-\nu_\tau$ is

$$B_{\phi K}^{} = (4.05 \pm 0.25) \times 10^{-5},$$

(3)

where the uncertainty is due to the statistical uncertainty in the $N_{\phi K}^{}$ and $N_{\phi \pi}^{}$ terms. The uncertainty in the detection efficiencies, $\epsilon$'s, will be taken into account in the systematic error. The result for $B_{\phi \pi}^{}$ is $B_{\phi \pi}^{} = (6.05 \pm 0.71) \times 10^{-5}$; however, small background from $\tau^- \rightarrow \phi \pi^- (n\pi)\nu_\tau$ ($1 \leq n \leq 4$) decays is included and must be subtracted to obtain the final branching fraction.

The systematic uncertainties are estimated as follows: the uncertainties in the integrated luminosity, $\tau^3\tau^-$ cross-section and trigger efficiency are 1.4%, 1.3% and 1.1%, respectively. Track finding efficiency has an uncertainty of 4.0%. Uncertainties in lepton and kaon identification efficiencies and fake rate are evaluated, respectively, to be 3.2% and 3.1% by averaging the estimated uncertainties depending on momentum and polar angle of each charged track. To evaluate the systematic uncertainty of fixing $\Gamma_\phi$ in the BW fit, we calculate the change in the signal yield when $\Gamma_\phi$ is varied by $\pm 0.05$ MeV/c$^2$ (the uncertainty quoted by the PDG) [1]: the result is 0.2%. The branching fraction for $\phi \rightarrow K^+K^-$ gives an uncertainty of 1.2% [1]. The signal detection efficiency $\epsilon_{\phi K}^{}$ has an uncertainty of 0.5% due to MC statistics. A total systematic uncertainty of 6.5% is obtained by adding all uncertainties in quadrature. The resulting branching fraction is then

$$B(\tau^- \rightarrow \phi K^-\nu_\tau) = (4.05 \pm 0.25 \pm 0.26) \times 10^{-5}.$$  

(4)

Finally, we consider the possibility that a resonant state contributes to the final $\phi K^-$ hadronic system. We generate a resonant MC with the KKMC simulation program. The weak current is generated with a $V-A$ form while the $\phi K^-$ system is assumed to be produced from a 2-body decay of a resonance. In Fig. 4(a), the $\phi K^-$ mass distribution for data is compared to MC; the combinatorial background is subtracted using the $K^+K^-$ sideband. The MC distributions correspond to $(M, \Gamma) = (1650, 100)$ MeV/c$^2$, $(M, \Gamma) = (1570, 150)$ MeV/c$^2$, and also non-resonant phase space. Fig. 4(b) shows the $\phi$'s angular distribution in the $\phi K^-$ rest frame ($\cos \alpha$), where the negative of the lab frame direction in the $\phi K^-$ frame is taken as the reference axis. It indicates an isotropic distribution in the $\phi K^-$ system. For both the invariant mass and angular distributions of the $\phi K^-$ system, the phase space MC reproduces the signal distribution well. We therefore neglect systematic uncertainty due to possible resonant structure. On the other hand, the 1650 MeV/c$^2$ state assumed in the CLEO search [4], indicated by the dotted histogram in Fig. 4(a), clearly cannot account for the entire signal. If production via a single resonance is assumed, the best agreement with data is found for a mass and a width of $\simeq 1570$ MeV/c$^2$ and $\simeq 150$ MeV/c$^2$, respectively, as shown by the dot-dashed histogram. However, since the shape of the resonant MC is similar to the phase-space-distributed MC, we cannot draw any strong conclusions about an intermediate resonance with $\Gamma \sim O(100$ MeV/c$^2$) in this narrow mass range of $\sim 250$ MeV/c$^2$.

5. Conclusion

Using 401 fb$^{-1}$ of data, we make the first observation of the rare decay $\tau^- \rightarrow \phi K^-\nu_\tau$. The measured branching fraction is

$$B(\tau^- \rightarrow \phi K^-\nu_\tau) = (4.05 \pm 0.25 \pm 0.26) \times 10^{-5}.$$  

(5)

Acknowledgements

We gratefully acknowledge the essential contributions of Mari Kitayabu, which are described in her bachelor thesis at Nagoya University. We thank the KEKB group for the excellent operation of the accelerator, the KEK cryogenics group for the efficient operation of the solenoid, and the KEK computer group and the National Institute of Informatics for valuable computing and Super-SINET network support. We acknowledge support from the Ministry of Education, Culture, Sports, Science, and Technology of Japan and the Japan Society for the Promotion of Science; the Australian Research Council and the Australian Department of Education, Science and Training;

Fig. 4. (a) Invariant mass and (b) angular distributions for the $\phi K^-$ system. The non-$\phi$-resonant backgrounds are subtracted using the sideband spectra. Points with error bars indicate the data. The open histogram shows the phase-space-distributed signal MC, and the dotted and dot-dashed histograms indicate the signal MC mediated by a resonance with $M = 1650$ MeV/c$^2$ and $\Gamma = 100$ MeV/c$^2$ and $M = 1570$ MeV/c$^2$ and $\Gamma = 150$ MeV/c$^2$, respectively. In the MC, a branching fraction of $4 \times 10^{-5}$ is assumed. (b) $\phi$'s angular distribution in the $\phi K^-$ rest frame, where the negative of the lab frame direction in the $\phi K^-$ frame is taken as the reference axis.
the National Science Foundation of China and the Knowledge Innovation Program of the Chinese Academy of Sciences under contract No. 10575109 and IHEP-U-503; the Department of Science and Technology of India; the BK21 program of the Ministry of Education of Korea, the CHEP SRC program and Basic Research program (grant No. R01-2005-000-10089-0) of the Korea Science and Engineering Foundation, and the Pure Basic Research Group program of the Korea Research Foundation; the Polish State Committee for Scientific Research; the Ministry of Science and Technology of the Russian Federation; the Slovenian Research Agency; the Swiss National Science Foundation; the National Science Council and the Ministry of Education of Taiwan; and the US Department of Energy.

References

Energy density of the glasma

T. Lappi

Physics Department, Brookhaven National Laboratory, Upton, NY 11973, USA

Received 27 June 2006; received in revised form 3 October 2006; accepted 9 October 2006
Available online 23 October 2006
Editor: J.-P. Blaizot

Abstract

The initial energy density produced in an ultrarelativistic heavy ion collision can, in the color glass condensate framework, be factorized into a product of the integrated gluon distributions of the nuclei. Although this energy density is well defined without any infrared cutoff besides the saturation scale, it is apparently logarithmically ultraviolet divergent. We argue that this divergence is not physically meaningful and does not affect the behavior of the system at any finite proper time.

© 2006 Elsevier B.V. All rights reserved.
PACS: 24.85.+p; 25.75.-q; 12.38.Mh

1. Introduction

The matter produced at central rapidities in a heavy ion collision is dominated by the small $x$ partons in the wave function of the high energy nuclei. These degrees of freedom can, because of their high occupation numbers, be described as a classical Weizsäcker–Williams color field. The source for this field is formed by the large $x$ partons, which are seen by the small $x$ ones as classical color charges. The nonlinear interactions between the small $x$ gluons give rise to gluon saturation, and the wavefunction is described by an energy (or $x$) dependent saturation scale. This way of understanding the small $x$ wavefunction is known as the color glass condensate. A model incorporating these physical features was written down by McLerran and Venugopalan (MV) [1–3].

The initial transverse energy and gluon multiplicity in a collision of two sheets of color glass in the MV model has been calculated to all orders in the gluon field already some time ago [4–9]. Recently there has been a renewed interest in the very early time behavior of these classical "glasma" gluon fields [10, 11], in the context of pair production from the classical background field [12–19] and parity violation through the Chern–Simons charge density of the fields [20,21]. More attention has also been paid to the 3-dimensional energy density (instead of the energy per unit rapidity) of these field configurations as a quantity that could be directly related to the initial conditions of hydrodynamical calculations [22,23].

The purpose of this Letter is to clarify some properties of the initial energy density of the gauge fields in the MV model. We shall first, in Section 2, demonstrate that the initial energy density can be completely factorized into the product of the gluon distribution functions of the colliding nuclei. Going to finite proper times, or looking at the multiplicity, will change this factorization into a convolution of the unintegrated gluon distributions. We shall then, in Section 3, go on to discuss the known properties of the correlator of two pure gauge fields involved in the initial energy density and, in Section 4, try to understand the behavior of large $p_T$ modes with the help of the lowest order perturbative solution of the classical field equations.

2. Initial condition

Because of their high speed and Lorentz time dilation, the large $x$ degrees of freedom are seen by the low $x$ fields as slowly evolving in light cone time. They form classical, static (in light cone time) sources on the light cones:

$$J^\mu = \delta^\mu + \rho(1)(x_T, x^-) + \delta^\mu - \rho(2)(x_T, x^+).$$

(1)
where the support of the sources around the light cone must be understood as being very close to a delta function: \( \rho_{(1,2)}(x_T, x^\pm) \sim \delta(x^\pm) \). We shall work here in the Schwinger gauge \( A_\tau = (x^+ A^- + x^- A^+)/\tau = 0 \), in which the current (1) is not rotated by the soft classical field; in more general gauges Eq. (1) should be dressed by Wilson lines to maintain its covariant conservation. The Weizsäcker–Williams fields describing the lighter degrees of freedom can then be computed from the classical Yang–Mills equation

\[
[D_\mu, F^{\mu\nu}] = j^\nu.
\]

In the light cone gauge the field of one nucleus is a pure gauge outside the light cone (see Fig. 1)

\[
A_{(1,2)}^i(x_T) = i g U_{(1,2)}(x_T) \partial_i U_{(1,2)}^\dagger(x_T),
\]

where the SU(3) matrices \( U_{(n)}(x_T) \) are determined from the color sources as

\[
U_{(1,2)}(x_T) = P \exp \left\{ -ig \int dx^\pm \frac{1}{\Delta_T} \rho_{(1,2)}(x_T, x^\pm) \right\}.
\]

In the original MV model, which we shall be using here, the color charge densities are stochastic Gaussian random variables on the transverse plane: \( \rho_{(1)}^a(x_T, x^-) = \delta(x^-) \rho_{(1)}^a(x_T) \) with

\[
\langle \rho^a(x_T) \rho^b(x_T) \rangle = g^2 \mu^2 \delta^{ab} \delta^2(x_T - y_T),
\]

where the density of color charges \( g^2 \mu \) is, up to a numerical constant and a logarithmic uncertainty, related to the saturation scale \( Q_s \).

The initial conditions for the fields in the future light cone between the two colliding sheets were derived and the equations of motion solved to lowest order in the fields in [24–26] (see also Ref. [27] for the same calculation in covariant gauge and Ref. [11] for another formulation). This initial condition has a simple expression in terms of the pure gauge fields of the two colliding nuclei (3):

\[
A^i = A^i_{(1)} + A^i_{(2)},
\]

where \( A^i = A^i_{(1)} + A^i_{(2)} \). Note that the metric in the \((\tau, \eta, x_T)\) coordinate system is \( g_{\mu\nu} = \text{diag}(1, -\tau^2, -1, -1) \) so that \( A_\eta = -\tau^2 A^\eta \). In the Schwinger gauge \( A^i = 0 \) the \( \pm \) components of the gauge field are related by \( A^\pm = \pm x^\pm A^0 \). Because of the explicit time dependence in the metric \( A^0 \) corresponds, at \( \tau = 0 \), to the \( z \)-component of the chromoelectric field. At the initial time the only nonzero components of the field strength tensor are the longitudinal electric and magnetic fields and consequently the energy density is given by

\[
\varepsilon(\tau = 0) = \lim_{\tau \to 0^+} \frac{1}{\tau} \frac{dE}{d^2x_T d\eta} = \frac{1}{2} \text{Tr} F_{ij} F_{ij} + 4 \text{Tr}(A^0)^2.
\]

Let us introduce a shorter notation for the correlation function of the pure gauge field of the nucleus when averaged with the distribution (5). We shall define the correlation function \( G(p_T) \) by

\[
\langle A_i^{(m)a} (p_T) A_j^{(n)b} (q_T) \rangle = (2\pi)^3 \delta^{mn} \delta^{ab} \langle p_T^2 + q_T^2 \rangle \frac{p_i p_j}{p_T^2} G(p_T).
\]

The index in parentheses \((m)\) refers to the two colliding nuclei, which are, naturally, independent of each other, thus the \( \delta^{mn} \) in the correlator. The correlator must also be diagonal in the color index \((\delta^{ab})\) because there is no preferred direction in color space present in the problem. We are assuming translational invariance on the transverse plane \((\delta^2(p_T + q_T))\), which is justified because we are only interested in momentum scales much larger than the nuclear geometry effects which break this invariance (meaning that we are assuming \(|p_T| \gg 1/R_A\)). The transverse spatial index structure \( p_i p_j \) is the only one consistent with rotational invariance on the transverse plane (again, at momentum scales \(|p_T| \gg 1/R_A\) there is no preferred direction in the system to break this invariance).

Using the notation of Eq. (11) the two terms in the energy density (10) become

\[
\int d^2x_T \frac{1}{2} \text{Tr} F_{ij} F_{ij} = \frac{1}{2} g^2 N_c (N_c^2 - 1) \pi R_A^2
\]

\[
\times \int \frac{d^2k_T d^2p_T}{(2\pi)^2 (2\pi)^2} \left[ \frac{p_T^2 k_T^2}{(p_T \cdot k_T)^2} - (p_T \cdot k_T)^2 \right] \frac{G(p_T) G(k_T)}{p_T^2 k_T^2},
\]

and the final result factorizes into

\[
\varepsilon_{T=0} = \frac{g^2}{2} N_c (N_c^2 - 1) \left[ \int \frac{d^2p_T}{(2\pi)^2} G(p_T) \right]^2.
\]
Eq. (14) is our main result. The initial energy density factorizes completely into a product of two terms, both of which only depend on the properties of one single nucleus. This happens only strictly at \( \tau = 0 \).

Note that due to rotational invariance the initial energy density in the magnetic field, Eq. (12) and the electric field, Eq. (13) are equal. The discretized version of the computation on a transverse lattice breaks this rotational invariance (for an explicit expression see the lattice perturbation theory result in Appendix B of Ref. [4]). This violation is largest for the momentum modes near the edges of the Brillouin zone. As we shall discuss in the following, for larger proper times, these modes do not affect the energy density any more, and the energy densities in the longitudinal electric and magnetic fields approach each other, as can be seen in Fig. 2.

3. Properties of the gauge field correlator

Let us then recall some known properties of \( G(p_T) \), the correlation function of the pure gauge fields defined in Eq. (11). In light cone quantization it is related to the unintegrated gluon distribution function

\[
xG(x, Q^2) = R_A^2 (N_c^2 - 1) \int \frac{d^2k_T}{(2\pi)^2} G(k_T).
\]

Note, however, that our \( G(p_T) \) is equivalent to the unintegrated gluon distribution used to compute gluon production in \( pA \)-collisions only in the weak field limit. We refer the reader to Refs. [29–31] for a discussion of the difference. Our \( G(p_T) \) is equivalent, up to the normalization, to \( \phi^{WW} \) of Ref. [29].

The correlator \( G(p_T) \) has been analytically evaluated in several papers [32–34]. The result is expressed in closed form in coordinate space as

\[
G(x_T) = \frac{4}{g^2 N_c x_T^2} \left( 1 - e^{-\frac{\mu}{2g}} \right)^2 \ln \frac{1}{|x_T|^2}.
\]

Here \( \Lambda \) is an infrared cutoff that one must introduce in order to invert the 2-dimensional Laplace operator. The same function can also be measured in the numerical setup used to compute the glasma fields [4–8]. The numerical procedure used is not exactly equivalent to the calculation leading to Eq. (16), because the source in Eq. (4) is taken as exactly a delta function on the light cone and the infrared divergence in inverting the Laplace operator \( \Delta \) is effectively regulated by the size of the lattice. These differences can, however, be absorbed into the infrared cutoff \( \Lambda \), and the numerical evaluation (see in particular Fig. 3 of Ref. [35]) of the correlator agrees with the behavior of Eq. (16). Note that to derive the correct initial conditions, Eqs. (3), (4), (6) and (7), it is essential to consider the source as spread out in the longitudinal coordinate. Only when this is done can one, in practice, take the source as a delta function on the light cone when evaluating the Wilson line, Eq. 4.

Let us then estimate the behavior of the correlator \( G(p_T) \) in momentum space. For small momenta \( G(p_T) \) diverges, but the divergence is only logarithmic and thus integrable. This is the essential feature of gluon saturation; bulk quantities that are sensitive to the harder modes in the spectrum, such as the energy density, are infrared finite when the nonlinear interactions are taken into account fully.

For large momenta \( G(p_T) \) has a perturbative tail behaving as \( (p_T^2)^{-\alpha} \), meaning that the integral \( \int d^2p_T \, G(p_T) \) and thus the initial energy density are seemingly ultraviolet divergent. This can be seen equivalently as the logarithmic ultraviolet divergence in Eq. (16),

\[
\int \frac{d^2p_T}{(2\pi)^2} G(p_T) = \lim_{|x_T| \to 0} G(x_T) = \frac{1}{2\pi g^2 (g^2 \mu)^2} \lim_{|x_T| \to 0} \ln \frac{1}{|x_T|^2}.
\]

We must emphasize that although Eq. (17) involves, for dimensional reasons, the infrared cutoff \( \Lambda \), it corresponds to a divergence from large transverse momentum, or small distance, modes.

The initial energy density of the glasma is infrared finite, but seemingly logarithmically ultraviolet divergent. There are two reasons why this divergence is fundamentally not a problem for the physical picture of the glasma. The first reason is that, as can be seen from Eq. (15), the divergence corresponds to large \( Q^2 \) and thus, for a fixed energy, large \( x \) modes in the wavefunction. These are degrees of freedom that were, by our initial assumptions, not meant to be included in the classical field in the first place. It would therefore be physically well motivated to regulate them with an ultraviolet cutoff \( A_{UV} \gtrsim Q_s \), and then match this cutoff with whatever way one treats these hard collisions. This is indeed the approach advocated e.g. in Ref. [11]. The other reason for not worrying about the ultraviolet divergence is that, as we shall argue in the following, the energy density of the system at later proper times \( \tau \sim 1/Q_s \) has a finite \( A_{UV} \to \infty \) limit. Thus if one regulates the ultraviolet divergence in any convenient way and proceeds to solve the equations forward in time, the cutoff no longer significantly influences the later time evolution of the glasma fields.

1 The notation and numerical constants at this point are very confusing. Here an attempt is made to follow Ref. [28] (in particular Section 2.4) and Ref. [11].
4. Perturbative comparison and the apparent UV divergence

In the perturbative (lowest order in the source charge densities) solution [24–27] the field amplitudes behave, in the two-dimensional Coulomb gauge, like Bessel functions

\[ A_i(\tau, k_T) = A_i(\tau = 0, k_T) J_0(|k_T| \tau), \]

\[ A_\eta(\tau, k_T) = -\frac{2\tau}{|k_T|} A^\eta(\tau = 0, k_T) J_1(|k_T| \tau). \]

The energy density corresponding to this perturbative solution (this time dependence is also derived in Refs. [36,37]) is

\[ \varepsilon(\tau \to \infty) = \frac{1}{2} g^2 N_c (N_c^2 - 1) \]

\[ \int \frac{d^2 k_T}{(2\pi)^2} \frac{d^2 p_T}{(2\pi)^2} G(p_T) G(k_T) \times \left[ J_0^2(|p_T + k_T| \tau) + J_1^2(|p_T + k_T| \tau) \right]. \]

To lowest order in the sources the pure gauge field correlator is

\[ G(p_T) = \frac{1}{g^2} \frac{(g^2 \mu)^2}{p_T^2}. \]

and, using the asymptotic behavior of the Bessel functions in Eq. (20), the energy density for large times reduces to

\[ \varepsilon(\tau \to \infty) = \frac{1}{\tau} \int d^2 k_T \frac{dN}{dyd^2 k_T} |k_T| \]

with the Bertsch–Gunion [38] type multiplicity resulting from the lowest order solution [24–27]:

\[ \frac{dN}{dyd^2 k_T} = \frac{\pi R_A^2 N_c (N_c^2 - 1) g^2 \mu^4}{\pi} \frac{1}{R_A^2} \times \int \frac{d^2 p_T}{(2\pi)^2} \frac{1}{p_T^2 (k_T - p_T)^2}. \]

Also in the full nonperturbative solution the most ultraviolet modes behave in the same way [7,8,39]. Let us now assume that we have computed the initial energy density with some ultraviolet cutoff \( \Lambda_{UV} \) (such as the inverse lattice spacing in the numerical calculation). The initial energy density Eq. (14) depends on this cutoff as \( \ln^2 \Lambda_{UV} \), as can be seen in the numerical result in Fig. 3. After a time \( \tau \gtrsim 1/\Lambda_{UV} \), however, the time dependence of the modes near the cutoff changes from the initial \( A_1(\tau) = A_i(0)(1 + \mathcal{O}(\tau^2)) \) to the asymptotic regime \( A_1 \sim 1/|k_T| \tau \) and the contribution of these modes to the total energy density is suppressed by an additional power of the momentum, making the energy density finite in the limit \( \Lambda_{UV} \to \infty \). After a time \( \tau \gtrsim 1/\Lambda_{UV} \), or \( \tau \gtrsim a \) in the lattice computation, the energy density then converges to a value that is independent of the lattice spacing, as shown in Fig. 4. The larger the proper time that one is looking at, the better the convergence in the continuum limit. Fig. 5 shows the continuum extrapolations of the energy densities at \( g^2 \mu \tau = 0.2, 0.5 \) and 1.0.

By turning this argument around one can see that if one takes first the continuum limit (\( \Lambda_{UV} \to \infty \) or \( a \to 0 \)) for a finite
of the continuum extrapolated energy density will behave as \( \ln^2 1/\tau \). Thus if the limits are taken in this order, the energy density is indeed finite for all \( \tau > 0 \), but diverges logarithmically at \( \tau = 0 \). Note that this divergence is so weak that the energy per unit rapidity \( (\tau \varepsilon) \) is still zero at \( \tau = 0 \). Thus we see that the solution of the field equations is well defined as an initial value problem only in the presence of an ultraviolet cutoff in the transverse momenta, but for later times the system loses memory of this cutoff. Incidentally, because of this feature one can argue that introducing of a finite initial time (such as done in Refs. [40,41] to avoid the singularity resulting from broken boost invariance of the field configurations) does not really add another physical parameter into the model.

5. Energy density in physical units and conclusion

For concreteness let us finally try to express the results shown in Figs. 3, 4 and 5 in physical units. Due to the difficulty in fixing exactly the right value of the color charge density parameter \( g^2 \mu \) this is not necessarily straightforward. For RHIC energies one can argue, based on both the gluon multiplicity found in the numerical calculation [6–8] and counting the number of large x degrees of sources in the wavefunction from conventional parton distribution functions [26] that the relevant value would be \( g^2 \mu = 2 \text{ GeV} \). Other estimates, e.g. [15], give a smaller value, so \( g^2 \mu \approx 2 \text{ GeV} \) should be considered an upper bound for RHIC. For LHC energies the estimate based on parton distributions [26] gives \( g^2 \mu \approx 4 \text{ GeV} \), but this is most certainly an overestimate, since the calculation in Ref. [26] is based on parton distributions in the proton and does not take into account shadowing corrections. Another way of estimating the color charge density is based on the small x scaling properties observed in deep inelastic scattering data [42–46]. The saturation scale in this scaling can then be related to the MV model color charge density [47,48]. This line of thought leads to a scaling \( Q_s^2 \sim (g^2 \mu)^2 \sim \sqrt{\lambda} \), where \( \lambda = \lambda/(1+\lambda/2) \) and a fit to the HERA data [42] gives \( \lambda = 0.288 \) and thus \( \lambda \approx 0.25 \). The result for the color charge density at the LHC would be \( g^2 \mu \approx 3 \text{ GeV} \), which is the value we will use in the following.

As we have seen, the energy density strictly at \( \tau = 0 \) is not the best quantity to look at. Let us instead estimate the energy density at the time \( \tau = 1/g^2 \mu \). This is when, as can be seen from Fig. 4, the \( 1/\tau \)-decrease of the energy density seems to start. The simple linear continuum extrapolation of Fig. 5 yields \( \varepsilon(\tau = 1/g^2 \mu) = 0.26(g^2 \mu)^4/g^2 \). Using \( g = 2 \) we then get the estimates \( \varepsilon(\tau = 0.1 \text{ fm}) \approx 130 \text{ GeV}/\text{fm}^3 \) for RHIC and \( \varepsilon(\tau = 0.07 \text{ fm}) \approx 700 \text{ GeV}/\text{fm}^3 \) for the LHC. This estimate agrees with the values given in Ref. [39] for \( \tau = 3/g^2 \mu \) when the ~1/\tau behavior of the energy density is taken into account. The uncertainty due to the unprecise value of \( g^2 \mu \) in these numbers is quite large because of the power law dependence \( \varepsilon \sim (g^2 \mu)^4 \).

In conclusion, we have shown that the initial 3-dimensional energy density of the glasma fields in the MV model can be expressed as a product of the (integrated) gluon distribution functions of the colliding nuclei. Only the energy density at later times and the multiplicity involve convolutions of the unintegrated gluon distributions, probing the \( k_T \)-distributions in the wavefunctions of the colliding nuclei in more detail. We have recalled the known properties if the pure gauge field correlator Eq. (11) appearing in the initial energy density. As expected from general gluon saturation arguments, the initial energy density is infrared finite when the gluon fields are solved to all orders in the source. The energy density strictly at \( \tau = 0 \) is, however, ultraviolet divergent in the MV model. We show, both by a direct numerical calculation and by examining the time dependence of the ultraviolet modes, that this divergence does not persist when the equations of motion are solved to times larger than the inverse ultraviolet cutoff.

Acknowledgements

The author would like to thank K. Kajantie, L. McLerran and R. Fries for discussions that led to writing this Letter and R. Venugopalan for urging to actually write it up and comments on the manuscript. This manuscript has been authorized under Contract No. DE-AC02-98CH10886 with the US Department of Energy.

References

Exclusive electroproduction and the quark structure of the nucleon

Adam P. Szczepaniak * 1, J.T. Londergan 2

Department of Physics and Nuclear Theory Center, Indiana University, Bloomington, IN 47405, USA

Received 24 June 2006; received in revised form 21 September 2006; accepted 22 October 2006

Available online 30 October 2006

Abstract

The natural interpretation of deep inelastic scattering is in terms of hard scattering on QCD constituents of the target. We examine the relation between amplitudes measured in exclusive lepto-production and the quark content of the nucleon. We show that in the Bjorken limit, the natural interpretation of amplitudes measured in these hard exclusive processes is in terms of the quark content of the meson cloud and not the target itself. In this limit, the most efficient representation of these exclusive processes is in terms of leading Regge amplitudes.

© 2006 Published by Elsevier B.V.

PACS: 13.40.-f; 13.60.-r; 12.40.Nn

Keywords: Generalized parton distributions; Inclusive reactions; Exclusive reactions; Regge phenomenology

1. Introduction

Recent interest in hard exclusive lepto-production, in particular deeply virtual Compton scattering (DVCS) and meson production, has been stimulated by the idea that these processes may give new insights into the quark structure of the nucleon [1–10]. The connection between hard exclusive amplitudes and quark distributions in the nucleon, commonly referred to as generalized parton distributions, is formally analogous to that between the deep inelastic scattering cross-section and the structure functions. As shown by Feynman [11], structure functions can be interpreted in terms of quark probability distributions in the nucleon. Duality teaches us that, at least in principle, it is possible to use any channel to describe the scattering amplitude. The parton basis of deep inelastic scattering (DIS) is an example of a process that is most efficiently interpreted in the s-channel representation. The basis of quasi-free QCD constituents is the natural choice for expressing structure functions in the Bjorken scaling regime, $Q^2 \to \infty$ and finite $x_{BJ}$. In this regime the relevant matrix elements are diagonal in the parton Fock space basis. However even in the case of DIS the s-channel parton representation becomes less useful in the limit $x_{BJ} \to 0$. In this weel parton regime it becomes more efficient to parametrize structure functions in terms of amplitudes associated with t-channel processes. The physical interpretation of the structure functions changes in between these two regimes. As $x_{BJ} \to 0$ the structure function evolves to represent ladders of partons originating from t-channel meson exchanges.

As in the case of DIS, a factorization theorem in exclusive lepto-production enables one to separate the hard quark-photon (alternatively, $W$ or $Z$ boson) scattering from the target (nucleon) contribution [12]. The latter contribution is typically parametrized by the generalized parton distributions or GPDs [1,3,5]. In analogy with deep structure functions the GPDs are often interpreted as corresponding to some quark distributions of the nucleon [10]. Just as in the case of DIS, one can interpret hard exclusive lepto-production either in terms of s-channel exchanges or via t-channel exchanges.

Recently Mueller and Schaefer [13] produced a conformal spin expansion of GPDs. As part of their study they investigated the extent to which the GPDs displayed the characteristics of their leading Regge trajectories. When they examined the effective slope parameters for amplitudes corresponding to $\omega$ and...
ρ exchange, they found them to be extremely close to the phenomenological slopes for those trajectories, a result they called “quite astonishing”. Also, a recent analysis of ω electroproduction with the CLAS detector at Jefferson Laboratory [14] found that their data agreed quite well with standard Regge phenomenology. The purpose of this Letter is to show that in the Bjorken limit exclusive lepto-production amplitudes are most naturally described in terms of t-channel processes. Our results will demonstrate why one should expect that interpretation of the quark content of exclusive lepto-production processes will be most effectively discussed in the context of the parton content of reggeons, rather than of the nucleon.

Consider the case of exclusive vector meson production at high-s and low-t. As shown by a large body of evidence [15–18], such processes can be described by t-channel exchanges, where sums over exchanged mesons with all possible spins can be described by Regge trajectories [19]. The amplitude for a given Regge trajectory has the behavior s^{α(t)}. At asymptotic energies, W ≥ 10 GeV pomeron exchange dominates [20,21] since it has the largest intercept α(p)(0) ~ 1 and the process is predominantly s-channel helicity conserving.

In this Letter, we present simple arguments to justify our claim that hard exclusive processes are most naturally understood in terms of t-channel exchanges. For comparison purposes, this involves a review of some very well known material in both deep inelastic scattering and Regge phenomenology. Such a review is necessary in order to compare and contrast the underlying mechanisms that drive inclusive and exclusive lepto-production in the Bjorken limit, and to clarify the differences between the quark/nucleon amplitudes that can be extracted from these reactions.

2. The hadronic tensor in inclusive and exclusive lepto-production

Consider a deep inclusive reaction on a nucleon, a*(q) + N(p) → X. Here a*(q) is a virtual photon or weak gauge boson with momentum q, N(p) is a nucleon with momentum p, and X is the final state. To calculate the DIS cross section, one takes the square of this amplitude and sums over all unobserved final states X. As is well known [22], the resulting inclusive cross section can be obtained from the discontinuity across the right hand energy cut of the forward virtual Compton amplitude. This is a special case of the general exclusive amplitude

\[ a^*(q) + N(p) → b(q') + N(p'), \]

where a*(q) is a virtual boson (γ, W or Z) with momentum q, N(p) is a nucleon with momentum p, and X is the final state. To calculate the DIS cross section, one takes the square of this amplitude and sums over all unobserved final states X. As is well known [22], the resulting inclusive cross section can be obtained from the discontinuity across the right hand energy cut of the forward virtual Compton amplitude. This is a special case of the general exclusive amplitude

Under such approximations the absorptive part of the hadronic contribution to the cross section is determined from the hadronic tensor (a scalar function under the above approximations),

\[ T = \int d^4z e^{i\frac{2mN}{2}}e^{i\frac{2mN}{2}}(p')|T \left( \frac{z}{2} \right) |p). \]

In Eq. (2) j(z) = \Phi^+(z)\Phi(z) represents a (scalar) quark current in the Heisenberg picture which couples to the external fields representing the a and b particles. The Heisenberg nucleon states represent fully interacting nucleons; in particular, they include the meson cloud contribution. To study the valence and sea parton content of the nucleon the bare nucleon is often introduced within models that separate the QCD interactions among partons from chiral meson–nucleon interactions [23,24]. The xBJ = O(1) region is then found to be dominated by the bare nucleon and the sea contributes in the limit xBJ → 0, as expected.

From Lorentz symmetry it follows that the amplitude T in Eq. (2) is a function of four independent Lorentz scalars, T = T(Q^2, v, t, q^2) = T(Q^2, xBJ, t, q^2), with, v = p · q/m_N = Q^2/(2xBJm_N), and t = (p' − p)^2 = (q − q')^2. For inclusive processes we require the forward amplitude characterized by t = 0, 0 > q^2 = −Q^2, while the kinematics for exclusive processes require t < 0, 0 ≤ q^2 ≤ m_N^2. Using the operator product expansion to leading order in QCD one finds that the matrix elements of the time-ordered product of the quark currents can be replaced by the product of two quark field operators and the quark propagator:

\[ T(Q^2, v, t, q^2) = i \int d^4z d^4k \frac{e^{-ikz}}{(2\pi)^4} \frac{(q' + k)^2}{(q + q')^2} \langle p'|T \left( \frac{z}{2} \right) \Phi^+(z)\Phi\left(\frac{-z}{2}\right) |p). \]

Using Wick’s theorem the normal ordered product of fields (in the interaction picture with the interaction arranged as a power series of the QCD coupling) was replaced by the time-ordered product since the c-number difference between the two types of ordering does not contribute to connected matrix elements. The integral in Eq. (3) is dominated by points on the light cone with z^2 ~ O(1/Q^2). It is convenient to use light cone coordinates, A^μ = (A^+, A^-, A_⊥) with A^± = A^0 ± A_⊥ and to choose the frame in which q^+ = 0, q_⊥^2 = Q^2, p_⊥ = 0.

For inclusive reactions where q^2 = q, the quark propagator in Eq. (3) becomes

\[ \left( \frac{q + q'}{2} + k \right)^2 = Q^2 \left( \frac{x}{xBJ} - 1 \right) - q_⊥ · k_⊥ + k^2 \]

\[ \sim Q^2 \left( \frac{x}{xBJ} - 1 \right), \]

where x ≡ k_⊥/p_+. The fraction of the nucleon longitudinal momentum carried by the struck quark. The approximation is based on the observation that the matrix element in Eq. (3) does not involve hard scales and thus on average k_⊥, k_⊥^2 ≪ |Q|^2. Under such approximations the absorptive part of the hadronic

\[ \rho \]
tensor, \( W \equiv T(v + i\epsilon) - T(v - i\epsilon) \) which determines the DIS cross section is given by

\[
W(Q^2, x_{\text{BJ}}) = \frac{1}{Q^2} \int dx \, x \delta(x - x_{\text{BJ}}) F(x, Q),
\]

where \( F(x, Q) \) is the structure function,

\[
F(x, Q) = \frac{1}{2} p^+ \int dz e^{-i x P^+ z^- / 2} \times \langle p|T \left[ \phi^\dagger \left( \frac{z}{2} \right) \phi \left( \frac{-z}{2} \right) \right]|p\rangle_{z^+ = 0, |z^-| < 1/\sqrt{s}}.
\]

For exclusive production with \( Q^2 \gg q^2 \gg 0, (p - p')^2 = t < 0 \), again using light cone coordinates, to leading order in \( O(Q^2) \) the quark propagator can be approximated by,

\[
\left( \frac{q + q'}{2} + k \right)^2 = \frac{Q^2}{2} \left( \frac{x}{\xi} - 1 \right).
\]

In the Bjorken limit \( \xi = x_{\text{BJ}}/(2 - x_{\text{BJ}}) \), and the longitudinal component of quark momentum in this case is \( x = k^+/P^+ \) with \( P^+ = (p^+ + p'^+)/2 \). The hadronic tensor for exclusive processes in the Bjorken limit is therefore given by,

\[
T(Q^2, v, t, q^2) = \frac{p^+}{Q^2} \int dz^- dx \, i \xi e^{-i P^+ x z^-} \times \langle p'|T \left[ \phi^\dagger \left( \frac{z}{2} \right) \phi \left( \frac{-z}{2} \right) \right]|p\rangle_{z^+ = 0, |z^-| < 1/\sqrt{s}}.
\]

The positive energy cut contribution to the hadronic tensor, which determines the inclusive cross section, forces \( x = x_{\text{BJ}} > 0 \). This is not the case in exclusive processes; here the full amplitude \( T \) is needed to determine the cross section, so it contains an integral over both positive and negative \( x \). Defining the free quark and anti-quark creation and annihilation operators in the standard way in terms of field operators, it is possible to reinterpret the integration over the negative-\( x \) region of the quark (anti-quark) operator matrix elements in terms of the positive-\( x \) region of the anti-quark (quark) operator matrix elements [8].

Thus in the quark representation the matrix element in Eq. (8) receives contributions from pair creation and pair annihilation operators which mix different Fock space sectors of the nucleon wave function. Thus unlike DIS the DVCS matrix elements require non-diagonal overlaps of light front wave functions [25, 26]. More detailed analysis of the correspondence between current matrix elements and the light cone wave function representation is given in [25]. We also note that calculations of exclusive cross sections based on GPD models that employ the quark handbag phenomenology also include explicit contributions from meson exchanges, most notably an elementary \( t \)-channel pion exchange [27,28].

When an observable becomes sensitive to mixing between elements of a particular basis, it makes it difficult to interpret the internal structure of a state. This suggests that for hard exclusive processes there may be a more efficient representation of the matrix elements defining the observable. In the following we will show that just as a hierarchy of \( t \)-channel processes naturally explains the low-\( x \) behavior of DIS structure functions, the same is true for the amplitudes representing exclusive reactions in the entire kinematical region of Bjorken-\( x_{\text{BJ}} \).

### 3. \( t \)-channel dominance of exclusive lepto-production

Duality implies that all Feynman diagrams contributing to the hadronic tensor can be classified as either \( s \)-channel exchanges with baryon quantum numbers, or \( t \)-channel exchanges with meson quantum numbers. For large \( s \) and small \( t \), \( t \)-channel exchange of a particle with spin \( J \) is proportional to \( \beta_s J \) with the residue \( \beta \) depending on \( t \) and particle masses (which in our case includes the large virtual photon mass, \( Q^2 \)). Formally, in the limit where \( s/|t| \gg 1 \) the singularity \( J(\alpha(M^2)) = \alpha_0 + \alpha' M^2 \) and the sum over all exchanged mesons leads to an amplitude proportional to \( \beta_s q^{(t)} \). Such a description successfully reproduces the experimentally observed shrinkage of the forward peak with increasing energy [19]. In general in the regime where \( s/|t| \gg 1 \) the singularity is cut right to the rightmost complex angular momentum plane, determines the leading power of the energy dependence. We are concerned with reactions that obey the constraints \( s/|t| \gg 1 \). In this kinematic region one would expect \( t \)-channel exchanges to accurately parametrize these amplitudes. It is well known that in the case of DIS the Regge parametrization is relevant when \( x_{\text{BJ}} \to 0 \); however, for finite \( x_{\text{BJ}} \), the \( t \)-channel exchange description becomes inefficient. This occurs because away from the forward region, all singularities in the complex angular momentum plane i.e. all daughter trajectories contribute equally to the amplitude as the rightmost singularity, which defines the leading Regge trajectory. For DIS processes as one goes away from the region \( x_{\text{BJ}} \sim 0 \), it very quickly becomes more efficient to represent the amplitude by \( s \)-channel exchanges of quasi-free partons. However, we will show that the conditions that characterize exclusive production are quite different from the conditions governing the inclusive processes.

The contribution to the hadronic tensor from \( t \)-channel exchange of a spin-\( J \) meson is proportional to

\[
T_J = \frac{\beta_J^t(t) \beta_J^s(q^2, q'^2, t)}{t - M_J^2} \times \sum_{\lambda = -J}^J \left[ \left( p^+ + p \right)^{\mu_1} \cdots \left( p' + p \right)^{\mu_J} \right] \epsilon_{\mu_1 \cdots \mu_J} \left( p' - p \right) \left[ \left( q'^+ + q \right)^{\nu_1} \cdots \left( q' + q \right)^{\nu_J} \right] \epsilon_{\nu_1 \cdots \nu_J} \left( p' - p \right).
\]

In Eq. (9), \( \epsilon \) is the spin-\( J \) polarization vector, and \( \beta_J^t \) and \( \beta_J^s \) are the residue functions at the lower and upper vertex, respectively. This is shown schematically in Fig. 1. In the Bjorken limit, \( s \to Q^2(1 - x_{\text{BJ}})/x_{\text{BJ}} \) and the amplitude reduces to

\[
T_J = \frac{\beta_J^t(t) \beta_J^s(q^2, q'^2, t)}{t - M_J^2} \left( \frac{Q^2}{2x_{\text{BJ}}} \right)^J.
\]

The key question is how the upper residue function depends on the large variables \( Q^2 \) and \( -q'^2 = Q^2 \) in the case of inclusive processes, and \( Q^2 \) for the exclusive amplitudes. It is well
known that for kinematics that are relevant to inclusive scattering, the upper residue function behaves as \((1/Q^2)^{J+1}\), modulo logarithmic corrections, so that the amplitude scales, \(Q^2 T_J \propto (1/x_{BJ})^J\), as expected [29–32]. Summing over all spins leads to the Regge behavior, \(Q^2 T = \sum J T_J \propto (1/x_{BJ})^{a(0)}\). The leading Regge trajectory with \(a(0) > 0\) will dominate the \(x_{BJ} \to 0\) behavior of the hadronic tensor, while all other trajectories with \(a_n(0) < a(0)\) are subleading as \(x_{BJ} \to 0\). For finite \(x_{BJ}\), however, the leading Regge trajectories are no longer suppressed, and as a result the Regge description becomes ineffective while the \(s\)-channel parton model description becomes natural.

We will now show that for exclusive amplitudes the upper vertex scales with a finite power of \(1/Q^2\) instead of being suppressed for high spins. Thus after summing over all spins it gives an amplitude that behaves as \(1/Q^2\), as expected. However for exclusive amplitudes, \(s\)-channel parton model description becomes natural.

To show that, we first rewrite the contribution of a \(t\)-channel, spin-\(J\) exchange to the matrix element in Eq. (3) in terms of the two-current correlation in the exchanged meson,

\[
\beta^u_J(n) = \int d^4 z e^{-ikz} \langle p'|T \left[ \phi \left( \frac{z}{2} \right) \phi \left( -\frac{z}{2} \right) \right]|p \rangle
\]

\[
\times \int_J \frac{1}{t - M_{Jn}^2} \left[ k^{\mu_1} \cdots k^{\mu_J} \epsilon_{\mu_1 \cdots \mu_J}^\ast (p') \right] \left[ (p' + p)^{\mu_1} \cdots (p' + p)^{\mu_J} \right] .
\]

In Eq. (11), \(n\) refers to other quantum numbers that distinguish between exchanged mesons (after reggeization \(J\) distinguishes between the leading and daughter trajectories), and \(\Phi_{Jn}(\Delta, k)\) is the covariant Bethe–Salpeter amplitude of a meson with momentum \(\Delta\), where \(\Delta^2 = t\). Finally \(k\) is the relative momentum between the quarks, as shown in Fig. 1. Using dispersion relations, the (unnormalized) Bethe–Salpeter amplitude can be represented as [33],

\[
\Phi_{Jn}(\Delta, k) = i \int d\mu^2 \frac{g_{Jn}(x, \mu^2)}{(k - \frac{i}{2} \Delta^2 - \mu^2 + i\epsilon)^{n+J}} ,
\]

where the spectral density \(g_{Jn}\) is related to the parton distribution amplitude in a meson and can in principle be constrained from electromagnetic data [34] and QCD asymptotics [35].

What is important for our argument are the following model independent features of the amplitude in Eq. (12). The magnitude of the relative momentum \(k\) is of the order of the hadronic scale \(\mu_{m}\). Secondly, in the infinite momentum frame, \(k \propto \Delta\), \(\xi \equiv (1 \pm \Delta)/2\) represents the fraction of the longitudinal momentum carried by the quark (anti-quark), and \(q\) becomes the parton distribution function. In the Bjorken limit the leading, helicity-zero component of the meson distribution amplitude has \(J\)-independent behavior near \(\xi \to 1\) [36]. Finally the power dependence of the relative momentum is constrained by the angular momentum i.e. the power of the denominator in Eq. (12) increases with \(J\). Inserting the analytical expression for the Bethe–Salpeter amplitude of Eq. (12) into Eq. (11) and then into Eq. (3), one obtains the final expression for the contribution of spin-\(J\) exchange to the hadronic tensor. It is given in terms of an integral over \(k\) (see Eq. (3)) of the product of the quark propagator, the Bethe–Salpeter amplitude of Eq. (12), and a polynomial in \(k\) originating from the coupling to the spin-\(J\) polarization vectors (Eq. (11)). The polynomial is responsible for the \(s^2 \sim (Q^2)^2\) behavior of the amplitude. The integral can easily be evaluated using the Feynman parametrization which introduces an integral over the parameter \(\alpha\).

Ignoring terms of order \(m_{\pi}^2/Q^2\) and \(t/Q^2\), up to constant numerical factors one finds

\[
\beta^u_J = \int d\mu^2 g_{Jn}(x, \mu^2) \times \int d\alpha \frac{\alpha^J}{1 - x - O(\mu^2/Q^2)},
\]

For inclusive amplitudes, \(q^2 = q^2 = -Q^2\) the \(x\) disappears from the denominator and the integration over \(\alpha\) is dominated by \(\alpha \sim \mu^2/Q^2\). As a result, the entire integral is of order \((\mu^2/Q^2)^{J+1}\), as expected. However for exclusive amplitudes, \(q^2 \sim 0\) the integrand is dominated by the region \(1 - x = O(\mu^2/Q^2)\), and finite \(\alpha\). The endpoint behavior of the distribution amplitudes \(g_{Jn}\) is spin independent, and for leading-twist amplitudes \(g_{Jn}(x \to 1) \sim (1 - x)\). This leads to a \(J\)-independent suppression of the upper vertex with \(Q^2\), \(\beta^u_J \sim O(\mu^2/Q^4)\) which is independent of the spin of the exchanged meson. This is our main result. As discussed above, upon summing over all spins from a single trajectory one determines that the hadronic tensor is proportional to \((Q^2/x_{BJ})^{0(J)} \sim (Q^2/x_{BJ})^{0(J)}\), for small \(t\). Thus, in the Bjorken limit hard exclusive processes should be dominated by a single, leading Regge trajectory for all \(x_{BJ}\), and not just for \(x_{BJ} \to 0\). We argue that this is the most efficient way to interpret hard exclusive processes.
processes. We also note that the Regge approach to exclusive deeply virtual production was previously considered in [37], where a different $Q^2$ dependence was obtained for the full exclusive amplitude. However, those authors assumed a Regge-like amplitude with a particular $Q^2$ dependence, rather than deriving the behavior from a sum of $t$-channel poles as was done here. We also note that our analysis respects QCD factorization and is based on the same OPE that can be applied to the standard "handbag" picture. QCD corrections will in general induce corrections to the $Q^2$ dependence but in leading twist will not change the power behavior, and thus it will not alter our conclusions.

As we mentioned earlier, a recent experimental analysis of $\omega$ electroproduction at Jefferson Laboratory [14] showed that their data was in good agreement with predictions from models of generalized parton distributions [38]. Note that our results were derived in the Bjorken limit with analyses based on models of generalized parton distributions [40].

We would like to thank Stan Brodsky, Jean-Marc Laget, Wally Melnitchouk, Dieter Mueller, Anatoly Radyushkin, and Christian Weiss for useful comments and discussions.

Acknowledgements

We would like to thank Stan Brodsky, Jean-Marc Laget, Wally Melnitchouk, Dieter Mueller, Anatoly Radyushkin, and Christian Weiss for useful comments and discussions.

References

Evidence for the absence of gluon orbital angular momentum in the nucleon

Stanley J. Brodsky \(^a\), Susan Gardner \(^b,\ast\)

\(^a\) Stanford Linear Accelerator Center, Stanford University, Stanford, CA 94309, USA
\(^b\) Department of Physics and Astronomy, University of Kentucky, Lexington, KY 40506-0055, USA

Received 27 August 2006; received in revised form 13 October 2006; accepted 18 October 2006
Available online 27 October 2006
Editor: B. Grinstein

Abstract

The Sivers mechanism for the single-spin asymmetry in unpolarized lepton scattering from a transversely polarized nucleon is driven by the orbital angular momentum carried by its quark and gluon constituents, combined with QCD final-state interactions. Both quark and gluon mechanisms can generate such a single-spin asymmetry, though only the quark mechanism can explain the small single-spin asymmetry measured by the COMPASS Collaboration on the deuteron, suggesting the gluon mechanism is small relative to the quark mechanism. We detail empirical studies through which the gluon and quark orbital angular momentum contributions, quark-flavor by quark-flavor, can be elucidated.

The nucleon is a composite particle with spin \(\frac{1}{2}\). There is little doubt that the theory of quantum chromodynamics (QCD) describes the manner in which the nucleon’s spin is carried by its constituents, yet clarifying the details of this picture has incited intense theoretical and experimental activity [1]. Much has been made of the empirical fact that the spin of the nucleon is not given by the net helicity of its valence quarks [2]; however, this is not so much a “crisis” for QCD as it is for the non-relativistic quark model, since the latter rationalizes the charges, spins, and magnetic moments of the baryons in terms of the properties of its constituent quarks. The rich structures revealed through deeply inelastic scattering experiments on the proton [3] and through Drell–Yan production of massive lepton pairs with hadron beams and targets [4], compellingly demonstrate the limitations of such a simple picture.

The nucleon contains both quark and gluon components in QCD, so that its spin of \(\frac{1}{2}\) must follow from the sum of the spin and orbital angular momenta carried by these constituents:

\[
\frac{1}{2} = L_q^{\text{net}} + \frac{1}{2} \Delta \Sigma + L_g + \Delta g,
\]

where we write \(\Delta \Sigma\) for the net helicity of the quarks. Our decomposition is referenced to a polarization axis, so that \(L_q^{\text{net}}\) and \(L_g\) are the components of the orbital angular momentum, due to quark and gluon constituents, respectively, with respect to that axis. The decomposition is not unique [5]. In what follows, we will use the decomposition based on the angular momentum tensor in light-cone gauge, \(A^+ = 0\) [5], so that the gluon constituents have physical polarization \(S^z = \pm 1\), and there are no ghosts [6]. Our conclusions concerning the decomposition of the nucleon spin will thus be specific to light-cone gauge, giving us a natural connection to the physics of light-front wave functions [6], which are invariably defined in this gauge. We note that a manifestly gauge-invariant decomposition is also possible [7].

\(\ast\) Work supported in part by the US Department of Energy under contracts DE-AC02-76SF00515 and DE-FG02-96ER40989.
\ast Corresponding author.

E-mail address: gardner@pa.uky.edu (S. Gardner).

© 2006 Elsevier B.V. All rights reserved.
Light-front wave functions (LFWFs) enjoy many advantages: they are frame independent, and the spins of the constituents satisfy $J^z = \sum_{i=1}^n S_i^z + \sum_{i=1}^n L_i^z$. Fock state by Fock state for a polarization axis $z$—we emphasize that there are only $n - 1$ internal orbital angular momenta for a given Fock state with $n$ constituents. The LFWFs are the eigensolutions of the QCD Hamiltonian defined at fixed light-front time $\tau = t + z/c$. Indeed, LFWFs are the natural way to understand the structure of hadrons as probed through lepton scattering experiments. For example, the computation of the electromagnetic elastic form factors in the light-front formalism [8,9] yields the insight [10] that the anomalous magnetic moment defined at fixed light-front time $n$ components carry non-zero transverse orbital angular momentum, i.e., if $S_1^z \cdot L_1^z \neq 0$. We neglect fundamental T violation, so that $\kappa$ is real [11]. Since the proton’s anomalous magnetic moment is nearly twice that of its Dirac magnetic moment, a “spin crisis” in DIS could have been altogether expected.

In this Letter we will study constraints on the orbital angular momentum of the nucleon’s constituents using the azimuthal single-spin asymmetries produced from a target polarized transverse to the reaction plane. Such asymmetries have been seen in a variety of reactions, although we focus on those observed in semi-inclusive deeply inelastic lepton scattering (SIDIS), as in, e.g., $\ell p \rightarrow \ell' p' X$. In general, the single-spin asymmetry is proportional to the invariant form $\epsilon_{\mu\nu\sigma} P^{\mu} S_{n}^{\nu} p_{\pi}^{\sigma} q^{T}$ where $S_n$, the nucleon spin, satisfies $S_P^z = -1$ and $S_P \cdot P = 0$, where $P$ is the nucleon momentum and $q = P_P - p_P$ is the momentum transfer, with $Q^2 = -q^2$. The correlation is proportional to $S_P \cdot p_{\pi} \times q$ in the target rest frame, and since it is of leading twist, it obeys Bjorken scaling [12]. This pseudo-T-odd correlation is engendered by final-state interactions (FSI) of the struck quark, and thus it does not reflect a fundamental violation of time-reversal invariance [12,13]. The discrete symmetry transformations in the light-front formalism are studied in detail in Ref. [11].

The azimuthal single-spin asymmetry (SSA) for $x^\pm$ production in SIDIS from a unpolarized beam and transversely polarized target, is defined as [14]

$$A_{UT}^{\pm}(\phi, \phi_s) \equiv \frac{1}{|\langle S_P \rangle|} \left( \frac{N_{\pi^+}^{1\pm}(\phi, \phi_s) - N_{\pi^-}^{1\pm}(\phi, \phi_s)}{N_{\pi^-}^{1\pm}(\phi, \phi_s)} \right) = A_{UT}^{\pm}(\phi + \phi_s) + A_{UT}^{\pm}(\phi - \phi_s) + \cdots,$$

where the “C” and “S” superscripts refer to the asymmetries generated by the Collins and Sivers effects, respectively, and the definition of the azimuthal angles $\phi$ and $\phi_s$ are as in Ref. [14]. In the Collins mechanism [15], the asymmetry is formed through the product of the transversity distribution and a pseudo-T-odd spin-dependent fragmentation function which describes the correlation of the transverse polarization of the struck quark with the transverse momentum of the produced hadron. In the Sivers mechanism [16], the asymmetry arises from the product of a pseudo-T-odd distribution function which describes the correlation of the transverse momentum of the struck quark with the transverse nucleon spin and a spin-independent fragmentation function [17,18].

The asymmetries arising from the two mechanisms differ in their precise azimuthal angle dependence and can be separated as in Eq. (2) [14]. If the thrust [19] of the quark jet can be empirically determined, then the Sivers asymmetry can be measured directly, where we replace $p_{\pi}$ with $p_s$. We consider the Sivers effect exclusively.

As discussed in Ref. [12], a non-zero Sivers SSA follows from the interference of two amplitudes $\mathcal{M}[\gamma^* p(p^\circ) \rightarrow F]$ of differing nucleon spin $J^P = \pm 1/2$ which couple to the same final-state $F$. In particular, the quantum numbers of the struck quark are the same in each case. The polarization axis $y$ is chosen transverse to the scattering plane. The two amplitudes must also differ in their strong phase to generate a non-zero SSA, so that the SSA is proportional to $\text{Im}(\mathcal{M}[J^P = +1/2]^\circ\mathcal{M}[J^P = -1/2])$. The photon cannot flip the helicity of the struck quark, so that the two amplitudes differ by $|\Delta L^z| = 1$ to yield a non-zero result. The requisite matrix element is related, but not identical, to the matrix element which generates the anomalous magnetic moment [12]. In particular, the presence of a strong phase engendered by FSI is essential to generating a non-zero SSA.

The Sivers SSA $A_{UT}^{\pm}(x, k_{\perp})$ is determined by the function $f_{1T}^{\pm}(x, k_{\perp}^2)$, which is subsumed in $f_{q/p}^{\perp}(x, k_{\perp})$, the distribution of unpolarized quarks in a transversely polarized proton of spin $S$ and mass $M$; we define [20]

$$f_{q/p}^{\perp}(x, k_{\perp}) = f_{1}^{q}(x, k_{\perp}^2) - f_{1T}^{q}(x, k_{\perp}^2) \frac{\epsilon_{\mu\nu\sigma} P_{\mu} S_{n}^{\nu} p_{\pi}^{\sigma}}{M (P \cdot n)} = f_{1}^{q}(x, k_{\perp}^2) - f_{1T}^{q}(x, k_{\perp}^2) \frac{(\hat{P} \times k_{\perp}) \cdot S}{M},$$

in a frame where $\hat{P}$ and $n$ are auxiliary lightlike vector, point in opposite directions. We thus have

$$A_{UT}^{\pm} = \frac{2}{M} \frac{\langle \sum_{q} |k_{\perp}| e_{q} f_{1T}^{q}(x, k_{\perp}^2) \rangle D(z, p_{\pi}, k_{\perp}) \sin^2(\phi - \phi_s)}{\langle \sum_{q} e_{q} f_{1}^{q}(x, k_{\perp}^2) \rangle D(z, p_{\pi}, k_{\perp})},$$

where $D(z, p_{\pi}, k_{\perp})$ contains the fragmentation function $D_{q}^{\pm}(z, p_{\pi})$ and (\cdots) refer to the appropriate angle and $k_{\perp}$ integrals [21]. The quark electric charge $e_q$ dependence follows since $f_{q/p}^{\perp}(x, k_{\perp})$ comes from the square of the scattering amplitude. The function $f_{1T}^{q}(x, k_{\perp}^2)$ can be extracted from $f_{q/p}^{\perp}$, which is defined in a gauge-invariant way via [22,23]

$$f_{q/p}^{\perp}(x, k_{\perp}) = \int \frac{d^2z}{16\pi^3} e^{-ik_{\perp} \cdot \xi \cdot z} \langle P | \bar{\psi}(\xi, \bar{\phi}) | \infty, \infty, \xi^- , \xi^\perp \rangle \bar{\gamma}^{\gamma^+} [\infty, \infty, 0^- , 0^\perp \rangle C \psi(0, 0^\perp) | P \rangle,$$
where \([\cdot \cdot \cdot]_C\) denote gauge links stretched in both light-like and transverse directions [22], capturing the final-state interactions necessary for the SSA. As we shall see, the latter imply that the Sivers function can only be related, rather than identical to, the matrix element for the anomalous magnetic moment.

The \(\triangle L^\gamma\) needed for a non-zero single-spin asymmetry can stem from two distinct physical sources: \(\triangle L^\gamma\) can arise from either quark or gluon degrees of freedom [24,25]. In the first case the virtual photon strikes a quark in a Fock component of the nucleon’s light-front wave function, whereas in the second case, the virtual photon fuses with a gluon in some \(\{qqgqg \cdots\}\) Fock state of the nucleon to produce a \(q \bar{q}\) pair. Analogues of both mechanisms can also contribute to the nucleon’s anomalous magnetic moment \(\kappa\), as we shall detail. However, the empirical anomalous magnetic moments of the proton and neutron sum to nearly zero, suggesting that the gluon contribution to the anomalous magnetic moment is small. We shall argue that a similar cancellation observed in the SSA data from a deuteron target allows us to conclude that the gluon mechanism is small in this case as well. The two mechanisms we discuss are distinct, in part, because the virtual photon interacts with either an “intrinsic” quark, namely, a multiply-connected constituent of a multi-parton Fock state, or an “extrinsic” quark, produced as a member of a \(qq\) pair from photon–gluon fusion. The intrinsic quark carries an orbital angular momentum \(L_q\), whereas the extrinsic quark carries, in part, the orbital angular momentum of the gluon constituent \(L_g\) which spawns it. We can similarly distinguish intrinsic from extrinsic gluons, so that an intrinsic gluon is also a multiply-connected constituent of a multi-parton Fock state. Our separation can be clouded by QCD evolution effects, for an extrinsic gluon spawned by the DGLAP (Dokshitzer–Gribov–Lipatov–Altarelli–Parisi) splitting \(q \to qg\) can fuse with the virtual photon to generate extrinsic quarks and a SSA. This mechanism thus serves to dilute the correlation between the gluon dynamics and the intrinsic gluon orbital angular momentum contribution to the proton spin.

We emphasize that the two mechanisms, quark and gluon, differ in their isospin character and, indeed, that SSA data on the proton and deuteron can be used to distinguish them. We note that a large SSA for \(p^+\) production from a transversely polarized proton has been observed by the HERMES collaboration [14]; however, when an analogous observable is measured by the COMPASS collaboration from a deuterium target [26], the SSA is consistent with zero. The polarization of the deuteron itself is used to define the spin correlation. The deuteron with spin \(S^\gamma = +1\) normal to the scattering plane has both nucleon spins aligned \(S^\gamma_p = +1/2\) and \(S^\gamma_n = +1/2\). Since the deuteron is a weakly bound state, the SSA from the deuteron is the sum of single-spin asymmetries (SSAs) for the proton and neutron to a very good approximation. The small SSA observed with a \(^3\)H target is, in fact, natural if the matrix element is related to that of the anomalous magnetic moment—the empirical \(p\) and \(n\) anomalous magnetic moments sum to nearly zero. However, we shall show that this connection can only occur with the \(L_q\) mechanism; the isospin structure of the \(L_g\) mechanism is altogether distinct. In this regard whether the gluon-mediated SSA emerges from intrinsic or extrinsic gluons is without consequence. We note, in passing, that the use of other polarized nuclear targets can give empirical checks of these observations; for example, polarized \(^3\)He offers an effective neutron target, up to “spin dilution” corrections of some 10% [27]. We shall now develop these ideas in detail.

To set the stage, we review the manner in which the anomalous magnetic moment of the nucleon is connected to the quark orbital angular momentum in the light-front formalism. Working in the interaction picture for the electromagnetic current \(J^\mu(0)\) and the \(q^+ = 0\) frame [8,9], we have [10]

\[
\kappa = -\sum_a \left[ \sum_j e_j \int [d^2k_{\perp}] \Psi^a_\perp(x_j, k_{\perp j}, \lambda_i) S_{\perp} \cdot L_{\perp j}^q \psi_a(x_j, k_{\perp j}, \lambda_i) \right] = \sum_q e_q a_q,
\]

where we write \(\kappa\) in units of \(e/2M\) and define \(S_{\perp} \cdot L_{\perp j}^q \equiv (S_+ L_{\perp j}^q + S_- L_{\perp j}^q)/2\) with \(S_\pm = S_1 \pm i S_2\) and \(L_{\perp j}^q = \sum_{q\neq j} \lambda_i (\partial/\partial k_{\perp i}) = i\delta/\delta k_{\perp j}\)—the last sum is over quark flavor \(q\). Consequently, the orbital angular momentum contribution \(L_{\perp j}^q\) associated with a struck quark \(j\) in Fock state \(a\) is not an independent variable, but, rather, is determined by the sum of the orbital angular momenta of all the spectator partons in that Fock state. This notion also gives rise to the vanishing anomalous gravitomagnetic moment for composite systems, Fock state by Fock state [28]. Although we regard \(L_{\perp j}^q\) as the transverse orbital angular momentum associated with the struck quark \(j\), the explicit sum over \(i \neq j\) makes it apparent that the transverse orbital angular momenta carried by gluon spectators implicitly contributes to its definition. Moreover, both quark and gluon contributions from the parent nucleon Fock state are captured by the matrix element of the \(L_{\perp j}^q\) operator, as the gluon can fluctuate to a \(q \bar{q}\) pair, to which the photon can couple. It is these quark- and gluon-mediated contributions which we distinguish as the “quark” and “gluon” mechanisms. We note that the light-front formalism in \(A^+ = 0\) gauge permits a simple kinematic operator representation of the \(L_\perp\) operator; this, in turn, permits \(S_{\perp} \cdot L_{\perp}^q\) to act as a ladder operator which raises or lowers the value of \(L_\perp\) in this representation. In the last equality of Eq. (6) we subsume the Fock-state sum to define the contribution to the anomalous magnetic moment, quark flavor by quark flavor; the \(a_j\) are real. The phase-space integration is given by

\[
\int [d^2k_{\perp}] \sum_{\lambda_i, c_i, f_i} \prod_{j=1}^{n} \left( \int \frac{d^2k_{\perp j}}{(2\pi)^3} \right) |16\pi^3 \delta \left( 1 - \sum_{i=1}^{n} x_i \right) \delta^{(2)} \left( \sum_{i=1}^{n} k_{\perp i} \right)|,
\]

where \(n\) denotes the number of constituents in Fock state \(a\), and we sum over the possible \(\{\lambda_i\}, \{c_i\}\), and \(\{f_i\}\) in state \(a\). The summations of Eq. (6) are over all contributing Fock states \(a\) and struck constituent charges \(e_j\); we refrain from including the
constituents’ color and flavor dependence in the arguments of the light-front wave function (LFWF) $\psi^S_a$, which we define in the $A^+ = 0$ gauge, with the principal-value prescription for singularities in $k^+$. We emphasize that both quark and gluon degrees of freedom in the nucleon’s LFWF contribute to Eq. (6). Either an intrinsic or extrinsic gluon constituent in the nucleon can couple to a photon via a $q\bar{q}$ pair. The lowest-order effective $γ^*gg$ vertex, a contribution to the anomalous magnetic moment, is forbidden by C invariance, though the radiation of an extra gluon from the effective vertex removes this constraint and makes it finite. We note the gluon mechanism should contribute to $κ_p$ and $κ_n$ with equal weight, up to isospin-breaking effects, yet the empirical isoscalar magnetic moment of the nucleon, $κ_s \equiv (κ_n + κ_p)/2 = -0.06$, is numerically very small relative to $|κ_n,p|$—suggesting that the gluon contribution to the nucleon anomalous magnetic moments is itself small.

Returning to the Sivers function, we define $f^q_{l/p\perp}(η,1_L) = \tilde{u}(P,λ')\left[f^q_T(η,1_L) + f^q_{1T}(η,1_L)iσ_{+a}1/2 M\right]u(P,λ)$, where $u(P,λ)$ is a Dirac spinor associated with a spin-$1/2$ state of momentum $P$ and helicity $λ$ [29], with $y$ the polarization axis—only the $α = 1, 3$ matrix elements are non-zero. We identify $f^q_{l/p\perp}(η,1_L)$ as the function $q(η,1_L)$ of Ref. [22], where we work in leading twist and ignore all QCD evolution effects. Turning to the light-front formalism in $A^+ = 0$ gauge, with boundary conditions appropriate to SIDIS, a Fock component of the proton’s LFWF has the form $\tilde{ψ}_a^S = ψ_a^S \exp(iφ_a^S)$; we emphasize that the LFWFs are complex in this case [22]. Note that we contrast the LFWF for a proton in isolation to that for a proton immersed in an external electromagnetic gauge field. For simplicity we assume the LFWFs differ only in the phase $φ_a^S$. Working in the $q^+ = 0$ frame we thus identify

$$f_{l/p\perp}^q(η,1_L)\frac{l_1}{M} = -\frac{i}{2} \sum_a \sum_{i=1}^n q_{bij} \int \! [d^2k] \left[\frac{1}{2i(1-x_i)} [\tilde{ψ}_a^S(x_i',k_{i\perp},λ_i)\tilde{S}_{lT} \cdot \tilde{L}_{qT}^j \tilde{ψ}_a(x_i,k_{i\perp},λ_i)] - \bar{ψ}_a^S(x_i',k_{i\perp},λ_i)\bar{S}_{lT} \cdot \bar{L}_{qT}^j \bar{ψ}_a(x_i,k_{i\perp},λ_i)] \right],$$

(9)

where $k_{i\perp} = k_{i\perp} + l_{1\perp}$ and $x'_i = x_j + η$ for the struck constituent $j$ and $k_{j\perp} = k_{j\perp} - x_1l_{1\perp}/(1 - x_j)$ and $x'_j = x_j[1 - η/(1 - x_j)]$ for each spectator $i$, where $i ≠ j$. The existence of a term in $1_L$ mandates not only orbital angular momentum [30] but also the imaginary parts in the right-hand side (RHS) of Eq. (9)—a FSI phase must be present to incur a SSA. We have suppressed the arguments of $φ_a^S$ but assert that it depends on the magnitude and not the direction of $k_{lT}^j$, so that the $|ΔL^j| = 1$ structure of the matrix element entails the $l_1$ dependence, specifically that the RHS $\sim -(i/2)[(l_3 - il_1) - (l_1 + il_1)] = -l_1$, as found in explicit model calculations [12,23]. The Kronecker $δ$ ensures that the struck quark is of flavor $q$. Following the development of Eq. (6) in Ref. [10], we find that $f_{l_{1T}}^q(x_i,1_L)$ can also be written in terms of the matrix element of a spin-orbit operator, if we consider the leading terms $1_L \rightarrow 0$:

$$f_{l_{1T}}^q(η,0)\frac{l_1}{M} = \sum_a \sum_{i=1}^n q_{bij} \int \! [d^2k] \left[2i(l - x_i) \left[\tilde{ψ}_a^S(x_i',k_{i\perp},λ_i)\tilde{S}_{lT} \cdot \tilde{L}_{qT}^j \tilde{ψ}_a(x_i,k_{i\perp},λ_i) - \bar{ψ}_a^S(x_i',k_{i\perp},λ_i)\bar{S}_{lT} \cdot \bar{L}_{qT}^j \bar{ψ}_a(x_i,k_{i\perp},λ_i)] \right].$$

(10)

where $\tilde{ψ}_a^S = ψ_a^S \exp(-iφ_a^S)$, $\tilde{S}_{lT} \cdot \tilde{L}_{qT}^j = (S_{lT}L_{qT}^j - S_{qT}L_{lT}^j)/2$, $S_{lT} = S_3 \pm iS_1$, and $L_{qT}^j = \sum_{i≠j} x_i(∂/∂k_{3i} ± i∂/∂k_{1i})$. We define

$$f_{l_{1T}}^q(η,1_L)\frac{l_1}{M} = -\tilde{a}_q(η,1_L),$$

(11)

where the $\tilde{a}_j$ are real. A comparison of Eqs. (6) and (10) prompts us to include a minus sign in the definition of $\tilde{a}_q(x_i,1_L)$. We note in passing that the generalized parton distribution $E(x, ζ, τ)$ probed in virtual Compton scattering (VCS) can also be connected to the nucleon’s orbital angular momentum. The generalized form factors in VCS, $γ^* (q) + p(P) → γ^* (q') + p(P')$ with $t = Δ^2$ and $Δ = P - P' = (τ P^++ Λ_Δ, (τ + Δ^2)/τ P^+)$, have been constructed in the light-front formalism [32]. Using Eq. (40) of Ref. [32], and the procedure and syntax of Eq. (6), noting $q_{l→Δ_1}$, we determine, for $ζ ≤ x ≤ 1$,

$$\frac{E(x,ζ,0)}{2M} = \sum_a (\sqrt{1-ζ})^{1-n} \sum_j δ(x - x_j) \int \! [d^2k] [\tilde{ψ}_a^S(x_i',k_{i\perp},λ_i)\tilde{S}_{lT} \cdot \tilde{L}_{qT}^j \tilde{ψ}_a(x_i,k_{i\perp},λ_i)],$$

(12)

with $x'_j = (x_j - ζ)/(1 - ζ)$ for the struck parton $j$ and $x'_i = x_i/(1 - ζ)$ for the spectator parton $i$. We emphasize that the LFWFs for Fock-state $a$ with spin up or down for fixed struck quark helicity differ by $|ΔL^j| = 1$ because $L_{qT}^j$ contains ladder operators.

---

1 An overall sign should appear in Eq. (21) in Ref. [12], as noted in Ref. [31].
The SSAs and the anomalous magnetic moments

We see that the matrix element, Eq. (10), which drives the Sivers SSA bears similarity to that of the anomalous magnetic moment [12,31,33–37]. To understand the consequences of this in a transparent way, we recall that under an assumption of isospin symmetry, we have \( a^u_d = a^u_u = a^d_d, \) \( a^u_d = a^d_u, \) \( a^u_d = a^u_u, \) and \( a^d_d = a^d_u, \) for sea quarks of other flavors. Isospin symmetry acts at the level of the hadronic matrix elements; the quark charges are not isospin mirrors. Neglecting the contributions of anti-quarks, we have

\[
\begin{align*}
\kappa_p &= 1.79 = (+2/3)a^p_u + (-1/3)a^p_d, \\
\kappa_n &= -1.91 = (-1/3)a^n_u + (+2/3)a^n_d = (+2/3)a^p_d + (-1/3)a^p_u,
\end{align*}
\]

to yield \( a^u_d = a^u_u = -2.03 \) and \( a^d_u = a^d_d = 1.67. \) Concommitant isospin relations follow for the \( \tilde{a}_q \) of Eq. (11) as well. We also neglect the contributions of the anti-quarks to the existing SSA data, which is consistent with recent models [21,38]. In what follows, if we conjecture that the isospin structure of the empirical anomalous magnetic moments is that of the Sivers SSA, then we find the relative strength of the various \( \tilde{a}_q \) to be isoscalar, the extra factor of \( \frac{1}{\sqrt{2}} \) in the matrix element of Eq. (10) could change the relative strength of the \( u \) and \( d \) contributions. Nevertheless, considering the consequences of this simple picture for the deuteron data, we note that \( a^u_d + a^u_u + a^d_d = -0.360, \) implying that the SSA for \( u \)-quark fragmentation from a proton is increasing, so that the gluon mechanism is distinctive, for a gluon in a nucleon Fock state will produce \( uu \) or \( dd \) pairs with equal weight—up to tiny differences driven by the \( u-d \) mass difference. Thus the \( u \)-quark and \( d \)-quark SSAs add constructively in SIDIS from a \( ^2 \)H target; there is no cancellation of this \( 1 \)-dependence.

\[ (L_\perp)^2 \]

Thus, we can estimate the relative strength of the two mechanisms. We find that the gluon mechanism is smaller than the quark mechanism by a factor of 4, since the asymmetry is controlled by \( e \) production should indeed be small, and this is observed [14]. Recent model fits [21,38] are consistent with these trends. To be more specific, we note the fits of Ref. [38] yield \( S_u = -0.81 \pm 0.07 \) and \( S_d = 1.86 \pm 0.28, \) where \( S_p \) is to be compared to \( -\langle \tilde{a}_q \rangle, \) with \( \langle \tilde{a}_q \rangle \) defined to be the average value of \( \tilde{a}_q (q, \underline{F}_1). \) This implies that the SSA asymmetry from \( u \)-quark fragmentation is smaller than that that we have

\[
\tilde{a}_q (q, \underline{F}_1) = \frac{1}{\sqrt{2}} \left( \frac{1}{3} a^p_q + \frac{2}{3} a^n_q \right).
\]

As to the magnitudes of the SSAs, although \( |a^p_u| < |a^p_d|, \) the SSA engendered by \( u \)-quark fragmentation from a proton is enhanced by a factor of 4, since the asymmetry is controlled by \( e^+ \). In the absence of anti-quark contributions, the SSA for \( \pi^- \) production should indeed be small, and this is observed [14]. Recent model fits [21,38] are consistent with these trends. To be more specific, we note the fits of Ref. [38] yield \( S_u = -0.81 \pm 0.07 \) and \( S_d = 1.86 \pm 0.28, \) where \( S_p \) is to be compared to \( -\langle \tilde{a}_q \rangle, \) with \( \langle \tilde{a}_q \rangle \) defined to be the average value of \( \tilde{a}_q (q, \underline{F}_1). \) This implies that the SSA asymmetry from \( u \)-quark fragmentation is smaller than that that we have estimated from the use of the empirical anomalous magnetic moments alone. Although we expect the strong phase from the Wilson line to be isoscalar, the extra factor of 1/(1 \( -x \jmath \) ) in the matrix element of Eq. (10) should change the relative strength of the \( u \) and \( d \) contributions. Nevertheless, considering the consequences of this simple picture for the deuteron data, we note that \( a^u_d + a^u_u + a^d_d = -0.360, \) implying that the SSA for \( u \)-quark fragmentation to leading positively charged hadrons, as well as \( d \)-quark fragmentation to leading negatively charged hadrons, ought be small. This is borne out by the recent COMPASS data in SIDIS from a deuteron target [26]. Such a cancellation is consequent to the differing signs of \( \tilde{a}_u (p,n) \) and \( \tilde{a}_d (p,n) \) [21,38].

Quark or gluon orbital angular momentum?

We now wish to use the differing isospin structure of the \( L_q \) and \( L_\perp \) mechanisms to infer the relative size of these contributions to the total orbital angular momentum of the nucleon. The isospin structure of the \( L_q \) mechanism is distinctive, for a gluon in a nucleon Fock state will produce \( uu \) or \( dd \) pairs with equal weight—up to tiny differences driven by the \( u-d \) mass difference. Thus the \( u \)-quark and \( d \)-quark SSAs add constructively in SIDIS from a \( ^2 \)H target; there is no cancellation of this \( 1 \)-dependence. However, the \( L_\perp \) mechanism cannot always generate a leading hadron, i.e., a contribution which survives in the projectile-like hadron in SIDIS at high \( \eta \) and \( \theta \) values. As noted in Refs. [31,34,35], both predictions are consistent with HERMES data at sufficiently large \( \eta \) [14].

As to the magnitudes of the SSAs, although \( |a^p_u| < |a^p_d|, \) the SSA engendered by \( u \)-quark fragmentation from a proton is enhanced by a factor of 4, since the asymmetry is controlled by \( e^+ \). In the absence of anti-quark contributions, the SSA for \( \pi^- \) production should indeed be small, and this is observed [14]. Recent model fits [21,38] are consistent with these trends. To be more specific, we note the fits of Ref. [38] yield \( S_u = -0.81 \pm 0.07 \) and \( S_d = 1.86 \pm 0.28, \) where \( S_p \) is to be compared to \( -\langle \tilde{a}_q \rangle, \) with \( \langle \tilde{a}_q \rangle \) defined to be the average value of \( \tilde{a}_q (q, \underline{F}_1). \) This implies that the SSA asymmetry from \( u \)-quark fragmentation is smaller than that that we have estimated from the use of the empirical anomalous magnetic moments alone. Although we expect the strong phase from the Wilson line to be isoscalar, the extra factor of 1/(1 \( -x \jmath \) ) in the matrix element of Eq. (10) should change the relative strength of the \( u \) and \( d \) contributions. Nevertheless, considering the consequences of this simple picture for the deuteron data, we note that \( a^u_d + a^u_u + a^d_d = -0.360, \) implying that the SSA for \( u \)-quark fragmentation to leading positively charged hadrons, as well as \( d \)-quark fragmentation to leading negatively charged hadrons, ought be small. This is borne out by the recent COMPASS data in SIDIS from a deuteron target [26]. Such a cancellation is consequent to the differing signs of \( \tilde{a}_u (p,n) \) and \( \tilde{a}_d (p,n) \) [21,38].

---

2 Preliminary SSA data in \( K^+ \) production on the proton from the HERMES Collaboration is much larger than that for \( \pi^+ \) production. This suggests, however, that the role of anti-quarks may be larger than assumed. N. Makins, talk at the NNPS, July, 2006, http://www.physics.indiana.edu/~nnpss/Makins.pdf.

3 In this study [26] the COMPASS Collaboration defines the leading hadron as the most energetic hadron with \( \eta > 0.25 \) which originates from the reaction vertex.
experimentally, one needs to work at large $z$ where the jet tagging is reliable, i.e., where the hadron type tags the flavor of the “struck” quark.

We have argued that the $L_g$ mechanism is small and cannot always produce a leading hadron, so that one is left to ponder how current empirical constraints can be bettered. It strikes us as efficacious to study SSAs associated with produced hadrons of non-valence quark content. The $\gamma^* g \rightarrow s\bar{s} \rightarrow K^- K^+ + X$ reaction is one such possibility. In principle, one can trace the SSA of the $K^- K^+$ to the gluon’s orbital angular momentum $L_g$. One can also consider the $\gamma^* g \rightarrow s\bar{s} \rightarrow \phi + X$ reaction: The SSA in $\phi$ production. Both reactions are important tests for the $L_g$ mechanism, since the gluon contributions of the two nucleons to the SSA add. One can consider these processes as aspects of the gluon jet. In this, we ignore the possibility of intrinsic strangeness in the nucleon’s non-perturbative structure, since parity-violating electron scattering experiments show the strangeness contribution to the proton’s anomalous magnetic moment to be small [43]. We note that Anselmino et al. have discussed accessing the Sivers gluon distribution through open charm production [25]; this is similar in conception to the suggestions we offer here.

The Sivers asymmetry from gluons can also be studied directly, if the empirical thrust [19] of the gluon jet axis can be determined. If this could be done, the study of the $L_g$ mechanism would be much facilitated, as the Collins mechanism would no longer contribute. To extract detailed numerical information about the gluon mechanism from the SSAs, one ultimately requires information about the strong phase from FSI; nevertheless, the experiments we suggest do serve to bound the size of the gluon’s orbital angular momentum contribution to the nucleon’s spin.

Let us conclude with a brief summary.

- A non-zero SSA follows from the interference of two amplitudes of differing nucleon spin, but of common quark helicity.
- The two amplitudes must also differ in their strong phase to generate a non-zero SSA, so that the matrix element which yields the Sivers function cannot be identical to that for the anomalous magnetic moment. Nevertheless, the signs of the SSAs from leading $u$- and $d$-quark fragmentation from the proton correlate with those of the $u$ and $d$-quark contributions to the anomalous magnetic moment, analyzed under an assumption of isospin symmetry.
- The anomalous magnetic moment, the Sivers function, and the generalized parton distribution $E$ can all be connected to matrix elements involving the orbital angular momentum of the nucleon’s constituents. To be specific, the matrix elements are between LFWFs that differ by one unit of orbital angular momentum along the polarization axis; this follows as we compute matrix elements of the angular momentum lowering or raising operator.
- The SSA can be generated by either a quark or gluon mechanism, and the isospin structure of the two mechanisms is distinct. The approximate cancellation of the SSA measured on a deuterium target suggests that the gluon mechanism, and thus the orbital angular momentums carried by gluons in the nucleon, is small.

Note added

After the completion of our paper, we became aware of work contemporary to ours by Anselmino et al., Ref. [44], which draws similar conclusions from the measured SSA in $p^\uparrow p \rightarrow \pi^0 X$ scattering.

Acknowledgements

S.G. thanks Wolfgang Korsch for bringing the single spin asymmetry data on $^2$H to her attention and the SLAC theory group for gracious hospitality.

References

On the strong energy dependence of the $e^+e^- \leftrightarrow p\bar{p}$ amplitude near threshold

J. Haidenbauer a, H.-W. Hammer b,*, Ulf-G. Meißner a, b, A. Sibirtsev b

a Institut für Kernphysik (Theorie), Forschungszentrum Jülich, D-52425 Jülich, Germany
b Helmholtz-Institut für Strahlen- und Kernphysik (Theorie), Universität Bonn, Nußallee 14-16, D-53115 Bonn, Germany

Received 14 June 2006; received in revised form 7 September 2006; accepted 3 October 2006
Available online 19 October 2006
Editor: J.-P. Blaizot

Abstract

We study the energy dependence of the $e^+e^- \rightarrow p\bar{p}$ cross section close to the two-nucleon threshold, recently reported by the BaBar Collaboration. Our analysis also includes the $p\bar{p} \rightarrow e^+e^-$ data collected by PS170 Collaboration and the $e^+e^- \rightarrow N\bar{N}$ data from the FENICE Collaboration. We show that the near-threshold enhancement in the $e^+e^- \rightarrow p\bar{p}$ cross section can be explained by the final-state interaction between proton and antiproton in the $^3S_1$ partial wave, utilizing the Jülich nucleon–antinucleon model. As a consequence, the strong dependence of the proton electromagnetic form factors on the momentum transfer close to the two-nucleon threshold is then likewise driven by this final-state interaction effect. This result is in line with our previous studies of the near-threshold enhancement of the $p\bar{p}$ invariant mass spectrum seen in the $J/\Psi \rightarrow \gamma p\bar{p}$ decay by the BES Collaboration and in the $B^+ \rightarrow p\bar{p}K^+$ decay by the BaBar Collaboration.

© 2006 Elsevier B.V. All rights reserved.


The observation of a steep energy dependence of the proton electromagnetic form factors (EMFF) in the time-like region at momentum transfers $q^2 \approx (2m_p)^2$, where $m_p$ is the proton mass, was first reported by the PS170 Collaboration [1], based on a measurement of the $p\bar{p} \rightarrow e^+e^-$ reaction cross section close to the $p\bar{p}$ threshold at LEAR. Later the FENICE Collaboration at Frascati measured the cross section for the time-reversed process $e^-e^- \rightarrow p\bar{p}$ [2,3]. However, their data were taken at energies not close enough to the threshold in order to confirm this strong energy dependence and, furthermore, had very large uncertainties. The FENICE Collaboration also made the first and only measurement of the $e^+e^- \rightarrow n\bar{n}$ cross section [3] which turned out, within the large experimental errors, to be close to the $e^+e^- \rightarrow p\bar{p}$ one. Only recently the BaBar Collaboration reported very precise data on the $e^+e^- \rightarrow p\bar{p}$ cross section down to energies very close to the $p\bar{p}$ threshold [4]. The form factor deduced from those data substantiates the finding of the PS170 Collaboration.

A steep dependence of the proton EMFF on the momentum transfer simply reflects the fact that the underlying (measured) $e^+e^- \rightarrow p\bar{p}$ cross section shows a significant enhancement near the $p\bar{p}$ threshold. It is interesting that a near-threshold enhancement was also reported recently in an entirely different reaction involving the $p\bar{p}$ system, namely the radiative decay $J/\Psi \rightarrow \gamma p\bar{p}$ [5]. For the latter case several explanations have been put forth, including scenarios that invoke $N\bar{N}$ bound states or so far unobserved meson resonances. However, it was also shown that a rather conventional but plausible interpretation of the data can be given in terms of the final-state interaction (FSI) between the produced proton and antiproton [6–10]. Specifically, in the calculation of our group [6] utilizing the Jülich $N\bar{N}$ model [11,12], the mass dependence of the $p\bar{p}$ spectrum close to the threshold could be nicely reproduced by the $S$-wave $p\bar{p}$ FSI in the isospin $I = 1$ state within the Watson–Migdal [13] approach.

* Corresponding author.
E-mail address: hammer@itkp.uni-bonn.de (H.-W. Hammer).
The success of those investigations suggests that the same effects, namely the FSI between proton and antiproton, could be also responsible for the near-threshold enhancement in the $e^+e^- \rightarrow p\bar{p}$ cross section and, accordingly, for the strong momentum-transfer dependence of the proton EMFF in the time-like region near $q^2 \approx (2m_p)^2$. In the present Letter, we report results of a corresponding calculation, utilizing again the scattering amplitudes predicted by the Jülich $N\bar{N}$ model A(OBE) [11] for the $S_1^1$ partial wave, multiplied by appropriate phase-space factors.

The effective reaction for the extraction [14] of the proton EMFF, the strong energy dependence of the two-body phase space. The solid line is the scattering amplitude squared predicted by the Jülich $N\bar{N}$ model A(OBE) [11] for the $S_1^1$ partial wave.

\[ \sigma(e^+e^- \rightarrow p\bar{p}) \simeq \left[ 1 - \frac{4m^2_p}{M^2(p\bar{p})} \right] \sigma(p\bar{p} \rightarrow e^+e^-), \] (1)

where we neglect the electron mass. Although there seems to be a systematical difference between the $e^+e^- \rightarrow p\bar{p}$ and $p\bar{p} \rightarrow e^+e^-$ cross section data, the latter are by a factor of about 1.3 smaller, their energy dependence is very similar. The dashed line in Fig. 1 shows the energy dependence due to the two-body phase space given by

\[ \sigma(e^+e^- \rightarrow p\bar{p}) = \frac{|A|^2}{16\pi M^2(p\bar{p})} \left[ 1 - \frac{4m^2_p}{M^2(p\bar{p})} \right]^{1/2}, \] (2)

where the constant Lorenz invariant amplitude $A$ was normalized to the data at the excess energy of 136 MeV, $|A|^2 = 46\text{ MeV}^2\text{fm}^2$. The experimental results clearly exhibit an energy dependence that differs from the phase space especially at excess energies below 50 MeV. This implies that the transition amplitude $A$ must vary substantially for energies close to the $p\bar{p}$ threshold.

To illustrate this conjecture more transparently we extract the squared invariant amplitude $|A|^2$ from the near-threshold data [1,4] by dividing out the phase space factor according to Eq. (2). The corresponding results are shown in Fig. 2. They clearly indicate that the squared transition amplitude depends rather strongly on the energy within the range $M(p\bar{p}) - 2m_p \leq 50\text{ MeV}$, say.

Since the $e^+e^- \rightarrow p\bar{p}$ and $p\bar{p} \rightarrow e^+e^-$ data are used for the extraction [14] of the proton EMFF, the strong energy dependence of the transition amplitude is reflected in the behaviour of the EMFF in the time-like region close to threshold. Phenomenological models such as vector dominance model (VDM), which assumes that the photon couples to hadrons through intermediate vector mesons [15,16], fail to describe that steep energy dependence. To resolve this discrepancy the VDM was extended to include also heavier vector mesons [16,17] besides the light $\rho$, $\omega$ and $\phi$ mesons. Taking the couplings of the heavy vector mesons to the proton as free parameters it was possible to reproduce the steep dependence of the $p\bar{p} \rightarrow e^+e^-$ cross section close to $p\bar{p}$ threshold. For a discussion of this issue in the context of dispersion relations, see [18,19].

On the other hand, the success of $p\bar{p}$ FSI effects in explaining the near-threshold enhancement in the $p\bar{p}$ mass spectrum of $J/\psi \rightarrow \gamma p\bar{p}$ suggests that the same mechanisms could be also responsible for the behaviour of the EMFF. Indeed FSI effects have been already considered before [20,21] to describe the near-threshold energy dependence of the $p\bar{p} \rightarrow e^+e^-$ reaction by the $p\bar{p}$ initial-state-interaction, though at a time when
only the less accurate LEAR data were available. Based on the usual assumption that one-photon exchange constitutes the main reaction mechanism the reaction can only proceed from the $J^{PC} = 1^{--}$ state. Then $\bar{p}p \rightarrow e^+e^-$ as well as the time-reversed reaction $e^-e^+ \rightarrow \bar{p}p$ can only involve a single partial wave, namely the coupled $^3S_1 - ^3D_1 \bar{p}p$ state. Obviously, close to the $\bar{p}p$ threshold the reaction amplitude will be dominated by the $^3S_1$ component. Invoking the Watson–Migdal prescription for the treatment of final-state effects [13] and using the scattering length approximation with keeping only the term linear in the antiproton momentum in the center-of-mass system, the squared transition amplitude should behave like

$$|A|^2 \approx N/(1 - \text{Im} a \sqrt{M^2(p\bar{p}) - 4m_p^2}),$$

(3)

where $\text{Im} a$ is the imaginary part of the $^3S_1$ scattering length and $N$ a normalization constant.

Eq. (3) has the advantage that one can obtain a rough but model-independent estimate of the FSI effects by utilizing available experimental values for the $p\bar{p}$ scattering lengths extracted from $1s$ level shifts and widths of antiprotonic hydrogen atoms [28,29]. The most recently published value for the imaginary part of the pure strong-interaction spin-averaged scattering length is $\text{Im} a = (0.03 \pm 0.03)$ fm [29]. The corresponding result, the dashed line in Fig. 2, is in line with the trend shown by the data and, therefore, definitely an indication that FSI effects might be responsible for the near-threshold enhancement in the $\bar{p}p \rightarrow e^+e^-$ amplitude. One should say, however, that there are uncertainties in using the experimental $\text{Im} a$ since the value extracted from $p\bar{p}$ atoms is, in fact, an average of the $^3S_1$ and $^1S_0$ states and not the one corresponding to the $^3S_1$ alone. Moreover, only data extremely close to the threshold are expected to be in line with Eq. (3), i.e. to exhibit a linear dependence on the antiproton momentum $k$ in the center-of-mass system. To include higher orders $\sim k^2$ would require the real part of the $p\bar{p}$ scattering length, but also the (complex) effective range which is not known experimentally.

Therefore, in our analysis we use explicitly the full $^3S_1, p\bar{p}$ amplitude of the Jülich $N\bar{N}$ model A(OBE) [11]. This model is constrained by the available data on $N\bar{N}$ interactions and it will be interesting to see whether it can reproduce the strong energy dependence of $|A|^2$. The purely nuclear $p\bar{p}$ scattering length predicted by this model for the $^3S_1$ partial wave is $a = (0.96 - 1.08)$ fm. The value for the imaginary part is in reasonable agreement with the experimental information, cited above, considering the fact that the latter is actually a spin-averaged result. As already mentioned above, in a previous study [6] we have demonstrated that the near-threshold enhancement in the $p\bar{p}$ invariant mass spectrum from the $J/\psi \rightarrow \gamma p\bar{p}$ decay observed by the BES Collaboration [5] is presumably due to the FSI between the outgoing proton and antiproton, utilizing this $N\bar{N}$ model. Similar conclusions on the origin of the near-threshold enhancement in the $p\bar{p}$ mass spectrum were drawn by other groups, employing the Paris $N\bar{N}$ model [7] but also within the effective range approximation [8–10].

The solid line in Fig. 2 is the $p\bar{p}$ isospin-averaged scattering amplitude squared predicted by the $N\bar{N}$ model A(OBE) [11] for the $^3S_1$ partial wave. It is normalized to the low-energy data in order to facilitate the comparison with the $e^+e^- \rightarrow p\bar{p}$ amplitude. The same result is also shown in Fig. 1, multiplied by appropriate phase-space factors, cf. Eq. (2), in order to enable a comparison with the $e^+e^- \rightarrow p\bar{p}$ cross section. It is obvious that the energy dependence of the $e^+e^- \rightarrow p\bar{p}$ transition amplitude squared for energies $E(p\bar{p}) - 2m_p < 50$ MeV is indeed rather similar to that of the $N\bar{N}$ scattering amplitude. This result strongly suggests that, like for $J/\psi \rightarrow \gamma p\bar{p}$, the FSI in the $p\bar{p}$ system is predominantly responsible for the near-threshold enhancement observed in the $e^+e^- \rightarrow p\bar{p}$ cross section. The same argument immediately applies to the EM form factors of the nucleon. It is useful to consider the effective form factor $F_p$ [4], since most experiments are analyzed under the assumption $G_E = |G_M| = |F_p|$ due to the lack of precise angular distributions. This assumption is exactly satisfied at threshold and approximately valid in the threshold region. The effective form factor squared $|F_p|^2$ in the time-like region is proportional to the squared transition amplitude $|A|^2$ shown in Fig. 2. The relation near threshold is given by

$$|F_p|^2 = \frac{3M^2(p\bar{p})}{64\pi^2\alpha^2 C(M^2(p\bar{p}) + 2m_p^2)}|A|^2,$$

(4)

where we have neglected the electron mass [4]. Moreover, $\alpha$ denotes the fine structure constant and $C$ is the Coulomb correction factor. As a consequence, the strong dependence of the proton EM form factor on the momentum transfer near $q^2 \approx (2m_p)^2$ must be driven by the same FSI effect in the $p\bar{p}$ system.

We want to mention that we also performed analogous calculations utilizing other $N\bar{N}$ models of the Jülich group, specifically the potentials A(BOX) and D, which are described in Refs. [11,12]. In all these cases the obtained results were rather similar to the ones for the model A(OBE) and, therefore, we refrain from showing them here. Note that the disagreement with the experiment at higher excess energies is not a reason of concern and, in particular, does not discredit the interpretation of the data in terms of FSI effects. We have omitted the contribution from the $^3S_1D_1$ state in our calculation, which is negligible in the near-threshold region. However, at energies around $M(p\bar{p}) - 2m_p \approx 100–150$ MeV its contribution is presumably no longer small and, therefore, most likely responsible for the underestimation of the experimental cross section by our model analysis in this energy range.

In summary, we have analyzed the energy dependence of the squared transition amplitudes for the $\bar{p}p \rightarrow e^+e^-$ [1] and $e^+e^- \rightarrow p\bar{p}$ [4] reactions utilizing the Jülich $N\bar{N}$ model [11,12]. Our investigation demonstrates that the strong energy dependence of the $e^+e^- \rightarrow p\bar{p}$ cross section is driven by the initial or final-state-interaction in the $^3S_1$ partial wave of the $p\bar{p}$ system. This explanation is in line with our previous stud-
ies [6,30] of the near-threshold enhancement in the \( p\bar{p} \) invariant mass spectrum from the \( J/\Psi \rightarrow \gamma p\bar{p} \) decay observed by the BES Collaboration [5] and the \( B^+ \rightarrow p\bar{p}K^+ \) decay reported by the BaBar Collaboration [31]. As a consequence, the steep dependence of the proton electromagnetic form factor on the momentum transfer in the time-like region near \( q^2 \approx (2m_p)^2 \) is likewise a reflection of this initial or final-state-interaction effect. This leaves not much room for other non-standard dynamics in the time-like EMFF close to threshold, such as the narrow resonance scenario put forth in Refs. [3,32]. Nevertheless, the dynamics of the EMFF in the time-like region is far from well understood and many important problems, such as the asymptotic ratio of the space-like and time-like form factors or the reliable separation of electric and magnetic form factors, remain.

**Note added**

After submission of our paper, a similar study by Dmitriev and Milstein appeared which confirmed our results using the Paris \( N\bar{N} \) potential [33].

**Acknowledgements**

This work was partially supported by the DFG (SFB/TR 16, “Subnuclear Structure of Matter”) and by the EU Integrated Infrastructure Initiative Hadron Physics Project under contract No. RII3-CT-2004-506078. A.S. acknowledges support by the JLab grant SURA-06-C0452 and the COSY FFE grant No. 41760632 (COSY-085).

**References**

    A.B. Migdal, JETP 1 (1955) 2.
**X(1576) as diquark–antidiquark bound state**

Gui-Jun Ding *, Mu-Lin Yan

Interdisciplinary Center for Theoretical Study, University of Science and Technology of China, Hefei, Anhui 230026, China

Received 27 July 2006; received in revised form 23 September 2006; accepted 11 October 2006

Available online 26 October 2006

© 2006 Elsevier B.V. All rights reserved.

**Abstract**

We propose that the broad $1^{-+}$ resonance structure recently discovered by BES in $J/\psi \to K^+ K^- \pi^0$ is the P-wave excitation of a diquark–antidiquark bound state. This interpretation implies that there exists a negative parity, vector nonet. A crucial prediction is that $\Delta F=0$. The observation of $I_3=1$ or $I_3=-1$ states which predominantly decays into strange mesons could provide another important test to our proposal. To search the charged $1^{-+}$ states which predominantly decays into strange mesons could provide another important test to our proposal. To search the charged $1^{-+}$ states which predominantly decays into strange mesons could provide another important test to our proposal. To search the charged $1^{-+}$ states which predominantly decays into strange mesons could provide another important test to our proposal.

**1. Introduction**

A broad $1^{-+}$ resonant structure $X(1576)$ in $J/\psi \to K^+ K^- \pi^0$ has been reported by the BES Collaboration recently [1]. Its pole position is determined to be $(1576^{+49+98}_{-35-91})$ MeV $- i(409^{+11+32}_{-12-67})$ MeV, and the product branching ratio $\text{Br}(J/\psi \to X(1576)\pi^0)\text{Br}(X(1576)\to K^+ K^-) = (8.5 \pm 0.6) \times 10^{-4}$. Therefore the branching fraction of $J/\psi \to X(1576)\pi^0$ should be larger than $O(10^{-4})$. Considering the branching ratio of the $J/\psi$ electromagnetic decay is usually of the order $O(10^{-4})$, so we suggested that the decay $J/\psi \to X(1576)\pi^0$ is mainly hadronic, where both isospin and $G$-parity are conserved. Then $X(1576)$ is of even $G$-parity and its isospin $I = 1$. In other words, the assignment of $I = 1$ for $X(1576)$ is a rather reasonable assumption. Consequently, the quantum numbers of this structure are $I^G(J^{PC}) = 1^+(1^{-+})$ [2,3]. There is no obvious standard $q\bar{q}$ candidate for this state.

Since the decay products $K^+ K^-$ contain a pair of strange quark, it may contain a pair of hidden strange quark, and the isospin triplet nature of this resonance requires that at least contains additionally a pair of nonstrange quark, so it is reasonable to expect that $X(1576)$ is a diquark–antidiquark bound state. The combined effects of the negative parity and the total angular momentum $J = 1$ require a unit of orbital angular momentum excitation. Thus we are lead to the following assumption:

$$X(1576) = \frac{1}{\sqrt{2}} \left(\langle |s\bar{u}| \bar{s} \bar{d}\rangle_{p\text{-wave}} - \langle |s\bar{u}| \bar{s} \bar{d}\rangle_{p\text{-wave}}\right).$$  \hspace{1cm} (1)

In Ref. [4] Maiani et al. pointed out that the exotic states $X(3872)$ and $X(3940)$ can be well explained if they are S-wave diquark–antidiquark bound state $\langle |c\bar{q}| \bar{c} \bar{q}\rangle_{s\text{-wave}}$. Furthermore, they proposed that the new state $Y(4260)$ may be the first orbital excitation of a diquark–antidiquark bound state [5], $Y(4260) = \langle |c\bar{q}| \bar{c} \bar{q}\rangle_{p\text{-wave}}$. If these are really what happens in nature, it is

---

* Corresponding author.

E-mail addresses: dinggj@mail.ustc.edu.cn (G.-J. Ding), mlyan@ustc.edu.cn (M.-L. Yan).

0370-2693/ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.physletb.2006.10.022
reasonable to expect the P-wave excitation of the four quark state \((|q_1q_2\rangle[\bar{q}_3\bar{q}_4]\rangle)_{\text{P-wave}}\) \((q_i\) is light quark with \(i = 1–4)\) should be seen experimentally, i.e., the P-wave excitation of the nonet of light scalar \((J^{PC} = 0^{++})\) mesons \(a_0(600), f_0(980), a(980), \kappa(800)\).

In our scheme, \(X\) (1576) is exactly the P-wave excitation of \(a_0(980)\), and there exists analogously an nonet of vector mesons with \(J^{PC} = 1^{-+}\). Henceforth, this nonet is denoted by \(X\). In this Letter, we would like to give a rough mass estimate of these states, and the prediction about the mass of \(X\) (1576) is consistent with its experimental value. The decay properties of these states are discussed, which can decay into two pseudoscalars or one pseudoscalar plus one vector meson, and some distinctive predictions are given.

2. Mass spectrum of the vector nonet with \(J^{PC} = 1^{-+}\)

The weight diagram for the nonet is shown in Fig. 1, and we define \([q_1q_2] \equiv \frac{1}{2}(q_1q_2 - q_2q_1)\), then the composition of the states of the nonet is as followings:

\[
X^+_a = ([su][\bar{d}\bar{s}])_{\text{P-wave}}, \quad X^-_a = ([ds][\bar{s}\bar{u}])_{\text{P-wave}}, \quad X^0_a = \frac{1}{\sqrt{2}}\left(([ds][\bar{d}\bar{s}])_{\text{P-wave}} - ([su][\bar{s}\bar{u}])_{\text{P-wave}}\right),
\]

\[
X^+_\kappa = ([ud][\bar{d}\bar{s}])_{\text{P-wave}}, \quad X^-_\kappa = ([ds][\bar{u}\bar{d}])_{\text{P-wave}}, \quad X^0_\kappa = \frac{1}{\sqrt{2}}\left(([ds][\bar{d}\bar{s}])_{\text{P-wave}} + ([su][\bar{s}\bar{u}])_{\text{P-wave}}\right),
\]

\[
X^0_\kappa = ([ud][\bar{d}\bar{s}])_{\text{P-wave}}, \quad X^0_\sigma = ([ud][\bar{u}\bar{d}])_{\text{P-wave}},
\]

where for the two isosinglets, the states with definite strange quark pair are introduced by assuming ideal mixing. The physical states \(X_f\) and \(X_\sigma\) are mixing of \(X^0_f\) and \(X^0_\sigma\) with mixing angle \(\theta\),

\[
X_f = \cos \theta X^0_f + \sin \theta X^0_\sigma, \quad X_\sigma = -\sin \theta X^0_f + \cos \theta X^0_\sigma.
\]

We will assume that the quarks prefer to form the “good” diquark when possible. States dominated by that configuration should be systematically lighter, more stable, and therefore more prominent than the states formed from other types of diquarks. The residual QCD interaction and the spin–orbit interaction will mix the \(S = 0\) “good” diquark with \(S = 1\) “bad” diquark (“good” and “bad” diquarks in Jaffe’s terminology [7]), and a more sophisticated treatment would have to consider these effects quantitatively. However the effects only give a second order correction to the mass and other properties, so we restrict to the “good” diquark in this first analysis.

Most quark model treatments of multiquark spectroscopy use the colormagnetic short range hyperfine interaction as the dominate mechanism for possible binding [8–10]. Here we follow the same procedure, and the colormagnetic hyperfine interaction is:

\[
H' = -\sum_{i>j} C_{ij} \vec{\sigma}_i \cdot \vec{\lambda}_i \vec{\sigma}_j \cdot \vec{\lambda}_j.
\]

Here \(\vec{\sigma}\) and \(\vec{\lambda}\) are the Pauli and Gell-Mann matrices, \(i\) and \(j\) run over the constituent quarks and antiquarks. The coefficient \(C_{ij}\) are dependent on the quark masses and properties of the spatial wave functions of the quarks and antiquarks in the system. In the
SU(3) flavor symmetry limit, $C_{ij} \equiv C$, and the standard treatment using the color-spin $SU(6)_{cs}$ algebra gives the hyperfine energy contribution [12,13]:

$$E' = \frac{C}{2} \left[ D(\text{tot}) - 2D(Q) - 2D(\bar{Q}) + 16N \right],$$  \hspace{1cm} (5)$$

where $D = C_6 - C_3 - \frac{8}{3}S(S + 1)$, and $D(\text{tot})$, $D(Q)$, $D(\bar{Q})$ denote the $D$ of the total system, the subsystem of the quarks and the antiquarks respectively. $C_6$ and $C_3$ are the quadratic Casimir operators of $SU(6)_{cs}$ and $SU(3)$, respectively, $S$ is the spin and $N$ is the total number of the quarks and antiquarks. Rich phenomenology based on the colormagnetic hyperfine interaction have been developed [10–13], and a fit of charmed baryons gives the consistent quark mass:

$$m_q \approx m_d \approx 360 \text{ MeV}, \quad m_s \approx 540 \text{ MeV}, \quad m_c \approx 1710 \text{ MeV}$$  \hspace{1cm} (6)$$

and the strength factors

$$C_{qq} = 20 \text{ MeV}, \quad C_{qs} = 12.5 \text{ MeV}, \quad C_{ss} = 10 \text{ MeV}.$$  \hspace{1cm} (7)$$

Because the diquark and antidiquark are in P-wave and are separated by a distance larger than the range of the colormagnetic force, the color hyperfine interaction operates only within the diquark (antidiquark), but is not felt between the clusters.

There are three contributions to the mass of the states, i.e., the masses of the constituent quarks, the colormagnetic hyperfine interaction energy, and the energy due to the P-wave excitation. We estimate the contribution of the constituent quark mass from

$$m_{K}, A, B$$  \hspace{1cm} (8)$$

where the second term is the spin–orbit interaction and the third term is the mass contribution of the orbital angular momentum. Then we find

$$B = \frac{m_{K_1} + m_{K_2} - 2m_{K^*}}{2} \approx 458 \text{ MeV}.$$  \hspace{1cm} (9)$$

Then the mass contribution of the P-wave excitation for this set of mesons equals $B$ which is approximately 458 MeV. The mass of the charm mesons $D^*(2007)^0$, $D_1(2420)^0$, $D_s^*(2460)^0$ can also be described by the formula Eq. (8) with different parameters $K$, $A$, $B$. In this case, the parameter $B$ is about 433 MeV, then the P-wave excitation energy for this set of charm mesons approximately is 433 MeV. From above, we can see that the P-wave excitation energy changes slowly with the meson mass variations within the range about $1 \to 2.5$ GeV. For simplicity, we approximately take the P-wave excitation energy of the nonet $X^*$ as follows:

(1) The mass of $X_d(I = 1)$ and $X^0_f$

$$m_{X_d} = m_{X^0_f} = 2(m_K + 16C_{qs}) + E_P - 8C_{qs} - 8C_{ss} = 2m_K + E_P + 16C_{qs} \approx 1632 \text{ MeV}.$$  \hspace{1cm} (10)$$

This prediction is consistent with the experiment data, the pole position of $X(1576)$ is: $(1576^{+49+98}_{-35-91}) \text{ MeV} - i(409^{+11+32}_{-12-67}) \text{ MeV}$. Because of the large decay width, it is very difficult to precisely determine the mass of this resonance by experiments.

(2) The mass of $X_k$ ($X^\pm_k, X^0_k$ and $\bar{X}^0_k$)

$$m_{X_k} = m_{X_a} - 8C_{qq} + 8C_{qs} + m_q - m_s = m_{X_a} - 240 \approx 1392 \text{ MeV}.$$  \hspace{1cm} (11)$$

If we use the experiment central value for the mass of $X_a$, the peak mass of $X_k$ is 1336 MeV.

(3) The mass of $X^0_{\sigma}$

$$m_{X^0_{\sigma}} = m_{X_a} - 16C_{qq} + 16C_{qs} + 2m_q - 2m_s = m_{X_a} - 480 \approx 1152 \text{ MeV}.$$  \hspace{1cm} (12)$$

If the experimental value for the mass of $X_a$ is input, the peak mass of $X^0_{\sigma}$ is 1096 MeV. The spectrum is similar to the that of the light scalar nonet, which is inverted with respect to the $q\bar{q}$ nonet.
3. The decay of the vector nonet $X$

The dominant decay mode of the four quark states is that they dissociate into two colorless $q\bar{q}$ mesons [5,13], which means that a quark–antiquark pair is switched between the diquark and antidiquark, then a pair of colorless $q\bar{q}$ states are formed. This mechanism has successfully described the decay of the scalar nonet [6], also has been used to discuss the decay of other four quark states, and the predictions for the decay width are close to the experiment [4–6]. The nonet $X$ can decay into two pseudoscalars or one pseudoscalar and one vector meson. In the exact SU(3) flavor limit, the decay amplitude can be described with a single parameter $g$, which describes the tunneling from the bound diquark–antidiquark pair configuration to the meson–meson pair. The parameters for the two pseudoscalars channel and one pseudoscalar and one vector meson channel should be different, are denoted as $g_1$ and $g_2$ respectively.

3.1. $X \rightarrow$ pseudoscalar + pseudoscalar

We can describe the decay process by a single switch amplitude, e.g., the decay of $X^+_a$

\[
[su]^3[\bar{s}\bar{d}]_c \rightarrow (s\bar{d})_1_c (u\bar{d})_1_c - (s\bar{d})_1_c (u\bar{s})_1_c,
\]

where the subscripts indicate color configuration. Taking into account the conservation of $C$-parity and $G$-parity, we can further write out the invariant three mesons effective coupling:

\[
ig_1X^+_a[\bar{K}^{-}\partial_\mu K^0 - K^0\partial_\mu K^{-}].
\]

Here the coupling constant $g_1$ is dimensionless, and we will introduce $\eta_q$ and $\eta_s$ in the following, which are defined by $\eta_q = \sqrt{\frac{2}{3}}\eta_1 + \frac{1}{\sqrt{3}}\eta_8, \eta_s = \frac{1}{\sqrt{3}}\eta_1 - \sqrt{\frac{2}{3}}\eta_8$. The physical states $\eta, \eta'$ are related to $\eta_8$ and $\eta_1$ via the usual mixing formula $\eta_8 = \eta \cos \theta_p - \eta' \sin \theta_p, \eta_1 = \eta \sin \theta_p + \eta' \cos \theta_p$ with the mixing angle $\theta_p = 16.9^0 \pm 1.7^0$ [14]. From the effective Lagrangian (14), we find the decay amplitude:

\[
\mathcal{M}(X^+_a \rightarrow K^+ \bar{K}^0) = g_1 \epsilon^\mu(X^+_a) \left[p_\mu (K^+-) - p_\mu (\bar{K}^0)\right].
\]

Here $\epsilon^\mu(X^+_a)$ is the polarization vector of $X^+_a$, $p_\mu (K^+)\ is the four momentum vector of $K^+$. The decay of the other member of the nonet can be investigated in the same way, and the effective Lagrangian for the relevant decays is as follows,

\[
\mathcal{L}_{\text{eff}} = ig_1 \left\{ X^+_a \left[ K^{-}\partial_\mu K^0 - K^0\partial_\mu K^{-} \right] + X^-_a \left[ -K^+\partial_\mu \bar{K}^0 + \bar{K}^0\partial_\mu K^+ \right] 
\right.
\]

\[
\left. + X^0\partial_\mu \pi^--\pi^-\partial_\mu \frac{1}{\sqrt{2}} K^- \partial_\mu (-\pi^0 + \eta_q) - \frac{1}{\sqrt{2}} (-\pi^0 + \eta_q) \partial_\mu K^- \right] 
\]

\[
+ X_\pi^+ \left[ -K^0 \partial_\mu \pi^+ + \pi^+ \partial_\mu K^0 \right] - \frac{1}{\sqrt{2}} K^0 \partial_\mu (-\pi^0 + \eta_q) + \frac{1}{\sqrt{2}} (-\pi^0 + \eta_q) \partial_\mu K^0 
\]

\[
+ X_\pi^0 \left[ -K^- \partial_\mu \pi^+ + \pi^+ \partial_\mu K^- \right] + \frac{1}{\sqrt{2}} (\pi^0 + \eta_q) \partial_\mu K^0 - \frac{1}{\sqrt{2}} K^0 \partial_\mu (\pi^0 + \eta_q) 
\]

\[
+ X_\pi^0 \left[ K^+ \partial_\mu \pi^- - \pi^- \partial_\mu K^+ \right] + \frac{1}{\sqrt{2}} (\pi^0 + \eta_q) \partial_\mu K^0 - \frac{1}{\sqrt{2}} K^0 \partial_\mu (\pi^0 + \eta_q) 
\]

\[
+ X_f \left[ K^\pm \partial_\mu K^\mp - K^- \partial_\mu K^+ + K^+ \partial_\mu K^- \right] - \frac{1}{\sqrt{2}} \left[ K^\pm \partial_\mu K^\mp - K^- \partial_\mu K^+ + K^+ \partial_\mu K^- \right] 
\]

\[
\left. + \frac{1}{\sqrt{2}} \left[ -K^+ \partial_\mu K^- + K^- \partial_\mu K^+ \right] \right].
\]

From the above Lagrangian, we can calculate the width of various decay channels following the standard procedure. The decay width is expressed as

\[
\Gamma(X \rightarrow P_1 + P_2) = \frac{g^2}{6\pi} C_{X \rightarrow P_1 P_2} \frac{2}{(1-\beta)\sqrt{2}\pi} \int_{m_X^\pm\delta}^{m_X^\pm\delta} d\bar{m} \frac{|\vec{p}|^3}{m^2} \exp \left[ -\frac{(m - m_X)^2}{2(\Gamma_X/2)^2} \right],
\]

where because of the large decay width, the mass distribution has been considered by using an exponential function [15]. $|\vec{p}|$ is the decay momentum $|\vec{p}| = \sqrt{(m^2 - (m_1 + m_2)^2)(m^2 - (m_1 - m_2)^2)}, \delta = 1.64\frac{\Gamma_X}{2}, \beta = 10\%$ [15], $m_1$ and $m_2$ are respectively the mass of two pseudoscalars $P_1$ and $P_2$, $C_{X \rightarrow P_1 P_2}$ is a numerical coefficient which can be found from the effective Lagrangian (16), and coefficient
However, since \(X_0 \to K^-K^0\) is the isospin partner of \(X_0\), it is unlikely to decay into \(\pi^+\pi^-\). So this prediction cannot distinguish the different models about \(X(1576)\). Generally the width of these resonances is very large, so it is likely that some members of the vector nonet disappear into the continuum and cannot be observed.

### Table 1

The numerical coefficient entering Eq. (17) for the decay of the nonet \(X\)

<table>
<thead>
<tr>
<th>Decay</th>
<th>(C_{X \to p_1 p_2})</th>
<th>Decay</th>
<th>(C_{X \to p_1 p_2})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X_0^+ \to K^+K^0)</td>
<td>1</td>
<td>(X_0^+ \to \pi^+K^0)</td>
<td>1</td>
</tr>
<tr>
<td>(X_1^+ \to K^+\pi^0)</td>
<td>(\frac{1}{2})</td>
<td>(X_1^+ \to K^+\eta)</td>
<td>(\frac{1}{2}(\sqrt{7} \sin \theta_p + \sqrt{2} \cos \theta_p)^2)</td>
</tr>
<tr>
<td>(X_2^+ \to K^+\eta')</td>
<td>(\frac{1}{2}(\sqrt{7} \cos \theta_p - \sqrt{2} \sin \theta_p)^2)</td>
<td>(X_2^0 \to \pi^-K^+)</td>
<td>1</td>
</tr>
<tr>
<td>(X_0^0 \to \pi^0K^0)</td>
<td>(\frac{1}{2})</td>
<td>(X_0^0 \to K^0\eta)</td>
<td>(\frac{1}{2}(\sqrt{7} \sin \theta_p + \sqrt{2} \cos \theta_p)^2)</td>
</tr>
<tr>
<td>(X_1^0 \to K^0\eta')</td>
<td>(\frac{1}{2}(\sqrt{7} \cos \theta_p - \sqrt{2} \sin \theta_p)^2)</td>
<td>(X_1^0 \to K^+K^0)</td>
<td>(\frac{1}{2})</td>
</tr>
<tr>
<td>(X_0^0 \to K_LKS)</td>
<td>(\frac{1}{2})</td>
<td>(X_1^0 \to K^+K^0)</td>
<td>(\frac{1}{2})</td>
</tr>
<tr>
<td>(X_0^0 \to K_LKS)</td>
<td>(\frac{1}{2})</td>
<td>(X_1^0 \to K^+K^0)</td>
<td>(\frac{1}{2})</td>
</tr>
</tbody>
</table>

\(C_{X \to p_1 p_2}\) for various decays are listed in Table 1. In this table, we have not shown the decay channels which can be obtained from the channels appearing in the table by making charge conjugation, e.g., for \(X_0^0 \to K^-K^0\), the corresponding numerical coefficient is 1. From Table 1, we can see that the dominant decay modes of \(X^0_0(X(1576))\) are \(K^+K^-\) and \(K_LKS\), and

\[
\frac{\Gamma(X^0_0(X(1576)) \to K^+K^-)}{\Gamma(X^0_0(X(1576)) \to K_LKS)} \approx 1. \tag{18}
\]

However, \(X^0_0(X(1576))\) cannot decay into \(\pi^+\pi^-\),

\[
\Gamma(X^0_0(X(1576)) \to \pi^+\pi^-) \approx 0. \tag{19}
\]

Dominant \(K^+K^-\) and \(K_LKS\) decays is a distinctive signature of the validity of the present model. Some interesting relations can be found, such as:

\[
\Gamma(X^+_0 \to K^+K_L) \approx \Gamma(X^+_0 \to K^+K_S) \approx \Gamma(X^0_0 \to K^+K^-) \approx \Gamma(X^0_0 \to K_LKS).
\]

\[
\Gamma(X^+_0 \to K^-K^-) \approx \Gamma(X^+_0 \to K_LKS) \approx \Gamma(X^0_0 \to K^+K^-),
\]

\[
\Gamma(X^+_0 \to K^+\pi^0) \approx 2\Gamma(X^+_0 \to K^+\eta) \approx \Gamma(X^+_0 \to K^+\eta') \approx 2\Gamma(X^+_0 \to K^0\pi^0),
\]

\[
\Gamma(X^+_0 \to K^-\eta) \approx \Gamma(X^0_0 \to K^0\eta), \Gamma(X^+_0 \to K^-\eta') \approx \Gamma(X^+_0 \to K^0\eta'),
\]

\[
\Gamma(X^+_0 \to \pi^-K^+) \approx \Gamma(X^+_0 \to \pi^+K^0) \approx \Gamma(X^+_0 \to K^+\eta) + \Gamma(X^+_0 \to K^+\eta'),
\]

\[
\Gamma(X^0_0 \to \pi^-K^+) \approx \Gamma(X^0_0 \to \pi^+K^0) + \Gamma(X^0_0 \to K^0\eta) + \Gamma(X^0_0 \to K^0\eta'), \tag{20}
\]

where \(\tilde{\Gamma}\) denotes the decay width neglecting phase space correction (i.e., ignoring the effect of the factor \(|\tilde{p}|^3\) in Eq. (17)). It can be easily checked that the first four equations are consistent with the isospin symmetry, and the last two equations in Eq. (20) express the flavor cross symmetry [3]. The effective Lagrangian (16) describes the decays allowed by the OZI rule, and the contributions of the other couplings which violate the OZI rule are neglectable in the first order. Since \(X^+_0\) and \(X^0_0\) form an isospin triplet, the pole position of these states should be approximately equal. Under this approximation and using Eq. (17), we can further obtain the following ratio:

\[
\Gamma(X^+_0 \to K^+K_L) : \Gamma(X^+_0 \to K^+K_S) \approx 1:1. \tag{21}
\]

We can search the other members of the nonet \(X\) in \(J/\psi\) decay, e.g., we can search \(X^+_1\) which is the \(I_3 = 1\) isospin partner of \(X^0_0(X(1576))\) in \(J/\psi \to X^+_1\pi^- \to K^+K_L\pi^-\) or \(J/\psi \to X^+_1\pi^- \to K^+K_S\pi^-\). However, since \(X^0_0(X(1576)) \to K^+K^-\) has been observed, this prediction is naturally the outcome of isospin conservation, and any rational proposal about the nature of \(X(1576)\) should produce this result. So this prediction cannot distinguish the different models about \(X(1576)\), and we should search some particular signals which are almost unique in our model. With this idea in mind, we will investigate another strong decay mode \(X \to \text{pseudoscalar} + \text{vector}\). Generally the width of these resonances is very large, so it is likely that some members of the vector nonet disappear into the continuum and cannot be observed.

### 3.2. \(X \to \text{pseudoscalar} + \text{vector}\)

The OZI allowed decays can be described by the effective Lagrangian:

\[
\mathcal{L}_{\text{eff}} = g_2 \varepsilon^{\mu
u\alpha\beta} \left[ (X^+_0)^\mu \left[ \rho_{\alpha\beta} \eta_{\nu} + \phi_{\alpha\beta} \pi^- - K_{\alpha\beta}^{a0} K^0 - K_{\alpha\beta}^{s0} K^+ \right] 
+ (X^-_0)^\mu \left[ \rho_{\alpha\beta} \eta_{\nu} + \phi_{\alpha\beta} \pi^+ - K_{\alpha\beta}^{s+} \tilde{K}^0 - K_{\alpha\beta}^{a+} K^+ \right] \right],
\]

where \(g_2\) is the coupling constant.
\[ \Gamma(X \rightarrow P + V) = \frac{4g_\rho^2}{3\pi} D_{X \rightarrow PV} \frac{2}{(1 - \beta)2\pi F_X m_{X + \delta}} \int dm |\vec{p}|^3 \exp \left[ -\frac{(m - m_X)^2}{2(F_X/2)^2} \right]. \] 

Here \(|\vec{p}|\) is the momentum of the vector meson \(V\) or that of the pseudoscalar \(P, |\vec{p}| = \sqrt{(m^2 - (m_p + m_V)^2)(m^2 - (m_p - m_V)^2), \delta = 1.64 F_X, \beta = 10\% [15]. m_p, m_V are respectively the masses of the pseudoscalar \(P\) and the vector meson \(V. \) Being similar to \(C_{X \rightarrow \rho_1 \rho_1}, D_{X \rightarrow PV}\) is also a numerical coefficient, which can be read from the Lagrangian (22), and \(D_{X \rightarrow PV}\) for various decay channels are listed in Table 2.

**Table 2**

<table>
<thead>
<tr>
<th>Decay</th>
<th>(D_{X \rightarrow PV}) entering Eq. (23) for the decay of the nonet (X)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X_\rho^+ \rightarrow K^{*+} K^0)</td>
<td>(X_\rho^+ \rightarrow K^{*+} K^0)</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow \rho^+ \eta)</td>
<td>((\sqrt{2} \sin \theta_\rho - \sqrt{2} \cos \theta_\rho)^2)</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow \phi \pi^+)</td>
<td>(X_\rho^+ \rightarrow \phi \pi^+)</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow K^{0(\pi^0)})</td>
<td>(X_\rho^+ \rightarrow K^{0(\pi^0)})</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow K^{*+} \eta)</td>
<td>(X_\rho^+ \rightarrow K^{*+} \eta)</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow \rho^0 K^+)</td>
<td>(X_\rho^+ \rightarrow \rho^0 K^+)</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow K^{*+} \pi^-)</td>
<td>(X_\rho^+ \rightarrow K^{*+} \pi^-)</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow \rho^0 K^0)</td>
<td>(X_\rho^+ \rightarrow \rho^0 K^0)</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow K^{0(\pi^0)} \eta)</td>
<td>(X_\rho^+ \rightarrow K^{0(\pi^0)} \eta)</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow K^{*0} \eta)</td>
<td>(X_\rho^+ \rightarrow K^{*0} \eta)</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow \rho^0 \eta)</td>
<td>(X_\rho^+ \rightarrow \rho^0 \eta)</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow \phi \pi^0)</td>
<td>(X_\rho^+ \rightarrow \phi \pi^0)</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow K^{*0} \pi^0)</td>
<td>(X_\rho^+ \rightarrow K^{*0} \pi^0)</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow K^{*0} K^0)</td>
<td>(X_\rho^+ \rightarrow K^{*0} K^0)</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow K^{*0} \pi^0)</td>
<td>(X_\rho^+ \rightarrow K^{*0} \pi^0)</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow K^{*0} K^0)</td>
<td>(X_\rho^+ \rightarrow K^{*0} K^0)</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow \rho^0 \eta)</td>
<td>(X_\rho^+ \rightarrow \rho^0 \eta)</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow \phi \pi^0)</td>
<td>(X_\rho^+ \rightarrow \phi \pi^0)</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow K^{*0} \pi^0)</td>
<td>(X_\rho^+ \rightarrow K^{*0} \pi^0)</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow K^{*0} K^0)</td>
<td>(X_\rho^+ \rightarrow K^{*0} K^0)</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow \rho^0 \eta)</td>
<td>(X_\rho^+ \rightarrow \rho^0 \eta)</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow \phi \pi^0)</td>
<td>(X_\rho^+ \rightarrow \phi \pi^0)</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow K^{*0} \pi^0)</td>
<td>(X_\rho^+ \rightarrow K^{*0} \pi^0)</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow K^{*0} K^0)</td>
<td>(X_\rho^+ \rightarrow K^{*0} K^0)</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow \rho^0 \eta)</td>
<td>(X_\rho^+ \rightarrow \rho^0 \eta)</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow \phi \pi^0)</td>
<td>(X_\rho^+ \rightarrow \phi \pi^0)</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow K^{*0} \pi^0)</td>
<td>(X_\rho^+ \rightarrow K^{*0} \pi^0)</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow K^{*0} K^0)</td>
<td>(X_\rho^+ \rightarrow K^{*0} K^0)</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow \rho^0 \eta)</td>
<td>(X_\rho^+ \rightarrow \rho^0 \eta)</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow \phi \pi^0)</td>
<td>(X_\rho^+ \rightarrow \phi \pi^0)</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow K^{*0} \pi^0)</td>
<td>(X_\rho^+ \rightarrow K^{*0} \pi^0)</td>
</tr>
<tr>
<td>(X_\rho^+ \rightarrow K^{*0} K^0)</td>
<td>(X_\rho^+ \rightarrow K^{*0} K^0)</td>
</tr>
</tbody>
</table>
The decay channels which can be obtained from the channels appearing in Table 2 by making charge conjugation, are not shown. From this table, we can learn that $X^0_a(\chi(1576))$ can decay into $K^{*+}K^-, K^{*-}K^+, K^{0}\bar{K}_L, K^{0}\bar{K}_S, K^{*0}\bar{K}_L, K^{*0}\bar{K}_S, \rho^0\eta, \rho^0\eta'$, $\phi\pi^0$. Since the pole position of $X^0_a(\chi(1576))$ is below the threshold of $\rho^0\eta'$, the process $X^0_a(\chi(1576)) \to \rho^0\eta'$ only occurs from the tail of its mass distribution. Some interesting relations can be obtained,

\begin{align}
\Gamma(\chi^+_a \to K^{*+}\bar{K}^0) & \approx \Gamma(\chi^+_a \to \bar{K}^{*0}K^+) \approx 2\Gamma(\chi^+_a \to K^{*+}K^-) \approx 2\Gamma(\chi^+_f \to K^{*+}K^-), \\
\Gamma(\chi^+_0 \to K^{*+}K^-) & \approx \Gamma(\chi^+_a \to K^{*+}K^-) \approx \Gamma(\chi^+_a \to K^{*0}\bar{K}^0) \approx \Gamma(\chi^+_a \to \bar{K}^{*0}K^0), \\
\Gamma(\chi^+_0 \to K^{*+}K^-) & \approx \Gamma(\chi^+_f \to K^{*+}K^-) \approx \Gamma(\chi^+_f \to K^{*0}\bar{K}^0) \approx \Gamma(\chi^+_f \to \bar{K}^{*0}K^0), \\
\Gamma(\chi^+_a \to \rho^+\eta') & \approx \Gamma(\chi^+_a \to \rho^0\eta') \approx \Gamma(\chi^+_f \to \omega\eta), \\
\Gamma(\chi^+_a \to \rho^+\eta') & \approx \Gamma(\chi^+_0 \to \rho^0\eta') \approx \Gamma(\chi^+_0 \to \omega\eta'), \\
\Gamma(\chi^+_a \to \phi\pi^0) & \approx \Gamma(\chi^+_0 \to \phi\pi^0).
\end{align}

(24)

\begin{align}
\Gamma(\chi^+_a \to \rho^+K^0) & \approx \Gamma(\chi^+_a \to \rho^-K^+), \\
\Gamma(\chi^+_a \to \rho^0K^0) & \approx \Gamma(\chi^+_a \to \omega K^0) \approx \frac{1}{2}\Gamma(\chi^+_a \to \rho^+K^+), \\
\Gamma(\chi^+_0 \to K^{0}\pi^+) & \approx 2\Gamma(\chi^+_0 \to K^{*+}\pi^+) \approx \Gamma(\chi^+_0 \to K^{*+}\pi^-) \approx 2\Gamma(\chi^+_0 \to K^{*0}\pi^0), \\
\Gamma(\chi^+_a \to K^{*+}\eta) & \approx \Gamma(\chi^+_a \to K^{0}\eta), \Gamma(\chi^+_a \to K^{*+}\eta) \approx \Gamma(\chi^+_a \to K^{*0}\eta), \\
\Gamma(\chi^+_0 \to \rho^-\pi^+ & \approx \Gamma(\chi^+_a \to \rho^+\pi^-) \approx \Gamma(\chi^+_0 \to \rho^0\pi^0),
\end{align}

(25)

where $\tilde{\Gamma}$ denotes the partial decay width neglecting phase space. We can see that Eqs. (24) and (25) are consistent with the isospin symmetry. The first equation in Eq. (26) is exactly Eq. (4) of Ref. [3], and the equations in Eq. (26) reflect the flavor cross symmetry in the decay of the four quark states. Using Eq. (23) and the pole position of $\chi(1576)$: $(1576\pm49^{+98}_{-97})$ MeV $-i(409^{+11+12}_{-12-64})$ MeV, we can further obtain,

\begin{align}
\Gamma(X^+_a(\chi(1576)) \to K^{*+}\bar{K}^0) & : \Gamma(X^+_a(\chi(1576)) \to \bar{K}^{*0}K^+) : \Gamma(X^+_a(\chi(1576)) \to K^{*+}K^-) : \Gamma(X^+_f(\chi(1576)) \to \rho^+\eta) \\
: \Gamma(X^+_a(\chi(1576)) \to \rho^-\eta') : \Gamma(X^+_a(\chi(1576)) \to \phi\pi^+) \approx 1 : 1 : 0.47 : 0.175 : 1.24, \\
\Gamma(X^+_0(\chi(1576)) \to K^{*+}K^-) : \Gamma(X^+_0(\chi(1576)) \to K^{*0}\bar{K}^0) : \Gamma(X^+_0(\chi(1576)) \to \rho^0\eta) \\
: \Gamma(X^+_a(\chi(1576)) \to \rho^0\eta') : \Gamma(X^+_a(\chi(1576)) \to \phi\pi^+) \approx 1 : 1 : 0.94 : 0.35 : 2.48.
\end{align}

(27)

The above ratios shows that in our four quark state scenario, the decay $X^0_a(\chi(1576)) \to \phi\pi^0$ is favorable, which is the distinctive feature of our four quark state interpretation. We expect the $I=1$ state $X^+_a$ should also appear in $J/\psi \to X^+_a\pi^- \to \phi\pi^+\pi^-$, $J/\psi \to X^+_a\pi^- \to K^{*0}\bar{K}^0\pi^+$ and $J/\psi \to X^+_a\pi^- \to K^{*+}\bar{K}^0\pi^-$. Experimental search of the channel $X^0_a(\chi(1576)) \to \text{pseudoscalar} + \text{vector}$ is necessary so that the existence of $X(1576)$ can be reexamined.

4. Conclusion and discussion

We propose that $\chi(1576)$ recently reported by BES Collaboration can be interpreted as the diquark–antidiquark bound state in P-wave excitation. This implies that there exists a vector nonet $\chi$, and $\chi(1576)$ is a member of the nonet. We estimate the mass spectrum of the nonet by considering both the colormagnetic hyperfine interaction energy and the P-wave excitation energy. The theoretical prediction for the mass of $\chi(1576)$ is about 1632 MeV, which is consistent with the experimental data: $(1576\pm49^{+98}_{-97})$ MeV $-i(409^{+11+12}_{-12-64})$ MeV. The strong coupling of $\chi(1576)$ to its decay channel $K^{*+}K^-$ may affect both the imag-
inary part of the pole position and its real part [16], this effect is ignored in the work, which is need to be studied further. Because the experimental error on the pole position is large, we expect the prediction will also be consistent with the experimental data if this effect is taken into account. The diquark here is taken as “good” diquark, generally the “bad” diquark is involved. However, the lowest lying and more stable states are dominated by the “good” diquark configuration [7,13]. Dealing with the mixing effects exactly from quark model is in progress.

OZI allowed strong decay of the nonet are investigated in detail. Both two pseudoscalars decay channel and one pseudoscalar plus one vector meson channel are discussed. We find out that in our four quark state scheme, the dominant decay modes of $X_0^{(1)}(X(1576))$ are $K^+K^-$, $K_2K_S$, $f\pi$, but not $\pi^+\pi^-$, and this is a important test for our proposal. We predict that the positive and negative charged isospin partner of $X_0^{(1)}(X(1576))$ dominantly decay into strange mesons. Since these two states are connected by charge conjugation, we concentrate on the positive charged $I_3=1$ states $X_0^+$. In order to search these states, we suggest to analyze the $J/\psi$ decay data in $J/\psi \rightarrow K^+K^-\pi^+$, $J/\psi \rightarrow K^+K_S\pi^-$ and $J/\psi \rightarrow f\pi^+\pi^-$. The observation of $X_0^+$ is another crucial test of our scheme. The decays of the other members of the nonet are also discussed, which can provide important clue to the experimental search of these states. Similar to $X_0^{(1)}$ (i.e., $X(1576)$), the width of these states should be broad too, and hence it is also difficult to observe them experimentally.

Diquark is in $3_c$ configuration, so diquark and antidiquark cannot be observed individually. As the distance between diquark and antidiquark gets large, a $q\bar{q}$ will be created from the vacuum, then the state decays into baryon–antibaryon. But the central values of the mass distribution of the nonet are generally bellow the threshold of the baryon–antibaryon pair, the decay width should be small. These states can mix with the ordinary $q\bar{q}$ states, if they have the same quantum numbers (e.g., $X_0^{(1)}(X(1576))$ can mix with $\rho(1450)$, $\rho(1700)$ and so on). The mixing effects interfere in the spectrum and the decay properties, and a full consideration including mixing effects is notoriously a difficult problem in exotic hadron spectrum. More sophisticated treatment of $X(1576)$ which takes these effects into account is expected. However, as a first step to understand $X(1576)$, we expect that such mixing effects in most cases are small from previous work on multiquark states [7,13], and the results obtained in this Letter are at least correct qualitatively.

Recently BES performs a partial wave analysis of $J/\psi \rightarrow \phi\pi\pi$ and $J/\psi \rightarrow \phi^+\phi^-$ from a sample of 58 $M$ $J/\psi$ events in the BESII detector. There is a strong peak in the $\phi\pi$ mass distribution which centers at 1500 MeV/c$^2$ with a full-width of 200 MeV/c$^2$, however, this peak disappears after making a cut against $K^*(890)$ [17]. Moreover, a $\phi\pi$ peak with $J^P=1^-$ at about 1500 MeV/c$^2$ was also reported about twenty years ago [18]. Maybe more experimental work is needed to make sure if the $\phi\pi$ peak exists, and whether there is some component of $X_0^+$ in this $\phi\pi$ peak. If the $\phi\pi$ peak is confirmed, it will be a great support to our picture of $X(1576)$. Finally, we would like to mention that if our predictions are not consistent with future experimental results, $X(1576)$ should have a different structure. More experimental facts about $X(1576)$ are needed in order to clarify these issues.

Acknowledgements

This work is partially supported by National Natural Science Foundation of China under Grant Numbers 90403021, and by the PhD Program Funds 20020358040 of the Education Ministry of China and KJCX2-SW-N10 of the Chinese Science Academy.

References

On the extraction of the quark mass ratio \((m_d - m_u)/m_s\) from
\[ \frac{\Gamma(\eta' \to \pi^0 \pi^+ \pi^-)}{\Gamma(\eta' \to \eta \pi^+ \pi^-)} \]

B. Borasoy \(^a\), Ulf-G. Meißner \(^a, b\), R. Nißler \(^a, *\)

\(^a\) Helmholtz-Institut für Strahlen- und Kernphysik (Theorie), Universität Bonn, D-53115 Bonn, Germany
\(^b\) Institut für Kernphysik (Theorie), Forschungszentrum Jülich, D-52425 Jülich, Germany

Received 6 September 2006; accepted 13 October 2006
Available online 20 October 2006
Editor: G.F. Giudice

Abstract

The claim that the light quark mass ratio \((md - mu)/ms\) can be extracted from the decay width ratio \(\Gamma(\eta' \to \pi^0 \pi^+ \pi^-)/\Gamma(\eta' \to \eta \pi^+ \pi^-)\) is critically investigated within a \(U(3)\) chiral unitary framework. The influence of the recent VES data on the \(\eta' \to \eta \pi^+ \pi^-\) decay is also discussed. © 2006 Elsevier B.V. All rights reserved.

1. Introduction

The light quark masses \(mu, md, ms\) are fundamental parameters of quantum chromodynamics and ought to be constrained as accurately as possible. The determination of the light quark mass ratios has been the goal of a variety of investigations in low-energy hadron physics, see e.g. [1–5]. Of particular interest is the quark mass difference \(md - mu\) which induces isospin breaking in QCD. Moreover, the possibility \(mu = 0\) would provide an explanation for the strong \(CP\) problem.

An accurate way of extracting \(md - mu\) is given by the isospin-violating decays \(\eta, \eta' \to \pi^0 \pi^+ \pi^-\) and \(\eta, \eta' \to 3\pi^0\). While for most processes isospin-violation of the strong interactions is masked by electromagnetic effects, these corrections are expected to be small for the three pion decays of the \(\eta\) and \(\eta'\) (Sutherland’s theorem) [6] which has been confirmed in an effective Lagrangian framework [7]. Neglecting electromagnetic corrections the decay amplitudes are directly proportional to \(md - mu\).

For this reason, it has been claimed in [8] that the branching ratio \(r = \Gamma(\eta' \to \pi^0 \pi^+ \pi^-)/\Gamma(\eta' \to \eta \pi^+ \pi^-)\) can be utilized in a very simple manner to extract the light quark mass difference \(md - mu\). To this aim, it is assumed that

\[(a)\] the amplitude \(A(\eta' \to \pi^0 \pi^+ \pi^-)\) is determined by the corresponding amplitude \(A(\eta' \to \eta \pi^+ \pi^-)\) via \(A(\eta' \to \pi^0 \pi^+ \pi^-) = \epsilon A(\eta' \to \eta \pi^+ \pi^-)\) with \(\epsilon = (\sqrt{3}/4)(md - mu)/(ms - \tilde{m})\) the \(\pi^0 - \eta\) mixing angle and \(\tilde{m} = (md + ms)/2\). (Note that in [8] the difference \(ms - \tilde{m}\) has been approximated by \(ms\) in the denominator of \(\epsilon\). Eq. (1) implies that the decay \(\eta' \to \pi^0 \pi^+ \pi^-\) proceeds entirely via \(\eta' \to \eta \pi^+ \pi^-\) followed by \(\pi^0 - \eta\) mixing;

\[(b)\] both amplitudes are “essentially constant” over phase space (see the remark in front of Eq. (19) of Ref. [8]).

Based on these two assumptions one arrives at the relation
\[ r = \frac{\Gamma(\eta' \to \pi^0 \pi^+ \pi^-)}{\Gamma(\eta' \to \eta \pi^+ \pi^-)} \simeq (16.8) \frac{3}{16} \left( \frac{md - mu}{ms} \right)^2, \quad (2) \]
where the factor 16.8 represents the phase space ratio. Comparison with experimental data—for which, so far, only an upper limit exists—would then lead to a prediction for the quark mass ratio \((md - mu)/(ms - \tilde{m}) \simeq (md - mu)/ms\). The purpose of the present work is to critically examine these two assumptions which lead to the simple relation in Eq. (2). Such an investigation is very timely in view of the recent and ongoing exper-

---

\* Corresponding author.

E-mail addresses: borasoy@itkp.uni-bonn.de (B. Borasoy), meissner@itkp.uni-bonn.de (U.-G. Meißner), missler@itkp.uni-bonn.de (R. Nißler).

0370-2693/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.physletb.2006.10.020
imental activities on $\eta$ and $\eta'$ decays at the WASA facility at COSY [9], MAMI-C [10], KLOE at DAΦNE [11] and by the VES Collaboration [12,13].

An appropriate theoretical framework to investigate low-energy hadronic physics is provided by chiral perturbation theory (ChPT) [14], the effective field theory of QCD. In ChPT Green functions are expanded perturbatively in powers of Goldstone boson masses and small momenta. However, final-state interactions in $\eta \to 3\pi$ have been shown to be substantial both in a complete one-loop calculation in $SU(3)$ ChPT [15] and using a dispersive framework [16,17]. It is hence important to include final-state interactions in a non-perturbative fashion.

In $\eta'$ decays final-state interactions are expected to be even more important due to larger phase space and the presence of nearby resonances. In this investigation, we apply the framework of $U(3)$ chiral effective field theory in combination with a relativistic coupled-channels approach developed in [18,19] in order to calculate the ratio $r$. Final-state interactions are included by deriving $s$- and $p$-wave interaction kernels for meson–meson scattering from the chiral effective Lagrangian and iterating them in a Bethe–Salpeter equation. The infinite iteration of the chiral effective potentials generates resonances dynamically. Very good overall agreement with currently available data on $\eta$, $\eta'$ decay widths and spectral shapes has been achieved in [18,19].

In the next section, we will investigate in our approach if both the assumptions (a) and (b) are justified. The inclusion of the recent VES data for $\eta \to 3\pi$ which is significantly lower than the previous upper limit of 5% [24]. This tighter bound translates to an upper limit of 3.8 keV for the partial decay width and reduces the upper limit for $r$ from 10% (as quoted by the PDG) to 4.1%. The pertinent results for the four fit clusters of [19] are well below these new upper limits and can be utilized without modification. As already reported in [19], the fit to the data does not allow for conclusions on the size of violations to Dashen’s theorem since $Q^2$ is treated as an input parameter and variations of $Q$ in the range of 20 . . . 24 lead to equally good fits within our approach. Hence, we will set $Q = 24.1$ in our calculations—the value predicted by Dashen’s theorem. The results for the branching ratios obtained by employing assumption (a) and $Q = 24.1$ are shown in Table 1. The ratios are obtained by explicitly performing the integration of the amplitudes over phase space. Obviously, assumption (a) is not justified—at least for the neutral decay where, in particular, clusters 1 and 2 are in clear disagreement with experiment.

Next, we employ in addition assumption (b). This is achieved by averaging the $\eta' \to \pi^0\pi^+\pi^-$ amplitudes over phase space which are then employed for $\eta' \to 3\pi$ by means of assumption (a). The results are displayed in Table 2. One observes that for clusters 1 and 2 the decay widths into $3\pi$ and hence the ratios $r$, $r_2$ increase, while the changes for clusters 3 and 4 are rather moderate. However, recall that the Dalitz plot parameters of the approach [19] clearly indicate that the assumption of a constant amplitude is not justified for $\eta' \to \pi^0\pi^+\pi^-$, particularly for clusters 3 and 4. The partial compensation of the effects of assumption (a) in clusters 1, 2 and the moderate changes in clusters 3, 4 are therefore purely accidental.

We conclude that both assumptions (a) and (b) are not justified. This is further substantiated by comparison of $r$ and $r_2$ in Table 2 with the respective values from the full chiral unitary approach shown in Table 3. The values are in clear disagreement and, hence, both assumptions are not appropriate—at least within the chiral unitary approach.
Finally, we would like to investigate the differences which result if assumption (a) is replaced by the decay mechanism where \( \eta' \to 3\pi \) occurs due to \( \pi^0 - \eta' \) mixing followed by a (virtual) transition \( \pi^0 \to 3\pi \). Employing the relation \( A(\eta' \to 3\pi) = e'A(\pi^0 \to 3\pi) \) with \( e' \) being the \( \pi^0 - \eta' \) mixing angle [18] we find the values shown in Table 4. Assuming the \( \eta' \to 3\pi \) decays to proceed via this mechanism introduces a huge uncertainty and leads to different ratios \( r \) and \( r_2 \). This underlines the observation that the decays \( \eta' \to 3\pi \) cannot be expected to simply proceed either \( \pi^0 - \eta \) or \( \pi^0 - \eta' \) mixing. In particular, the isospin-breaking transition due to the quark mass difference \( m_d - m_u \) cannot be completely assigned to \( \pi^0 - \eta \) mixing as done in assumption (a). Despite its appealing simplicity, the crude estimate given in Eq. (2) is certainly not suited to precisely determine the double quark mass ratio \( Q^2 \). In fact, even at leading chiral order the \( \eta' \to 3\pi \) decay amplitude is not entirely due to \( \pi^0 - \eta \) mixing. There is also a contribution from an isospin-violating \( \eta' 3\pi \)-vertex from the explicit chiral symmetry breaking part of the Lagrangian at second chiral order, see e.g. Ref. [18].

On the other hand, employing the chiral unitary approach of [19] does not lead to a conclusive extraction of \( Q \) due to the present experimental situation. From a fit to the data in [24] supplemented by Dashen’s theorem one obtains the decay width ratio \( r = (0.35 \ldots 1.5\%) \) which is larger than the value of 0.18% quoted in [8]. Note also that there is a tendency to even larger values of \( r \) if \( Q \) is lowered, e.g., for \( Q = 22 \) we obtain the range \( r = (0.4 \ldots 2.8\%) \).

### 3. Inclusion of the VES data for \( \eta' \to \eta\pi^\pm\pi^- \)

In this section we study the changes in our results that occur if the recent VES data on the spectral shape of \( \eta' \to \eta\pi^\pm\pi^- \) [13] are taken into account. Note that the most recent analysis of the VES Collaboration [13] has not yet been included in Ref. [24]. The VES data have much higher statistics on the Dalitz slope parameters than previous experiments and by including them in the fit we obtain the results shown in Table 5. Since the amplitudes for \( \eta' \to \eta\pi^\pm\pi^- \) and \( \eta' \to \eta\pi^0\pi^0 \) are equal in the isospin limit and deviations are thus isospin-breaking and small in our approach, we only include the leading Dalitz parameter \( a \) of \( \eta' \to \eta\pi^0\pi^0 \) [25] and omit the higher ones which are—assuming only small isospin-violating contributions—not quite compatible with the new results of the
VES experiment for \( \eta' \to \eta \pi^+ \pi^- \). Our results are in good agreement with the Dalitz plot parameters extracted from the VES experiment. In Table 5 only the best least-squares fit is shown which is sufficient to discuss the qualitative changes of the results compared to those of Section 2. Note also that we have supplemented our fitting routine by a conjugate gradient method [26] and hence the numerical values have improved with respect to [19].

Our results for the \( \eta' \to 3 \pi \) decay widths and width ratios are displayed in Table 6. It is important to emphasize that the inclusion of the VES data reduces the number of fit clusters to one and we observe indeed one global minimum. There is, however, a strong tendency of the fits towards the upper limit \( \Gamma(\eta') \) and hence the numerical values have improved ago in [27].

The reason for both the large decay width and the strong fluctuations over phase space are mainly due to a large contribution from isospin \( I = 1 \) \( p \)-waves in the final-state interactions of the decay. While for \( I = 1 \) \( p \)-waves the uncharged two-particle channels are \( C \)-even and, due to \( C \)-invariance, do not couple to \( C \)-odd channels related to the \( \rho^0(770) \) as already pointed out in [19], the coupling of charged channels to the \( \rho^\pm(770) \) is not forbidden. In fact, an important feature of the fits including the VES data compared to those without these is the large enhancement of the \( \eta' p^\pm \rightarrow \rho^0 p^\pm \) coupling which also determines the importance of the \( \rho^\pm(770) \) in this decay. The pertinent Dalitz plot is shown in Fig. 1 and exhibits signatures of the \( \rho^\pm(770) \). Note, however, that these resonances do not appear as bands of increased amplitude at fixed two-particle energies (dotted lines in Fig. 1), since the \( p \)-wave contributions have a kinematical zero in the middle of these bands as indicated in Fig. 1 (dashed lines). Thus the amplitude only peaks at the edge of the Dalitz plot. Moreover, due to the symmetry of the amplitude under \( \pi^+ \leftrightarrow \pi^- \) exchange (\( C \)-invariance) the \( \rho^+, \rho^- \) peaks interfere constructively on the symmetry axis producing a pronounced peak structure at the top of the Dalitz plot, where the invariant mass of the \( \pi^+ \pi^- \) system is minimal. These features of the Dalitz plot of a pseudoscalar meson decaying into three pions have been pointed out long ago in [27].

4. Conclusions

In this work, we have critically investigated the claim of Ref. [8] that the light quark mass ratio \( (m_d - m_u)/(m_s - \hat{m}) \) can be extracted from the decay width ratio \( r = \Gamma(\eta' \to \pi^0 \pi^+ \pi^-)/\Gamma(\eta' \to \eta \pi^+ \pi^-) \). In order to study this issue we have employed a \( U(3) \) chiral unitary framework developed in [18,19] which is in very good agreement with the \( \eta, \eta' \) data on widths and spectral shapes. Our results clearly indicate that the two underlying assumptions of [8] in order to arrive at a relation between \( r \) and \( (m_d - m_u)/(m_s - \hat{m}) \), i.e., that (a) the decay \( \eta' \to \pi^0 \pi^+ \pi^- \) proceeds entirely via the decay \( \eta' \to \eta \pi^+ \pi^- \) followed by \( \rho^0 \rightarrow \eta \) mixing and that (b) the decay amplitudes are constant over phase space, are not justified at all. The results from the full chiral unitary approach are in plain disagreement with these two assumptions. Moreover, the present experimental situation which is used as input to fit the parameters of the
chiral unitary approach does not allow for a precise determination of the double quark mass ratio \( Q_2 \) from \( r \).

Inclusion of the recent VES data on the \( \eta' \to \eta \pi^+ \pi^- \) spectral shape reduces the uncertainty of the fit results to some extent. In this case, the overall fit to \( \eta \) and \( \eta' \) data yields for \( \eta' \to \pi^0 \pi^+ \pi^- \) a large contribution from the isospin \( I = 1 \) \( p \)-wave in the final-state interactions which can be attributed to a large coupling to the \( \rho^\pm (770) \) resonances while contributions related to the \( \rho^0 (770) \) are forbidden by \( C \)-invariance.

More precise data on \( \eta \) and \( \eta' \) decays are needed in order to eventually clarify this issue. An improvement of the experimental situation is foreseen in the near future due to the upcoming data from WASA at COSY [9], MAMI-C [10] and KLOE at DAΦNE [11].

Acknowledgements

We thank Heiri Leutwyler for reading the manuscript and useful comments. This research is part of the EU Integrated Infrastructure Initiative Hadron Physics Project under contract number RII3-CT-2004-506078. Work supported in part by DFG (SFB/TR 16, “Subnuclear Structure of Matter”, and BO 1481/6-1).

References

The QCD transition temperature: Results with physical masses in the continuum limit

Y. Aoki, Z. Fodor, S. D. Katz, K. K. Szabó

Abstract

The transition temperature ($T_c$) of QCD is determined by Symanzik improved gauge and stout-link improved staggered fermionic lattice simulations. We use physical masses both for the light quarks ($m_{ud}$) and for the strange quark ($m_s$). Four sets of lattice spacings ($N_t = 4, 6, 8$ and $10$) were used to carry out a continuum extrapolation. It turned out that only $N_t = 6, 8$ and $10$ can be used for a controlled extrapolation, $N_t = 4$ is out of the scaling region. Since the QCD transition is a non-singular cross-over there is no unique $T_c$. Thus, different observables lead to different numerical $T_c$ values even in the continuum and thermodynamic limit. The peak of the renormalized chiral susceptibility predicts $T_c = 151(3)(3)$ MeV, whereas $T_c$-s based on the strange quark number susceptibility and Polyakov loops result in $24(4)$ MeV and $25(4)$ MeV larger values, respectively. Another consequence of the cross-over is the non-vanishing width of the peaks even in the thermodynamic limit, which we also determine. These numbers are attempted to be the full result for the $T \neq 0$ transition, though other lattice fermion formulations (e.g. Wilson) are needed to cross-check them.

1. Introduction

The $T \neq 0$ QCD transition plays an important role in the physics of the early Universe and of heavy ion collisions (most recently at RHIC at BNL; future plans exist for the LHC at CERN and FAIR at GSI). In this Letter we study the absolute scales of the transition at vanishing chemical potential ($\mu = 0$), which is of direct relevance for the early universe ($\mu$ is negligible there) and for present heavy ion collisions (at RHIC $\mu \lesssim 40$ MeV, which is far less than the typical hadronic scale). The transition is known to be a cross-over [1] (at least using staggered fermions, for a discussion about the fourth-root trick, which is usually applied see e.g. [2] and references therein). There are several results in the literature for $T_c$ using both staggered and Wilson fermions [3–8]. Note however, that these results have typically four serious limitations. (a) The first one is related to the unphysical spectrum. (b) Another question is how to extrapolate to vanishing lattice spacing ($a \to 0$), approaching the continuum limit. (c) The third problem is how to set the absolute scale for a question, which needs an answer of a few percent accuracy. (d) Finally the fourth problem is related to some implicit assumptions about a real singularity, thus ignoring the analytic cross-over feature of the finite temperature QCD transition.

Our goal is to eliminate all these limitations and give the full answer.

ad a. All previous calculations were carried out with unphysical spectra. On the one hand, results with Wilson fermions were obtained with pion masses $m_\pi \gtrsim 560$ MeV when approaching the thermodynamical limit (since lattice QCD can give only dimensionless combinations, it is more precise to say that $m_\pi/m_\rho \gtrsim 0.7$, where $m_\rho$ is the mass of the rho meson). The transition is related to the spontaneous breaking of the chiral symmetry (which is driven by the pion sector) and the three physical pions have masses smaller than the transition temperature, thus the numerical value of $T_c$ could be sensitive to the unphysical spectrum. Though at $T = 0$ chiral perturbation the-
Ad b. Lattice QCD uses a discretized version of the Lagrangian and approaches the continuum limit by taking smaller and smaller lattice spacings. For lattice spacings which are smaller than some approximate limiting value the dimensionful predictions with a few percent accuracy. As we already emphasized lattice QCD predicts dimensionless combinations of physical observables. For dimensionful predictions one calculates an experimentally known dimensionful quantity, which is used then to set the overall scale. In many analyses the overall scale is related to some quantities which strictly speak-

ory provides a technique to extrapolate to physical $m_\pi$, no such controllable method exists around $T_c$. On the other hand, results with staggered fermions suffer from taste violation. There is one lightest pion state and a large (usually several 100 MeV) unphysical mass splitting between this lightest state and the higher lying other pion states. This mass splitting results in an unphysical spectrum. The artificial pion mass splitting disappears only in the continuum limit. For some choices of the actions the restoration of the proper spectrum happens only at very small lattice spacings, whereas for other actions somewhat larger lattice spacings are already satisfactory.

Our solution for problem (a) is threefold. First of all, we use physical quark masses in our staggered analysis (or equivalently we fix $m_K/f_K$ and $m_K/m_\pi$ to their experimental values, here $m_K$ and $f_K$ are the mass and decay constant of the kaon). Secondly, our choice of action (stout link improved fermions) leads to smaller pion mass splittings than other choices (see Fig. 1 of Ref. [9]). The third ingredient is the continuum limit extrapolation which removes the pion mass splitting completely.

Ad c. An additional problem appears if we want to give dimensionful predictions with a few percent accuracy. As we already emphasized lattice QCD predicts dimensionless combinations of physical observables. For dimensionful predictions one calculates an experimentally known dimensionful quantity, which is used then to set the overall scale. In many analyses the overall scale is related to some quantities which strictly speak-

Ad d. The QCD transition at non-vanishing temperatures is an analytic cross-over [1]. Since there is no singular temperature dependence different definitions of the transition point lead to different values. The most famous example for this phenomenon is the water-vapor transition, for which the transition temperature can be defined by the peaks of $dp/dT$ (temperature derivative of the density) and $c_p$ (heat capacity at fixed pressure). For pressures ($p$) somewhat less than $p_c = 22.064$ MPa the transition is of first order, whereas at $p = p_c$ the transition is second order. In both cases the singularity guarantees that both definitions of the transition temperature lead to the same result. For $p > p_c$ the transition is a rapid cross-over, for which e.g. both $dp/dT$ and $c_p$ show pronounced peaks as a function of the temperature, however these peaks are at different temperature values. Fig. 1 shows the phase diagram based on [12].

---

1 Note, that outside the scaling region even a seemingly small lattice spacing dependence can lead to an incorrect result. An infamous example is the Naik action [10] in the Stefan-Boltzmann limit: $N_t = 4$ and 6 are consistent with each other with a few % accuracy, but since they are not in the scaling region they are 20% off the continuum value.
Analogously, there is no unique transition temperature in QCD. Therefore, we determine $T_c$ using the sharp changes of the temperature ($T$) dependence of renormalized dimensionless quantities obtained from the chiral condensate ($\langle \bar{\psi} \psi \rangle$), quark number susceptibility ($n_q$) and Polyakov loop ($P$).

The Letter is organized as follows. In Section 2 we define our action, discuss the simulation techniques and list our simulation points at $T = 0$, which will be used to carry out our continuum extrapolation procedure unambiguously. Section 3 deals with the different definitions of observables, which are used to locate the transition point at $T \neq 0$. Having located the transition in the lattice parameter space we make a connection to dimensionful physical quantities, thus determine the overall scale and carry out the continuum extrapolation. In Section 4 we conclude.

2. Lattice action, simulations at $T = 0$ and setting the scale

In this Letter we use a tree-level Symanzik improved gauge, and a stout-improved staggered fermionic action (for the detailed form of our action see Eqs. (2.1)–(2.3) of Ref. [9]). The stout-smearing [13] reduces the taste violation, a lattice artefact of the staggered type of fermions. In a previous study we showed that this sort of smearing has the smallest taste violation among the ones used in the literature for large scale thermodynamical simulations.

We have not improved the fermion action, however due to the order of magnitude smaller costs (compared to e.g. the p4fat3 action) we could afford to take smaller lattice spacings ($N_t = 4, 6, 8$ and $10$). This turned out to be extremely beneficial, when converting the transition temperature into physical units. In particular the $T_c$ values were used. Simulations at $T = 0$, which will be used to carry out our continuum extrapolation procedure unambiguously. Section 3 deals with the different definitions of observables, which are used to locate the transition point at $T \neq 0$. Having located the transition in the lattice parameter space we make a connection to dimensionful physical quantities, thus determine the overall scale and carry out the continuum extrapolation.

We have to tune the lattice spacings and their continuum extrapolation. Our three $N_t$ values were varied along the LCP at each $\beta$. The lattice sizes were chosen to satisfy the $m_qN_t \geq 4$ condition. However, when calculating the systematic uncertainties of meson masses and decay constants, we have taken finite size corrections into account using continuum finite volume chiral perturbation theory [14] (these corrections were around or less than 1%). We have simulated between 700 and 3000 RHMC trajectories for each point in Table 1.

Chiral extrapolation to the physical pion mass led to $m_K/f_K$ and $m_K/m_N$ values, which agree with the experimental numbers on the 2% level. (Differences resulting from various fitting forms and finite volume corrections were included in the systematics.) This is the accuracy of our LCP.

In order to be sure that our results are safe from ambiguous determination of the overall scale, and to prove that we are really in the $a^2$ scaling region, we carried out a continuum extrapolation for three additional quantities which could be similarly good to set the scale (we normalized them by $f_K$ for $f_K$ determination in staggered QCD see [15]). Fig. 2 shows the measured values of $m_K/f_K$, $f_\pi/f_K$ and $r_0 f_K$, at different lattice spacings and their continuum extrapolation. Our three continuum predictions are in complete agreement with the experimental results (note, that $r_0$ cannot be measured directly in experiments; in this case the original experimental input is the spectrum of the $T$ resonance, which was used by the MILC, HPQCD and UKQCD Collaborations to calculate $r_0$ on the lattice [15,16]).

It is important to emphasize that at lattice spacings given by $N_t = 4$ and $6$ the overall scales determined by $f_K$ and $r_0$ are different quantities obtained from the chiral condensate ($\langle \bar{\psi} \psi \rangle$), quark number susceptibility ($n_q$) and Polyakov loop ($P$).

The Letter is organized as follows. In Section 2 we define our action, discuss the simulation techniques and list our simulation points at $T = 0$, which will be used to carry out our continuum extrapolation procedure unambiguously. Section 3 deals with the different definitions of observables, which are used to locate the transition point at $T \neq 0$. Having located the transition in the lattice parameter space we make a connection to dimensionful physical quantities, thus determine the overall scale and carry out the continuum extrapolation. In Section 4 we conclude.

2. Lattice action, simulations at $T = 0$ and setting the scale

In this Letter we use a tree-level Symanzik improved gauge, and a stout-improved staggered fermionic action (for the detailed form of our action see Eqs. (2.1)–(2.3) of Ref. [9]). The stout-smearing [13] reduces the taste violation, a lattice artefact of the staggered type of fermions. In a previous study we showed that this sort of smearing has the smallest taste violation among the ones used in the literature for large scale thermodynamical simulations.

We have not improved the fermion action, however due to the order of magnitude smaller costs (compared to e.g. the p4fat3 action) we could afford to take smaller lattice spacings ($N_t = 4, 6, 8$ and $10$). This turned out to be extremely beneficial, when converting the transition temperature into physical units. In particular the $T_c$ values were used. Simulations at $T = 0$, which will be used to carry out our continuum extrapolation procedure unambiguously. Section 3 deals with the different definitions of observables, which are used to locate the transition point at $T \neq 0$. Having located the transition in the lattice parameter space we make a connection to dimensionful physical quantities, thus determine the overall scale and carry out the continuum extrapolation. In Section 4 we conclude.

Table 1 Lattice parameters and sizes of our zero temperature simulations. The strange quark mass is varied along the LCP as $\beta$ is changed. The light quark masses, listed at each $(\beta, m_s)$ values, correspond approximately to $m_\pi$ values of 250 MeV, 320 MeV, 380 MeV and 430 MeV.

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$m_\pi$</th>
<th>$m_{sd}$</th>
<th>Lattice size</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.330</td>
<td>0.23847</td>
<td>0.02621</td>
<td>$12^3 \times 24$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.04368</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.06115</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.07862</td>
<td></td>
</tr>
<tr>
<td>3.450</td>
<td>0.15730</td>
<td>0.01729</td>
<td>$16^3 \times 32$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.02881</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.04033</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05186</td>
<td></td>
</tr>
<tr>
<td>3.550</td>
<td>0.10234</td>
<td>0.01312</td>
<td>$16^3 \times 32$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.01874</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.02624</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.03374</td>
<td></td>
</tr>
<tr>
<td>3.670</td>
<td>0.06331</td>
<td>0.00928</td>
<td>$24^3 \times 32$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.01391</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.01739</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.02203</td>
<td></td>
</tr>
<tr>
<td>3.750</td>
<td>0.05025</td>
<td>0.00736</td>
<td>$24^3 \times 32$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.01104</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.01473</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.01841</td>
<td></td>
</tr>
</tbody>
</table>

somewhat larger than the physical one (the $m_N$ values were approximately 250 MeV, 320 MeV, 380 MeV and 430 MeV), whereas the strange quark mass was fixed by the LCP at each $\beta$. The lattice sizes were chosen to satisfy the $m_NN_t \geq 4$ condition. However, when calculating the systematic uncertainties of meson masses and decay constants, we have taken finite size corrections into account using continuum finite volume chiral perturbation theory [14] (these corrections were around or less than 1%). We have simulated between 700 and 3000 RHMC trajectories for each point in Table 1.

Chiral extrapolation to the physical pion mass led to $m_K/f_K$ and $m_K/m_N$ values, which agree with the experimental numbers on the 2% level. (Differences resulting from various fitting forms and finite volume corrections were included in the systematics.) This is the accuracy of our LCP.

In order to be sure that our results are safe from ambiguous determination of the overall scale, and to prove that we are really in the $a^2$ scaling region, we carried out a continuum extrapolation for three additional quantities which could be similarly good to set the scale (we normalized them by $f_K$ for $f_K$ determination in staggered QCD see [15]). Fig. 2 shows the measured values of $m_K/f_K$, $f_\pi/f_K$ and $r_0 f_K$, at different lattice spacings and their continuum extrapolation. Our three continuum predictions are in complete agreement with the experimental results (note, that $r_0$ cannot be measured directly in experiments; in this case the original experimental input is the spectrum of the $T$ resonance, which was used by the MILC, HPQCD and UKQCD Collaborations to calculate $r_0$ on the lattice [15,16]).

It is important to emphasize that at lattice spacings given by $N_t = 4$ and $6$ the overall scales determined by $f_K$ and $r_0$ are different quantities obtained from the chiral condensate ($\langle \bar{\psi} \psi \rangle$), quark number susceptibility ($n_q$) and Polyakov loop ($P$).

The Letter is organized as follows. In Section 2 we define our action, discuss the simulation techniques and list our simulation points at $T = 0$, which will be used to carry out our continuum extrapolation procedure unambiguously. Section 3 deals with the different definitions of observables, which are used to locate the transition point at $T \neq 0$. Having located the transition in the lattice parameter space we make a connection to dimensionful physical quantities, thus determine the overall scale and carry out the continuum extrapolation. In Section 4 we conclude.

2. Lattice action, simulations at $T = 0$ and setting the scale

In this Letter we use a tree-level Symanzik improved gauge, and a stout-improved staggered fermionic action (for the detailed form of our action see Eqs. (2.1)–(2.3) of Ref. [9]). The stout-smearing [13] reduces the taste violation, a lattice artefact of the staggered type of fermions. In a previous study we showed that this sort of smearing has the smallest taste violation among the ones used in the literature for large scale thermodynamical simulations.

We have not improved the fermion action, however due to the order of magnitude smaller costs (compared to e.g. the p4fat3 action) we could afford to take smaller lattice spacings ($N_t = 4, 6, 8$ and $10$). This turned out to be extremely beneficial, when converting the transition temperature into physical units. In particular the $T_c$ values were used. Simulations at $T = 0$, which will be used to carry out our continuum extrapolation procedure unambiguously. Section 3 deals with the different definitions of observables, which are used to locate the transition point at $T \neq 0$. Having located the transition in the lattice parameter space we make a connection to dimensionful physical quantities, thus determine the overall scale and carry out the continuum extrapolation. In Section 4 we conclude.
ferring by ∼ 20−30%, which is most probably true for any other staggered formulation used for thermodynamical calculations. Since the determination of the overall scale has a ∼ 20−30% ambiguity, the value of \( T_c \) cannot be determined with the required accuracy.

For the simulations we were using the RHMC algorithm with multiple time scales [17]. The time consuming parts of the computations were carried out in single precision, however the exact reversibility of the algorithm was achieved. On one of our largest lattices we have cross-checked the results with a fully double precision calculation.

### 3. \( T \neq 0 \) simulations, transition points for different observables

The \( T \neq 0 \) simulations (cf. Table 2) were carried out along our LCP (that is at physical strange and light quark masses, which correspond to \( m_K = 498 \text{ MeV} \) and \( m_\pi = 135 \text{ MeV} \)) at four different sets of lattice spacings (\( N_t = 4, 6, 8 \) and 10) and on three different volumes (\( N_s / N_t \) ranging between 3 and 6). We have observed moderate finite volume effects on the smallest volumes for quantities which are supposed to depend strongly on light quark masses (e.g. chiral susceptibility). To determine the transition point we used \( N_s / N_t \geq 4 \), for which we did not observe any finite volume effect. The number of RHMC trajectories were between 1500 and 8000 for each parameter set (the integrated autocorrelation time was smaller or around 10 for all our runs).

We considered three quantities to locate the transition point: the chiral susceptibility, the strange quark number susceptibility and the Polyakov-loop. Since the transition at vanishing chemical potential is a cross-over, we expect that all three quantities result in different transition points (similarly to the case of the water, cf. Fig. 1).

#### Chiral susceptibility

The chiral susceptibility of the light quarks (\( \chi \)) is defined as

\[
\chi_{\bar{\psi}\psi} = \frac{T}{V} \frac{\partial^2}{\partial m_{ud}^2} \log Z = -\frac{\partial^2}{\partial m_{ud}^2} f,
\]

where \( f \) is the free energy density. Since both the bare quark mass and the free energy density contain divergences, \( \chi_{\bar{\psi}\psi} \) has to be renormalized [1].

The renormalized quark mass can be written as \( m_{R,ud} = Z_m \cdot m_{ud} \). If we apply a mass independent renormalization then we have

\[
m_{ud}^2 \frac{\partial^2}{\partial m_{ud}^2} = m_{R,ud}^2 \frac{\partial^2}{\partial m_{R,ud}^2}.
\]

The free energy has additive, quadratic divergencies. They can be removed by subtracting the free energy at \( T = 0 \) (this is the usual renormalization procedure for the free energy or pressure), which leads to \( f_R \). Therefore, we have the following identity:

\[
m_{ud}^2 \frac{\partial^2}{\partial m_{ud}^2} (f(T) - f(T = 0)) = m_{R,ud}^2 \frac{\partial^2}{\partial m_{R,ud}^2} f_R(T),
\]

the right-hand side contains only renormalized quantities, which can be determined by measuring the susceptibilities of the left-hand side (for the above expression we use the shorthand notation \( m_{ud}^2 \cdot \Delta \chi_{\bar{\psi}\psi} \)). In order to obtain a dimensionless quantity it is natural to normalize the above quantity by \( T^4 \) (which minimizes the final errors). Alternatively, one can use combinations of \( T \) and/or \( m_\pi \) to construct dimensionless quantities (though these conventions lead to larger errors). Since the transition is a cross-over (cf. discussion d of our Introduction) the maxima of \( m_{ud}^2 / m_\pi^2 \cdot \Delta \chi_{\bar{\psi}\psi} / T^2 \) or \( m_{ud}^2 / m_\pi^4 \cdot \Delta \chi_{\bar{\psi}\psi} \) give somewhat different values for \( T_c \).

The upper panel of Fig. 3 shows the temperature dependence of the renormalized chiral susceptibility for different temporal extensions (\( N_t = 4, 6, 8 \) and 10). For small enough lattice spacings, thus close to the continuum limit, these curves should
Fig. 3. Temperature dependence of the renormalized chiral susceptibility \( m^2 \Delta \chi_{\bar{\psi}\psi}/T^4 \), the strange quark number susceptibility \( \chi_s/T^2 \) and the renormalized Polyakov-loop (\( P_R \)) in the transition region. The different symbols show the results for \( N_t = 4, 6, 8 \) and 10 lattice spacings (filled and empty boxes for \( N_t = 4 \) and 6, filled and open circles for \( N_t = 8 \) and 10). The vertical bands indicate the corresponding critical temperatures and its uncertainties coming from the \( T \neq 0 \) analyses. This error is given by the number in the first parenthesis, whereas the error of the overall scale determination is indicated by the number in the second parenthesis. The orange bands show our continuum limit estimates for the three renormalized quantities as a function of the temperature with their uncertainties.

As it can be seen, the \( N_t = 4 \) result has considerable lattice artefacts, however the two smallest lattice spacings (\( N_t = 8 \) and 10) are already consistent with each other, suggesting that they are also consistent with the continuum limit extrapolation (indicated by the orange band). The curves exhibit pronounced peaks. We define the transition temperatures by the position of these peaks. We fitted a second order expression to the peak to obtain its position. The slight change due to the variation of the fitting range is taken as a systematic error. The left panel of Fig. 4 shows the transition temperatures in physical units for different lattice spacings obtained from the chiral susceptibility. As it can be seen \( N_t = 6, 8 \) and 10 are already in the scaling region, thus a safe continuum extrapolation can be carried out. The extrapolations based on \( N_t = 6, 8, 10 \) fit and \( N_t = 8, 10 \) fit are consistent with each other. For our final result we use the average of these two fit results (the difference between them are added to our systematic uncertainty). Our \( T = 0 \) simulations resulted in a 2% error on the overall scale. Our final result for the transition temperature based on the chiral susceptibility reads:

\[
T_c(\chi_{\bar{\psi}\psi}) = 151(3)(3) \text{ MeV},
\]

where the first error comes from the \( T \neq 0 \), the second from the \( T = 0 \) analyses.

We use the second derivative of the chiral susceptibility \( \chi'' \) at the peak position to estimate the width of the peak \( (\Delta T_c)^2 = -\chi(T_c)/\chi''(T_c) \). For the continuum extrapolated width we obtained:

\[
\Delta T_c(\chi_{\bar{\psi}\psi}) = 28(5)(1) \text{ MeV}.
\]

Note, that for a real phase transition (first or second order), the peak would have a vanishing width (in the thermodynamic limit), yielding a unique value for the critical temperature. Due to the crossover nature of the transition there is no such value, there is a range \( (151 \pm 28 \text{ MeV}) \) where the transition phenomena takes place. Other quantities than the chiral susceptibility could result in transition temperatures within this range.

The MILC Collaboration also reported a continuum result on the transition temperature based on the chiral susceptibility [7]. Their result is 169(12)(4) MeV. Note, that their lattice spacings were not as small as ours (they used \( N_t = 4, 6 \) and 8), their aspect ratio was quite small \( (N_s/N_t = 2) \), they used non-physical quark masses (their smallest pion mass at \( T \neq 0 \) was \( \approx 220 \text{ MeV} \)), the non-exact R-algorithm was applied for the simulations and they did not use the renormalized susceptibility, but they looked for the peak in the bare \( \chi_{\bar{\psi}\psi}/T^2 \). Using \( T^4 \) as a normalization prescription (as we did) the transition temperature would decrease their \( T_c \) values by approximately 9 MeV. Note, that their continuum extrapolation resulted in a quite large error. Taking into account their uncertainties our result and their result agree on the 1-sigma level.

Quark number susceptibility

For heavy-ion experiments the quark number susceptibilities are quite useful, since they could be related to event-by-event fluctuations. Our second transition temperature is obtained from
the strange quark number susceptibility, which is defined via [7]

\[
\frac{\chi_s}{T^2} = \frac{1}{TV} \frac{\partial^2 \log Z}{\partial \mu_s^2} \bigg|_{\mu_s=0},
\]

where \(\mu_s\) is the strange quark chemical potential (in lattice units). Quark number susceptibilities have the convenient property, that they automatically have a proper continuum limit, there is no need for renormalization.

The middle panel of Fig. 3 shows the temperature dependence of the strange quark number susceptibility for different temporal extensions (\(N_t = 4, 6, 8\) and 10). For small enough lattice spacings, thus close to the continuum limit, these curves should coincide again (our continuum limit estimate is indicated by the orange band).

As it can be seen, the \(N_t = 4\) results are quite off, however the two smallest lattice spacings (\(N_t = 8\) and 10) are already consistent with each other, suggesting that they are also consistent with the continuum limit extrapolation. This feature indicates, that they are closer to the continuum result than our statistical uncertainty.

We defined the transition temperature as the peak in the temperature derivative of the strange quark number susceptibility, that is the inflection point of the susceptibility curve. The position was determined by two independent ways, which yielded the same result. In the first case we fitted a cubic polynomial on the susceptibility curve, while in the second case we determined the temperature derivative numerically from neighboring points and fitted a quadratic expression to the peak. The slight change due to the variation of the fitting range is taken as a systematic error. The middle panel of Fig. 4 shows the transition temperatures obtained from the renormalized chiral susceptibility (\(\chi_s/T^2\)) and renormalized Polyakov-loop (\(P_R\)).

where the first (second) error is from the \(T \neq 0\) (\(T = 0\)) temperature analysis (note, that due to the uncertainty of the overall scale, the difference is more precisely determined than the uncertainties of \(T_c(\chi_s)\) and \(T_c(\chi_{\bar{q}q})\) would suggest). Similarly to the chiral susceptibility analysis, the curvature at the peak can be used to define a width for the transition.

\[
\Delta T_c(\chi_s) = 42(4)(1) \text{ MeV.} \tag{8}
\]

Polyakov loop

In pure gauge theory the order parameter of the confinement transition is the Polyakov-loop:

\[
P = \frac{1}{N_t^3} \sum_x \text{tr} \left[ U_{4}(x, 0)U_{4}(x, 1) \cdots U_{4}(x, N_t - 1) \right]. \tag{9}
\]

\(P\) acquires a non-vanishing expectation value in the deconfined phase, signaling the spontaneous breakdown of the \(Z(3)\) symmetry. When fermions are present in the system, the physical interpretation of the Polyakov-loop expectation value is more complicated (see e.g. [19]). However, its absolute value can be related to the quark–antiquark free energy at infinite separation:

\[
\langle \langle P \rangle \rangle^2 = \exp(-\Delta F_{q\bar{q}}(r \to \infty)/T). \tag{10}
\]

\(\Delta F_{q\bar{q}}\) is the difference of the free energies of the quark–gluon plasma with and without the quark–antiquark pair.

The absolute value of the Polyakov-loop vanishes in the continuum limit. It needs renormalization. This can be done by renormalizing the free energy of the quark–antiquark pair [20]. Note, that QCD at \(T \neq 0\) has only the ultraviolet divergencies which are already present at \(T = 0\). In order to remove these divergencies at a given lattice spacing we used a simple renormalization condition [21]:

\[
V_R(r_0) = 0, \tag{11}
\]

where the potential is measured at \(T = 0\) from Wilson-loops. The above condition fixes the additive term in the potential at a given lattice spacing. This additive term can be used at the same lattice spacings for the potential obtained from Polyakov loops, or equivalently it can be built in into the definition of the

\[2\] A continuum extrapolation using only the two coarsest lattices (\(N_t = 4\) and 6) yielded \(T_c \sim 190\) MeV [18], where an approximate LCP was used, if the lattice spacing is set by \(r_0\).
renormalized Polyakov-loop.

\[ \langle P_R \rangle = \langle P \rangle \exp \left( V(r_0)/(2T) \right), \]

where \( V(r_0) \) is the unrenormalized potential obtained from Wilson-loops.

The lower panel of Fig. 3 shows the temperature dependence of the renormalized Polyakov-loops for different temporal extensions \( (N_t = 4, 6, 8 \text{ and } 10) \). The two smallest lattice spacings \( (N_t = 8 \text{ and } 10) \) are approximately in 1-sigma agreement (our continuum limit estimate is indicated by the orange band).

Similarly to the strange quark susceptibility case we defined the transition temperature as the peak in the temperature derivative of the Polyakov-loop, that is the inflection point of the Polyakov-loop curve. To locate this point and determine its uncertainties we used the same two methods, which were used to determine \( T_c(\chi_s) \). The right panel of Fig. 4 shows the transition temperatures in physical units for different lattice spacings obtained from the Polyakov-loop. As it can be seen \( N_t = 6, 8 \text{ and } 10 \) are already in the scaling region, thus a safe continuum extrapolation can be carried out. The extrapolation and the determination of the systematic error were done as for \( T_c(\chi_s) \). The continuum extrapolated value for the transition temperature based on the renormalized Polyakov-loop is significantly higher than the one from the chiral susceptibility. The difference is 25(4) MeV. For the transition temperature in the continuum limit one gets:

\[ T_c(P) = 176(3)(4) \text{ MeV}, \]

where the first (second) error is from the \( T \neq 0 \) (\( T = 0 \)) temperature analysis (again, due to the uncertainties of the overall scale, the difference is more precisely determined than the uncertainties of \( T_c(\chi_s) \) and \( T_c(P) \) suggest). Similarly to the chiral susceptibility analysis, the curvature at the peak can be used to define a width for the transition.

\[ \Delta T_c(P) = 38(5)(1) \text{ MeV}. \]

4. Conclusions

We determined the transition temperature of QCD by Symanzik improved gauge and stout-link improved staggered fermionic lattice simulations. We used an exact simulation algorithm and physical masses both for the light quarks and for the strange quark. The parameters were tuned with a quite high precision, thus at all lattice spacings the \( m_K/f_K \) and \( m_K/m_\pi \) ratios were set to their experimental values with an accuracy better than 2%. Four sets of lattice spacings \( (N_t = 4, 6, 8 \text{ and } 10) \) were used to carry out a continuum extrapolation. It turned out that only \( N_t = 6, 8 \text{ and } 10 \) can be used for a controlled extrapolation, \( N_t = 4 \) is out of the scaling region. Lattice spacings obtained at \( N_t = 6, 8 \text{ and } 10 \) transition points still result in different values for different physical inputs, but they are already in the scaling region, and an \( a^2 \) type extrapolation can be used. Any extrapolation merely based on \( N_t = 4 \text{ and } 6 \) would contain an unknown systematic error. We demonstrated, that our result is independent of the choice of the physical quantity, which is used to set the overall scale. We calculated three additional quantities, which would give the same dimensionful result for \( T_c \), since we reproduced their experimental values in the continuum limit. (These ambiguities, related to setting the scale, are serious drawbacks of the analyses which can be found in the literature.)

Since the QCD transition is a non-singular cross-over [1] there is no unique \( T_c \). We illustrated this well-known phenomenon on the water–vapor phase diagram. Different observables lead to different numerical \( T_c \) values even in the continuum and thermodynamic limit also in QCD. We used three observables to determine the corresponding transition temperatures. The peak of the renormalized chiral susceptibility predicts \( T_c = 151(3)(3) \text{ MeV} \), whereas \( T_c(\chi) \) based on the strange quark number susceptibility resulted in 24(4) MeV larger value. Another quantity, which is related to the confining phase transition in the large quark mass limit is the Polyakov loop. Its behavior predicted a 25(4) MeV larger transition temperature, than that of the chiral susceptibility. Another consequence of the cross-over is the non-vanishing widths of the peaks even in the thermodynamic limit, which we also determined. For the chiral susceptibility, strange quark number susceptibility and Polyakov-loop we obtained widths of 28(5)(1) MeV, 42(4)(1) MeV and 38(5)(1) MeV, respectively. These numbers are attempted to be the full result for the \( T \neq 0 \) transition, though other lattice fermion formulations (e.g. Wilson) are needed to cross-check them.

Note

After finishing the simulations of the present Letter and preparing our manuscript an independent study on \( T_c \) based on large scale simulations of the Bielefeld–Brookhaven–Columbia–Riken group appeared on the archive [8]. The p4fat3 action was used, which is designed to give very good results in the \( (T \rightarrow \infty) \) Stefan–Boltzmann limit (their action is not optimized at \( T = 0 \), which is needed e.g. to set the scale). The overall scale was set by \( r_0 \). The \( T_c \) analysis based on the chiral susceptibility peak gave in the continuum limit \( T_c(\chi) = 192(7)(4) \text{ MeV} \). (The second error, 4 MeV, estimates the uncertainty of the continuum limit extrapolation, which we do not use in the following, since we attempt to give a more reliable estimate on that.) This result is in obvious contradiction with our continuum result from the same observable, which is \( T_c(\chi) = 151(3)(3) \text{ MeV} \). For the same quantity (position of chiral susceptibility peak with physical quark masses in the continuum limit) one should obtain the same numerical result independently of the lattice action. Since the chance probability that we are faced with a statistical fluctuation and both of the results are correct is small, we attempted to understand the origin of the discrepancy. We repeated some of their simulations and analyses. In these cases a complete agreement was found. In addition to their \( T = 0 \) analyses we carried out an \( f_K \) determination, too. This \( f_K \) was used to extend their work, to use an LCP based on \( f_K \) and to determine \( T_c \) in physical units.

We summarize the origin of the contradiction between our findings and theirs. The major part of the difference can be explained by the fact, that the lattice spacings of [8] are too large
that the lattice spacings \( \gtrsim \) cannot give a consistent continuum limit for scaling regime. Thus, results obtained with their lattice spacings cannot give a consistent continuum limit for scaling regime. We used the renormalized chiral susceptibility normalized by \( T^2 \) of Ref. [8].

Acknowledgements

We thank F. Csikor, J. Kuti and L. Lellouch for useful discussions. We thank P. Petreczky for critical reading of the manuscript. This research was partially supported by OTKA Hungarian Science Grants Nos. T34980, T37615, M37071, T032501, AT049652, by DFG German Research Grant No. FO 502/1-1 and by the EU Research Grant No. RI3-CT-20040506078. The computations were carried out on the 370 processor PC cluster of Eötvös University, on the 1024 processor PC cluster of Wuppertal University, on the 107 node PC cluster equipped with Graphical Processing Units at Wuppertal University and on the BlueGene/L at FZ Jülich. We used a modified version of the publicly available MILC code [22] with next-neighbor communication architecture for PC-clusters [23].

References


Fig. 5. Resolving the discrepancy between the critical temperature of Ref. [8] and that of the present work (see text). The major part of the difference can be traced back to the unreliable continuum extrapolation of [8]. Left panel: In Ref. [8] \( r_0 \) was used for scale setting (filled boxes), however using the kaon decay constant (empty boxes) leads to different critical temperatures even after performing the continuum extrapolation. Right panel: In our work the extrapolations based on the finer lattices are safe, using the two different scale setting methods one obtains consistent results (the errors contain uncertainties from the determination of the transition point, \( r_0 \) and \( f_K \)).
[22] MILC Collaboration, public lattice gauge theory code, see http://physics.indiana.edu/~sg/milc.html.
Negative horizon mass for rotating black holes

Jutta Kunz, Francisco Navarro-Lérida *

Institut für Physik, Universität Oldenburg, Postfach 2503, D-26111 Oldenburg, Germany
Received 3 October 2006; accepted 16 October 2006
Available online 27 October 2006
Editor: N. Glover

Abstract

Charged rotating black holes of Einstein–Maxwell–Chern–Simons theory in odd dimensions, \( D \geq 5 \), may possess a negative horizon mass, while their total mass is positive. This surprising feature is related to the existence of counterrotating solutions, where the horizon angular velocity \( \Omega \) and the angular momentum \( J \) possess opposite signs. Black holes may further possess vanishing horizon angular velocity while they have finite angular momentum, or they may possess finite horizon angular velocity while their angular momentum vanishes. In \( D = 9 \) even non-static black holes with \( \Omega = J = 0 \) appear. Charged rotating black holes with vanishing gyromagnetic ratio exist, and black holes need no longer be uniquely characterized by their global charges.

© 2006 Elsevier B.V. All rights reserved.

1. Introduction

The Kerr–Newman (KN) family of black hole solutions of Einstein–Maxwell (EM) theory [1] possesses many special properties, not present in black hole solutions of theories with more general matter content. The inclusion of a dilaton, for instance, leads to deviations of the gyromagnetic ratio from the KN value [2], and when the dilaton coupling exceeds the Kaluza–Klein value counterrotating black holes appear, where the horizon rotates in the opposite sense to the total angular momentum [3]; while the EM uniqueness theorem or Israel’s theorem [4] do not hold in general in theories with non-Abelian fields [5,6].

In the context of supergravity or string theory, higher-dimensional black holes are of interest. The Myers–Perry solutions [7], representing higher-dimensional rotating vacuum black holes, have long been known, as well as various charged rotating black holes of supergravity and string theory [8]. In contrast, the higher-dimensional generalization of the KN black holes has so far resisted attempts to obtain them in analytical form [9,10].

In odd dimensions, \( D = 2N + 1 \), the EM action may be supplemented by an ‘\( A F^N \)’ Chern–Simons (CS) term. In 5 dimensions, for a certain value of the CS coefficient \( \lambda = \lambda_{SG} \), the theory corresponds to the bosonic sector of \( D = 5 \) supergravity, where rotating black hole solutions are known analytically [11,12]. In particular, extremal solutions exist, whose horizon angular velocity vanishes. Thus their horizon is non-rotating, although their angular momentum is non-zero. In these solutions a negative fraction of the total angular momentum is stored in the Maxwell field behind the horizon, and the effect of rotation on the horizon is not to make it rotate but to deform it into a squashed 3-sphere [13,14]. Here supersymmetry is associated with a borderline between stability and instability, since for \( \lambda > \lambda_{SG} \) a rotational instability arises, where counterrotating black holes appear [15]. Moreover, when the CS coefficient is increased beyond the \( 2\lambda_{SG} \), black holes (with horizon topology of a sphere [16]) are no longer uniquely characterized by their global charges [15].

Here we reanalyze 5-dimensional Einstein–Maxwell–Chern–Simons (EMCS) black holes and show, that they may possess a negative horizon mass for a CS coefficient beyond \( 2\lambda_{SG} \). We also investigate their higher-dimensional generalizations for arbitrary CS coefficient. While \( (D = 2N + 1) \)-dimensional black holes generically possess \( N \) independent angular momenta [7], we here focus on black holes with equal-magnitude angular mo-
menta [10], allowing for the reduction of the EMCS equations to a set of ordinary differential equations. In Section 2 we recall the EMCS action and present the appropriate Ansätze for the metric and the gauge potential. We discuss the black hole properties in Section 3. We present the black hole solutions in Section 4, and summarize in Section 5.

2. Action, metric and gauge potential

We consider the odd \(D\)-dimensional EMCS action with Lagrangian [13]

\[
L = \frac{1}{16\pi G_D} \sqrt{-g} \left( R - F_{\mu\nu} F^{\mu\nu} + \frac{8}{D+1} \hat{\lambda} \epsilon^{\mu_1 \mu_2 \cdots \mu_{D-2} \mu_D} F_{\mu_1 \mu_2} \cdots F_{\mu_{D-2} \mu_D} A_{\mu_D} \right),
\]

with curvature scalar \(R\), \(D\)-dimensional Newton constant \(G_D\), and field strength tensor \(F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu\), where \(A_\mu\) denotes the gauge potential. \(\hat{\lambda}\) corresponds to the CS coupling constant.

The field equations are obtained from the action Eq. (1) by taking variations with respect to the metric and the gauge potential, yielding the Einstein equations

\[
G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 2T_{\mu\nu},
\]

with stress-energy tensor

\[
T_{\mu\nu} = F_{\mu\rho} F_{\nu\sigma} - \frac{1}{4} g_{\mu\nu} F_{\rho\sigma} F^{\rho\sigma},
\]

and the gauge field equations

\[
\nabla_\nu F^{\mu\nu} = \hat{\lambda} \epsilon^{\mu_1 \mu_2 \cdots \mu_D} F_{\mu_1 \mu_2} \cdots F_{\mu_D}. \tag{4}
\]

In order to obtain stationary black hole solutions of Eqs. (2)–(4), representing charged generalizations of the odd \(D\)-dimensional Myers–Perry solutions [7], we consider black hole space–times with \(N\)-azimuthal symmetries, implying the existence of \(N + 1\) commuting Killing vectors, \(\xi \equiv \partial_t\) and \(\eta_{(k)} \equiv \partial_{\phi_k}\), for \(k = 1, \ldots, N\), where \(N\) is defined by \(D = 2N + 1\).

While generic EMCS black holes possess \(N\) independent angular momenta, we now restrict to black holes whose angular momenta have all equal magnitude. The metric and the gauge field parametrization then simplify considerably. In particular, for such equal-magnitude angular momenta black holes, the general Einstein and gauge field equations reduce to a set of ordinary differential equations [10], since the angular dependence can be treated explicitly.

Restricting to black holes with spherical horizon topology, we parametrize the metric in isotropic coordinates, which are well suited for the numerical construction of rotating black holes [9,15,17],

\[
ds^2 = -\frac{f}{f} \, dt^2 + \frac{m}{f} \left[ dr^2 + r^2 \sum_{i=1}^{N-1} \left( \prod_{j=0}^{i-1} \cos^2 \theta_j \right) d\theta_i^2 \right]
\]

\[
+ \frac{n}{f} r^2 \sum_{k=1}^{\frac{N}{2}} \left( \prod_{l=0}^{k-1} \cos^2 \theta_l \right) \sin^2 \theta_k \left( \epsilon_k \partial \phi_k - \frac{\omega}{r} dt \right)^2
\]

\[
+ \frac{m-n}{f} r^2 \left( \sum_{k=1}^{\frac{N}{2}} \left( \prod_{l=0}^{k-1} \cos^2 \theta_l \right) \sin^2 \theta_k \partial \phi_k^2
\]

\[\quad - \left[ \sum_{k=1}^{\frac{N}{2}} \left( \prod_{l=0}^{k-1} \cos^2 \theta_l \right) \sin^2 \theta_k \epsilon_k \partial \phi_k \right]^2 \right], \tag{5}
\]

where \(\theta_0 \equiv 0, \theta_i \in [0, \pi/2]\) for \(i = 1, \ldots, N - 1, \theta_N \equiv \pi/2, \phi_0 \in [0, 2\pi]\) for \(k = 1, \ldots, N\), and \(\epsilon_k = \pm 1\) denotes the sense of rotation in the \(k\)th orthogonal plane of rotation.

The parametrization for the gauge potential, consistent with Eq. (5), is

\[
A_\mu \, dx^\mu = a_0 \, dt + a_\phi \sum_{k=1}^{\frac{N}{2}} \left( \prod_{l=0}^{k-1} \cos^2 \theta_l \right) \sin^2 \theta_k \epsilon_k \partial \phi_k. \tag{6}
\]

It is remarkable that, independent of the odd dimension \(D \geq 5\), this parametrization involves only four functions \(f, m, n, \omega\) for the metric and two functions \(a_0, a_\phi\) for the gauge field, which all depend only on the radial coordinate \(r\).

3. Black hole properties

3.1. Boundary conditions

By substituting Eqs. (5), (6) in Eqs. (2)–(4), a system of 6 ordinary differential equations is obtained. In order to generate black hole solutions, appropriate boundary conditions have to be imposed.

For asymptotically flat solutions, the metric functions should satisfy at infinity the boundary conditions

\[
f|_{r=\infty} = m|_{r=\infty} = n|_{r=\infty} = 1, \quad \omega|_{r=\infty} = 0. \tag{7}
\]

For the gauge potential we choose a gauge, in which it vanishes at infinity

\[
a_0|_{r=\infty} = a_\phi|_{r=\infty} = 0. \tag{8}
\]

The horizon is located at \(r = r_H\), and is characterized by the condition \(f(r_H) = 0\) [17]. Requiring the horizon to be regular, the metric functions must satisfy the boundary conditions

\[
f|_{r=r_H} = m|_{r=r_H} = n|_{r=r_H} = 0, \quad \omega|_{r=r_H} = r_H \Omega, \tag{9}
\]

where \(\Omega\) is (related to) the horizon angular velocity, defined in terms of the Killing vector

\[
\chi = \xi + \Omega \sum_{k=1}^{N} \epsilon_k \eta_{(k)}, \tag{10}
\]

which is null at the horizon. The gauge potential satisfies

\[
\chi A_\mu|_{r=r_H} = \Phi_H = (a_0 + \Omega a_\phi)|_{r=r_H}, \quad \frac{da_\phi}{dr}|_{r=r_H} = 0, \tag{11}
\]

with constant horizon electrostatic potential \(\Phi_H\).
3.2. Global charges

Since the space–times we are considering are stationary, (multi-)axisymmetric, and asymptotically flat we may compute the mass \( M \) and the \( N \) angular momenta \( J_{(k)} \) of the black holes by means of the Komar expressions associated with the respective Killing vector fields

\[
M = \frac{-1}{16\pi G_D} \frac{D - 2}{D - 3} \int_{S^{D-2}} \alpha, \quad J_{(k)} = \frac{1}{16\pi G_D} \int_{S^{D-2}} \beta_{(k)},
\]

with \( \alpha_{\mu_1 \cdots \mu_{D-2}} \equiv \epsilon_{\mu_1 \cdots \mu_{D-2} \rho \sigma} \nabla^\rho \epsilon^\sigma \) and \( \beta_{(k)\mu_1 \cdots \mu_{D-2}} \equiv \epsilon_{\mu_1 \cdots \mu_{D-2} \rho \sigma} \nabla^\rho \eta^\sigma_{(k)} \). For equal-magnitude angular momenta \( J_{(k)} = \epsilon_k J, k = 1, \ldots, N \).

The electric charge \( Q \) associated with the Maxwell field can be defined by

\[
Q = \frac{-1}{8\pi G_D} \int_{S^{D-2}} \hat{F},
\]

with \( \hat{F}_{\mu_1 \cdots \mu_{D-2}} \equiv \epsilon_{\mu_1 \cdots \mu_{D-2} \rho \sigma} F^{\rho \sigma} \).

These global charges and the magnetic moment \( \mu_{\text{mag}} \) can be obtained from the asymptotic expansions of the metric and the gauge potential

\[
f = 1 - \frac{\hat{M}}{r^{D-3}}, \ldots, \quad \omega = \frac{j}{r^{D-3}}, \ldots,
\]

\[
a_0 = \frac{\hat{Q}}{r^{D-3}}, \ldots, \quad a_\phi = -\frac{\mu_{\text{mag}}}{r^{D-3}}, \ldots,
\]

where

\[
\hat{M} = \frac{16\pi G_D}{(D - 2)A(S^{D-2})} M, \quad \hat{J} = \frac{8\pi G_D}{A(S^{D-2})} J,
\]

\[
\hat{Q} = \frac{4\pi G_D}{(D - 3)A(S^{D-2})} Q, \quad \hat{\mu}_{\text{mag}} = \frac{4\pi G_D}{(D - 3)A(S^{D-2})} \mu_{\text{mag}},
\]

and \( A(S^{D-2}) \) denotes the area of the unit \((D - 2)\)-sphere. The gyromagnetic ratio \( g \) is then defined via

\[
g = 2\frac{M_{\mu_{\text{mag}}}}{Q J}.
\]

3.3. Mass formulae

The surface gravity \( \kappa \) of the black holes is defined by

\[
\kappa^2 = \frac{1}{2} \left( \lim_{r \to r_H} (\nabla_\mu \chi_\nu)(\nabla^\mu \chi^\nu) \right).
\]

For equal-magnitude angular momenta, the area of the horizon \( A_H \) reduces to

\[
A_H = \frac{D - 2}{H} A(S^{D-2}) \lim_{r \to r_H} \sqrt{\frac{m^{D-3} n}{f^{D-2}}}. \tag{18}
\]

The horizon mass \( M_H \) and horizon angular momenta \( J_{H(k)} \) are given by

\[
M_H = -\frac{1}{16\pi G_D} \frac{D - 2}{D - 3} \int_{\mathcal{H}} \alpha, \quad J_{H(k)} = \frac{1}{16\pi G_D} \int_{\mathcal{H}} \beta_{(k)},
\]

where \( \mathcal{H} \) represents the surface of the horizon, and for equal-magnitude angular momenta \( J_{H(k)} = \epsilon_k J_H, k = 1, \ldots, N \). The mass \( M \) and angular momenta \( J_{(k)} \) may thus be reexpressed in the form \([13,14]\)

\[
M = M_H + M_\Sigma, \quad J_{(k)} = J_{H(k)} + J_{\Sigma(k)},
\]

where \( J_{\Sigma(k)} \) is a ‘bulk’ integral over the region of \( \Sigma \) outside the horizon, i.e., \( \Sigma \) is a space-like hypersurface with boundaries at spatial infinity and on the horizon.

The black holes satisfy the horizon mass formula

\[
\frac{D - 3}{D - 2} M_H = \frac{\kappa A_H}{8\pi G_D} + N \Omega J_H.
\]

They further satisfy the Smarr-like mass formula \([13]\),

\[
M = \frac{D - 2}{D - 3} \frac{\kappa A_H}{8\pi G_D} + \frac{D - 2}{D - 3} N \Omega J + \Phi_H Q + \frac{D - 5}{D - 3} \tilde{\lambda}, \tag{22}
\]

where \( I \) denotes the integral

\[
I = -\frac{1}{4\pi G_D} \int_{\Sigma} dS_\sigma \chi^\nu F_{\nu\rho} J^{\rho\sigma},
\]

and \( J^{\rho\sigma} \) is defined by

\[
J^{\rho\sigma} = -\epsilon^{\rho\sigma \mu_1 \mu_2 \cdots \mu_D - 3 \mu_D - 2} A_{\mu_1} F_{\mu_2 \mu_3} \cdots F_{\mu_D - 3 \mu_D - 2}.
\]

3.4. Scaling symmetry

The system of ODEs is invariant under the scaling transformation

\[
r_H \to \gamma r_H, \quad \Omega \to \Omega / \gamma, \quad \tilde{\lambda} \to \gamma^{N - 2} \tilde{\lambda},
\]

\[
Q \to \gamma^{D - 3} Q, \quad a_\phi \to \gamma a_\phi,
\]

with \( \gamma > 0 \). We note that in two cases the scaling transformation does not change the CS coupling constant, namely, in the case \( \tilde{\lambda} = 0 \) for arbitrary dimension \( D \) (i.e., in pure EM theory), and in the case \( D = 5 \) for arbitrary \( \tilde{\lambda} \). This is in accordance with the mass formula, Eq. (22). In both cases, the mass formula reduces to the standard Smarr formula, since the last term vanishes. Indeed, this is not accidental, but both features rely on the fact that the CS coupling constant is dimensionless only for \( D = 5 \), unless it vanishes \([13]\).

4. Numerical results

Apart from the case \( D = 5 \), \( \tilde{\lambda} = \tilde{\lambda}_{SG} = 1/(2\sqrt{3}) \) \([11,12]\), no charged rotating EMCS black hole solutions with spherical horizon topology are known analytically. Here we address the problem of finding such solutions numerically, and discuss their properties.
4.1. Numerical procedure

Owing to the existence of the first integral of the system of ODE’s

\[
\frac{r^{D-2}m^{N-2}}{f^{N-1}} \sqrt{\frac{m}{f}} \left( \frac{dd_0}{dr} + \omega \frac{da_\psi}{dr} \right) - \lambda \delta_D 2^{D-2}(N-1)d_\psi = -\frac{4\pi G_D}{A(SD-2)} Q, \tag{26}
\]

we eliminate \(a_0\) from the equations, replacing it in terms of the electric charge. This leaves a system of one first order equation (for \(n\)) and four second order equations (for \(f, m, \omega, \) and \(a_\psi\)). In Eq. (26), \(\delta_D\) is just a sign, depending on the dimension \(D\), and given by the expression

\[
\delta_D \equiv (-1)^{N(N+1)/2} = (-1)^{(D^2-1)/8}. \tag{27}
\]

For the numerical calculations we then introduce the compactified radial coordinate \(\tilde{r} = 1 - r_{H}/r\) [17], and we take units such that \(G_D = 1\). We employ a collocation method for boundary-value ordinary differential equations, equipped with an adaptive mesh selection procedure [18]. Typical mesh sizes include \(10^3\)–\(10^4\) points. The solutions have a relative accuracy of \(10^{-8}\). The set of numerical parameters to be fixed for a particular solution is \([r_H, \Omega, Q, \lambda]\). By varying these parameters we generate families of EMCS black holes.

The scaling symmetry Eq. (25) leads us to consider \(D = 5\) black holes and black holes in \(D > 5\) dimensions separately. Since odd-dimensional EM black holes were investigated previously [10], we here concentrate on \(\tilde{\lambda} \neq 0\).

Moreover, since the action Eq. (1) is invariant under the transformation

\[
[\tilde{\lambda}, Q] \rightarrow \{-1\}^{N-1} [\tilde{\lambda}, -Q], \tag{28}
\]

it is sufficient to consider just two of the four possible sign combinations of \([\tilde{\lambda}, Q]\).

4.2. \(D = 5\) EMCS black holes

For convenience we redefine the CS coupling constant,

\[
\lambda = 2\sqrt{3} \tilde{\lambda}, \tag{29}
\]

for \(D = 5\). Additionally, owing to Eq. (28), without loss of generality we may choose \(\lambda \geq 0\).

\(D = 5\) EMCS black holes have been considered before for CS coupling \(\lambda_{SG}\) [11–14] and for general coupling [15]. To briefly recall some of their main features, we exhibit in Fig. 1(a) the domain of existence of these black holes. Here the scaled angular momentum \(|J|/M^{3/2}\) of extremal EMCS black holes is shown versus the scaled charge \(Q/M^{1}\) for four values of \(\lambda\): the pure EM value \(\lambda_{EM} = 0\), the supergravity value \(\lambda_{SG} = 1\), the second critical value \(\lambda_c = 2\), and a value beyond this critical value, \(\lambda = 3\). For a given value of \(\lambda\), black holes exist only in the regions bounded by the \((J = 0)\)-axis and by the respective curves. For fixed finite \(\lambda\), there is an explicit asymmetry between solutions with positive and negative electric charge. The properties of EMCS black holes with positive \(Q\) are similar to those of EM black holes [10], whereas for EMCS black holes with negative \(Q\) surprising features are present.

The figure also shows stationary black holes with nonrotating horizon, i.e., black holes with horizon angular velocity \(\Omega = 0\) but finite total angular momentum. Such solutions are present only for \(\lambda \geq 1\). For \(\lambda = 1\) they are extremal solutions, and form the vertical part of the \(Q < 0\) borderline of the domain of existence. For \(\lambda > 1\), however, they are (with the exception of endpoints) non-extremal black holes, located in the interior of the respective domains of existence. In fact, the \(\Omega = 0\) lines divide these domains into two parts. The right part contains ordinary black holes, where the horizon rotates in the same sense as the angular momentum, while the left part contains black holes with unusual properties. When \(1 < \lambda < 2\), all black holes in this region are counterrotating, i.e., their horizon rotates in
the opposite sense to the angular momentum [15]. When \( \lambda > 2 \), further intriguing features appear.

As long as \( 1 < \lambda < 2 \) only static \( J = 0 \) solutions exist, and consequently, the static extremal black holes are located on the borderline of the domain of existence. (Recall that static black holes are independent of \( \lambda \).) However, beyond \( \lambda = 2 \), a continuous set of rotating \( J = 0 \) solutions appears [15], whose existence relies on a special partition of the total angular momentum \( J \), where the horizon angular momentum \( J_H \) is equal and opposite to the angular momentum \( J_S \) in the Maxwell field outside the horizon.\(^2\) The static extremal black holes then no longer mark the lower left border of the domain of existence, which is now formed by extremal rotating \( J = 0 \) black holes, as seen in Fig. 1(a).

Moreover, beyond \( \lambda = 2 \) the set of extremal solutions not only forms the boundary of the domain of existence, but continues well within this domain, until a static extremal black hole is reached in a complicated pattern of bifurcating branches. Insight into this set is gained in Fig. 1(b), where (almost) extremal solutions are exhibited for CS coefficient \( \lambda = 3 \), together with non-static \( \Omega = 0 \) solutions. Note, that all this new structure arises well within the counterrotating region, in the vicinity of the static extremal black holes.

To explore the properties of \( \lambda > 2 \) EMCS black holes further, let us now consider non-extremal black holes. We exhibit in Fig. 2 a set of solutions for \( \lambda = 3 \), possessing constant charge \( Q = -10 \) and constant (isotropic) horizon radius \( r_H = 0.1 \). Fig. 2(a) and 2(b) show the total angular momentum \( J \) and the horizon angular momentum \( J_H \), respectively, as functions of the horizon angular velocity \( \Omega \). Fig. 2(c) and 2(d) show the corresponding masses \( M \) and \( M_H \), and Fig. 2(e) and 2(f) the gyromagnetic ratio \( g \) and the horizon area \( A_H \).

Fig. 2(a) exhibits the four types of rotating black holes, as classified by their total angular momentum \( J \) and horizon angular velocity \( \Omega \): Type I black holes correspond to the corotating regime, i.e., \( \Omega J > 0 \), and \( \Omega = 0 \) if and only if \( J = 0 \) (static). Type II black holes possess a static horizon (\( \Omega = 0 \)), although their total angular momentum does not vanish (\( J \neq 0 \)). Type III black holes are characterized by counterrotation, i.e., the horizon angular velocity and the total angular momentum have opposite signs, \( \Omega J < 0 \). Type IV black holes, finally, possess a rotating horizon (\( \Omega 
eq 0 \)) but vanishing total angular momentum (\( J = 0 \)).

As seen in Fig. 2(b), the horizon angular momentum \( J_H \) of these black holes need not have the same sign as the total angular momentum \( J \), and neither does the ‘bulk’ angular momentum \( J_S \). In fact, as the horizon of the black hole is set into rotation, angular momentum is stored in the Maxwell field both behind and outside the horizon, yielding a rich variety of configurations. Starting from the static solution, a corotating branch evolves, along which \( J_H \) and \( J_S \) have opposite signs. After the first bifurcation \( \Omega \) moves back towards zero and so does \( J \), but both \( J_H \) and \( J_S \) remain finite, retaining part of their built up angular momentum and thus their memory of the path, like in a hysteresis. This is important, since when moving \( \Omega \) continuously back to and beyond zero, the total angular momentum follows and changes sign as well. The horizon angular momentum, however, retains its sign. Thus the product \( \Omega J_H \) turns negative and remains negative up to the next bifurcation and still further, until \( \Omega \) reaches zero again.

The observation, that solutions with \( \Omega J_H < 0 \) can be present, is crucial to explain the occurrence of black holes with negative horizon mass, \( M_H < 0 \), exhibited in Fig. 2(d). The correlation between \( \Omega J_H < 0 \) and \( M_H < 0 \) black holes is evident here. In fact, the sets of \( \Omega J_H < 0 \) and \( M_H < 0 \) black holes almost coincide, when \( \kappa A_H \) is small, as seen from the horizon mass formula Eq. (21). The angular momentum stored in the Maxwell field behind the horizon is thus responsible for the negative horizon mass of the black holes. The total mass is always positive, however, as seen in Fig. 2(c).

The correlation between \( \Omega J_H < 0 \) and \( M_H < 0 \) is further demonstrated in Fig. 3, where we exhibit \( J_H \) and \( M_H \) for black holes with smaller as well as larger (isotropic) horizon radii \( r_H \). Whereas for larger values of \( r_H \) the set of negative horizon mass black holes decreases, it increases for smaller values of \( r_H \). In fact, as \( r_H \) decreases, more branches of solutions appear in the vicinity of the static solution, giving rise to more branches of negative horizon mass black holes.

Another interesting feature of these charged rotating EMCS black holes is their gyromagnetic ratio \( g \), exhibited in Fig. 2(e). The gyromagnetic ratio is unbounded, reaching any real value including zero. The main consequence of this is that, contrary to pure EM theory, a vanishing total angular momentum does not readily imply a vanishing magnetic moment and vice versa.

4.3. \( D > 5 \) EMCS black holes

Unlike in \( D = 5 \), the CS coupling constant is dimensionful when \( D > 5 \), and consequently changes under scaling transformations Eq. (25), unless \( \lambda = 0 \). Thus any feature present for a certain non-vanishing value of \( \lambda \) will be present for any other non-vanishing value, when the charge is scaled correspondingly. To classify the critical behavior of the solutions we therefore consider classes of solutions labeled by the scale invariant ratio \( \lambda_Q \),

\[
\lambda_Q = \lambda/Q^{(D-5)/(D-3)}.
\]  

(30)

Otherwise, black holes of \( D > 5 \) EMCS theories exhibit very similar properties to \( D = 5 \) EMCS black holes. Besides regions containing black holes with features resembling those of EM black holes, there are regions where black holes reside, possessing all those new features. Assuming \( Q < 0 \), the sign of \( \lambda \) associated with the appearance of these new types of black holes coincides with \( -\tilde{E}_D \) (while for \( Q > 0 \) one needs to employ Eq. (28)).

We exhibit in Fig. 4(a) sets of extremal 7D solutions, choosing the same value of the charge in all sets, while varying \( \lambda \). In 7D, the first critical value is \( \lambda_Q = -0.0414 \). Here the first \( \Omega = 0 \) solution appears. Counterrotating black holes then exist only for values of the ratio \( \lambda_Q \) above this critical value. The

\(^2\) The numerical data indicate, that at \( \lambda = 2 \) a continuous set of extremal rotating \( J = 0 \) black holes with constant mass is present [15].
second critical value characterizes the limit above which non-static \( J = 0 \) solutions are present, and where solutions are no longer uniquely specified by their global charges. In 7D this value is \( \lambda Q = -0.2101 \). We note, that the sets of \( \Omega = 0 \) solutions, also exhibited in the figure, always connect extremal solutions, whose mass assumes an extremal value as well, in accordance with the first law. At the second critical value of \( \lambda Q \), the lower endpoint of the \( \Omega = 0 \) set of solutions precisely reaches a static extremal black hole.

In Fig. 5 we demonstrate that the four types of rotating black holes, as classified by their total angular momentum \( J \) and horizon angular velocity \( \Omega \), are also present for \( D > 5 \). Interestingly, in \( D = 9 \), beside I corotating black holes, II non-static \( \Omega = 0 \) black holes, III counterrotating black holes, IV non-
Fig. 3. Properties of 5D non-extremal $\lambda = 3$ EMCS black holes with charge $Q = -10$ and horizon radii $r_H = 0.03, 0.1, 0.2$ and 0.5. (a) Horizon angular momentum $J_H$, (b) horizon mass $M_H$ versus horizon angular velocity $\Omega$.

Fig. 4. (a) Scaled angular momentum $J/M^{5/4}$ versus scaled charge $Q/M$ for (almost) extremal 7D EMCS black holes and $\Omega = 0$ solutions for $\lambda_Q = 0, -0.0414, -0.0562, -0.2101, -0.3374$ and charge $Q = 10$. (b) Magnetic moment $\mu_{\text{mag}}$ versus total angular momentum $J$ for 9D EMCS black holes with $\Omega = 0$ for $\tilde{\lambda} = 0.5$ and charge $Q = 10$.

Fig. 5. Angular momentum $J$ of 7D non-extremal $\tilde{\lambda} = -0.6$ EMCS black holes with charge $Q = 10$ (a) versus horizon angular velocity $\Omega$ for horizon radius $r_H = 0.03$, (b) versus mass $M$ for horizon radii $r_H = 0.03, 0.7$. 
static $J = 0$ black holes, a further type of black holes appears: V non-static $\Omega = J = 0$ black holes. This is seen in Fig. 4(b), where the magnetic moment is exhibited for a set of 9D $\Omega = 0$ black holes. Clearly, there appears a non-static $\Omega = J = 0$ solution with finite magnetic moment. We further illustrate in Fig. 5(b) that these EMCS black holes are no longer uniquely characterized by their global charges, i.e., the uniqueness conjecture does not hold in general for $D \geq 5$ EMCS stationary black holes with horizons of spherical topology.

We illustrate in Fig. 6 that $D > 5$ EMCS black holes can also possess negative horizon mass $M_H$, by exhibiting sets of $D = 7$ and $D = 9$ black holes. As in $D = 5$ dimensions, the angular momentum stored in the Maxwell field behind the horizon is responsible for this phenomenon. Finally, the gyromagnetic ratio $g$ is unbounded as well for $D > 5$ EMCS black holes, reducing the correlation between angular momentum and magnetic moment.

5. Conclusions

We have considered charged rotating black holes of EMCS theory in odd dimensions, which are asymptotically flat and possess a regular horizon of spherical topology. When all their angular momenta have equal-magnitude, a system of 5 ordinary differential equations results [10], which we have solved numerically.

EMCS black holes exhibit remarkable features, not present for EM black holes. Classifying the EMCS black holes by their total angular momentum $J$ and horizon angular velocity $\Omega$, we find several types of black holes: I corotating black holes, i.e., $\Omega J \geq 0$, II black holes with static horizon and non-vanishing total angular momentum, i.e., $\Omega = 0, J \neq 0$, III counterrotating black holes, where the horizon angular velocity and the total angular momentum have opposite signs, i.e., $\Omega J < 0$, and IV black holes with rotating horizon and vanishing total angular momentum, i.e., $\Omega \neq 0, J = 0$. Furthermore, in 9D Type V black holes appear: non-static black holes with static horizon and vanishing total angular momentum, i.e., $\Omega = J = 0$.

As the horizon of static EMCS black holes is set into rotation, angular momentum is stored in the Maxwell field both behind and outside the horizon. Following paths through configuration space, the horizon angular momentum $J_H$ and the 'bulk' angular momentum $J_2$ can retain part of the angular momentum built up in the Maxwell field, even when the horizon angular velocity vanishes again. Thus they retain the memory of the path, like a hysteresis. Consequently, solutions with $\Omega J_H < 0$ appear, which possess a negative horizon mass, $M_H < 0$, as long as $\kappa A_H$ is sufficiently small. The angular momentum stored in the Maxwell field behind the horizon is therefore responsible for the negative horizon mass of these black holes. Their total mass is always positive, however.

Moreover, EMCS black holes are no longer uniquely characterized by their global charges, i.e., the uniqueness conjecture does not hold in general for $D \geq 5$ EMCS stationary black holes with horizons of spherical topology, and the gyromagnetic ratio of these black holes may take any real value, including zero.

Acknowledgement

F.N.-L. gratefully acknowledges Ministerio de Educación y Ciencia for support under grant EX2005-0078.

References


3 Previous counterexamples involved black rings [16].

4 The occurrence of a negative horizon mass has been also reported recently in the context of 4D black holes surrounded by perfect fluid rings [19].
Logarithmic terms in brick wall model

Zhong-Heng Li

Department of Physics, Zhejiang University of Technology, Hangzhou 310032, China
Department of Physics, Zhanjiang Normal College, Zhanjiang 524048, China

Received 19 October 2005; received in revised form 18 September 2006; accepted 17 October 2006
Available online 7 November 2006
Editor: W. Haxton

Abstract

We clarify the relation between black hole entropy and quantum field spin. Starting from the Newman–Penrose formalism, we derive the Teukolsky-type master equation governing massless fields of arbitrary spin $s \leq 2$ in the Kerr–Newman–de Sitter spacetime. Then using the 't Hooft’s brick wall model, we calculate the entropy of the black hole due to the quantum fields by the WKB approximation on the Teukolsky-type master equation. In particular, we carefully deal with the logarithmic term contribution to the entropy. It is shown that the term not only depends on the characteristics of the metric and the quadratic term of quantum field spin but also on the linear term of quantum field spin. This is very different from the result obtained earlier. All results of the present work are valid for the Schwarzschild, Reissner–Nordström, Kerr, Kerr–Newman, and de Sitter spacetime background, or any combination of these.

© 2006 Elsevier B.V. All rights reserved.
PACS: 04.70.Dy; 04.62.+v; 97.60.Lf

1. Introduction

It is now over 20 years since 't Hooft [1,2] introduced the brick wall model. In this model, the black hole entropy includes the entropy of the quantized fields in its neighborhood. The calculations for various black holes were given that a quantum field has an entropy which is proportional to the event horizon area (see for example Refs. [3–11] and references cited therein). In fact, this is the case if one considers only the leading term. In 1994, Solodukhin [12], starting from the one-loop effective action for scalar matter, investigated the quantum corrections to Rindler space and Schwarzschild black-hole entropy by means of the path integral approach of Gibbons and Hawking [13], and discovered first the divergent logarithmic correction to the entropy. Then, Solodukhin [14] studied the general divergence structure of the entropy for static black holes. While the divergences of a charged Kerr black hole due to scalar field were obtained by Mann and Solodukhin [15]. On the other hand, we [16] studied the quantum corrections to the entropy of the Reissner–Nordström black hole due to spin fields by using the brick wall model, and found that the entropy of spin fields has both linearly and logarithmically divergent terms, where linear divergence can reduce to the form proportional to the event horizon area, while the logarithmic divergences explicitly depend on the spin of the field considered. Therefore the whole expression of the entropy does not take the form of the scalar field.

Although logarithmic corrections to the black hole entropy due to arbitrary spin fields have been studied by different authors in different types of spacetime so far, there has been some uncertainty for spin-dependent terms. Jing and Yan [17–19], and López-Ortega [20] calculated the entropy up to subleading terms of massless fields of spin $s = 1/2, 1, 3/2$, and 2 for the Kerr–(Newman) black hole and pointed out that the subleading logarithmic terms depend on the spins of the fields just in quadratic term $s^2$. However the coefficients of the quadratic spin terms in the expressions of entropy presented in Refs. [17–19] and [20] are incompatible with each other (spin-dependent terms will decrease the entropy in Refs. [17–19]; and increase the entropy in Ref. [20]). Wu
and Yan [21] investigated the entropy of arbitrary spin fields on the Kerr–de Sitter black hole background and showed that the logarithmic correction to the entropy depends quadratically on the spins of the fields, as in Refs. [17–20]. But the coefficient of the spin-dependent term does not reduce to the case discussed by López-Ortega or Jing and Yan.

In this Letter, we extend the brick wall model to the Kerr–Newman–de Sitter black hole case and show logarithmic terms not only depend on the quadratic term of the spins of fields but also on the linear terms of the spins of fields. We also analyze the reason why linear spin-dependent terms disappear in the results obtained earlier, and give some arguments about the above-mentioned difference. The Letter organized as follows: In Section 2 we study the main properties of the Kerr–Newman–de Sitter spacetime, including its Petrov classification. In Section 3 using the Newman–Penrose formalism, we derive a master equation describing spin \( s \leq 2 \) test fields. In Section 4 the entropies of the spin fields for nonextremal Kerr–Newman–de Sitter black hole background are studied in the brick wall model, while we conclude in Section 5.

2. Kerr–Newman–de Sitter spacetime

The Kerr–Newman–de Sitter spacetime can be expressed in Boyer–Lindquist-type coordinates as [22,23]

\[
ds^2 = \frac{\rho \bar{\rho} \Delta_r}{\Xi^2} (dt - a \sin^2 \theta d\varphi)^2 - \frac{\rho \bar{\rho} \Delta_\theta \sin^2 \theta}{\Xi^2} \left[ a dt - (r^2 + a^2) d\varphi \right]^2 - \frac{1}{\rho \bar{\rho}} \left( \frac{dr^2}{\Delta_r} + \frac{d\theta^2}{\Delta_\theta} \right),
\]

where

\[
\rho = -\frac{1}{r - ia \cos \theta}, \quad \Delta_r = (r^2 + a^2) \left( 1 - \frac{A}{3} r^2 \right) - 2Mr + Q^2, \quad \Delta_\theta = 1 + \frac{A}{3} a^2 \cos^2 \theta, \quad \Xi = 1 + \frac{A}{3} a^2.
\]

Here \( A \) is the cosmological constant, \( M \) is the mass of the black hole, \( Q \) is its charge and \( a \) is the angular momentum per unit mass. The metric determinant is \( g = -\left( \rho \bar{\rho} \right)^{-2} \Xi^{-4} \sin^2 \theta \). The inverse metric \( g^{\mu \nu} \) is given by

\[
g^{\mu \nu} = -\rho \bar{\rho} \left[ \Delta_r \delta_1^\mu \delta_1^\nu + \Delta_\theta \delta_2^\mu \delta_2^\nu + \frac{\Xi^2}{\sin^2 \theta} \left( \frac{1}{\Delta_\theta} - \frac{a^2 \sin^2 \theta}{\Delta_r} \right) \delta_3^\mu \delta_3^\nu \\
- a \Xi^2 \left( \frac{r^2 + a^2}{\Delta_r} - \frac{1}{\Delta_\theta} \delta_0^\mu \delta_3^\nu + \delta_3^\mu \delta_0^\nu \right) - \Xi^2 \left( \frac{(r^2 + a^2)^2}{\Delta_r} - \frac{a^2 \sin^2 \theta}{\Delta_\theta} \delta_0^\mu \delta_0^\nu \right) \right].
\]

The event horizon is located at \( r = r_H \), the larger of the three positive roots of the polynomial \( \Delta_r \). The other two roots may be interpreted as the Cauchy horizon and cosmological horizon. The area of the event horizon is

\[
A_H = \frac{4\pi (r_H^2 + a^2)}{\Xi}.
\]

The surface gravity \( \kappa_H \) and Hawking inverse temperature \( \beta_H \) are given by

\[
\beta_H = \frac{2\pi}{\kappa_H} = \frac{4\pi \Xi (r_H^2 + a^2)}{|\Delta_{r_H}|},
\]

where the prime denotes the partial differential with respect to its argument. The angular velocity of the event horizon is defined by

\[
\Omega_H = -\frac{g_{\varphi \varphi}}{g_{\varphi \varphi}} \bigg|_{r=r_H} = \frac{a}{r_H^2 + a^2}.
\]

The null tetrad vectors for the Kerr–Newman–de Sitter spacetime are chosen as

\[
l^\mu = \left[ \frac{(r^2 + a^2) \Xi}{\Delta_r}, 1, 0, \frac{a \Xi}{\Delta_r} \right],
\]

\[
n^\mu = \frac{\rho \bar{\rho}}{2} \left[ \frac{(r^2 + a^2) \Xi}{-\Delta_r}, 0, 0, \frac{a \Xi}{\sin \theta} \right],
\]

\[
m^\mu = -\frac{\bar{\rho}}{\sqrt{2} \Delta_\theta} \left[ i a \Xi \sin \theta, 0, \Delta_\theta, \frac{i \Xi}{\sin \theta} \right],
\]
where \( l^\mu \) and \( n^\mu \) are real null vectors, and \( m^\mu \) with its complex conjugate \( \tilde{m}^\mu \) are complex null vectors with the normalization condition \( l_\mu n^\mu = -m_\mu \tilde{m}^\mu = 1 \). Using these null tetrad vectors we calculate the spin coefficients and the components of the Weyl tensor as follows:

\[
\begin{align*}
\kappa &= \epsilon = \lambda = \sigma = v = 0, \\
\rho &= -\frac{1}{r - ia \cos \theta}, \\
\tau &= -i \sqrt{\frac{\Delta_\theta}{2}} \rho \tilde{\rho} a \sin \theta, \\
\mu &= \frac{1}{2} \rho^2 \tilde{\rho} \Delta_r, \\
\beta &= -\frac{\sqrt{\Delta_\theta} \tilde{\rho}}{2 \sqrt{2}} (\frac{\Delta'_\theta}{2 \Delta_\theta} + \cot \theta), \\
\gamma &= \frac{\rho \tilde{\rho}}{4} \Delta'_r + \mu, \\
\pi &= i \rho^2 a \sqrt{\frac{\Delta_\theta}{2}} \sin \theta, \\
\alpha &= \pi - \tilde{\pi}, \\
\Psi_0 &= \Psi_1 = \Psi_3 = \Psi_4 = 0, \\
\Psi_2 &= \rho^2 (M + \tilde{\rho} Q^2).
\end{align*}
\]

Eq. (8) tells us that the Kerr–Newman–de Sitter metric is of Petrov type D.

3. Teukolsky-type master equation

In a type-D spacetime, the equations of the Weyl neutrino \((s = 1/2)\), electromagnetic \((s = 1)\), massless Rarita–Schwinger \((s = 3/2)\), and gravitational \((s = 2)\) fields for the source free case can be combined into [16,24]

\[
\begin{align*}
&[[D - (2s - 1)\epsilon + \tilde{\epsilon} - 2s \rho - \tilde{\rho}] (\Delta - 2s \gamma + \mu) - [\delta + \tilde{\pi} - \tilde{\alpha} - (2s - 1)\beta - 2s \tau] (\delta + \pi - 2s \alpha) \\
&- (2s - 1)(s - 1)\Psi_2)\Phi_{s+0} = 0, \\
&[[\Delta + (2s - 1)\gamma - \tilde{\gamma} + 2s \mu + \tilde{\mu}] (D + 2s \epsilon - \rho) - [(\delta - \tilde{\epsilon} + \tilde{\beta} + (2s - 1)\alpha + 2s \tau)] (\delta - \tau + 2s \beta) \\
&- (2s - 1)(s - 1)\Psi_2)\Phi_{-s} = 0,
\end{align*}
\]

where

\[
D = l^\mu \partial_\mu, \quad \Delta = n^\mu \partial_\mu, \quad \delta = m^\mu \partial_\mu.
\]

In Eq. (9) the first equation is for spin states \( p = s \), while the other one is for \( p = -s \). Eq. (9) is also valid when \( s = 0 \); either coincide with the (conformally invariant) massless scalar field equation

\[
(\nabla^\mu \nabla_\mu + \frac{1}{6} R) \Phi_0 = 0,
\]

with \( \Phi_0 = \Phi_{s+0} = \Phi_{-s} \). Here \( R \) is the scalar curvature of the spacetime.

A straightforward computation, using the spin coefficients, the components of the Weyl tensor and the directional derivatives written down in Eqs. (8) and (10), and making the transformations

\[
\Phi_{s+0}, \Phi_{-s} = \rho^s \Psi_p,
\]

Eq. (9) can be written as

\[
\begin{align*}
&\left\{ \Delta_r^{-p} \frac{\partial}{\partial r} \left( \Delta_r^{p+1} \frac{\partial}{\partial r} \right) + \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left( \Delta_\theta \sin \theta \frac{\partial}{\partial \theta} \right) - \Xi^2 \left( \frac{a^2}{\Delta_r} - \frac{1}{\Delta_\theta \sin^2 \theta} \right) \frac{\partial^2}{\partial \phi^2} \\
&+ 2p \Xi \left[ a \Delta'_r \frac{\Delta_\theta}{2 \Delta_r} + \frac{i}{\sin \theta} \left( \frac{\Delta'_\theta}{2 \Delta_\theta} + \cot \theta \right) \right] \frac{\partial}{\partial \phi} - 2a \Xi \left( \frac{a^2 + r^2}{\Delta_r} - 1 \right) \frac{\partial}{\partial \theta} \frac{\partial}{\partial \phi} \\
&- \Xi^2 \left( \frac{a^2 + r^2}{\Delta_r} - \frac{a^2 \sin^2 \theta}{\Delta_\theta} \right) \frac{\partial^2}{\partial \theta^2} + 2p \Xi \left[ ia \sin \theta \left( \frac{\Delta'_\theta}{2 \Delta_\theta} - \cot \theta \right) + \frac{\Delta'_r}{2 \Delta_r} (a^2 + r^2) - 2a \right] \frac{\partial}{\partial r} \\
&+ \frac{1}{2} \rho \Delta'' - p^2 \Delta_\theta \left( \frac{\Delta'_\theta}{2 \Delta_\theta} + \cot \theta \right)^2 - \frac{2}{3} \Lambda (2p^2 + 1) (a^2 + r^2 \cos^2 \theta) \right\} \Psi_p = 0.
\end{align*}
\]

In 1973, Teukolsky [25], using the Newman–Penrose formalism, succeeded in disentangling the perturbations of the Kerr metric, and wrote a master equation for the massless scalar, Weyl neutrino, electromagnetic and gravitational fields. Here, in the Kerr–Newman–de Sitter spacetime, we have derived the master equation (13) governing not only these spin fields, but the massless Rarita–Schwinger field as well. It should be noted that when \( Q = 0 \), we recover the Kerr–de Sitter case [21], and when \( Q = 0 \), \( \Lambda = 0 \), our results reduce to the case discussed by Teukolsky. Therefore Eq. (13) can be named as the Teukolsky-type master equation.
At the WKB level, the solution of the field equation of arbitrary spin may be written as

$$\Phi_{+s}, \Phi_{-s} = e^{iS(t,r,\theta,\phi,p)}.$$  

(14)

Applying Eq. (14) to the Teukolsky-type master equation (13) and separating the real and imaginary parts, we obtain

$$g^{\mu\nu} P_\mu P_\nu - 2B^{\mu} P_\mu + V = 0,$$

(15)

where

$$B^t = p \rho a \sin \theta \left( \frac{\Delta_{\theta}^t}{2\Delta_{\theta}} - \cot \theta \right),$$

$$B^r = (s - p)(\rho \tilde{p})^2 \Delta_{\theta} a \cos \theta,$$

$$B^\theta = (s - p)(\rho \tilde{p})^2 \Delta_{\theta} a \sin \theta,$$

$$B^\phi = \frac{p \rho a \sin \theta}{\sin \theta} \left( \frac{\Delta_{\theta}^\phi}{2\Delta_{\theta}} + \cot \theta \right),$$

$$V = (s - p)(\rho \tilde{p})^2 \left[ (s - p - 1) \rho \tilde{p} (r^2 - a^2 \cos^2 \theta) (\Delta_r - \Delta_{\theta} a^2 \sin^2 \theta) + (p + 1) \Delta_{\theta}^r - 2\Delta_{\theta} a^2 \sin \theta \cos \theta \left( \frac{\Delta_{\theta}^\phi}{2\Delta_{\theta}} + \cot \theta \right) \right]$$

$$\quad + \frac{1}{2} p \rho a \Delta_{\theta}^r - p^2 \chi_{\theta} \Delta_{\theta} \left( \frac{\Delta_{\theta}^\phi}{2\Delta_{\theta}} + \cot \theta \right)^2 - \frac{2}{3} \Lambda (2p^2 + 1),$$

(16)

and $P_\mu = \partial_\mu S$ are the conjugate momenta. Since the existence of two Killing vectors $\xi_t \equiv \partial_t$ and $\xi_\phi \equiv \partial_\phi$ in the metric (1), we have $P_t = \text{const} = -E$, $P_\phi = \text{const} = m$. $E$ and $m$ are explained as the total energy of the particle and axial component of its angular momentum, respectively. It is well known that the Hamilton–Jacobi equation for a massless free particle is given as

$$g^{\mu\nu} P_\mu P_\nu = 0.$$  

(17)

Obviously, the difference between Eq. (15) and Eq. (17) is due to the presence of the spins and cosmological constant.

4. Entropy of spin fields

In this section, we discuss the entropy due to arbitrary spin fields in a neighborhood of the Kerr–Newman–de Sitter black hole by using the Teukolsky-type master equation (13), emphasizing in particular that in addition to the quadratic term of the spins in the entropy (and free energy), there is also the linear term of the spins. We introduce a six-dimensional phase space with $r, \theta, \phi, P_r, P_\theta, P_\phi$ as coordinate axes. Then the number of quantum states of the system in a given region of phase space read

$$g(E) = \frac{1}{(2\pi)^3} \sum_p \int dr d\theta d\phi \, dp_r \, dp_\theta \, dp_\phi.$$  

(18)

According to ‘t Hooft [1,2], the region of integration in coordinate space takes the spherical shell with radii $r_H + \epsilon_H$ and $r_H + l_H$, where $0 < \epsilon_H \ll l_H \ll r_H$. In momentum space, the integrals are taken only over those values for which the square roots in Eq. (15) exist. The $P_r$ and $P_\phi$ integrals can be performed explicitly to give

$$g(E) = \int dm \, g_m(E) = \frac{1}{8\pi^2} \sum_p \int dm \, \frac{d\theta d\phi}{\sqrt{g^{\phi\phi}}} \left[ \tilde{g}^{\phi\phi} (m - m_0)^2 + \tilde{g}^{tt} (E - m\Omega_H - E_0)^2 + V_0 \right],$$  

(19)

where

$$\tilde{g}^{\phi\phi} = g^{\phi\phi} - 2g^{t\phi} \Omega_H + g^{tt} \Omega_H^2,$$

$$\tilde{g}^{tt} = g^{tt} - (g^{\phi\phi})^{-1} (g^{t\phi} - g^{tt} \Omega_H)^2,$$

$$m_0 = (\tilde{g}^{\phi\phi})^{-1} \left[ (B^{\phi} - B' \Omega_H + (g^{t\phi} - g^{tt} \Omega_H)(E - m\Omega_H)) \right],$$

$$E_0 = (\tilde{g}^{tt})^{-1} \left[ (g^{\phi\phi})^{-1} (g^{t\phi} - g^{tt} \Omega_H)(B^{\phi} - B' \Omega_H) - B'. \right],$$

$$V_0 = V - \tilde{g}^{tt} E_0^2 - (g^{rr})^{-1} (B^{rr})^2 - (g^{\theta\theta})^{-1} (B^{\theta\theta})^2 - (\tilde{g}^{\phi\phi})^{-1} (B^{\phi} - B' \Omega_H)^2.$$  

(20)

The free energy $F$ at the inverse temperature $\beta$ for rotating black hole background is given by

$$\beta F = (-1)^{2s} \sum_q \ln \left[ 1 - (-1)^{2s} e^{-\beta(E_q - m\Omega_H)} \right].$$  

(21)
where \( q \) represents the set of quantum numbers. In Eq. (21), we have assumed that the spin fields are rotating with angular velocity \( \Omega_H \) [7,17–21]. Using Eq. (19) to determine the density of states, the free energy can be rewritten as

\[
F = -\int dm \int_{m\Omega_H}^{\infty} dE \frac{g_m(E)}{e^{\beta(E - m\Omega_H)} - (-1)^{2s}} = -\frac{1}{6\pi^2} \sum_{s} \int dr \, d\theta \, d\varphi \sqrt{-g} \left( \hat{g}^{\mu\nu} \right)^2 \int_{0}^{\infty} dE \frac{\left[E - E_0 \right] + (\hat{g}^{\mu\nu})^{-1}V_0}{e^{\beta E} - (-1)^{2s}}.
\]

(22)

Note that we do not consider here the contribution from the superradiant modes which are characterized by the condition of \( 0 < E < m\Omega_H \) and \( m > 0 \). There are two reasons to support this: (1) superradiance is nonthermal and is independent of black hole temperature [26], and (2) some fields do not superradiate, such as Weyl neutrinos [27]. Performing the remaining integrations, we find the leading behavior of the free energy:

\[
F \approx -\frac{15\omega + (-1)^{2s}}{16} \frac{\Gamma(4)(4)}{6\pi^2\beta^4} I_1 - \frac{3 + (-1)^{2s}}{4} \frac{\Gamma(2)(2)}{2\pi^2\beta^2} \sum_p I_2,
\]

(23)

where

\[
I_1 = \int dr \, d\theta \, d\varphi \sqrt{-g} \left( \hat{g}^{\mu\nu} \right)^2 \approx \frac{15\omega}{4\kappa_H^3} \frac{1}{e^\beta} + \frac{4\pi \kappa_H}{2} \left[ 1 - \frac{A}{2} \left( \frac{r_H^2 + a^2}{3} \right) - \frac{3Q^2}{4r_H^2} \left( 1 + \frac{r_H^2 + a^2}{ar_H} \right) \right] \ln \frac{L}{\epsilon},
\]

\[
I_2 = \int dr \, d\theta \, d\varphi \sqrt{-g} \left( \hat{g}^{\mu\nu} \right)^2 \left( E_0^2 + V_0^2 \right) \pi \kappa_H \left[ \frac{2}{3} \right] \left( \frac{2\kappa_H}{3} \right) \left( \frac{(s-p)^2}{2} \right) \left[ 1 - \frac{\Delta r_H^2}{r_H^2} \left( \frac{r_H^2 + a^2}{r_H^2 + 3\Delta r_H^2} \right) + \frac{5\Delta r_H^2}{r_H^2} - \frac{a^2}{r_H^2} \left( 1 + \frac{r_H^2 + a^2}{r_H^2} \right) \right] \ln \frac{L}{\epsilon}.
\]

(24)

In Eq. (23), \( \omega \) is the degeneracy due to the spin. For the scalar field we have \( \omega = 1 \); for the Weyl neutrino, electromagnetic, Rarita–Schwinger and gravitational fields we have \( \omega = 2 \). In Eq. (24), \( \epsilon = \sqrt{15\epsilon_H^2/2} \), \( L = \sqrt{L_H^2/\epsilon_H} \) are, respectively, the ultraviolet cutoff parameter and infrared cutoff parameter in Ref. [14], where \( \epsilon_H = 2\sqrt{(r_H^2 + a^2 \cos^2 \theta)^2} \) is the proper distance of the brick wall from event horizon. Here we have assumed that the proper cutoff is constant. Of course, one can also choose that the coordinate cutoff is constant as Wu and Yan did in Ref. [21]. But the choice does not influence the final result.

From Eq. (23) one can easily obtain the corresponding entropy of the spin fields

\[
S = \beta^2 \frac{\partial F}{\partial \beta} = \frac{15 + (-1)^{2s}}{16} \frac{\Gamma(4)(4)}{5\pi^2\beta^3} I_1 + \frac{3 + (-1)^{2s}}{4} \frac{\Gamma(2)(2)}{2\pi^2\beta^2} \sum_p I_2.
\]

(25)

Choosing the inverse temperature \( \beta \) to correspond to the Hawking inverse temperature \( \beta_H \), we find that the entropy of the nonextremal Kerr–Newman–de Sitter black hole due to arbitrary spin fields can be written as

\[
S = \frac{15 + (-1)^{2s}}{16} \left[ \frac{A_H}{48\pi \epsilon} \frac{1}{4\xi} \left[ \frac{3Q^2}{4r_H^2} \left( 1 + \frac{A_H}{4\pi ar_H} \arctan \frac{a}{r_H} \right) \right] \ln \frac{L}{\epsilon} \right] + \frac{3 + (-1)^{2s}}{4} \left[ \frac{A_H}{36\pi} \right] \ln \frac{L}{\epsilon}.
\]

(26)

In Eq. (26), the first term (in the first brace) is agreement with similar calculations using the Hamilton–Jacobi equation (17), while the second term (in the second brace) come from the effect of the spins and cosmological constant.
Eq. (26) can be rewritten more explicitly as

\[
S(s = 0) = \frac{A_H}{48\pi e^2} + \left[ \frac{1}{45} - \frac{11\Lambda A_H}{360\pi} - \frac{Q^2}{60\Xi r_H^2} \left( 1 + \frac{\Xi A_H}{4\pi a r_H} \arctan \frac{a}{r_H} \right) \right] \ln \frac{L}{\epsilon} \\
= \frac{A_H}{48\pi e^2} + \left[ \frac{A_H}{240\pi r_H^2} - \frac{23\Lambda A_H}{720\pi} - \frac{M}{30\Xi r_H} + \frac{1}{45} \right] \ln \frac{L}{\epsilon},
\]

(27)

\[
S\left( s = \frac{1}{2} \right) = \frac{7A_H}{192\pi e^2} + \left[ \frac{13A_H}{640\pi r_H^2} - \frac{457\Lambda A_H}{5760\pi} - \frac{M}{10\Xi r_H} + \frac{73}{720} \right] \ln \frac{L}{\epsilon},
\]

(28)

\[
S\left( s = 1 \right) = \frac{A_H}{24\pi e^2} + \left[ \frac{11A_H}{120\pi r_H^2} - \frac{31\Lambda A_H}{120\pi} - \frac{M}{15\Xi r_H} + \frac{2}{45} \right] \ln \frac{L}{\epsilon},
\]

(29)

\[
S\left( s = \frac{3}{2} \right) = \frac{7A_H}{192\pi e^2} + \left[ \frac{179A_H}{1920\pi r_H^2} - \frac{1117\Lambda A_H}{5760\pi} + \frac{M}{15\Xi r_H} - \frac{107}{720} \right] \ln \frac{L}{\epsilon},
\]

(30)

\[
S\left( s = 2 \right) = \frac{A_H}{24\pi e^2} + \left[ \frac{3A_H}{10\pi r_H^2} - \frac{47\Lambda A_H}{90\pi} + \frac{3M}{5\Xi r_H} - \frac{43}{45} \right] \ln \frac{L}{\epsilon},
\]

(31)

It should be noted that when \( \Lambda = 0 \), Eq. (27) reduces to the case discussed by Mann and Solodukhin [15], which was obtained by a different method.

5. Discussion and conclusions

We have derived the quantum corrections to the entropy of a nonextremal Kerr–Newman–de Sitter black hole due to fields of arbitrary spin \( s \leq 2 \). In the expression for the entropy there is a leading divergence, which goes as the inverse of the square of the proper distance of the brick wall from event horizon, and a logarithmic divergence too. The leading term of Eq. (26) is in agreement with similar calculations in the literature [17–21]. As regards the logarithmic contribution, they are completely different.

As already mentioned in the introduction, the literature [17–21] showed that the logarithmic terms depend on the spins of the fields just in quadratic term \( s^2 \), but we found that the linear terms of \( s \) are not eliminated by each other when we sum over all entropies corresponding to the spin states to get the total entropy. Therefore we conclude that the subleading logarithmic correction of quantum entropies not only depend on the characteristics of the metric and the quadratic term of the spins but also the linear term of the spins.

In fact, the difference arises from the use of the WKB approximation, for instance, if Eq. (14) is replaced by \( \Psi_p = \exp[iS(t, r, \theta, \varphi, p)] \), the quantum entropy is compatible with the result given in Ref. [21]. Of course, different forms of \( \Psi_p \) lead to different results. This is the reason why the difference appears in the previously published results [17–21]. We think that the WKB approximation can only apply to the wave functions \( \Phi_{\pm s} \). For a part of the wave function, such as \( \Psi_p \), the WKB approximation is not applicable. Note that, at the WKB level, usually one also writes \( \Phi_{\pm s} = f(x) \exp[iS(x)] \), but its validity requires that \( f(x) \) must be a slowly varying amplitude, i.e. \( \partial_x f \approx 0 \). Obviously, the factors \( \rho^{(s^2-p)} \) in Eq. (14) do not satisfy the requirement.

It is worth pointing out that all results of the present Letter are valid for arbitrary spin fields within the cosmological horizon neighborhood if the quantities of the black hole horizon are replaced by the corresponding quantities of the cosmological horizon. (This is the reason why we used the absolute value of \( \Delta r_H \) in Eq. (5) and the expression of the proper distance \( \epsilon_{HP} \).) Therefore, our study includes as special cases Schwarzschild, Reissner–Nordström, Kerr, Kerr–Newman and de Sitter spacetime background, or any combination of these.
Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant No. 10375051.

References

Cosmological constant from gauge fields on extra dimensions

Hing-Tong Cho\textsuperscript{a,b}, Choon-Lin Ho\textsuperscript{a}, Kin-Wang Ng\textsuperscript{b,*}

\textsuperscript{a} Department of Physics, Tamkang University, Tamsui, Taipei County, Taiwan 251, ROC
\textsuperscript{b} Institute of Physics, Academia Sinica, Nankang, Taipei, Taiwan 115, ROC

Received 21 August 2006; received in revised form 13 October 2006; accepted 17 October 2006
Available online 7 November 2006
Editor: T. Yanagida

Abstract

We present a new model of dark energy which could explain the observed accelerated expansion of our Universe. We show that a five-dimensional Einstein–Yang–Mills theory defined in a flat Friedmann–Robertson–Walker universe compactified on a circle possesses degenerate vacua in four dimensions. The present Universe could be trapped in one of these degenerate vacua. With the natural requirement that the size of the extra dimension could be of the GUT scale or smaller, the energy density difference between the degenerate vacua and the true ground state can provide us with just the right amount of dark energy to account for the observed expansion rate of our Universe.

© 2006 Elsevier B.V. All rights reserved.

PACS: 98.80.Cq; 98.80.Es; 11.25.Mj; 11.15.Ex

It is now generally agreed that the biggest unsolved problem in astronomy and cosmology is the newly discovered fact that our Universe is undergoing a stage of accelerated expansion. Recent astrophysical and cosmological observations such as structure formation, type Ia supernovae, gravitational lensing, and cosmic microwave background anisotropies have concordantly predicted a spatially flat universe containing a mixture of matter and a dominant smooth component, dubbed the “dark energy”, which provides an anti-gravity force to accelerate the cosmic expansion [1]. The simplest candidate for this invisible component carrying a sufficiently large negative pressure for the anti-gravity is a true cosmological constant. Current observational data, when fitted to dark energy models with a static $w$, favor the so-called $\Lambda$CDM model with vacuum energy $\Omega_\Lambda = 0.7$ and cold dark matter $\Omega_{\text{CDM}} = 0.3$, while constraining $w = -1.02^{+0.13}_{-0.19}$ at the 95% confidence level. Furthermore, joint constraints on both $w(z)$ and its time evolution $dw(z)/dz$ at $z = 0$ are consistent with the value of the equation of state expected of a static cosmological constant [3]. However, the data are not precise enough to pinpoint whether the dark energy is truly static or dynamical. Most likely, new kinds of measurements or next-generation experiments are needed to reveal the nature of dark energy. In this Letter, we will address the dark energy problem by considering non-Abelian gauge theories on compact extra dimensions.

Interest in theories of extra dimensions has been quite immense in recent years [4]. New models such as brane scenarios, large extra dimensions, and warped extra dimensions have not only revolutionized the Kaluza–Klein theory, but also shed new light on some long-standing problems in particle physics and cosmology. Interestingly, theories with large extra dimensions can be even tested by future collider experiments. Recently cosmological models involving extra dimensions have been constructed to account for the current cosmic acceleration or to accommodate the dark energy [5]. Here we shall show how a

\textsuperscript{*} Corresponding author.
\textit{E-mail address: nkw@phys.sinica.edu.tw} (K.-W. Ng).

0370-2693/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.physletb.2006.10.042
five-dimensional (5d) gauge theory based on the Hosotani symmetry-breaking mechanism [6] could naturally give rise to a small but finite cosmological constant that plays the role of dark energy.

The Hosotani mechanism is a non-Higgs-type symmetry-breaking mechanism which has been widely discussed in the literature [7]. Its main idea is that in a multiply connected spacetime manifold, the vanishing of the field strength $F_{MN} = 0$ of gauge fields $A_M$ in a vacuum does not necessarily imply the vanishing of the gauge fields, and $A_M \neq 0$ will imply gauge symmetry breaking in general. The relevant order parameter is the path-ordered Wilson lines $U_n = P \exp(fC_n A_M dx^M)$, where $C_n$ represent non-contractible loops in the manifold. If $U_n$ do not belong to the center of the group, then the original gauge group is broken to some subgroup that commutes with all $U_n$. This mechanism has been extensively employed in superstring phenomenology [8], and applied to Kaluza–Klein cosmology in connection with the problem of vacuum stability [9]. Recently, a new extranatural inflation model in which the inflaton is the fifth component of a gauge field in a 5d theory compactified on a circle was presented [10], and it was shown that the fifth component may also be a good candidate for quintessence if the quintessential potential is provided by massive bulk fields with bare masses of order of the GUT scale [11].

In Ref. [9], the authors considered a 5d Einstein–Yang–Mills theory, with massless fermions, defined in a flat Friedmann–Robertson–Walker universe compactified on a circle. The spacetime metric is given by

$$ds^2 = g_{MN} dx^M dx^N \quad (M = 0, 1, 2, 3, 5)$$

and the action is consisted of gravity, $SU(2)$ gauge fields, and a fermionic sector $L_f$ which contains $N_f$ massless gauged fermions and $n_f$ massless free fermions:

$$S = -\int d^5 x \sqrt{|g_{MN}|} \left( \frac{R}{16\pi G} + \frac{1}{2} \text{Tr} F_{MN}^2 + L_f \right).$$

Here the gauge bosons and the fermions can be considered as fields in some hidden sector of a grand unified theory. Dynamics of the gauge fields and the fermion fields here determines the structure of the vacuum. The Casimir energy of the system was computed by evaluating the one-loop effective potentials of the gauge fields and the fermions in the backgrounds defined by the metric (1) and the classical gauge fields of the form

$$A_\mu = 0, \quad A_5 = \phi(t) \sigma^3, \quad \mu = 0, 1, 2, 3,$$

where $A_5$ is the fifth component along the circle and $\sigma$ is the Pauli matrix. This amounts to a total effective potential of the system denoted by $V(b R_0, \phi)$, where $R_0$ is the final radius of the circle. It was shown that the Einstein’s equations for $a(t)$, $b(t)$ and $\phi(t)$, derived from the action (2), admit static vacuum solutions with stable compactification. These solutions with $\dot{a}_0 = \dot{b}_0 = \phi_0 = 0$ are given by the global minima of $V(b R_0, \phi)$ determined by

$$2g_5 b_0 R_0 \phi_0 = \frac{r}{2} \quad (\text{mod } r),$$

where $g_5$ is the 5d $SU(2)$ gauge coupling and $r = 1$ and $r = 2$ correspond respectively to periodic fermions in the adjoint and fundamental representations. In the case of adjoint fermions, the vacuum states correspond to $U(1)$ symmetry, hence $SU(2)$ gauge symmetry is dynamically broken. For fundamental fermions the gauge symmetry is unbroken as the Wilson line is an element of $Z_2$. To obtain a zero cosmological constant for these vacuum states, an appropriate number ($n_f$) of free fermions with assigned boundary conditions along the circle has been chosen such that $V(b R_0, \phi_0) = 0$. A salient feature of this model is that only a small number of fermions is required to stabilize the vacuum. In fact, $N_f = 1$ gauged fermion will suffice to do the job. On the contrary, in the other Kaluza–Klein-type theories one usually needs to add a large number (of the order of $10^4$) of matter fields for the same purpose.

Now we turn to the dark energy problem. Naively, one would expect that the vacuum energy density $\rho_A$ is of order $M_p^4$, where the reduced Planck mass is given by $M_p = (8\pi G)^{-1/2} = 2.44 \times 10^{18}$ GeV, since the Planck scale is the natural cutoff scale of zero-point energies of each quantum field. But the observed value for the vacuum-like energy density is $\rho_A \approx 0.7 \rho_c \approx 1.6 h^2 \times 10^{-120} M_p^4$, where $h \approx 0.7$ is the present Hubble parameter defined by $H_0 = 8.76 h \times 10^{-61} M_p$ [1], and so the naive estimate is larger than the observed value by a factor of $10^{20}$. Many solutions have been proposed about a vanishingly small cosmological constant in some ultimate ground state [12]. Here we assume that the cosmological constant absolutely vanishes in a true ground state with lowest possible energy density. Then, we will show that the vacuum states in the model considered above are indeed metastable due to quantum tunneling effects and will eventually settle down to this true ground state. If the present Universe is still in one of these quasiground states, then the energy density difference above the true ground state can provide us with a small but finite cosmological constant. This idea has already been put forth by considering the topological vacua in a 4d $SU(2)$ Yang–Mills–Higgs theory which is spontaneously broken to $U(1)$ via the Higgs mechanism with a Higgs potential [13]. Unfortunately, the existence of massless or light gauged fermions would suppress the tunneling [14] and thus spoil the idea. Although the fermions can get masses via the Higgs mechanism, the Yukawa couplings are quite arbitrary and may be very small. In contrast, our scenario has new merits. Firstly, we do not need to introduce an ad hoc Higgs field, which is here replaced by the fifth component of the gauge field. Secondly, the gauge symmetry breaking needs not be introduced arbitrarily, but instead is determined dynamically by the Casimir energy of the non-integrable phase of the gauge field on the compact dimension through the Hosotani mechanism. Thirdly, the gauged fermions naturally get huge masses of order of the inverse of the size of the compact extra dimension. Lastly, the cosmological constant in our model, as we will show below, is naturally related to the size of the compact extra dimension.

Without loss of generality, let us consider a universe corresponding to the vacuum state in the case of periodic fundamental fermions with $2g_5 R_0 \phi_0 = 1$, where we have set $a_0 = b_0 = 1$ and $r = 2$ in Eq. (4). Presumably, the universe has undergone
the compactification of the circle with initial conditions at some early time $t_1$ (for instance, given by $a(t_1)$, $b(t_1)$, and $\phi(t_1) \cong 0$), and rolls down the effective potential $V$ to this vacuum state. The actual evolution is very interesting and it warrants a detailed study of the Einstein’s equations. In fact, this scenario has been discussed in the context of the extranatural inflationary model [10]. Here we only concern about the vacuum state and the dark energy problem. At energies below $1/R_0$, the 4d effective action for the zero Fourier modes of gauge fields in Eq. (2) is given by

$$S_{\text{eff}}^{\text{gauge}} = -\int d^4x \left(\frac{1}{2} \text{Tr} \tilde{F}_{\mu
u} \tilde{F}^{\mu\nu} + \text{Tr} \tilde{F}_{\mu5} \tilde{F}^{\mu5}\right).$$

Note that we have rescaled $A_M = \tilde{A}_M / \sqrt{2\pi R_0}$ and $g_5 = g_4 \sqrt{2\pi R_0}$, where $g_4$ is the dimensionless 4d SU(2) gauge coupling. Since $\tilde{A}_\mu$ is independent of $\phi_5$ in the effective action, so $\partial_5 \tilde{A}_\mu = 0$ and $\tilde{F}_{\mu5}$ reduces to the covariant derivative of $\tilde{A}_\nu$. Hence, by rewriting $\tilde{A}_5 = \Phi$, we have

$$S_{\text{eff}}^{\text{gauge}} = \int d^4x \left(-\frac{1}{2} \text{Tr} F_{\mu\nu} F^{\mu\nu} + \text{Tr}(D_\mu \Phi)(D^\mu \Phi)\right).$$

This resembles the SU(2) Yang–Mills–Higgs system without a Higgs potential. Although the 5d gauge symmetry is unbroken for the vacuum states given by Eq. (4), the fifth component obtains a non-zero vacuum expectation value $\langle \Phi \rangle = i\phi_0 \sigma^3$, where $\phi_0 = 1/(2g_4 R_0)$ [9]. Thus, the zero-mode gauge fields and gauged fermions acquire huge mass of the mass of $1/(2R_0)$ [15]. In the case of periodic adjoint fermions, the system is spontaneously broken to $U(1)$ by the Wilson loop along the small circle with $\langle \Phi \rangle = i\phi_0 \sigma^3$, where $\phi_0 = 1/(4g_4 R_0)$ [9]. Nevertheless, some fermion modes still remain massless [15]. Since these massless fermions suppress the vacuum tunneling [14], we will not consider this case.

It is well known [14,16] that a non-Abelian gauge theory, such as that described by Eq. (6), bears degenerate perturbative vacua classified in terms of the winding number $n$ and denoted by $|n\rangle$. Before we go on, we should emphasize that there are no instantons nor topological numbers for the 5d gauge fields in Eq. (2). Even in the massless sector (given by Eq. (6)), we still do not have exact instantons but only “constrained ones”. However, these constrained instantons are sufficient in describing the tunneling phenomenon [16,17] as we shall do in due course. Thus let us consider the 4d SU(2) Yang–Mills theory with gauge coupling $g$. While the degenerate vacua are each separated by an energy barrier, the change of the winding number can take place through quantum tunneling from one vacuum to another. As such, the true ground state, so-called the $\theta$ vacuum, can be constructed as

$$|\theta\rangle = \sum_{n=-\infty}^{\infty} e^{in\theta} |n\rangle,$$  

where $\theta$ is a real parameter. Using the dilute instanton gas approximation [14], it was found that the expectation value of the Hamiltonian $\mathcal{H}$ over the Euclidean spacetime volume $V_T$ is given by

$$\langle \theta | e^{-\mathcal{H}T} | \theta \rangle \propto \sum_{n,n'} \langle n'|e^{-\mathcal{H}T}|n\rangle e^{(n-n')\theta}$$

$$= \exp\left[2KV_T e^{-S_0} \cos \theta\right],$$

where $S_0 = 8\pi^2/g^2$ is the Euclidean action for an instanton solution which corresponds to the quantum tunneling from $|n\rangle$ to $|n \pm 1\rangle$. Here $K$ is a positive determinantal prefactor. Eq. (8) shows that the energy density difference between each $\theta$ vacuum and the perturbative vacuum is given by

$$\langle \theta | \rho_\Lambda | \theta \rangle - \langle n | \rho_\Lambda | n \rangle = -2K e^{-S_0} \cos \theta.$$  

Therefore, the $\theta$ vacuum with the lowest energy is given by $\theta = 0$. Let us assume that the $\theta = 0$ vacuum is the true ground state where the vacuum energy vanishes. In fact, it has been proposed [18] that the effect of wormholes [19] drives the Universe to this CP-symmetric $\theta = 0$ vacuum state with zero cosmological constant. (More discussions about this can be found in Ref. [13] and references therein.) Then, after normalizing $\langle \theta | \rho_\Lambda | \theta \rangle - \langle n | \rho_\Lambda | n \rangle = 0$, we find that the vacuum energy density in each perturbative vacuum is

$$\langle n | \rho_\Lambda | n \rangle = 2K e^{-S_0}.$$  

This vacuum energy will manifest as a cosmological constant provided that we still live in one of these perturbative vacua. To guarantee this condition, the tunneling probability from any one of these perturbative vacua to the true ground state in the current horizon volume in the cosmic age must satisfy $\Gamma H_{\text{obs}}^4 \lesssim 1$, where $\Gamma$ is the tunneling rate per unit volume per unit time given by $\Gamma \propto \rho_\Lambda e^{-S_0}$.

Unfortunately, for a pure SU(2) gauge theory which is scale-invariant, the size of the instanton $\rho$ is arbitrary, and so the prefactor $K$ which involves an integral over instanton sizes diverges as $\rho$ goes to infinity. However, if the SU(2) gauge field is coupled to a Higgs scalar with isospin $q$ and vacuum expectation value $v$, the quantum tunneling will proceed via a constrained instanton whose size is cut off at a scale $v^{-1}$ [16,17]. The Higgs contribution to the Euclidean action of the constrained instanton is approximately given by $S_H = 4\pi^2 q^2 v^2$, rendering $K$ a finite quantity. Including $N_f$ gauged fermions with mass $m_f$, we find that [16]

$$K = \frac{4\pi^2}{\alpha^4} \int_0^\infty \frac{dp}{\rho^5} (m_f \rho)^N \exp\left[-S_H + c_1 \ln(\rho v) + c_2\right]$$

$$= 2\pi^2 e^{c_2} \frac{\Gamma((N_f + c_1 - 4)/2)}{(4\pi^2 q)^{(N_f + c_1 - 4)/2}} \left(\frac{m_f}{v}\right)^{N_f} \left(\frac{v}{\alpha}\right)^4,$$  

(11)

\footnote{It has been proved [20] that the $|n\rangle$ vacua are not physical because they do not satisfy cluster decomposition: $|\langle n | e^{-\mathcal{H}(T_1+T_2)} | n + \nu \rangle\rangle = \sum_{m}(\langle n | e^{-\mathcal{H}T_1} | m \rangle \langle m | e^{-\mathcal{H}T_2} | n + \nu \rangle) v$ for large times $T_1$ and $T_2$. Here, however, the tunneling time is longer than the age of the Universe. So, observationally we have $|\langle n | e^{-\mathcal{H}T} | p \rangle\rangle \approx \delta_{np}$, which implies the cluster decomposition, $|\langle n | e^{-\mathcal{H}(T_1+T_2)} | n \rangle\rangle = |\langle n | e^{-\mathcal{H}T_1} | n \rangle\rangle |\langle n | e^{-\mathcal{H}T_2} | n \rangle\rangle$.}
where \( \alpha = g^2 / (4\pi) \) is the gauge coupling strength at the energy scale \( v \), \( c_1 = 20/3 - 2N_f / 3 \), and \( c_2 = 5.96 - 0.36N_f \). Therefore (cf. [13]), if \( m_f = gv \), we will have

\[
\rho_\Lambda \simeq c^2 \left( \frac{v}{\alpha} \right)^4 \alpha^\nu_f e^{-2\pi / \alpha},
\]

(12)

\[
\Gamma \simeq c^2 \left( \frac{v}{\alpha} \right)^4 \alpha^\nu_f e^{-4\pi / \alpha},
\]

(13)

where

\[
c^2 = 4\pi^2 e^{c^2 + 4\pi} \left( \frac{v}{\alpha} \right)^4 \Gamma \left( N_f + c_1 - 4 / 2 \right) \left( 4\pi^2 g \right)^{(N_f + c_1 - 4) / 2}.
\]

(14)

The requirements that \( \rho_\Lambda = 1.6h^2 \times 10^{-12} M_P^4 \) and \( \Gamma H_0^4 \lesssim 1 \) give

\[
\left( \frac{v}{M_P} \right) + \left( 1 - \frac{N_f}{8} \right) \ln \alpha = \ln \left( \frac{v}{M_P} \right) + 30 \ln 10 - \frac{1}{2} \ln \left( \frac{12.7h}{c} \right),
\]

(15)

\[
\frac{v}{M_P} \geq 1.45\alpha^\nu_f \sqrt{c}.
\]

(16)

For the minimal value of \( v \) in Eq. (16) and \( h = 0.7 \), we obtain \( \alpha \simeq 1/44.25 \). Let us assume that \( q = 1 \). Then, we have \( v \simeq 3.8 \times 10^{16} \) GeV for \( N_f = 1 \) and \( v \simeq 8.3 \times 10^{16} \) GeV for \( N_f = 3 \). If we take \( v = M_P \) in Eq. (15), we will obtain \( \alpha \simeq 1/46.9 \) for \( N_f = 1 \) and \( \alpha = 1/46.4 \) for \( N_f = 3 \).

To apply the above results to our model (6), we make the replacements \( g = g_4 \) and \( v = 0 \). Thus, we find that \( R_0 \simeq 4.1 \times 10^{16} \) GeV, which indicates that the maximal size of the compact extra dimension is of the order of the GUT scale. This result is obtained without fine-tuning. The smallness of the cosmological constant is related to the tunneling rate between the various vacua of the theory which is also very tiny. This idea was first put forth generically in Ref. [13] based on the Hosotani mechanism, but Eq. (11) shows that the existence of massless or light fermions with \( m_f < v \) will suppress the prefactor \( K \). Consequently, the \( |n> \) vacua do not satisfy cluster decomposition due to the chirality consideration [14]. Here we revolutionize the idea by replacing the Higgs field with the gauge field in the extra dimension through the Hosotani mechanism. The vacuum state is chosen by the minimum of the Casimir energy instead of the rather ad hoc Higgs potential. An important advancement is that the gauged fermions in our model naturally obtain a huge and definite mass given by \( m_f = gv \). To our knowledge, the Hosotani mechanism is the only way that does the job.

In summary, we have constructed a new model of dark energy based on a 5d Einstein–Yang–Mills theory. The presence of dark energy comes out rather naturally in our model just from the assumption that there exist extra dimensions. Symmetry breaking by the Hosotani mechanism and the constrained instantons related to the vacuum structure of the gauge field are both immediate consequences. As long as the size of the extra dimension is between the GUT and the Planck scales, the resulting cosmological constant will then be just of the right amount to account for the dark energy content in our Universe. No other ingredients are needed to achieve this goal.

Acknowledgements

This work was supported in part by the National Science Council, Taiwan, ROC under the Grants NSC 94-2112-M-032-009 (H.T.C.), NSC 94-2112-M-032-007 (C.L.H.), and NSC 94-2112-M-001-024 (K.W.N.).

References

[15] See, e.g., Hosotani’s papers in Ref. [6].


Modified gravity inspired DGP brane cosmology

K. Atazadeh, H.R. Sepangi *

Department of Physics, Shahid Beheshti University, Evin, Tehran 19839, Iran

Received 1 August 2006; accepted 21 October 2006
Available online 30 October 2006
Editor: T. Yanagida

Abstract

We consider a DGP brane scenario where a scalar field is present on the brane through the introduction of a scalar potential, itself motivated by the notion of modified gravity. This theory predicts that the mass appearing in the gravitational potential is modified by the addition of the mass of the scalar field. The cosmological implications that such a scenario entails are examined and shown to be consistent with a universe expanding with power-law acceleration.

© 2006 Elsevier B.V. All rights reserved.

1. Introduction

The idea that extra dimensions can be probed by gravitons and eventually non-standard matter has been the dominant trend in the recent past. These models usually yield the correct Newtonian \(1/r\) potential at large distances because the gravitational field is quenched on sub-millimeter transverse scales. This quenching appears either due to finite extension of the transverse dimensions [1,2] or due to sub-millimeter transverse curvature scales induced by negative cosmological constants [3–8]. A feature common to these type of models is that they predict deviations from the usual 4D gravity at short distances. The model proposed by Dvali, Gabadadze and Porrati (DGP) [9,10] is very different in that it predicts deviations from the standard 4D gravity over large distances. The transition between four and higher-dimensional gravitational potentials in the DGP model arises because of the presence of both the brane and bulk Einstein terms in the action. An interesting observation was made in [11,12] where it was shown that the DGP model allows for an embedding of the standard Friedmann cosmology in the sense that the cosmological evolution of the background metric on the brane can entirely be described by the standard Friedmann equation plus energy conservation on the brane. This was later generalized to arbitrary number of transverse dimensions in [13]. For a comprehensive review of the phenomenology of DGP cosmology, the reader is referred to [14].

An interesting observation made a few years ago is that the expansion of our universe is currently undergoing a period of acceleration which is directly measured from the light-curves of several hundred type Ia supernovae [15,16] and independently from observations of the cosmic microwave background (CMB) by the WMAP satellite [17] and other CMB experiments [18]. However, the mechanism responsible for this acceleration is not well understood and many authors introduce a mystery cosmic fluid, the so-called dark energy, to explain this effect [19]. Recently, it has been shown that such an accelerated expansion could be the result of a modification to the Einstein–Hilbert action [20]. One such modification has been proposed in [21] where a term of the form \(R^{-1}\) was added to the usual Einstein–Hilbert action. It was then shown that this term could give rise to accelerating solutions of the field equations without dark energy.

In this Letter, we focus attention on the DGP brane model and introduce a scalar field on the brane. The potential describing such a scalar field is taken to be that appearing in modified theories of gravity when the term \(R^{-1}\) is added to the usual Einstein–Hilbert action. This model predicts that for such a potential, the mass density should be modified by the addition of the mass density of the corresponding scalar field on the brane. We obtain the evolution of the metric on the spacetime by solving the field equations in the limit of small curvature, predicting...
a power-law acceleration on the brane. The components of the metric in Gaussian normal coordinates are also calculated and presented.

2. DGP model with a brane scalar field

We start by writing the action for the DGP model with a scalar field on the brane part of the action

\[ S = \frac{m_4^3}{2} \int d^5x \sqrt{-g} \mathcal{R} + \int d^4x \sqrt{-g} \left[ \frac{m_3^2}{2} R - \frac{1}{2} (\nabla \Phi)^2 - V(\Phi) \right] + S_m[q_{\mu\nu}, \psi_m], \]

(1)

where the first term in (1) corresponds to the Einstein–Hilbert action in 5D for the 5-dimensional bulk metric \( g_{AB} \), with the Ricci scalar denoted by \( \mathcal{R} \). Similarly, the second term is the Einstein–Hilbert action for a scalar field \( \Phi \) corresponding to the induced metric \( q_{\mu\nu} \) on the brane, where \( R \) is the relevant scalar curvature and \( m_3 \) and \( m_4 \) are reduced Planck masses in four and five dimensions respectively and \( S_m \) is the matter action on the brane with matter field \( \psi_m \). The induced metric \( q_{\mu\nu} \) is defined as usual from the bulk metric \( g_{AB} \) by

\[ q_{\mu\nu} = \delta_{\mu\nu}^{AB} g_{AB}. \]

(2)

It would now be possible to write the field equations resulting from this action, yielding, in \( d - 1 \) spatial dimensions

\[ m_3^2 \left( R_{AB} - \frac{1}{2} g_{AB} \mathcal{R} \right) + m_3^2 \delta^{AB} \left( R_{(d-1)}^{(d-1)} - \frac{1}{2} q_{\mu\nu} R_{(d-1)}^{(d-1)} \right) \delta(y) = \delta_{\mu\nu}^{AB} \left( T_{\mu\nu} + \tau_{\mu\nu} \right) \delta(y), \]

(3)

where \( T_{\mu\nu} \) is the energy–momentum tensor in the matter frame, and\n
\[ \tau_{\mu\nu} = \partial_\mu \Phi \partial_\nu \Phi - \frac{1}{2} q_{\mu\nu} q^{\alpha\beta} \partial_\alpha \Phi \partial_\beta \Phi - V(\Phi) q_{\mu\nu}. \]

(4)

with the equation of motion for \( \Phi \) becoming

\[ \nabla_\mu \nabla^\mu \Phi = \frac{d V(\Phi)}{d \Phi}. \]

(5)

Note that \( \Phi \) lives on the brane. The corresponding junction conditions, relating the extrinsic curvature to the energy–momentum tensor, become \([11,12]\)

\[ \lim_{\epsilon \rightarrow +0} \left[ K_{\mu\nu} \right]_{y=\epsilon} = \frac{1}{m_4^3} \left( \tau_{\mu\nu} - \frac{1}{d-1} q_{\mu\nu} q^{\alpha\beta} \tau_{\alpha\beta} \right) \bigg|_{y=0} = \frac{m_3^2}{m_4^3} \left( R_{(d-1)}^{(d-1)} - \frac{1}{2(d-1)} q_{\mu\nu} q^{\alpha\beta} R_{\alpha\beta}^{(d-1)} \right) \bigg|_{y=0}. \]

(6)

In order to get a qualitative picture of how gravity works for the DGP braneworld, let us take small metric fluctuations around flat, empty space and look at gravitational perturbations, \( h_{AB} \),

that is

\[ g_{AB} = \eta_{AB} + h_{AB}, \]

(7)

where \( \eta_{AB} \) is the five-dimensional Minkowski metric. Choosing the harmonic gauge in the bulk

\[ \partial_\mu h_{AB} = \frac{1}{2} \delta_\mu h^A_B, \]

(8)

the \( \mu 5 \)-components of this gauge condition lead to \( h_{\mu 5} = 0 \), so that the surviving components are \( h_{\mu\nu} \) and \( h_{55} \). The latter component is solved by the following equation

\[ \Box^{(5)} h^5_\mu = \Box^{(5)} h_\mu, \]

(9)

where \( \Box^{(5)} \) is the five-dimensional d’Alembertian. The \( \mu\nu \)-component of the field equations (3) become, after a little manipulation [10]

\[ m_3^2 \Box^{(4)} h_{\mu\nu} + m_3^2 \left( \Box^{(4)} h_{\mu\nu} - \delta_\mu \partial_\nu h_\delta^5 \right) \delta(y) = -2 \delta(y) \left[ T_{\mu\nu} + \tau_{\mu\nu} - \frac{1}{d-1} \eta_{\mu\nu} h^{\alpha\beta}(T_{\alpha\beta} + T_{\alpha\beta}) \right], \]

(10)

where \( \Box^{(4)} \) is the four-dimensional (brane) d’Alembertian, and we take the brane to be located at \( y = 0 \). This yields the equation for the gravitational potential of mass densities \( \rho(r) = M \delta(\bar{r}) \) and \( \rho_\Phi(r, \bar{y}) = M_\Phi \delta(\bar{r}) \delta(y) \).

\[ m_3^2 \left( \Box^{(4)} + \partial_\gamma^2 \right) U(\bar{r}, y) + m_3^2 \delta(y) \Box^{(4)} U(\bar{r}, y) = \frac{2}{3} (M + M_\Phi) \delta(y) \delta(\bar{r}). \]

(11)

Therefore, in Eq. (11), the mass is modified by the addition of the mass of the scalar field. The resulting gravitational potential for \( r \ll \ell_{DGP} \) is given by \([9,11]\)

\[ U(\bar{r}) = -\frac{(M + M_\Phi)}{6 \pi m_3 \bar{r}} \left[ 1 + \left( \frac{y}{\frac{\sqrt{3}}{2}} \right) \frac{r}{\ell_{DGP}} + \frac{r}{\ell_{DGP}} \ln \left( \frac{r}{\ell_{DGP}} \right) + \mathcal{O} \left( \frac{r}{\ell_{DGP}} \right)^2 \right], \]

(12)

and

\[ U(\bar{r}) = -\frac{(M + M_\Phi)}{6 \pi^2 m_4^3} \left[ 1 - 2 \left( \frac{r}{\ell_{DGP}} \right)^{-2} + \mathcal{O} \left( \frac{r}{\ell_{DGP}} \right)^{-4} \right], \]

(13)

for \( r \gg \ell_{DGP} \), where \( \gamma = 0.577 \) is the Euler constant and \( \ell_{DGP} = \frac{m_3^2}{2m_4^3} \) is the transition scale between the four and five-dimensional behavior of gravitational potential that the DGP scenario predicts.

3. DGP cosmology with a brane scalar field

Although the DGP model predicts deviations to gravity at large distances, it could account for the standard cosmological equations of motion at any distance scale on the brane. It is therefore appropriate to start by writing the from of the line element in brane gravity, that is

\[ ds^2 = q_{\mu\nu} dx^\mu dx^\nu + b^2(y, t) dy^2 = -n^2(y, t) dt^2 + a^2(y, t) \gamma_{ij} dx^i dx^j + b^2(y, t) dy^2, \]

(14)
where $\gamma_{ij}$ is a maximally symmetric 3-dimensional metric where $k = -1, 0, 1$ parameterizes the spatial curvature. Building on the results of [22], the cosmological evolution equations of a 3-brane in a 5-dimensional bulk resulting from Eqs. (3) and (6) were presented in the first two references in [20]. Here, we will follow [11,12] and only give the results relevant to the present work for a brane of dimension $v + 1$ and the scalar field $\Phi$. A detailed discussion on the derivation of these results can be found in the said references. Adopting the Gaussian normal system gauge

$$b^2(y, t) = 1,$$

the field equations on the brane for metric (14) and $d = v + 1$ spatial dimensions are

$$G_{00}^{(v)} = \frac{1}{2} v(v - 1) n^2 \left( \frac{\dot{a}^2}{n^2 a^2} + \frac{k}{a^2} \right) = \frac{1}{m_v^2} T_{00},$$

$$G_{ij}^{(v)} = (v - 1) \left( \frac{\dot{n} a}{n^2 a} - \frac{\dot{a}}{n^2 a} \right) q_{ij}$$

$$- \frac{1}{2} (v - 1) (v - 2) n^2 \left( \frac{\dot{a}^2}{n^2 a^2} + \frac{k}{a^2} \right) q_{ij}$$

$$= \frac{1}{m_v^2} T_{ij},$$

$$\nabla_\mu \nabla^\mu \Phi = \frac{dV(\Phi)}{d\Phi}.$$  (18)

The junction conditions (6) for an ideal fluid on the brane, given by

$$T_{\mu\nu} = (\rho + p) u_\mu u_\nu + p q_{\mu\nu},$$

together with energy conservation resulting from the vanishing of

$$G_{05} = v \left( n' \frac{\dot{a}}{n a} - \frac{\dot{n}}{a} \right) = 0,$$

in the bulk, leads to

$$(\dot{\rho} + \dot{\rho}_\Phi) a |_{y=0} = -v(\rho + \rho_\Phi + p + p_\Phi) \dot{a} |_{y=0},$$

where

$$\rho_\Phi = \left[ \frac{1}{2} \dot{\Phi}^2 + n^2 V(\Phi) \right] |_{y=0},$$

$$p_\Phi = a^2 \left[ \frac{1}{2n^2} \dot{\Phi}^2 - V(\Phi) \right] |_{y=0}. $$

One may now proceed to obtain the cosmological equations by taking the gauge

$$n(0, t) = 1,$$

and performing the transformation

$$t = \int_0^t n(0, \tau) d\tau,$$

of the time coordinate. This gauge is convenient because it gives the usual cosmological time on the brane. Consequently, we find that our basic dynamical variable is $a(y, t)$ with $n(y, t)$ given by

$$n(y, t) = \frac{\dot{a}(y, t)}{a(0, t)}.$$  (26)

The basic set of cosmological equations in the present setting for a brane scalar field without a cosmological constant in the bulk now become

$$\lim_{\epsilon \to 0^+} [\partial_a a]_{y=\epsilon(t)} = \frac{m_v^2}{2m_v^{v+1}} (v - 1) \left[ \frac{\dot{a}^2(0, t)}{a(0, t)} + \frac{k}{a(0, t)} \right] |_{y=0}$$

$$- (\rho + \rho_\Phi) a(0, t) \right|_{y=0},$$

$$I^+ = [\dot{a}^2(0, t) - a^2(y, t) + k] a_v^{-1}(y, t) |_{y>0},$$

$$I^- = [\dot{a}^2(0, t) - a^2(y, t) + k] a_v^{-1}(y, t) |_{y<0},$$

$$\nabla_\mu \nabla^\mu \Phi = \frac{dV(\Phi)}{d\Phi},$$

$$n(y, t) = \frac{\dot{a}(y, t)}{a(0, t)}.$$  (31)

It is appropriate at this point to discuss the cosmology in the DGP model by taking $I^+ = I^-$. The cosmological equations in this framework for a $(v - 1)$-dimensional space are given by

$$\dot{a}^2(0, t) + k = \frac{2(\rho + \rho_\Phi)}{a^2(0, t)} v(\rho - 1)m_v^{-1},$$

$$\dot{\phi} + \frac{3}{2} \dot{a}(0, t) \dot{\phi} + \frac{dV(\Phi)}{a(0, t)} = 0,$$

$$I = [\dot{a}^2(0, t) - a^2(y, t) + k] a_v^{-1}(y, t),$$

$$n(y, t) = \frac{\dot{a}(y, t)}{a(0, t)}.$$  (35)

Eqs. (34) and (35) may now be used, taking $v = 3$, to obtain the components of the metric

$$a^2(y, t) = a^2(0, t) + (\dot{a}^2(0, t) + k)y^2$$

$$+ 2\sqrt{(\dot{a}^2(0, t) + k)a^2(0, t) - I y},$$

$$n(y, t) = \left[ a(0, t) + \dot{a}(0, t)y^2$$

$$+ a(0, t) \frac{\dot{a}(0, t) + 2\sqrt{(\dot{a}^2(0, t) + k)a^2(0, t) - I}}{\sqrt{2a^2(0, t) + k} a(0, t) y} \right] \frac{1}{a(y, t)},$$

$$The form of these equations becomes particularly simple for $I = 0$, that is

$$a(y, t) = a(0, t) + \sqrt{\dot{a}^2(0, t) + ky},$$

$$n(y, t) = 1 + \frac{\dot{a}(0, t)}{\sqrt{\dot{a}^2(0, t) + ky}}.$$  (39)

4. $R^{-1}$ inspired DGP scenario

To progress further, the form of the potential $V(\Phi)$ should be specified. Motivated by theories of modified gravity where a $R^{-1}$ term is present in the action and with an eye on the effects
that such a modification may have on the DGP brane scenarios we take the potential as [21]
\[ V(\Phi) \approx \mu^2 m_3^2 \exp(-\sqrt{3/2}\Phi/m_3). \]  
(40)

This is the form of the potential one encounters when studying theories of modified gravity with a Lagrangian of the form \( \mathcal{L}(R) = R - \frac{L}{R^2} \) and we shall concentrate in the limit of small \( R \) in the Einstein frame. Our aim is to obtain explicitly the components of the metric on the brane, i.e. Eqs. (36) and (37) in the limit of small curvature, and show that the results are consistent with an accelerating universe. Note that one may obtain an effective mass for the scalar field, \( \mu \), which is the cosmological constant in \( 4D \) gravity [21]. Hence, for models of modified gravity with any function of the Ricci scalar, \( \mathcal{L}(R) \), within the framework of the discussion at hand, we must take into account the contribution of the vacuum energy (cosmological constant) to the gravitational potential.

To proceed, we consider the evolution of the scale factor with time on the brane. From Eqs. (32) and (33) for the spatially flat FRW metric and setting \( \nu = 3 \), we can write
\[ 3H_0^2 = \frac{1}{m_3^2} (\rho + \rho_\phi), \]  
(41)

and
\[ \dot{\Phi} + 3H_0 \Phi + \frac{dV(\Phi)}{d\Phi} = 0, \]  
(42)

where \( H_0 \equiv \frac{\dot{a}(0,t)}{a(0,t)} \) and for \( n(0,t) = 1 \) (on the brane) we have
\[ \rho_\phi = \frac{1}{2} \dot{\Phi}^2 + V(\Phi). \]  
(43)

We must now solve the system of Eqs. (41) and (42) with \( \rho = 0 \). Substituting potential (40) into Eqs. (41) and (42), we obtain the evolution of the scale factor on the brane
\[ a(0,t) \propto t^{4/3}, \]  
(44)

together with that of the scalar field
\[ \dot{\Phi} \propto -\frac{4}{3} \ln t. \]  
(45)

As can be seen, Eq. (44) predicts a power-law acceleration on the brane. This result is consistent with the observational results similar to quintessence with the equation of state parameter \(-1 < w_{DE} < -\frac{1}{3} \) [23].

To continue, we find the evolution of \( a(y,t) \) and \( n(y,t) \) everywhere in spacetime. Thus, using Eqs. (44) and substituting into Eqs. (36) and (37), one finds
\[ a^2(y,t) = C \left[ y^{8/3} + \left( \frac{16}{9} \right) y^{2/3} y^2 \right] + 2 \sqrt{\left( \frac{16}{9} \right) C} t^{4/3} y - I y \]  
(46)

and
\[ n(y,t) = C \left[ t^{4/3} + \left( 4 \right) y^{1/3} \right] \]  
(47)

In the particular case \( I = 0 \), we obtain
\[ a(y,t) = C \left[ t^{4/3} + 3 t^{1/3} \right], \]  
(48)

and
\[ n(y,t) = C \left[ 1 + \frac{y}{3y} \right], \]  
(49)

where \( C \) is the proportionality constant. Note that for \( y = 0 \), Eqs. (48) and (49) reduce to (44) and \( n(0,t) = 1 \) respectively. There appears coordinate singularities on the space-like hyper-cone \( y = \pm 3t \). This is presumably a consequence of the fact that the orthogonal geodesics emerging from the brane (which we used to set up our Gaussian normal system, \( b^2 = 1 \)) do not cover the full five-dimensional spacetime.

5. Conclusions

Brane models provide a somewhat exotic, yet interesting extension of our parameter space for gravitational theories. In this work we have considered the DGP model with a brane scalar field. The scalar field was motivated and inspired by a desire to study the effects of modified gravity, represented by a term like \( R^{-1} \). We have shown that this model predicts that the mass in the gravitational potential is modified by the addition of the mass of such a scalar field. The cosmological evolution of this model was also studied by solving the relevant dynamical equations and the components of the metric was obtained in the limit of small curvature. The evolution of the universe in such a scenario seems to be consistent with the present observations, predicting an accelerated expansion.

References

E. Floratos, C. Kokorelis, hep-th/0607217.
Gamma ray burst constraints on ultraviolet Lorentz invariance violation

Tina Kahniashvili a,b,*, Grigol Gogoberidze b, Bharat Ratra a

a Department of Physics, Kansas State University, 116 Cardwell Hall, Manhattan, KS 66506, USA
b National Abastumani Astrophysical Observatory, 2A Kazbegi Ave., GE-0160 Tbilisi, Georgia

Received 28 September 2006; accepted 18 October 2006
Available online 7 November 2006
Editor: M. Cvetić

Abstract

We present a unified general formalism for ultraviolet Lorentz invariance violation (LV) testing through electromagnetic wave propagation, based on both dispersion and rotation measure data. This allows for a direct comparison of the efficacy of different data to constrain LV. As an example we study the signature of LV on the rotation of the polarization plane of γ-rays from gamma ray bursts in a LV model. Here γ-ray polarization data can provide a strong constraint on LV, 13 orders of magnitude more restrictive than a potential constraint from the rotation of the cosmic microwave background polarization proposed by Gamboa, López-Sarrión, and Polychronakos [J. Gamboa, J. López-Sarrión, A.P. Polychronakos, Phys. Lett. B 634 (2006) 471].

Available online 7 November 2006
Published by Elsevier B.V.
PACS: 11.30.Cp; 98.70.Rz; 98.70.Vc

Lorentz invariance violation (LV) has been proposed as a possible modification of the standard model of particle physics and cosmology (for recent reviews see Refs. [2–4]). Various LV mechanisms have been considered, including those motivated by phenomenological quantum gravity, string theory, non-commutative geometry, and through a Chern–Simons coupling (for a review see Section 2 of Ref. [3]). LV can influence particle propagation (the dispersion relation), result in rotation of linear polarization (birefringence), and affect the interaction of particles (including resulting in photon decay and vacuum Čerenkov radiation) [4]. These effects can be used to probe LV; for reviews of current and future tests see Refs. [2–4].

The assumed LV mechanism determines the kind of measurements required to test the model. Here we study frequency-dependent Faraday-like rotation of gamma ray burst (GRB) γ-ray and X-ray photon polarization in the context of ultraviolet LV. For discussions of such high energy LV see Refs. [5–7]. Refs. [1,8,9] study a generalized electromagnetism motivated by this kind of LV. On the other hand, LV associated with a Chern–Simons interaction [10,11] affects the complete spectrum of electromagnetic radiation, not just the high-frequency part, and induces a frequency-independent polarization-plane rotation (see Section 4 of Ref. [12]).

In this Letter we present a general formalism for LV testing that encompasses both rotation measure (RM) and photon dispersion measure (DM)1 observations. This formalism is based on an analogy with electromagnetic (EM) wave propagation in a magnetized medium, and extends previous works [5,10,15,20]. We show that the Gamboa et al. (GLP), [1], LV model is more tightly constrained by RM data than by DM data. The LV model of Myers and Pospelov (MP), [7], can be tightly constrained by GRB γ-ray DM and RM observations. The highly-variable γ-ray flux of energetic GRB photons propagating over cosmological distances make GRBs a powerful cosmological probe [13] (for reviews of cosmological tests involving GRBs, see Refs. [3,21]). Testing LV through RM observations of GRB 0370-2693/S – see front matter Published by Elsevier B.V.
doi:10.1016/j.physletb.2006.10.041

* Corresponding author.
E-mail address: tinatin@phys.ksu.edu (T. Kahniashvili).

1 The DM test is based on the LV effect of a phenomenological energy-dependent photon speed [13] or a modified electron dispersion relation. See Ref. [14] for reviews and Refs. [15–18] for recent studies of this effect; related early discussion include Ref. [19]. (Refs. [13,16,18] consider LV models in which rotational and translational invariance are preserved but boost invariance is broken.)
polarization was proposed, [22,23], after the reported observation of highly linearly polarized γ-rays from GRB021206 [24]; this measurement has been strongly contested [25]. On the other hand, Ref. [26] recently presented evidence that the γ-ray flux from GRB 930131 and GRB 960924 is consistent with polarization degree > 35% and > 50%, respectively. Since the issue of polarization of GRB γ-rays still remains uncertain, we also discuss using future X-ray RM observations. See Refs. [31] for other RM tests.

We first consider the ultraviolet LV model of GLP [1]. Breaking Lorentz invariance leads to a modification of the Maxwell equations [7,9], and in vacuum they become [1]

\[
\nabla \cdot \mathbf{B} = 0, \quad \nabla \times \mathbf{B} = \dot{\mathbf{E}}, \\
\nabla \cdot \mathbf{E} = 0, \quad \nabla \times \mathbf{E} + (\mathbf{g} \cdot \nabla)\dot{\mathbf{E}} = -\dot{\mathbf{B}}. 
\]

(1)

Here an overdot represents a derivative with respect to conformal time \( t \), \( \mathbf{g} \) is the LV vector related to the non-zero commutator of gauge potentials [1], \( \mathbf{B} \) is the magnetic field, and \( \mathbf{E} \) is the electric field that couples to matter in the usual way but is not related to the gauge potential in the usual way [1]. To account for the expansion of the Universe, we have to specify how \( \mathbf{g} \) scales in the expanding Universe. In conventional electrodynamics the expansion of the Universe is accounted for by a conformal rescaling of physical quantities, i.e. \( \mathbf{B} \rightarrow \mathbf{B}, \mathbf{E} \rightarrow a^2 \mathbf{E} \), where \( a \) is the scale factor [32]. Assuming that the GLP model is conformally invariant, the expansion may be accounted for by rescaling \( \mathbf{g} \rightarrow \mathbf{g}/a \), while the components of the physical electric and magnetic field are diluted as \( 1/a^2 \). On the other hand if the GLP model also violates conformal invariance, it is due to a small effect and so the expansion can be accounted for as above. So GLP LV results in only the Bianchi identity being modified.

In this model the equations for EM wave propagation in vacuum are

\[
(\omega^2 - k^2)\delta_{ij} - i\omega k_\epsilon \epsilon_{ij} k \cdot \mathbf{g}] E_j(k) = 0, \\
k_j E_j(k) = 0. 
\]

(2)

Here \( \epsilon_{ij} \) is the totally antisymmetric symbol, Latin indices denote space coordinates, \( i \in (1, 2, 3) \), \( \omega \) is the angular frequency of the EM wave measured today, and \( k \) is the wavevector. When transforming between position and wave-number spaces we use

\[
E_j(k) = e^{i\omega t} \int d^3x \ e^{i k \cdot x} E_j(x, t), \\
e^{i\omega t} E_j(x, t) = \int \frac{d^3k}{(2\pi)^3} e^{-i k \cdot x} E_j(k). 
\]

The \( e^{i\omega t} \) prefactor describes rapidly varying (compared to the cosmological expansion time) EM waves.

A linearly polarized wave can be expressed as a superposition of left (L) and right (R) circularly polarized (CP) waves.

Using the polarization basis of Section 1.1.3 of Ref. [33], Eqs. (2) become, for LCP (\( E^+ \)) and RCP (\( E^- \)) waves,

\[
(\omega^2 - k^2 \mp \omega k^2 \hat{\mathbf{k}} \cdot \mathbf{g}) E^\pm = 0.
\]

(4)

A similar dispersion relation, in a D-brane recoil model, has been obtained in Ref. [17]. To account for the phenomenological LV of an energy-dependent photon speed [3,4,7,12,22], we add photon-spin-dependent \( \mp \gamma(k) k^2 E^\pm(k) \) to the left-hand side of Eqs. (2) [23]. Here (Eq. (5) of Ref. [18])

\[
\gamma(k) = \left( \frac{\hbar k}{\xi_{\text{mpl}}} \right)^q,
\]

(5)

where \( m_{\text{pl}} \) is the Planck mass, \( \hbar \) is Planck’s constant, \( \xi \) is a dimensionless constant that determines the LV energy scale, and \( q \) is a model-dependent number. This modification may be viewed as an effective photon “mass” that makes the photon speed less (greater) than the low energy speed of light \( c \) for the RCP (LCP) waves.

To keep the formalism simple we consider an EM wave propagating in the \( z \) direction with \( k = (0, 0, k) \), and with the LV vector oriented along the \( z \) axis, i.e., \( g = (0, 0, g) \). Eqs. (4) lead to the dispersion relations

\[
\omega^2 = k^2 [1 \pm \gamma(k) \pm g\omega],
\]

(6)

and in this case \( E^\pm = (E_x \pm i E_y)/\sqrt{2} \). We now draw an analogy with the propagation of a high-frequency EM wave in a magnetized plasma. High-frequency RCP and LCP waves propagating in the \( z \) direction in an homogeneous magnetic field directed along the \( z \) axis obey [36]

\[
1 - \frac{\varepsilon_1}{n^2} E_x(k) - i \frac{\varepsilon_2}{n^2} E_y(k) = 0, \\
i \frac{\varepsilon_2}{n^2} E_x(k) + \left( 1 - \frac{\varepsilon_1}{n^2} \right) E_y(k) = 0.
\]

(7)

(8)

Here \( n = k/\omega \) is the refractive index and \( \varepsilon_1 \) and \( \varepsilon_2 \) are components of the electric permittivity or dielectric tensor \( \epsilon_{ij} \),

\[
\varepsilon_1 = \varepsilon_{xx} = \varepsilon_{yy} = 1 + \frac{\omega_p^2}{\omega_e^2 - \omega^2}, \\
\varepsilon_2 = \varepsilon_{yx} = -\varepsilon_{xy} = \frac{\omega_e}{\omega_e^2 - \omega^2},
\]

(9)

where \( \omega_p \) and \( \omega_e \) are the plasma and electron cyclotron angular frequencies (see Section 4.9 of Ref. [36]).

In the magnetized plasma case an homogeneous magnetic field induces a phase velocity difference between LCP and RCP waves and so causes rotation of the polarization plane. Also, in this case, the group velocity of an EM wave differs from \( c \) and so results in time delay. These two independent

---

2 For a review of models for generating polarized γ-rays from GRBs see Sections V.F and V.L of Ref. [21]; more recent discussions include Ref. [27]. For discussions of hard X-ray and γ-ray polarimetry see Refs. [28,29].

3 Ref. [30] predicts linearly polarized X-rays from flares following prompt GRB γ-ray emission.

4 In this case the modification of Maxwell equations does not preserve conformal invariance [34].

5 Ref. [35] argues that the much-studied \( q = 1 \) case is almost ruled out by Crab nebula X-ray polarimetry data.

6 This is motivated by the fact that LV generates an homogeneous magnetic field [9,34].
DM and RM effects can be expressed in terms of refractive indices, $n_{l,R} = (\epsilon_1 \mp \epsilon_2)^{1/2}$, where the sum (lower sign) corresponds to the RCP wave [36]. As a consequence the LCP and RCP wavevectors are $k_{l,R} = \omega n_{l,R}$. Both DM and RM effects depend on the photon travel distance $\Delta l$ and are expressed through

$$\Delta l_{l,R} = \Delta l \left(1 - \frac{\partial k_{l,R}}{\partial \omega}\right),$$

(10)

$$\Delta \phi = \frac{1}{2}(k_L - k_R)\Delta l.$$  

(11)

Here $\Delta l_{l,R}$ is the difference between the LCP (RCP) photon travel time and that for a “photon” which travels at $c$, and $\Delta \phi$ is the polarization-plane rotation angle.

We can rewrite Eqs. (2) and (3) for the LV case in a form similar to Eqs. (7) and (8) for a magnetized plasma. Define two dimensionless quantities $\epsilon = (\epsilon_1 \mp \epsilon_2)^{1/2}$ in these quantities. To linear order, $\Delta l$ becomes

$$\Delta l_{l,R} \approx \frac{\Delta l}{2}(1 \mp 2/\omega \gamma(k)),$$  

(12)

$$\Delta \phi \approx \frac{\Delta l}{2\omega \gamma(k)}.$$  

(13)

These expressions agree with those obtained earlier in Refs. [18,22,23]. DM and RM measurements can be used to constrain $\gamma$. DM testing of LV through the time delay of GRBs photons has been widely discussed (for a recent review see Ref. [4]) and so is not discussed here.

When $\gamma(k) \ll g\omega$, as in the MP model [7], Eqs. (10) and (11) become

$$\Delta l_{l,R} \approx \frac{\Delta l}{2}(1 + g)\gamma(k),$$  

(12)

$$\Delta \phi \approx \frac{\Delta l}{2\omega \gamma(k)}.$$  

(13)

and Eq. (10) for the time delay gives

$$\Delta l_{l,R} \approx \frac{g\omega \Delta l}{2}.$$  

(15)

DM and RM measurements constrain the value of $\epsilon_2$ (or $g\omega$), but the dependence on frequency is different, with the constraints from the RM test being the strongest for high-frequency waves.

For “classical” Faraday rotation $\Delta \phi \sim \omega^{-2}$, [37], and the effect is the strongest for low-frequency waves. For GLP LV $\Delta \phi \sim \omega^2$ and the effect is the strongest for high-frequency waves. Ref. [1] suggests using cosmic microwave background (CMB) polarization data to test GLP LV, as was previously proposed to detect a primordial cosmological magnetic field [32, 37] and test for CPT violation [38]. We argue below that GRB $\gamma$-rays polarization measurements will give a much stronger bound on this kind of LV. On the other hand, lower-frequency CMB polarization data may be used to constrain LV induced by a Chern–Simons coupling since in this case the RM is frequency independent [10] (this will complement the limit obtained from radio galaxy RM data [10]).

It should be possible to measure a $\Delta \phi \sim 10^{-2}$ rad. For CMB radiation with $\omega \sim 10^{11}$ Hz and for photon travel distance $\Delta l \sim 1.3 \times 10^{10}$ y, the RM GLP LV constraint, Eq. (14), indicates that one may probe to $g_{\text{CMB}} \sim 10^{-18}$ GeV$^{-1}$.

For GRB $\gamma$-rays with $\omega \sim 10^{19}$ Hz and $\Delta l \sim 3 \times 10^9$ y, even with less accurate RM data with, say, $\Delta \phi \sim 1$ rad, Eq. (14) shows that there is detectable LV down to $g_{\text{GRB}} \sim 10^{-31}$ GeV$^{-1}$.

In the GLP model GRB $\gamma$-ray data can probe 13 orders of magnitude higher in energy than can CMB data. Note that synchrotron radiation RM data at $\omega = 340$ GHz [39] from Sagittarius A* at $\Delta l \approx 2.5 \times 10^4$ y with $\Delta \phi \approx 0.5$ rad gives the weaker constraint $g_{\text{Sag}} \approx 10^{-11}$ GeV$^{-1}$. The polarization data at the optical band from active galactic nuclei give 8 magnitudes weaker limits than GRB future data.

To compare the relative efficacy of RM and DM data at probing LV, we consider the ratios of the same-source DM and RM data LV limits for the two characteristic LV quantities $\xi^{-1}$ and $g$,

$$r_\xi = \frac{\xi_{\text{DM}}^{-1}}{\xi_{\text{RM}}^{-1}}, \quad r_g = \frac{g_{\text{DM}}}{g_{\text{RM}}}.$$  

(18)

The constraints on $\xi^{-1}$ in the case when $\gamma \approx g\omega$ and $k \sim \omega$ can be obtained from Eqs. (5), (12), and (13),

$$\xi_{\text{DM}} = \frac{h\omega}{m_p l} \left[\left(\frac{q + 1}{\omega} 2\Delta l_{l,R}\right)^{1/4},\right.$$

(19)

$$\xi_{\text{RM}} = \frac{h\omega l^{1+1/4}}{m_p l} \frac{\Delta l}{2\Delta \phi}.$$  

(19)

The constraints on $g$ when $\gamma \ll g\omega$ can be obtained from Eqs. (14) and (15),

$$\frac{\Delta l_{l,R}}{\omega \Delta l} = \frac{\Delta l_{l,R}}{2\Delta \phi}, \quad \frac{\Delta l_{l,R}}{\omega \Delta l} = \frac{2\Delta \phi}{\omega^2 \Delta l}.$$  

(20)

We first consider GLP LV where $\gamma \ll g\omega$. Using the GRB $\gamma$-ray parameters mentioned above, taking $|\Delta l_{l,R}| = 10^{-4}$ s as the current accuracy of time delay data [18], and assuming $\Delta \phi = 1$ rad, $|r_{\text{GRB}}| = |\Delta l_{l,R}|/(2\Delta \phi) \sim 10^{14}$. So in this case the limit from RM data is the strongest. If one wishes to constrain $g$ using GRB DM and CMB RM data, then $g_{\text{GRB}} \sim \sqrt{c_{\text{CMB}}/\xi_{\text{RM}}} \approx 0.2$, so both are almost equally good tests for LV.
In the opposite case when $\gamma \gg g\omega$, if DM and RM data from the same source are used, 
\[ r_{\xi}^{L,R} = \left[ \mp g\omega \Delta t_{\xi,R} \right]^{1/q} / (q+1) \Delta \phi \]  
(21)
Conventionally two cases are considered, the linear case with $q = 1$ [15,18], and the quadratic case with $q = 2$ [18,22]. For $q = 1$ Eq. (21) reads for GRB $\gamma$-rays $|r_{\xi}^{\text{GRB}}| = \omega|\Delta t_{\xi,R}|/(2\Delta \phi) \sim 10^{14}$. Note that our $\xi$ is the inverse of the $\xi$ of Ref. [15] and coincides with the $\xi$ of Ref. [18]. Using the GRB $\gamma$-ray parameters considered above, we see that CMB polarization RM data may slightly improve the $\xi$ limit obtained from GRB $\gamma$-ray DM data [18]. The improvement will be much more significant if GRB $\gamma$-ray RM data is used [4,22]. For the $q = 1$ MP model [7] $r_{\xi}$ and $r_{\phi}$ are the same order of magnitude; i.e., RM data used for frequency $\omega > 2\Delta \phi/(|\Delta t_{\xi,R}|)$ results in similar limits on $g$ and $\xi^{-1}$. With $q = 1$, as a consequence of the frequency dependence $|r_{\phi,\xi}| \propto \omega$, high-frequency data result in more restrictive constraints. For the $q = 2$ case, $r_{\xi} \propto \sqrt{\omega}$ and $r_{\phi} \propto \omega$, so the potential limit on $\xi^{-1}$ from GRB $\gamma$-ray RM data is 6–7 orders of magnitude better than that from DM data [22].

In summary, we present a unified general treatment of both LV DM and RM tests by analogy with EM wave propagation in a magnetized plasma. This treatment does not depend on the LV model, and allows simultaneous consideration of different LV mechanisms. We considered conventional ultraviolet LV, i.e., linear MP, quadratic MP, and GLP models. For these models, RM data provide better limits than DM data, (the improvement is ~ 100 for linear MP and GLP LV, ~ 10 for quadratic MP LV, if $\omega > 100$ kHz), and the improvement increases by using higher frequency EM wave data (for an arbitrary MP model $r_{\phi} \propto \omega^{1/3}$ and thus RM test efficacy decreases as $q$ increases). Future $\gamma$- and X-ray RM data from distant objects, such as GRBs, quasars, or blazars hold great promise for testing and strongly constraining LV.

Acknowledgements

It is our pleasure to thank G. Melikidze for fruitful comments and suggestions. We also thank G. Gabadadze, A. Kosowsky, I. Litvinyuk, I. Mitrofanov, A. Sakharov, L. Samushia, S. Sarkar, T. Vachaspati, and L. Weaver for helpful discussions. T.K. and B.R. acknowledge support from DOE grants DE-FG02-00ER45824 and DE-FG03-99EP41093. T.K. and G.G. acknowledge partial support from INTAS grant 06-1000017-9258.

References


Weak effects in proton beam asymmetries at polarised RHIC and beyond

S. Moretti * , M.R. Nolten, D.A. Ross

School of Physics and Astronomy, University of Southampton Highfield, Southampton SO17 1BJ, UK

Received 30 September 2005; received in revised form 20 June 2006; accepted 20 June 2006

Available online 2 November 2006

Editor: N. Glover

Abstract

We report on a calculation of the full one-loop weak corrections through the order $\alpha^2 S \alpha_W$ to parton–parton scattering in all possible channels at the Relativistic Heavy Ion Collider (RHIC) running with polarised $pp$ beams (RHIC-Spin). This study extends the analysis previously carried out for the case of $2 \rightarrow 2$ subprocesses with two external gluons, by including all possible four-quark modes with and without an external gluon. The additional contributions due to the new four-quarks processes are extremely large, of order 50 to 100% (of either sign), not only in the case of parity-violating beam asymmetries but also for the parity-conserving ones and (although to a more limited extent) the total cross section. Such $O(\alpha^2 S \alpha_W)$ effects on the CP-violating observables would be an astounding 5 times larger for the case of the LHC with polarised beams—which has been discussed as one of the possible upgrades of the CERN machine—whereas they would be much reduced for the case of the CP-conserving ones as well as the cross section.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Electroweak physics; Parity violation; Polarised $pp$ scattering; RHIC

The purely weak component of Electro-Weak (EW) interactions is responsible for inducing parity-violating effects in jet observables, detectable through asymmetries in the cross section, which are often regarded as an indication of physics beyond the Standard Model (SM) [1]. These effects are further enhanced if polarisation of the incoming beams is exploited, like at the BNL machine mentioned in the abstract [2,3]. There have also been some discussions [4,5] on the idea of polarising the proton beams at the Large Hadron Collider (LHC) as one of the possible upgrades of the CERN machine, though no proposal has been put forward yet. At either machine, comparison of theoretical predictions involving parity-violation with experimental data can be used as a powerful tool for confirming or disproving the existence of some beyond the SM scenarios, such as those involving right-handed weak currents [6], contact interactions [7] and/or new massive gauge bosons [8–10].

In view of all this, it becomes of crucial importance to assess the quantitative relevance of weak effects entering via $O(\alpha^2 S \alpha_W)$ the fifteen possible $2 \rightarrow 2$ partonic subprocesses responsible for jet production in hadronic collisions, namely:

\begin{align}
    gg & \rightarrow q\bar{q}, \\
    gq & \rightarrow gg, \\
    qg & \rightarrow qg, \\
    \bar{q}g & \rightarrow \bar{q}g, \\
    qq & \rightarrow qq, \\
    \bar{q}\bar{q} & \rightarrow \bar{q}\bar{q}, \\
    qQ & \rightarrow qQ \text{ (same generation)}, \\
    \bar{q}Q & \rightarrow \bar{q}Q \text{ (same generation)}, \\
    qQ & \rightarrow qQ \text{ (different generation)}, \\
    \bar{q}\bar{Q} & \rightarrow \bar{q}\bar{Q} \text{ (different generation)}, \\
    qq & \rightarrow q\bar{q}, \\
    q\bar{q} & \rightarrow Q\bar{Q} \text{ (same generation)}. \\
\end{align}

* Corresponding author.

E-mail address: stefano@hep.phys.soton.ac.uk (S. Moretti).

0370-2693/S – see front matter © 2006 Elsevier B.V. All rights reserved.

doi:10.1016/j.physletb.2006.06.084

1 Note that in our treatment we identify the jets with the partons from which they originate.
\[ q\bar{q} \rightarrow Q\bar{Q} \] (different generation), \hspace{1cm} (13)
\[ q\bar{Q} \rightarrow q\bar{Q} \] (same generation), \hspace{1cm} (14)
\[ q\bar{Q} \rightarrow q\bar{Q} \] (different generation), \hspace{1cm} (15)

with \( q \) and \( Q \) referring to quarks of different flavours, limited to \( u-, d-, s-, c- \) and \( b- \) type (all taken as massless). While the first four processes (with external gluons) were already computed in Ref. [3], the eleven four-quark processes are new to this study.\(^2\)

Besides, these four-quark processes can be (soft and collinear) infrared divergent, so that gluon bremsstrahlung effects ought to be evaluated to obtain a finite cross section at the considered order. In addition, for completeness, we have also included the non-divergent \( 2 \to 3 \) subprocesses

\[ qg \rightarrow q\bar{q}q, \] \hspace{1cm} (16)
\[ \bar{q}g \rightarrow \bar{q}\bar{q}q, \] \hspace{1cm} (17)
\[ qg \rightarrow q\bar{Q}\bar{Q} \] (same generation), \hspace{1cm} (18)
\[ \bar{q}g \rightarrow \bar{q}\bar{Q}\bar{Q} \] (same generation). \hspace{1cm} (19)

By recalling that at the typical RHIC-Spin energies (e.g., \( \sqrt{s} = 300 \) and 600 GeV) the quark luminosity is much larger than the gluon one, it is clear that are processes with oncoming quarks that dominate the phenomenology of jet production here. In contrast, at the LHC (\( \sqrt{s} = 14 \) TeV), gluon-induced processes are largely dominant, particularly at low Bjorken-\( x \). As for what concerns the processes with external gluons, it is worth noticing that no CP-violation occurs at tree-level, so that \( O(\alpha_s^2 \alpha_W) \) is the first non-trivial order at which parity violation is manifest. Regarding four-quark processes, the following should be noted. Parity-violating contributions to channels (5)–(15) are induced already at tree-level, through \( O(\alpha_s^2 \alpha_W) \).\(^1\) Besides, all four-quark channels also exist through the CP-violating \( O(\alpha_s \alpha_W) \) [11], although subprocesses (9), (10), (13) and (15) only receive Cabibbo–Kobayashi–Maskawa (CKM) suppressed contributions at this accuracy (i.e., they mainly proceed via CP-conserving \( O(\alpha_s^2) \) interactions). Furthermore, notice in general that through \( O(\alpha_s^2 \alpha_W) \) there are many more diagrams available for channels (1)–(15) than via \( O(\alpha_s^3) \) or indeed \( O(\alpha_s \alpha_W) \) and \( O(\alpha_s^2 \alpha_W) \). Therefore, in terms of parton luminosity, simple combinatorics and power counting, one should expect the impact of \( O(\alpha_s^2 \alpha_W) \) terms to be large, certainly in parity-violating observables and possibly in parity-conserving ones as well. This is what we set out to test in this Letter\(^2\) for the case of RHIC and LHC. (See Refs. [12,13] for an account of these effects at Tevatron.)

Before proceeding further we ought to clarify at this stage that we have only computed purely weak effects at one-loop level through \( O(\alpha_s^2 \alpha_W) \), while in the case of tree-level processes via \( O(\alpha_s \alpha_W) \) and \( O(\alpha_s^2 \alpha_W) \) also the Electro-Magnetic (EM) contributions are included (and so are the interference effects between the two). This is why we are referring in this paper to the purely weak terms by adopting the symbol \( \alpha_W \), while reserving the notation \( \alpha_{EM} \) for the full EW corrections. Here then, we will have \( \alpha_W \equiv \alpha_{EM}/\sin^2\theta_W \) (with \( \alpha_{EM} \) the Electro-Magnetic (EM) coupling constant and \( \theta_W \) the weak mixing angle) while \( \alpha_{EW} \) will refer to the appropriate composition of QED and weak effects as dictated by the SM dynamics.

We have not computed one-loop EM effects for two reasons. Firstly, their computation would be technically very challenging, because the photon in the loop can become infrared (i.e., soft and collinear) divergent, thus requiring also the inclusion of photon bremsstrahlung effects, other than of gluon radiation. Secondly, \( O(\alpha_s^2 \alpha_{EM}) \) terms (in the above spirit, \( \alpha_{EM} \) signifies here only the contribution of purely EM interactions) would carry no parity-violating effects and their contribution to parity-conserving observables would anyway be overwhelmed by the well-known \( O(\alpha_s^2) \) terms [14] (see also [15,16]). However, notice that we are not including these next-to-leading order (NLO) QCD corrections either, as we are mainly interested in parity-violating beam asymmetries.

Since we are considering weak corrections that may be identified via their induced parity-violating effects and since we wish to apply our results to the case of polarised proton beams, it is convenient to work in terms of helicity Matrix Elements (MEs). Here, we define the helicity amplitudes by using the formalism discussed in Ref. [17]. At one-loop level such helicity amplitudes acquire higher order corrections from: (i) self-energy insertions on the fermions and gauge bosons; (ii) vertex corrections and (iii) box diagrams. The expressions for each of the corresponding one-loop amplitudes have been calculated using FORM [18] and checked by an independent program based on FeynCalc [19]. Internal gauge invariance tests have also been performed. The full expressions for the contributions from these graphs are however too lengthy to be reproduced here.

As already mentioned, infrared divergences occur when the virtual or real (bremsstrahlung) gluon is either soft or collinear with the emitting parton and these have been dealt with by using the formalism of Ref. [20], whereby corresponding dipole terms are subtracted from the bremsstrahlung contributions in order to render the phase space integral free of infrared divergences. The integrations over the gluon phase space of these dipole terms were performed analytically in \( d \)-dimensions, yielding pole terms which cancelled explicitly against those of the virtual graphs. There remains a divergence from the initial state collinear configuration, which is absorbed into the scale dependence of the Parton Distribution Functions (PDFs) and must be matched to the scale at which these PDFs are extracted. Recall that the remnant initial state collinear divergence at \( O(\alpha_S) \) is absorbed by the LO \( Q^2 \) dependence of the PDFs. Therefore, to \( O(\alpha_s^2 \alpha_W) \), it is sufficient, for the purpose of matching these divergences, to consider the LO PDFs. It is also consistent to use the values of the running \( \alpha_S \) obtained form the one-loop \( \beta \)-function. In order to display the corrections due to genuine weak interactions the same PDFs and strong coupling are used in the LO and NLO observables.

\(^{2}\) Note that \( gg \to gg \) does not appear through \( O(\alpha_s^2 \alpha_W) \) nor do \( qq \to QQ \).
\(^{3}\) Subprocesses (16)–(19) turn out to be numerically negligible at both machines and whichever the observable, so that we will not consider them in the remainder.
The self-energy and vertex correction graphs contain ultra-violet divergences that have been subtracted here by using the ‘modified’ Dimensional Reduction (DR) scheme at the scale $\mu = M_Z$. The use of DR, as opposed to the more usual ‘modified’ Minimal Subtraction (MS) scheme, is forced upon us by the fact that the $W$- and $Z$-bosons contain axial couplings which cannot be consistently treated in ordinary-dimensional regularisation. Thus the values taken for the running $\alpha_S$ refer to the DR scheme whereas the EM coupling, $\alpha_{EM}$, has been taken to be 1/128 at the above subtraction point. The one exception to this renormalisation scheme has been the case of the self-energy insertions on external fermion lines, which have been subtracted on mass-shell, so that the external fermion fields create or destroy particle states with the correct normalisation.

The top quark entering the loops in reactions with external $b$’s has been assumed to have mass $m_t = 175$ GeV and width $\Gamma_t = 1.55$ GeV. The $Z$ mass used was $M_Z = 91.19$ GeV and was related to the $W$-mass, $M_W$, via the SM formula $M_W = M_Z \cos \theta_W$, where $\sin^2 \theta_W = 0.232$. (Corresponding widths were $\Gamma_Z = 2.5$ GeV and $\Gamma_W = 2.08$ GeV.) For the strong coupling constant, $\alpha_S$, we have used the one-loop expression with $\Lambda^{(n_f=4)}_{\text{MS}}$ chosen to match the value required by the (LO) PDFs used. The latter were Gehrmann–Stirling set A (GSA) [21] and Glück–Reya–Stratmann–Vogelsang standard set (GRSV–STN) [22].

The following beam asymmetries, e.g., can be defined, depending on whether one or both beams are polarised:

\[ A_{LL} d\sigma \equiv d\sigma_{++} - d\sigma_{--} + d\sigma_{+-} - d\sigma_{-+}, \]
\[ A_L d\sigma \equiv d\sigma_+ - d\sigma_-, \]
\[ A_{PV} d\sigma \equiv d\sigma_- - d\sigma_+. \]  

(20)

The first is parity-conserving while the last two are parity-violating.\(^4\)

Fig. 1 shows the size of the $\mathcal{O}(\alpha_S^2 \alpha_W)$ effects relatively to the well-known LO results, the latter being defined as the sum of all $\mathcal{O}(\alpha_S^2)$, $\mathcal{O}(\alpha_S \alpha_{EM})$ and $\mathcal{O}(\alpha_W^2)$ contributions, for the case of RHIC, for two reference energies. Both the differential cross section and the above beam asymmetries are plotted, each as a function of the jet transverse energy. The $\mathcal{O}(\alpha_S^2 \alpha_W)$ corrections are already very large at cross section level, by reaching $-5$ ($-9$)% at $\sqrt{s} = 300$ (600) GeV, in the vicinity of $E_T = 120$ (240) GeV. Effects onto the $A_{LL}$ asymmetry are even larger, with maxima of $\approx 25$ (60)% for $E_T \approx 70$ (140) GeV, again, in correspondence of $\sqrt{s} = 300$ (600) GeV. In the case of both $A_L$ and $A_{PV}$, in regions away from the threshold at $E_T \approx M_W/4$ (where resonance effects emerge), there is no local maximum for positive or negative corrections, as both grow monotonically to the level of $+100$% (at low $E_T$) and $-50$ to $-70$% (at high $E_T$ and with increasing collider energy). All such effects should comfortably be observable at RHIC, for the customary values of integrated luminosity.

\(^4\) In the numerical analysis which follows we will assume 100% polarisation of the beams.
of 200 and 800 pb\(^{-1}\), in correspondence of \(\sqrt{s} = 300\) and 600 GeV [1].

At the LHC with polarised beams (but standard energy \(\sqrt{s} = 14\) TeV), the \(O(\alpha_s^2\alpha_W)\) corrections to the total cross section as well as the CP-conserving asymmetry are reasonably under control. In fact, they grow monotonically and reach the \(\approx -3\)\% and \(\approx 4\)\% at the kinematic limit of the jet transverse energy (as defined by the PDFs), respectively. However, it is debatable as to whether these effects can actually be disentangled, as we expect systematic experimental uncertainties to be of the same order. Away from the threshold at \(E_T \approx M_W/2\), \(O(\alpha_s^2\alpha_W)\) effects onto the parity-violating asymmetries are instead enormous, as they yield a \(K\)-factor increasing from \(-2\) to \(-4.5\), as \(E_T\) varies from 80 to 500 GeV. Despite the absolute value of the CP-violating asymmetries is rather small in the above interval, the huge LHC luminosity (10 fb\(^{-1}\) per year should be feasible for, say, a 70\% polarisation per beam [23]) would render the above higher order corrections manifest.

It is intriguing to understand the different behaviours of the \(O(\alpha_s^2\alpha_W)\) effects depending on the observable and the collider being considered. To this end, we have presented in Tables 1–2 the contributions to the \(E_T\) dependent cross section of subprocesses (1)–(15) through \(O(\alpha_s^2\alpha_W)\) separately, at both RHIC-Spin and LHC. The purpose of these tables is to illustrate that the leading partonic composition of the \(O(\alpha_s^2\alpha_W)\) corrections is markedly different at the two machines. While at RHIC the

Table 1
The contributions of subprocesses (1)–(15) to order \(\alpha_s^2\alpha_W\) with respect to the full LO result for the total cross section at RHIC-Spin, at \(E_T = 70\) GeV for \(\sqrt{s} = 300\) GeV and \(E_T = 140\) GeV for \(\sqrt{s} = 600\) GeV. Here, we have paired together the channels with identical Feynman diagram topology. We use GSA as PDFs and \(\mu = E_T/2\) as factorisation/renormalisation scale. Column (a) is the percentage contribution from the \(O(\alpha_s^2\alpha_W)\) corrections, column (b) is the percentage correction to the tree-level partonic subprocess and column (c) is the percentage contribution at the tree-level of that partonic subprocess to the differential cross section at the relevant \(E_T\).

<table>
<thead>
<tr>
<th>Subprocess</th>
<th>(\sqrt{s} = 300) GeV</th>
<th></th>
<th>(\sqrt{s} = 600) GeV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(a)</td>
<td>(b)</td>
<td>(c)</td>
</tr>
<tr>
<td>(gg \to gg)</td>
<td>1.73E-05</td>
<td>0.0267</td>
<td>0.0650</td>
</tr>
<tr>
<td>(1)</td>
<td>4.72E-05</td>
<td>0.0253</td>
<td>0.188</td>
</tr>
<tr>
<td>(2)</td>
<td>-0.0144</td>
<td>-0.0589</td>
<td>24.7</td>
</tr>
<tr>
<td>(3)–(4)</td>
<td>-0.411</td>
<td>-0.399</td>
<td>46.1</td>
</tr>
<tr>
<td>(5)–(6)</td>
<td>-1.57</td>
<td>-6.64</td>
<td>23.8</td>
</tr>
<tr>
<td>(7)–(8)</td>
<td>0.00710</td>
<td>0.00234</td>
<td>0.00131</td>
</tr>
<tr>
<td>(qq' \to QQ') or (\bar{q}q' \to \bar{Q}Q')</td>
<td>0.00710</td>
<td>0.00234</td>
<td>0.00131</td>
</tr>
<tr>
<td>(qq' \to Q\bar{Q}')</td>
<td>0.00710</td>
<td>0.00234</td>
<td>0.00131</td>
</tr>
<tr>
<td>Total</td>
<td>-1.94</td>
<td>-5.11</td>
<td></td>
</tr>
</tbody>
</table>

Table 2
The contributions of subprocesses (1)–(15) to order \(\alpha_s^2\alpha_W\) with respect to the full LO result for the total cross section at LHC, at \(E_T = 300\) GeV for \(\sqrt{s} = 14\) TeV. Here, we have paired together the channels with identical Feynman diagram topology. We use GSA as PDFs and \(\mu = E_T/2\) as factorisation/renormalisation scale. Column (a) is the percentage contribution from the \(O(\alpha_s^2\alpha_W)\) corrections, column (b) is the percentage correction to the tree-level partonic subprocess and column (c) is the percentage contribution at the tree-level of that partonic subprocess to the differential cross section at the relevant \(E_T\).

<table>
<thead>
<tr>
<th>Subprocess</th>
<th>(\sqrt{s} = 14) TeV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(a)</td>
</tr>
<tr>
<td>(gg \to gg)</td>
<td>-0.0315</td>
</tr>
<tr>
<td>(1)</td>
<td>-0.00386</td>
</tr>
<tr>
<td>(2)</td>
<td>-0.455</td>
</tr>
<tr>
<td>(3)–(4)</td>
<td>-0.112</td>
</tr>
<tr>
<td>(5)–(6)</td>
<td>-0.431</td>
</tr>
<tr>
<td>(7)–(8)</td>
<td>-0.0330</td>
</tr>
<tr>
<td>(9)–(10)</td>
<td>-0.0328</td>
</tr>
<tr>
<td>(11)</td>
<td>0.0466</td>
</tr>
<tr>
<td>(12)</td>
<td>-0.00316</td>
</tr>
<tr>
<td>(13)</td>
<td>0.0131</td>
</tr>
<tr>
<td>(14)</td>
<td>-0.0325</td>
</tr>
<tr>
<td>(qq' \to QQ') or (\bar{q}q' \to \bar{Q}Q')</td>
<td>0.00221</td>
</tr>
<tr>
<td>(qq' \to Q\bar{Q}')</td>
<td>0.000979</td>
</tr>
<tr>
<td>Total</td>
<td>-1.075</td>
</tr>
</tbody>
</table>
key role is played by subprocess (7), at the LHC the conspicuous rise of the gluon-luminosity enhances in turn the yield of channel (3) to a level comparable to that of mode (7).5 The hierarchy among the subprocesses seen in Tables 1–2, for fixed jet transverse energy, is characteristic across most of the available $E_T$ range at both colliders.) The $O(\alpha_S^2\alpha_W)$ corrections are particularly large for subprocesses (7)–(8) and (12), mainly in virtue of the large combinatorics involved at loop level (as anticipated earlier), with respect to the LO case.

The different behaviours seen in Figs. 1–2 can easily be interpreted in terms of the LO contributions. In this respect, as already mentioned, Tables 1–2 clearly make the point that the jet phenomenology at RHIC-Spin is dominated by subprocesses initiated by quarks only while at the LHC gluon-induced channels are generally predominant. At RHIC energies, LO production through order $\alpha_S^2$ is dominated by channel (5) whereas at the LHC the overwhelmingly dominant $\alpha_S^2$ channels are $gg \to gg$ (which is not subject to $O(\alpha_S^2\alpha_W)$ corrections, as already mentioned) and subprocess (3). Besides, channel (5) through $O(\alpha_S^2)$ is mainly concentrated at low $E_T$ while with growing $E_T$ the $O(\alpha_S\alpha_W)$ and—particularly—$O(\alpha_W^2)$ terms gain in relative importance. Furthermore, $O(\alpha_S^2)$ terms entering channel (5) do not contribute, obviously, to the parity-violating asymmetries. Therefore, it should not be surprising to see at RHIC-Spin that our corrections are very large in the case of the latter, where the LO term is $O(\alpha_S\alpha_W)$, respect to which the corrections computed here are suppressed only by one power of $\alpha_S$. We attribute instead the size of the $O(\alpha_S^2\alpha_W)$ effects on the cross section and the parity-conserving asymmetry again to the fact that through $O(\alpha_S^2\alpha_W)$ there are many more diagrams available for such channels than via $O(\alpha_S^2)$ or indeed $O(\alpha_S\alpha_W)$ and $O(\alpha_W^2)$. As for the LHC, the fact that $gg \to gg$ and subprocess (3) vastly dominates through $O(\alpha_S^2)$ the $d\sigma/dE_T$ distribution explains why $O(\alpha_S^2\alpha_W)$ corrections are limited to the percent level. In $A_{LL}$, which has no $gg \to gg$ component, $O(\alpha_S^2\alpha_W)$ effects become somewhat more visible in comparison.

Furthermore, in the case of the LHC, one should note the monotonic rise of the corrections with increasing jet transverse energy, for all observables studied, which can be attributed to the so-called Sudakov (leading) logarithms [24,25] of the form $\alpha_W\log^2(E_T^2/M_W^2)$, which appear in the presence of higher order weak corrections. These 'double logarithms' are due to a lack of cancellation of infrared (both soft and collinear) virtual and real emission in higher order contributions due to $W$-exchange, arising from a violation of the Bloch–Nordsieck theorem occurring in non-Abelian theories. (In fact, if events with real $Z$ radiation are vetoed in the jet sample, $\alpha_W\log^2(E_T^2/M_Z^2)$ terms would also affect the corrections [12].) Clearly, at LHC energies, $E_T$ can be very large, thus probing the kinematic regime of these logarithmic effects, which instead affected RHIC only very mildly. Combine then the effects of such large logarithms with the fact that $A_L$ and $A_{PV}$ receive no pure QCD contributions, and one can explain the enormous (and increasing with $E_T$) $O(\alpha_S^2\alpha_W)$ corrections to these two observables. In fact, recall that another way of viewing the $O(\alpha_S^2\alpha_W)$ terms computed here is as first order QCD corrections to the $O(\alpha_S\alpha_W)$ terms, which are the leading order contributions to $A_L$ and $A_{PV}$. From this perspective then, the large results reported here can be understood as large $O(\alpha_S)$ corrections. Furthermore, also recall here the following two aspects, already mentioned. Firstly, there are several partonic processes which are CKM suppressed at $O(\alpha_S\alpha_W)$ but which occur without CKM suppression at $O(\alpha_S^2\alpha_W)$. Secondly, processes involving external gluons and weak interactions occur for the first time at $O(\alpha_S^2\alpha_W)$.

As one of the purposes of polarised colliders is to measure polarised structure functions, in the ultimate attempt to reconstruct the proton spin, it is of some relevance to see how the $O(\alpha_S^2\alpha_W)$ results obtained so far for GSA compare against GRSV–STN. This is done in Figs. 3–4, where we have also adopted the different choice $\mu = E_{cm}(E_T)$ as factorisation/renormalisation scale, i.e., the centre-of-mass energy at parton level $\sqrt{s}$ (jet transverse energy), for the GSA (GRSV–STN) set. A comparison between the GSA curves in Figs. 3–4 and those in Figs. 1–2 reveals that the scale dependence of our corrections is not very substantial for a given PDF set (the same is true for the case of GRSV–STN). In contrast, depending on the choice of PDFs, corrections through $O(\alpha_S^2\alpha_W)$ can be very different for each observables studied at both RHIC-Spin and LHC, with the exception of the cross section in either case.

Altogether, the results presented here point to the extreme relevance of one-loop $O(\alpha_S\alpha_W^2)$ weak contributions for precision analyses of jet data produced in polarised proton–proton scattering at RHIC. We have confirmed that this would be the case also at a polarised LHC, which has been discussed as one of the possible upgrades of the CERN collider. The size of the afore-mentioned corrections, relative to the lowest order results, is rather insensitive to the choice of factorisation/renormalisation scale, yet it shows some sizable dependence on the polarised PDFs used. EM effects were neglected here because they are not subject to parity-violating effects. However, their computation is currently in progress. The inclusion of NLO terms from pure QCD, through $O(\alpha_S^3)$, is also in order, as they can produce effects of order 100%, and the parity-violating asymmetries, through in this case they will only amount to a rescaling (within a factor of 2 at the most) of the normalisation, not to a change in shape. We are now working towards the full $O(\alpha_S^3)$ results including beam polarisation effects [26].

Finally, extrapolation of our results to other collider energies, chiefly for RHIC, for operation at $\sqrt{s} = 200$ and 500 GeV, is straightforward. As expected, these results do not differ dramatically from those obtained for 300 and 600 GeV, respectively. Fig. 1 also shows the size of the corrections at these two additional energies, for our usual observables and default choice of PDFs and factorisation/renormalisation scale. For

---

5 The relevance of the latter throughout originates from the combination of a always sizable valence quark luminosity and a large Feynman diagram combinatorics, as opposed to, e.g., a gluon luminosity steeply increasing with the collider energy but combined with a small numbers of graphs [3].
Fig. 2. The dependence of the cross section as well as of the beam asymmetries on the jet transverse energy at tree-level (large frames) and the size of the one-loop weak corrections (small frames), at the LHC energy $\sqrt{s} = 14$ TeV. Notice that the pseudo-rapidity range of the jets is limited to $|\eta| < 2.5$ and the standard jet cone requirement $\Delta R > 0.7$ is imposed as well (although we eventually sum the two- and three-jet contributions). We use GSA as PDFs and $\mu = E_T / 2$ as factorisation/renormalisation scale.

Fig. 3. The dependence of the corrections to the cross section as well as the beam asymmetries on the jet transverse energy for two sets of PDFs, GSA and GRSV–STN, at the two RHIC-Spin energies $\sqrt{s} = 300$ (curves extending to 150 GeV) and 600 GeV (curves extending to 300 GeV). Notice that the pseudo-rapidity range of the jets is limited to $|\eta| < 1$ and the standard jet cone requirement $\Delta R > 0.7$ is imposed as well (although we eventually sum the two- and three-jet contributions). We use $\mu = E_{cm}$ ($E_T$) as factorisation/renormalisation scale in the case of GSA (GRSV–STN).
the purpose of emulating the effects of $O(\alpha_s^2)$ terms at whichever collider and energy, we make available our code on request.

Acknowledgements

We thank Ezio Maina for collaborative work in the early stages of this analysis.

References

The extraction of the bare triple-pomeron vertex; a crucial ingredient for diffraction

V.A. Khoze\textsuperscript{a,b}, A.D. Martin\textsuperscript{a,*}, M.G. Ryskin\textsuperscript{a,b}

\textsuperscript{a} Department of Physics and Institute for Particle Physics Phenomenology, University of Durham, DH1 3LE, UK
\textsuperscript{b} Petersburg Nuclear Physics Institute, Gatchina, St. Petersburg 188300, Russia

Received 11 October 2006; accepted 11 October 2006
Available online 30 October 2006
Editor: N. Glover

Abstract

The triple-pomeron coupling lies at the heart of the predictions for high energy diffractive processes. We explain why the existing determinations, which use single-particle inclusive hadronic data \(a+b\rightarrow a+Y\), underestimate the value of the bare coupling, due to the neglect of soft rescattering which populates the rapidity gaps. We describe how data for the process \(\gamma+p\rightarrow J/\psi+Y\) can be used to give a much more reliable estimate of the bare coupling. We use the existing, fragmentary, \(J/\psi\) data from HERA to show that the triple-pomeron coupling is probably about 3 times larger than the previous determinations. We emphasize the importance of an explicit measurement of the mass spectrum of the \(Y\) system which accompanies \(J/\psi\) production at HERA. The consequences for ultra high energy cosmic ray showers are mentioned.

The triple-pomeron coupling is a crucial ingredient for understanding diffraction. The value and the \(t\)-dependence of the vertex determine the asymptotic high energy behaviour of total cross sections and the cross section of diffractive dissociation. Moreover, in turn, the calculation of the rapidity gap survival factors, \(S^2\), for diffractive processes depends on the properties of the triple-pomeron vertex. The factor \(S^2\) depends on the particular diffractive process, as well as the values of its kinematic variables, and is found to be of size from about 0.01 to 1. An example of an analysis of soft high energy hadronic scattering, with the consequent calculation of the \(S^2\) factors, and their relation to the triple-pomeron vertex can be found in [1].

The existing determinations of the vertex, which were made many years ago,\textsuperscript{1} were based on the analyses of the cross sections for single particle hadronic inclusive processes

\(a+b\rightarrow c+Y\),

in the triple pomeron region where the leading hadron, \(c\), carries a large fraction \(x_L\) of the incoming momentum, see Fig. 1.

The production of the leading hadron with \(x_L\rightarrow 1\) selects pomeron exchange with trajectory \(\alpha_P(t)\), providing the quantum numbers of hadrons \(a\) and \(c\) are the same. For simplicity we take \(c=a\), which is generally the case. In addition the production of a system \(Y\) of high mass \(M_Y\) is described by pomeron exchange with trajectory \(\alpha_P(0)\); leading to a diagram with a triple-pomeron vertex, which is shown on the right-hand side of Fig. 1. In this formalism the \(a+b\rightarrow a+Y\) cross section is given by [4,5]

\[
M_Y^2 \frac{d\sigma^{SD}}{dt dM_Y^2} = \frac{S^2}{16\pi^2} \frac{G_a^2(t)G_b(0)g_{3P}(t)}{s} \left( \frac{M_Y^2}{s} \right)^{1+\alpha_P(0)-2\alpha_P(t)} \left( \frac{s}{s_0} \right)^{\alpha_P(0)-1} + \text{RRP + PPR terms}
\]

with \(s_0\equiv 1\ \text{GeV}^2\) and where \(\sqrt{s}\equiv W\) is the centre-of-mass energy of the incoming \(ab\) system. Here we use SD to denote single-particle diffractive dissociation. The triple-pomeron term (PPP) is shown explicitly in (2). There are also contributions RRP, PPR, ... in which secondary reggeons R replace the upper and/or lower pomeron exchanges. The RRP term may give some contribution as \(x_L\) decreases away from 1, and the PPR

\textsuperscript{*} Corresponding author.
E-mail address: a.d.martin@durham.ac.uk (A.D. Martin).
\textsuperscript{1} See, for example, [2], or the review [3].
term may give some effect if the mass $M_Y$ is not sufficiently large.

Unfortunately, the determinations of the triple-pomeron coupling $g_{3P}(0)$ from data for the hadronic $a + b \rightarrow c + Y$ processes [2] are, at best, just estimates. The problem is that no account is taken of the soft rescattering. In such a rescattering the leading hadron will produce new secondary particles and, as a consequence, its value of $x_L$ will be diminished. Nowadays this effect is called the rapidity gap survival factor $S^2$, since the secondaries populate and destroy the rapidity gap, $\Delta \eta \simeq \ln 1/(1 - x_L)$, between the leading hadron $c$ and the hadrons in the system $Y$. Since the $S^2$ was not accounted for explicitly, the determinations give only an effective value of the triple-pomeron vertex, which implicitly embodies an $S^2$ factor. However the probability of soft rescattering, and hence the $S^2$ factor, depends on the nature of the incoming and leading hadrons. Moreover the value of $S^2$ depends on the kinematical variables of the process [6]. Clearly it is important to find a way to measure the universal bare triple-pomeron vertex, free from these rescattering effects.

A good illustration of the situation is an analogous problem which arises in the analysis of leading neutron production at HERA [7,8]. There, to understand the process we need to allow for rescattering (absorptive) corrections to pure Regge $\pi$ exchange. The rescattering modifies the predictions of not only the absolute value, but also the $t$ dependence of leading neutron production. Quantitatively it reduces the cross section by a factor of 0.4, and changes the $t$ slope, $B_t$, by 1 GeV$^{-2}$, where $d\sigma/dt \sim \exp(B_t)$.

The $t$ dependence of the triple-pomeron vertex plays a crucial role in the solution of the so-called Finkelstein–Kajantie problem [9,10]. That is, if we neglect the $t$ dependence of the vertex and the gap survival factor, then it turns out, even in the case $\alpha_F(0) = 1$, that the cross section for multigap events grows as a power of the energy, and so violates the Froissart limit. Each additional gap brings a $\ln s$ factor arising from the integral over the gap size. The sum of these $\ln s$ factors leads to the power behaviour.

Two solutions were proposed. First, the weak coupling solution in which the vertex vanishes as $t \rightarrow 0$ [11]. In this case the $\ln s$ factor arising from the gap size is compensated by a $1/\ln s$ caused by the shrinkage of the $t$ distribution with increasing energy. Interestingly, it predicts that, at ultra high energies, the total cross sections $\sigma_{ab}$ will tend to a universal constant independent of $a$ and $b$ [12]. An alternative possibility is to suppress the multigap events by a strong absorptive correction, that is by decreasing the gap survival factors. In other words a large multigap cross section, calculated using a bare triple-pomeron vertex which is non-vanishing as $t \rightarrow 0$, is multiplied by a gap survival factor which decreases faster with energy than the bare cross section. This leads to the Froissart-like black disc limit, and is called the strong coupling solution [13].

The old hadron–hadron data on (1) show no indication that the triple-pomeron vertex vanishes as $t$ $\rightarrow 0$. However these data are inconclusive regarding the $t$ behaviour. The reasons are as follows. First, if it were to vanish, this behaviour would only be apparent at rather small $t$, $|t| \lesssim 1/R^2$, where $R$ is the typical transverse size of the incoming hadrons. Unfortunately, the data are not precise enough in this small $t$ region. Second, the effect is masked by the pomeron cut contribution—soft rescattering washes out the momentum transferred through the individual pomeron, and zero $p_t$ of the leading hadron does not mean that the momentum transfer in the bare triple-pomeron vertex is $< 1/R^2$. It has been shown [14] that after accounting for multipomeron effects, the old hadron–hadron data cannot eliminate the possibility that the bare triple-pomeron vertex vanishes as $t \rightarrow 0$.

It is clearly important to study the triple-pomeron interaction in a process where the rescattering effects are absent or strongly suppressed. Such a process, which is experimentally accessible, is proton dissociation in diffractive $J/\psi$ photo- (or electro-) production

$$\gamma + p \rightarrow J/\psi + Y. \quad (3)$$

The process is shown schematically in Fig. 2. By detecting the $J/\psi$ in the final state, we select the charm component of the photon wave function, which has a small size of $\sim 1/m_c$. Such a component has a small absorptive cross section. Therefore the probability of an additional rescattering is much weaker. Even in the case of a nuclear target the probability of $J/\psi$ rescattering is small [15,16]. Of course, there may be some ‘enhanced’ absorptive corrections (of the type indicated by the dotted pomeron-exchange line on the diagram on the right-hand side of Fig. 2). However the phenomenological analysis of leading neutron data [7] demonstrates that such corrections are small at the available HERA energies. Otherwise the probability to observe a leading neutron would strongly depend on the initial $\gamma p$ energy, $W_{\gamma p}$, in contradiction with the data. Moreover, to provide enough phase space for such an ‘enhanced’ correction, we would need to have sufficiently large rapidity in-
tervals (on either side of the dotted line) between the different vertices [17]. This corresponds to an extremely small value of \(1 - x_L \lesssim 10^{-2}\), which is beyond the reach of the present experiments.

Another advantage of the process \(\gamma + p \rightarrow J/\psi + Y\) is that, thanks to the small size of the \(J/\psi\) meson, the vanishing of the triple-pomeron vertex (if true) would reveal itself over a more extended region of \(t\). This possibility is practically already eliminated by the present data, see Fig. 4 of [18]. Thus we can conclude that the strong coupling solution of the triple-pomeron interaction is confirmed by the HERA \(J/\psi\) data.

Though the process \(\gamma + p \rightarrow J/\psi + Y\) is experimentally accessible at HERA, so far no dedicated measurements are available in the literature. However some information does exist, as the process has been considered as a background to elastic \(J/\psi\) production, \(\gamma + p \rightarrow J/\psi + p\). This somewhat fragmentary information allows us to make the following observations. We use \(\sigma_Y\) and \(\sigma_{el}\) to denote the two \(J/\psi\) production cross sections.

- The energy dependence of \(J/\psi\) production with proton dissociation and for the elastic process are consistent with each other. If we write
  \[
  \sigma_t \propto W^{\delta_Y},
  \]
  where \(W\) is the photon–proton centre-of-mass energy, then [19–22]
  \[
  \delta_Y \simeq 0.7 \pm 0.2, \quad \delta_{el} \simeq 0.7 \pm 0.05.
  \]
  Next, the \(M_Y\) dependence [19] for a fixed value of the ratio \(M_Y^2/W^2\) is consistent with the behaviour of hadronic total cross sections, that is
  \[
  \sigma_Y \propto (M_Y^2)^{\epsilon} \quad \text{with} \quad \epsilon \simeq 0.08.
  \]
  The above results are the properties expected from the triple-pomeron term \(^3\) in (2).

- The \(t\) slope of \(J/\psi\) production by proton dissociation is observed [19,21] to be rather small, \(b_Y \simeq 0.7 \pm 0.2\ \text{GeV}^{-2}\). In some sense, we can regard \(b_Y\) to be driven by the pomeron form factor. On the other hand, the slope for elastic \(J/\psi\) production, which is driven by the proton form factor, is measured [23,24] to be \(b_{el} \simeq 4.5 \text{GeV}^{-2}\). This indicates that the size of the pomeron and the triple-pomeron vertex are much less than that for the proton. This conclusion about the small size of the triple-pomeron vertex is confirmed by the data on \(\rho\) production, where the difference of slopes, \(\Delta b = b_{el} - b_Y \simeq 4 \text{GeV}^{-2}\) [25]. These properties justify the idea that the pure triple-pomeron coupling may be extracted from a triple-Regge analysis of data for \(J/\psi\) production with proton dissociation. The small size of the pomeron indicates that the rescattering corrections are indeed small. This information comes from a consideration of proton dissociation as a background process in the presentation of elastic \(J/\psi\) production data. However there is more detailed information on the \(t\) dependence of the proton dissociation process in Fig. 4 of [18], that we mentioned previously.

- A comparison of the cross sections is also revealing. For \(J/\psi\) production, an approximate estimate gives [18,21]
  \[
  \sigma_Y/\sigma_{el} \simeq 1 \pm 0.3,
  \]
  while for \(\rho\) production we have [24,26]
  \[
  \sigma_Y/\sigma_{el} \simeq 0.6 \pm 0.2.
  \]
  This difference in the values of the ratios is anticipated because strong absorption is expected in the case of \(\sigma_Y\) for \(\rho\) production.

The above summary of the existing, admittedly fragmentary, data on \(\gamma + p \rightarrow J/\psi + Y\) certainly supports a dominant triple-pomeron behaviour. So we may attempt an extraction of the bare triple-pomeron coupling, \(g_{3P}(0)\). We assume that the data for \(\sigma_Y\), collected [18] in the interval \(^4\) of \(M_Y\) from 2.5 GeV up to \(M_Y^2 = 0.1W^2\), is described just by the triple-pomeron term in (2).

\(^2\) We thank Michele Arneodo and Alessia Bruni for discussions concerning the data.

\(^3\) Note that the pomeron trajectories \(a_P(0)\) and \(a_P(t)\) in (2), that is in the triple-pomeron diagram in Fig. 2, are not the same. The lower pomeron \(a_P(0)\) in Fig. 2 is the usual ‘soft’ pomeron; whereas the upper ones, with \(a_P(t)\), include DGLAP evolution from a low initial scale \(\mu = \mu_0\) up to a rather large scale \(\mu \sim M_Y/\psi\) at the \(J/\psi\) production vertex. The summation of the double logarithms \((a_P(0) \ln(1/x) \ln(\mu^2/\mu_0^2))^\delta\) leads to a steeper \(x\)-dependence and hence to a larger effective intercept for the trajectory \(a_P(t)\) of the upper ‘hard’ pomeron. Thus in (5) and (6) we have \(\delta/4 > \epsilon\), where \(\delta = 4(\alpha_P^{\text{hard}}(0) - 1)\) and \(\epsilon = \alpha_P^{\text{soft}}(0) - 1\).

\(^4\) We thank Alessia Bruni for informing us about the interval of \(M_Y\).
in (2). Then
\[
\int dM_{Y}^{2} \frac{d\sigma}{dt} \frac{d\sigma_{d}}{dt} = \frac{1}{16\pi^{2}} g_{3P}^{\text{bare}}(0) \cdot J \int \frac{1}{16\pi} g_{N}(0), \tag{9}
\]
where we estimate that the integration over \( M_{Y} \) gives the factor \( J \simeq 0.6 \). In this way we find
\[
g_{3P}^{\text{bare}}(0)/g_{N}(0) \simeq 1/3. \tag{10}
\]
This has to be compared with the ratio extracted [2] many years ago from hadron–hadron initiated data
\[
g_{3P}^{\text{effective}}(0)/g_{N}(0) \simeq 1/10. \tag{11}
\]
The distortion of the coupling in the old estimate, due to the survival factors, is clearly evident.\(^5\) The fact that including absorptive effects gives a larger value of the triple-pomeron coupling, is not new. Larger values of \( g_{3P} \) were obtained, for example, in [14,27]. However all the previous estimates are model dependent. We cannot sum up all the reggeon diagrams which describe the absorptive effects. Therefore the value of \( g_{3P} \) depends on the class of multi-pomeron diagrams chosen for resummation, on the assumptions about the behaviour of the multi-pomeron vertices, and sometimes on the threshold factor in the multi-pomeron vertices, etc. The advantage of the present evaluation of \( g_{3P} \) is that we study a process where the absorptive corrections are small, and so affect the final result much less.

Of course, in the integration over \( M_{Y} \), we cannot neglect the possible contributions of secondary reggeons in (2), such as the RRP term near the upper limit of the integral and the PPR term for \( M_{Y} \) close to \( M_{Y} \) (min). The value of \( g_{3P}^{\text{bare}}(0) \) is therefore expected to be a bit smaller than in (10). It is thus crucially important to have \( \gamma + p \rightarrow J/\psi + Y \) with an explicit measurement of the \( M_{Y} \) spectrum in order to perform a full triple-Regge analysis in which we quantify the different triple-Regge contributions.

The discussion in this Letter has implications for all diffractive processes. We mention two topical examples, both of which illustrate the importance of the energy dependence of the survival factor \( S^{2} \). In the existing Monte Carlos for high energy multiparticle production, the distributions of leading particles are parametrized in the triple-Regge form. In comparison with the limiting fragmentation or Feynman scaling hypothesis \([5,28]\), where the normalized cross section does not depend on energy, we now have
\[
\frac{1}{\sigma_{\text{tot}}} \frac{d\sigma^{\text{SD}}}{dt} \frac{d\sigma_{\ell}}{dt} = g_{2}^{\gamma}(t) g_{3P}(t) \left(1 - x_{L}\right)^{\alpha_{2\ell}(0) - 2\ell a_{2}(t)} S_{\text{SD}}^{2}(s,t), \tag{12}
\]
which depends, not only on \( x_{L} \), but on the energy.\(^6\) The energy dependence comes entirely from the survival factor \( S_{\text{SD}}^{2} \). So,

\[\text{Fig. 3. The energy dependence of the survival factor, } S_{\text{SD}}^{2}, \text{ for single (soft and hard) diffractive dissociation, via the triple-pomeron interaction. The GZK energy, } 2.3 \times 10^{20} \text{ eV, is the theoretical upper (Greisen–Zatsepin–Kuzmin [29]) limit on the energy of cosmic rays which arises from the large cross section for the interaction of the protons with relic photons due to } \Delta \text{ isobar production.}\]

\[\text{before we discuss the two examples, we compute the energy dependence of } S_{\text{SD}}^{2}. \text{ As before, we use SD to denote single-particle diffractive dissociation.}\]

In Fig. 3 we show the expected energy dependence of the survival factor \( S_{\text{SD}}^{2} \) for the cross section integrated over \( t \). This was calculated in a simplified two-channel eikonal model\(^7\) with the \( t \) dependence of the elastic pomeron–proton vertex, \( g_{N}(t) \), given by the proton electromagnetic form factor \( F_{1} \), that is \( g_{N}(t) = g_{N}(0) F_{1}(t) \). The parameters of the pomeron trajectory were tuned to describe the CERN-ISR, CERN-SPS and Tevatron elastic data. It is clear from the figure that the energy dependence of \( S_{\text{SD}}^{2} \) is quite pronounced. Moreover, we find that the predictions for \( S_{\text{SD}}^{2} \) obtained using a two-channel eikonal are stable\(^8\) to changing the details of the model, after the ‘soft’ parametrization is tuned to fit the elastic data.

The continuous curve in Fig. 3 is for soft diffractive dissociation. Here the process takes place at relatively large impact parameters \( b_{v} \), where the absorption is not so strong. Suppose, instead, we were to consider hard diffractive dissociation, for example, where a pair of high \( E_{T} \) jets [6] or a heavy quark pair or \( W \) or \( Z \) boson or some other heavy object is produced within the system \( Y \). In these cases the interaction takes place at much smaller \( b_{v} \). Here, in the central \( b_{v} \) region, the absorption is stronger and the corresponding \( S_{\text{SD}}^{2} \) factor is smaller, espe-

\[\text{\(^{5}\) Interestingly, a similar distortion had been already allowed for in the global analysis of ‘soft’ hadronic data that was performed in [1].}\]

\[\text{\(^{6}\) Note, however, that as it stands, formula (12) does not allow for the effects of migration which, as we have seen in the case of the leading neutrons \([7]\), give a noticeable contribution for } x_{L} < 0.7–0.8. \text{ Therefore (12) is only valid in its present form for } x_{L} \geq 0.8. \text{ Also (12) contains just the PPP-contribution. For } x_{L} < 0.8–0.9 \text{ allowance should be made for a RRP contribution with its own } S_{\text{RRP}}^{2}.\]

\[\text{\(^{7}\) The details of the two-channel eikonal model will be presented elsewhere.}\]

\[\text{\(^{8}\) If we try to mimic the two-channel eikonal by an ‘enhanced’ one-channel eikonal, then the predictions for } S_{\text{SD}}^{2} \text{ decrease and the values at ultra high energies become too low and unstable.}\]
cially at ultra high energies where the proton opacity becomes very close to the black disc limit. Therefore, for completeness, we also show the predictions for the energy dependence of $S_{SD}^2$ for hard diffractive dissociation, by the dashed curve in Fig. 3.

Our first topical example, where the energy dependence of $S_{SD}^2$ plays an important role, is the evaluation of the so-called pile-up backgrounds to forward physics studies at the LHC, see, for instance, [30]. At high LHC luminosities, the soft pile-up single (and even double) diffractive processes can fake the signal from hard diffractive events due to overlap with the non-diffractive hard processes observed in the central detector. However, when we allow for the energy dependence of $S_{SD}^2$ (for soft diffraction), we decrease the predicted leading proton spectra at the LHC energies, as compared to the leading proton spectrum measured at HERA [31]. A preliminary estimate gives a reduction by about a factor of 3. Therefore the impact of pile-up is to be suppressed by the energy dependence of $S_{SD}^2$.

The second example illustrates the importance of the energy dependence of $S_{SD}^2$ to the analysis of ultra high energy cosmic ray data. The observed structure of the extensive cosmic ray air showers depends sensitively on the spectrum of the leading hadrons. From the continuous curve in Fig. 3 we see that in going from HERA energies to the GZK cut-off energy [29], the survival probability $S_{SD}^2$ decreases by almost an order of magnitude. Of course, the small values of $S_{SD}^2$ at ultra high energies do not mean that the leading hadron disappears. Instead, after the rescattering, the large $x_L$ hadron migrates to a lower $x_L$ and larger $p_t$ domain [7], thereby changing the shape of the extensive air shower.

Acknowledgements

We thank Michele Arneodo, Alessia Bruni and Risto Orava for valuable discussions. M.G.R. and V.A.K. thank, respectively, the IPPP at the University of Durham and the Aspen Center for Physics for hospitality. This work was supported by the Royal Society, by INTAS grant 05-103-7515, by grant RFBR 04-02-16073 and by the Federal Program of the Russian Ministry of Industry, Science and Technology SS-1124.2003.2.

References

K.A. Ter-Martirosyan, Nucl. Phys. B 68 (1964) 591;
[16] V.A. Abramovsky, Pis'ma JETP (1976) 228;

9 The reduction by about 3 is only an illustrative factor, as it is based on $S^2 \sim 0.4$ found [8] for leading neutrons at HERA.
Effective restoration of chiral symmetry in excited mesons

R.F. Wagenbrunn, L.Ya. Glozman *

Institute for Physics, Theoretical Physics Branch, University of Graz, Universitätsplatz 5, A-8010 Graz, Austria

Received 8 August 2006; received in revised form 17 October 2006; accepted 18 October 2006

Available online 7 November 2006

Editor: J.-P. Blaizot

Abstract

A fast restoration of chiral symmetry in excited mesons is demonstrated. A minimal “realistic” chirally symmetric confining model is used, where the only interaction between quarks is the linear instantaneous Lorentz-vector confining potential. Chiral symmetry breaking is generated via the nonperturbative resummation of valence quarks self-energy loops and the meson bound states are obtained from the Bethe–Salpeter equation. The excited mesons fall into approximate chiral multiplets and lie on the approximately linear radial and angular Regge trajectories, though a significant deviation from the linearity of the angular trajectory is observed.

© 2006 Elsevier B.V. All rights reserved.

PACS: 11.30.Rd; 12.38.Aw; 14.40.-n

There are certain phenomenological evidences that in highly excited hadrons, both in baryons [1–3] and mesons [4,5] chiral and $U(1)_A$ symmetries are approximately restored, for a short overview see [6]. This “effective” restoration of chiral and $U(1)_A$ symmetries should not be confused with the chiral symmetry restoration at high temperatures and/or densities. What actually happens is that the excited hadrons gradually decouple from the quark condensates. Fundamentally it happens because in the high-lying hadrons the semiclassical regime is manifest and semiclassically quantum fluctuations of the quark fields are suppressed relative to the classical contributions which preserve both chiral and $U(1)_A$ symmetries [6,7]. The microscopical reason is that in high-lying hadrons a typical momentum of valence quarks is large and hence they decouple from the quark condensate and consequently their Lorentz-scalar dynamical mass asymptotically vanishes [1,3,7,10,14]. Restoration of chiral symmetry requires a decoupling of states from the Goldstone bosons [3,8–10] which is indeed observed phenomenologically since the coupling constant for $h^* \rightarrow h + \pi$ decreases fast higher in the spectrum.

At the moment there are two main paths to understand this phenomenon. In the first one tries to connect the high-lying states to the short-range part of the two-point correlation function where the operator product expansion is valid [2,11]. However, the OPE is an asymptotic expansion. Then, while the correct spectrum of QCD must be consistent with the OPE, there is an infinite amount of incorrect spectra that can also be in agreement with the OPE. Hence the results within the present approach crucially depend on additional assumptions [12,13].

In the second approach the authors try to understand this phenomenon within the microscopical models [7,8,10,14]. There are also interesting attempts to formulate the problem on the lattice [16,17], though extraction of the high-lying states on the lattice is a task of future.

In the absence of the controllable analytic solutions to QCD an insight into phenomenon can be achieved only through models. Clearly the model must be field-theoretical (in order to be able to exhibit the spontaneous breaking of chiral symmetry), chirally symmetric and contain confinement. In principle any possible gluonic interaction can contribute to chiral symmetry breaking and it is not known which specific interaction is the most important one in this respect. However, at the first stage it is reasonable to restrict oneselfs to the simplest possible model that contains all three key elements. Such a model is known, it is a generalized Nambu and Jona-Lasinio model (GNJL) with the instantaneous Lorentz-vector confining kernel [18–20]. This model is similar in spirit to the large $N_c$ ’t Hooft model (QCD in $1 + 1$ dimensions) [21]. In both models the
only interaction between quarks is the instantaneous infinitely raising Lorentz-vector linear potential. Then chiral symmetry breaking is described by the standard summation of the valence quarks self-interaction loops in the rainbow approximation (the Schwinger–Dyson or gap equations), while mesons are obtained from the Bethe–Salpeter equation for the quark–antiquark bound states, see Fig. 1.

Conceptually the underlying physics is very clear in the ‘t Hooft model in the sense that once the proper gauge is chosen, the linear Lorentz-vector confining potential appears automatically as the Coulomb interaction in 1+1 dimensions. In 3+1 dimensions, once the Coulomb gauge is used for the gluonic field [22], an almost linearly raising confinement potential has been obtained [23–25].

An obvious advantage of the GNJL model is that it can be applied in 3+1 dimensions to systems of arbitrary spin. In 1+1 dimensions there is no spin, the rotational motion of quarks is impossible, and the states are characterized by the only quantum number, which is the radial quantum number. Then it is known that the spectrum represents an alternating sequence of positive and negative parity states and chiral multiplets never emerge. This happens because in 1+1 dimensions the valence quarks can perform only an oscillatory motion. In 3+1 dimension, on the contrary, the quarks can rotate and hence can always be ultrarelativistic and chiral multiplets should emerge naturally [3].

Restoration of chiral symmetry in excited heavy-light mesons has been studied with the quadratic confining potential [14] and it was also mentioned in a model with the instantaneous potential of a more complicated form [15]. Here we report our results for excited light-light mesons with the linear confining potential, $V_0(r) = \sigma r$.

The Schwinger–Dyson equation for the self-energy operator

$$\Sigma(p) = [A_p - m] + (\gamma \cdot \vec{p})B_p - p,$$

where, due to the instantaneous nature of the interaction the time-component of the Dirac operator is not dressed. The Lorentz-scalar dynamical mass $A_p$ as well as the Lorentz-vector spatial part $B_p$ contain both the classical and quantum contributions, the latter coming from loops [7]:

$$A_p = m + \frac{1}{2} \int \frac{d^3k}{(2\pi)^3} V(\vec{p} - \vec{k}) \sin \varphi_k,$$

$$B_p = p + \frac{1}{2} \int \frac{d^3k}{(2\pi)^3} (\vec{p} \cdot \vec{k}) V(\vec{p} - \vec{k}) \cos \varphi_k,$$

where $\tan \varphi_p = \frac{A_p}{B_p}$.

Solution of the Schwinger–Dyson equation (3) with the linear potential is well known, see e.g. [27], and the mass-gap equation has a nontrivial solution which breaks chiral symmetry, by generating a nontrivial dynamical mass function $A_p$. This dynamical mass is a very fast decreasing function at larger momenta. Then the quark condensate is given as

$$\langle \bar{q} q \rangle = -\frac{N_C}{\pi^2} \int_0^\infty dp p^2 \sin \varphi_p.$$

The homogeneous Bethe–Salpeter equation for the quark–antiquark bound state with mass $M$ in the rest frame, i.e. with the four momentum $P^\mu = (M, \vec{P} = 0)$, is

$$\chi(\vec{p}, M) = -i \int \frac{d^3k}{(2\pi)^3} V(\vec{p} - \vec{k}) \gamma_0 S(k_0 + M/2, \vec{k}) \chi(\vec{k}, M) S(k_0 - M/2, \vec{k}) \gamma_0,$$

where $\chi(\vec{p}, M)$ is the mesonic Bethe-Salpeter amplitude in the rest frame. Eq. (9) is written in the ladder approximation for the vertex which is consistent with the rainbow approximation for the quark mass operator and which is well justified in the large-$N_C$ limit.

The Bethe–Salpeter amplitude can be decomposed into two components for mesons with $J^{PC} = (2n)^{+-}, (2n+1)^{+-}, (2n+1)^{++},$
(2n + 2)−−, or 0++ and into four components for mesons with $J^{PC} = (2n + 1)^{−−}$ or $(2n + 2)++$, respectively. Here J is the spin, P the parity and C the charge conjugation parity of the meson and $n \in \mathbb{N}_0$. In that way the Bethe–Salpeter equation becomes a system of coupled integral equations for the components which we solve by expanding them into a finite number of properly chosen basis functions. This leads to a matrix eigenvalue problem which can be solved by standard linear algebra methods. We vary the meson mass until one of the eigenstates is equal to one.

In the gap as well as in the Bethe–Salpeter equations the infrared divergences are removed by introducing a finite “mass” into a confining potential, which is a standard trick. Then in the infrared limit (“mass” goes to 0) the quark propagator consists of a finite and diverged parts, while the mesons masses are finite. Recently it was demonstrated that also the masses of quark–quark subsystems in the color-antitriplet state go to infinity in this limit and hence are removed from the physical spectrum [28]. The results presented in the following were obtained by calculating the infrared limit numerically, i.e. the quoted meson masses were extrapolated to the infrared limit from a few points with a very small but finite mass of the infrared regulator. It turned out that in this region of the small mass of the infrared regulator the squares of the meson masses depend almost linearly on the mass of the infrared regulator making the extrapolation reliable. The presented results are accurate within the quoted digits at least for states with small J and for states with higher J but small n. At larger J for larger n numerical errors accumulate in the second digit after comma.

By definition an effective chiral symmetry restoration means that (i) the states fall into approximate multiplets of $SU(2)_L \times SU(2)_R$ and the splittings within the multiplets ($\Delta M = M_+ - M_-$) vanish at $n \to \infty$ and/or $J \to \infty$; (ii) the splitting within the multiplet is much smaller than between the two subsequent multiplets [4–6].

The condition (i) is very restrictive, because the structure of the chiral multiplets for the $J = 0$ and $J > 0$ mesons is very different [4,5]. For the $J > 0$ mesons chiral symmetry requires a doubling of states with some quantum numbers in contrast to the $J = 0$ states. Given the complete set of standard quantum numbers $I, J^{PC}$, the multiplets of $SU(2)_L \times SU(2)_R$ are

$$J = 0$$

$$\begin{align*}
(1/2, 1/2)_a: & \quad 1, 0^{−−} \leftrightarrow 0, 0^{++}, \\
(1/2, 1/2)_b: & \quad 1, 0^{++} \leftrightarrow 0, 0^{−−}.
\end{align*}$$

(10)

$$J = 2k, \quad k = 1, 2, \ldots$$

$$\begin{align*}
(0, 0): & \quad 0, J^{−−} \leftrightarrow 0, J^{++}, \\
(1/2, 1/2)_a: & \quad 1, J^{++} \leftrightarrow 0, J^{++}, \\
(1/2, 1/2)_b: & \quad 1, J^{−−} \leftrightarrow 0, J^{++}, \\
(0, 1) \oplus (1, 0): & \quad 1, J^{++} \leftrightarrow 1, J^{−−}.
\end{align*}$$

(11)

$$J = 2k − 1, \quad k = 1, 2, \ldots$$

$$\begin{align*}
(0, 0): & \quad 0, J^{++} \leftrightarrow 0, J^{−−}, \\
(1/2, 1/2)_a: & \quad 1, J^{++} \leftrightarrow 0, J^{−−}, \\
(1/2, 1/2)_b: & \quad 1, J^{−−} \leftrightarrow 0, J^{++}, \\
(0, 1) \oplus (1, 0): & \quad 1, J^{−−} \leftrightarrow 1, J^{++}.
\end{align*}$$

(12)

Note that within the present model the axial anomaly is absent. Even so there are no exact $U(1)_A$ multiplets, because this symmetry is broken not only by the anomaly, but also by the chiral condensate of the vacuum. Then the mechanism of the $U(1)_A$ symmetry breaking and restoration is exactly the same as of $SU(2)_L \times SU(2)_R$. Hence the effective restoration of $SU(2)_L \times SU(2)_R$ would automatically imply restoration of $U(1)_A$ and of $U(2)_L \times U(2)_R$ and vice versa. An effective restoration of the $U(1)_A$ symmetry would mean an approximate degeneracy of the opposite spatial parity states with the same isospin from the distinct $(1/2, 1/2)_a$ and $(1/2, 1/2)_b$ multiplets of $SU(2)_L \times SU(2)_R$.

Note that within the present model there are no vacuum fermion loops. Then since the interaction between quarks is flavor-blind the states with the same $J^{PC}$ but different isospins from the distinct multiplets $(1/2, 1/2)_a$ and $(1/2, 1/2)_b$ as well as the states with the same $J^{PC}$ but different isospins from $(0, 0)$ and $(0, 1) \oplus (1, 0)$ representations are exactly degenerate. Hence it is enough to complete a set of the isovector (or isoscalar) states.

In Table 1 we present our results for the spectrum for the two-flavor ($u$ and $d$) mesons in the chiral limit. Clearly the model should not be taken seriously for the low-lying states where other gluonic interactions as well as the $1/N_c$ corrections should be important. The purpose of the study is, however, to demonstrate that a solvable field-theoretical model

<table>
<thead>
<tr>
<th>Chiral multiplet</th>
<th>$J^{PC}$</th>
<th>Radial excitation $n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(0, 1) \oplus (1, 0)$</td>
<td>$1^{++}$</td>
<td>0.00 2.93 4.35 5.49 6.46 7.31 8.09</td>
</tr>
<tr>
<td>$(1/2, 1/2)_a$</td>
<td>$0^{++}$</td>
<td>1.49 3.38 4.72 5.80 6.74 7.57 8.33</td>
</tr>
<tr>
<td>$(1/2, 1/2)_b$</td>
<td>$1^{−−}$</td>
<td>2.68 4.03 5.15 6.14 7.01 7.80 8.53</td>
</tr>
<tr>
<td>$(0, 1) \oplus (1, 0)$</td>
<td>$1^{++}$</td>
<td>1.55 3.28 4.56 5.64 6.57 7.40 8.16</td>
</tr>
<tr>
<td>$(0, 1) \oplus (1, 0)$</td>
<td>$1^{−−}$</td>
<td>2.20 3.73 4.95 5.98 6.88 7.69 8.43</td>
</tr>
</tbody>
</table>

Table 1
Masses of isovector mesons in units of $\sqrt{\pi}$
in $3 + 1$ dimensions does exhibit the effective restoration of the chiral symmetry at large radial excitations $n$ and large spins. The excited mesons fall into approximate chiral and $U(1)_A$ multiplets and all conditions of the effective symmetry restorations are satisfied. We observe a very fast restoration of both $SU(2)_L \times SU(2)_R$ and $U(1)_A$ symmetries with increasing $J$ and essentially more slow restoration with increasing of $n$.

When the chiral symmetry breaking Lorentz-scalar dynamical mass of quarks is zero, then there are independent Bethe–Salpeter amplitudes just according to the chiral representations (10)–(12). A finite dynamical mass plays a role of the off-diagonal matrix element and mixes the otherwise independent chiral Bethe–Salpeter amplitudes for the states $1^-, 2^{++}, 3^{--}, \ldots$. A key feature of this dynamical mass is that it is strongly momentum-dependent and vanishes very fast once the momentum is increased. When one increases excitation energy of a hadron, one also increases a typical momentum of valence quarks. Consequently, the chiral symmetry violating dynamical mass of quarks becomes small. Hence the mixing of the independent chiral Bethe–Salpeter amplitudes becomes small. A given state in the table is then assigned to the chiral representation according to the chiral Bethe–Salpeter amplitude that dominates in the given state.

In Fig. 2 the rates of the symmetry restoration against the radial quantum number $n$ and spin $J$ are shown. It is seen that with the fixed $J$ the splitting within the multiplets $\Delta M$ decreases asymptotically as $1/\sqrt{n}$, dictated by the asymptotic linearity of the radial Regge trajectories. This property is consistent with the dominance of the free quark loop logarithm at short distances.

In Fig. 3 the angular and radial Regge trajectories are shown. Both kinds of trajectories exhibit deviations from the linear behavior. This fact is obviously related to the chiral symmetry breaking effects for lower mesons. Note, that the chiral symmetry requires a doubling of some of the radial and angular Regge trajectories for $J = 1, 2, \ldots$. This is a highly nontrivial prediction of chiral symmetry. For example, some of the rho-mesons lie on the trajectory that is characterized by the chiral index $(0, 1) + (1, 0)$, while the other fit the trajectory with the chiral index $(1/2, 1/2)_b$. The intercepts of the asymptotic angular Regge trajectories for mesons in the given and different chiral representations coincide. Hence the asymptotic rate of the symmetry restoration with $J$ is faster than $1/\sqrt{J}$.

The numerical result for the quark condensate is $\langle \bar{q}q \rangle = (-0.231 \sqrt{\sigma})^3$, which agrees with the previous studies within the same model. If we fix the string tension from the phenomenological angular Regge trajectories, then $\sqrt{\sigma} \approx 300\text{–}400$ MeV and hence the quark condensate is between $(-70 \text{ MeV})^3$ and $(-90 \text{ MeV})^3$ which obviously underestimates the phenomenological value. Probably this indicates that other gluonic interactions could also contribute to chiral symmetry breaking. Notice that the string tension in Coulomb gauge can be larger than the asymptotic one. Lattice results suggest a value about twice the value obtained here [29,30]. This would increase the value for the condensate but on the other hand lead to an unrealistically small pion decay constant [31,32].
Fig. 3. Angular (top) and radial (bottom) Regge trajectories for isovector mesons with $M^2$ in units of $\sigma$. Mesons of the chiral multiplet $(1/2, 1/2)_a$ are indicated by circles, of $(1/2, 1/2)_b$ by triangles, and of $(0, 1) \oplus (1, 0)$ by squares ($J^{++}$ and $J^{--}$ for even and odd $J$, respectively) and diamonds ($J^{--}$ and $J^{++}$ for even and odd $J$, respectively).

In the limit $n \to \infty$ and/or $J \to \infty$ one observes a complete degeneracy of all multiplets, which means that the states fall into

$$\left[(0, 1/2) \oplus (1/2, 0)\right] \times \left[(0, 1/2) \oplus (1/2, 0)\right]$$

representation that combines all possible chiral representations for the systems of two massless quarks [5]. This means that in this limit the loop effects disappear completely and the system becomes classical [6,7].

Acknowledgements

R.F.W. acknowledges helpful discussions with R. Alkofer, M. Kloker, and A. Krassnigg. This work was supported by:

Deutsche Forschungsgemeinschaft (project Al 279/5-1) and the Austrian Science Fund (projects P16823-N08 and P19168-N16).

References

Duality and the Transition to pQCD, World Scientific, Singapore, 2006,  
1233;  
137.  
Pairing in fermion systems with unequal masses: Nonperturbative renormalisation group approach

Boris Krippa

School of Physics and Astronomy, University of Manchester, Manchester M13 9PL, UK

Received 29 May 2006; received in revised form 21 September 2006; accepted 3 October 2006

Abstract

The application of the nonperturbative renormalisation group approach to a system with two fermion species is studied. Assuming a simple ansatz for the effective action with effective bosons, describing pairing effects we derive a set of approximate flow equations for the effective coupling including boson and fermionic fluctuations. The case of two fermions with different masses but coinciding Fermi surfaces is considered. The phase transition to a phase with broken symmetry is found at a critical value of the running scale. The large mass difference is found to disfavour the formation of pairs. The mean-field results are recovered if the effects of boson loops are omitted. While the boson fluctuation effects were found to be negligible for large values of $p_F a$ they become increasingly important with decreasing $p_F a$ thus making the mean field description less accurate.

© 2006 Published by Elsevier B.V.

Keywords: Nonperturbative renormalisation group; EFT; Broken phase; Superfluidity

The properties of asymmetric many fermion systems have recently attracted much attention (see, for example Ref. [1] and references therein) driven by the substantial advance in experimental studies of trapped fermionic atoms. This asymmetry can be provided by unequal masses, different densities and/or chemical potentials. Understanding the pairing mechanism in such settings would be of immense value for different many fermion systems from atomic physics to strongly interacting quark matter. The important theoretical issue to be resolved here is the nature of the ground state. Several competing states have been proposed so far. These include: LOFF [2] phase, breached-pair (BP) superfluidity [3] (or Sarma phase) and mixed phase [4]. Establishing the true ground state is still an open question. It was shown, for example, that LOFF and mixed phases are more stable then the Sarma phase in the systems of fermions with the mismatched Fermi surfaces and with both equal and different masses [1,4,5]. All these studies, however, have been performed within the mean field approximation (MFA). In spite of the fact that in many cases MFA is quite reliable it is important to understand better the limits of applicability of MFA in the context of the fermion systems with a certain type of asymmetry (masses and/or densities) and work out the physical regimes where the MFA is too crude or even inadequate. The convenient way to estimate the corrections to MFA is provided by the nonperturbative renormalisation group (NRG) approach [6] which was successfully applied to the standard pairing problem with one type of fermions [7–10]. The main element of NRG is the effective average action $\Gamma_k$ which is a generalisation of the standard effective action $\Gamma$, the generating functional of the 1PI Green functions. The only difference between them is that $\Gamma_k$ includes only quantum fluctuations with momenta larger then the infrared scale $k$. The evolution of the system as the function of the scale $k$ is described by the nonperturbative flow equations. When $k \to 0$ all fluctuations are included and full effective action is recovered. Similarly, at starting scale $k = K$ no fluctuations are included so $\Gamma_{k=K}$ can be associated with the classical action $S$ therefore $\Gamma_k$ provides an interpolation between the classical and full quantum effective actions.
The dependence of $I_2$ from the infrared scale $k$ is given by the nonperturbative renormalisation group equation (NRGE)
\[ \partial_k I_2 = - \frac{i}{2} \text{Tr} \left[ (\partial_k R) (I^{(2)}_B - R)^{-1} \right]. \tag{1} \]
Here $I^{(2)}_B$ is the second functional derivative of the effective action taken with respect to all types of field included in the action and $R(q, k)$ is a regulator which should suppress the contributions of states with momenta less than or of the order of running scale $k$. To recover the full effective action we require $R(q, k)$ to vanish as $k \to 0$ whereas for $q \ll k$ the regulator behaves as $R(q, k) \simeq k^2$. The above written equation is, in general, the functional equation. For a practical applications it needs to be converted to the system of partial or ordinary differential equations so that approximations and truncations are required.

We consider a nonrelativistic many-body system at zero temperature with two types of the fermion species $a$ and $b$ interacting through a short-range attractive interaction and introduce a boson field $\phi$ describing the pair of interacting fermions. The ansatz for $I$ takes the form
\[ I[\psi, \bar{\psi}, \phi, \bar{\phi}, \mu, k] = \int d^4x \left[ \left( \frac{\mu_a + \mu_b}{2m} \right) \phi(x) - U(\phi, \bar{\phi}) + \frac{1}{2} \sum_{i=a}^b \left( \frac{Z_{M,i} \phi_i}{M_i} \right)^2 \right] \]
\[ - Z_m \left( \frac{1}{2} \left( \frac{\mu_a + \mu_b}{2m} \right) \phi \right) \]
\[ \times \left[ \text{tr} \left( \partial_k R \right) \right]. \tag{2} \]
Here $M_i$ is the mass of the fermion in vacuum and the factor $1/2m$ with $m = M_a + M_b$ in the boson kinetic term is chosen simply to make $Z_m$ dimensionless. The coupling $Z_m$, the wave-function renormalisations factors $Z_{\phi, \bar{\phi}}$ and the kinetic-mass renormalisations factors $Z_m, M_i$ all run with $k$, the scale of the regulator. Having in mind the future applications to the crossover from BCS to BEC (where chemical potential becomes negative) we also let the chemical potentials $\mu_a$ and $\mu_b$ run, thus keeping the corresponding densities (and Fermi momenta $p_{F,i}$) constant. The bosons are, in principle, coupled to the chemical potentials via a quadratic term in $\phi$, but this can be absorbed into the potential by defining $U = U - (\mu_1 + \mu_2) Z_{\phi} \phi^2 \bar{\phi}$. The evolution equations include running of chemical potentials, effective potential and all couplings ($Z_{\phi, \bar{\phi}}, Z_m, Z_{M,i}, Z_{\phi_i, \phi}$). However, in this Letter we allow to run only $Z_{\phi}$, parameters in the effective potential ($u$’s and $\rho_0$) and chemical potentials since this is the minimal set needed to include the effective boson dynamics.

We expand the effective potential about its minimum, $\phi^\dagger \phi = \rho_0$, so that the coefficients $u_i$ are defined at $\rho = \rho_0$,
\[ \bar{U}(\rho) = u_0 + u_1 (\rho - \rho_0) + \frac{1}{2} u_2 (\rho - \rho_0)^2 + \frac{1}{6} u_3 (\rho - \rho_0)^3 + \cdots, \tag{3} \]
where we have introduced $\rho = \phi^\dagger \phi$. Similar expansion can be written for the renormalisation factors. The coefficients of the expansion run with the scale. The phase of the system is determined by the coefficient $u_1$. We start evolution at high scale where the system is in the symmetric phase so that $u_1 > 0$. When the running scale becomes comparable with the pairing scale (close to average Fermi momentum) the system undergoes the phase transition to the phase with broken symmetry, energy gap etc. The point of the transition corresponds to the scale $u_1 = 0$. The bosonic excitations in the gapped phase are gapless Goldstone bosons. Note, that in this phase the minimum of the potential will also run with the scale $k$ so that the value $\rho_0(k \to 0)$ determines the physical gap.

The evolution equation takes the following general form
\[ \partial_k I_2 - \frac{i}{2} \text{Tr} \left[ (\partial_k R)_B (I^{(2)}_B - R)_B \right] \tag{4} \]
where $I^{(2)}_B$ is the matrix of the second functional derivatives of the effective action taken with respect to boson (fermion) fields included in the action and $R_B$ ($R_F$) is the boson (fermion) regulator which should suppress the contributions of states with momenta less than or of the order of running scale $k$. The boson regulator has the structure
\[ R_B = R_B \text{diag}(1, 1). \tag{5} \]

The fermion regulator for both types of fermions has the structure
\[ R_{F,i} = \text{sgn}(\epsilon_i(q) - \mu_i) R_{F,i}(q, \mu_i, 1) \text{diag}(1, -1). \tag{6} \]
Note that this regulator is positive for particle states above the Fermi surface and negative for the hole states below the Fermi surface.

Calculating the second functional derivatives, taking the matrix trace and carrying out the pole integration in the loop integrals we get the evolution equation for $U$ at constant chemical potentials
\[ \partial_k U = \frac{1}{V_4} \partial_k I_2 \]
\[ = \frac{1}{2} \int \frac{d^3q}{(2\pi)^3} \frac{E_{FS}}{\sqrt{E_{FS}^2 + \Delta^2}} \times \left[ \text{sgn}(q - p_{u,a}) \partial_k R_{F,a} + \text{sgn}(q - p_{u,b}) \partial_k R_{F,b} \right] + \frac{1}{2Z_{\phi}} \int \frac{d^3q}{(2\pi)^3} \frac{E_B}{\sqrt{E_B^2 - V_B^2}} \partial_k R_B. \tag{7} \]
Here
\[ E_S = (E_{F,a} + E_{F,b})/2, \]
\[ E_A = (E_{F,a} - E_{F,b})/2, \tag{8} \]
and
\[ E_B(q, k) = \frac{Z_m}{2m} q^2 + u_1 + u_2 (2\phi^\dagger \phi - \rho_0) + R_B(q, k), \]
\[ V_B = u_2 \phi^\dagger \phi. \tag{9} \]
\[ E_{F,i}(q, p_{\mu,i}, k) = \frac{1}{2M_{i}}q^{2} - \mu_{i} + R_{F}(q, k) \sgn(q - p_{\mu,i}), \]
\[ \Delta^{2} = g^{2}\phi^{4} \phi \] (10)
and we have introduced \( p_{\mu,i} = \sqrt{2M_{i}\mu_{i}} \), the Fermi momentum corresponding to the (running) value of \( \mu_{i} \). It is worth mentioning that poles in the fermion propagator occur at
\[ \bar{q}_{0}^{1,2} = -\Delta_{A} \pm \sqrt{E_{S}(q, k)^{2} + \Delta^{2}}. \] (11)
At \( k = 0 \) (\( R_{F} = 0 \)) in the condensed phase, these become exactly the dispersion relations obtained in [3] where the possibility of having the gapless excitations has been discussed. The ordinary BCS spectrum can easily be recovered when the asymmetry of the system is vanishing (\( E_{A} \rightarrow 0 \)). The first term in the evolution equation for the effective potential describes the evolution of the system related to the fermionic degrees of freedom whereas the second one takes into account the bosonic contribution. The mean field results can be recovered if the second term is omitted. In this case the equation for the effective potential can be integrated analytically.
\[ \bar{U}(\rho, \mu, K) = \bar{U}(\rho, \mu, K) \] (12)
At starting scale \( K \) the potential has the form
\[ \bar{U}(\rho, \mu, K) = u_{0}(K) + u_{1}(K)\rho. \] (13)
The renormalised value of \( u_{1}(K) \) can be related to the scattering length.
\[ \frac{u_{1}(p_{F}, K)}{g^{2}} = -\frac{M}{2\pi a} + \frac{1}{2} \int \frac{d^{3}q}{(2\pi)^{3}} \left[ \frac{1}{E_{S}(q, 0, 0, 0)} - \frac{1}{E_{S}(q, \mu_{a}, \mu_{b}, K)} \right]. \] (14)
Here \( M \) is the reduced mass and the dependence of \( E_{S} \) on the chemical potentials has been made explicit.
Differentiating the effective potential with respect to \( \rho \), setting the derivative to zero and taking the limit \( K \rightarrow \infty \), we arrive at the equation
\[ -\frac{M}{2\pi a} + \frac{1}{2} \int \frac{d^{3}q}{(2\pi)^{3}} \left[ \frac{1}{E_{S}(q, 0, 0, 0)} - \frac{1}{\sqrt{E_{S}(q, \mu_{a}, \mu_{b}, k)^{2} + \Delta^{2}}} \right] = 0. \] (15)
Taking the physical limit \( (k = 0) \) we obtain the gap equation identical to that derived in the mean field approximation [1,4].
We now turn to the full set of the evolution equations which includes the effects of the bosonic fluctuations. In this Letter we consider the case of two fermion species with the different masses and the same Fermi momenta. It implies that the chemical potentials are different. In this situation the Sarma phase does not exist and the system experiences the BCS pairing depending however on the mass asymmetry. The general case of the mismatched Fermi surfaces will be discussed in the subsequent publication.

The derivation of the evolution equations was discussed in details in Ref. [7] so that here we just mention the main points. Within the above described approximation (fixed couplings \( Z_{m}, Z_{M,i}, Z_{\Phi,i}, Z_{g} \)) all of these can be obtained from the evolution of the effective potential, for example
\[ \partial_{k}Z_{\phi} = -\frac{1}{2} \frac{\partial^{2}}{\partial \mu \partial \rho} \left( \partial_{k}Z_{\phi} \right)_{\rho = \rho_{0}}, \] (16)
where \( \mu = \mu_{a} + \mu_{b} \). The alternative way of getting the ERG flow equation for \( Z_{\phi} \) is to consider a time-dependent background field. Taking
\[ \phi(x) = \phi_{0} + \eta e^{-ip\phi}, \] (17)
where \( \eta \) is a constant, we can get the evolution of \( Z_{\phi} \) from
\[ \partial_{k}Z_{\phi} = \frac{\partial}{\partial \rho_{0}} \left( \frac{\partial^{2}}{\partial \eta \partial \eta} \partial_{k}Z_{\phi} \right)_{\eta = 0} = 0. \] (18)
It can be shown that both methods lead to identical analytic expressions for \( \partial_{k}Z_{\phi} \). Substituting the expansion for the effective potential on the left-hand side of the evolution equation leads to a set of ordinary differential equations for the running minimum \( \rho_{0} \) and coefficients \( u_{n} \). These equations have a generic form
\[ \partial_{k}u_{n} - u_{n+1} \partial_{k} \rho = \frac{\partial}{\partial \rho_{0}} \left( \partial_{k}Z_{\phi} \right)_{\rho = \rho_{0}}. \] (19)
One can see from this equation that some sort of closure approximation is needed as the equation for \( u_{n} \) always include \( u_{n+1} \) coefficient etc. In this Letter we calculated \( \rho_{n+2} \) in the MFA with the effective potential given by Eq. (12). As already mentioned we follow the evolution of the chemical potential keeping density fixed. Defining the total derivative
\[ \frac{d}{dk} = \partial_{k} + \frac{d\rho_{0}}{dk} \frac{\partial}{\partial \rho_{0}} \] (20)
and applying it to the \( \frac{\partial U}{\partial \rho} \) (or to \( \frac{\partial U}{\partial \mu} \)) we obtain the following set of equations
\[ -2z_{\phi 0} \frac{d\rho_{0}}{dk} + \chi \frac{d\mu}{dk} = -\frac{\partial}{\partial \mu} \left( \partial_{k}Z_{\phi} \right)_{\rho = \rho_{0}}, \] (21)
where \( z_{\phi 0} \) is the coefficient in the leading term of the expansion for \( Z_{\phi} \) similar to Eq. (3), and
\[ \frac{du_{0}}{dk} + n \frac{d\mu}{dk} = \partial_{k}Z_{\phi} \big|_{\rho = \rho_{0}}. \] (22)
\[ -u_{2} \frac{d\rho_{0}}{dk} + 2z_{\phi 0} \frac{d\mu}{dk} = \frac{\partial}{\partial \rho} \left( \partial_{k}Z_{\phi} \right)_{\rho = \rho_{0}}, \] (23)
\[ \frac{du_{2}}{dk} - u_{3} \frac{d\rho_{0}}{dk} + 2z_{\phi 0} \frac{d\mu}{dk} = \frac{\partial^{2}}{\partial \rho^{2}} \left( \partial_{k}Z_{\phi} \right)_{\rho = \rho_{0}}, \] (24)
\[ \frac{dz_{\phi 0}}{dk} - z_{\phi 0} \frac{d\rho_{0}}{dk} + \frac{1}{2} \frac{d^{2}}{dk^{2}} \frac{d\mu}{dk} = -\frac{1}{2} \frac{\partial^{2}}{\partial \mu \partial \rho} \left( \partial_{k}Z_{\phi} \right)_{\rho = \rho_{0}}, \] (25)
where we have defined

\[ \chi' = \frac{\delta^3\bar{U}}{\delta \mu^2 \delta \rho} \bigg|_{\rho = \rho_0}, \quad z \phi_1 = -\frac{1}{2} \frac{\delta^3 U}{\delta \mu \delta \rho} \bigg|_{\rho = \rho_0}. \]  

(26)

These functions have also been calculated in the MFA. The set of evolution equations in symmetric phase can easily be recovered using the fact that chemical potentials do not run in symmetric phase and that \( \rho_0 = 0. \)

Let us now turn to the results. For simplicity we consider the case of the hypothetical “nuclear” matter with short range attractive interaction between two types of fermions, light and heavy, and study the behaviour of the energy gap as the function of the mass asymmetry. We choose the Fermi momentum to be \( p_F = 1.37 \, \text{fm}^{-1}. \) One notes that the formalism is applicable to any type of a many-body system with two fermion species from quark matter to fermionic atoms so that the hypothetical asymmetrical “nuclear” matter is simply chosen as a study case. We assume that \( M_a < M_b, \) where \( M_a \) is always the mass of the physical nucleon.

In this Letter we use a sharp cutoff function chosen in the form which makes the loop integration as simple as possible

\[ R_{F,i} = \frac{1}{2M_i} [(k + p_{\mu,i})^2 - q^2] \theta(p_{\mu,i} + k - q) 
+ [(k + p_{\mu,i})^2 + q^2 - 2p_{\mu,i}^2] \theta(q - p_{\mu,i} + k), \]  

(27)

and similarly for the boson regulator

\[ R_B = \frac{1}{2m}(k^2 - q^2) \theta(k - q). \]  

(28)

Here \( \theta(x) \) is the standard step-function. This type of boson regulator was also used in Ref. [11] (see also Ref. [12]).

The use of a sharp cutoffs can be potentially dangerous as it may generate the artificial singularities when calculating the flow of the renormalisation constants \( (Z \text{'s}) \) but seem to be harmless when all the evolution parameters are related to the effective potential RG flow as is the case here.

As we can see the fermion sharp cutoff consists of two terms which result in modification of the particle and hole propagators respectively. The hole term is further modified to suppress the contribution from the surface terms, which may bring in the dangerous dependence of the regulator on the cutoff scale even at the vanishingly small \( k. \) We found that the value of the gap practically does not depend on the starting point provided \( M_{a,b} \ll K. \) As expected, the system undergoes the phase transition to the gapped phase at some critical scale which depends on the value assumed for the parameter \( p_F a \) where \( a \) is the scattering length in vacuum. One notes that the critical scale does not depend on the mass asymmetry.

First we consider the case of the unitary limit where the scattering length \( a = -\infty. \) The results of our calculations for the gap are shown in Fig. 1.

We see from this figure that increasing mass asymmetry leads to a decreasing gap that seems to be a natural result. However, the effect of the boson loops is found to be small. We found essentially no effect in symmetric phase, 2–4% corrections for the value of the gap in the broken phase and even smaller corrections for the chemical potential so that one can conclude that the MF approach indeed provides the reliable description in the unitary limit for both small and large mass asymmetries. It is worth mentioning that the boson contributions are more important for the evolution of \( u_2 \) where they drive \( u_2 \) to zero as \( k \rightarrow 0 \) making the effective potential convex in agreement with the general expectations. This tendency retains in the unitary regime regardless of the mass asymmetry.

We have also considered the behaviour of the gap as the function of the parameter \( p_F a \) for the cases of the zero asymmetry \( M_a = M_b \) and the maximal asymmetry \( M_b = 10M_a. \) The results are shown in Fig. 2.

One can see from Fig. 2 that in the case of zero (or small) asymmetry the corrections stemming from boson loops are
small at all values of the parameter $p_{F\alpha}$ considered here (down to $p_{F\alpha} = 0.94$). On the contrary, when $M_b = 10M_a$ these corrections, being rather small at $p_{F\alpha} \geq 2$ becomes significant ($\sim 40\%$) when the value of $p_{F\alpha}$ decreases down to $p_{F\alpha} \sim 1$. We found that at $p_{F\alpha} \sim 1$ the effect of boson fluctuations becomes $\sim 10\%$ already for $M_b = 5M_a$. One can therefore conclude that the regime of large mass asymmetries, which starts approximately at $M_b > 5M_a$, moderate scattering length and/or the Fermi momenta is the one where the MF description becomes less accurate so that the calculations going beyond the MFA are needed. One might expect that the deviation from the mean field results could even be stronger in a general case of a large mass asymmetry and the mismatched Fermi surfaces but the detailed conclusion can only be drawn after the actual calculations are performed.

We were not able to follow the evolution of the system at small gap (or small $p_{F\alpha}$) because of the nonanalyticity of the effective action in this case. This nonanalyticity of the effective action can explicitly be demonstrated in the mean-field approximation. The flow equations can be solved analytically in this case and one can see from the solution, which has a closed-form expression in terms of an associated Legendre function, $P_l^m(y)$ that the fermion loops contain a term $\phi^\dagger \phi \log(\phi^\dagger \phi)$ at $k = 0$. It remains to be seen whether, within the given ansatz, the full solution of the system of the partial differential equations for the effective potential and running couplings is required to trace the evolution of the system in the case of small gaps.

To summarise, we have studied the pairing effect for the asymmetric fermion matter with two fermion species as a function of fermion mass asymmetry. We found that regardless of the size of the fermion mass asymmetry the boson loop corrections are small at large enough values of $p_{F\alpha}$ so that the MFA provides a consistent description of the pairing effect in this case. However, when $p_{F\alpha} \sim 1$ these corrections become significant at large asymmetries ($M_b > 5M_a$) making the MFA inadequate. In this case it seems to be necessary to go beyond the mean field description.

There are several ways where this approach can further be developed. The next natural step would be to consider the case of the mismatched Fermi surfaces taking into account the possibility of formation of Sarma, mixed and/or LOFF phases and exploring the importance of the boson loop for the stability of those phases and applying the approach to the real physical systems like fermionic atoms, for example. Work in this direction is in progress. The other important extension of this approach would be to include running of all couplings of the effective action and use different type of cutoff function, preferably the smooth one. The three body force effects [13], when the correlated pair interact with the unpaired fermion may also turn out important, especially for nondilute systems.

Acknowledgements

The author thanks Mike Birse, Niels Walet and Judith McGovern for numerous very helpful discussions.

References

Baryon–baryon bound states from first principles in 3 + 1 lattice QCD with two flavors and strong coupling

Paulo A. Faria da Veiga*, Michael O’Carroll

Departamento de Matemática Aplicada e Estatística, ICMC-USP, C.P. 668, 13560-970 São Carlos, SP, Brazil

Received 27 April 2006; received in revised form 10 October 2006; accepted 11 October 2006
Available online 2 November 2006
Editor: J.-P. Blaizot

Abstract

We determine baryon–baryon bound states in (3 + 1)-dimensional SU(3) lattice QCD with two flavors, 4 × 4 spin matrices, and in an imaginary time formulation. For small hopping parameter κ > 0 and large glueball mass (strong coupling), we show the existence of three-quark isospin 1/2 particles (proton and neutron) and isospin 3/2 baryons (delta particles), with asymptotic masses −3lnκ and isolated dispersion curves. Baryon–baryon bound states of isospin zero are found with binding energy of order κ^2, using a ladder approximation to a lattice Bethe–Salpeter equation. The dominant baryon–baryon interaction is an energy-independent spatial range-one attractive potential with an O(κ^2) strength. There is also attraction arising from gauge field correlations associated with six overlapping bonds, but it is counterbalanced by Pauli repulsion to give a vanishing zero-range potential. The overall range-one potential results from a quark, antiquark exchange with no meson exchange interpretation; the repulsive or attractive nature of the interaction depends on the isospin and spin of the two-baryon state.

© 2006 Published by Elsevier B.V.

PACS: 11.15.Ha; 02.30.Tb; 11.10.St; 24.85.+p

Keywords: Spectral analysis; Bethe–Salpeter equation; Excitation spectrum; Lattice QCD

1. Introduction

One fundamental problem in particle physics is to determine the low-lying energy–momentum (EM) spectrum of Quantum Chromodynamics (QCD). A convenient ultraviolet cutoff version is given by the Wilson lattice QCD model [1–4]. In the strong coupling regime, the infinite volume limit can be reached, hadrons are seen as tightly-bound bound states of quarks, and confinement is manifested.

In a recent series of papers [5–11], we started a research program aiming at understanding, from first principles, the hadronic particles and their bound states, in the context of imaginary-time formulation of SU(3) lattice QCD with strong coupling (small hopping parameter 0 < κ ≪ 1, and large glueball mass). Our goal is to bridge the gap between QCD and nuclear physics, to understand when and how bound states occur and how their binding is related to the effective Yukawa meson-exchange theory. We determined the low-lying EM spectrum for increasingly complex SU(3) QCD lattice models, using an imaginary-time formulation. The simplest case in which a two-baryon bound state appears is in the two-flavor total isospin I = 0, 1 sectors with 2 × 2 spin matrices in 2 + 1 dimensions; there are no bound states for I = 2, 3. However, this model is not complex enough to accommodate protons and neutrons in the one-particle spectrum.

Here, we consider the more realistic two-flavor case (with up and down quarks) in 3 + 1 dimensions with 4 × 4 Dirac spin matrices, which has a global SU(2) isospin symmetry. The following two ingredients are basic in our analysis: (i) derivation of...
spectral representations for the two- and four-baryon correlation functions, via Feynman–Kac formulas, which allow us to relate complex momentum singularities with the one- and two-baryon EM spectrum; (ii) a lattice version of the Bethe–Salpeter (B–S) equation in a ladder approximation. The spectral representations are new. Other approaches to the particle spectrum, both theoretical or numerical, are given in Refs. [1,12–24].

It must be pointed out that the detection of particle (and their bound states) masses from the exponential decay rate of suitable correlations, without a spectral representation, is meaningless, and the resulting values may be far from the correct ones, especially in cases where degeneracies are broken with small separations.

Concerning our results, we first show the existence of twenty, three-quark, one-particle states with isolated dispersion curves (upper gap property), and also their associated antiparticles, which includes the proton (p), the neutron (n) and the delta (Δ) particles. These one-baryon spectral results are exact and, making the lower order explicit, their asymptotic masses are \(-3\ln x - 3\kappa^2/4 + O(\kappa^3)\); if there is mass splitting it is due to contributions of \(O(\kappa^3)\) or higher. The upper gap property in the Hamiltonian formulation is unknown.

Next, we determine the two-baryon bound states in the \(I = 0\) sector and below the two-baryon threshold, which is given by twice the smallest of the baryon masses. We find several bound states with binding energies of order \(\kappa^2\). The most strongly bound, bound states are given by \(\Delta - \Delta\), total spin \(S = 3\) states, and also by a superposition of \(p-n\) and \(\Delta - \Delta\) total spin \(S = 1\) states. The more weakly bound, bound states are associated with a superposition of \(p-n\) and \(\Delta - \Delta\) total spin \(S = 0\) states, and also with \(\Delta - \Delta\), \(S = 2\) bound states. In contrast to the \(I = 0\) states treated here, we have found that for the maximum isospin \(I = 3\) sector there are bound states in the lowest total spin sectors \(S = 0, 1\) and no bound states if \(S = 2, 3\). These results are in agreement with our previous results of Ref. [10] that the attraction between the two particles decreases with increasing \(I\). Moreover, as before, there are two sources of attraction, namely, (i) the exchange of a quark and an antiquark, which is not a meson particle exchange, and (ii) gauge field correlation effects associated with six overlapping bonds. We point out that this work provides the main ingredients for a rigorous treatment of the model, going beyond the ladder approximation (see Refs. [25,26]).

2. The model and the one-baryon spectrum

We now introduce our SU(3) QCD lattice model and show how our results are obtained. The partition function is given formally by \(Z = \int e^{-S(\psi,\bar{\psi},g)} d\psi d\bar{\psi} d\mu(g)\), where \(S(\psi,\bar{\psi},g)\) is the Wilson action

\[
S = \frac{k}{2} \sum_{\alpha, \beta} \bar{\psi}_{\alpha f}(u) \Gamma^{\alpha \beta}_{\alpha \beta}(g_{u,u+\sigma+f}) \psi_{\beta f}(u) + \sum_{u \in Z^3_2} \bar{\psi}_{\alpha f}(u) M_{\alpha \beta} \psi_{\beta f}(u) - \frac{1}{g_0^2} \sum_{p} \chi(g_p).
\]

Here, besides the sum over repeated indices \(\alpha, \beta = 1, 2, 3\) (spin), \(a = 1, 2, 3\) (color) and \(f = +1/2, -1/2\) \(\equiv +, -\) (isospin), the first sum runs over \(u = (u^0, \bar{u}) = (u^1, u^2, u^3) \in \mathbb{Z}^3_\tau \equiv \{\pm 1/2, \pm 3/2, \pm 5/2, \ldots\} \times \mathbb{Z}^3, \sigma = \pm 1\) and \(\mu = 0, 1, 2, 3\). For \(F(\bar{\psi}, \psi, g)\), the normalized expectations are denoted by \(\langle F \rangle\). For more details about notation and on the treatment of symmetries such as gauge, SU(2) isospin, charge conjugation, parity, time-reversal and rotational symmetry, we refer to Refs. [6,7,10,11].

Letting \(e^\mu, \mu = 0, 1, 2, 3\), denote the unit lattice vectors, there is a gauge group matrix \(U(g_{u,u+e^\mu}) = U(g_{u,u+e^\mu})^{-1}\) associated with the directed bond \(u, u + e^\mu\). We take \(0 \leq g_0^2 < \kappa \ll 1\). The parameter \(m > 0\) is fixed such that \(M_{\alpha \beta} = M_{\delta \delta \beta}\) (\(\delta\) denoting the Kronecker delta) and \(M \equiv M(\kappa) = m + 2\kappa = 1\). The choice of the shifted lattice for the time direction, avoiding the zero-time coordinate, is so that, in the continuum limit, two-sided equal time limits of quark Fermi fields correlations can be accommodated.

The action \(S\) is one of a family of actions which has no spectral doubling for the free fermions, and the free fermion dispersion curve is increasing in each momentum component, convex for small momenta. The SU(2) spin symmetry is also recovered, for \(\kappa = 0\), for the upper (\(\alpha = 1, 2\)) and, separately, for the lower (\(\alpha = 3, 4\)) fermion spin components.

By polymer expansion methods (see Refs. [3,27]), the thermodynamic limit of correlations exists and truncated correlations have exponential tree decay. The limiting correlation functions are lattice translational invariant. Furthermore, the correlation functions extend to analytic functions in the coupling parameters \(\kappa\) and \(\beta \equiv 1/g_0^2\).

The quantum mechanical Hilbert space \(\mathcal{H}\) and the EM operators \(H\) and \(P_j\), \(j = 1, 2, 3\) are defined as in Refs. [3,6,7], as well as the Feynman–Kac formula (\(G, \bar{T}_0^0, \bar{T}_1, \bar{T}_2^0, \bar{T}_3^1\) F), \(\mathcal{H} = \langle \{T_0^0 \bar{T}_1^1 \bar{T}_2^2 \bar{T}_3^3\} \theta(G) \rangle, \mathcal{T}_0^0 = e^{-2H}, \mathcal{T}_j = e^{P_j}\), where \(T_{\mu = 0,1,2,3}^{e^\mu}\) denote translation of the functions of Grassmann and gauge variables by \(x^0 > 0, x = (x^1, x^2, x^3) \in \mathbb{Z}^3, \) respectively, for \(\mu = 0, 1, 2, 3\) and \(\Theta\) is an anti-linear operator which involves time reflection. We refer to each point in the EM spectrum associated to zero-momentum as mass. Here, we analyze the one- and two-baryon sectors of \(\mathcal{H}\). Points in the EM spectrum are detected as singularities in complex momentum space spectral representations of two and four-baryon correlations.

In order to determine two-baryon bound states, it is essential to carefully understand first the one-baryon sector. To do this, we use the decoupling of hyperplane method (see Refs. [6,7]) which has the nice feature that it reveals the explicit form of the fields that create the particles and that they are gauge-invariant and composite, so that there is no guesswork in the choice of the fields. The gauge-invariant barred fields listed below create baryon particles labeled by one-baryon total isospin (1st index), the one baryon z-component isospin (2nd index) and z-component of total spin (3rd index). Letting the dot ‘ mean the presence or absence of a bar
for all $\psi$’s, they are given by (with all fields evaluated at the same point)

$$
\tilde{B}_{i/2} \pm \pm \pm = \frac{\epsilon_{abc}}{3\sqrt{2}} (\bar{\psi}_{a\pm} \bar{\psi}_{b\pm} - \bar{\psi}_{a\mp} \bar{\psi}_{b\mp}) \psi_{c\pm}, \quad \tilde{B}_{i/2} \pm \pm \mp = \frac{\epsilon_{abc}}{6} \left( 2(-\sigma) \bar{\psi}_{a\pm} \bar{\psi}_{b-} + 2\sigma \bar{\psi}_{a-} \bar{\psi}_{b+} \right) \psi_{c\pm},
$$

$$
\tilde{B}_{i/2} \mp \mp \pm = \frac{\epsilon_{abc}}{2\sqrt{3}} \bar{\psi}_{a\pm} \bar{\psi}_{b\pm} \psi_{c\pm}, \quad \tilde{B}_{i/2} \mp \mp \mp = \frac{\epsilon_{abc}}{2\sqrt{3}} \bar{\psi}_{a\pm} \bar{\psi}_{b\pm} \psi_{c\pm} - (\sigma),
$$

(1)

where the index $\alpha$ in $\tilde{\psi}_{\alpha f}$ is a lower spin component $\alpha = 3, 4 \equiv +, -$. Also, $\sigma = \pm 2 + (2-\sigma)$ means a 2 is present for the upper (lower) sign; absent for the lower (upper) sign.

The fields obey the normalization $\langle B_i, B_j \rangle^{(0)} = -\delta_{ij}$, where $i, j$ are collective indices and the superscript $(0)$ denotes $\kappa = 0$ in the hopping terms in the action. In general, below, the superscript $(0)$ means the coefficient of $\kappa^n$, $n = 0, 1, \ldots$. In analogy with the two-flavor SU(3) gauge QCD, with only up ($f = +1/2$) and down ($f = -1/2$) quarks, we identify the particles associated with the fields in Eq. (1) as proton, neutron, $\Delta^{++}$, $\Delta^{-}, \Delta^{+}, \Delta^{-}$ and $\Delta^{0}$.

We treat the one-baryon states adapting the methods of Refs. [6,10]. The baryon–baryon correlation function is defined by $\chi$ here denotes the characteristic function and * complex conjugation

$$
G_{\ell_1 \ell_2}(u, v) = \langle B_{\ell_1}(u) B_{\ell_2}(v) \rangle_{\chi_{u0} \leq \psi, 0} - \langle \hat{B}_{\ell_1}(u) B_{\ell_2}(v) \rangle_{\chi_{u0} > \psi, 0} \equiv \tilde{G}_{\ell_1 \ell_2}(u - v),
$$

where $\ell = (I, I_z, s)$ is a collective index. It is important to stress that the apparently awkward form of this correlation emerges naturally from the two time orderings in the Feynman–Kac formula. By isospin symmetry, $G_{\ell_1 \ell_2}(x = u - v)$ diagonal in $I, I_z$ and, for $I$ fixed, independent of $I_z$. Also, the $z$-component of spin $s$ takes values $\pm 1/2$ for $I = \pm 1/2$ or $\pm 3/2$, $\pm 1/2$ for $I = 3/2$.

Setting $x^0 \equiv u^0 - v^0 \neq 0$ and dropping the isospin indices, from the Feynman–Kac formula and taking the Fourier transform $\tilde{G}_{\ell_1 \ell_2}(p) = \int_{\mathbb{R}^2} G_{\ell_1 \ell_2}(x) e^{-i p \cdot x}$, we obtain the following spectral representation for $\tilde{G}$

$$
\tilde{G}_{\ell_1 \ell_2}(p) = \tilde{G}_{s_{1/2} \ell_2}(\tilde{p}) - (2\pi)^3 \int_{-1}^{1} f(p^0, \lambda^0) d\lambda^0 \tilde{\alpha}_{\ell_1 \ell_2 / \lambda^0}(\lambda^0),
$$

(2)

where $\tilde{\alpha}_{\ell_1 \ell_2 / \lambda^0}$ has the approximate decomposition $\tilde{d}_{\ell_1 \ell_2 / \lambda^0} = Z_{s_{1/2}}(\tilde{p}) \delta(\lambda^0) - e^{-w(p)} d_{\nu_{1/2}}(\lambda^0, \tilde{p})$, where the first term corresponds to the one-particle contribution. Letting $\tilde{G}_{s_{1/2} \ell_2}(p)$ denote the analytic extension of $\tilde{G}_{s_{1/2} \ell_2}(p)\{^{-1}$ up to near the two-baryon threshold, we have $Z_{s_{1/2}}(\tilde{p})\}^{-1} = -2(2\pi)^3 e^{w(p)} \frac{\partial^2 Z_{s_{1/2}}(p)}{\partial^2 x} (p^0 = i \chi, \tilde{p})|_{\chi = w(\tilde{p})}$, and the $\lambda^0$ support of $d_{\nu_{1/2}}(\lambda^0, \tilde{p})$ is contained in $|\lambda^0| \leq |\chi|^{1-\epsilon}$.

3. Baryon–baryon bound states

Here we consider the subspace of the physical Hilbert space consisting of two-baryon states generated by the product of one-baryon fields, $B_{\ell_1}(x_1) B_{\ell_2}(x_2)$, and determine baryon–baryon bound states below the two-baryon threshold. The method employs a B–S equation for a suitable set of four-point functions.

If we ignore possible linear dependencies, the above subspace has dimension $400 = 20 \times 20$, which decomposes into two-baryon total isospin $I = 0, 1, 2, 3$ sectors, of dimensions 20, 108, 160, 112, respectively. Note that, because of Pauli exclusion, the dimension of this subspace is reduced if $x_1 = x_2$. Hence, any analysis that only considers coincident points does not give information on the whole two-baryon subspace.

To reduce the algebraic complexity of the analysis, hereafter we restrict our attention to the $I = 0$ sector, where we expect to find a deuteron, $n$–$p$ bound state. The $I = 0$ states are given by the isospin Clebsch–Gordan (C–G) linear combinations

$$
\frac{1}{2} s_{1/2}(x_1, x_2) = \sum_{l_1+l_2=0} \frac{1}{2} l_{1/2} l_{1/2} l_{1/2} l_{1/2} l_{1/2} s_{1/2}(x_1) l_{1/2} l_{1/2} s_{1/2}(x_2),
$$

for the coupling of two 1/2 isospin baryons, and similarly, with $(1/2 \rightarrow 3/2, s_j \rightarrow t_j)$, for the coupling of two 3/2 isos. We use $s_j (t_j)$ for isospin 1/2 (3/2) baryons and recall that the 1/2, 3/2 coupling does not give $I = 0$. We refer to this description of the two-particle states as the individual spin basis. It turns out that the dominant interaction between baryons admits a simpler
description in terms of total spin. The total spin states are obtained by taking C–G combinations of the ħ’s. Namely, \( \tilde{T}^{1/2}_{SSz}(x_1, x_2) = \sum_{s_1+s_2=S} a_{s_1s_2}^{1/2} \tilde{T}^{1/2}_{s_1s_2}(x_1, x_2) \) and similarly, with \((1/2 \to 3/2, S \to T, s_1 \to t_1)\), where the \(a\)’s are the spin C–G coefficients. Using their properties, we find that \( \tilde{T}^{1/2}_{SSz}(x_1, x_2) = (-1)^{S+1} \tilde{T}^{1/2}_{s_1s_2}(x_2, x_1) \) and \( \tilde{T}^{3/2}_{Tz}(x_1, x_2) = (-1)^{T+1} \tilde{T}^{3/2}_{t_1t_2}(x_2, x_1) \), so that for coincident points the states vanish for \(S = 0\), and for \(T = 0, 2\), a kinematic statistical effect; they are space symmetric (antisymmetric) for \(S = 1, T = 1, 3\) (\(S = 0, T = 0, 2\)).

We order the individual spin pairs as \((s_1, s_2) = (\frac{1}{2}, -\frac{1}{2}), (\frac{1}{2}, \frac{1}{2}), (\frac{1}{2}, -\frac{1}{2}) (t_1, t_2) = (\frac{3}{2}, \frac{3}{2}), (\frac{3}{2}, \frac{1}{2}), (\frac{3}{2}, -\frac{1}{2}), (\frac{3}{2}, -\frac{3}{2}), (\frac{1}{2}, \frac{3}{2}), (\frac{1}{2}, -\frac{3}{2}), (\frac{1}{2}, \frac{3}{2}), (\frac{1}{2}, -\frac{3}{2})\), and the total spin pairs as \((S, S_z) = (1, 1), (T, T_z) = (1, 1), (S, S_z) = (1, 0), (T, T_z) = (1, 0), (S, S_z) = (1, -1), (T, T_z) = (1, -1), (S, S_z) = (0, 0), (T, T_z) = (0, 0), (T, T_z) = (2, 2), (2, 1), (2, -2)\). The individual and the total spin basis are related by a real orthogonal transformation whose entries are, C–G coefficients.

As we do not know a priori which linear combination of states corresponds to a bound state, we consider the 20 × 20 matrix of (truncated) four-baryon functions

\[
M(x_1, x_2; x_3, x_4) = \begin{pmatrix}
M_{11} & M_{12} \\
M_{21} & M_{22}
\end{pmatrix}
(x_1, x_2; x_3, x_4),
\]

where, in the individual spin basis, suppressing the lattice coordinates \((x_1, x_2; x_3, x_4)\), and for \(x_1^0 = x_2^0\), \(x_3^0 = x_4^0\), \(x_1^0 \leq x_3^0\), \(M_{11,1,1,1}(x_1, x_2; x_3, x_4) = (\frac{1}{\sqrt{2}}(x_1, x_2; x_3, x_4)), M_{12,1,1,1}(x_1, x_2; x_3, x_4) = (\frac{1}{\sqrt{2}}(x_1, x_2; x_3, x_4)), M_{21,1,1,1}(x_1, x_2; x_3, x_4) = (\frac{1}{\sqrt{2}}(x_1, x_2; x_3, x_4)), M_{22,1,1,1}(x_1, x_2; x_3, x_4) = (\frac{1}{\sqrt{2}}(x_1, x_2; x_3, x_4)), M_{11,1,2,2}(x_1, x_2; x_3, x_4) = (\frac{1}{\sqrt{2}}(x_1, x_2; x_3, x_4)), M_{12,1,2,2}(x_1, x_2; x_3, x_4) = (\frac{1}{\sqrt{2}}(x_1, x_2; x_3, x_4)), M_{21,1,2,2}(x_1, x_2; x_3, x_4) = (\frac{1}{\sqrt{2}}(x_1, x_2; x_3, x_4)), M_{22,1,2,2}(x_1, x_2; x_3, x_4) = (\frac{1}{\sqrt{2}}(x_1, x_2; x_3, x_4))\).

For small \(κ\), \(M\) is symmetric (antisymmetric) for \(S = 1, T = 1, 3\) (\(S = 0, T = 0, 2\)). Due to the linear dependence relation at coincident points \(T_{SSz}(x_1, s_1; x_2, s_2) = 0, S = T = -1, 0, 1\), \(M\) still has a null space in the symmetric subspace of \(\ell_2\). This for this reason, we are forced to reduce the symmetric subspace for coincident points only. For coincident points (denoted by zero argument), in the total spin basis, by a lengthy computation we find \(M^{(0)}(0) = M^{(1)}(0) \oplus M^{(1)}(0) \oplus M^{(1)}(0) \oplus M^{(0)}(0)\), where \(M^{(1)}(0) = -\frac{1}{2}(-8^{2/3}, -8^{2/3}, 16)\) and \(M^{(0)}(0)\) acts as \(-4\) on the components from 7 to 13. The eigenvalues of \(M^{(1)}(0)\) are \(-4\) and 0, and we take the restriction of \(M^{(0)}(0)\) to the orthogonal complement of the null space in the symmetric subspace. So, for coincident points we have a 10-dimensional space. For non-coincident points, \(M^{(0)}(0)\) agrees with \(M^{(0)}(0)\), so that \(K\) is well defined. As in Ref.\[10\], we modify \(M\) at coincident points by a multiplicative constant 1/2. Precisely, we modify \(M\) by replacing \(M(x_1, x_2; x_3, x_4)\) by \(M'(x_1, x_2; x_3, x_4) = h(x_1, x_2) M(x_1, x_2; x_3, x_4) h(x_3, x_4)\), where \(h(x, y) = [1 - \delta(y - x)] + \frac{1}{\sqrt{2}}(x - y)\) and \(M\) becomes \(M' = M_0 + M'_0 K M\). For simplicity of notation, we drop the prime and keep using \(M\) and \(K\). By doing this, we improve the temporal decay properties of \(K\) and avoid having to deal with energy-dependent potentials (see Ref.\[11\]). \(K\) admits a Neumann expansion representation. With \(\delta M = M^{(0)}(0) - \delta M_0 = M_0 - M^{(0)}(0)\), we have \(K = \sum_{n=0}^{\infty} (-1)^n [(M^{(0)}(0) - \delta M_0)^n (M^{(0)}(0) - \delta M)^n - (M^{(0)}(0) - \delta M)^n]\). For small \(k\), the dominant contributions to \(K\) come from coincident points at \(k = 0\) and a space range one potential of order \(k^2\), and \(K^{(0)}(0) = (M^{(0)}(0) - \delta M)^n (M^{(0)}(0) - \delta M)^n = 0\), and, since \(\delta M = O(k^2)\) and \(\delta M_0 = O(k^2)\), the ladder approach to \(K\) is \(L \equiv K^{(2)}(k^2) = (M^{(0)}(0) - \delta M)^n (M^{(0)}(0) - \delta M^\dagger)^n k^2\). Its kernel is given by \(K^{(2)}(x_1, x_2; x_3, x_4) = \frac{1}{4}\delta(x_2 - x_1 - \sigma e^\dagger)\delta(x_2 - x_1 - \sigma e^\dagger)\delta(x_2 - x_1)\).
\[ \chi_i^0(x_1 - x_3) \delta(x_2 - x_4) \] where, in the individual spin basis, \( \nu^{(2)} \) is

\[ \nu^{(2)}_{s_1s_2s_3s_4} = \sum \left( C_{I_1I_2}^{1/2} \frac{1}{2} \frac{1}{2} (B_4^{I_1I_2} B_{4I_34} B_{I_2I_3} \bar{\psi} a_{1a_1f_1} \psi a_{2a_2f_2} )^{(0)} \right) \left( B_4^{I_1I_2} B_{4I_34} B_{I_2I_3} \bar{\psi} a_{1a_1f_1} \psi a_{2a_2f_2} )^{(0)} \right) \] with all fields at the same site, and where the prime means the sum over \( I_1, \ldots, I_4 \), with \( I_1 + I_2 = 0 = I_3 + I_4 \). Here, we have made use of the translation and rotational symmetries. A similar expression holds for the components \( t_1t_2s_3s_4, s_1s_2t_3t_4 \) and \( t_1t_2s_3t_4 \). By inspection, since the \( B \) fields have lower spin indices, the \( \psi \) and \( \bar{\psi} \) in the above means must have both spin indices either lower or upper. This is a quark–antiquark exchange, and since a meson particle has one upper and one lower spin index (see Refs. [7,11]), this is not a meson particle exchange, and we refer to it as a quasi-meson exchange. By a lengthy computation, it turns out that \( \nu^{(2)} \) has an especially simple, almost diagonal form in the total spin basis given by \( \nu^{(2)} = W \Theta W \Theta W \Theta W_{ld} + W_{2d} + W_{2d} \), with \( W = \frac{1}{\sqrt{2}} \left( \begin{array}{cc} -1 & -1 \end{array} \right) \) (with eigenvalues \( \pm 3 \)), \( W_{ld} = \left( \begin{array}{cc} 1 & 1 \end{array} \right) \) (with eigenvalues \( -1 \) and \( 9 \)), \( W_{ld} = -3 P_{13}, W_{2d} = -P_{22}, \) where \( P_{13} \) is the orthogonal projection on the 7th through 13th components and \( P_{22} \) on the 16th through 20th.

We return to the determination of bound states and solve the B–S equation using \( L \) (see Refs. [10,11] for details). In \( M \) and \( M_0 \), we first pass to the lattice relative coordinates \( \xi = x_2 - x_1, \eta = x_4 - x_3, \) and \( \tau = x_3 - x_2 \), take the Fourier transform in the variable \( \tau \), with dual variable \( (k^0, \vec{k}) \), and set the system momentum \( \vec{k} = 0 \). The new quantities are denoted with a hat and have arguments \( (\xi, \eta, k^0) \). (Note that \( M \) and \( M' \) have the same \( k^0 \) singularities!) In our ladder approximation \( L = \kappa^2 \nu^{(2)} \sum_{\kappa j} \frac{1}{2} \delta(\vec{q} - \vec{\xi}) \delta(\sigma \vec{e}^j - \vec{\xi}) \), and the B–S equation has the solution

\[ \hat{M}(\vec{\xi}, \vec{\eta}, k^0) = \hat{M}_0(\vec{\xi}, \vec{\eta}) + \sum_{\sigma_1, \sigma_2, j, j_2} \hat{M}_0(\sigma_1, \sigma_2, j, j_2) \kappa^2 \nu^{(2)} (1 - \kappa^2) \hat{M}_0(\sigma_1, \sigma_2, j, j_2) \hat{M}_0(\sigma_2, j, j_2), \] where the \( \hat{M}_0 \) lattice indices are restricted to the spatial nearest neighbors of zero, and we suppress the \( k^0 \) dependence and spin indices on the r.h.s. Reinstating the spin indices, the resulting matrix is \( 120 \times 120 \).

We determine bound states below the two-baryon threshold as singularities in \( k^0 = i(2 \vec{m} - \epsilon) \), where here \( \vec{m} \) denotes the minimum of the baryon masses and \( \epsilon > 0 \) is the bound state binding energy. To go further, we need properties of \( M_0 \). Starting from the spectral representation for the two-baryon function, and making a spin diagonal approximation to \( G \), the dominant contribution to \( \hat{M}_0 \) is given by, in the total spin basis,

\[ \hat{M}_{0,SS'S'}(\vec{\xi}, \vec{\eta}, k^0) = \left[ -2(2\pi)^3 \right] \int \hat{G}_{S'S'}(\vec{p}) \hat{G}_{S'S'}(\vec{p}) d\vec{p} - 2(2\pi)^3 \times \left[ \int \frac{1}{2} f(k_0, \lambda^0, \lambda^0) d_{\lambda^0} \alpha_{\vec{p}, SS} (\lambda^0) d_{\lambda} \alpha_{\vec{p}, S'S'} (\lambda^0) \right] \delta_{SS} \delta_{S'S'}, \] where \( \delta_{SS} \cos \beta \sin \beta \cos \vec{p} \cdot \vec{\xi} \cos \vec{p} \cdot \vec{\eta} + \delta_{SS} \sin \beta \sin \vec{p} \cdot \vec{\xi} \sin \vec{p} \cdot \vec{\eta} \); and the same for \( T = 1, 3 \) (\( T = 0 \)) replacing \( S = 1 \) (\( S = 0 \)). We further make the following approximations to \( \hat{M}_0 \): (a) retain only the product of one-particle contributions; (b) \( u(\vec{p}) - m \approx \kappa^2 \rho^2 / 8 \); (c) \( \vec{m} \approx -3 \ln \kappa \) and (d) \( Z(\vec{p}) \approx -2(2\pi)^3 e^{-\vec{u}(\vec{p})} \approx -2(2\pi)^3 \kappa^3 \). With these approximations, the singularities of \( \hat{M}_0 \) below the two-baryon threshold, occur as zeroes of the determinant of the \( 120 \times 120 \) matrix \( [1 - \kappa^2 \nu^{(2)} \hat{M}_0(\sigma_1, \sigma_2, j, j_2) \hat{M}_0(\sigma_2, j, j_2)] \). Anticipating that the binding energy is of order \( \kappa^2 \), we also approximate the denominator of \( \hat{M}_0 \) to get \( \hat{M}_{0,SS'S'}(\xi, k^0) \approx -2(1 - e^{-\epsilon})^{-1}(2\pi)^3 f_{\kappa^2} \hat{Z}(\vec{p}, \vec{\xi}, \vec{\eta}) d\vec{p} \), and the same for \( T \)'s. Note that \( \hat{M}_0(\sigma_1, \sigma_2, j, j_2) \) is diagonal in \( j_1 \) and \( j_2 \), and that the integral takes the values \( \pm (2\pi)^3 / 2 \). Thus, the bound state condition becomes \( \det[1 + 4(1 - e^{-\epsilon})^{-1}\nu^{(2)}\kappa^2] = 0 \), for a \( 20 \times 20 \) matrix. Next, using the spectral representation for \( \nu^{(2)} \), if \( \lambda \) is a negative eigenvalue, then the bound state condition is \( 1 + 4\kappa(1 - e^{-\epsilon})^{-1}\kappa^2 = 0 \), with approximate solution \( \epsilon \approx 4\kappa^2 \), which is the bound state binding energy. Hence, we have a bound state for each negative eigenvalue of \( \nu^{(2)} \), with multiplicity given by the space dimension \( d = 3 \). For these bound states, we have attractive potential wells at \( \pm e^j, j = 1, \ldots, d \), which would give us a multiplicity \( 2d \). However, we only have symmetric or antisymmetric wave functions, depending on the total spin through \( \hat{M}_0 \), which reduces the multiplicity to \( d \).

Referring to the spectral decomposition of \( \nu^{(2)} \) above, and recalling the ordering of the total spin basis, we see that the most strongly bound, bound states \( (\lambda = -3) \) are associated with a superposition of \( p-n \) and \( \Delta-\Delta \) total spin 1 states; and also with \( \Delta-\Delta \) total spin 3 states. The more weakly bound, bound states \( (\lambda = -1) \) are associated with a superposition of \( p-n \) and \( \Delta-\Delta \) total spin 0 states; and also with \( \Delta-\Delta \) total spin 2 states.

Similarly, applying our method to the maximum \( I = 3 \) sector (freezing all isospins to +), we find no bound states if \( S = 1, 2 \), but bound states do appear if \( S = 0, 1 \). Moreover, the binding for the \( S = 0 \) case is stronger than the one for \( S = 1 \). This is similar to the bound state results of Ref. [10] regarding \( 2 \times 1 \) dimensions, \( 2 \times 2 \) spin matrices and two flavors, i.e. with interchanged roles of spin and isospin. We are now working on the more complex \( I = 2, 3 \) sectors.
4. Final remarks

We note that although our bound state results are obtained using quite complicated machinery, in the end a simple picture emerges for the formation of a baryon–baryon bound state. The two-baryon dynamics in relative coordinates behaves approximately like that of a non-relativistic one-particle lattice Hamiltonian $T + V$ with lattice kinetic energy $-\kappa^3\Delta/8$, where $\Delta$ is the spatial lattice Laplacian and the potential energy $V$ is $\kappa^2 V'$, the quasi-meson exchange space range-one potential which dominates the kinetic energy for small $\kappa$. The attractive or repulsive nature of the interaction depends on the isospin, spin spectral structure of $V$ at a single site of space-range one. Because of the $\kappa^3$ dependence of the kinetic energy $T$ and the $\kappa^2$ dependence of the potential energy, there is no minimal critical value of the interaction strength needed for the presence of a bound state.

To conclude, if we consider contributions to the $B$–$S$ kernel comprised of linear chains of quark, antiquark pairs they result in an exponentially decreasing potential with decay rate $-2\ln\kappa$, as for the Yukawa theory. It would be interesting to look for the expected Ornstein–Zernicke like correction to this potential and to determine the spin and isospin dependence of the binding energy, and the effect of the number of flavors. Especially, the analysis of the SU(3) flavor case can shed some light in understanding bound states when strangeness and, consequently, $\Lambda$ particles are present. Also, there is the problem of determining bound states of more than two baryons. However, lattice effects are expected to be relevant and unrealistic in determining the resulting geometric spatial configuration of possible bound states. Finally, and more importantly, we would like to know how the baryon spectrum and the bound state binding energies behave near the scaling limit.

Acknowledgements

This work was supported by CNPq and FAPESP. Algebraic computations were checked with Maple and Matlab, for which we thank Louis O’Carroll.

References

[24] Ph. de Forcrand, S. Kim, hep-lat/0608012.
Complex CKM from spontaneous CP violation without flavor changing neutral current

G.C. Branco\textsuperscript{a,b,*}, R.N. Mohapatra\textsuperscript{a,c,d}

\textsuperscript{a} Theoretische Physik, Technische Universität München, Garching, Germany
\textsuperscript{b} Departamento de Fisica and CFTP, Instituto Superior Técnico, Lisboa, Portugal
\textsuperscript{c} Department of Physics, University of Maryland, College Park, MD 20742, USA
\textsuperscript{d} Sektion Physik, Universität München, München, Germany

Received 4 August 2006; received in revised form 19 October 2006; accepted 20 October 2006
Available online 7 November 2006
Editor: G.F. Giudice

Abstract

We analyse the general constraints on unified gauge models with spontaneous CP breaking that satisfy the conditions that (i) CP violation in the quark sector is described by a realistic complex CKM matrix, and (ii) there is no significant flavor changing neutral current effects in the quark sector. We show that the crucial requirement in order to conform to the above conditions is that spontaneous CP breaking occurs at a very high scale by complex vevs of standard model singlet Higgs fields. Two classes of models are found, one consisting of pure Higgs extensions and the other one involving fermionic extensions of the standard model. We give examples of each class and discuss their possible embeddings into higher unified theories. One of the models has the interesting property that spontaneous CP violation is triggered by spontaneous P violation, thereby linking the scale of CP violation to the seesaw scale for neutrino masses.

© 2006 Published by Elsevier B.V.

1. Introduction

It is now becoming increasingly clear that the dominant contribution to low energy CP violation arises from the complex CKM matrix which parameterizes the weak quark current coupling to the W-boson. Indeed the recent measurement [1] of the angle \( \gamma = -\text{Arg}(V_{ud}V_{cb}V_{cd}^* V_{ub}^*) \) provides evidence [2] for a complex CKM matrix even if one allows for New Physics (NP) contributions to \( B_d - \bar{B}_d \) mixing and \( B_s - \bar{B}_s \) mixings.

However, this cannot be the full story of CP violation in elementary particle interaction [3] since it is believed that the explanation of the only cosmic manifestation of CP nonconservation i.e. the asymmetry between matter and anti-matter must come from sources other than the CKM CP violation; similarly the solution to the QCD \( \theta \) problem may also imply new forms of CP violating interactions. Moreover, there is the fundamental question of the origin and nature of CP violation and its relation to other constituents and forces.

Even before the full story of CP violation is clear, one can ask the question as to whether the observed CKM CP violation is spontaneous in origin [4] or intrinsic to the Yukawa couplings in the theory. This question has nontrivial cosmological implications since spontaneous CP violation will lead to domain walls and in order to avoid conflict with observations such as WMAP data, one must have the scale of this breaking to be above that of the inflation reheating, thus imposing constraints on both cosmological as well as particle physics aspects of models.

* Corresponding author.
E-mail address: gbranco@mail.ist.utl.pt (G.C. Branco).

0370-2693/$ – see front matter © 2006 Published by Elsevier B.V.
doi:10.1016/j.physletb.2006.10.039
In practical construction of models with spontaneous CP breaking, one must have one or more Higgs fields to have complex vevs [4]. It is obvious that implementing this requires extending the standard model, by having either more Higgs/or fermion fields plus Higgs because gauge invariance allows no room for Higgs vevs to be complex in the standard model. Furthermore, since spontaneous CP violation (SCPV) requires nontrivial constraints on the realistic gauge models, it is not surprising that the process of implementing it can lead to unpleasant side effects. One such unpleasant effect is the plethora of flavor changing neutral current (FCNC) effects induced in the process of obtaining spontaneous CP breaking.

Therefore, the challenge in constructing realistic models with spontaneous CP violation is twofold:

(i) One should achieve genuine spontaneous CP violation and assure that the vacuum phase does lead to a non-trivially complex CKM matrix. This is not an easy task since CP invariance of the Lagrangian requires the Yukawa couplings to be real.

(ii) One should find a natural suppression mechanism for FCNC in the Higgs sector. Again, this is a challenging task, since there is in general a close connection [5] between the appearance of FCNC and the possibility of generating a complex CKM matrix through CP violating vacuum phases.

The above link between SCPV and FCNC can be seen by considering a two Higgs extension \((\phi_{1,2})\) of the standard model to implement SCPV. It is well known (and we repeat the derivation in Section 2 and in Appendix A) that general two Higgs models have FCNC mediated by neutral Higgs fields. In order to suppress these FCNC effects one may consider two possibilities. One consists of the introduction of extra symmetries which eliminate FCNC and guarantee natural flavour conservation (NFC) [6] in the Higgs sector. It is well known that the introduction of such symmetries in the two Higgs doublet framework eliminates the possibility of having spontaneous CP violation [5]. With three Higgs doublets one can have NFC and yet achieve spontaneous CP violation but the resulting CKM matrix is real, in contradiction with recent data. Above we have considered the case where FCNC are avoided through the introduction of extra symmetries, not by fine-tuning. It has been shown that even if one considers elimination of FCNC through fine-tuning, for three generations one cannot generate a realistic complex CKM matrix [7]. The other possibility for suppressing FCNC effects is by choosing a large mass for the neutral Higgs which violate flavour. Indeed the strength of FCNC effects is proportional to \(1/M_H^2\) where \(H\) denotes the new neutral Higgs field (we will denote the standard model Higgs by \(h\)). So clearly, suppression of FCNC effects require that \(M_H\) become very large. On the other hand, as we show below, the magnitude of the CP phase (denoted by \(\delta\) in the text) in this model is given by \(\delta \sim M_H/M_H^\prime\) so that as \(M_H \rightarrow \) very large, \(\delta \rightarrow 0\) and the theory becomes almost CP conserving. Note that to obtain CKM CP violation, we need \(\delta \sim 1\). We will thus show that in the context of models with SCPV at the electroweak scale, it is not possible to obtain a complex CKM matrix while suppressing FCNC effects. In this class of SCPV models, obtaining a large CP phase and having significant FCNC seem to go together.

In this Letter, we discuss the conditions under which this connection can be avoided. We point out that the crucial point is to have CP broken at a high energy scale. We present two classes of models: one where the extension involves only the Higgs sector of the standard model and another one which involves the fermion sector as well. In the latter case, there is a small departure from unitarity of the CKM matrix.

Several of the models we discuss have already been considered in the literature. We present a systematic classification of these models, adding some new ones and sharpening the connection between SCPV and FCNC. In particular, we present criteria for constructing realistic SCPV models free of FCNC constraints.

This Letter is organized as follows: in Section 2, we discuss the connection between SCPV and FCNC in doublet Higgs extension of the SM. In Section 3, we discuss spontaneous CP breaking at high scale in a pure Higgs extension and show how one can avoid the FCNC effects in this case. In Section 4, we present a fermionic extension of the SM with spontaneous CP breaking at high scale. In Section 5, we discuss these two classes models into a left–right model and discuss two models one of which has the interesting property that spontaneous CP violation is triggered by spontaneous P violation. In Section 6, we briefly comment on how our ideas can be extended to supersymmetric models and finally in Section 7, we present our conclusions. In the Appendices A and B we present a detailed demonstration of the results of sections in Sections 2 and 3.

2. Two Higgs doublet model for SCPV and FCNC

The simplest extension of the standard model that can accommodate spontaneous CP violation is the two Higgs doublet model. If we denote the two Higgs doublets as \(\phi_{1,2}\), and define \(V_0(x, y) = -\mu_1^2 x - \mu_2^2 y + \lambda_1 x^2 + \lambda_2 y^2 + \lambda_3 x y\), we can write the potential as follows:

\[
V(\phi_{1,2}) = V_0(\phi_1^+, \phi_2^+, \phi_2^- \phi_1^-) + V_{12},
\]

where

\[
V_{12}(\phi_1, \phi_2) = \mu_{12}^2 \phi_1^+ \phi_2^- + \lambda_4 (\phi_1^+ \phi_2^-)^2 + \lambda_5 \phi_1^+ \phi_2^+ \phi_1^- + \lambda_6 \phi_1^+ \phi_2^- \phi_2^+ + \text{h.c.} + \lambda_7 \phi_1^+ \phi_2^+ \phi_2^- \phi_1^-. \tag{2}
\]

We can now write down the potential in terms of the electrically neutral components of the doublets. It looks exactly the same as the above potential as long as we understand the various fields as the neutral components of the fields.
In order to discuss spontaneous CP violation [8], we look for a minimum of the form:

\[ \langle \phi_1 \rangle = \left( \frac{0}{\sqrt{2} v_1} \right), \quad \langle \phi_2 \rangle = \left( \frac{0}{\sqrt{2} v_2 e^{i \delta}} \right) \]

(3)

The potential at this minimum looks like

\[ V(v_1^2, v_2^2, \delta) = V_0(v_1^2, v_2^2) + \frac{1}{4} \lambda_3 v_1^2 v_2^2 + \mu_{12} v_1 v_2 \cos \delta + \frac{1}{2} \lambda_4 v_1^2 v_2^2 \cos 2\delta + \frac{1}{2} (\lambda_5 v_1^2 + \lambda_6 v_2^2) v_1 v_2 \cos \delta. \]

(4)

The three extremum equations are:

\[ \begin{align*}
-\mu_1^2 + \lambda_1 v_1^2 + (\lambda_3 + \lambda_3^3) v_2^2 + \lambda_4 v_2^2 \sin 2\delta & = 0, \\
-\mu_2^2 + \lambda_2 v_2^2 + \frac{1}{2} (\lambda_3 + \lambda_3^3) v_1^2 + \lambda_4 v_1^2 \sin 2\delta & = 0, \\
-\sin \delta (\mu_{12} v_1 v_2 + 2 \lambda_4 v_1^2 v_2^2 \cos \delta + v_1 v_2 (\lambda_5 v_1^2 + \lambda_6 v_2^2)) & = 0.
\end{align*} \]

(5)

Now let us study the implications of the extremum equations for SCPV and FCNC. Writing the Yukawa couplings as \( \phi \) from the doublet \( hu, \) the two neutral Higgs fields mix and therefore the \( \mu \) mass matrix will have an admixture of the light Higgs \( h \) but mixing is always proportional to the mass ratio \( m_h^2/M_H^2 \) and \( M_H \) is much larger than the neutral Higgs which mediate FCNC. We will consider that both possibilities do not work as far as generating a viable complex CKM matrix, but the discussion is useful in order to motivate the breaking of CP at a high energy scale which will be considered in Sections 3 and 4.

2.1. Eliminating FCNC through extra symmetries

It is well known that it is possible to avoid FCNC by introducing for example a \( Z_2 \) symmetry which restricts the Yukawa couplings so that only one Higgs doublet gives mass terms to the down quarks while the other doublet gives mass to the up quarks. However, it has been shown [5] that the same symmetry which leads to these selective Yukawa couplings prevents the occurrence spontaneous CP breaking. A possible way out of this difficulty involves the introduction of a third Higgs doublet. In this case it is possible to obtain a CP violating vacuum [5] but the CKM matrix is real, in conflict with the recent experimental findings. The reason why CKM matrix is real in this case has to do with the fact that due to the selective Yukawa couplings, the vacuum phase which appears in the quark mass matrices can be eliminated by rephasing right handed quark fields.

2.2. Suppressing FCNC effects through large Higgs masses

It is straightforward to see that we could diagonalize one set of Yukawa couplings \( h^{u,d,1} \) so that the neutral Higgs \( h \) coming from the doublet \( \phi_1 \) has flavor conserving couplings whereas that from \( \phi_2 \) \((H)\) has flavor violating couplings. In general of course the two neutral Higgs fields mix and therefore the \( h^{u,d} \) coupling which in the symmetry limit involves only the \( H \) Higgs field will have an admixture of the light Higgs \( h \) but mixing is always proportional to the mass ratio \( m_h^2/M_H^2 \) assuming \( M_H \approx m_h \).

Thus FCNC processes will arise via the tree level exchange of \( H \) boson and will be proportional to \( M_H^{-2} \) and a contribution from the mixing term which due to the mixing will also have the same kind of power dependence on \( M_H \). Therefore in order to suppress FCNC interactions, we must demand that \( M_H \) be very large. This can be achieved by making \( -\mu_1^2 > 0 \) and \( |\mu_2^2| \gg v_{uk} \).

Let us now study Eq. (6). This equation tells us the scale of the vev \( v_2 \) which depends on the scale of the mixing term \( \mu_{12} \). (Note that getting the correct weak scale fixes \( \mu_1^2 \) to be of order \( v_{uk} \) and stopping FCNC tells us that \( |\mu_2^2| \gg v_{uk}^2 \) but so far \( \mu_{12} \) remains a free parameter.) We have two cases: (i) \( \mu_{12}^2 \sim v_{uk}^2 \) and (ii) \( \mu_{12}^2 \sim M_H^2 \sim |\mu_2^2| \gg v_{uk}^2 \). In case (i), it is easy to see using the middle equation above that:

\[ v_2 \sim \lambda_5 \frac{v_1^3}{|\mu_2^2|} \ll v_1 \]

(8)

i.e. the vev of \( \phi_2 \) is highly suppressed in the limit of no FCNC. Note that the mass of the second neutral Higgs is not of order \( v_2 \) since in this case the vev is induced by a tadpole like diagram. Substituting this small value of \( v_2 \) in Eq. (8), we then see that for natural values of the parameters (\( \lambda_i \)), the only solution for the CP violating phase is \( \delta = 0, \pi, \ldots \).
We therefore conclude that in this simple model, the requirement of suppression of the neutral current effects implies no SCPV. The main point is that to get a large enough SCPV phase, Eq. (7) tells us that $v_2$ must be comparable in magnitude to $v_1$. For this to happen, we must have $|\mu_{23}^2| \sim v_{uk}^2$, which again means that there must be large FCNC effects at low energies.

The above result can also be seen as follows: In a two Higgs doublet theory, one can change the basis of Higgs bosons to pass to a basis where the new doublets are $\Phi_1 = (v_2 e^{i\delta} \phi_1 - v_1 \phi_2)/\sqrt{v_1^2 + v_2^2}$ and $\Phi_2$ is the orthogonal combination to $\Phi_1$, where we have anticipated the vevs of the fields in the original basis, as discussed above. Now we see that $\langle \Phi_1 \rangle = 0$ while $\langle \Phi_2 \rangle \neq 0$ and it leads to the same mass matrices for quarks as before. Now we can choose parameters of the Higgs potential such that the mass of $\Phi_1$ is very large to avoid FCNC effects. In this case, the effective theory below the mass of $\Phi_1$ i.e. $M_{\Phi_1}$ is same as the standard model up to zeroth order in $M_W/M_{\Phi_1}$. Therefore, to this order, the vev of $\Phi_2$ (which is the equivalent of the standard model Higgs) will be real, and there will be no spontaneous CP violation in the theory (to order $M_W/M_{\Phi_1}$). This again proves that in the limit of zero FCNC, there will be no SCPV. In Appendix A, we give explicit calculations in the mass basis that substantiates this conclusion.

This result can be generalized to the case of arbitrary number of Higgs doublets. For example for the case of three doublets, the argument is that as long as all the doublets couple to quark fields, at least two of the neutral Higgs bosons i.e. $H_{1,2}$ must be heavy in order to avoid large FCNC effects and this implies that $|\mu_{23}^2| \gg v_{uk}^2$; in that case their vev’s must be suppressed and of order $v_{uk}^2/|\mu_{23}^2|$ and therefore small. The potential will then be forced to choose the minimum such that all SCPV phases are zero.

3. High scale spontaneous CP violation leading to complex CKM while avoiding FCNC: Model with extra Higgs only

In this section, we show how the FCNC problem is avoided if spontaneous violation of CP symmetry arises at a high scale. First we discuss this using a model with two $SU(2)_L \times U(1)_Y$ Higgs doublets $\phi_{1,2}$ as before and a complex singlet $\sigma$. The potential for this case can be written as follows:

$$V_{\Phi_1,\Phi_2,\sigma} = V(\phi_{1,2}) + V(\sigma) + V(\phi,\sigma),$$

where $V(\phi_{1,2})$ is defined in Eqs. (1), (3) and the other two terms are given by

$$V(\sigma) = -M_0^2 \sigma^* \sigma + M_1^2 \sigma^2 + \lambda_\sigma (\sigma^* \sigma)^2 + \lambda'_\sigma \sigma^4 + \lambda'' \sigma^3 \sigma^* + \text{h.c.}$$

and

$$V(\phi,\sigma) = M_{2,ab} \phi^*_a \phi_b \sigma^* + \kappa_{1,ab} \phi^*_a \phi_b \sigma^2 + \kappa_{2,ab} \phi^*_a \phi_b \sigma^* \sigma + \text{h.c.}$$

It is clear that the minimum of the potential $V(\sigma)$ corresponds to $\langle \sigma \rangle = A e^{i\alpha}$, where $\Lambda \sim M_{0,1,2} \gg v_{uk}$ and $\alpha$ can be large. Substituting this vev in the potential, we can write the effective tree level potential for the $\phi_{1,2}$ fields at low energies to be:

$$V_{\text{eff}}(\phi_1,\phi_2) = V(\phi_{1,2}) + V_{\text{new}},$$

where $V_{\text{new}} = (M_{2,ab} A e^{i\alpha} + \kappa_{1,ab} A^2 e^{2i\alpha} + \kappa_{2,ab} A^2) \phi^*_a \phi_b + \text{h.c.} \equiv A^2 (\lambda_{11} \phi^*_1 \phi_1 + \lambda_{22} \phi^*_2 \phi_2 + \lambda_{12} e^{i\beta} \phi^*_1 \phi_2) + \text{h.c.}$ If we keep only the neutral components of the Higgs doublets, then the form of the potential is

$$V_{\text{eff}} = A^2 (\lambda_{11} \phi^*_1 \phi_1 + \lambda_{22} \phi^*_2 \phi_2 + \lambda_{12} e^{i\beta} \phi^*_1 \phi_2 + \text{h.c.}) + \sum_{abcd} \lambda_{abcd} \phi^*_a \phi_b \phi_c \phi_d + \text{h.c.},$$

where $A \gg v_{uk}$. It is clear that although CP is spontaneously broken at a high scale $A$, at low energies, one has CP explicitly softly broken [9] by the bilinear terms in $\lambda_{12}$. Note that both the fields $\phi_{1,2}$ have Yukawa couplings and we can make a redefinition of the phase of one of the doublet fields (say $\phi_2$) i.e. $\phi_2 \rightarrow e^{-i\beta} \phi_2$ so that all the bilinear and $O(A^2)$ terms in the potential become phase independent but the Yukawa couplings become complex. Thus the effective theory at low energies looks naively like hard CP violation, even though it is spontaneous CP violation at a very high scale. The Yukawa coupling Lagrangian looks like

$$L_Y = \bar{Q}_{La} (h^{a1}_{ub} \phi_1 + h^{a2}_{ub} e^{-i\beta} \phi_2) u_{R,b} + \bar{Q}_{La} (h^{d1}_{ub} \phi_1 + h^{d2}_{ub} e^{i\beta} \phi_2) d_{R,b} + \text{h.c.}$$

This still does not imply a viable complex CKM matrix; to achieve that, we must show that the vev of $\phi_2$ where the phase resides, does not become very tiny when we demand the suppression of FCNC. In order to show this, let us write down the extremization of the potential as in Section 2. For simplicity, we keep only the $\lambda_{1111}, \lambda_{2222}$ and $\lambda_{1122}$ terms in the potential but our results follow in general:

$$-\mu_1^2 + A^2 \lambda_{11} + \lambda_{1111} v_1^2 + \lambda_{1122} v_2^2 v_1 + v_2 (A^2 \lambda_{12}) = 0,$$

$$-\mu_2^2 + A^2 \lambda_{22} + \lambda_{2222} v_2^2 + \lambda_{1122} v_1^2 v_2 + v_1 (A^2 \lambda_{12}) = 0.$$
Of course we do not need to make the rephasing $\phi_2 \rightarrow e^{-i\beta}\phi_2$ and eliminate the phase from the bilinear terms. If we do not do rephasing, the extremum equation of the Higgs potential would look like:

$$-\lambda_2 \sin(\beta + \delta) - \sin \delta [2\lambda_4 v_1^2 v_2^2 \cos \delta + v_1 v_2 (\lambda_5 v_2^2 + \lambda_6 v_1^2)] = 0.$$ (17)

Since $\Lambda^2 \gg v^2$, it is clear that to an excellent approximation one has:

$$\beta = -\delta.$$ (18)

The phase $\delta$ would then appear in the quark mass matrices which will be nontrivially complex, thus leading to a complex CKM matrix. In Appendix B, we discuss how the fine-tuning needed to keep the standard model Higgs at the electroweak scale does not prevent the components of the extra Higgs become superheavy in order to suppress the FCNC effects.

4. SCPV without FCNC problem in fermionic extensions of standard model

In this section, we turn to SCPV models which extend the fermionic sector of the standard model to solve the flavor changing neutral current problem while giving complex CKM at the weak scale [10,11]. Typical features of these models are new SM singlet quarks and SM singlet Higgs fields, the latter used to generate SCPV. We briefly review the model in [10] which is a typical model of this type to illustrate the main points of our discussion i.e. spontaneous violation of CP at high scale without FCNC problem but with complex CKM.

In the model of Ref. [10], the new SM singlet vector like fermion are of down type: $(D_{L,R})$ with $U(1)_R$ quantum number $-2/3$ and a complex singlet Higgs field $\sigma$ as in Section 3. The potential for the $\sigma$ field is the same as in Eq. (11). As a result, the $\sigma$ field has a complex vev leading to high scale spontaneous CP violation (since $\langle \sigma \rangle = \Lambda \gg v_{ewk}$).

The CP violation is transmitted to the weak scale via its couplings given below:

$$\mathcal{L}_\sigma = \sum_a \bar{D}_L d_{a,R} (g_a \sigma + g'_a \sigma^*) + (f \sigma + f' \sigma^*) \bar{D}_L D_R + \text{h.c.},$$ (19)

where $g_a, g'_a, f, f'$ are real due CP conservation. But after symmetry breaking, the mass matrix contains terms mixing the heavy $D$ quarks with the light $d$ quarks [10]. This can be seen by writing down the full down quark mass matrix (in the notation $\tilde{\psi}_L M_{dD} \psi_R$):

$$M_{dD} = \begin{pmatrix} m_d & 0 \\ \Lambda (g e^{i\delta} + g' e^{-i\delta}) & \Lambda (f e^{i\delta} + f' e^{-i\delta}) \end{pmatrix},$$ (20)

where $g$ and $g'$ denote the row vectors $(g_1, g_2, g_3)$ and $(g'_1, g'_2, g'_3)$. Diagonalizing $M_{dD} M_{dD}^T$, we can get the generalized $4 \times 4$ CKM matrix which indeed has a complex phase in the $3 \times 3$ sector involving the standard model quarks even in the limit of heavy $D$ quark masses. This is an example of a breakdown of the decoupling theorem [10]. Clearly since there is only one neutral Higgs boson coupling to the effective down quark mass matrix, there is no FCNC effects at the tree level as in the case of the standard model. Clearly, if the masses of the vectorlike quarks were at the weak scale, the mixing between the light $d$ quarks and $D$ would be significant and lead to large FCNC effects at low energies.

This provides a second way to introduce spontaneous CP violation without simultaneously having flavor changing neutral current effects. Note that the common thread between the examples in Sections 3 and 4 is the fact that CP is violated spontaneously at high scale, which highlights the main point of this Letter. In the remainder of this Letter, we show how these ideas can be embedded into extended models on the way towards a possible grand unified scheme where spontaneous CP violation occurs at the GUT scale.

5. Embedding high scale SCPV into left–right symmetric models

The left–right symmetric models are based on the gauge group $SU(2)_L \times SU(2)_R \times U(1)_{B-L} \times SU(3)_c$ with fermions assigned in a left–right symmetric manner [14] and Higgs belonging to bidoublet field $\Phi(2, 2, 0)$ and a pair of fields of either $(\chi_L(2, 1, -1) \oplus \chi_R(1, 2, -1))$ type (called $\chi$-type below) or $(\Delta_L(3, 1, +2) \oplus \Delta_R(1, 3, +2))$ type (called $\Delta$-type below). The left–right symmetric models are ideally suited to embed the first class of high scale SCPV models since the bidoublet Higgs field already contains the necessary two standard model doublet Higgs fields in it. All we have to do is to embed the high scale singlet field into a left–right Higgs field. We present two different ways to do this embedding in the two subsections below.

5.1. Left–right SCPV: Model I

The first way to implement high scale SCPV is by choosing two pairs of $\chi$ type or $\Delta$-type fields. Two pairs are needed since with a single pair, constraint that $W_R$ scale must be much higher than $W_L$ scale suppresses the SCPV phase by a factor $M_{W_L}/M_{W_R}$ [15]. The two $\Delta$ type model has been discussed in [12] where at the high scale, the $\Delta_R$’s have vevs as follows: $\langle \Delta_{1,R}^0 \rangle = v_{1,R}$ and
The coupling of the form \( \text{Tr}(\Phi^\dagger \tau_2 \Phi^* \tau_2) \text{Tr}(\Delta_{1,R}^\dagger \Delta_{2,R}) \) then induces the term \( \lambda_{12} e^{i\delta} A^2 \phi_1^* \phi_2 \) at low energies and the rest of the discussion is as in Section 3 above.

Let us now turn our attention to embedding of the model of Ref. [10] into the left–right model. We consider the left–right model without the bidoublet but with the \( (\chi_L (2, 1, -1) \oplus \chi_R (1, 2, -1)) \) pair and three pairs of \( SU(2)_L \times SU(2)_R \) singlet vector-like quarks \( (P_{L,R} (1, 1, 4/3) \oplus N_{L,R} (1, 1, -2/3)) \). Such models were extensively studied in the early 90’s but not from the point of view of spontaneous CP violation [13]. We take a complex singlet Higgs field \( \sigma \) as before and assume the theory to be CP conserving prior to symmetry breaking so that all couplings in the theory are real. Again, we assume the potential for the \( \sigma \) field to be as in Eq. (11) so that its minimum corresponds to a complex vev for \( \langle \sigma \rangle = A e^{i\delta} \) as before. The vevs for the fields \( \chi_{L,R} \) are real.

To study the implications of the theory for low energy quark mixings, let us write down the quark Yukawa couplings:

\[
\mathcal{L}_Y = h^u_{ab} \bar{Q}_{L,a} \chi_L P_{R,b} + \bar{Q}_{R,a} \chi_R P_{L,b} + h^d_{ab} [\bar{Q}_{L,a} \chi_L N_{R,b} + \bar{Q}_{R,a} \chi_R N_{L,b}] + \text{h.c.}
\]

\[
+ [f^u_{ab} \sigma + f^{d,\dagger} \sigma^*] P_{L,a} P_{R,b} + [f^d_{ab} \sigma + f^{u,\dagger} \sigma^*] N_{L,a} N_{R,b} + \text{h.c.}
\]

(21)

After spontaneous symmetry breaking we have \( \langle \sigma \rangle = A e^{i\delta} \), \( \langle \chi_{L,R} \rangle = v_{L,R} \) with \( v_R \sim A \gg v_L \). This leads to the mass matrix of the form:

\[
\mathcal{M}_{uP} = \begin{pmatrix} 0 & h^u_{ab} v_L M_P \\ h^u_{ab} v_R & M_P \end{pmatrix},
\]

\[
\mathcal{M}_{dN} = \begin{pmatrix} 0 & h^d_{ab} v_L M_N \\ h^d_{ab} v_R & M_N \end{pmatrix}.
\]

Left–right symmetry requires that \( M_{P,N} = M_{P,N}^\dagger \) whereas the matrices \( h^{u,d} \) are real. After diagonalization, the effective up and down quark mass matrices become:

\[
M_{u,d} \simeq v_L v_R h^{u,d}_T M_{P,N}^{-1} h^{u,d}.
\]

(24)

These matrices are Hermitian and therefore lead to equal left and right handed CKM matrices as in the usual left–right models with bi-doublets and lead to complex CKM matrices. In fact one can write the rotation matrices for both the up and down sector as follows in a basis where the couplings \( h^{u,d} \) are diagonal:

\[
V_{u,d} = M_{P,N}^{-1/2} h^{u,d}_T U_{P,N} M_{P,N}^{-1/2} v_L v_R.
\]

(25)

Clearly since \( U_{P,N} \) is a unitary matrix with complex phases, \( V_{u,d} \) will lead to complex CKM matrix i.e. \( U_{\text{CKM}} = V_u V_d^\dagger \).

As far as the FCNC effects are concerned, they arise only in order \( m_{u,d}/M_{P,N} \) and therefore suppressed when \( A \to \) large values. Note however that the quark mixing effects arise in zeroth order of this parameter.

5.2. Left–right SCPV model II: Connecting the CP violation and seesaw scales

In this subsection, we present a more economical left–right embedding of the high scale spontaneous CP violation with suppressed FCNC. The model consists of the usual left–right assignment of the fermions [14] and Higgs system consists of a single bidoublet \( \phi (2, 2, 0) \) and the \( \chi (2, 1, -1) \oplus \chi (1, 2, -1) \). Here spontaneous CP violation is implemented via the vev of a CP and P odd real singlet scalar field \( \eta \) [16]. The CP invariant Higgs potential for the theory can be written as:

\[
V(\chi_{L,R}, \eta, \phi) = V_0(\phi) + i \mu \eta \text{Tr}(\phi_1^\dagger \phi_2) + M' \chi^\dagger_L \phi \chi_R + V_2(\eta, \chi_{L,R}),
\]

(26)

where

\[
V_0(\phi) = -\mu^2_{\phi} \text{Tr}(\phi_1^\dagger \phi_2) + \sum_{a,b,c,d} \kappa_{abcd} \text{Tr}(\phi_1^a \phi_b \phi_2^c \phi_d^c) + \kappa'_{abcd} \text{Tr}(\phi_1^a \phi_b) \text{Tr}(\phi_2^c \phi_d) + \text{h.c.}
\]

(27)

with \( (a, b) \) going over \( (1, 2) \) with \( \phi_1 = \phi \) and \( \phi_2 = \tau_2 \phi^* \tau_2 \).

\[
V_2(\eta, \chi_{L,R}) = M_0^2 \eta^2 + \lambda_\eta \eta^4 - M_X^2 (\chi^\dagger_L \chi_L + \chi^\dagger_R \chi_R) + \lambda_X (\chi^\dagger_L \chi_L + \chi^\dagger_R \chi_R)^2 + \lambda'_X (\chi^\dagger_L \chi_L - \chi^\dagger_R \chi_R)^2 + M'_0 \eta (\chi^\dagger_L \chi_L - \chi^\dagger_R \chi_R).
\]

(28)

We have assumed that under CP transformation \( \eta \to -\eta \) and \( \chi_L \to \chi^\dagger_R \) and \( \phi \to \phi^* \). Invariance under this transformation requires that all parameters in the potential be real (except for one imaginary coupling shown explicitly in the above equation).

Note now that if the term in the potential connecting \( \eta \) and \( \chi \) fields was absent, we would have \( \langle \eta \rangle = 0 \) since \( M_0^2 > 0 \). However as soon as \( SU(2)_R \) symmetry is broken by \( \langle \chi_R^0 \rangle \neq 0 \), the \( M_0^2 \) term in the potential introduces a tadpole term for \( \eta \) thereby generating

\[
\langle \eta \rangle \simeq \frac{+M_0^2 v_\eta^2}{2M_0^2}.
\]

(29)
Since $\eta$ is CP odd, this breaks CP spontaneously. The way it manifests is that the $i\mu(\eta) \text{Tr}(\phi_1^\dagger \phi_2)$ term now combines with the $\mu_{\eta}^2 \text{Tr}\phi_1^\dagger \phi_2$ to generate at low energies an effective soft CP breaking term as in Eq. (13) where $\phi_{1,2}$ are the two doublets contained in the bidoublet $\phi$ of the left–right model. The same arguments as in the Appendix B then guarantee that in this model the FCNC can be suppressed by making one of the left–right Higgs doublets superheavy.

This can also be seen in an alternative manner by minimizing the potential, noting that there is a range of values of the parameters in the potential for which we have $\langle \chi_R^0 \rangle = v_R \neq 0; \langle \eta \rangle \neq 0; \langle \chi_L \rangle = 0$ provided $M^2(\eta) > 2\lambda^2 v^2$. The vevs of $\chi_R$ and $\eta$ fields are much larger than the weak scale.

An interesting point worth stressing is that in this model, the scale of CP violation and the seesaw [17] scale for neutrino masses are connected. To see this, note that the right handed neutrino masses come from the higher dimensional term $(L_R \bar{\chi}_R)^2/M_P$ leading to seesaw right handed neutrino masses given by $M_{\text{seesaw}} \simeq v_R^2 M_P$ and from Eq. (29), the CP violating scale $\langle \eta \rangle$ and $M_{\text{seesaw}}$ owe their origin to the same scale $v_R$ i.e. violation of parity. Since in grand unified theories, $v_R$ can be identified with the GUT scale, one would therefore relate several scales of the theory i.e. $M_{\text{SCPV}}, M_{\text{seesaw}}$ and $M_{\text{GUT}}$.

6. Possible extensions to supersymmetry and SUSY CP problem

As is well known, generic minimal supersymmetric extensions of the standard model (MSSM) are plagued with the SUSY CP problem. There have been many solutions suggested to solve this problem [18]. A simple solution to this problem would of course be to have CP spontaneously broken. However, in MSSM, CP cannot be spontaneously broken. Furthermore, it has also been pointed out that [19] it is particularly hard to have spontaneous CP breaking by considering multi-Higgs generalizations of the MSSM. A possibility for achieving spontaneous CP breaking within SUSY involves the introduction of singlet chiral fields [20]. As far as the FCNC effects are concerned, in these models one may fine tune the $\mu$ terms to make of the extra Higgs doublets heavy thereby eliminating large FCNC effects. However, the early versions of these models are no longer viable since they had a real CKM matrix, in contradiction with recent experimental data.

Therefore, the ideas described in this Letter may be particularly useful if one wants to solve the SUSY CP problem by spontaneous CP violation in a viable scenario, where vacuum phases do lead to a complex CKM matrix, while at the same time suppressing FCNC effects. In fact, recently it has been suggested one such model which includes two singlet Higgs superfields and adds an extra vector like singlet fermion to MSSM [21] to break CP spontaneously and generate a complex CKM matrix. One can embed this scheme into the SUSY left–right model. Detailed analysis of SUSY models that exploit the ideas of this Letter is under way and will be taken up in a forthcoming publication.

7. Conclusion

We have emphasized the close connection between spontaneous CP violation and FCNC effects in theories where CP breaking vev is at the weak scale. We have also shown that in order to avoid FCNC effects while at the same time generating a complex CKM matrix through vacuum phases, one is naturally led to have spontaneous CP breaking at a high energy scale, well above the electroweak scale. We then describe two classes of models one without and one with extra heavy fermions where having a high vev break CP spontaneously leads to complex CKM matrix as given by experiment without simultaneously having large FCNC effects. We then show how these models can be embedded into the high scale left–right models where parity violation and neutrino mass are connected via the seesaw mechanism. We find one particular model where spontaneous parity violation triggers the spontaneous CP violation thus connecting the three scales: Seesaw scale for neutrino masses, scale of spontaneous parity and CP violation.

In conclusion, if our view on the origin of CP violation is correct, then small neutrino masses and CP violation at low energies would have in common the fact that they are both manifestations of physics occuring at very high energy scale.

Acknowledgements

The work of R.N.M. is supported by the National Science Foundation grant No. Phy-0354401 and the work of G.C.B. is supported by Fundacao para a Ciencia e a Tecnologia (FCT, Portugal), through the projects POCTI/FNU/44409/2002, PDCT/FP/FNU/50250/2003, POCI/FP/63415/2005, POCTI/FP/FNU/50167/2003, which are partially funded through POCTI (FEDER). Both the authors are very grateful for the Alexander von Humboldt Senior Research Award which made this collaboration possible. They are also grateful to A. Buras and M. Lindner at TUM and R.N.M. to H. Fritzsch at LMU for kind hospitality when the work was done.
Appendix A

In this appendix, we elaborate on the connection between SCPV and FCNC and complex CKM in a two Higgs doublet model. For this purpose, we write the Yukawa Lagrangian as:

$$\mathcal{L}_Y = \sum_{a,b} (h_{ab}^{\nu} \tilde{Q}^0_{La} \phi_i u^0_{R,b} + u \rightarrow d) + \text{h.c.}$$  \hfill (30)

It can be readily seen \([5,22]\) that in the quark mass eigenstate basis, the scalar coupling can be written as:

$$\mathcal{L}_{\text{scalar}} = [\bar{u} D_u u + \bar{d} D_d d]^T \frac{H}{v} \left[ \bar{u} (N_u P_R + N_u^T P_L) u + \bar{d} (N_d P_R + N_d^T P_L) d \right] \frac{R}{v}$$

$$+ i \left[ \bar{u} (N_u P_R - N_u^T P_L) u - \bar{d} (N_d P_R - N_d^T P_L) d \right] \frac{I}{v},$$  \hfill (31)

where

$$H = \frac{1}{v} [v_1 R_1 + v_2 R_2], \quad R = \frac{1}{v} [v_2 R_1 - v_1 R_2], \quad I = \frac{1}{v} [v_2 I_1 - v_1 I_2]$$  \hfill (32)

with \(\phi_i^0 = \frac{1}{\sqrt{2}} [v_1 + R_1 + i I_1], \phi_2^0 = \frac{1}{\sqrt{2}} e^{i \delta} [v_2 + R_2 + i I_2], \) where

$$N_d = U^T_{dL} \left[ \frac{v_2}{\sqrt{2}} Y_d^T - \frac{v_1}{\sqrt{2}} e^{i \delta} Y_2^T \right] U_{dR},$$  \hfill (33)

where \(U_{dL,R}\) are the unitary matrices which diagonalize the down quark mass matrix \(M_d\). Analogous expressions are there for \(N_u\).

It is clear that \(N_{d,u}\) are in general not diagonal and therefore \(R\) and \(I\) mediate FCNC.

The quark mass matrices are in the form

$$M_d M_d^T = H_{\text{real}} + 2 i v_1 v_2 \sin \delta (Y^{d1} Y^{d1T} - Y^{d2} Y^{d2T}),$$  \hfill (34)

where \(H_{\text{real}}\) is a symmetric real matrix. It is clear that \(M_d M_d^T\) (and similarly \(M_u M_u^T\)) is an arbitrary complex matrix and therefore CKM is a complex matrix.

If one fine tunes such that \(Y_1^{d1} \propto Y_2^{d1}, N_d\) would be diagonal and FCNC would be eliminated. But in that case, Eq. (34) implies that \(M_d M_d^T\) becomes real. This illustrates the connection between FCNC and the possibility of generating a complex CKM by a vacuum phase.

Appendix B

In this appendix, we discuss how the extra neutral Higgs fields in the model of Section 3 that are potential mediators of FCNC effects can be made heavy while at the same time the SM Higgs can be kept light by one fine tuning. We will work with the potential in Eqs. (13), (14). Clearly, the minimum of this potential corresponds to:

$$\langle \phi_1 \rangle = \left( \begin{array}{c} 0 \\ v_1 \end{array} \right), \quad \langle \phi_2 \rangle = \left( \begin{array}{c} 0 \\ v_2 e^{i \delta} \end{array} \right).$$  \hfill (35)

Let us work in a basis in which

$$\begin{pmatrix} H_1 \\ H_2 \end{pmatrix} = \frac{1}{v} \begin{pmatrix} v_1 & v_2 \\ v_2 & -v_1 \end{pmatrix} \begin{pmatrix} \phi_1 \\ e^{-i \delta} \phi_2 \end{pmatrix}. $$  \hfill (36)

The potential in Eq. (13) looks as follows:

$$V(H_{1,2}) = \Lambda^2 (\lambda_{11} H_1^1 H_1 + \lambda_{22} H_2^1 H_2 + (\lambda_{12} H_1^1 H_2 + \text{h.c.})) + \lambda_1 (H_1^1 H_1)^2 + \lambda_2 (H_2^1 H_2)^2$$

$$+ \lambda_3 (H_1^1 H_1)(H_2^1 H_2) + \lambda_4 (H_1^1 H_1)(H_2^1 H_2) + [\lambda_5 H_1^1 H_2 + \lambda_6 H_1^1 H_1 + \lambda_7 H_2^1 H_2] H_1^1 H_2 + \text{h.c.}$$  \hfill (37)

Even though we use the same \(\lambda\)’s in both Eq. (13) and here, they are different and in fact now \(\lambda_{12}, \lambda_{5,6,7}\) are in general complex while the other \(\lambda\)’s are real.

Now we can write the \(H_{1,2}\) in terms of their components:

$$H_1 = \left( \frac{1}{\sqrt{2}} (v + H + i G) \right), \quad H_2 = \left( \frac{C^{+}}{\sqrt{2}} \right).$$  \hfill (38)
As already discussed in [3], the stability of vacuum demands that, the coefficients of the linear terms in \((H, S, P)\) vanish and gives
\[
A^2 \lambda_{11} + 2 \lambda_1 v^2 = 0, \quad A^2 \lambda_{12} + \lambda_6 v^2 = 0. \tag{39}
\]
These are the fine tuning conditions in the \((H_1, z)\) basis to have SM Higgs field light and have the correct electroweak symmetry breaking. We can now write down the mass matrix for the other neutral Higgs fields \((H, S, P)\) as follows [3]:
\[
M_{H,S,P} = \begin{pmatrix}
4v^2 \lambda_1 & 2v^2 \text{Re} \lambda_6 & -2v^2 \text{Im} \lambda_6 \\
2v^2 \text{Re} \lambda_6 & \lambda_2 A^2 + (\lambda_3 + \lambda_4 + 2 \text{Re} \lambda_5) v^2 & -2v^2 \text{Im} \lambda_5 \\
-2v^2 \text{Im} \lambda_6 & -2v^2 \text{Im} \lambda_5 & \lambda_2 A^2 + \lambda_3 v^2 + (\lambda_4 - 2 \text{Re} \lambda_5) v^2
\end{pmatrix} \tag{40}
\]
From this expression, we can explicitly see that beyond the standard model neutral Higgs particles \((S, P)\) have masses of order \(\Lambda\) whereas the SM Higgs field has mass of order of the electroweak scale. Also the mixings of the SM Higgs which can generate FCNC effects are of order \(v^2/\Lambda^2\) and hence very small as \((S, P)\) are made heavy. Also \(\lambda_2 A^2 + \lambda_3 v^2\) gives the mass of the charged Higgs field from the second Higgs field \(H_2\). Thus we have complex CKM from SCPV while at the same time suppressing the FCNC effects.

References

B. Aubert, et al., BaBar Collaboration, hep-ex/0507101.


T. Yanagida, in: O. Sawada, A. Sugamoto (Eds.), Proceedings of the Workshop on the Unified Theory and the Baryon Number in the Universe, KEK, Tsukuba, Japan, 1979, p. 95;


B. Aubert, et al., BaBar Collaboration, hep-ex/0507101.
Eluding the BBN constraints on the stable gravitino

W. Buchmüller a, K. Hamaguchi a,b, M. Ibe b,*, T.T. Yanagida a,b

a Deutsches Elektronen-Synchrotron DESY, 22603 Hamburg, Germany
b Department of Physics, University of Tokyo, Tokyo 113-0033, Japan

Received 15 July 2006; received in revised form 24 July 2006; accepted 25 July 2006
Available online 22 August 2006
Editor: M. Cvetic

Abstract

We investigate how late-time entropy production weakens the big-bang nucleosynthesis (BBN) constraints on the gravitino as lightest superparticle with a charged slepton as next-to-lightest superparticle. We find that with a moderate amount of entropy production, the BBN constraints can be eluded for most of the parameter space relevant for the discovery of the gravitino. This is encouraging for experimental tests of supergravity at LHC and ILC.

© 2006 Elsevier B.V. All rights reserved.

1. Introduction

The gravitino $\tilde{G}$ is a unique and inevitable prediction of supergravity (SUGRA) [1], and hence the discovery of the gravitino would provide unequivocal evidence for SUGRA. It has been pointed out that this test of SUGRA may be possible at LHC or ILC, if the gravitino is the lightest superparticle (LSP) and the long-lived next-to-lightest superparticle (NLSP) is a charged slepton [2].

From an experimental point of view, a relatively large gravitino mass $m_{\tilde{G}}$ comparable to the slepton mass $m_{\tilde{\ell}}$, $m_{\tilde{G}} \gtrsim \mathcal{O}(0.1)m_{\tilde{\ell}}$, is particularly interesting [2,3]. This is because in such a gravitino mass region the kinematical reconstruction of the gravitino mass becomes possible, which leads to a determination of the “Planck scale”, and even the gravitino spin might become measurable.

However, such a parameter region is strongly constrained by cosmology. In particular, the BBN constraints on a late decaying particle [4,5] lead to an upper bound on the gravitino mass for a given slepton mass [6,7], which makes the SUGRA test at collider experiments very challenging.

It is, however, easy to evade the BBN constraints if late-time entropy production occurs after the slepton decoupling (and before BBN). In this Letter we explicitly show how much late-time entropy production weakens the BBN constraints on the NLSP decay into the gravitino. We find most of the relevant parameter space to survive for a moderate amount of entropy production. This is very encouraging with respect to experimental tests of SUGRA at LHC and ILC. It has also interesting implications for leptogenesis, which will be discussed elsewhere [8].

2. BBN constraint with late-time entropy production

For concreteness, we assume that the NLSP is the superpartner of the tau lepton, stau ($\tilde{\tau}$). In the early universe, the stau NLSP is in thermal equilibrium until its decoupling at $T_{d} \sim m_{\tilde{\tau}}/20$. If the stau particle decays during or after BBN, $T_{BBN} \sim 1$ MeV, it may spoil the successful BBN predictions [4,5]. In the model with stau NLSP and gravitino LSP, this leads to severe constraints on the parameter space ($m_{\tilde{\ell}}, m_{\tilde{G}}$), in particular to upper bounds on the gravitino mass for a given stau mass [6,7].

If there is no entropy production after the stau decoupling, the thermal relic abundance of the stau before its decay is given by [6]

$$Y_{\tilde{\tau}}^{\text{thermal}} \equiv \frac{n_{\tilde{\tau}}}{s} = \kappa \times 10^{-13} \left( \frac{m_{\tilde{\tau}}}{100 \text{ GeV}} \right),$$

(1)
where \( n_\tau \) and \( s \) are the stau number density and the entropy density, respectively. Here, \( \kappa = (0.7-1) \) is a numerical coefficient which depends on the model parameters (e.g., \( \tan \beta \)). In the following we take \( \kappa = 0.7 \) as a representative value. The decay rate of the stau NLSP is given by \([2]\)

\[
\Gamma_\tilde{\tau} (\tilde{\tau} \rightarrow G \tau) = \frac{m_\tilde{\tau}^5}{48\pi m_\tilde{G}^2 M_P^2} \left(1 - \frac{m_\tilde{G}^2 + m_\tau^2}{m_\tilde{\tau}^2}\right)^4 \times \left[1 - \frac{4m_\tilde{G}^2 m_\tau^2}{(m_\tilde{\tau}^2 - m_\tilde{G}^2 - m_\tau^2)^2}\right]^{3/2},
\]

\[
\simeq (6 \times 10^6 s)^{-1} \left(\frac{m_\tilde{\tau}}{100 \text{ GeV}}\right)^5 \left(\frac{10 \text{ GeV}}{m_\tilde{G}}\right)^2 \left(1 - \frac{m_\tilde{G}^2}{m_\tilde{\tau}^2}\right)^4,
\]

where in the second equation we have neglected the mass of the tau-lepton, \( m_\tau \).

The energetic tau-lepton produced by the stau decay causes the electromagnetic (EM) cascade, which results in destructions or overproductions of light elements (D, \(^3\)He, \(^4\)He, etc.). However, the tau-lepton itself decays before interacting with background photons, and hence, some of the energy carried by the tau-lepton is lost to neutrinos. Thus, the electromagnetic energy released by the stau decay is given by

\[
E_\text{EM} Y_\tilde{\tau} = \xi_\text{EM} \left(\frac{m_\tilde{G}^2 - m_\tau^2}{2m_\tilde{\tau}}\right) Y_\tilde{\tau},
\]

where \( \xi_\text{EM} \lesssim 1 \) denotes the suppression factor due to the energy loss in the tau decay into neutrinos. In the following, we take \( \xi_\text{EM} \approx 0.5 \) as a representative value \([7]\). On the other hand, the hadronic contribution relevant for BBN dominantly comes from the three- and four-body decay of the stau, and therefore the hadronic branching ratio of the stau NLSP is suppressed, \( B_h \lesssim 10^{-3} \) \([9]\). Thus, the hadronic energy released by the stau decay is suppressed by a factor \( \lesssim 10^{-3} \) compared to the EM energy.

Without late-time entropy production, the parameter region with a relatively large gravitino mass, \( m_\tilde{G} \gtrsim \mathcal{O}(0.1)m_\tilde{\tau} \), which is wanted for tests of SUGRA, is severely constrained by the above BBN constraints. However, if an adequate entropy production occurs after the stau decoupling, the stau abundance is diluted by a factor \( \Delta \),

\[
Y_\tilde{\tau} = \frac{1}{\Delta} Y_\text{thermal},
\]

and the BBN constraints can be easily eluded.

For example, we can consider the entropy produced by the late-time decay of a long-lived particle \( \varphi \). In this case, the dilution factor \( \Delta \) is given by \( \Delta \simeq \left( T_d/T_\varphi \right)^3 \) in terms of the decay temperature of \( \varphi \), \( T_\varphi \) (assuming that the direct production of \( \tilde{\tau} \) from \( \varphi \) is negligible). Thus, for instance, the dilution factor is \( \Delta \sim 10^3 \) for \( T_\varphi \sim 1 \text{ GeV} \) and \( m_\tilde{\tau} \sim 200 \text{ GeV} \).

In Fig. 1, we plot the BBN constraints in the \((m_\tilde{G}, m_\tilde{\tau})\) plane for a given dilution factor. Here, we have used the constraint from the \(^3\)He/D bound, and neglected other photo-dissociation and hadro-dissociation effects on the light elements. This is because, in the region \( m_\tilde{\tau} Y_\tilde{\tau} \lesssim 10^{-11} \text{ GeV} \), the most stringent constraint comes from the \(^3\)He overproduction for a late decaying particle with \( B_h \lesssim 10^{-3} \) (see Figs. 41 and 42 in Ref. \([4]\)).

Thus, for our purposes, the \(^3\)He/D bound is the most stringent BBN constraint for \( m_\tilde{\tau} \lesssim 500 \text{ GeV} \) and \( \Delta \gtrsim 10^3 \).

As Fig. 1 demonstrates, the \(^3\)He/D bound severely constrains the parameter space for \( \Delta = 1 \), while the most of the interesting region is allowed for \( \Delta = 10^3 \). Therefore, the bulk of the relevant parameter space for the SUGRA test survives with a moderate amount of entropy production. Consequences of the present analysis for leptogenesis will be discussed in Ref. \([8]\).

Acknowledgements

M.I. thanks the Japan Society for the Promotion of Science for financial support. This work is partially supported by Grand-in-Aid Scientific Research (s) 14102004. The work of T.T.Y. has been supported in part by a Humboldt Research Award.

References


---

1 In Fig. 1, we have used the constraints in Fig. 41 of Ref. [4]. Our \( \Delta = 1 \) line in Fig. 1 almost reproduces the \(^3\)He/D bound in Fig. 3 of Ref. [7].
Constraints on hybrid inflation from flat directions in supersymmetry

Fuqiang Xu \textsuperscript{a}, Jin Min Yang \textsuperscript{b,a,*}

\textsuperscript{a} Institute of Theoretical Physics, Academia Sinica, Beijing 100080, China
\textsuperscript{b} CCAST (World Laboratory), P.O.Box 8730, Beijing 100080, China

Received 30 August 2006; received in revised form 5 October 2006; accepted 16 October 2006
Available online 30 October 2006
Editor: M. Cvetič

Abstract

We examine the constraints on F-term hybrid inflation by considering the flat directions in the Minimal Supersymmetric Standard Model (MSSM). We find that some coupling terms between the flat direction fields and the field which dominates the energy density during inflation are quite dangerous and can cause the no-exit of hybrid inflation even if their coupling strength is suppressed by Planck scale. Such couplings must be forbidden by imposing some symmetry for a successful F-term hybrid inflation. At the same time, we find that in the D-term inflation these couplings can be avoided naturally. Further, given the tachyonic preheating, we discuss the feasibility of Affleck–Dine baryogenesis after the F-term and D-term inflations.

© 2006 Elsevier B.V. All rights reserved.

1. Introduction

Hybrid inflation is now the standard paradigm of inflation and has proved very fruitful since it was introduced [1]. In the models of hybrid inflation, most of the energy density is provided by an auxiliary field $\sigma$ instead of the slowly rolling inflaton field $\Phi$. When $\sigma$ falls below a critical value $\sigma_c$, the inflation ends, which gives a natural exit of inflation. But when a particle physics model is tried for hybrid inflation, e.g., F-term hybrid inflation, there always exist a mass term of inflaton which can be of the same order as the Hubble scale and destroy the slow-rolling condition (the so-called $\eta$ problem). So far various approaches have been proposed to avoid this $\eta$ problem [2].

The flat directions, abundant in the MSSM, have many distinctive features and may thus play an important role in the early universe. Such flat directions are protected from perturbative quantum corrections and only lifted by the soft breaking terms which are insignificant in the early universe. During the inflation these flat directions can get large vacuum expectation values (vev) which naturally make an initial condition for many phenomena. From the potential of a flat direction $\varphi$ during inflation [11]

$$V(\varphi) = (m_0^2 + c_H H_I^2)|\varphi|^2 + \frac{A\lambda H_I|\varphi|^n e^{i\theta_0}}{n M^{n-3}} + \text{h.c.} + \lambda^2 |\varphi|^{2(n-1)} M^{2(n-3)},$$

we get the large vevs

$$\varphi_0 \sim (H_I M^{-3} / \lambda)^{1/n-2}.$$  

Here $n$ is an integer ($\geq 4$), $H_I$ is the Hubble parameter during inflation, $m_0$ is the soft mass term, $c_H (< 0)$ and $A$ are constants of $O(1)$, and $M$ is usually the Planck scale.\textsuperscript{1} Note that the large vevs of the flat directions can also cause some problems in various inflation models [3,4], especially they can kinematically block the resonance preheating in chaotic inflation due to the invalidation of the non-adiabaticity condition [5].

As noticed in [5], the large vevs of the flat directions can also cause no exit of hybrid inflation. In this work we will scrutinize this problem and find that such a problem can be caused

\textsuperscript{1} In fact if $M$ is the GUT or smaller scale, the problem of no preheating in chaotic inflation (mentioned below) will be alleviated. The same thing will happen in hybrid inflation considered in our work below. But to be conservative, we consider the worst case where $M$ is taken as the Planck scale.
by some coupling terms between the flat direction fields and the field which dominates the energy density during inflation. Such couplings must be forbidden by imposing some symmetry for a successful F-term hybrid inflation. Then we point out that in the D-term inflation these couplings can be avoided naturally and thus the problem of no-exit of hybrid inflation does not exist, and further the no-preheating problem can also be avoided.

We will also discuss the feasibility of Affleck–Dine baryogenesis [12] for both F-term and D-term inflations, given the tachyonic preheating. It is well known that after chaotic inflation ends, the process of resonance preheating occur at once [6]. This preheating leads to the large oscillation of the fields which couples with inflaton, and can make the symmetries restored for a moment which can make the baryon number produced after chaotic inflation [7]. Noticing that similar processes may occur during tachyonic preheating [15] after hybrid inflation, we find that it is possible the baryon number produced in these phases is copious due to the large amplitude of inflaton.

2. Flat directions constraints in F-term hybrid inflation

So far in the literature the couplings of inflaton to matter fields have not been intensively studied and only some toy models have been considered which have no relevance to SM particles [6,16]. In [17] the importance of gauge invariance was first highlighted and recently the strength of the couplings of inflaton to matter fields was studied [5]. In the following we will examine the couplings between the MSSM fields and the auxiliary field dominating the energy density where the gauge symmetry is also important.

In the simplest supersymmetric hybrid inflation model, the superpotential contains the terms

\[ W \supset g\hat{\phi}(\hat{\sigma}^2 - \sigma_0^2), \]

where \( g \) is a coupling constant, \( \hat{\phi} \) is the inflaton superfield and \( \hat{\sigma} \) is the superfield containing the scalar field \( \sigma \). Large vevs of flat directions can induce a mass \( \lambda_i \phi_0 \) for the \( \sigma \) field if we assume the existence of the \( \lambda_i^2|\phi|^2|\sigma|^2 \) interaction with \( \lambda_i \) being a coupling constant (such a term is assumed to have the same sign as the interaction terms from the superpotential \( W \)). Then if

\[ \lambda_i \phi_0 > \sqrt{2}g\sigma_0, \]

the mass-square of \( \sigma \) will remain positive even for \( \langle \phi \rangle < \phi_c \) (\( \phi_c \) is the critical value at which the hybrid inflation naturally exits), which leads to \( \langle \sigma \rangle = 0 \) and no tachyonic preheating and no exit from hybrid inflation.

In fact, the above problem always exists in hybrid inflation if \( \lambda_1 > \sqrt{2}g \), which can be seen clearly by replacing

\[ H_1 = \sqrt{V_0/3M_p^2}, \]

with \( V_0 = g^2\sigma_0^4 \) into Eq. (2). We find that if we ignore the coupling constants in Eq. (1), then we have \( \phi_0 \geq \sigma_0 \), where \( \sigma_0 \) decides the dominant energy density in hybrid inflation even for \( n = 4 \).

Then a natural idea for solving this problem is to suppress the dangerous large coupling \( \lambda_1 \) to invalidate Eq. (4). But it is difficult to find out a way in generic hybrid inflation. For example, let us consider the matter fields contained in the MSSM flat directions besides the two scalar fields \( \sigma \) and \( \phi \) in hybrid inflation. With the addition of \( \sigma \) field, the superpotential is obtained by multiplying the \( \sigma \) superfield with the following MSSM gauge invariant terms

\[ H_u H_d, \quad H_u L, \quad (5) \]

and

\[ H_u Q_u, \quad H_d L_e, \quad H_d L_e, \quad Q L_d, \quad u_d d, \quad L L_e, \quad (6) \]

where \( H_u \) and \( H_d \) are the two Higgs doublets, \( L \) and \( Q \) denote respectively a doublet of lepton and quark, and \( e, u \) and \( d \) denote respectively a singlet of lepton, up-quark and down-quark. Here we drop out the higher order gauge-invariant terms, which are negligible, as shown in our following analysis. We will label a generic MSSM flat direction by \( \varphi \). Examples of such flat directions are given by \( L H_u \sim \phi^3 \) in terms of doublet components \( L = (\varphi, 0) \) and \( H_u = (0, \varphi) \), and by \( Q L_d, d_2 \sim \phi^3 \). Then we obtain interactions like

\[ \lambda_1^2|\varphi|^2 \sigma^2 + \lambda_2^2|\varphi|^4 \sigma^2/M_p^2, \]

where \( \lambda_i \) (\( i = 1, 2 \)) are coupling constants.

Let us consider the renormalizable terms, i.e., \( \sigma H_u H_d \) and \( \sigma H_u L \). It is natural to assume their coupling constants to be \( \mathcal{O}(1) \). On the other hand, we know that these gauge invariant terms correspond to the D-flat directions of the MSSM [8]. These terms \( \sigma H_u H_d \) and \( \sigma H_u L \) can lead to \( \lambda_{1,2} \sim \mathcal{O}(1) \). Since \( \phi_0 \geq \sigma_0 \) even for \( n = 4 \) (D-flat \( H_u H_d \) and \( H_u L \) directions are lifted by \( n = 4 \) non-renormalizable terms), it is impossible to avoid Eq. (6) unless \( g \gg \mathcal{O}(1) \).

Actually, the coupling terms in Eq. (7) with even a small \( \lambda \) cannot be neglected because the vevs of the flat directions which are lifted by a large \( n \) (\( n \) can be as large as 9 [8]) are much larger than \( \sigma_0 \). From some calculations we find that there are two other dangerous flat directions: \( L L_e \) and \( u_d d \), which are lifted up by \( n = 6 \) non-renormalizable terms.

Some global symmetry\(^2\) like R-symmetry \( U(1)_R \) [18] must be imposed to forbid the 4 dangerous terms \( H_u H_d, H_u L, L L_e \) and \( u_d d \) in Eq. (6). If we use R-symmetry, we must make special assignment for the R-charges of involved superfields, e.g., the R-charge must be zero for \( \sigma \) field as can be seen from Eq. (3).

Now we discuss the feasibility of Affleck–Dine baryogenesis during tachyonic preheating after the F-term hybrid inflation. After the inflation the A-term in the potential in Eq. (1) becomes small as the Hubble parameter \( H \) decreases, and there will appear a CP-violating interaction [7]

\[ c(\phi^2/M_p^{n-2})\phi^n + \text{h.c.} \]

\[ (8) \]

\(^2\) Note that unlike a discrete symmetry [20], such a global symmetry may have a potential problem since it may be violated by quantum gravity effects [19].
During tachyonic preheating, the large oscillation of inflaton $\phi$ can make the flat directions go to zero and
\[ \phi^2 \sim (\phi^2) e^{2\theta_\phi}. \] (9)

Through Affleck–Dine mechanism, this process will produce baryon number
\[ n_B \sim 2|\theta_\phi|^2. \] (10)

The initial value of $\theta_\phi$ is determined by the $H_3$-dependent $A$-term in Eq. (1). Such baryon number will be released in the ensuing decay of $\phi$. Because $\theta_\phi$ is so large, this baryon number generated during the tachyonic preheating after hybrid inflation cannot be neglected. This is quite similar to the case of resonance preheating after chaotic inflation considered in [7].

3. D-term inflation and its consequence

In the preceding section we have shown that it is not easy to obtain a negative mass-square term due to the large vevs of flat directions, which will lead to no elegant exit of hybrid inflation in the F-term inflation. Here we show that such a problem can be solved naturally in D-term inflation.

D-term inflation can preserve the flat directions of global supersymmetry and, in particular, keep the inflation potential flat provided that one of the contributions to the potential $V_D$ contains a Fayet–Iliopoulos term [13] as in Eq. (12). This was first pointed out in [10] and significantly improved in [9].

Consider a toy model of D-term inflation, whose field content includes a inflaton chiral superfield $\phi$ and auxiliary superfields $\sigma_\pm$ with charges $\pm 1$ under an anomalous $U(1)_X$ symmetry. From the superpotential
\[ W = c\phi\sigma_+\sigma_- , \] (11)

and the minimal Kahler potential, we obtain the tree-level scalar potential
\[ V = |c|^2(\langle \sigma_+\sigma_- \rangle^2 + |\phi\sigma_+|^2 + |\phi\sigma_-|^2) \]
\[ + \frac{\xi^2}{2}(\langle \sigma_+^2 - |\sigma_-|^2 + \xi^2 \rangle^2) . \] (12)

Here $c$ is a coupling constant and the Fayet–Iliopoulos term $\xi^2$ is assumed to be positive. The role of $\sigma$ field in Eq. (3) is now replaced by the $\sigma_-$ field in this model.

Now we argue that in the D-term inflation the no-exit problem happened in generic F-term inflation discussed in the preceding section does not exist. The magnitude of the vevs of the MSSM flat directions are primarily fixed by higher dimension operators in the Kahler potential that couple the flat direction to other fields like
\[ \Delta L = \int d^4\theta \left( c_1|\phi|^2 + c_2|\sigma_+|^2 + c_3|\sigma_-|^2 \right) \frac{\phi^2 + \phi^2}{M^2} . \] (13)

where $c_i$ $(i = 1, 2, 3)$ are coupling constants. These induced mass terms for flat direction $\phi$ always vanish because $\langle \sigma_- \rangle = 0$ and $\langle \sigma_+ \rangle = 0$ during D-term inflation. Here the zero vev of $\sigma_\pm$ can be obtained by minimizing the potential of Eq. (12) when $\phi > \phi_c = g\xi/c$ with $\phi_c$ being the critical value at which inflation ends. Then the flat directions can only get vev after inflation where $U(1)_X$ is broken and we get
\[ \langle \sigma_+ \rangle = 0, \quad \langle \sigma_- \rangle \sim \xi, \quad \langle F_{\sigma_\pm} \rangle \sim \xi \phi . \] (14)

After the quick tachyonic preheating, the inflaton $\phi$ fall down to its global minimum and equal to zero. So the $\langle F_{\sigma_\pm} \rangle \sim 0$, but there always exist the kinetic energies in Eq. (13) which will lead to the negative mass-square term of flat directions fields. Of course, such delayed appearance of the vevs of flat directions will not cause the problem of no-exit of inflation.

On the other hand, as pointed in [5], the large vev of flat directions will prohibit the tachyonic preheating in F-term inflation. However, we should point out that this problem does not happen in D-term inflation because the direct coupling terms between flat directions and $\sigma_\pm$ are forbidden by $U(1)_X$ gauge symmetry. Then the preheating will proceed despite of the large vevs of the flat directions. The ensuing Affleck–Dine baryogenesis can occur as indicated in [14] after tachyonic preheating, but the decays of inflaton and $\sigma_\pm$ must be quite different with those in no-preheating case.

4. Conclusion

We discussed the effects of large vevs of MSSM flat directions in hybrid inflation in detail. We have shown that some dangerous coupling terms between the flat direction fields and the field which dominate the energy density during inflation must be forbidden for a successful inflation even if their coupling constants are small. Such couplings must be forbidden by imposing some symmetry for a successful F-term hybrid inflation. At the same time, we found that in the D-term inflation these couplings can be avoided naturally. Further, given the tachyonic preheating, we discussed the feasibility of Affleck–Dine baryogenesis after the F-term and D-term inflations.

Acknowledgements

This work was supported in part by National Natural Science Foundation of China (NNSFC) under number No. 10475107.

References


A note on the Cardy–Verlinde formula

Boliang Lin\textsuperscript{a}, Rong-Gen Cai\textsuperscript{b,}\textsuperscript{*}

\textsuperscript{a} State Key Lab of Rail Control and Safety, Beijing Jiaotong University, Beijing 100044, China
\textsuperscript{b} Institute of Theoretical Physics, Chinese Academy of Sciences, PO Box 2735, Beijing 100080, China

Received 24 July 2006; received in revised form 20 October 2006; accepted 20 October 2006
Available online 7 November 2006

Editor: T. Yanagida

Abstract

We study static, spherically symmetric, and asymptotically AdS black hole spacetimes with matter distribution satisfying $T_{tt}=T_{rr}$. The first law of black hole thermodynamics is checked. We show that the Cardy–Verlinde formula, which is supposed to be an entropy formula of CFTs in any dimensions, is just a rewritten form of Einstein equations on the black hole horizon. Further, the entropy of dual CFT to the black hole solution is found to indeed satisfy the Cardy–Verlinde formula and the energy of matter outside the black hole should be subtracted from the total energy of the dual CFT in the Cardy–Verlinde formula. The implications of our results are discussed.

© 2006 Elsevier B.V. All rights reserved.

A remarkable property of black hole is that black hole is not ‘black’, but emits radiation like a black body [1]. The Hawking temperature is determined by surface gravity of black hole and black hole entropy is proportional to horizon area [2]. It is very interesting to notice that these thermodynamic properties of black hole are completely fixed by geometric quantities like surface gravity and horizon area, while the latters are governed by Einstein equations.

Since black hole was found to be of thermodynamic properties like an ordinary thermodynamic system, people have been speculating that there are some relations between thermodynamics and Einstein equations (determining the dynamics of spacetime). Indeed, based on black hole thermodynamics, ’t Hooft [3] and Susskind [4] proposed the holographic principle of gravity, which says that independent degrees of freedom of a gravitational system are proportional to its surface area, not its volume. In particular, using the first law of thermodynamics to all Rindler causal horizons through each spacetime point, Jacobson [5] was indeed able to derive Einstein equations. Applying the first law of thermodynamics to apparent horizon of Friedmann–Robertson–Walker (FRW) universe, we can derive the Friedmann equations governing the dynamics of the universe [6]. Indeed more and more evidence has been accumulated, which shows the close relation between thermodynamics and Einstein equations, see, for instance, Ref. [7] and references therein.

On the other hand, Verlinde [8] found that for a radiation dominated FRW universe, the Friedmann equation can be rewritten to the so-called cosmological Cardy formula, a same form as the Cardy–Verlinde formula, the latter is an entropy formula for a conformal field theory in a higher-dimensional spacetime. Note that the radiation can be described by a conformal field theory. Therefore, the entropy formula describing the thermodynamics of radiation in the universe has the same form as that of the Friedmann equation, which describes the dynamics of spacetime. In particular, when the so-called Hubble entropy bound is saturated, these two equations coincide with each other [9] (for a more or less complete list of references on this topic see, for example, [10]; see also [11] for a discussion of the Cardy formula in a cosmological setup with cosmological constant). Therefore, Verlinde’s observation further indicates some relation between thermodynamics and Einstein equations. More concretely, it reveals the relation between the entropy formula of CFTs in the FRW universe and Friedmann equations describing the evolution of the universe.

Suppose there is a CFT residing on an $n$-dimensional sphere with radius $R$. 

---

\textsuperscript{*} Corresponding author.
E-mail addresses: bllin@center.bjtu.edu.cn (B. Lin), cairg@itp.ac.cn (R.-G. Cai).
\[ ds^2 = -e^{2\delta(r)} \mu(r) \, dt^2 + \frac{1}{\mu(r)} \, dr^2 + r^2 \, d\Omega_n^2, \]

Verlinde argued that by generalizing the well-known Cardy formula, the entropy of the CFT can be expressed by

\[ S = \frac{2\pi R}{n} \sqrt{E_c (2E - E_c)}, \]

where \( E \) is the total energy of the CFT and \( E_c \) is the Casimir energy, nonextensive part in the total energy, which is defined by \( E_c = n(E - TS + pV) \) (note that when some chemical potentials are present, the definition of the Casimir energy should be generalized). While the Cardy–Verlinde formula is very interesting one since the well-known Cardy formula holds only in two dimensions, its validity should be checked. In the spirit of the AdS/CFT correspondence, the thermodynamics of a certain CFT at high temperature can be identified with the thermodynamics of the dual black hole in AdS space. Therefore the Cardy–Verlinde formula can be checked in the AdS/CFT correspondence. Indeed this formula has been found to hold in various cases, for example, Schwarzschild–AdS black holes [8], rotating Kerr–AdS black holes [12,13], charged AdS black holes [14], Taub–Bolt–AdS spacetimes [15]. Kerr–Newmann–AdS black holes [16] and Born–Infeld–AdS black holes [17], etc. Of interest is that similar formula also holds for CFTs dual to asymptotically de Sitter spaces [18]. Clearly, these checks have been made in case by case. On the other hand, these checks only show that this formula holds for CFTs with AdS duals, while for weak coupled CFTs this formula seems no longer valid [19].

In this Letter we would like to go a further step. On one hand, we want to see the validness of the Cardy–Verlinde formula for CFTs, which have gravity dual descriptions in the AdS/CFT correspondence, as general as possible. On the other hand, we are particularly interested in finding some relation between the Cardy–Verlinde formula and Einstein equations together with black hole horizon feature, like the relation between the first law of thermodynamics and Einstein equations made by Jacobson [5], or the relation between Friedmann equations and the first law of thermodynamics [6]. In a word, the general aim of this Letter is to add a piece of evidence linking thermodynamics and Einstein equations.

Let us start with a general \((n + 2)\)-dimensional static, spherically symmetric black hole spacetime

\[ ds^2 = -e^{2\delta(r)} \mu(r) \, dt^2 + \frac{1}{\mu(r)} \, dr^2 + r^2 \, d\Omega_n^2, \]

where \( \delta \) and \( \mu \) are two functions of the radial coordinate \( r \) and \( d\Omega_n^2 \) is the line element of an \( n \)-dimensional unit sphere with volume \( \omega_n = 2\pi^{(n+1)/2} / \Gamma[(n+1)/2] \). The Einstein equations with a negative cosmological constant \( \Lambda = -n(n+1)/2l^2 \) can be written down as

\[ R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R - \frac{n(n+1)}{2l^2} g_{\mu\nu} = 8\pi G T_{\mu\nu}, \]

where \( G \) is the gravitational constant in \((n + 2)\) dimensions and \( T_{\mu\nu} \) is the energy–momentum tensor of matter.

Suppose that \( \delta \) is a continuous function of \( r \), and set

\[ \mu(r) = 1 - \frac{16\pi G m(r)}{n \omega_n r^{n-1}} + \frac{r^2}{l^2}, \]

where the mass function \( m(r) \) is a function of \( r \) only. For the spacetime (3), the Einstein equations reduce to

\[ \delta'(r) = \frac{8\pi G r}{n \mu(r)} (T'_t - T'_r), \]

and

\[ \frac{n}{2r} \mu'(r) = \frac{n(n-1)}{2r^2} (1 - \mu(r)) - \frac{n(n+1)}{2l^2} = 8\pi GT'_t, \]

or

\[ m'(r) = -\omega_n r^n T'_t, \]

where a prime denotes the derivative with respect to \( r \). Integrating (8) yields

\[ m(r) = M - \tilde{m}(r), \quad \tilde{m}(r) = \omega_n \int_r^{\infty} r^n \rho(r) \, dr, \]

where \( \rho(r) \) is the energy density of matter and \( M = m(r)|_{r \to \infty} \) is an integration constant. Since the spacetime (3) is supposed to be an asymptotically anti-de Sitter one, the integration constant \( M \) is just the mass of the configuration. On the other hand, in order the spacetime (3) to be asymptotically AdS, it can be seen from (9) that the energy density \( \rho \) has to fall off at least as

\[ \rho \sim \frac{1}{r^{n+1 + \epsilon}}, \]

where \( \epsilon \) is a small positive quantity. For instance, for a Maxwell field, one has \( T_{\mu\nu} = F_{\mu\lambda} F_{\nu \lambda} - \frac{1}{4} g_{\mu\nu} F^2 \), where \( F_{\mu\nu} \) is the field strength. Further, for the electric field produced by a static point source with charge \( q \), one has the energy density, \( \rho \sim q^2 / r^{2n} \), which satisfies the requirement (10). If the spacetime (3) describes a black hole, its event horizon is then determined by \( \mu(r)|_{r=r_+} = 0 \). In order the horizon to be regular, we can see from (6) that the following condition has to be satisfied

\[ T'_t - T'_r = 0, \]

on the horizon \( r = r_+ \) because \( \mu(r_+) = 0 \), which implies that the energy density of matter must equal to the radial stress on the black hole horizon. For simplicity, in this Letter, we consider the case with \( T'_t - T'_r = 0 \) in the whole spacetime. (In general case, \( T'_t - T'_r = 0 \) holds only on the horizon, see for example, black holes discussed in [20] and references therein.) The static electric field discussed above in the spacetime (3), for instance, satisfies this requirement that \( T'_t - T'_r = 0 \) in the whole spacetime. To see this, let us note that the only nonvanishing component of the Maxwell field in this case is \( F_{tr} = q e^{-\delta} / r^n \), where \( q \) is an integration constant related to the electric charge. Therefore one has \( T'_t = T'_r = -q^2 / (2r^{2n}) \). This clearly satisfies the requirement (11).

The resulting solution of Einstein equations (4) is just the Reissner–Nordström–AdS black hole spacetime. The charge \( q \) can be explained as the R-charge in the dual CFT [21]. Here we also mention that the Born–Infeld–AdS black hole solution also satisfies the condition (11) [17].

With the condition \( T'_t = T'_r \) in the whole spacetime, we have \( \delta = \text{const} \). Without loss of generality, we may take \( \delta = 0 \);
otherwise, we always can absorb the constant $e^{2\delta}$ to the coordinate time $t$. In other words, if $T'_{\\gamma \gamma} - T'_{\gamma r} \neq 0$ in the whole spacetime, then $\delta$ must be a function of $r$.

The Hawking temperature of the black hole can be obtained as follows. Analytically continuing the solution (3) to its Euclidean sector by Wick rotation, requiring the Euclidean solution to be regular on the horizon leads to a fixed period for the Euclidean time. The inverse of the fixed period is just the Hawking temperature. This is the well-known method to get Hawking temperature of a black hole in the path integral approach of Euclidean quantum gravity [22]. Of course, we can also obtain the Hawking temperature by computing the surface gravity of the black hole. Both methods give the same result

$$T_{BH} = \frac{1}{4\pi} \mu' |_{r^+}$$

$$= - \frac{16\pi Gr_+ \rho(r_+)}{n} + \frac{1}{4\pi r_+} \left( (n-1) + \frac{(n+1)r_+^2}{l^2} \right).$$

(12)

Since we are considering black hole solutions in Einstein general relativity, the black hole entropy then obeys the area formula. That is, the entropy of the black hole is given by a quarter of horizon area

$$S = \frac{\omega_n r_+^n}{4G}.$$

(13)

Further, the mass of the black hole can be expressed in terms of the horizon radius $r_+$

$$M = \frac{n \omega_n r_+^{n-1}}{16\pi G} \left( 1 + \frac{r_+^2}{l^2} \right) + \tilde{m}(r_+),$$

(14)

where

$$\tilde{m}(r_+) = \omega_n \int_{r^+}^{\infty} r^n \rho \, dr.$$

(15)

Clearly here $\tilde{m}(r_+)$ is just the total energy of matter outside the black hole horizon.

The first law of thermodynamics for black holes with matter distribution outside the black hole can be written down as [23]

$$dM = T_{BH} dS + \int_\Sigma d^{n+1}x \sqrt{-g} \delta \rho,$$

(16)

where $\Sigma$ denotes the hypersurface with boundary consisting of black hole horizon and spatial infinity. For our black hole solution, using the relation (14), we have

$$dM = \frac{n \omega_n r_+^{n-2}}{16\pi G} \left( (n-1) + \frac{(n+1)r_+^2}{l^2} \right) dr_+ + \delta \tilde{m}(r_+),$$

(17)

with

$$\delta \tilde{m}(r_+) = - \omega_n r_+^n \rho(r_+) \, dr_+ + \omega_n \int_{r^+}^{\infty} r^n \delta \rho \, dr.$$  

(18)

On the other hand, using (12) and (13), the right-hand side of Eq. (16) turns out

$$T_{BH} dS + \int_\Sigma d^{n+1}x \sqrt{-g} \delta \rho$$

$$= - \omega_n r_+^n \rho(r_+) \, dr_+$$

$$+ \frac{n \omega_n r_+^{n-2}}{16\pi G} \left( (n-1) + \frac{(n+1)r_+^2}{l^2} \right) dr_+$$

$$+ \omega_n \int_{r^+}^{\infty} r^n \delta \rho \, dr.$$  

(19)

This showed that the first law of black hole thermodynamics holds for our black hole solution.

Next let us consider the mass expression (14) of the black hole. Multiplying both sides of (14) by a dimensionless factor $l/R$, where $R$ is a length scale, one can rewrite (14) as

$$2(M - \bar{m}(r_+)) \frac{l}{R} = \frac{2n \omega_n r_+^{n-1}}{16\pi GR} \left( 1 + \frac{r_+^2}{l^2} \right) \equiv E_c + 2E_e,$$

(20)

where

$$E_c = \frac{2n \omega_n r_+^{n-1}}{16\pi GR}, \quad E_e = \frac{n \omega_n r_+^{n+1}}{16\pi G R l}.$$

(21)

If we define further

$$E = M/l, \quad E_m = \bar{m}(r_+)l/R,$$

(22)

then we find that Eq. (20) can be rewritten as

$$S = \frac{2\pi R}{n} \sqrt{E_c(2(E - E_m) - E_c)},$$

(23)

where $S$ is the black hole entropy (13). This is nothing but the Cardy–Verlinde formula for a CFT residing on an $n$-dimensional sphere with radius $R$. Compared to the standard Cardy–Verlinde formula (2), a new term $E_m$ appears here. Its meanings will be given shortly. Furthermore, it is quite interesting to note that here we have only used the mass function (14) in the deriving the Cardy–Verlinde formula (23), and have not used any thermodynamic quantities such as temperature, entropy, and pressure, etc. In other words, the Cardy–Verlinde formula (23) is just a rewritten expression of the metric function $\mu(r)$ on the horizon; the metric function $\mu$ of course satisfies the Einstein equations (see (7)). Therefore, in our setup, the Cardy–Verlinde formula is just a rewritten expression of the Einstein equations on the black hole horizon. This provides a relation between the entropy formula of CFTs and Einstein equations in the setup of black hole spacetime. Note that in the setup of FRW universe, the Friedmann equation can be rewritten as the cosmological Cardy formula, which has a same form as the Cardy–Verlinde formula, for details see [8]. The difference between the cosmological Cardy formula and our Cardy–Verlinde formula (23) is that the cosmological Cardy formula always holds for whole evolution of the FRW universe, while the Cardy–Verlinde formula holds only on the black hole horizon, although both are rewritten expressions of Einstein equations in different setups. In addition, let us stress that the above discussions also hold for the case with $E_m = 0$, namely the Schwarzschild–AdS black hole spacetime. Finally, let us
stress that from (23), one can see it is valid only for the case with \( E = E_m > 0 \). To see this clearly, with the help of (22), (15) and (9), \( E - E_m = 0 \) corresponds to \( m(r_+) = 0 \) (in this case \( r_+ = 0 \)), which means that the solution is nothing, but a pure AdS space; when \( E - E_m < 0 \), it corresponds to a naked singularity. In that case, there is no associated thermodynamics with the gravitational configuration, therefore the above discussion loses its meanings.

Next we consider the CFT dual to the AdS black hole solution (3) and show that the entropy of dual CFT indeed satisfies the Cardy–Verlinde formula (23). The dual CFT resides on the boundary of AdS space, whose metric can be determined by the bulk metric, up to a conformal factor. According to the bulk metric (3), the boundary metric can be written down as

\[
ds^2_{\text{CFT}} = \frac{R^2}{r^2} d\tau^2 - \frac{R^2}{l^2} dt^2 + R^2 d\Omega_n^2,
\]

where \( R \) is a length scale. Rescaling the time coordinate \( R\tau/\ell \to \tau \), the boundary metric (24) is then the same as the one introduced in (1), which implies that the dual CFT resides on an \( n \)-dimensional sphere with radius \( R \). Due to the rescaling of the time coordinate, with help of the AdS/CFT correspondence, the dual CFT has energy \( E \) and temperature \( T \)

\[
E = \frac{1}{R} M, \quad T = \frac{1}{R} T_{\text{BH}}.
\]

And its entropy is just the black hole entropy \( S \) given in (13). In addition, let us stress here that for the black hole solution (9), there is an energy contribution from matter outside the black hole. This part of energy should be subtracted in the dual CFT thermodynamics, as noted for the charged black hole discussed in [14]. In the dual CFT, this part should be rescaled as \( E_m = l\hat{m}(r_+)/R \) like the total energy. Thus by definition, the Casimir energy of dual CFT is

\[
E_c = n(E - E_m + TS - pV).
\]

Note that we are considering CFTs, for which the pressure has the expression \( p = (E - E_m)/nV \). Using (25), we find that the Casimir energy is

\[
E_c = \frac{2n\omega_0 \pi^{n+1} l}{16\pi G R},
\]

which is just the one defined in (21). The extensive part of the total energy, by definition [8], is \( E_c = (E - E_m) - E_c/2 \), and has the expression in (21) in terms of black hole horizon radius. Then the entropy \( S \) of dual CFT has the form

\[
S = \frac{2\pi R}{n} \sqrt{E_c(2(E - E_m) - E_c)}.
\]

Thus we showed that the entropy of CFTs dual to AdS black holes with matter distribution outside black hole can be expressed by the Cardy–Verlinde formula. One point which should be stressed here is that the matter energy outside the black hole should be subtracted from the total energy of the dual CFT. In addition, let us stress that for a pure AdS black hole (Schwarzschild–AdS black hole), the term \( E_m \) will be absent in formulas (23) and (28); for AdS black holes with

References


Gauge coupling unification in a 6D SO(10) orbifold GUT

Hyun Min Lee

Deutsches Elektronen-Synchrotron DESY, 22603 Hamburg, Germany

Received 11 September 2006; received in revised form 3 October 2006; accepted 10 October 2006
Available online 3 November 2006
Editor: G.F. Giudice

Abstract

We consider the gauge coupling running in a six-dimensional SO(10) orbifold GUT model. The bulk gauge symmetry is broken down to the standard model gauge group with an extra U(1)X by orbifold boundary conditions and the extra U(1)X is further broken through the U(1)B–L breaking with bulk hyper multiplets. We obtain the corrections of Kaluza–Klein massive modes to the running of the gauge couplings and discuss their implication to the successful gauge coupling unification.

© 2006 Elsevier B.V. All rights reserved.

Grand unified theories (GUTs) have been revived recently in the models of extra dimensions which are compactified on orbifolds, the so-called GUT orbifolds [1,2]. Thanks to orbifold boundary conditions in extra dimensions, a GUT gauge symmetry can be broken down to the Standard Model (SM) gauge group without the need of a GUT Higgs field in the large representation and the doublet-triplet splitting problem can be solved easily.

On an orbifold M/Γ with M a compact manifold and Γ a point group, there are fixed points which transform into themselves under Γ. When the orbifolding breaks the gauge symmetry, there are some of fixed points where the active gauge symmetry is reduced. Although the non-universal gauge kinetic terms localized at the fixed points can be introduced at tree level and generated even by loop corrections [3–5], those effects may be ignored by making the strong coupling assumption at the GUT scale with a large volume of extra dimensions [6]. Thus, due to contributions coming from Kaluza–Klein (KK) massive modes, the GUT orbifolds can provide a minimal setup to predict the QCD coupling for a successful gauge coupling unification.

In this Letter, we consider the running of the gauge couplings in the six-dimensional SO(10) orbifold GUT model proposed in Ref. [7]. This is the minimal setup to break SO(10) down to the SM gauge group up to a U(1) factor only by orbifold boundary conditions without obtaining massless modes from the extra component of gauge bosons. We compute the threshold corrections due to KK massive modes to the gauge coupling running for a number of hyper multiplets with arbitrary parities.

In our case, after the orbifolding, on top of the SM gauge group, there is an extra U(1)X gauge symmetry which has to be broken by a usual Higgs breaking of the U(1)B–L [8,9]. In so doing, 16 Higgs multiplets are introduced in the bulk, so one ends up with extra color triplets as zero modes. Although the extra color triplets can get masses of order the B–L breaking scale MB–L at the fixed points, they could give a large threshold correction to the gauge couplings. We show that the KK threshold corrections can come with opposite sign to the threshold corrections of the color triplets. Thus, even if MB–L is much smaller than the GUT scale, we can get the successful gauge coupling unification due to the cancellation between the large threshold corrections. In this case, the volume of extra dimensions can be large enough for satisfying the strong coupling assumption.

We take some specific examples of embedding hyper multiplets
to show explicitly that this is the case for $M_{B-L}$ being smaller than the compactification scale. There are an extensive list of references [10] where related discussions on the gauge coupling unification have been done mainly in the context of a 5D $SO(10)$ orbifold GUT.

Two extra dimensions are compactified on a torus and they are identified by a $Z_2$ reflection symmetry to make up a $T^2/Z_2$ orbifold. For the extra coordinates $z = \pm x^5 + i x^6$, there are double periodicities in extra dimensions such as $z \sim z + 2\pi R_5 \sim z + 2i\pi R_6$. Due to the orbifold action, there are four fixed points or branes, $z_0 = 0$, $z_1 = \pi R_5$, $z_2 = i\pi R_6$ and $z_3 = \pi R_5 + i\pi R_6$.

A bulk vector multiplet is composed of a vector multiplet $V$ and an adjoint chiral multiplet $\Sigma$ in 4D $\mathcal{N} = 1$ language. In order to break the bulk gauge symmetry down to the SM gauge group, we introduce a nontrivial boundary condition at each fixed point for a bulk vector multiplet by the parity matrices [7].

$$P_i V (z + z_i) P_i^{-1} = V (z + z_i), \quad i = 0, 1, 2, 3,$$

where

$$P_0 = I_{10 \times 10}, \quad P_1 = \text{diag}(-1, -1, 1, 1, 1) \times \sigma^0,$$

$$P_2 = \text{diag}(1, 1, 1, 1, 1) \times \sigma^2,$$

and $P_3 = P_1 P_2$ from the consistency condition on the orbifold. Then, the parity operations $P_1$, $P_2$ break $SO(10)$ down to its maximal subgroups, Pati–Salam group $SU(4) \times SU(2)_L \times SU(2)_R$ and Georgi–Glashow group $SU(5) \times U(1)_X$, respectively.

A bulk hyper multiplet is composed of two chiral multiplets with opposite charges ($H, H'$) and it satisfies the orbifold boundary conditions

$$\eta_i P_i H (-z + z_i) = -H (z + z_i), \quad i = 0, 1, 2, 3,$$

with $\eta_i^2 = 1$. Here $\eta_0 = 1$ and $\eta_1 = \eta_2 \eta_3$, independent of the representation of the hyper multiplet. We consider a set of hyper multiplets, $N_{10} 10'$s and $N_{16} 16'$s satisfying $N_{10} = 2 + N_{16}$ for no irreducible anomalies [11,12]. We also note that both $N_{10}$ and $N_{16}$ have to be even for the absence of localized anomalies unless there are split multiplets at the fixed points [12].

In a 6D non-Abelian gauge theory on orbifolds, where there is no orbifold breaking of the gauge symmetry, the one-loop effective action for the gauge field has been obtained [4]. The analysis has been extended to 6D GUTs with the orbifold breaking of GUT symmetry [5]. By using the general result in the latter analysis, we study the running of the 4D effective gauge couplings of the SM gauge group much below the compactification scale in 6D $SO(10)$ GUTs. After including all possible contributions, the running of the low-energy gauge couplings are governed in dimensional regularization by

$$\frac{4\pi}{g_{\text{eff},a}^2(k^2)} = \frac{4\pi}{g_{a}^2(k^2)} + \frac{1}{4\pi} b_a \ln \frac{M_a^2}{M_{B-L}^2} + \frac{1}{4\pi} b'_a \ln \frac{M_{B-L}^2}{k^2} - \frac{1}{4\pi} \left( \sum_{\pm} b_{a,\pm} L_{\pm\pm} + \sum_{\pm} b'_{a,\pm} L_{\pm\mp} \right),$$

where $M_a$ is the 6D fundamental scale, $M_{B-L}$ is the $B-L$ breaking scale, $g_a$ is the universal renormalized gauge coupling1, and $\Delta_{B-L}$ are corrections due to renormalized gauge couplings localized at the Pati–Salam and flipped $SU(5)$ fixed points. $\Delta_{B-L}$ stands for the effect due to the modification of the KK masses due to the $B-L$ breaking brane-localized mass terms. Note further that $b_{a,\pm} = (33, 5, 1, -3)$ is the beta function in the MSSM as given below the $B-L$ breaking scale while $b_a$ is the beta function above the $B-L$ breaking scale. More importantly, $L_{\pm\pm}(L_{\pm\mp})$ are the logarithmic KK threshold corrections with the corresponding beta functions $b_{a,\pm}(b'_{a,\mp})$. These are purely bulk contribution [5].

Here we present the details of the beta functions in Eq. (8). We split $b_a$ into $b_a = b_a - c_a + b_{a,\pm}$. Here $b_a$ is the contribution from zero modes which are distributed both in the bulk and at the fixed points [4,5]. It is given by

$$b_a = b_a^V + b_{a,10}^V + b_{a,16}^V + b_a^1,$$

with

$$b_a^V = (0, -6, -9),$$

$$b_{a,10}^V = \frac{1}{4} N_{10}(1, 1, 1) + \frac{1}{4} \sum_{i=10}^{12} d_{10}^i \left( \frac{1}{5}, 1, -1 \right),$$

$$b_{a,16}^V = \frac{1}{4} \sum_{i=16}^{10} d_{16}^i \left( 1, 1, 1 \right) + \frac{1}{4} \sum_{i=16}^{10} d_{16}^i \left( -\frac{6}{5}, 2, 0 \right) + \frac{1}{4} \sum_{i=16}^{10} d_{16}^i \left( \frac{7}{5}, -1, -1 \right),$$

where $d_{10}^i$ and $d_{16}^i$ with $(d_{10}^i)^2 = (d_{16}^i)^2 = 1$ are the parities for $10$ and $16$, respectively. $c_a$ is the beta function for vector-like massless modes which will get tree-level brane

\footnote{1 Although there are also power-like threshold corrections in the cutoff regularization [4,13], they do not contribute to the differential running of gauge couplings. Nevertheless, the power-like contributions may have the net effect of placing an upper limit on the possible volume of the extra dimensions [14].}
masses of the order of the GUT scale. Moreover, \( b_m^a \) is the beta function for the brane-localized fields. Depending on the parities, we get the different logarithms for the KK threshold corrections as

\[
L_{++} = \ln \left[ 4e^{-2} \eta (iu)^4 u VM_{0}^2 \right],
\]

\[
L_{+-} = \ln \left[ e^{-2} \eta \left( \frac{1}{2} iu \right)^4 u VM_{0}^2 \right],
\]

\[
L_{-+} = \ln \left[ e^{-2} \eta \left( - \frac{1}{2} iu \right)^4 u VM_{0}^2 \right],
\]

\[
L_{--} = \ln \left[ e^{-2} \eta \left( \frac{1}{2} - \frac{1}{2} iu \right)^4 u VM_{0}^2 \right],
\]

where \( u = R_0/R_5 \), \( V = 4\pi^2 R_5 R_6 \), \( \eta \) and \( \theta_1 \) are the Dedekind eta function and the Jacobi theta function, respectively. The beta function for KK massive modes is

\[
b_{a+} = \frac{1}{4} (-8 + N_{10} + 2 N_{16})(1, 1, 1),
\]

\[
b_{a+} = \frac{1}{4} \left( \frac{12}{5}, 4, 0 \right) + \frac{1}{4} \sum_{16} \eta^{10} \left( \frac{1}{5}, 1, -1 \right)
\]

\[
+ \frac{1}{4} \sum_{16} \eta^{16} \left( - \frac{6}{5}, 2, 0 \right),
\]

\[
b_{a+} = \frac{1}{4} \left( 2 + \sum_{16} \eta^{16} \right) (-1, -1, -1),
\]

\[
b_{a+} = \frac{1}{4} \left( \frac{38}{5}, -2, -2 \right) + \frac{1}{4} \sum_{16} \eta^{16} \left( \frac{7}{5}, -1, -1 \right).
\]

Compared to Eq. (9), we obtain the relation between beta functions as

\[
b_a = (0, -4, -6) + b_{a+} + b_{a+} + b_{a-} + b_{a-},
\]

where the first term is due to the difference between the beta functions of \( N = 1 \) vector multiplets and \( N = 2 \) vector multiplets for the SM gauge group. Consequently, from the beta functions (11), (12), (18) and (20), one can find that the part proportional to \( \eta_i^R \) or \( \eta_i^R \eta_i^R \) is non-universal. So, because of the orbifold actions associated with Pati–Salam and flipped SU(5) gauge groups, both massless and massive mode contributions can affect the differential running of the gauge couplings.

For a number of hyper multiplets with arbitrary parities, we assume that both vector-like particles (getting brane masses of order the GUT scale) and brane-localized particles fill GUT multiplets, i.e. \( c_a \) and \( b_{a}^m \) are universal. In this case, those particles do not affect the unification of the one-loop gauge couplings. Then, we get the general formula for the differential running of gauge couplings as

\[
\frac{1}{g_3^2} = \frac{12}{7} \frac{1}{g_2^2} + \frac{5}{7} \frac{1}{g_1^2}
\]

\[
= \frac{1}{8\pi^2} \left( \tilde{b} \ln \frac{M_s}{M_{B-L}} - \frac{1}{2} b^{++} L_{--} - \frac{1}{2} b^{--} L_{--} \right. \]

\[
\left. + \tilde{\alpha}^l + \tilde{\alpha}^{B-L} \right),
\]
see that the individual logarithm can be large, being compatible with the gauge coupling unification due to a cancellation. We will focus on this possibility later on. The case with the anisotropic compactification \( u \gg 1 \) will be discussed in detail elsewhere in Ref. [5].

Now we are in a position to apply our general formula (27) to particular cases for the unification of the SM gauge couplings. To this purpose, we consider some known \( SO(10) \) models of embedding the MSSM into the extra dimensions. In the minimal model: (model I) [8] that contains Higgs fields in the bulk for breaking \( U(1)_{B-L} \) and the SM gauge group,\(^2\) there are 4 \( 10 \)'s with parities \( (\eta_1, \eta_2) \) such as \( H_1 = (+, +), \ H_2 = (+, -), \ H_3 = (-, +) \) and \( H_4 = (-, -) \), and one pair of \( 16 \) and \( \overline{16} \) with parities \( \Phi = (-, +), \ \Phi^c = (-, +) \). Then, the resulting massless modes are two doublet Higgs fields \( H_1^c \) and \( H_2 \) from \( H_1 \) and \( H_2 \), and \( G_5^c, G_4, (D^c, N^c), (D, N) \) from \( H_3, H_4, \Phi \) and \( \Phi^c \) in order. Moreover, each family of quarks and leptons is introduced as a \( 16 \) being localized at the fixed point without \( SO(10) \) gauge symmetry. After the \( B-L \) breaking via the bulk \( 16 \)'s with \( \langle N \rangle = \langle N^c \rangle \neq 0 \), neutrino masses are generated at the fixed points by a usual see-saw mechanism. Moreover, \( G_5^c, G_4, (D^c, N^c), (D, N) \) can acquire masses of order the \( B-L \) breaking scale by the brane superpotential [8,9]

\[
W = \lambda N D G_5^c + \lambda' N^c D^c G_4 \quad \text{for} \quad \langle N \rangle = \langle N^c \rangle \neq 0.
\]

In this case, since \( \sum_{10} \eta_1^{10} = 0, \sum_{16} \eta_1^{16} = \sum_{16} \eta_1^{16} \eta_1^{16} = 2, \) we get the values \( b = \frac{18}{7}, b^{-+} = \frac{2}{7} \) in Eq. (27).

We consider another 6D \( SO(10) \) GUT model where the realistic flavor structure of the SM was discussed: (model II) [9]. In this case, on top of the minimal model, there are more hyper multiplets; 2 \( 10 \)'s such as \( H_5 = (-, +) \) and \( H_6 = (-, -) \), and one pair of \( 16 \) and \( \overline{16} \) with \( \Phi = (+, +), \ \Phi^c = (+, +) \). Then, there are additional zero modes \( G_5^c, G_6, L, L^c \) from \( H_5, \ H_6, \Phi, \\Phi^c \) in order. They are assumed to get brane masses of order the GUT scale. Thus, the running of gauge couplings between the GUT scale and the \( B-L \) breaking scale is the same as in the minimal model. In this case, since \( \sum_{10} \eta_1^{10} = -2, \sum_{16} \eta_1^{16} = \sum_{16} \eta_1^{16} \eta_1^{16} = 0, \) we get the values \( b = \frac{18}{7}, b^{-+} = \frac{2}{7} \) in Eq. (27).

Consequently, in both cases, we can see that logarithmic contributions of zero modes and those of KK massive modes appear with opposite signs so that there is a possibility of having the large volume of extra dimensions consistent with perturbativity and gauge coupling unification. From the data of the electroweak gauge couplings at the scale of the Z mass, one can compare the predicted value of the QCD coupling in a theory to a measure one [15] \( \alpha_{\exp}^s = 0.1176 \pm 0.0020 \). In the 4D supersymmetric GUT's, the prediction without threshold corrections for the QCD coupling is \( \alpha_{\exp}^s = 0.130 \pm 0.004 \). Thus, in this case, there is a discrepancy from the experimental data as \( \Delta \alpha_L = \alpha_{\exp}^s - \alpha_{\exp}^s = 0.0124 \pm 0.0045 \). For the models that we considered above, ignoring the unknown brane-localized gauge couplings and the \( B-L \) breaking effect, we depict in Fig. 1 the parameter space of \( (M_c, M_{B-L}) \) with \( M_c \equiv 1/\sqrt{V} \) and \( u \sim 1 \), being compatible with the experimental data. Taking \( M_c/M_c \sim 63 \) for strong coupling assumption at the 6D fundamental scale, the correction due to the brane-localized gauge couplings is \( \Delta^L = O(1) \) so it is negligible to the KK threshold corrections which is of order \( \ln(M_c/M_c) \sim 3 \). For \( M_{B-L} < M_c \), it has been shown [17] that the KK massive modes of the color triplets are modified to \( m_n^3 \approx (n_s/2R_5)^2 + (n_s/2R_6)^2 + c M_{B-L}^2 \) where \( c \) is of order unity independent of the KK level for \( R_5 \neq R_6 \). In this case, the \( B-L \) breaking effect to the differential running (22) is estimated as \( \Delta^{B-L} \sim M_{B-L}^2/M_c^2 \). In the model I(II), for \( M_s/M_c \sim 63/\sqrt{V} \sim 22 \) for the group theory factor \( C = 8, M_{B-L}/M_c \) can be as small as 0.23(0.12) at the 2\( \sigma \) level so that the \( B-L \) breaking can be suppressed compared to the KK threshold corrections. Apart from the two models, we can consider other possibilities of embedding the matter representations into extra dimensions, like in the field-theory limit of a successful string orbifold compactification [16] where there are two families at the fixed points and one family in the bulk. In view of the general formula (27), however, as far as an extra particle contributes to the running of the gauge couplings above the \( B-L \) breaking scale, \( M_{B-L} \) tends to be close to \( M_c \) for the success of the gauge coupling unification, independent of the details of the model.

To conclude, we have obtained the KK massive mode corrections as a dominant contribution to the gauge coupling running

\[ M_{\text{LL}}/M_c \]

\[ M_{\text{BL}}/M_c \]
in a six-dimensional SO(10) orbifold GUT model. The shape dependent correction of the KK massive modes can be dominant in the anisotropic compactification of the extra dimensions. Compared to the 5D case, the 5D limit of our computation shows that the 5D power-like threshold corrections can be computed to be non-universal for the SM gauge couplings. Focusing on the isotropic compactification of the extra dimensions, we have shown that there is a generic cancellation between the dominant logarithmic corrections to the differential logarithmic running of the SM gauge couplings: one is the contribution of the extra particles above the $B-L$ scale and the other is the KK massive mode contribution. In the models that we considered, extra color triplets contribute to the running of the gauge couplings above the $B-L$ scale but the KK threshold corrections can be large enough to cancel the contribution of the extra color triplets for the large volume of extra dimensions. Therefore, the $B-L$ scale can be much smaller than the GUT scale.

Since the $B-L$ breaking scale tends to be close to or larger than the compactification scale as shown in the allowed parameter space of Fig. 1, it may be also important to see how much the modified KK massive modes of the color triplets due to the brane-localized mass terms can affect the running of the gauge couplings. On the other hand, one can look for a consistent model where the color triplets make up GUT multiplets together with some extra doublets, i.e. $b = 0$. Then, the $B-L$ breaking would not be relevant for the gauge coupling unification any more. In this case, the extra dimensions could be also large enough for the successful gauge coupling unification, independent of the details of the model with hyper multiplets. We leave the relevant issues in a future publication.

Acknowledgements

The author would like to thank W. Buchmüller, L. Covi and A. Hebecker for useful comments and discussions.

References


Constraining superWIMPy and warm subhalos with future submillilensing

Junji Hisano a, Kaiki Taro Inoue b, Tomo Takahashi c,*

a Institute for Cosmic Ray Research, University of Tokyo, Kashiwa 277-8582, Japan
b Department of Science and Engineering, Kinki University, Higashi-Osaka 577-8502, Japan
c Department of Physics, Saga University, Saga 840-8502, Japan

Received 19 August 2006; received in revised form 17 September 2006; accepted 28 October 2006
Available online 7 November 2006
Editor: T. Yanagida

Abstract
We propose to observe QSO-galaxy strong lens systems to give a new constraint on the damping scale of the initial fluctuations. We find that the future observation of submilliarc scale astrometric shifts of the multiple lensed images of QSOs would find $\sim 10^{3} - 10^{9} M_{\odot}$ subhalos inside the macrolens halo. The superweakly interacting massive particles (superWIMPs) produced from a WIMP decay and the warm dark matter (WDM) particles that predict a comoving damping scale larger than $\sim 2$ kpc can be constrained if $\sim 10^{3} M_{\odot}$ subhalos are detected.

© 2006 Elsevier B.V. All rights reserved.

1. Introduction

There has been mounting evidence that most of the matter in the universe is not luminous but dark. Current observations such as the cosmic microwave background (CMB) suggest that the dark matter (DM) makes up about 25% of the universe [1]. The dark matter is usually assumed to be a collisionless cold component (nonrelativistic at the time of freeze-out), called as the cold dark matter (CDM). Weakly-interacting massive particles (WIMPs), such as the lightest neutralino in the supersymmetric standard model (SUSY SM) [2], and the lightest Kaluza–Klein particle in the universal extra dimension (UED) [3], are the popular CDM candidates. The predicted thermal relic abundances and the large-scale structure ($\geq 1$ Mpc) are in good agreement with the observed values.

However, the recent high-resolution $N$-body simulations on the CDM-based structure formation revealed various discrepancies on smaller scales ($\lesssim 1$ Mpc). The first one is so-called the “missing satellite problem” [4]: the $N$-body simulations of the CDM particles predict significantly more virialized dark objects with mass $M \lesssim 10^{9} M_{\odot}$ (or subhalos) in the galaxy-sized halos with $M \sim 10^{12} M_{\odot}$ than those observed around the Milky Way.

The other one is called the “cusp problem” [5]: the CDM-based models also predict a cuspy profile for mass density distributions for the CDM halos [6] although the measurements of the rotation curves imply the presence of cores in the centers of the halo.

Although these discrepancies may be circumvented by some baryonic processes [7], it may be worthwhile to consider the other kinds of DM particles with different clustering properties. For instance, the superweakly interacting massive particles (superWIMPs) [8] or the warm dark matter (WDM) [9] particles can have large velocity dispersion at the epoch of radiation–matter equality. If the DM consists of the superWIMPs or the WDM particles, the number of less massive $\lesssim 10^{9} M_{\odot}$ subhalos is significantly reduced, and the cusp formation is also suppressed because the primordial fluctuations at the small scales ($\lesssim 1$ Mpc) are damped.

In this Letter, we consider the possibilities of probing such subhalos with $M \lesssim 10^{9} M_{\odot}$ via strong lensing in which the image separations are on submilliarcsecond scales, called as “submillilensing”. Recently it has been pointed out that the future submillilensing observations of multiply-imaged QSO-galaxy lens systems can directly probe the mass scale of subhalos in the parent galaxy halo [10]. The submillilensing observation might resolve whether above problems originate from various baryonic contrivances or the nature of the DM particles. In Section 2, we begin with the discussion of submillilensing and

* Corresponding author.
E-mail address: tomat@cc.saga-u.ac.jp (T. Takahashi).

0370-2693/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.physletb.2006.10.050
consider the possibility of detecting the small-mass subhalos via submillilensing in the next decade. In Section 3, we study its implications to the superWIMP and the WDM scenarios. Section 4 is devoted to conclusions and discussion.

2. Submillilensing

In what follows, we discuss the possibilities of direct detection of subhalos with a mass of \( M \gtrsim 10^9 \, M_\odot \) via substructure lensing which is defined as lensing by \( M \lesssim 10^6 \, M_\odot \) subhalos that perturb a “simple” strong lensing by \( \sim 10^{12} \, M_\odot \) parent galaxy halo. To date, about 10 quadruply-imaged gravitational lenses with flux ratio “anomalies” have been detected [11–13]. Here “anomaly” refers to an observed image flux ratio that does not agree with the ratio predicted by standard macrolens models with a smooth gravitational potential. From the radio and the mid-infrared observation, some of those lens systems are described by tidally-cut singular isothermal spheres (SISs) [16]. Using the equations for SIS at \( \theta E \) is approximately given by

\[
\theta E \sim 10 \left( \frac{\sigma_{\text{SIS}}}{20 \, \text{km s}^{-1}} \right)^2 \text{mas}. \tag{2.2}
\]

Thus, observation with angular resolution of submilliarcsecond scales (\( \sim 0.01 \, \text{mas} \)), which will be achieved in the next generation satellite VLBI mission such as the VSO2P [20], can reveal subhalos with one-dimensional velocity dispersion \( \sigma_{\text{SIS}} \gtrsim 0.6 \, \text{km s}^{-1} \). It corresponds to \( M \gtrsim 10^3 \, M_\odot \) at the distance equal to the Einstein radius \( r = r_E \) of the macrolens, assuming that the velocity dispersion of the parent halo is \( \sigma \sim 200 \, \text{km s}^{-1} \).

From the astrometric shifts of the multiple extended images perturbed locally by a subhalo with respect to an unperturbed macrolensed image, we can break the degeneracy between the subhalo mass and the distance in the line of sight to the images if resolved at scale of an Einstein radius of the perturber [17,19]. This is of great importance because otherwise we cannot determine whether the flux ratio anomaly is caused by more massive intergalactic halos in the line of sight or by less massive subhalos within the macrolens halo. Even if the density profile of the perturber is shallower than an SIS, we can make a distinction between models with different density profiles from astrometric shifts of the surface brightness profile within the source [19]. A direct detection of less massive \( 10^3 \, M_\odot \lesssim M \lesssim 10^9 \, M_\odot \) subhalos within the parent halo will give a stringent constraint on the superWIMP and the WDM scenarios, which will be discussed in the next section.

3. Implications to superWIMP and WDM scenarios

There exist many well motivated models for superWIMPs and WDM from particle physics. The natural candidates for the superWIMPs are gravitino and axino, that are superpartners of graviton and axion, respectively [8]. The right-handed sneutrino is also the candidate when neutrino masses are Dirac-type [21]. Others are Kaluza–Klein graviton and axion states in the UED [8]. As for the WDM, a light gravitino and sterile neutrinos have been discussed as such candidates. Thus, it is interesting to study the feasibility of probing or constraining those models in the near future experiments.

In the following, we refer “superWIMPs” as the particles whose interactions are weaker than the weak interaction, such as the gravitino which couples to other particles only gravita-
tionally. SuperWIMPs are also assumed to be produced by the decays of heavier particles whose interactions are weak (e.g. WIMPs).\footnote{Usually conventional WDM particles do not satisfy both of these properties. In this sense, superWIMPs can be distinguished from conventional WDM particles.}

In the CDM-based structure formation models, the structure of the universe forms in a hierarchical manner. Protohalos, which are the first virialized objects, appear first after the mass density fluctuation becomes nonlinear, and larger objects form successively via their merger. In the ordinary WIMP models, the comoving damping scale of the power spectrum is typically (0.01–10) pc, depending on the Hubble radius at the kinetic decoupling temperature \[T_\text{dec} \simeq 10^{-3} \text{K}\] and the protohalo mass is \[(10^{-12}–10^{-4}) M_\odot\]. \[23\]

In the superWIMP or WDM scenarios, the primordial fluctuations on even larger scales can be damped because they have large velocity dispersion at the decoupling. Therefore, the protohalos become massive, and the total amount of subhalo mass inside the parent halo can be significantly reduced.

When the comoving damping scale of the DM power spectrum is \(R_{\text{cut}}\), the protohalo mass is roughly given by
\[
M_{\text{cut}} \simeq 1 \times 10^{11} M_\odot \left( \frac{R_{\text{cut}}}{1 \text{ Mpc}} \right)^3. \tag{3.1}
\]

Here, the present matter density \(\Omega_{\text{m,0}} = 0.24\) and the Hubble constant \(h = 0.73\) are assumed [1]. In the previous section, we showed that the sensitivity of direct detection of subhalos inside the parent halo with submillilensing may reach intermediate mass \(\sim 10^4 M_\odot\) scales. Thus, when \(R_{\text{cut}} \gtrsim 2\) kpc, the future submillilensing experiments may directly detect such protohalos.

First, we discuss the damping scale in the superWIMP scenario and compare it with the sensitivities of future submillilensing. In the scenario, some WIMPs freeze out from the thermal equilibrium as the usual WIMP DM models, and superWIMPs are nonthermally produced from the WIMP decay, since the superWIMPs interact superweakly with the thermal bath. The scenario retains the property of the ordinary WIMPs that the observed relic density is naturally achieved. Furthermore, they are produced by decay of some long-lived massive particle \(X\). Since they can have large velocities at the production epoch, the free-streaming damps the small-scale inhomogeneities. The superWIMP scenario is a possible explanation for the small-scale structure [24,25].

The comoving damping scale is typically given by the free-streaming length of the superWIMP, \(R_{fs}\), at the matter-radiation equality \(t_{eq}\). When the long-lived particle has lifetime \(\tau_X\) and the superWIMP is produced with three-momentum normalized by its mass \(u(= p/m)\), the comoving free-streaming scale \(R_{fs}\) is given by
\[
R_{fs} = \frac{2u_{eq}}{a(t_{eq}) u_{eq}} \left[ \log \left( \frac{1}{u_{eq}} + \sqrt{1 + \frac{1}{u_{eq}^2}} \right) - \log \left( \frac{1}{u} + \sqrt{1 + \frac{1}{u^2}} \right) \right]. \tag{3.2}
\]

where \(u_{eq} = \langle a(\tau_X)/a(t_{eq}) \rangle u\) and \(a(t)\) is the scale factor as a function of \(t\).

When the superWIMPs are produced from decay of electrically-charged particles,\footnote{It is recently pointed out in Ref. [26] that long-lived charged particles \((\tau_X \gtrsim 10^3 s)\) lead to overproduction of \(^6\)Li in the Big Bang nucleosynthesis due to the catalytic enhancement of the production. On the other hand, it is argued in Ref. [27] that there are a lot of ambiguities in the derivation.} the small-scale power may be further suppressed [28]. The charged particles are coupled to the photon-baryon fluid which oscillates on subhorizon scale. If the scale in question enters the horizon before the decay, the density fluctuation of such DMs cannot grow but oscillates, which gives a rise to the suppression on small scales. The damping scale is typically given by the Hubble radius at the decay time, \(H^{-1}(\tau_X)\).

\[
R_{ch} = \frac{1}{a(\tau_X)} H^{-1}(\tau_X). \tag{3.3}
\]

In Fig. 1 we show the damping scales \(R_{\text{cut}}\) of the charged and neutral \(X\) cases as functions of \(\tau_X\) and \(u\). When \(X\) is neutral, the damping scale is determined by the free streaming scale \(R_{fs}\). Longer lifetime \(\tau_X\) implies a larger free streaming scale, since the velocity dispersion at the radiation-matter equality time, which is proportional to \(a(\tau_X)/a(t_{eq})\), becomes larger. When \(X\) is charged, the damping scale \(R_{\text{cut}}\) corresponds to the larger one of \(R_{fs}\) and \(R_{ch}\). For \(u \ll 1\), \(R_{\text{cut}}\) is given by \(R_{ch}\), since \(R_{fs}\) is \(\sim u \times R_{ch}\). The hatched region with damping scales larger than \(\sim 1\) Mpc is constrained from Lyman alpha clouds [29]. The future submillilensing experiments may cover regions above the bold lines, which correspond to \(R_{\text{cut}} \gtrsim 2\) kpc. When the \(X\) lifetime is longer than \(\sim 1\) s and the superWIMPs are produced with relativistic momentum \((u \gtrsim 1)\), the damping scale is larger than \(\sim 1\) kpc, which may be constrained by future observation if the subhalos with \(M \sim 10^3 M_\odot\) were discovered. When \(X\) is charged, the region with \(\tau_X \gtrsim 400\) s may be also covered even in the small \(u\) cases.

Following Ref. [24], we indicated the parameter region which is suitable to solve the discrepancies in small-scale structure in Fig. 1. The region with \(R_{\text{cut}} \simeq (0.4 – 1.0)\) Mpc is suitable to solve the “missing satellite problem”, while the gray region with \(1 \lesssim u^{-3}(\tau_X/10^3 \text{ s})^{3/2} \lesssim 4\) is favored to solve the “cusp problem”, which requires a large DM velocity dispersion [30]. As we can see in Fig. 1, the model parameters corresponding to these regions can be well constrained by the future submillilensing observation.

For the illustrative purpose, we consider a superWIMP gravitino model, in which the gravitinos are produced by slepton or sneutrino decay. sleptons or sneutrinos are assumed to be the lightest SUSY particles in the SUSY SM.\footnote{The lightest neutralino is also one of the candidates for the lightest SUSY particle in the SUSY SM. However, the decay produces hadronic shower to spoil the BBN if it is not photino-like. The long-lived slepton and sneutrino are not strongly constrained by the BBN, since their hadronic branching ratios are small. See Ref. [31] for the constraints on the superWIMP gravitino model.} The slepton and
sneutrino lifetimes are given by [32]

$$\tau_X = 48\pi M_*^2 m_X^2 \left(1 - \frac{m_X^2}{m_*^2}\right)^{-1},$$  \hspace{1cm} (3.4)

where $M_* = 2.4 \times 10^{18}$ GeV, $m$ is the gravitino mass, and $m_X$ is the mass of the parent WIMP $X$. In Fig. 2, $R_{fs}$ (solid lines) is shown as a function of $m$ and the mass difference between the gravitino and the parent particle ($\Delta m = m_X - m$). Dashed lines indicate the damping scale for which $R_{ch} > R_{fs}$. The gray region which corresponds to explain the “cusp problem” [24] may be excluded when the future submillilensing experiments find subhalos with mass smaller than $1 \times 10^{7.8-8} M_\odot$. It is known that the lifetime and the mass of the long-lived particles whose decay produces the superWIMPs, are constrained by the Big Bang nucleosynthesis (BBN) and by the CMB Planckian spectrum, depending on the decay channels. When the hadronic modes are dominant in the decay, the BBN constrains the lifetime to be shorter than $\sim 1$ s [33]. Even if the hadronic shower is suppressed, the electromagnetic energy injection from the long-lived particle decay to the thermal bath is also constrained from the BBN while the constraint is weaker. In addition to it, the CMB Planckian spectrum also gives a constraint when the lifetime is longer than $\sim 10^6$ s [32].

The future submillilensing experiments are complementary to those constraints, since the damping scale $R_{cut}$ is independent of the decay channels. For example, when the superWIMP gravitinos are produced from sneutrino decay, the energy injection to thermal bath is very tiny so that the constraints from the BBN and the CMB Planckian spectrum are very weak. Even in such a case, the submillilensing will still constrain the model.

Next, we discuss the damping scale in the WDM models. Among many WDM candidates discussed so far, the light gravitino with mass of order from $10^{-6}$ eV up to 1 keV, which is likely to be the lightest SUSY particle in gauge-mediated models of supersymmetry breaking, is one of the well-motivated models from the viewpoint of particle physics [34,35]. Such gravitinos are in thermal equilibrium at early times but decouple when the degrees of freedom $g_{*}(T_{D})$ is $\mathcal{O}(100)$, where $T_{D}$ is the decoupling temperature. Because the decoupling temperature of such species is higher than that of (active) massive neutrinos which play roles of hot dark matter, their velocity dispersion is not so large compared to that of massive neutrinos but nonnegligible at the time of structure formation. Thus, they can act as the WDM. The comoving damping scale for the free-streaming for such gravitinos or any other thermal relic can be written as

$$R_{fs} \sim 0.84 \text{ Mpc} \left(\frac{g_{*}(T_{D})}{10.75}\right)^{1/3} \left(\frac{1 \text{ keV}}{m_{WDM}}\right) \left(\frac{\langle p/T \rangle}{3.15}\right),$$  \hspace{1cm} (3.5)

where $\langle p/T \rangle$ is the mean momentum over the temperature. For thermally decoupled species, this factor gives almost unity, i.e., $\langle p/T \rangle/3.15 \sim 1$. The requirement from Lyman alpha clouds, $R_{fs} \lesssim 1$ Mpc, implies $m_{WDM} \gtrsim 1$ keV. On the other hand, the energy density of WDM can be written as $\Omega_{DM} h^2 = (m_{WDM}/94 \text{ eV})(10.75/g_{*}(T_{D}))$. Assuming $\Omega_{DM} h^2 \sim 0.10$, the mass of WDM should be $m_{WDM} \sim 0.1$ keV even for $g_{*}(T_{D}) \sim 100$. We need to introduce more extra degree of freedom around $T_{D}$ as $g_{*}(T_{D}) \sim \mathcal{O}(10^3)$. When the constraint on $R_{fs}$ is improved to be $\lesssim 1$ kpc, the lower bound for the mass can reach $m_{WDM} \sim 1$ MeV for $g_{*}(T_{D}) \sim 100$. Thus, the WDM scenario may face further difficulties if future submillilensing experiments would find small-mass subhalos.
Another well-motivated candidate for WDM is the sterile neutrinos [36]. Because they directly couple to the active neutrinos alone, they can be produced via neutrino oscillation. Although the evaluation of their energy density requires a numerical integration of the Boltzmann equation, some useful fitting formulae are available. The present energy density of sterile neutrinos can be written as \[ \Omega_{\nu_s} h^2 \sim 0.3 \left( \frac{\sin^2 2\theta}{10^{-10}} \right) \left( \frac{m_s}{100 \text{ keV}} \right)^2, \] (3.6)

where \( \theta \) is the mixing angle between the active neutrinos and sterile neutrinos and \( m_s \) is the mass of sterile neutrinos. The temperature at the time when the production is most efficient is \[ T \sim 130 \text{ MeV} \left( \frac{m_s}{3 \text{ keV}} \right)^{1/3}. \] (3.7)

The sterile neutrinos can damp the small-scale inhomogeneities via the free streaming in the same manner as the thermally decoupled WDM particles do. The free-streaming scale for a sterile neutrino WDM can also be obtained using Eq. (3.5) with different values for \( \langle p/T \rangle \) from that of the thermally decoupled ones. Because the sterile neutrinos are not in the thermal bath at early times, their distribution function deviates from that of a thermal one and the above factor can be \( \langle p/T \rangle /3.15 \sim 0.9 \) for the standard production mechanism [39]. Although there are some differences between the thermally decoupled WDM and the sterile neutrino WDM, they give the same predictions for the damping of matter power spectrum by identifying their masses as [40,41]

\[ m_s = 4.71 \text{ keV} \left( \frac{m_{\text{thermal}}}{1 \text{ keV}} \right)^{4/3} \left( \frac{0.10}{\Omega_{\text{DM}} h^2} \right)^{1/3}, \] (3.8)

where \( m_{\text{thermal}} \) is the mass of thermally decoupled relics such as a light gravitino, which is denoted as \( m_{\text{WDM}} \) in Eq. (3.5). Because the shape of matter power spectrum is determined by the ratio \( m/T \) and the density parameter \( \Omega_{\text{DM}} h^2 \), the constraint on the sterile neutrino mass bound would be different from that on the thermally decoupled WDM mass bound. Using the above formula and assuming the damping scale as \( R_{\text{cut}} \lesssim 1 \text{ kpc} \) which can be reached by future submillilensing experiments, we can expect that the mass of sterile neutrinos can be constrained to be \( m_s \gtrsim 40 \text{ MeV} \), which will be in conflict even with the current constraint \( m_s \lesssim 10 \text{ keV} \) [39,42,43].

Here some comments on the mixed dark matter scenario are in order. It is possible that, for example, gravitinos are produced not only from the decay of the next-lightest supersymmetric particles (NLSP) but also from thermal plasma. In this case, the present DM is composed of CDM and superWIMP. In such mixed DM scenarios, the damping of the small-scale structure is less significant in comparison with the models in which the superWIMPs make up all of the DM. To what extent the amplitude at small scales can be reduced depends on the ratio of the energy density of CDM and superWIMP DM. Detailed discussion on this issue is beyond the scope of this Letter. Some discussions on the matter power spectrum in such mixed models can be found in Refs. [44,45]. [44] analyzed models in which the superWIMPs are produced from the charged particle decay while [45] considered models in which the superWIMPs are produced from the neutral particle decay.

4. Conclusions and discussion

We have shown that future observation of multiply-imaged QSO-galaxy lens systems with a high angular resolution \( \sim 0.01 \text{ mas} \) will prove the small-scale clustering properties of the DM halos down to \( \sim 10^5 M_{\odot} \). The presence of \( \sim 10^3 M_{\odot} \) subhalos implies the comoving damping scale of the primordial fluctuations \( \sim 2 \text{ kpc} \) assuming that the observed subhalos retain the original mass during the merger process. The superWIMP and the warm DM scenarios that predict a larger damping scale (0.002–1) Mpc in comparison with (0.01–10) pc in the ordinary WIMP scenarios would be strongly constrained if the presence of such subhalos were proved.

In the superWIMP scenario, the superWIMP DM is produced from the long-lived particle \( X \) decay, and free-streaming of the superWIMPs damps the small-scale inhomogeneities. When the lifetime of \( X \) is longer than \( \sim 1 \text{ s} \), the damping scale is larger than \( \sim 1 \text{ kpc} \) unless the superWIMP and DM masses are degenerate and the superWIMPs are nonrelativistic at the production epoch. In addition, when \( X \) is a charged particle, it is coupled to the oscillating photon–baryon fluid before the decay. Therefore, the small-scale inhomogeneities inside the sound horizon cannot grow. The damping scale becomes larger than \( \sim 1 \text{ kpc} \times (\tau_X/100 \text{ s})^{1/2} \) in the case of charged \( X \). One of the natural superWIMP candidates is the gravitino produced from the slepton or sneutrino. In this case, the damping scale is typically larger than \( 10^2 \text{ kpc} \). The future submillilensing experiments, which cover the DM subhalos with mass \( \gtrsim 10^5 M_{\odot} \), are important tests for probing such superWIMP scenarios. For the WDM scenarios, such as the light gravitino and the sterile neutrinos, the WDM mass bound would be further constrained.

The survivability of the protohalos during the merger process is under debate now. In the ordinary WIMP scenarios, it is claimed that most of the earth-mass protohalos are stable against the tidal stripping [46]. It is also discussed whether those halos are disrupted by interaction with stars [47]. On the one hand, subhalos that cross the galactic disk nearly perpendicularly or that fall off to the center of the parent galaxy are strongly disrupted by the tidal force, leading to a significant decrease in the total mass. On the other hand, subhalos that orbit on the plane nearly parallel to the disk or that reside in the low density region survive more-or-less intact.

For simplicity, we have assumed that the observed mass scale of the subhalo is equivalent to the lower bound on the protohalo mass scale. In practice, however, the observable mass

---

4 The properties of the sterile neutrinos can also be constrained from the measurement of the X-ray flux, since they can contribute to the X-ray flux due to radiative decay [42]. See also Refs. [39,43].
using gravitational lensing is limited to the one within a certain radius centered at the line of sight. As a result, there remains a certain ambiguity in estimating the mass scale of the observed subhalo. Observation of astrometric shifts of lensed QSO images with a substructure in the surface brightness may help to reconstruct the subhalo mass density profile, thereby reducing the ambiguity [10].

It has been argued that the superWIMP and WDM scenarios can resolve the small-scale $\lesssim 1$ Mpc discrepancies, such as the “missing satellite problem” and the “cusp problem” if the damping scale is as large as $(0.4-1.0)$ Mpc. Therefore, future submillilensing experiments will shed a new light on these small-scale structure problems once the above ambiguities are removed.

Acknowledgements

We would like to thank Masashi Chiba for useful comments. This work is supported in part by the Grant-in-Aid for Science Research, Ministry of Education, Science and Culture, Japan (Nos. 18034002 and 15540255 for J.H.).

References

A. Boyarsky, A. Neronov, A. Neronov, O. Ruchayskiy, M. Shaposhnikov, astro-ph/0603568;
Study of leading hadrons in gluon and quark fragmentation

DELPHI Collaboration

Abstract

The study of quark jets in $e^+e^-$ reactions at LEP has demonstrated that the hadronisation process is reproduced well by the Lund string model. However, our understanding of gluon fragmentation is less complete. In this study enriched quark and gluon jet samples of different purities are selected in three-jet events from hadronic decays of the $Z$ collected by the DELPHI experiment in the LEP runs during 1994 and 1995. The leading systems of the two kinds of jets are defined by requiring a rapidity gap and their sum of charges is studied. An excess of leading systems with total charge zero is found for gluon jets in all cases, when compared to Monte Carlo simulations with JETSET (with and without Bose–Einstein correlations included) and ARIADNE. The corresponding leading systems of quark jets do not exhibit such an excess. The influence of the gap size and of the gluon purity on the effect is studied and a concentration of the excess of neutral leading systems at low invariant masses ($\lesssim 2\text{ GeV}/c^2$) is observed, indicating that gluon jets might have an additional hitherto undetected fragmentation mode via a two-gluon system. This could be an indication of a possible production of gluonic states as predicted by QCD.

1. Introduction

The study of quark jets provides us with remarkable insights into the mechanism of hadronisation. It gives strong evidence for chain-like charge ordered particle production in excellent agreement with string Monte Carlo models like JETSET [1]. This is shown e.g. by several contributions [2–5] of the DELPHI experiment at LEP, where the compensation of quantum numbers, in particular that of charge, has been extensively studied. Much less is, however, known about the behaviour of gluon jets. On the theoretical side, besides the fragmentation via two strings as implemented in JETSET/PYTHIA and ARIADNE [6], the direct neutralisation of the colour octet field by another gluon with the creation of a two-gluon system has been considered by Minkowski and Ochs [7,8] and also by Spiesberger and Zerwas [9]. Older references exist by Montvay [10] and Peterson and Walsh [11]. Additional references can be found in [7] where it is also emphasised that an experimental study of the gluon corner in three-jet events could contributevaluably to the question of the existence of glueballs, an early expectation of QCD [12]. No quantitative prediction however exists up to now. This has triggered an experimental investigation by the DELPHI Collaboration on gluon fragmentation in a leading system defined by a rapidity gap [13,14]. The preliminary results revealed that electrically neutral systems of leading particles in gluon jets occur more often than predicted by JETSET, in agreement with the expectations of the above theoretical arguments, while there was no disagreement observed in quark jets. This phenomenon, experimentally observed for the first time, has meanwhile also been seen by ALEPH and OPAL [15,16].

The JETSET (ARIADNE) model of a $q\bar{q}g$ event stretches a string from the $q$ to the $g$ and on to the $\bar{q}$. The string fragments for example by the creation of $q\bar{q}$ pairs, similar to what happens for quark fragmentation (Fig. 1(a)). Thus the JETSET (and ARIADNE) model regards gluon fragmentation as a double colour triplet fragmentation (most clearly sketched in Fig. 1 of Ref. [7]) and the leading system can obtain the charge ±1 or 0 in the limiting configuration. The process proposed by Minkowski and Ochs, namely the octet neutralisation of the gluon field by another gluon has the signature of an uncharged leading system due to the requirement that the sum of charges ($S^Q$) of the decay products of a two-gluon system is zero (Fig. 1(b)). In [7,8] it is also proposed to enhance the possible
contribution of this process by selecting events where a leading particle system is separated from the rest of the low energy particles by a large rapidity gap, empty of hadrons. In this situation of a hard isolated gluon the octet field is expected not to have been distorted by multiple gluon emission and by related colour neutralisation processes of small rapidity ranges [7]. The price to pay for such a selection is, however, a strong reduction of the number of events because of the Sudakov form factor [17].

A different mechanism—colour reconnection [18]—can produce similar effects. Two experiments, however, agree that the present colour reconnection models, as implemented in some versions of Monte Carlo simulations, cannot reproduce quantitatively the observed excess of $S\Omega = 0$ systems [15,19].

The present study aims to consolidate the results of the preceding analyses [13–16] by studying the dependence of the excess of neutral leading systems in enriched gluon jets on the gap size and gluon content, by studying their mass spectra and by investigating if there are possible trivial origins for the observed effect. This is especially important, since a significant failure of the string model to describe gluon jets might generally reveal the presence of hitherto undetected processes. The size of the effect for a pure gluon jet is estimated. As a cross-check, the same investigation is done for quark jets.

2. Data sample and 3-jet event selection

The data sample used has been collected by the DELPHI experiment at the LEP collider at the peak of the $Z$ resonance during 1994 and 1995. Three-jet events have been selected by using the appropriate cuts for track quality and for the hadronic event type [20] as well as applying a $k_t$ cluster algorithm (Durham) [21] to all observed charged and neutral particles with $y < 3.5$. The hadronic event type as well as applying a selected by using the appropriate cuts for track quality and for

![Fig. 1. Diagrams to illustrate the processes of colour triplet fragmentation (a) and colour octet fragmentation (b). The dashed lines represent the colour triplet strings and the helixes represent the colour octet strings.](Image)

consist of $\geq 2$ particles and the jets must be at least $30^\circ$ away from the beam direction [22–24]. About 314000 events meet all these conditions.

Without any additional tag the jet with the highest energy $E_1$ (jet1) is in most cases a quark jet and that with the smallest energy $E_3$ (jet3) the gluon jet. The measured mean jet energies are: $E_1 = 41.4$, $E_2 = 32.2$ and $E_3 = 17.7$ GeV. In the first data sample (sample 1), where the gluon and quark jet identification is based on energy ordering only, events are required not to exhibit any $b$-signal (235 080 events). Monte Carlo simulations show for the above mentioned conditions a quark jet contribution of about 90% for jet1 and a gluon jet contribution of about 70% for jet3. In a more detailed study of the gluon purity a second independent sample (sample 2) is selected, where jet1 and jet2, contrary to jet3, are required to exhibit a $b$-signal [24,25] (Section 4.3). This additional tag results in a gluon purity of jet3 of about 90% and consists of 31 400 events. A third sample (sample 3) is selected to enable purity unfolding for special cases. It is defined by the requirement that jet3 has a $b$-tag. For this jet3 sample consisting of 12 200 events the gluon content is very much diminished (about 26%).

3. Monte Carlo models

For comparisons a suitable number of Monte Carlo simulations using JETSET 7.3 [1] and ARIADNE [6] have been performed. In contrast to JETSET, ARIADNE incorporates dipole radiation of gluons instead of the parton shower used by JETSET. Since Bose–Einstein correlations (BEC) are present in nature, like-charged particles will stick together in momentum space and local charge compensation is expected to be diminished. The implementation of Bose–Einstein correlations into the Monte Carlo simulation, however, is highly problematic and the magnitude of the effect on charge compensation is unknown. Nevertheless, the possible effect of BEC has to be investigated and the possible uncertainties have to be considered.

Three different Monte Carlo event samples have been created by using different generators:

- Model (1): JETSET with BEC included (BE32 [26]);
- Model (2): JETSET without BEC;
- Model (3): ARIADNE without BEC.

The number of events generated for each sample corresponds roughly to that of the data.

The data are compared to these Monte Carlo event samples with full simulation of the DELPHI detector. The same reconstruction and analysis chain has been applied to the data and Monte Carlo (MC) samples.

---

1. This value has been obtained from a study optimising simultaneously purity and statistics [22].

2. Although the mean energies of jet1 and jet3 differ by more than a factor 2, the maximum possible rapidities and mean multiplicities for charged particles differ much less (e.g. $\langle n_{chjet} \rangle = 6.04$, $\langle n_{chjet} \rangle = 7.36$).

3. All quark jet selections (jet1) shown in the figures for comparisons are defined by sample 1.
4. Analysis

4.1. The sum of charges in the leading system with a rapidity gap (sample 1)

After the selection of 3-jet events and the determination of enriched quark and gluon jet samples, the leading system of a jet is defined by requiring that all charged particles assigned to the jet must have a rapidity \( y \) with respect to the jet axis of \( y \geq \Delta y \). The quantity \( \Delta y \) represents a lower limit and defines a rapidity gap extending at least up to \( y = \Delta y \). By this requirement, also jets are discarded, if they include some particles (fraction \( 10^{-3} \)) with negative rapidity. The size of the demanded gap below this leading system is a compromise between the requirement of a gap as large as possible and the considerable loss of statistics at a larger gap. The requirement that the rapidity interval \( \delta y \geq \Delta y \) (with \( \Delta y = 1.5 \)) below the leading system be empty of charged particles reduces the number of jets appreciably. About 38 000 enriched gluon jets and 39 000 quark jets meet this condition. This reduction rate \( f_1 \) of enriched gluon jets is quite well reproduced by the three Monte Carlo event samples: \( f_1(\text{data}) = 0.160 \), \( f_1(\text{MC1}) = 0.169 \), \( f_1(\text{MC2}) = 0.158 \), \( f_1(\text{MC3}) = 0.157 \), with the mean value \( f_1(\text{MC-mean}) = 0.161 \). In principle, there could be neutral hadrons in the gap. It has been verified that removing in addition topologies where observed neutrals are contained in the gap (mainly \( \gamma \)'s from the decay of \( \pi^0 \)'s), leads to results that are fully consistent with the ones presented here, but with about 15% larger statistical errors.

The sum of charges \( (S_Q) \) of the particles belonging to the leading system defined as above is shown in Fig. 2(a) for enriched gluon jets and in Fig. 2(b) for quark jets and compared to ARIADNE. \( P(S_Q, \Delta y) \) is generally defined as the fraction of a jet sample with a gap and a given value of \( S_Q \),

\[
P(S_Q, \Delta y) = \frac{N(S_Q, \Delta y)}{N(\Delta y)}
\]

and is an estimate for the probability of a jet with a gap to have a certain \( S_Q \). The \( S_Q \) distribution of the leading system for the gluon jet (Fig. 2(a)) exhibits for \( S_Q = 0 \) a significant enhancement of the data over the Monte Carlo. This effect is predicted, if the process of colour octet neutralisation is present [7,8]. On the other hand, there is no significant difference between the data and the Monte Carlo simulation in the case of quark jets (Fig. 2(b)).

The lower parts of Fig. 2 show quantitatively the differences of the \( P(S_Q, 1.5) \) between the data and the Monte Carlo simulation. This difference for the gluon jet (Fig. 2(c)) amounts to about 4 standard deviations (statistical errors only), for the quark jet (Fig. 2(d)) this difference is compatible with zero.

4.2. The dependence on the size of the rapidity gap

Fig. 3(a), (b) shows, for neutral leading systems \( (S_Q = 0) \), the dependence of the relative deviation \( R(\Delta y) \) on the size of

\[
R(\Delta y) = \frac{N(S_Q = 0, \Delta y)}{N(\Delta y)}
\]

and is defined in Eq. (2). (c), (d): \( R'(\Delta y) \) as defined in Eq. (3). Because of the nature of the cut \( \Delta y \), the bins are correlated.
the lower limit ($\Delta y$) of the rapidity gaps considered:

$$R(\Delta y) = \frac{P(0, \Delta y)_{\text{data}} - P(0, \Delta y)_{\text{MC}}}{P(0, \Delta y)_{\text{MC}}}.$$  \hspace{1cm} (2)

For all three types of Monte Carlo simulations (models (1), (2) and (3), see Section 3), $R(\Delta y)$ ($\Delta y > 0.5$) is positive and increasing with $\Delta y$ for jet3 (Fig. 3(a)). This clearly shows that the surplus of neutral leading systems in the data, compared to models. In the case of jet1 (Fig. 3(b)) all values are scattered essentially independent of $\Delta y$, together with the statistical errors (symbols with error bars) and (3), which are presented in Fig. 3(c) for sample 1, are drawn and shown in Fig. 3(c), (d). The residual spread between the models is considered as systematic uncertainty.

4.3. The dependence of $R'(\Delta y)$ on the gluon purity

In Fig. 4(a) the mean values of $R'$ with models (1), (2) and (3), which are presented in Fig. 3(c) for sample 1, are drawn together with the statistical errors (symbols with error bars) and systematic uncertainty (shaded area). As a cross-check, an independent second sample (sample 2) of gluon jets with a much higher purity is selected. The dependence of $R'(\Delta y)$ on $\Delta y$ for this sample is given in Fig. 4(b). Although the statistics are smaller in Fig. 4(b) (3870 jet3 at $\Delta y = 1.5$, which is only about 1/10 of sample 1), the effect is increased, which is expected if it is connected to the gluon jet only. At $\Delta y = 1.5R'(\Delta y)$ is about 0.09 ± 0.02 (statistical) because of the higher purity.

To estimate the amount of disagreement between data and Monte Carlo in pure gluon jets the gluon purity has to be estimated for jet3 at the gap size $\Delta y$ for both data selections in Fig. 4. In principle, it can be directly obtained from the Monte Carlo. At the same scale, gluon and quark jets exhibit different rapidity distributions, i.e. gluon jets emit more particles per unit at small rapidity. Demanding a gap reduces therefore not only the number of jets, but also the gluon content in a mixed sample of gluon and quark jets. This is observed in the MC. An estimation of the gluon purity at gap $\Delta y$ however depends on the correct modelling of the rapidity distribution of pure gluon jets. Therefore another method has been used in addition. It uses the measured reduction rates of the number of jets $f_1(\Delta y)$ in sample-1 by demanding a gap (see Section 4.1) and from the MC only the composition at gap zero. Let us define $N_1(\Delta y) = f_1(\Delta y)N_1(0)$ in sample 1, and $N_2(\Delta y) = f_2(\Delta y)N_2(0)$ in sample 2, where $N_1(\Delta y)$ ($N_2(\Delta y)$) is the number of jets counted at $\Delta y$ in sample 1 (sample 2). Since these samples are an admixture of pure $q$ (= light quark) jets, $g$ (= gluon) jets and $b (= b$-quark) jets, the corresponding $f_1(\Delta y)$, $f_2(\Delta y)$ (and also $f_3(\Delta y)$ for sample 3) are also an admixture of the reduction rates $f_g(\Delta y)$, $f_b(\Delta y)$, $f_b(\Delta y)$ of the pure light-quark, pure gluon and pure $b$-quark subsamples, e.g.

$$f_1(\Delta y) = a_{1q}f_q(\Delta y) + a_{1g}f_g(\Delta y) + a_{1b}f_b(\Delta y)$$  \hspace{1cm} (4)

with two analogous equations for $f_2(\Delta y)$ and $f_3(\Delta y)$. The resulting system of three linear equations can be written in short:

$$F = AF_{\text{pure}}$$  \hspace{1cm} (5)

with the solution

$$F_{\text{pure}} = A^{-1}F.$$  \hspace{1cm} (6)

The vector $F(f_1(\Delta y), f_2(\Delta y), f_3(\Delta y))$ is measured, and the vector $F_{\text{pure}}(f_g(\Delta y), f_b(\Delta y), b(\Delta y))$ is the solution. The matrix $A$ represents the $q$, $g$, $b$ compositions for the three selections at gap = 0, estimated from Monte Carlo (e.g. $a_{1g}$ is the gluon purity of jet3 in sample 1, $a_{2g}$ that of sample 2, and $a_{3g}$ that of sample 3 and so on). With the solution of Eq. (6), the numbers of true gluon jets at $\Delta y$ can be determined in sample 1 and sample 2:

$$N_1^{\text{gluon}}(\Delta y) = f_g(\Delta y)a_{1g}N_1(0),$$  \hspace{1cm} (7)

$$N_2^{\text{gluon}}(\Delta y) = f_g(\Delta y)a_{2g}N_2(0).$$  \hspace{1cm} (8)

The fraction of gluon jets $c_{g}^{\Delta y}$ at $\Delta y$ is given by:

$$c_{g}^{\Delta y} \text{ (sample 1)} = N_1^{\text{gluon}}(\Delta y)/N_1(\Delta y) = a_{1g}f_g(\Delta y)/f_1(\Delta y),$$  \hspace{1cm} (9)

$$c_{g}^{\Delta y} \text{ (sample 2)} = N_2^{\text{gluon}}(\Delta y)/N_2(\Delta y) = a_{2g}f_g(\Delta y)/f_2(\Delta y).$$  \hspace{1cm} (10)
Applied to the two data sets in Fig. 4(a), (b) the following numbers for the gluon content have been obtained:

(1) The sample in Fig. 4(a) (sample 1): $c_g^0 = 0.65$, $c_g^{1.5} = 0.46$ (from Eq. (9)), and $c_g^{1.5} = 0.45$ (directly from the Monte Carlo at $\Delta y = 1.5$).

(2) The sample in Fig. 4(b) (sample 2): $c_g^0 = 0.88$, $c_g^{1.5} = 0.80$ (from Eq. (10)), and $c_g^{1.5} = 0.82$ (directly from the Monte Carlo at $\Delta y = 1.5$).

The statistical errors on these numbers are below 1%, systematic errors can be obtained by comparing the estimates with different Monte Carlo models (1), (2) and (3). They are $\leq 2.6\%$.

The purity estimates obtained above allow the determination of the excess of neutral systems $R'_g$ in pure gluon jets by dividing by $c_g^{1.5}$. The following values of $R'_g$ have been obtained for the two samples defined above:

- Sample 1: $R'_g(1.5) = 0.100 \pm 0.012$ (stat) $\pm 0.025$ (syst).
- Sample 2: $R'_g(1.5) = 0.107 \pm 0.022$ (stat) $\pm 0.028$ (syst).

The following sources of systematic errors have been considered:

(a) Quality of event reconstruction. Bad reconstructions and losses of particles (mainly neutrals) in the detector and wrong assignments to the jets can lead to differences of several GeV between the jet energy calculated from the angles between jets ($E_{\text{calc}}$) [24] and the sum of energies of all particles assigned to the jet ($E_{\text{sum}}$). Improving the quality by cutting away about 1/3 of the jets with the largest difference $E_{\text{calc}} - E_{\text{sum}}$ does not significantly change the signals at $SQ = 0$ both in gluon and quark jets;

(b) The dependence of the effect on the polar angle of the jet with respect to the collision axis has been investigated: the effect is stable;

(c) The influence of track finding efficiency in the detector. In order to investigate the influence of track finding efficiency the effect of a reduction of the efficiency by 1% has been simulated. No significant change in the signals at $SQ = 0$ has been observed. Since $R$ is a ratio with respect to the Monte Carlo simulation (including the detector effects) and the deviation which has been observed between data and MC below 10%, it can be expected that efficiency effects cancel to a large extent;

(d) To investigate whether the good agreement between data and Monte Carlo in quark jets is only due to the larger particle momenta, in a test-run only particles with momenta less than 30 GeV/c have been accepted in jet1. The agreement with the Monte Carlo remains.

(e) The estimations leading to (11) and (12) assume that quark jets, also at the lower energies of jet3, do not exhibit any excess of neutral leading systems. This is further tested by measuring the excess in sample 3 which exhibits at $\Delta y = 1.5$ an admixture of only 20% gluon jets. As expected, the signal is reduced, and is even negative with large error: $R'_g(1.5) = −0.02 \pm 0.02$. Adopting the same procedure as in Section 4.3 by using matrix inversion with the measured values $R'_g(1.5)$, $i = 1, 2, 3$ for the 3 selected samples as input, the resulting excess for pure quark and gluon jets could be estimated: $R'_g(1.5) = 0.00 \pm 0.02$, $R'_g(1.5) = 0.11 \pm 0.03$ and $R'_g(1.5) = −0.06 \pm 0.04$. These results do not show any evidence that pure quark jets exhibit an excess of neutral leading systems for the lower jet3 energies;

(f) At the generator level of JETSET and for pure gluon jets the effect of changing parameters within limits has been studied. For example, different DELPHI tunings have been used, the DELPHI tuning [27] has been replaced by that of OPAL [28] and by the JETSET default, $^5$ and the popcorn parameter has been varied. Some changes of $P(SQ, \Delta y = 1.5)$ at $SQ = 0$ are revealed in gluon jets and to a lesser extent in pure quark jets. At a gap size of $\Delta y = 1.5$ a maximum variation of $R(\Delta y)$ of 0.027 is observed.

The systematic error from (f) amounting to 18% is taken into account. This is a conservative estimate with a factor 0.68 of the maximum variation, corresponding to 1 $\sigma$ of a Gauss distribution. The contributions from (a)–(e) are negligible. The systematic errors for samples 1 and 2 are estimated as follows:

- Sample 1
  (a) from the spread in Fig. 4(a) at $\Delta y = 1.5$: $\Delta R' = 0.017$;
  (b) uncertainty in purity: 0.0026 (see Section 4.3);
  (c) uncertainty from (f): 0.018.

$^5$ All these studies have been done with BE correlations included.
Fig. 5. (a), (c): Invariant mass distribution \( P(M) \) of the leading system \( SQ = 0 \) (considering only charged particles) (a) for gluon-enriched jets (sample 1), (c) for quark jets. \( P \) is defined in Eq. (15). The dots with error bars are the data, the histograms are the mean values of the three Monte Carlos. (b), (d): \( P_{data}(M) - P_{MC}(M) \). By quadratically adding all 3 contributions a systematic error of 0.025 is obtained.

Sample 2
(a) from the spread in Fig. 4(b) at \( \Delta y = 1.5; \Delta R' = 0.021 \);
(b) uncertainty in purity: 0.0028 (see Section 4.3);
(c) uncertainty from (f): 0.019.

By quadratically adding all 3 contributions a systematic error of 0.028 is obtained.

5. Mass spectra

Colour octet neutralisation of the gluon field could produce a resonance spectrum which differs from that of colour triplet fragmentation [29] implemented in JETSET. In order to investigate in which region of the mass spectrum the observed excess of the leading neutral systems is located, the invariant mass \( (M) \) distributions

\[
P(M) = \frac{N(M, SQ = 0, \Delta y)}{N(\Delta y)}
\]  

(15)

of the leading systems with total charge zero at \( \Delta y = 1.5 \) have been calculated and compared with the mean values of models 1–3 for the two cases:

(a) \( M \) is computed using only charged particles with a momentum \( p \geq 0.2 \text{ GeV}/c \) and assuming pion mass. This distribution is shown in Fig. 5(a) for gluon enriched jets and in Fig. 5(c) for quark jets. Two broad bumps can be observed in Fig. 5(a), (c) which are the result of a superposition of a rapidly decreasing two-particle spectrum which dominates for \( M \lesssim 1 \text{ GeV}/c^2 \), with an increasing spectrum consisting of 4 and more particles. The latter dominates and peaks at \( \sim 1.5 \text{ GeV}/c^2 \). The region below 0.8 \text{ GeV}/c^2 consists only of two particles. One peak around \( M \sim 0.8 \text{ GeV}/c^2 \) can be attributed to the \( \rho \) resonance, another at \( M \lesssim 0.5 \text{ GeV}/c^2 \) to a reflection of \( \eta, \eta' \) and \( \omega \). The latter statement is corroborated by the fact that in events with no neutrals, the peak at \( M \leq 0.5 \text{ GeV}/c^2 \) vanishes. Only in Fig. 5(a) a third peak is indicated by the data points just below 1 GeV/c^2, in the region of the \( f_0(980) \) resonance. In this region \( 0.9 \leq M \leq 1 \text{ GeV}/c^2 \) the two particle contribution amounts to about 70%. Fig. 5(b), (d) shows the difference between the data and the Monte Carlo event sample. For gluon enriched jets the distribution in Fig. 5(b) exhibits possible evidence for a mass enhancement\(^6\) in the region \( 0.9 \leq M \leq 0.95 \text{ GeV}/c^2 \) but no signal is seen in this region for quark jets in Fig. 5(d). This peak is the remaining part of the original peak in Fig. 5(a) after the subtraction of a small and narrow but significant signal in the MC at 0.95–1.0 \text{ GeV}/c^2. Without emphasising too much this remaining narrow peak in Fig. 5(b), it has to be noted that it survived a quality cut (by accepting only jets with a polar angle \( \geq 50^\circ \)) well above 3\( \sigma \), whereas all other deviations from zero in isolated bins in Fig. 5(b), (d) were decreased. Whether this narrow signal in gluon enriched jets can be attributed to \( f_0(980) \) production remains an open interest-

---

\(^6\) The experimental mass resolution is below 10 MeV/c^2, the bin width in Fig. 5 is 50 MeV/c^2.
Fig. 6. (a), (d): Invariant mass distributions $P(M)$ of the leading system ($\mathcal{Q} = 0$) (with charged and neutral particles) for gluon-enriched jets, samples 1 and 2 respectively. (g): Same for quark jets (jet1, sample 1). The dots with error bars are the data, the histograms are the mean values of the three Monte Carlos. The quantity $P(M)$ is defined in Eq. (15). Second row (b), (e), (h): $P_{\text{data}}(M) - P_{\text{MC}}(M)$; Last row (c), (f), (i): $P_{\text{data}}(M) - P_{\text{MC}}(M)$ for 2 bins: (0.25–2 GeV/$c^2$) and (2–3.75 GeV/$c^2$).

An important remark concerns all mass spectra in Fig. 6. As stated in Section 3, all the comparisons to Monte Carlo event samples are done with full detector simulation. Consequently, due to the loss of neutral particles, all spectra and in particular the excess spectra in Fig. 6(b), (e) are shifted by about 0.3 to 0.5 GeV/$c^2$ to lower mass values. It has been verified, however, by a special Monte Carlo study using the detector response matrix (not shown here) that, after the correction of all spectra for this shift, the excess is still clearly concentrated in the low mass region. On the other hand, the spectra of leading systems consisting only of charged particles (Fig. 5) are not affected by shifts.

The observations in Fig. 5 confirm the first preliminary results presented in 2001 [13] for leading gluonic systems by considering charged particles only. In 2002 the OPAL Collaboration reported [16] a $2\sigma$ excess in the mass distribution of neutral leading systems, consisting of charged and neutral particles, between 1 and 2.5 GeV/$c^2$ in gluon jets.

The observation that the excess of neutral leading systems in gluon jets is limited to the low mass region (Figs. 5 and 6) supports arguments in favour of gluonic states. The
existence of glueballs, i.e. bound states of two or more gluons, is a prediction of QCD [12]. The experimental results and their interpretations, however, are still controversial [30]. Theoretically there is general agreement that the lightest glueball should be in the scalar channel with $J^{PC} = 0^{++}$. Quantitative results are derived from the QCD lattice calculations [31] which predict the lightest glueball to be around 1600 MeV/$c^2$, or from QCD sum rules [32], which predict also a possible gluonic state near 1 GeV/$c^2$. Alternatively, it could also be a very broad object [33,34]. The gluonic state could mix with ordinary $0^{++}$ states, like the $f_0(980)$, $f_0(1370)$, $f_0(1500)$ or $f_0(1710)$. As an example there are two scenarios, where the largest gluonic component is included in the $f_0(1500)$ [35], or alternatively in the $f_0(1710)$ [36]. Recent discussions with various references can be found in [30,37].

6. Summary

In the present study the leading systems defined by a rapidity gap have been investigated for gluon and quark jets. The statistics of 1994 and 1995 at $\sqrt{s} = 91.2$ GeV obtained by the DELPHI Collaboration is used to select 3-jet events and to single out quark jets (purity $\sim 90\%$) and gluon enriched jets (purity $\sim 70\%$) by energy ordering (sample 1). For the (enriched) gluon jets a higher rate of neutral leading systems than predicted by the Lund string model JETSET/ARIADNE (with and without Bose–Einstein correlations) is observed but no such enhancement is seen for the quark jets. Various checks have been performed which suggest that this effect is not a spurious one. An increase of the effect with increasing gluon purity, obtained by a tagging procedure in a second sample (sample 2), is observed corroborating that it is indeed connected with the gluon jets.

The excess of neutral leading systems in pure gluon jets at a gap size $\Delta y = 1.5$ has been measured to be about 10%, with a significance of 3.6σ. It is expected to be of the order of 0.5% in pure gluon jets without any charge or gap selection.

The mass spectra of the neutral leading systems of gluon jets, both with and without including neutral particles have been studied. Mass spectra which include charged and neutral particles, show clearly that the excess mentioned above is concentrated at low invariant masses ($\lesssim 2$ GeV/$c^2$). The significance is enhanced there and amounts to about 5σ (statistical) in sample 1 and the excess is increased roughly proportionally to the gluon purity in sample 2.

The corresponding mass spectra of leading systems in quark jets do not exhibit any excess in the low mass regions.

The observed excess of neutral systems in gluon jets and its increase with the gap size and with the gluon purity is in agreement with expectations, if the hitherto unobserved but predicted process of octet neutralisation of the gluon field takes place in nature. Although colour recombination could in principle alternatively explain the excess, the specific mass concentration at low mass seems to favour the first case and could be a signal of gluonic states predicted by QCD.

Acknowledgements

We thank W. Ochs for encouraging us to start the above study and for discussions in the course of it and O. Klapp for programming support. We thank P. Minkowski, G. Rudolph and T. Sjöstrand for valuable discussions and clarifications. Technical support by G. Walzel is gratefully acknowledged. We are greatly indebted to our technical collaborators, to the members of the CERN-SL Division for the excellent performance of the LEP collider, and to the funding agencies for their support in building and operating the DELPHI detector. We acknowledge in particular the support of: Austrian Federal Ministry of Education, Science and Culture, GZ 616.364/2-III/2a/98, FNRS–FWO, Flanders Institute to encourage scientific and technological research in the industry (IW), Belgium, FINEP, CNPq, CAPES, FUJB and FAPERJ, Brazil, Czech Ministry of Industry and Trade, GA CR 202/99/1362, Commission of the European Communities (DG XII), Direction des Sciences de la Matière, CEA, France, Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie, Germany, General Secretariat for Research and Technology, Greece, National Science Foundation (NWO) and Foundation for Research on Matter (FOM), the Netherlands, Norwegian Research Council, State Committee for Scientific Research, Poland, SPUB-M/CERN/PO3/DZ296/2000, SPUB-M/CERN/PO3/DZ297/2000, 2P03B 104 19 and 2P03B 69 23 (2002–2004) FCT—Fundação para a Ciência e Tecnologia, Portugal, Vedecka grantova agentura MS SR, Slovakia, Nr. 95/5195/134, Ministry of Science and Technology of the Republic of Slovenia, CICYT, Spain, AEN99-0950 and AEN99-0761, The Swedish Research Council, Particle Physics and Astronomy Research Council, UK, Department of Energy, USA, DE-FG02-01ER41155, EEC RTN contract HPRN-CT-00292-2002.

References


A consistent description of neutron stars with quark cores

Vikram Soni a,*, Dipankar Bhattacharya b

a National Physical Laboratory, K.S. Krishnan Marg, New Delhi 110012, India
b Raman Research Institute, Bangalore 560080, India

Received 5 August 2006; received in revised form 29 September 2006; accepted 2 October 2006
Available online 11 October 2006

Editor: M. Cvetić

Abstract

We consider the equation of state of hadroninc and quark matter at finite density in mean field theory, through an effective chiral Lagrangian whose parameters (coupling constants) are all fixed by hadronic data. Between three to seven times nuclear density, for charge neutral quark matter in \( \beta \) equilibrium, we find the ground state to be a neutral pion condensate. With increasing baryon density we then expect nuclear matter, followed by pion condensed quark matter at intermediate density, and finally the diquark colour-flavour CFL condensate. These are all states with chiral spontaneous symmetry breaking (SSB). We find another remarkable feature and this is that the scalar (pseudoscalar) coupling, \( \lambda \), has a crucial and unexpected influence on the physics of neutron stars. Neutron stars with pion condensed quark matter cores exist only in a small window, between \( 5.7 < \lambda < 6.45 \). Interestingly, this range is consistent with the value of \( \lambda \) derived from \( \pi, \pi \) scattering data and such stellar cores may carry magnetar strength magnetic fields.

© 2006 Elsevier B.V. All rights reserved.

1. Introduction

Neutron stars have been a subject of abiding interest for several decades. There are a variety of astrophysical phenomena that arise from the physics of neutron stars. These include supernovae, pulsars, accreting binary X-ray sources and magnetars, which have super-strong magnetic fields. Most of these phenomena require us to understand the physics of matter at very high density, which governs the mass and the size of neutron stars. In other words, one needs to have a clear understanding of the equation of state (EOS) of superdense matter. Although much effort has gone into this enterprise over the last four decades it still remains poorly understood. Why?

Central densities of neutron stars are high, more than \( \sim 5 \) times nuclear density \( \rho_{\text{nuc}} = 0.17 \text{ fm}^{-3} \). For a single species, neutrons, this naively translates into a Fermi gas with typical Fermi momentum, \( k_F \sim 600 \text{ MeV} \). On the other hand nucleons have structure and a typical size of the order of a Fermi (200 MeV)\(^{-1}\). It is clear that at such high densities nucleons (neutrons) cannot be treated as elementary. They are composite and resolved. They are colour singlet bound states of three valence quarks. At such high densities, therefore, treating nucleons as point particles interacting via two body (or more) forces will be inadequate. Yet most available equations of state adopt this approach and therefore fail to capture the correct physics at high density.

On the other hand, if we use quarks as the elementary degrees of freedom, we are presently bound by the fact that only perturbative calculations can be done for QCD. This implies that calculations can be done in QCD only at very high density when the theory is approximately in an asymptotically free (AF) phase. However, at intermediate and low density (close to nuclear density), where a nucleonic description is valid, we cannot use perturbative QCD as the coupling becomes strong and the physics nonperturbative and intractable. This is the dilemma.

There are attempts to model the physics by a two phase structure—a quark matter core with a hadronic/nuclear exterior shell and crust. Since there is no simple way to link the two phases without using separate parameters for both, this description is somewhat arbitrary. Further, the nature of the quark
matter state is not clear—for example, if it is in a spontaneous chiral symmetry broken state.

Can we find a single theory that connects both these domains? For this we use an effective chiral Lagrangian, \( L \), that receives broad support from many contexts. This Lagrangian has quarks, gluons and a chiral multiplet of \([\bar{\pi}, \sigma]\) that flavour-couples only to the quarks [1–6].

\[
L = -\frac{1}{4} G_{\mu \nu} G^{\mu \nu} - \sum \bar{\psi} (\not{\partial} + g_\gamma \not{\gamma} + g_\sigma \not{\sigma} \cdot \vec{\pi}) \psi \\
- \frac{1}{2} (\partial_\mu \sigma)^2 - \frac{1}{2} (\partial_\mu \bar{\pi})^2 - \frac{1}{2} \mu^2 (\sigma^2 + \bar{\pi}^2) \\
- \frac{\lambda^2}{4} (\sigma^2 + \bar{\pi}^2)^2 + \text{const.}
\]

The masses of the scalar (pseudoscalar) and fermions follow from the minimization of the potentials above. This minimization yields

\[ \mu^2 = -\lambda^2 \langle \sigma \rangle^2. \]

It follows that

\[ m^2_\pi = 2\lambda^2 \langle \sigma \rangle^2. \]

Experimentially, in vacuum, \( \langle \sigma \rangle = f_\pi \), the pion decay constant. This theory is an extension of QCD in that it additionally couples the quarks to a chiral multiplet, \([\bar{\pi}, \sigma]\) [1–4].

We now summarize some interesting physics that supports this Lagrangian at the mean field level (MFT) [4,7].

In this chiral, effective \( L \), we do not consider vector gluon mean fields which would spontaneously break colour symmetry and Lorentz invariance. Thus, at the level of mean field theory (MFT) our model reduces to a linear sigma model with quarks. As it stands, there is no confinement in this model at the level of MFT, but may be dynamically generated in the full theory as in QCD.

(i) It provides a model in which the nucleon is realized as a soliton with quarks being bound in a skyrmion configuration for the chiral field expectation values (EV) [1,4,7]. This model with composite nucleons gives a good account of the static properties of nucleons and nucleon–nucleon interaction potentials [8]. The model provides a natural explanation for the ‘Proton spin puzzle’ [9]. Such a Lagrangian also seems to naturally produce the Gottfried sum rule [10]. The Nambu–Jona-Lasinio model, that can be recast as a chiral sigma model [11], at the level of MFT, also yields a quark soliton nucleon. This gives structure functions for the nucleon which are close to the experimental ones.

(ii) In a finite temperature mean field theory such an effective Lagrangian also yields screening masses that match with those of a finite temperature QCD simulation with dynamical quarks [12]. This work does not show any parity doubling for the hadronic states.

(iii) The theory gives a qualitatively consistent description for the transition from hadronic matter to quark matter at high density and temperature [4,5,13,14].

This \( L \) has a single dimensional parameter, \( f_\pi \), that is the pion decay constant, and three couplings, \( g_3 \), the QCD coupling, \( g_\gamma \), the Yukawa coupling between quarks and mesons, that will be determined from the nucleon mass and the meson–meson coupling, \( \lambda \), which, for this model, can be determined from meson–meson scattering [15]. No further phenomenological input will be used.

Once the couplings of \( L \) are determined from the hadronic sector, the same effective Lagrangian describes the physics of the quark matter sector.

We now consider the question of the scales in QCD and the scale of validity of this effective \( L \).

(i) Constituent quarks vs. current quarks: to begin with let us consider the two main features of the strong interactions (QCD) at low energy. These are (a) that quarks are confined as hadrons and (b) chiral symmetry is spontaneously broken (SSB) with the pion as an approximate Goldstone boson. There is no specific reason that these two phenomena should occur at an identical temperature scale, though QCD lattice simulations show that for flavour \( SU_3(L) \times SU_3(R) \) they are close. The problem in giving an unequivocal answer to this question is that we are yet to find a solution to many of the nonperturbative aspects of QCD.

An interesting question arises: Is the quark matter in a chiral SSB state with constituent quarks or is it, as is usually assumed, in a chirally restored state with current quarks?

At finite density [6,13], we shall see in what follows, that quark matter is in a chiral SSB state. Also, if the chiral symmetry restoration (energy/temperature) scale was lower than the confinement scale we would expect hadrons to show parity doubling below the confinement scale but above the chiral SSB scale. This is not seen in finite temperature lattice simulations.

(ii) Compositeness scale: actually, QCD can have multiple scales [5]. Apart from a confinement scale and a chiral symmetry restoration scale we also have a compositeness scale for the pion. We find, somewhat in analogy with the top quark (large Yukawa coupling) composite Higgs picture, that we can get a compositeness scale for the scalars (pseudoscalars) in this model by using renormalization group (RNG) evolution. This is given by the scale at which the wavefunction renormalization for the mesons—the coefficient of the kinetic term—vanishes. Once this scale vanishes the meson fields are no longer bona fide degrees of freedom and can be eliminated using their field equation. We find that this scale, for the mesons, is inversely proportional to the running Yukawa coupling and thus naively vanishes when the Yukawa coupling blows up. For our theory such a ballpark scale falls between 700–800 MeV This also gives us an approximate idea of the range of validity of our effective Lagrangian, as, above the compositeness scale we lose the meson degrees of freedom.

An independent approach in setting a limit to the range of validity of nonasymptotically free (e.g. Yukawa) theories, like
ours, is the vacuum instability to small length scale fluctuations (or large momenta in quantum loop corrections), discovered by one of us [18]. The scale at which this occurs is of the same order as above. This is not very surprising since it is connected to non-AF character of the Yukawa coupling [18,19].

This discussion is to support the use of our effective Lagrangian up to a threshold scale in energy—the compositeness scale.

Given these facts we use the mean field theory to describe quark matter in the density regime bounded from above by the compositeness scale.

The plan of the Letter is as follows. In Section 2 we discuss how we fix the couplings for our Lagrangian from the hadronic sector and comment on the description of nuclear matter in this model. In Section 3 we present the equation of state (EOS) for 3 flavour quark matter using the lowest energy ground state we have found—the neutral pion condensed phase with chiral SSB. In Section 4 we use our quark matter EOS and the EOS for nuclear matter given by Akmal et al. (APR) [20] to make neutron stars. In Section 5 we discuss the phase diagram of QCD at finite density. We comment on other ground states and on the comparison between the pion condensed phase with the colour superconducting phase. Section 6 summarizes our results.

2. Preliminaries

2.1. The couplings of $L$

The nucleon, in our linear sigma model with quarks, is a colour singlet bound state of three valence quarks in a Skyrme background. The general solution follows on varying the Skyrme configuration $\pi$ and $\sigma$ and quark fields, that occur in the soliton field equation, independently. The quark soliton so obtained is projected to give a nucleon with good spin and isospin [1b].

The mass, $M$, of the nucleon depends just on $f_\pi$ (which is 93 MeV), $g_\gamma$ and $\lambda$. The dependence on $\lambda$ is marginal, so the only parameter that the mass depends on is the Yukawa coupling, $g_\gamma$. Fixing $M = M_{\text{nucleon}}$, yields a generally accepted value for $g_\gamma = 5.4$ [1b].

Recently, Schechter et al. [15] made a fit to scalar channel scattering data to see how it may be fitted with increasing $\sqrt{s}$ (centre of mass energy), using chiral perturbation theory and several resonances. They, further, looked at this channel using just a linear sigma model. Their results indicate that for centre of mass energy $\sqrt{s} < 800$ MeV, a reasonable fit to the data can be made using the linear sigma model with a ‘tree’ level sigma mass close to 800 MeV, or equivalently, $\lambda \sim 6$. We may add that the ‘tree’ level mass is a parameter for the ‘tree’ level $L$, and is equivalent to setting a value for $\lambda$—it is not the physical mass of the sigma (see [15] for details).

The range of validity of the linear sigma model is also consistent with the range of validity of our effective $L$ from the hadronic sector.

2.2. The nucleon or nuclear phase

In this phase, as the name suggests, the quarks are to be found as bound states in nucleons. Nucleon–nucleon interaction potentials like one pion exchange, tensor, etc., have been obtained from two skyrmion configurations [16]. These potentials have also been calculated for our quark solitons [8]. The simplest way to do nuclear matter is to use these potentials and to map on to nuclear physics as done with conventional nucleons.

Since we have a good description of nucleons from our theory we can in principle construct nuclear matter in terms of composite quark solitons. However, to construct the ground state of nuclear matter from composite nucleons is, to say the least, a formidable exercise. and can only be done using crude approximations.

More as a pedagogical exercise, we now indicate how we may look at nuclear matter. As the density increases we expect that the nucleons are no longer free to move around. If we make the assumption that they get localized into a “crystal” (of quark soliton nucleons) like configuration, we can then use a Wigner-Seitz approximation to convert to a single cell problem. Instead of imposing on the reader, we have moved this calculation to a preprint [17]. The EOS this yields [14,17] has many qualitatively correct features but nevertheless is far too stiff compared to most standard EOSs. For example, the APR [20] EOS. This underscores the difficulties in making an accurate many body calculation with our solitons.

It is then more reasonable to use the $N–N$ potentials in our model and map on to a potentials based many body nuclear matter calculation. We find it judicious to simply use a standard nuclear equation of state, like [20] to describe the nuclear matter sector.

At density higher than overlap, the nucleons dissolve into quarks. As stated earlier, our quark based, $L$, can be used to describe the quark matter ground state below.

3. The ground state for 3 flavour quark matter

For neutron stars it is the ground state of charge neutral quark matter in $\beta$ equilibrium at given density that is needed for the equation of state. We find, in what follows, the ground state to be the neutral pion condensate. What is also new is that the couplings of $L$ are fixed from the hadronic sector.

For the equation of state we need to calculate the free energy density at a given density. We first review the calculation for the energy density from a previous paper [6], which will be used to calculate the free energy.

In [6] we considered different patterns of symmetry breaking for the, $(\pi^+ \pi^- \sigma)$, fields and calculated the respective ground state energies. In particular, we considered two cases:

(i) The two and three-flavour (see below), $\pi_0$, pion condensed phase, where the $(\pi^+ \pi^- \sigma)$ expectation values are in a stationary wave configuration, with a wavevector, $\vec{q}$ (see below).
(ii) The space uniform symmetry broken state, which follows on putting $\vec{q} = 0$. At high density this state goes through a chiral restoration and essentially resembles conventional strange quark matter (SQM), that is chirally restored quark matter (CRQM).

The $\pi_0$ condensed ground state (PC) is found to have the lowest energy in the chiral limit $[6,13]$ (see Table 1). We refer the reader to $[6]$, where we have set up the machinery to describe this ground state which is built on nontrivial symmetry breaking in the presence of the $\pi_0$ condensate. This is given as follows,

For the $SU(3)$ flavour case we have a singlet $\xi_0$ and an $SU(3)$ octet $\xi_a$ of scalar fields and a singlet $\phi_0$ and an $SU(3)$ octet $\phi_a$ of pseudoscalar fields, with the expectation values $[6]$,

$$\langle \xi_0 \rangle = \sqrt{3/2} F (1 + 2 \cos(\vec{q} \cdot \vec{r})) / 3,$$

$$\langle \xi_a \rangle = -\sqrt{3} F (1 - \cos(\vec{q} \cdot \vec{r})) / 3,$$

$$\langle \phi_0 \rangle = 0,$$

$$\langle \phi_a \rangle = F (\sin(\vec{q} \cdot \vec{r}))$$

while all other fields have zero expectation value. On putting $\vec{q} = 0$, we get the vacuum (space uniform) symmetry broken state. This yields the simple mass relation for the strange quark, $M_s = g_s F + m_s$, where $m_s$ is the current mass and, $F = \sqrt{(\bar{r})^2 + (\sigma)^2}$, is the chiral order parameter ($F = f_N$, at zero density).

The ground state energy is obtained by summing all occupied single quasiparticle states, in the presence of the pion condensate, for the $u$ and $d$ quarks up to their Fermi energy. The quasiparticle states in the presence of this condensate have a spin-isospin alignment which gives the ground state a magnetic dipole moment. To this we add the sum over the plane wave states for the strange quarks of mass, $M_s$, up to the Fermi energy. Besides, we have, the gradient energy and the potential functional contributions from the meson sector. Charge neutrality requires us to include electrons as well. $\beta$-equilibrium is imposed and this implies several chemical potential relations between the different species (see [6]).

The ground state energy and the baryon density depend on the two variational parameters, the order parameter or the expectation value, $F$, and the condensate momentum, $[\vec{q}]$.

Ref. [6] provides the expressions for the baryon density, $n_b$, and the total energy density, $\epsilon$, of the PC in terms of the $u$, $d$, $s$ quark Fermi energies/chemical potentials.

For the EOS we need to construct the Gibbs free energy at a fixed baryon density

$$\Omega = \epsilon - n_b \mu_b.$$  \hfill (8)

The baryon chemical potential is defined as

$$\mu_b = \partial \epsilon / \partial n_b.$$  \hfill (9)

After meeting all the neutrality and equilibrium conditions above for fixed $F$ and $q$, we can write all the above variables as a function of a single variable, $\mu_u$.

We then minimize $\Omega$ independently with respect to $F$ and $q$. The energy per baryon, $E_b$, etc., then follow.

The results are presented in the table and in Fig. 1(a). Fig. 1(b) gives the EOS for all the phases considered so far.

As promised in the introduction

(i) We note that the SSB neutral pion condensed ground state is always lower energy than the CRQM state, indicating that quark matter is in the chiral symmetry broken state.

(ii) As is clear from the table we have stayed in range of validity, $\mu < 700$ MeV, of our effective L.

(iii) Another feature of this $\pi_0$ condensate is that we have a spin isospin polarization, ‘$u$’ and ‘$d$’ quark quasiparticles have opposite spin and opposite charge. We can then get a net magnetic moment in the ground state, as the magnetic moments of the $u$ and $d$ quarks add $[6]$.

4. Neutron stars from our EOS

In order to construct stars with this pion-condensed (PC) matter, we note that this state is thermodynamically stable only at densities above twice the nuclear density, so it is necessary to extend the equation of state to lower densities by interfacing with a nuclear equation of state. To describe the nuclear
Fig. 1. (a) Energy per baryon vs. baryon number density for 3-flavour pion-condensed phase for three values of assumed tree-level mass of the scalar meson $\sigma$. Charge neutrality and beta equilibrium are imposed. One-gluon exchange interaction is included using the prescription of Baym [21]. (b) Comparison of the equations of state of the 3-flavour space uniform phase and pion-condensed phase for $m_\sigma = 800$ MeV, with APR [20].

To construct the neutron star, we solve the Tolman–Oppenheimer–Volkoff (TOV) hydrostatic equilibrium equation [26] with the above equation of state. For a given $m_\sigma$, the PC core exists only if the central density of the star exceeds the APR-PC transition, which would happen above a threshold stellar mass $M_T$. With increasing $m_\sigma$, the APR to PC phase transition moves up to higher densities, and $M_T$ increases correspondingly. At $m_\sigma > 850$ MeV ($\lambda = 6.45$), the PC core cannot form since $M_T$ exceeds the maximum mass of the neutron star, which in this model works out to be about $1.6 M_\odot$ (see Fig. 3). On the other hand, at $m_\sigma < 750$ MeV ($\lambda = 5.7$) the Maxwell construction between the PC and the APR state is no longer possible. Neutron stars with quark matter PC cores can therefore exist only if $m_\sigma$ is in the range 750–850 MeV. The magnetic moment of such PC cores could lead to magnetic fields as strong as $\sim 10^{15}$ G at the surface of the neutron star and may be responsible for the strong fields found in magnetars.
5. Discussion

5.1. Quark matter ground states

(i) At issue is the question if the neutral pion condensate we have considered is the lowest energy ground state. It is to be noted from the lecture notes of Baym [21], that the neutral pion condensate is the preferred ground state over the charged pion condensate for charge neutral nuclear matter in $\beta$ equilibrium, in the nonrelativistic limit.

Let us first consider the question as to how the quark matter charged pion condensate ground state compares with our ground state. Ref. [13] finds these two condensates are related by a chiral rotation and are degenerate in energy, but this is only for isospin symmetric matter—that is in the absence of both charge neutrality and $\beta$ equilibrium.

We have considered this question for charge neutral quark matter in $\beta$ equilibrium analytically and find that the neutral pion condensate is the preferred ground state over the charged pion condensate, for nonrelativistic and point-like ($g_A = 1$) quarks. This is important as the charged pion condensate has no dipole magnetism [13].

(ii) The $K$ condensate in nucleon matter is a strong candidate for the ground state. In nuclear matter the term that gives rise to this is the chiral symmetry breaking, sigma term, which is proportional to the nucleon mass and first order in the symmetry breaking expansion parameter (or $m_\pi$). In the case of quark matter, with point like quarks, such a term is not proportional to the nucleon mass and is second order in $m_\pi$ and is thus unlikely to play a defining role. We think that this may rule against $K$ condensates in quark matter as opposed to nuclear matter, though we have not carried out this calculation.

(iii) It is worth pointing out that all these condensate states have lower energy than the chirally restored CRQM state, are chiral symmetry broken states.

5.2. Quark matter at even higher density

At very high density, QCD gluon interactions become weak and enter the asymptotically free regime. It is well known that the quark Fermi seas are unstable to the formation of the diquark condensate state, no matter how weak the gluon interaction is.

There is an important issue which then arises; at what density does the pion condensed quark matter state transit into the diquark condensate state? In this section we review some work [27,28] that addresses this question using the NJL model and which has implications for our work.

We have argued that the linear sigma model is a valid model till centre of mass energies/scales of less than, 700–800 MeV, the right procedure would be to take this model to describe physics up to this scale. But our model has only chiral condensates.

It is well known that there is an identity between the NJL model and the linear sigma model [29], and thus the NJL can be mapped to our linear sigma model [28] and the ground state thereof. The NJL model, which has a chiral symmetric four fermion interaction can, however, accommodate both chiral (quark–antiquark colour singlet) condensates and diquark condensates.

A comparison of these two states has been done by Sadzikowski [27,28,30] in the context of a NJL chiral symmetric model, for the case of 2 flavours—$SU(2)_L \times SU(2)_R$. What is done is at the level of mean field theory. The NJL model has four fermion interactions in terms of the quark bilinears corresponding to the $\sigma$ and $\pi$ field quantum numbers, with a common dimensional coupling, $G$. If we are interested in a ground state carrying sigma and/or pion condensates we can replace these quark bilinears by the corresponding $\sigma$ and $\pi$ EVs in the MFT. This yields the ground state energy of the space uniform SSB and the PC states. Further, these works calculate the ground state energy of the diquark condensate which is got from the NJL four Fermi interaction by Fierz transformation.

Although these results [27,28,30] are not for exactly the same parameters they provide a good sense of the physics. In this case the PC is the preferred ground state till $\mu$ well above 400 MeV. As the tables suggest such values of $\mu$, correspond to baryon density 5–6 times nuclear density—the central density in our stars. This makes the PC a likely state in our neutron star cores.

Of course, these works deal with the two-flavour case—where the colour diquark condensate is a chiral singlet. Realistically, we must consider 3 flavours, since the quark chemical potential is much greater than the strange quark mass. In this case we are likely to have the CFL state as the lowest energy state. The criterion for this is given in [31] and is $\Delta > m^2_s/4\mu$, which is easily satisfied. Furthermore, in this case the diquark condensate is a colour-flavour condensate which has both chiral SSB and colour SSB, albeit in a manner different to the PC. With increasing baryon density we then expect the following hierarchy. At nuclear density and above we have nuclear matter, followed by neutral pion condensed quark matter and finally a transition to the diquark CFL state which all have chiral SSB. These considerations indicate that, at any finite density ($T = 0$) chiral symmetry remains spontaneously broken. A similar result has also been obtained in 1 + 1 dimension [32].

6. Results

(i) For neutron stars it is charge neutral quark matter in $\beta$ equilibrium at given density that is needed for the equation of state. We have found that the neutral pion condensate is the ground state for such quark matter. This is the first such calculation for this EOS and what is also new is that all the parameters of $L$ are fixed from the hadronic sector.

(ii) We have made a strong case for the spontaneous breaking of chiral symmetry at all baryon density (at $T = 0$).

(iii) We have constructed neutron stars with pion condensed quark matter cores and found that such cores occur only under very particular constraints on the value of the scalar (pseudoscalar) coupling, $5.7 < \lambda < 6.45$ (or equivalently, when the ‘tree’ level sigma mass in this model is in a small window, 750–850 MeV). This window is consistent with, $\pi, \pi$ scattering data.
(iv) We also find that at the central density of such stars the pion condensed state is the most likely and that the quark cores can carry high magnetic fields.

Is this a coincidence that a single parameter in our effective $L$, the ‘tree’ level mass of the sigma or the value of $\lambda$, plays a crucial role? Is it fortuitous that the tree level sigma mass set by scattering experiments sits in a small window that simultaneously rules out SQM as the absolute ground state of matter [6] and also can provide us with neutron stars that can have magnetic PC cores?

The problem in sustaining a PC core with a nuclear exterior is that we have a stiff exterior with a soft interior—a rather unstable situation. It is thus not so surprising that very particular conditions must obtain for this to occur.

Clearly to get quantitative results we have to work with a specific model. In this case we have worked with a specific effective $L$ that is built on the two symmetries of the strong interaction, chiral symmetry and colour symmetry. Further, all its couplings are determined from experimental hadronic data. In this case we are working from low density (energy) to higher density (energy), till the compositeness scale. This is different from the vast literature on diquark condensation (CFL), which examines the problem from the high density end where is no experimental data. Though, we cannot claim any sanctity for our $L$, we have provided physical justification for it, its range of validity and determined its parameters from hadronic physics. Finally, proof can be provided only from testing our specific results.

**Acknowledgements**

This work began from a weekly series of lectures on the origin of magnetic fields in astrophysics organized by Siraj Hasan at the Indian Institute of Astrophysics in which, besides us, Pijush Bhattacharjee, C. Sivaram and C.S. Shukre participated. Thanks go to them all. We thank W. Broniowski for sharing some of their computer codes with us. V.S. thanks Manique Rho for his perennial interest and support. V.S. thanks J. Schechter for reading the manuscript, Brandon Carter, D. Sahadev, C.S. Aulakh and Nirvikar Prasad for discussions, and the Raman Research Institute for ample and steady hospitality.

**References**


Sigma exchange in the nuclear force and effective field theory

John F. Donoghue

Department of Physics, University of Massachusetts, Amherst, MA 01003, USA
Institut des Hautes Études Scientifiques, Bures-sur-Yvette, F-91440, France

Received 14 April 2006; received in revised form 11 August 2006; accepted 12 October 2006
Available online 7 November 2006
Editor: J-P. Blaizot

Abstract

In the phenomenological description of the nuclear interaction an important role is traditionally played by the exchange of a scalar $I = 0$ meson, the sigma, of mass 500–600 MeV, which however is not seen clearly in the particle spectrum and which has a very ambiguous status in QCD. I show that a remarkably simple and reasonably controlled combination of ingredients can reproduce the features of this part of the nuclear force. The use of chiral perturbation theory calculations for two pion exchange supplemented by the Omnes function for pion rescattering suffices to reproduce the magnitude and shape of the exchange of a supposed $\sigma$ particle. I also attempt to relate this description to the contact interaction that enters more modern descriptions of the internucleon interaction.

When describing QCD to non-physicists, we generally say that it is the theory that accounts for nuclear binding. However, in practice our understanding of the precise way that QCD leads to nuclear bound states is still not good. Nuclear binding is most commonly described by an internucleon potential which can be parameterized by the exchange of mesons [1–3]. A key feature is that there is an attractive component to the central potential with an intermediate range. This component is often parameterized by the exchange of a scalar isoscalar meson, the sigma, of mass around 500–600 MeV. While other exchanges in the potential are correlated with clear resonances seen in the particle spectrum, the sigma is a puzzle. It is not seen in the usual way in the spectrum and, after 40 years of debate, does not have a clear interpretation in terms of the quarks and gluons of QCD.1 It is unfortunate that this ingredient in the signature effect of the strong interactions has such an ambiguous status.

The expectation is that the sigma represents, in some way, the exchange of two pions. The quantum numbers certainly are correct for this. Sophisticated attempts that construct the potential from scattering data (e.g. [6]) have two pions as the lightest intermediate state. However, while phenomenologically useful, these are not able to answer the question of the fundamental nature of the central interaction. Modern descriptions of the internucleon interaction use chiral perturbation theory to calculate two pion exchange at low energy [7–11]. This description appears to provide a good description of the longest range part of the internucleon force and can be used to describe this component of nucleon–nucleon scattering. However, the chiral amplitude does not produce a potential in agreement with that expected for sigma exchange. As shown in Fig. 8 of Ref. [8], the resulting potential grows too strong at moderate distances. This problem is readily traceable to the fact that the chiral amplitudes grow monotonically with the energy and hence get very large at moderate energies. In this Letter I add a simple and well-motivated addition to the chiral description, i.e. the Omnes function describing pion rescattering. We will see that this will produce an interaction remarkably close in structure to the exchange of a 600 MeV sigma meson. It is clear that a sigma resonance is not the driving feature of this calculation, yet the needed properties of sigma exchange are reproduced.2

1 A discussion of the status of the sigma which is very much in the spirit of the present work can be found in [4]. A careful recent analysis of $\pi\pi$ scattering describing the sigma as a pole on second sheet, quite far from the real axis, is found in [5].

2 A few other attempts to describe the nuclear interaction without a sigma are seen in [12].
Recently, there have been successful applications of ideas of effective field theory in which the nuclear interaction is treated not by potentials but by contact interactions—delta function interactions [7,13]. At low energy (recall that the energy typical of nuclear binding is 10 MeV/nucleon) the result of the exchange of a heavy particle can be described by a local interaction. Mathematically, this is consistent with the potential description because, as the mass $m$ gets large, the Yukawa potential forms a representation of a delta function. Physically, this follows from the uncertainty principle, as the exchange of a heavy particle has a short range. Nonlocality, to the extent it is needed, can then be described by contact derivative interactions. This development greatly increases the generality of the description of nuclei, as it reduces multiple potentials with different functional forms to a small number of constants giving the strengths of the contact interactions. After discussing the potential treatment, I will also attempt to make contact with this effective field theory description.

As the primary tool, I will use the dispersion relation derived by Cottingham, Vinh Mau and others [6]. For textbook reviews, see [1,2]. The scattering amplitude for two nucleons obeys an unsubtracted $t$-channel dispersion relation.

$$M(s,t) = \frac{1}{\pi} \int_{4m^2}^{\infty} d\mu s \left( \frac{\text{Im } M(s,\mu^2)}{\mu^2 - t - i\epsilon} \right). \tag{1}$$

The imaginary part of this amplitude is connected to the on-shell amplitudes in the crossed channel $NN \rightarrow NN$ with the important intermediate state being that of two pions. The overall amplitude is decomposed into partial waves described by their spin and isospin quantum numbers. The greatest interest in this Letter will be on the scalar–isoscalar ($J = 0, I = 0$) channel. By taking the nonrelativistic limit and ignoring the energy dependence in the $S$ channel, one can define a momentum space potential that depends only on the momentum transfer $q^2$. For the scalar isoscalar central potential, let us define the corresponding spectral function

$$\rho_S(\mu) = \text{Im } M_S(q = i\mu). \tag{2}$$

In terms of this imaginary part the potential is defined by the dispersion relation

$$V_S(q^2) = \frac{2}{\pi} \int_{2m_N}^{\infty} d\mu \frac{\rho_S(\mu)}{\mu^2 + q^2}, \quad V_S(r) = \frac{1}{2\pi^2 r} \int_{2m_N}^{\infty} d\mu \mu e^{-\mu r} \rho_S(\mu). \tag{3}$$

Because $\rho_S$ describes physical on-shell intermediate states, this formalism provides a well defined tool either for the analysis of nucleon scattering or for theoretical attempts to describe the nuclear interaction. Much recent work (see [8–11] and references therein) has used this formalism to match to chiral descriptions.

At low energy these spectral functions can be rigorously calculated in chiral perturbation theory. The imaginary part of the Feynman diagrams describe the physical intermediate states and generate the spectral function $\rho_S$. These can be calculated either through the direct calculation of the Feynman diagram, or by appropriately multiplying together the relevant on-shell $\pi N \rightarrow \pi N$ scattering diagrams [14]. For the diagrams of Fig. 1a, b, c these imaginary parts are [8,9]

$$\rho_S^{a,b}(\mu) = \frac{3g^2_A}{64F^2} \left[ 4c_1m^2_{\pi} + c_3(\mu^2 - 2m^2_{\pi}) \right] \times \left( \frac{\mu^2 - 2m^2_{\pi}}{\mu} \right)^2 \Omega(\mu), \tag{4}$$

$$\rho_S^{c}(\mu) = -\frac{3}{32\pi F^2_{\pi}} \sqrt{1 - \frac{4m^2_{\pi}}{\mu^2}} \theta(\mu - 2m_{\pi}) \times \left[ \left( 4c_1m^2_{\pi} + \frac{c_2}{6}(\mu^2 - 4m^2_{\pi}) + c_3(\mu^2 - 2m^2_{\pi}) \right) \right]^2 + \frac{c_2^2}{45}(\mu^2 - 4m^2_{\pi})^2. \tag{5}$$

Here $c_1$, $c_2$, $c_3$ are parameters that describe the $NN\pi\pi$ vertex—these have been measured in pion nucleon interactions [8,9,11,15,16]. I will address the box and crossed box diagrams below. These spectral functions are valid in the low energy regime only, and one observes that they grow monotonically with the energy.

However, there is another ingredient which necessarily enters. In the description of the $\pi\pi$ system, unitarity requires the inclusion of $\pi\pi$ rescattering. For a single elastic partial wave, unitarity of the $S$ matrix and analyticity require a unique form of the solution, given originally by Omnes [17]. The amplitudes in the elastic region are described by a polynomial in the energy times the Omnes function

$$\Omega(\mu) = \exp \left[ \frac{\mu^2}{\pi} \int_{s}^{\infty} ds \frac{\delta(s)}{s - s - \mu^2} \right]. \tag{6}$$

Here $\delta$ is the $\pi\pi$ scattering phase shift, in our case for the $I = 0, J = 0$ channel. Chiral perturbation theory is consistent with this order by order in the energy expansion. Following Ref. [18], it is known how to match this general description to the results of chiral perturbation theory by appropriately identifying the polynomial. The elastic region in this channel extends effectively up to energies of 1000 MeV.

In practice there has been good success at using the lowest order chiral amplitudes, supplemented by the Omnes function.
Fig. 2. The left graph gives the pion phaseshifts that are the input into the Omnes function, while the right figure shows the real part and the absolute square of the Omnes function.

Fig. 3. The left figure shows our results for the spectral function $\rho(\mu)/\mu$ as well as the individual components of diagram 1 a, b, c. The right figure shows the coordinate space potential $rV(r)$. There are actually two curves in the figure on the right. One is the result of this calculation and the second is that of a narrow 600 MeV sigma, with normalization chosen to match. The curves cannot be distinguished.

An example which is close to the present problem is $\gamma\gamma \rightarrow \pi\pi$ in the S wave. Here the lowest order calculation results in an amplitude which also grows monotonically and which violates unitarity near 600 MeV [19]. However, the addition of the Omnes function [20] tames this runaway growth. When the Omnes function and the lowest order amplitude are combined the result is in close agreement with both experiment and with a two loop chiral calculation up to energies beyond 700 MeV [21]. This procedure is equally rigorous as the usual chiral method at low energies. At higher energies, it can be adapted order by order. The Omnes function captures some of the features that would emerge if chiral perturbation theory were applied at higher order—it captures a subset of diagrams that relate to unitarization. While a full description clearly requires a complete set of higher order calculations, the Omnes method can be useful in those cases where pion rescattering is strong. In practice, it is most important when the $\pi\pi$ system is in an S-wave.

I will adopt the Omnes solution matched to the leading order chiral result, and will explore possible modifications below. The description of the spectral function then becomes

$$\rho_S(\mu) = \rho_S^{a,b} \text{Re} \Omega(\mu) + \rho_S^c |\Omega(\mu)|^2.$$  \hspace{1cm} (7)

The phase shifts can be analyzed in chiral perturbation theory in combination with experiment, with the definitive treatment of Colangelo et al. (CGL) [22]. Their result for the $I = 0$, $J = 0$ phase shift is shown in Fig. 2, along with the resulting Omnes functions.\(^3\) Note that there is no sigma resonance visible in the phase shift near 500–600 MeV. A resonance in the elastic region is manifest by the phase shift passing through 90 degrees, which certainly does not happen near the sigma mass. If one explores the complex plane there is a pole on the second sheet very far from the real axis [5]. However, the resonance in not the driving force in the description of the $\pi\pi$ amplitude at these energies. Instead, the chiral amplitude can be parameterized by a few low energy constants, which in turn are more determined by the $\rho(770)$ than by the sigma [23].

With these ingredients, we can display the result for the scalar interaction. In Fig. 3, I show the result for $\rho$, along with the individual contributions of the diagrams of Fig. 1. If we had a pure sigma exchange this would be a delta function at the mass of the $\sigma$, or a Breit–Wigner shape corresponding to a narrow resonance. One could be forgiven for seeing this result

\(^3\) In producing the Omnes function, I had to extend the phase shifts above the $\mu = 850$ MeV endpoint of the CGL analysis in order that the principle value part of the Omnes function integral be well behaved near the upper end. I have explored several smooth extensions, with residual effects at the few percent level.
as a very broad resonance, even though no resonance exists in the formalism. The coordinate space potential is also shown in Fig. 3. Also shown for comparison is the potential of an infinitely narrow 600 MeV scalar with a normalization chosen to match. In practice these are hard to differentiate because the curves are nearly identical. The simple description of Eq. (5) reproduces closely the spatial variation of the sigma potential. This says that these results have a range which is capable of the describing the intermediate range attraction in the central potential which is needed for nuclear binding. The strength of the interaction will be addressed below.

One can address the robustness of this result by considering possible higher order modifications of the basic representation. The \( NN\pi\pi\) interaction has been described by the lowest order chiral Lagrangian. There are also energy dependent modifications to these low order results. In particular, when studying on-shell vertices such as those that go into the spectral function we most often find form factors that depend on the energy. It would be reasonable to expect that the \( NN\pi\pi\) interaction would be modified by a form factor such as

\[
c_3 \rightarrow \frac{c_3}{(\mu^2 + m^2)^n}. \tag{8}\]

Indeed, there is theoretical expectation that this should occur. A study of the ingredients of this coupling [24] suggests that 55–70% of \( c_3 \) arises from integrating out \( \Delta \) exchange in the t-channel. The \( \Delta \) propagator then leads to an energy dependence of the vertex. Following the techniques of [14], it is straightforward to incorporate the \( \Delta \) propagator into these diagrams.3 I have explored the effects of form factors and the \( \Delta \) propagator. While these influence the magnitude appreciably, as described below, there is not a qualitative change in the shape of the spectral function. For example, taking the extreme case that all of \( c_3 \) arises through \( \Delta \) exchange in the t-channel, the modification of spectral functions is shown in Fig. 4. While the relative contribution of the two types of diagram changes slightly, the energy variation and spatial variation are remarkably similar to the original case. The use of a pure monopole or a dipole form factor does not change this conclusion.

Let me address the strength of the interaction by considering the integral of the spectral function through the energy region under consideration,

\[
G_s = \frac{2}{\pi} \int_{2m_\pi}^{0.8 \text{ GeV}} \frac{d\mu}{\mu} \rho_S(\mu). \tag{9}\]

This is just the integral of the curves shown in Figs. 3a, 4. In potential models with the exchange of a narrow sigma, this has the value

\[
G_s^\sigma = \frac{g_\sigma^2}{m_\sigma^2}, \tag{10}\]

which numerically is often given as \( G_s^\sigma = 300–450 \text{ MeV}^{-2} \) [3,13]. The result depend most sensitively on the parameter \( c_3 \), which is not perfectly known. The phenomenological extraction of \( c_3 \) from \( \pi N \) and \( NN \) data has a large error bar, \( c_3 = -4.7^{+1.2}_{-1.0} \text{ GeV}^{-2} \) [8,11,15,16]. However, when using an Omnes representation, it is likely that this constraint is on the product \( c_3 \Omega(2m_\pi) \), in which case the value would be \( c_3 = -3.7^{+1.0}_{-0.8} \text{ GeV}^{-2} \). (The other parameter choices used were \( c_1 = -0.64 \text{ GeV}^{-2} \) and \( c_2 = 3.3 \text{ GeV}^{-2} \), although these have only a small impact on the results.) The result for \( G_s \) as a function of \( c_3 \) is shown in Fig. 5. There is rough agreement for the required range of magnitudes of \( G_s \) for the allowed values of \( c_3 \). Here the use of a form factor does make a difference. With the inclusion of the effect of the \( \Delta \) propagator, the result is 20% smaller. If we use a straight monopole form factor with a mass \( m = 800 \text{ MeV} \) the value of \( G_s \) is 40% smaller than without it for a given value of \( c_3 \). These examples show the model dependence of the higher order effects. At present understanding the differences may be accounted for by adjusting the value of \( c_3 \). These uncertainties in the appropriate values of \( c_3 \) and \( G_s \) keep us from using the magnitude of the potential as a precise test of the method.

Now consider the effective field theory description of the interaction. In a Wilsonian effective field theory treatment, one treats the light degrees of freedom (pions in this case) dynamically up to a scale \( \Lambda \). This means that we consider tree and loop diagrams with energies below this scale. Physics beyond this

---

4 The formulas of Section III of [14] apply directly with the modification \( \tau = (t + 2m_\Delta^2 - 2m_N^2 - 2m_\pi^2)/\sqrt{1 - 4m_\pi^2 \xi m_N} \) whenever there is a \( \Delta \) propagator.
scale is treated by a contact interaction, i.e. $G_s(\Lambda)$ — a local interaction that parameterizes the residual physics from energies above $\Lambda$. Epelbaum, Glöckle and Meißner (EGM) [10] have implemented such a treatment using the chiral amplitudes without the Omnes function in the spectral function with an energy cutoff. The ideal use of the present calculation would be to use the Omnes description in a treatment like EGM. The Omnes description appears to be a better representation of the long distance physics than just using the bare chiral amplitudes. There is no loss of rigor in adding the Omnes function to the chiral predictions at low energy as long as the matching is done correctly. In addition, the Omnes function tames the runaway growth of the bare chiral amplitudes. However, at higher energies, the present procedure is clearly not rigorous and there is inherently some uncertainty in the dynamical calculation, as we have demonstrated above. The contact interaction arising from the chiral Lagrangian would serve to correct any flaw in the dynamical calculation in order to fully agree with reality. To the extent that this dynamical calculation is a good one, the residual contact interaction would be small. Indeed, from the agreement seen above for the rough magnitude and shape of the potential, it plausible that the residual contact interaction could be negligible compared to the primary dynamical effect of two pion exchange.\(^5\)

However, the EGM approach is not the standard effective field theory treatment. More commonly, the usual pion exchange diagrams, Fig. 1d, e, are treated dynamically, but those of Fig. 1a, b, c are not explicitly considered. This is the case when dimensional regularization is used rather than an energy cutoff, because the former diagrams are finite and the latter are divergent when treated at any given order in the energy expansion. The effects of these diagrams are then included in the contact interaction. This contact interaction is given by the momentum space potential at zero momentum

$$ G_s^{\text{eft}}(q^2 = 0) = \frac{2}{\pi} \int_{2m_\pi}^\infty \frac{d\mu}{\mu} \rho_S(\mu). \quad (11) $$

To the extent that significant contributions do not come from energies above the end of our calculation (800 MeV), our calculation then provides an estimate of the strength of the contact interaction in the scalar–isoscalar channel.

In order to assess this result it is perhaps easiest to compare with Ref. [26] (EMGE). These authors have considered the extraction of the effective field theory coefficients from the data, including the effects of regularizing the calculations. They have also compared to the phenomenology of integrating over the potential or integrating out bosonic resonances. For potentials with a sigma effect, their results for this channel amounts to $G_s^{\text{EMGE}} = g_\sigma^2/m_\sigma^2$ plus smaller corrections from higher scalar resonances. While again the rather large uncertainties in the magnitude of the present calculation make it any precise conclusions possible, it then appears that the magnitude discussed above is also roughly appropriate for the effective field theory description. Note that present effective field theory treatments can differ in how the contact interaction is treated. In some applications, such as [13], the contact interaction can be directly used without further modifications. In others, such as [7], a smearing or renormalization of the contact interaction is implemented to deal with divergences in that calculation. In such situations, the appropriate renormalized value may be different and a matching of that calculation to the spectral description would be needed to compare the values.\(^6\)

Potentials can have different meanings in different contexts [27] and in different calculational schemes there can be different values of the contact interaction [28]. Therefore let me specify more fully the scheme of the present calculation. In the chiral treatment, one keeps pion exchange as an explicit degree of freedom while treating the shorter range interactions as contact terms. The $NN\pi$ vertex is the one in the usual baryon chiral Lagrangian. In such a treatment, we should treat the box and crossed-box diagrams of Fig. 1 dynamically, and they should therefore not be included into the contact interaction. For this reason I did not include these diagrams in the calculation of the integrand $\rho_S(\mu)/\mu$ whose integral gives the strength of the contact interaction. In frameworks other than the one considered here, it might be appropriate to include some of the box and/or crossed-box diagrams into the description of the contact interaction. However, it seems that for the scalar central potential the iteration of the one pion interaction is a numerically small compared to the irreducible two pion/sigma contribution, for example see [8] and Fig. 3.15 of Ref. [2]. With slight modifications, then, it appears that the present calculation could also be adapted to other frameworks.

Chiral perturbation theory plus the Omnes function give a quite simple description of the scalar central potential, with a result very similar to the exchange of a conventional sigma particle. However the physics important in this calculation is not a sigma resonance, but rather only chiral amplitudes and the Omnes function. This description appears to be robust, being qualitatively unchanged by the addition of higher order interactions. Besides elucidating a long standing puzzle, these results are useful because we have a reasonable control over all the main ingredients, the chiral amplitudes and the $\pi\pi$ phase shifts. The connection of the nuclear interaction to QCD becomes more under control.

Acknowledgements

I would like to thank especially Thibault Damour for many conversations during the course of this work. U. von Kolck for discussions about the iterated one pion interaction. R. Furnstahl and B. Serot for information on the nuclear interaction and J. Gasser, Barry Holstein and H. Leutwyler for continually informative discussions on the dispersion representation and the Omnes function. I also appreciate the hospitably the

\(^5\) This is further discussed in [25].

\(^6\) Experience from the meson sector [23] suggests that similar evaluations of chiral coefficients of that sector match well to dimensionally regularized coefficients with scales of 500–700 MeV.
IHES where most of this work was performed. This work has been partially supported by the US National Science Foundation.

References

Power-like corrections and the determination of the gluon distribution

F. Hautmann

Institut für Theoretische Physik, Universität Regensburg, 93040 Regensburg, Germany

Received 9 October 2006; received in revised form 24 October 2006; accepted 30 October 2006
Available online 7 November 2006
Editor: N. Glover

Abstract

Power-suppressed corrections to parton evolution may affect the theoretical accuracy of current determinations of parton distributions. We study the role of multigluon-exchange terms in the extraction of the gluon distribution for the large hadron collider (LHC). Working in the high-energy approximation, we analyze multigluon contributions in powers of $1/Q^2$. We find a moderate, negative correction to the structure function’s derivative $dF_2/d \ln Q^2$, characterized by a slow fall-off in the region of low to medium $Q^2$ relevant for determinations of the gluon at small momentum fractions.

© 2006 Elsevier B.V. All rights reserved.

The estimate of the theoretical accuracy on the determination of parton distributions is relevant for phenomenology at the large hadron collider (LHC) [1–3]. A potential source of theory uncertainty is given by power-suppressed contributions to parton evolution. In particular, the extraction of the gluon distribution for small momentum fractions depends on data in the region of high energies and moderate to low $Q^2$, in which power-like corrections from multiple parton scatterings are potentially significant.

The purpose of this note is to derive an estimate of these corrections, based on high-energy amplitudes for multigluon exchange [4]. We concentrate on the power correction $\Delta$ to the $Q^2$ derivative of the structure function $F_2$

$$
\frac{dF_2}{d \ln Q^2} = K \otimes G \left[ 1 + \Delta \right] + \text{quark term},
$$

(1)

where $K$ is the perturbative kernel, $G$ is the gluon distribution, and $\Delta$ is $O(1/Q^2)$. This correction contributes to the theory uncertainty on $G$, since below $x \lesssim 10^{-2}$ this is mostly determined from data for $dF_2/d \ln Q^2$.

The theoretical framework to treat multiple scatterings is based on the $s$-channel picture of DIS [4], and its basic degrees of freedom are described by matrix elements of eikonal lines. The corresponding predictions, incorporating nonperturbative dynamics, are valid down to small $Q^2$. To enforce consistency with the framework of standard parton analyses [2,3,5] we will expand the answer in powers of $1/Q^2$. We study the behavior of the power expansion at low $Q^2$ and $x$, and identify the power corrections by subtraction of the leading-power contribution.

We find moderate, but nonnegligible negative corrections, increasing in size as $x$ decreases. We find that with a physically natural choice of parameters in the eikonal matrix elements we can achieve a sensible description of data and still have power corrections to $dF_2/d \ln Q^2$ that do not exceed the leading power already at $Q^2 \gtrsim 0.5 \text{ GeV}^2$. However we find that for small $x$ the corrections have a slow fall-off with $Q^2$ in the region of intermediate $Q^2$, $Q^2 \simeq 1–10 \text{ GeV}^2$, behaving effectively like $1/Q^\lambda$ in this region, with $\lambda$ close to 1 for $x \lesssim 10^{-3}$. This behavior results from summing the power expansion. As a consequence the power corrections stay larger than 10% up to $Q^2$ of a few GeV$^2$ for $x$ below $10^{-3}$.

The contents of the Letter is as follows. We first give the $s$-channel results that provide the basic elements for the evaluation of the power expansion. Then we focus on the $Q^2$ derivative of the structure function. We examine the next-to-leading-power contribution and the sum of the power series, and present numerical results.

We begin by recasting the result [4] for multigluon contributions to DIS structure functions in the Mellin-transform representation. This is convenient to analyze the $1/Q^2$ expansion. In
the high-energy approximation we describe gluon exchange in terms of eikonal-line operators

\[
V(x) = \mathcal{P} \exp \left\{ -ig_s \int_{-\infty}^{+\infty} dz^- A^+_{\mu}(0, z^-, z) t_\mu \right\},
\]

where \( A \) is the color potential and \( \mathcal{P} \) is the path-ordering. Following [6] we note the matrix elements of eikonal operators at transverse positions \( \mathbf{b} \) and \( \mathbf{b} + \mathbf{z} \) as

\[
\mathcal{E}(\mathbf{z}, \mathbf{b}) = \int \left[ dP^\prime \right] \left[ P^\prime \right] \frac{1}{N_c} \text{Tr} \left\{ 1 - V^\dagger(\mathbf{b} + \mathbf{z}) V(\mathbf{b}) \right\} / |P^\prime|.
\]

Here \( [dP^\prime] = dP^\prime + d^2 P^\prime / 2 P^\prime (2\pi)^3 \), \( \mathbf{z} \) is the transverse separation between the eikonal lines, and \( \mathbf{b} \) is the impact parameter. The result [4] for the transverse structure function \( F_T \) can be written as

\[
x F_T(x, Q^2) = \int_{c-i\infty}^{c+i\infty} \frac{du}{2\pi i} \int d^2 b \frac{\mathcal{E}(u, \mathbf{b})}{\Phi(u)},
\]

where \( 0 < c < 1 \), \( \mathcal{E} \) is the Mellin transform of \( \mathcal{E} \)

\[
\mathcal{E}(u, \mathbf{b}) = \int \frac{d^2 z}{\pi^2} (z^2)^{u-1} \mathcal{E}(\mathbf{z}, \mathbf{b}),
\]

and \( \Phi \) is a calculable coefficient. To lowest order

\[
\Phi(u) = \sum_a e_a^2 \frac{N_c}{16\pi^2} \frac{1}{u} \left( \frac{Q^2}{2} \right)^u \frac{\pi}{4^u} \Gamma(3 - u) \Gamma(2 - u) \Gamma(1 - u) \Gamma(2 + u) \Gamma(5/2 - u) \Gamma(3/2 + u),
\]

where \( N_c = 3 \), and \( e_a \) is the electric charge of quark of type \( a \).

From this approach the parton-model framework is recovered through an expansion in powers of \( gA \), valid for small \( \mathbf{z} \). In particular, the coefficient of the quadratic term in the expansion of \( \mathcal{E} \) can be related to the gluon distribution \( G \), [4],

\[
\int d^2 b \mathcal{E} \simeq \frac{\pi s^2 T_R}{4N_c} (xG) z^2 \left( 1 + \mathcal{O}(|\mathbf{z}|) \right),
\]

where \( T_R = 1/2 \).

The quark distribution can be dealt with by the same method. The main difference is that while in the structure function case the ultraviolet region is naturally regulated by the physical scale \( Q^2 \), in the case of the quark distribution we need to treat the ultraviolet divergences. The result in dimensional regularization is [7]

\[
x q(x, \mu) = (\mu^2)^{-2x} \int d^2 z d^2 z' b w(z) \mathcal{E}(z, \mathbf{b}) - \text{UV},
\]

where

\[
w(z) = \frac{N_c}{3\pi^4} \frac{1}{z^4} \frac{\left( \mu^2 z^2 \right)^{2x}}{4\pi} \frac{\Gamma(2 - 2x)}{1 - 2x/3},
\]

and we have indicated by UV the ultraviolet subtraction. This is required since in Eq. (3) \( (1 - V^\dagger V) \to 0 \) for \( z \to 0 \) and \( \mathcal{E} \propto z^2 \).

Using Eqs. (7), (9) the \( x \ll 1 \) form of the evolution equation for the quark is reobtained from Eq. (8). Power-suppressed contributions arise from the difference between the terms left over from the ultraviolet regularization in the quark-distribution and structure-function case [7]. In general, these depend on the scheme used for the ultraviolet subtraction. This dependence does not enter in the correction to the structure function’s derivative which we consider next.

The matrix element \( \mathcal{E} \) is nonperturbative and is to be determined from experiment. For the numerical calculations that follow we model its functional form according to the model [8,9]

\[
\tilde{\mathcal{E}}(u, \mathbf{b}) = \frac{\Gamma(u)}{1 - u} \left( \frac{\mu_s^2(\mathbf{b})}{4} \right)^{1-u},
\]

where \( \mu_s \) is the saturation scale, with \( \mathbf{b} \) dependence as in [8]. The operator relation (7) implies for model (10) that

\[
\int d\mathbf{b} \mu_s^2 = \frac{4\pi^2 T_R}{N_c} \alpha_s(\mu_s) xG(x, \mu_f).
\]

In Eq. (11) we have indicated explicitly the dependence of the running coupling and gluon distribution on the renormalization/factorization scales \( \mu_r, \mu_f \). In the present context the choice of these scales amounts to specifying the model for \( \tilde{\mathcal{E}} \).

Consider now the derivative of \( F_T \) with respect to \( \ln Q^2 \), \( F'_T = dF_T / d\ln Q^2 \). Taking the derivative cancels the factor \( 1/u \) in Eq. (6). We determine the expansion of \( F'_T \) in powers of \( 1/Q^2 \) by closing the integration contour in the complex \( u \)-plane to the left and evaluating the residues at the poles of the integrand. First we verify that the result from the leading pole (LP) \( u = 0 \) coincides with that from Eq. (8) for the quark distribution. We get

\[
x F'_{T,LP} = \sum_a e_a^2 \frac{N_c}{3\pi} \int d^2 b \text{Re} s(u=0) \mathcal{E}.
\]

By inserting Eqs. (10), (11), Eq. (12) yields the perturbative leading-power coefficient. We identify the power-suppressed correction by subtracting off this contribution. Next we consider the contribution from the next-to-leading pole (NLP) \( u \to -1 \), \( F'_{T,NLP} \). This is proportional to the \( u = -1 \) residue of \( \mathcal{E} \).

In Fig. 1 we compute the ratio

\[
\delta^{(NLP)} = F'_{T,NLP} / (F'_{T,LP} + F'_{T,NLP})
\]

versus \( Q^2 \) at \( x = 10^{-2} \) and \( x = 10^{-4} \) for two different choices of \( \mu_r, \mu_f \). The natural scale for \( \mu_r \) and \( \mu_f \) should be set by the inverse of the mean transverse distance \( \mathbf{z} \). For the illustration in Fig. 1 we take this scale to be on the order of \( Q \), and plot results for \( \mu_f = Q, \mu_r = Q \) and \( \mu_f = 2Q, \mu_r = Q/2 \). We use the CTEQ parton distributions [3], supplemented with NLO evolution for scales below the input scale.

The ratio \( \delta^{(NLP)} \) goes like \( 1/Q^2 \). We see from Fig. 1 that the size of the NLP contribution is rather sensitive to the scales \( \mu_f \) and \( \mu_r \). This is not surprising. The definition of the model for \( \mathcal{E} \), embodied in the choice of \( \mu_f, \mu_r \), corresponds to defining the (otherwise arbitrary) separation of perturbative and nonperturbative effects. In particular, varying these scales amounts to effectively simulating contributions of higher perturbative order. The variation of the nonperturbative power correction in
Fig. 1 says that this is unambiguously defined only once we specify what we include in the perturbative part of the calculation. See [10] for an extensive study of the issue of scale-setting and running coupling effects in high-energy evolution. In what follows we simply set the scales from comparison with experimental data.

Beyond the next-to-leading power, the poles in the $u$-plane have multiplicity higher than 1, leading to $\ln Q^2$ enhancements of the power corrections. The order-$n$ term is of the form

$$C(n, \ln Q^2) \frac{\xi_n}{(Q^2)^n}, \quad (14)$$

where $\xi_n$ give the dimensionful nonperturbative scale in terms of the $b$-integral of the moments (5) of $\Xi$, and $C$ are coefficients determined from Eq. (6). Through the moments of $\Xi$ the correction (14) receives contribution from the exchange of any number of gluons via eikonal operators.

We now proceed to evaluate numerically the contribution of all powers in $1/Q^2$. We go to $z$-space via the inverse Mellin transform and let the scales $\mu_f$ and $\mu_r$ vary with the distance $z$. Details for the numerical integration in coordinate space are given in [7]. Adding to Eq. (4) the contribution of longitudinal $F_L$, we compute the structure function $F_2$ and tune the factorization/renormalization scales, $\mu_f = c_1/z$ and $\mu_r = c_2/z$, by comparing the answer with the experimental data [11]. The result of doing this is reported in Fig. 2, where $c_1 = 4$, $c_2 = 0.32$.

Using these values for the model parameters, we turn to the derivative $dF_2/d\ln Q^2$. We calculate the power correction by subtraction of the leading power as described around Eq. (12). In Fig. 3 we plot the result for the power correction normalized to the full answer and multiplied by $(-1)$. We see from Fig. 3 that above $Q^2 \approx 0.5$ GeV$^2$ for $x \approx 10^{-3}$ the correction is less than 10%. This is associated with the curves having a rather slow decrease in the range of medium $Q^2$ in the figure, much slower than the asymptotic $1/Q^2$. For instance, for $x \approx 10^{-3}$ the behavior in the region $Q^2 \approx 1–10$ GeV$^2$ is closer to an effective power $1/Q^\lambda$ with $\lambda \approx 1$. This slow fall-off results from summing the terms (14). We leave to a future study the question of whether this behavior can be interpreted in terms of an effective, $x$-dependent semihard scale on the order of the GeV.

In summary, the results above indicate that the power expansion should still work at $x$ and $Q^2$ as low as in the region of the data presently used for determinations of the gluon distribution $G$ for the LHC, but subleading corrections to $dF_2/d\ln Q^2$ from multigluon exchange are nonnegligible in this region and contribute to the theory uncertainty on $G$. We have also computed the analogous corrections for the derivative of the transverse structure function [7], and we find that these are generally

For $x \lesssim 10^{-3}$ it takes $Q^2$ of a few GeV$^2$ before the correction is less than 10%. This is associated with the curves having a rather slow decrease in the range of medium $Q^2$ in the figure, much slower than the asymptotic $1/Q^2$. For instance, for $x \approx 10^{-3}$ the behavior in the region $Q^2 \approx 1–10$ GeV$^2$ is closer to an effective power $1/Q^\lambda$ with $\lambda \approx 1.2$. This slow fall-off results from summing the terms (14). We leave to a future study the question of whether this behavior can be interpreted in terms of an effective, $x$-dependent semihard scale on the order of the GeV.

In summary, the results above indicate that the power expansion should still work at $x$ and $Q^2$ as low as in the region of the data presently used for determinations of the gluon distribution $G$ for the LHC, but subleading corrections to $dF_2/d\ln Q^2$ from multigluon exchange are nonnegligible in this region and contribute to the theory uncertainty on $G$. We have also computed the analogous corrections for the derivative of the transverse structure function [7], and we find that these are generally
smaller than in the case of $F_2$. This may be regarded as yet another motivation for the importance of a separate measurement of the longitudinal component $F_L$ [12].

The results in this Letter are obtained using NLO parton distributions. Similar to what observed above Eq. (14), it is natural to expect a change in the power correction when going from NLO to NNLO. The tendency toward a decrease in the small-$x$ gluon at NNLO [2] is consistent with the possibility that NNLO parton distributions have smaller power corrections. The detailed interpretation of this behavior, however, will be subtler than in the large-$x$ case, since distinctly different dynamics now drive the power-like and NNLO effects.

A word of caution is needed in interpreting the calculations above. Multigluon amplitudes are treated in the high-energy approximation. Also, the modeling of the nonperturbative matrix elements and the summation of the power series expansion call for a firmer understanding. Besides, power corrections from sources other than that considered here may be relevant as well. In particular, corrections from self-energy graph insertions are still largely unexplored for flavor-singlet observables [13]. Nevertheless, the method presented above allows one to obtain an estimate of multigluon corrections which can be made consistently with perturbative evolution order by order. It is based on subtraction of the leading pole in Mellin space. This serves to specify the definition of the power correction. In this work we have been concerned with the contribution to the $Q^2$ derivative of the structure function, relevant for the extraction of $G$. But the approach can in principle be extended to evaluate corrections to $F_2$ itself, and to processes directly coupled to gluons. The latter will be especially interesting for studying multiple-scattering effects in the production of jet final states.

Acknowledgement

I thank D. Soper for collaboration on topics related to this work.

References

Implications of Fritzsch-like lepton mass matrices

Monika Randhawa a, Gulsheen Ahuja b, Manmohan Gupta b,∗

a University Institute of Engineering and Technology, P.U., Chandigarh, India
b Department of Physics, Centre of Advanced Study, P.U., Chandigarh, India

Received 10 July 2006; received in revised form 15 September 2006; accepted 8 October 2006
Available online 10 November 2006
Editor: T. Yanagida

Abstract

Using seesaw mechanism and Fritzsch-like texture 6 zero and 5 zero lepton Dirac mass matrices, detailed predictions for cases pertaining to normal/inverted hierarchy as well as degenerate scenario of neutrino masses have been carried out. All the cases considered here pertaining to inverted hierarchy and degenerate scenario of neutrino masses are ruled out by the existing data. For the normal hierarchy cases, the lower limit of \(m_{\nu_1}\) and of \(s_{13}\) as well as the range of Dirac-like CP violating phase \(\delta\) would have implications for the texture 6 zero and texture 5 zero cases considered here.

© 2006 Elsevier B.V. All rights reserved.

In the last few years, apart from establishing the hypothesis of neutrino oscillations, impressive advances have been made in understanding the phenomenology of neutrino oscillations through solar neutrino experiments [1], atmospheric neutrino experiments [2], reactor based experiments [3] and accelerator based experiments [4]. At present, one of the key issues in the context of neutrino oscillation phenomenology is to understand the pattern of neutrino masses and mixings which seems to be vastly different from that of quark masses and mixings. In fact, in the case of quarks the masses and mixing angles show distinct hierarchy, whereas in the case of neutrinos the two mixing angles governing solar and atmospheric neutrino oscillations look to be rather large and may even be maximal. The third angle is very small compared to these and at present only its upper limit is known. Similarly, at present there is no consensus about neutrino masses which may show normal/inverted hierarchy or may even be degenerate. The situation becomes further complicated when one realizes that neutrino masses are much smaller than lepton and quark masses.

In the context of quark masses, it may be noted that texture specific mass matrices [5,6] seem to be very helpful in understanding the pattern of quark mixings and CP violation. This has motivated several attempts [7], in the flavor as well as the non-flavor basis, to consider texture specific lepton mass matrices for explaining the pattern of neutrino masses and mixings. In the absence of sufficient amount of data regarding neutrino masses and mixing angles, it would require a very careful scrutiny of all possible textures to find viable structures which are compatible with data and theoretical ideas so that these be kept in mind while formulating mass matrices at the GUT (Grand Unified Theories) scale. In this context, using seesaw mechanism as well as normal hierarchy of neutrinos, Fukugita, Tanimoto and Yanagida [8] have carried out an interesting analysis of Fritzsch-like texture 6 zero mass matrices [9]. It may be noted that when small neutrino masses are sought to be explained through seesaw mechanism [10] given by

\[
M_\nu = -M_{\nu_D}^T(M_R)^{-1}M_{\nu_D},
\]

where \(M_{\nu_D}\) and \(M_R\) are respectively the Dirac neutrino mass matrix and the right-handed Majorana neutrino mass matrix, then the predictions are quite different when texture is imposed on \(M_{\nu_D}\) or \(M_\nu\). For the normal hierarchy of neutrino masses, while Fukugita et al. [8] have imposed texture 6 zero structure on \(M_{\nu_D}\), Xing et al. [11] have considered several possible texture specific structures.

∗ Corresponding author.
E-mail address: mmgupta@pu.ac.in (M. Gupta).

0370-2693/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.physletb.2006.10.034
for \( M_v \). These attempts also use parallel texture structures for neutrinos and charged lepton mass matrices, compatible with specific models of GUTs [7] as well as these could be obtained using considerations of Abelian family symmetries [12]. In the absence of any clear signals from the data regarding the structure of mass matrices, it becomes desirable to carry out detailed and exhaustive studies related to any particular texture of lepton mass matrices.

Using seesaw mechanism and imposing Fritzsch-like texture structure on Dirac neutrino mass matrices, with charged leptons having Fritzsch-like texture structure as well as being in the flavor basis, the purpose of the present communication is to investigate large number of distinct possibilities of texture 6 zero and 5 zero mass matrices for normal/inverted hierarchy as well as degenerate scenario of neutrino masses. Further, detailed dependence of mixing angles on the lightest neutrino mass as well as the parameter space available to the phases of mass matrices have also been investigated for texture 6 zero as well as for texture 5 zero cases. Furthermore, several phenomenological quantities such as Jarlskog’s rephasing invariant parameter \( J \), the CP violating Dirac-like phase \( \delta \) and the effective neutrino mass \( \langle m_{ee} \rangle \), related to neutrinoless double beta decay \( (\beta\beta)_{0v} \), have also been calculated for different cases.

To begin with, we summarize the most recent (August 2006) 3\( \sigma \) values of the neutrino mass and mixing parameters [13],

\[
\begin{align*}
\Delta m_{12}^2 &= (7.1-8.9) \times 10^{-5} \text{eV}^2, \\
\Delta m_{23}^2 &= (2.0-3.2) \times 10^{-3} \text{eV}^2, \\
\sin^2 \theta_{12} &= 0.24-0.40, \\
\sin^2 \theta_{23} &= 0.34-0.68, \\
\sin^2 \theta_{13} &\leq 0.040.
\end{align*}
\]

To define the various texture specific cases considered here, we begin with the modified Fritzsch-like matrices, e.g.,

\[
M_l = \begin{pmatrix} A_l & B_l & 0 \\ A_l^* & 0 & C_l \\ 0 & B_l^* & 0 \end{pmatrix}, \quad M_{\nu D} = \begin{pmatrix} A_{\nu} & 0 & 0 \\ D_{\nu} & B_{\nu} & 0 \\ 0 & B_{\nu}^* & C_{\nu} \end{pmatrix},
\]

\( M_l \) and \( M_{\nu D} \) respectively corresponding to Dirac-like charged lepton and neutrino mass matrices. Both the matrices are texture 2 zero type with \( A_l^{(\nu)} = |A_l^{(\nu)}|e^{i\theta_l^{(\nu)}} \) and \( B_l^{(\nu)} = |B_l^{(\nu)}|e^{i\gamma_l^{(\nu)}} \), in case these are symmetric then \( A_l^{*(\nu)} \) and \( B_l^{*(\nu)} \) should be replaced by \( B_l^{(\nu)} \) and \( A_l^{(\nu)} \), as well as \( C_l^{(\nu)} \) and \( D_l^{(\nu)} \) should respectively be defined as \( C_l^{(\nu)} = |C_l^{(\nu)}|e^{i\omega_l^{(\nu)}} \) and \( D_l^{(\nu)} = |D_l^{(\nu)}|e^{i\epsilon_l^{(\nu)}} \). The matrices considered by Fukugita et al. are of symmetric kind and can be obtained from the above mentioned matrices by taking both \( D_l \) and \( D_{\nu} \) to be zero, which reduces the matrices \( M_l \) and \( M_{\nu D} \) to texture 3 zero type. Texture 5 zero matrices can be obtained by taking either \( D_l = 0 \) and \( D_{\nu} \neq 0 \) or \( D_l = 0 \) and \( D_{\nu} = 0 \), thereby, giving rise to two possible cases of texture 5 zero matrices, referred to as texture 5 zero \( D_l = 0 \) case pertaining to \( M_l \) texture 3 zero type and \( M_{\nu D} \) texture 2 zero type and texture 5 zero \( D_{\nu} = 0 \) case pertaining to \( M_l \) texture 2 zero type and \( M_{\nu D} \) texture 3 zero type.

To fix the notations and conventions as well as to facilitate the understanding of inverted hierarchy case and its relationship to the normal hierarchy case, we detail the essentials of formalism connecting the mass matrix to the neutrino mixing matrix. The mass matrices \( M_l \) and \( M_{\nu D} \) given in Eq. (4), for Hermitian as well as symmetric case, can be exactly diagonalized, details of Hermitian case can be looked up in our earlier work [6], the symmetric case can similarly be worked out. To facilitate diagonalization, the mass matrix \( M_k \), where \( k = l, \nu D \), can be expressed as

\[
M_k = Q_k M'_k P_k,
\]

where \( M'_k \) is a real symmetric matrix with real eigenvalues and \( Q_k \) and \( P_k \) are diagonal phase matrices, for the Hermitian case \( Q_k = P_k^\dagger \). In general, the real matrix \( M'_k \) is diagonalized by the orthogonal transformation \( O_k \), e.g.,

\[
M_k^{\text{diag}} = (Q_k O_k \xi_k)^\dagger M_k (P_k^\dagger O_k),
\]

wherein, to facilitate the construction of diagonalizing transformations for different hierarchies, we have introduced \( \xi_k \) defined as diag(1, e\( ^{i\eta} \), 1) for the case of normal hierarchy and as diag(1, e\( ^{i\gamma} \), e\( ^{i\theta} \)) for the case of inverted hierarchy.

The case of leptons is fairly straightforward, whereas in the case of neutrinos, the diagonalizing transformation is hierarchy specific as well as requires some fine tuning of the phases of the right-handed neutrino mass matrix \( M_R \). To clarify this point further, the matrix \( M_v \), given in Eq. (1), can be expressed as

\[
M_v = -P_{\nu D} O_{\nu D} M_v^{\text{diag}} O_{\nu D}^T Q_{\nu D}^{-1} Q_{\nu D} O_{\nu D} \xi_{\nu D} M_{\nu D}^{\text{diag}} O_{\nu D}^T P_{\nu D},
\]

wherein, assuming fine tuning, the phase matrices \( Q_{\nu D}^{\text{diag}} \) and \( Q_{\nu D} \) along with \( -M_R \) can be taken as \( m_R \) diag(1, 1, 1) as well as using the unitarity of \( \xi_{\nu D} \) and orthogonality of \( O_{\nu D} \), the above equation can be expressed as

\[
M_v = P_{\nu D} O_{\nu D} \frac{(M_{\nu D}^{\text{diag}})^2}{(m_R)^T} O_{\nu D}^T P_{\nu D}.
\]

The lepton mixing matrix in terms of the matrices used for diagonalizing the mass matrices \( M_l \) and \( M_v \), which can be obtained respectively from Eqs. (6) and (8), is expressed as

\[
U = (Q_l O_l \xi_l)^\dagger (P_{\nu D} O_{\nu D}).
\]
Eliminating the phase matrix $\xi_l$ by redefinition of the charged lepton phases, the above equation becomes

$$U = O_l^T Q_l^T P_{\nu D} O_{\nu D},$$

(10)

where $Q_l^T P_{\nu D}$, without loss of generality, can be taken as $(e^{i\phi_1}, 1, e^{i\phi_2})$, $\phi_1$ and $\phi_2$ being related to the phases of mass matrices and can be treated as free parameters.

For making the manuscript self contained as well as to understand the relationship between diagonalizing transformations for different hierarchies of neutrino masses and for the charged lepton case, we present here the essentials of these transformations. To begin with, we first consider the general diagonalizing transformation $O_k$, whose first element can be written as

$$O_k(11) = \frac{m_2 m_3 (m_3 + m_2 - D_l(\nu_1))}{(m_1 + m_2 + m_3 - D_l(\nu_1))(m_1 - m_3)(m_1 - m_2)},$$

(11)

where $m_1$, $m_2$, $m_3$ are eigenvalues of $M_k$. In the case of charged leptons, because of the hierarchy $m_e \ll m_\mu \ll m_\tau$, the mass eigenstates can be approximated respectively to the flavor eigenstates, as has been considered by several authors [8,11]. In this approximation, $m_{\ell 1} \simeq m_e$, $m_{\ell 2} \simeq m_\mu$, and $m_{\ell 3} \simeq m_\tau$, one can obtain the first element of the matrix $O_l$ from the above equation, Eq. (11), by replacing $m_1$, $m_2$, $m_3$ by $m_e, -m_\mu, m_\tau$, e.g.,

$$O_l(11) = \frac{m_\mu m_3 (m_3 - m_\mu - D_l)}{(m_e - m_\mu + m_\tau - D_l)(m_e - m_\mu)(m_e + m_\mu)},$$

(12)

Eq. (11) can also be used to obtain the first element of diagonalizing transformation for Majorana neutrinos, assuming normal hierarchy, defined as $m_{\nu 1} < m_{\nu 2} < m_{\nu 3}$, and also valid for the degenerate case defined as $m_{\nu 1} \lesssim m_{\nu 2} \sim m_{\nu 3}$, by replacing $m_1$, $m_2$, $m_3$ by $\sqrt{m_{\nu 1} m_R}, -\sqrt{m_{\nu 2} m_R}, -\sqrt{m_{\nu 3} m_R}$, e.g.,

$$O_\nu(11) = \frac{\sqrt{m_{\nu 1} m_R} \sqrt{m_{\nu 3}} (\sqrt{m_{\nu 3}} - \sqrt{m_{\nu 2}} - D_\nu)}{(\sqrt{m_{\nu 3}} - \sqrt{m_{\nu 2}} + \sqrt{m_{\nu 1}} - D_\nu)(\sqrt{m_{\nu 1}} - \sqrt{m_{\nu 2}} + \sqrt{m_{\nu 3}} + D_\nu)(\sqrt{m_{\nu 1}} + \sqrt{m_{\nu 2}} + \sqrt{m_{\nu 3}})}.$$

(13)

where $m_{\nu 1}$, $m_{\nu 2}$, and $m_{\nu 3}$ are neutrino masses. The parameter $D_\nu$ is to be divided by $\sqrt{m_R}$, however as $D_\nu$ is arbitrary therefore we retain it as it is.

In the same manner, one can obtain the elements of diagonalizing transformation for the inverted hierarchy case, defined as $m_{\nu 1} \ll m_{\nu 2} < m_{\nu 3}$, by replacing $m_1$, $m_2$, $m_3$ in Eq. (11) with $\sqrt{m_{\nu 1} m_R}, -\sqrt{m_{\nu 2} m_R}, -\sqrt{m_{\nu 3} m_R}$, e.g.,

$$O_\nu(11) = \frac{\sqrt{m_{\nu 2} m_R} \sqrt{m_{\nu 1}} (D_\nu + \sqrt{m_{\nu 2}} + \sqrt{m_{\nu 3}})}{(-\sqrt{m_{\nu 1}} + \sqrt{m_{\nu 2}} + \sqrt{m_{\nu 3}} + D_\nu)(\sqrt{m_{\nu 1}} + \sqrt{m_{\nu 2}} + \sqrt{m_{\nu 3}})(\sqrt{m_{\nu 1}} + \sqrt{m_{\nu 2}} + \sqrt{m_{\nu 3}})}.$$

(14)

The other elements of diagonalizing transformations in the case of neutrinos as well as charged leptons can similarly be found.

Assuming neutrinos to be Majorana-like, we have carried out detailed calculations pertaining to texture 6 zero as well as two possible cases of texture 5 zero lepton mass matrices, e.g., $D_l = 0$ case and $D_\nu = 0$ case. Corresponding to each of these cases, we have considered three possibilities of neutrino masses having normal/inverted hierarchy or being degenerate. In addition to these 9 possibilities, we have also considered those cases when the charged leptons are in the flavor basis. These possibilities sum up to 18, however, the texture 5 zero $D_\nu = 0$ case with charged leptons in the flavor basis reduces to the similar texture 6 zero case, hence the 18 possibilities reduce to 15 distinct cases.

Before discussing the results, we would like to mention some of the details pertaining to various inputs. The masses and mixing angles, used in the analysis, have been constrained by data given in Eqs. (2) and (3). For the purpose of calculations, we have taken the lightest neutrino mass, the phases $\phi_1$, $\phi_2$ and $D_{l,v}$ as free parameters, the other two masses are constrained by $\Delta m_{12}^2 = m_{\nu 2}^2 - m_{\nu 1}^2$ and $\Delta m_{23}^2 = m_{\nu 3}^2 - m_{\nu 2}^2$ in the normal hierarchy case and by $\Delta m_{23}^2 = m_{\nu 2}^2 - m_{\nu 3}^2$ in the inverted hierarchy case. It may be noted that lightest neutrino mass corresponds to $m_{\nu 1}$ for the normal hierarchy case and to $m_{\nu 3}$ for the inverted hierarchy case. In the case of normal hierarchy, the explored range for $m_{\nu 1}$ is taken to be 0.0001 eV–1.0 eV, which is essentially governed by the mixing angle $\theta_{12}$, related to the ratio $m_{\nu 2}/m_{\nu 3}$. For the inverted hierarchy case also we have taken the same range for $m_{\nu 3}$ as our conclusions remain unaffected even if the range is extended further. In the absence of any constraint on the phases, $\phi_1$ and $\phi_2$ have been given full variation from 0 to 2$\pi$. Although $D_{l,v}$ are free parameters, however, they have been constrained such that diagonalizing transformations, $O_l$ and $O_\nu$, always remain real, implying $D_l < m_{\nu 1} - m_{\nu 2}$ whereas $D_\nu < \sqrt{m_{\nu 1}^2 - m_{\nu 2}^2}$ for normal hierarchy and $D_\nu < \sqrt{m_{\nu 1}^2 - m_{\nu 2}^2}$ for inverted hierarchy.

Out of all the cases considered here, we first discuss those where $M_l$ is taken to be texture specific being the most general ones. We begin with the cases pertaining to inverted hierarchy or when neutrino masses are degenerate. Interestingly, we find that all the cases pertaining to inverted hierarchy and degenerate scenario of neutrino masses seem to be ruled out. For the texture 6 zero case, in Fig. 1, by giving full variations to other parameters, we have plotted the mixing angle $\theta_{23}$ against the lightest neutrino mass. The dotted lines and the dot-dashed lines depict the limits obtained assuming normal and inverted hierarchy respectively, the solid
For texture 5 zero cases, we first discuss the case when $D_l = 0$ and $D_v \neq 0$. Primarily to facilitate comparison with texture 6 zero case, in Fig. 2 we have plotted $s_{23}$ against the lightest neutrino mass for both normal and inverted hierarchy for a particular value of $D_v = \sqrt{m_{\nu_3}}$. Interestingly, we find texture 5 zero $D_l = 0$ case shows a big change in the behaviour of $s_{23}$ versus the lightest neutrino mass as compared to the texture 6 zero case shown in Fig. 1. A closer look at Fig. 2 reveals that the region pertaining to inverted hierarchy, depicted by dot-dashed lines, shows an overlap with the experimental limits on $s_{23}$, depicted by solid horizontal lines, around the region when neutrino masses are almost degenerate. This suggests that in case the degenerate scenario is ruled out, inverted hierarchy is also ruled out. To this end as well as for extending our results to other allowed values of $D_v$, in Fig. 3 we have plotted allowed parameter space for the three mixing angles in the $D_v$-lightest neutrino mass plane, for texture 5 zero $D_l = 0$ case. A closer look at the figure shows that the allowed parameter spaces of the three mixing angles show an overlap when $D_v \sim 0$, which leads to the present texture 6 zero case, wherein degenerate scenario has already been ruled out. The above analysis from Fig. 3 clearly indicates that inverted hierarchy as well as degenerate scenario is ruled out for texture 5 zero $D_l = 0$ case, not only for $D_v = \sqrt{m_{\nu_3}}$ but also for its other acceptable values. Coming to the texture 5 zero $D_v = 0$ and $D_l \neq 0$ case, a plot of $s_{23}$ against the lightest neutrino mass is very similar to Fig. 1 pertaining to the texture 6 zero case, therefore we have not presented it here. By similar arguments, this case is also ruled out for inverted hierarchy as well as for degenerate scenario.

Interestingly, we find that even if we give wider variations to all the parameters, all possible cases considered here pertaining to inverted hierarchy and degenerate scenario are ruled out. It may also be added that in the case when charged leptons are in the flavor basis, the mixing matrix becomes much more simplified and one can easily check that cases pertaining to inverted hierarchy as well as degenerate scenario for the texture 6 zero and 5 zero mass matrices are ruled out. Further, for the sake of completion, we have also investigated the cases when $M_l$ is texture specific or neutrinos are Dirac-like and find that inverted hierarchy and degenerate scenario are again ruled out, details regarding these have been not included here.

After ruling out the cases pertaining to inverted hierarchy and degenerate scenario, we now discuss the normal hierarchy cases.

For texture 6 zero as well as two cases of texture 5 zero mass matrices, in Table 1 we have presented the viable ranges of neutrino masses, mixing angle $s_{13}$, Jarlskog’s rephasing invariant parameter $J$, CP violating phase $\delta$ and effective neutrino mass $\langle m_{ee} \rangle$ related to neutrinoless double beta decay $\langle \beta\beta \rangle_{0\nu}$. The parameter $J$ can be calculated by using its expression given in [8], whereas $\delta$ can be determined from $J = s_{12}s_{23}s_{13}c_{12}c_{23}s_{13}^2\delta$ where $c_{ij} = \cos \theta_{ij}$ and $s_{ij} = \sin \theta_{ij}$, for $i, j = 1, 2, 3$. The effective Majorana mass, measured in $\langle \beta\beta \rangle_{0\nu}$ decay experiment, is given as

$$\langle m_{ee} \rangle = m_{\nu_1}U_{e1}^2 + m_{\nu_2}U_{e2}^2 + m_{\nu_3}U_{e3}^2.$$  (15)

Considering first the texture 6 zero case, the possibility when charged leptons are in flavor basis is completely ruled out, therefore the results presented in Table 1 correspond to the case when $M_l$ is considered texture specific. As can be checked from Table 1, the presently calculated values of parameters $m_{\nu_1}$, $s_{13}$, $J$ and $\langle m_{ee} \rangle$, found by using the latest data, are well within the ranges obtained by Fukugita et al., which are given as $m_{\nu_1} = 0.0004$–0.0030, $s_{13} = 0.04$–0.20, $J \leq 0.025$ and $\langle m_{ee} \rangle = 0.002$–0.007. Also, from the table, one finds the lower limit on $s_{13}$ is 0.066, therefore a measurement of $s_{13}$ would have implications for this case. Similarly, a measurement of effective mass $\langle m_{ee} \rangle$, through the $\langle \beta\beta \rangle_{0\nu}$ decay experiments, would also have implications for these kind of mass matrices. Besides the above mentioned parameters, we have also considered the implications of $s_{13}$ on the phases $\phi_1$ and $\phi_2$. To
therefore, lowering down of $s_{13}$ which we find as

\[ \langle m_{ee} \rangle = 0.0028 - 0.0062 \]

All masses are in eV.

Coming to the texture 5 zero cases, to begin with we consider the $D_l = 0$ case for the inverted hierarchy, with $D_l$ being varied from 0 to a value such that $D_l < \sqrt{m_{ee}^2 - m_{\nu_1}^2}$. Dotted lines depict allowed parameter space for $s_{12}$, dot-dashed lines depict allowed parameter space for $s_{23}$ and solid lines depict allowed parameter space for $s_{13}$.

In this end, in Fig. 4 we have drawn the contours for $s_{13}$ in $\phi_1 - \phi_2$ plane. From the figure it is clear that $s_{13}$ plays an important role in constraining the phases, in particular, we find that if lower limit of $s_{13}$ is on the higher side, then $\phi_1$ is restricted to I or IV quadrant.

Coming to the texture 5 zero cases, to begin with we consider the $D_l = 0$ case. Interestingly, results are obtained for both the possibilities of $M_l$ having Fritzsch-like structure as well as $M_l$ being in the flavor basis. When $M_l$ is assumed to have Fritzsch-like structure, one would like to emphasize a few points. A general look at the table reveals that the possibility of $D_l = 0$ considerably affects the viable range of $m_{\nu_1}$, particularly its lower limit. Similarly, the lower limit of $s_{13}$ is pushed higher. This can be easily understood by noting that $s_{13}$ is more sensitive to variations in $D_l$ than variations in $D_l$. Further, the lower limit of $s_{13}$ is pushed higher as the upper limit of $m_{\nu_1}$ now becomes somewhat lower as compared to the 6 zero case. Also, it may be of interest to construct the Pontecorvo–Maki–Nakagawa–Sakata (PMNS) mixing matrix [14] which we find as

\[
U = \begin{pmatrix}
0.7898 - 0.8571 & 0.5035 - 0.5971 & 0.0761 - 0.1600 \\
0.1845 - 0.4413 & 0.5349 - 0.7459 & 0.5725 - 0.8135 \\
0.3546 - 0.5615 & 0.3926 - 0.6689 & 0.5652 - 0.8107
\end{pmatrix}
\tag{16}
\]

When $M_l$ is considered in the flavor basis, we get a very narrow range of masses, $m_{\nu_1} \sim 0.00063$, $m_{\nu_2} = 0.0086 - 0.0088$ and $m_{\nu_3} = 0.0534 - 0.0546$, for which 5 zero matrices are viable. Also for this case, $s_{13}$ is almost near its upper experimental limit, therefore, lowering down of $s_{13}$ value would almost rule out this case.

Considering the texture 5 zero $D_l = 0$ case, we note that when $M_l$ is considered in the flavor basis, we do not find any viable solution, however when it has Fritzsch-like structure there are a few important observations. The range of $m_{\nu_1}$ gets extended as compared to the 6 zero case, whereas compared to the texture 5 zero $D_l = 0$ case, both the lower and upper limits of $m_{\nu_1}$ have higher values. Interestingly, this case has the widest $s_{13}$ range among all the cases considered here. The PMNS matrix corresponding to this case does not show any major variation compared to the earlier case, except that the ranges of some of the elements like $U_{\mu1}$, $U_{\mu2}$, $U_{\tau1}$ and $U_{\tau2}$ become little wider. This can be understood when one realizes that $D_l$ can take much wider variation compared to $D_v$.

A general look at the table reveals several interesting points. It immediately brings out the fact that the value of $\langle m_{ee} \rangle$, a measure of $(\beta\beta)_{00}$ decay, has more or less the same range for all the cases. This can be understood through Eq. (15) from which one finds that the major contribution to $\langle m_{ee} \rangle$ is given by the term proportional to $m_{\nu_2}$ as the first term gets suppressed by the small value of
whereas the third term gets suppressed by the small value of $U_{e3}^2$. Also, it must be noted that the calculated values of $\langle m_{ee} \rangle$ are much less compared to the present limits of $m_{ee}$ [15], therefore, these do not have any implications for texture 6 zero and texture 5 zero cases. However, the future experiments with considerably higher sensitivities, aiming to measure $\langle m_{ee} \rangle \simeq 3.6 \times 10^{-2}$ eV (MOON [16]) and $\langle m_{ee} \rangle \simeq 2.7 \times 10^{-2}$ eV (CUORE [17]), would have implications on the cases considered here.

In the absence of any definite information about $J$ as well as $\delta$, we find that the ranges corresponding to different cases are in agreement with other similar calculations, however, it is interesting to note that the ranges of $J$ and $\delta$ for the texture 5 zero $D_{13} = 0$ case are much wider than the other two cases. This, perhaps, is not due to any single factor, rather it is due to almost equal contribution of several terms in the case of Majorana neutrinos.

To summarize, using seesaw mechanism and Fritzsch-like texture 6 zero and 5 zero lepton Dirac mass matrices, detailed predictions for 15 distinct possible cases pertaining to normal/inverted hierarchy as well as degenerate scenario of neutrino masses have been carried out. Interestingly, all the presently considered cases pertaining to inverted hierarchy and degenerate scenario seem to be ruled out. Further, inverted hierarchy and degenerate scenario are also ruled out when $M_{13}$ and $M_{e}$ have Fritzsch-like textures.

In the normal hierarchy cases, when the charged lepton mass matrix $M_{l}$ is assumed to be in flavor basis, the texture 6 zero and the texture 5 zero $D_{1} = 0$ case are again ruled out. For the viable texture 6 zero and 5 zero cases, we find the lower limits of $M_{e3}$ and $s_{13}$ would have implications for the texture specific cases considered here. Interestingly, the lower limits of $s_{13}$ for the texture 5 zero $D_{1} = 0$ and $D_{3} = 0$ cases show an appreciable difference. Further, the phase $\phi_{1}$ seems to have strong dependence on the $s_{13}$ value for texture 6 zero as well as texture 5 zero mass matrices. Similarly, the Dirac-like CP violating phase $\delta$ shows very interesting behaviour, e.g., the texture 6 zero case and the texture 5 zero $D_{1} = 0$ case allow the range $0^\circ - 50^\circ$ whereas, the texture 5 zero $D_{1} = 0$ case allows comparatively a larger range $0^\circ - 90^\circ$. The restricted range of $\delta$, in spite of full variation to phases $\phi_{1}$ and $\phi_{2}$, seems to be due to texture structure, hence, any information about $\delta$ would have important implications. The different cases of texture 6 zero and texture 5 zero matrices do not show any divergence for the value of effective mass $\langle m_{ee} \rangle$.

Acknowledgements

The authors would like to thank S.D. Sharma and Sanjeev Verma for useful discussions. M.R. would like to thank the Director, UIET for providing facilities to work. G.A. would like to thank DAE, BRNS for financial support and the Chairman, Department of Physics for providing facilities to work in the department.

References

T. Yanagida, in: O. Sawada, A. Sugamoto (Eds.), Proc. of Workshop on Unified Theory and Baryon Number in the Universe, KEK, Tsukuba, 1979;
P. Ramond, Sanibel talk, retroprinted as hep-ph/9809459;


Fermion masses and mixings in a renormalizable $SO(10) \times \mathbb{Z}_2$ GUT

Walter Grimus *, Helmut Kühböck

Institut für Theoretische Physik, Universität Wien Boltzmanngasse 5, A-1090 Wien, Austria

Received 25 July 2006; received in revised form 20 September 2006; accepted 19 October 2006

Available online 10 November 2006

Abstract

We investigate a scenario in a supersymmetric $SO(10)$ Grand Unified Theory in which the fermion mass matrices are generated by renormalizable Yukawa couplings of the $10 \oplus 120 \oplus 126$ representation of scalars. We reduce the number of parameters by assuming spontaneous CP violation and a $\mathbb{Z}_2$ family symmetry, leading to nine real Yukawa coupling constants for three families. Since in the “minimal SUSY $SO(10)$ GUT” an intermediate seesaw scale is ruled out and our scenario lives in the natural extension of this theory by the $120$, we identify the vacuum expectation value (VEV) $w_R$ of $(10, 1, 3) \in 126$ with the GUT scale of $2 \times 10^{16}$ GeV. In order to obtain sufficiently large neutrino masses, the coupling matrix of the scalar $126$ is necessarily small and we neglect type II seesaw contributions to the light-neutrino mass matrix. We perform a numerical analysis of this 21-parameter scenario and find an excellent fit to experimentally known fermion masses and mixings. We discuss the properties of our numerical solution, including a consistency check for the VEVs of the Higgs-doublet components in the $SO(10)$ scalar multiplets.© 2006 Elsevier B.V. All rights reserved.

1. Introduction

The group $SO(10)$ is a favourite candidate for constructing grand unified theories (GUTs) [1]. The special interest in such theories also stems from the fact that they allow for type I [2] and type II [3] seesaw mechanisms (see also [4]) for the light neutrino masses. Confining oneself to renormalizable $SO(10)$ GUTs, the scalar representations coupling to the chiral fermion fields, which are all assembled for each family in the 16-dimensional irreducible representation (irrep), are determined by the relation [5,6]

\[ 16 \otimes 16 = (10 + 126)_S \oplus 120_{AS}, \tag{1} \]

where the subscripts “S” and “AS” denote, respectively, the symmetric and antisymmetric parts of the tensor product. The so-called “minimal SUSY $SO(10)$ GUT” (MSGUT) [7] makes use of one $10$ and one $126$ scalar irrep for the Yukawa couplings, to account for all fermion masses and mixings [8]. The MSGUT contains, in addition, one $210$ and one $126$ scalar irrep [7]. This model has built-in the gauge-coupling unification of the minimal SUSY extension of the Standard Model (MSSM). Detailed studies of this minimal theory have been performed [9–16]; in [9,14,15] small effects of the 120-plet were considered in addition. It turned out that the MSGUT works surprisingly well in the fermion sector, provided one neglects constraints on the overall scale of the light neutrino masses. This, however, proved to be crucial, since the natural order of the neutrino masses in GUTs is too low, namely $v^2/M_{\text{GUT}} \sim 1.5 \times 10^{-3}$ eV, with $v \sim 174$ GeV and the GUT scale $M_{\text{GUT}} \sim 2 \times 10^{16}$ GeV. Thorough studies of the heavy scalar states [17–22] have been used to show that this MSGUT is too constrained [23,24] and does not allow to enhance the neutrino mass scale to a realistic one [25,26], compatible with the results of the neutrino oscillation experiments (for a review see, e.g., [27]). One aspect of this problem is that a seesaw scale significantly lower than the GUT scale spoils the gauge coupling unification of the MSSM.

* Corresponding author.

E-mail address: walter.grimus@univie.ac.at (W. Grimus).

0370-2693/S – see front matter © 2006 Elsevier B.V. All rights reserved.

doi:10.1016/j.physletb.2006.10.038
An obvious attempt to loosen the corset of the minimal theory is to add the 120-plet of scalars. A study in that direction has been done in [28]. Earlier works considering a prominent 120-plet contribution to the fermion mass matrices are found in [29–32]. We note that $10 \oplus 120$ alone does not give a good fit in the charged fermion sector [33]. Thus the $126$ scalar irrep is not only needed in the neutrino sector but also for the charged fermion mass matrices. In that case, the mass matrices of the charged fermions and the neutrino Dirac-mass matrix are given, respectively, by

\begin{align}
M_d &= k_d H + \kappa_d G + v_d F, \\
M_u &= k_u H + \kappa_u G + v_u F, \\
M_e &= k_\ell H + \kappa_\ell G - 3v_d F, \\
M_D &= k_d H + \kappa_D G - 3v_u F.
\end{align}

The Yukawa coupling matrices $H$, $G$, $F$ belong to the scalar irreps $10$, $120$, $126$, respectively. The coefficients $k_d$, $k_\ell$, $v_d$ denote the vacuum expectation values (VEVs) of the Higgs doublet components in the respective $SO(10)$ scalar irreps which contribute to the MSSM Higgs doublet $H_d$, the rest of the coefficients refers to $H_u$. The light neutrino mass matrix is obtained as

$$
\mathcal{M}_\nu = M_L - M_D M_R^{-1} M_D^T \text{ with } M_L = w_L F, \; M_R = w_R F,
$$

with scalar triplet VEVs $w_L$ and $w_R$. The mass Lagrangian of the “light” fermions reads

$$
\mathcal{L}_M = -\bar{\psi}_L M_d d_R - \bar{u}_L M_u u_R - \bar{\ell}_L M_\ell \ell_R - \frac{1}{2} \bar{v}_L \mathcal{M}_\nu (v_L)^c + \text{H.c.},
$$

with $(v_L)^c$ being the charge-conjugate of $v_L$.

2. A renormalizable $SO(10)$ scenario

The goal of this Letter is a numerical study of the system of 3-generation mass matrices (2) to (6), taking into account the neutrino-mass suppression factor $v^2/M_{\text{GUT}}$. This system does not easily lend itself to such an investigation because it contains many parameters, thus we use some arguments to reduce their number. The scenario we want to investigate is defined by the following assumptions:

(i) The Yukawa coupling matrices $H$, $G$, $F$ are real.
(ii) We impose a $Z_2$ symmetry, which sets some of the Yukawa couplings to zero and which is spontaneously broken by the VEVs of the $120$, in particular, by $k_d$, $k_u$, $k_\ell$, $\kappa_D$ being non-zero.
(iii) We assume $w_R = M_{\text{GUT}}$, with $M_{\text{GUT}} = 2 \times 10^{16}$ GeV.
(iv) We set $w_L = 0$, i.e., we have pure type I seesaw mechanism.

Let us now comment on these items. Item (i) can be motivated by spontaneous CP violation. The $Z_2$ of item (ii) is given by

$$
\psi_2 \to -\psi_2, \quad \phi_{120} \to -\phi_{120}.
$$

where the $\psi_j$ ($j = 1, 2, 3$) denote the fermionic 16-plets and $\phi_{120}$ is the scalar 120-plet. All other multiplets, not mentioned in Eq. (8), transform trivially. With the $Z_2$ symmetry of Eq. (8), the coupling matrices have the form

$$
H = \begin{pmatrix}
h_{11} & 0 & 0 \\
0 & h_{22} & 0 \\
0 & 0 & h_{33}
\end{pmatrix}, \quad G = \begin{pmatrix} 0 & g_{12} & 0 \\
-g_{12} & 0 & g_{23} \\
0 & -g_{23} & 0
\end{pmatrix}, \quad F = \begin{pmatrix} f_{11} & 0 & f_{13} \\
0 & f_{22} & 0 \\
f_{13} & 0 & f_{33}
\end{pmatrix}.
$$

We have used the freedom of basis choice in the 1–3 sector to set $h_{13} = 0$. Of course, this $Z_2$ symmetry of Eq. (8) is an ad-hoc symmetry, but it enhances the importance of the $120$ because its Yukawa coupling matrix $G$ is now responsible for mixing of the second family with the other two.\footnote{In Eq. (8), for the definition of the $Z_2$ symmetry, all choices $\psi_j \to -\psi_j$ are equivalent. With choosing $\psi_2$, we anticipate the result of the fit of our scenario to the masses and mixings at the GUT scale. That fit gives a strong hierarchy of the elements of $H$, which—with Eq. (8)—can be formulated in the usual way as $|h_{11}| \ll |h_{22}| \ll |h_{33}|$. Furthermore, with the convention of (8) it is possible to have all diagonalizing matrices of the charged fermion masses in the vicinity of the unit matrix.} Item (iii) is motivated by the fact that the MSGUT does not allow to fix the problem of too small neutrino masses by taking $w_R$ significantly lower than the GUT scale [23–26,28]. Thus our scenario has built in that the natural neutrino mass scale in the MSGUT is too low. Consequently, the neutrino mass scale has to be enhanced by the smallness of the coupling matrix $F$ [28]. Item (iv) is a trivial consequence of the previous one: for small $F$, type II seesaw contribution to $\mathcal{M}_\nu$ is negligible.
Now we tackle the problem of parameter counting. Without loss of generality, we assume that $k_d$, $k_u$ and $w_R$ are real and positive. Then we define

$$H' = k_d H, \quad G' = |\kappa_d| G, \quad F' = |\nu_d| F.$$  

The primed matrices have the dimension of mass. The phases of the VEVs of the 120 and 126-plets cannot be removed. Thus we write the mass matrices as

$$M_d = H' + e^{i\delta_d} G' + e^{i\zeta_d} F', \quad (11)$$
$$M_u = r_H H' + r_d e^{i\delta_d} G' + r_F e^{i\zeta_d} F', \quad (12)$$
$$M_e = H' + r_e e^{i\delta_d} G' - 3 e^{i\zeta_d} F', \quad (13)$$
$$M_D = r_H H' + r_D e^{i\delta_d} G' - 3 r_F e^{i\zeta_d} F', \quad (14)$$
$$\kappa_v = r_R M_D F R^{-1} M_D^T.$$  

The ratios $r_H$, etc., are real by definition since we have extracted the phases from the VEVs. Now the counting is easily done. Since we have nine real Yukawa couplings, see Eq. (9), there are nine real parameters in $H'$, $G'$, $F'$. Furthermore, there are six phases and six (real) ratios of VEVs, altogether 21 real parameters. On the other hand, we have 18 observables we want to fit: nine charged-fermion masses, three mixing angles and one CP phase in the CKM matrix, two neutrino mass-squared differences $\Delta m^2_{\text{atm}}$ and $\Delta m^2_{\text{sol}}$, and three lepton mixing angles.

Suppose, we have obtained a good fit for the 18 observables. Then we still have to check if the fit allows for reasonable VEVs and Yukawa coupling constants. A detailed discussion of this issue is found in Appendix A. Here it is sufficient to note that

$$|\nu_d|^2 + |\nu_u|^2 = |\nu_d|^2 (1 + r_F^2) < v^2 \quad \text{with} \quad v = 174 \text{ GeV}$$  

for every fit. Clearly, this inequality holds at the electroweak scale, and we assume that approximately it is valid at the GUT scale too.

3. A numerical solution

To find a numerical solution, we employ the downhill simplex method [34] for minimizing a $\chi^2$-function of the parameters—for an explanation of the method see [26,33]. Actually, the $\chi^2$-function can be minimized analytically with respect to the parameter $r_R$ of Eq. (15), which results in a $\chi^2$-function depending on the remaining 20 parameters, and we apply our numerical method to that function. To build in the inequality (16) in our search for the minimum, we add a suitable penalty function to our

$$\kappa_v = r_R M_D F R^{-1} M_D^T.$$  

function. To build in the inequality (16) in our search for the minimum, we add a suitable penalty function to our

$$\kappa_v = r_R M_D F R^{-1} M_D^T.$$  

for every fit. Clearly, this inequality holds at the electroweak scale, and we assume that approximately it is valid at the GUT scale too.

Choosing the normal ordering $m_1 < m_2 < m_3$ of the neutrino masses ($\Delta m^2_\odot = m_2^2 - m_1^2$, $\Delta m^2_{\text{atm}} = m_3^2 - m_2^2$), we have found a fit with a $\chi^2 = 0.0087$, which is a perfect fit for all practical purposes. This fit is so good that it does not make sense to show the

<table>
<thead>
<tr>
<th>Quarks</th>
<th>Leptons</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_d$</td>
<td>$m_e$</td>
</tr>
<tr>
<td>1.5036 ± 0.2304</td>
<td>0.3585 ± 0.0003</td>
</tr>
<tr>
<td>$m_s$</td>
<td>$m_\mu$</td>
</tr>
<tr>
<td>29.9454 ± 4.444</td>
<td>75.6715 ± 0.501</td>
</tr>
<tr>
<td>$m_b$</td>
<td>$m_\tau$</td>
</tr>
<tr>
<td>1063.6 ± 141.4</td>
<td>1292.2 ± 0.012</td>
</tr>
<tr>
<td>$m_\tau$</td>
<td>$\Delta m^2_\odot$</td>
</tr>
<tr>
<td>0.7238 ± 0.1467</td>
<td>(7.9 ± 0.3) × 10^{-5}</td>
</tr>
<tr>
<td>$m_e$</td>
<td>$\Delta m^2_{\text{atm}}$</td>
</tr>
<tr>
<td>210.327 ± 21.2264</td>
<td>(2.5 ± 0.37) × 10^{-3}</td>
</tr>
<tr>
<td>$m_\mu$</td>
<td>$s_{12}$</td>
</tr>
<tr>
<td>8243.3 ± 1478.6</td>
<td>0.31 ± 0.025</td>
</tr>
<tr>
<td>$m_\tau$</td>
<td>$s_{23}$</td>
</tr>
<tr>
<td>210.327 ± 21.2264</td>
<td>0.50 ± 0.065</td>
</tr>
<tr>
<td>$s_{12}$</td>
<td>$s^2_{13}$</td>
</tr>
<tr>
<td>0.2243 ± 0.0016</td>
<td>&lt; 0.0155</td>
</tr>
<tr>
<td>$\delta_{\text{CKM}}^\mu$</td>
<td>$\delta_{\text{CKM}}^\tau$</td>
</tr>
<tr>
<td>60° ± 14°</td>
<td>60° ± 14°</td>
</tr>
</tbody>
</table>
Table 2
The values of the phases and ratios appearing in the mass matrices (11)–(15) in the case of our fit. Hyphens in the left two columns indicate that the ratio corresponding to the phase has been absorbed in one of the primed matrices, whereas hyphens in the right two columns signify that there is no physical phase associated with that ratio.

<table>
<thead>
<tr>
<th>r_H</th>
<th>91.0759</th>
<th>–</th>
<th>–</th>
</tr>
</thead>
<tbody>
<tr>
<td>r_F</td>
<td>297.758</td>
<td>ζ_d</td>
<td>–</td>
</tr>
<tr>
<td>–</td>
<td>–</td>
<td>ζ_d</td>
<td>–</td>
</tr>
<tr>
<td>r_u</td>
<td>7.14572</td>
<td>ξ_u</td>
<td>–</td>
</tr>
<tr>
<td>r_ℓ</td>
<td>1.33897</td>
<td>ξ_ℓ</td>
<td>–</td>
</tr>
<tr>
<td>r_D</td>
<td>3008.88</td>
<td>ξ_D</td>
<td>–</td>
</tr>
<tr>
<td>r_R</td>
<td>2.90553 × 10^{-17}</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Fig. 1. The $\chi^2$ as a function of $|v_d|\sqrt{1+r_F^2}$.

pulls. The matrices $H'$, $G'$ and $F'$ for our fit are given by

$$H' = \begin{pmatrix} 0.716986 & 0 & 0 \\ 0 & -40.6278 & 0 \\ 0 & 0 & 1114.41 \end{pmatrix}, \quad G' = \begin{pmatrix} 7.56737 & 0 & 36.8224 \\ 0 & -36.8224 & 0 \end{pmatrix},$$

$$F' = \begin{pmatrix} -0.0966851 & 0 & 4.25282 \\ 0 & 12.3136 & 0 \\ 4.25282 & 0 & -61.6491 \end{pmatrix},$$

where all numerical values are in units of MeV; the values of the ratios of VEVs and the phases are shown in Table 2.

The neutrino mass spectrum turns out to be hierarchical with $m_1 = 1.57 \times 10^{-3}$ eV $\ll m_2 = 9.03 \times 10^{-3}$ eV $\ll m_3 = 46.96 \times 10^{-3}$ eV, and the PMNS phase is $12^\circ$. We want to stress, however, that our fit solution is perhaps not unique, because with the numerical method used here we could miss other minima of $\chi^2$. For our fit, it turns out that $y \equiv |v_d|\sqrt{1+r_F^2} = 173.0$ GeV. This looks dangerously close to the upper bound of Eq. (16). To check if this danger is serious, we have plotted in Fig. 1 the minimal $\chi^2$ as a function of $y$. In order to pin $y$ down to a given value $\tilde{y}$ we have extended the $\chi^2$ function to $(\chi^2_y) = \chi^2 + \left( (y - \tilde{y})/(0.01\tilde{y}) \right)^2$, minimized $(\chi^2)_y$ and plotted $\chi^2$ at this minimum versus $\tilde{y}$—for previous uses of this method see, for instance, [26]. We read off from Fig. 1 that $\chi^2$ is minimal at $y = 173$ GeV, however, this minimum is rather flat; note that $\chi^2$ is plotted on a logarithmic scale. Thus we still obtain excellent fits if we go to lower values of $y$. In Appendix A, a consistency condition is worked out which the $SO(10)$ GUT has to fulfill in order to reproduce the VEV ratios of Table 2. There we also show that for our fit all Yukawa couplings stay in the perturbative regime.

For our fit, it turns out that $y \equiv |v_d|\sqrt{1+r_F^2} = 173.0$ GeV. This looks dangerously close to the upper bound of Eq. (16). To check if this danger is serious, we have plotted in Fig. 1 the minimal $\chi^2$ as a function of $y$. In order to pin $y$ down to a given value $\tilde{y}$ we have extended the $\chi^2$ function to $(\chi^2_y) = \chi^2 + \left( (y - \tilde{y})/(0.01\tilde{y}) \right)^2$, minimized $(\chi^2)_y$ and plotted $\chi^2$ at this minimum versus $\tilde{y}$—for previous uses of this method see, for instance, [26]. We read off from Fig. 1 that $\chi^2$ is minimal at $y = 173$ GeV, however, this minimum is rather flat; note that $\chi^2$ is plotted on a logarithmic scale. Thus we still obtain excellent fits if we go to lower values of $y$. In Appendix A, a consistency condition is worked out which the $SO(10)$ GUT has to fulfill in order to reproduce the VEV ratios of Table 2. There we also show that for our fit all Yukawa couplings stay in the perturbative regime.

In order to find out if our scenario makes a prediction for the PMNS phase $\delta$, we treat it in the same way as $y$ in the previous paragraph, i.e., we consider $(\chi^2)_{\delta} = \chi^2 + \left( (\delta - \tilde{\delta})/(0.01\tilde{\delta}) \right)^2$. Departing from $\tilde{\delta} = 12^\circ$ for our numerical solution given by Eq. (17) and Table 2, and going stepwise down close to $\delta = 0^\circ$ and up to $\delta = 360^\circ$, the quality of the fits remains excellent, with $(\chi^2)_{\delta}$ always below 0.3. Thus, in our scenario all values of the PMNS phase are possible.

\footnote{The largest pull is $5 \times 10^{-2}$ for $m_3$.}

\footnote{We use the same phase convention as for the CKM matrix in [36].}
One may ask the question how large enough neutrino masses and an atmospheric mixing angle which is close to maximal are accomplished with the numerical values given by Eq. (17) and Table 2. We concentrate on achieving $\sqrt{\Delta m_{\text{atm}}^2} \sim 0.05$ eV. With the value of $r_R$ we find that $r_R \times 10^9$ MeV $\simeq 0.029$ eV. Thus we take into account all contributions to $\mathcal{M}_v/r_R$ which are of order $10^9$ MeV. Rewriting Eq. (15) in the form

$$\mathcal{M}_v = r_R \left\{ r_H H' + r_D e^{i\beta} G' F^{-1} \left( r_H H' + r_D e^{i\beta} G' \right)^T - 6r_H r_D e^{i\xi_u} H' + 9r_D e^{2i\xi_u} F \right\},$$

(18)
a numerical analysis shows that all elements of the second and third term on the right-hand side are smaller than $1.8 \times 10^8$ MeV and $0.5 \times 10^8$ MeV, respectively. Thus the dominant matrix elements in $\mathcal{M}_v$ stem from the first term and are given by

$$\mathcal{M}_v^{(\text{dom})} = r_R r_D \frac{1}{d} \begin{pmatrix} 0 & 0 & 0 \\ 0 & r_D (2g'_{23}g'_{12}f'_{13} + f'_{33}(g'_{12})^2)/d & -r_H h'_{33}g'_{12}f'_{13}/d \\ 0 & -r_H h'_{33}g'_{12}f'_{13}/d & r_D (g'_{33})^2 / f'_{22} \end{pmatrix}.$$ (19)

Here, $d = f'_{11}f'_{33} - (f'_{13})^2$ is the determinant of the corresponding $2 \times 2$ submatrix of $F'$ and we have used the approximation $\xi_D = 180^\circ$. Apart from the common factor $r_R$, in this matrix there are four products of three matrix elements: one matrix element is always from $F'^{-1}$, the other two are either both from $r_D G'$ or one from $r_D G'$ and one from $r_H H'$. Looking at Eq. (17) and Table 2, we find that products of the largest elements, for instance $(g'_{23})^2 f'_{33}$, never occur, these would be too large. Plugging in the numerical values of the parameters, we find that all non-zero terms in $\mathcal{M}_v^{(\text{dom})}$ are similar in magnitude:

$$r_R \times 10^9 \text{ MeV} \begin{pmatrix} -1.77 + 2.64 & -0.81 \\ -0.81 & 1.00 \end{pmatrix}.$$ (20)

Due to the minus sign in the first term, we end up with $-1.77 + 2.64 = 0.87$, close to 1.00 and thus leading to nearly maximal mixing. In this crude approximation, which is only relevant for the largest mass and the atmospheric mixing angle, we obtain $m_3 \simeq 0.05$ eV and $\theta_{23} \simeq 43^\circ$. Apart from the smallness of $F'$, it is the large factor $r_D$ in $M_D$ which gives the correct magnitude of the neutrino masses. Maximal atmospheric neutrino mixing rather looks like a numerical contrivance in our scenario.

Finally, a word concerning the low-energy values of the quark masses is in order. Ref. [35] takes the quark mass values at $m_Z$ from Ref. [37] as input for the renormalization group evolution up to $M_{\text{GUT}}$, whereas Ref. [37] uses the input

$$m_u(1 \text{ GeV}) = 4.88 \pm 0.57, \quad m_d(1 \text{ GeV}) = 9.81 \pm 0.65, \quad m_s(1 \text{ GeV}) = 195.4 \pm 12.5$$ (21)

and

$$m_c(m_c) = 1.302^{+0.037}_{-0.038}, \quad m_b(m_b) = 4.34^{+0.07}_{-0.08}, \quad m_t(m_t) = 171 \pm 12,$$ (22)

see Tables I and II in [37]. The light quark masses are given in MeV, the heavy ones in GeV. Comparing these values with those given in the Review of Particle Properties of 2006 (RPP) [36], we find that the heavy quark masses are in reasonable agreement. However, in the last years the values of the light quark masses have become significantly lower [36]:

$$m_u(2 \text{ GeV}) = 1.5 \pm 3.0, \quad m_d(2 \text{ GeV}) = 3 \div 7, \quad m_s(2 \text{ GeV}) = 95 \pm 25.$$ (23)

Note that one has to take into account the scaling factor $m_i(1 \text{ GeV})/m_i(2 \text{ GeV}) = 1.35 \ (i = u, d, s)$ to compare Eq. (21) with Eq. (23) [36]. In order to assess the influence of lowering the light quark masses, we have performed a second fit, using the values of Eq. (23) as input, scaled to $M_{\text{GUT}}$ with the factor 0.200 for $m_u$ and 0.207 for $m_d$ and $m_s$ (see [35,37]), but leaving the previous values for the heavy quark masses. We found an excellent fit with $\chi^2 = 0.052$, which means that our scenario is able to reproduce the lower values of the light quark masses as well. The second fit has some qualitative differences in comparison with the first one, which reinforces the suspicion that, for given input values of the 18 observables, the fit solution in our scenario is not unique.

4. Summary

In this Letter we have investigated fermion masses and mixings in the $SO(10)$ MSGUT, augmented by a 120-plet of scalars. The main purpose was to show that in this setting it is possible to reconcile the type I seesaw mechanism (see Eq. (6)) with a triplet VEV $v_\mu$ equal to the GUT scale of $2 \times 10^{16}$ GeV, provided the theory admits that the MSSM Higgs doublet $H_D$ is composed mainly of the corresponding doublet components in the 126 and 210 scalar irreps—see Eq. (A.11); those are the irreps which have no Yukawa couplings. This reconciliation was feasible within the scenario defined in points (i)–(iv), in which we have used symmetries to significantly reduce the number of degrees of freedom in the Yukawa couplings—see Eq. (9). Within this scenario we were able to find an excellent fit for all fermion masses and mixings; in this fit we have a hierarchical neutrino mass spectrum.4

5 Note that there is also a small contribution from $M_t$.

We have also tried fits for the inverted ordering. In that case, the best fit we found has $\chi^2 = 1.8$ and $m_3 \simeq 7 \times 10^{-6}$ eV.
Thus we have obtained the following results for the minimal renormalizable $SO(10)$ GUT, with Yukawa couplings according to the relation (1):

- It is possible to reproduce the correct neutrino mass scale.
- Nevertheless, gauge coupling unification is not spoiled.
- The concrete $SO(10)$ scenario with type I seesaw mechanism, we treated in this Letter, has 21 parameters, just as the MSGUT with type I+II seesaw mechanism and general complex Yukawa couplings.

It remains to be studied if our scenario allows a sufficient suppression of proton decay. In [32] it was shown that the scalar 120-plet plays a crucial role for that purpose; a certain texture of the Yukawa coupling matrices—similar to our numerical solution (17)—enables that suppression even for large tan $\beta$.

Acknowledgements

W.G. thanks C.S. Aulakh for illuminating discussions and L. Lavoura for reading the manuscript.

Appendix A. The MSSM Higgs doublets and the mass matrices

The MSSM contains two Higgs doublets, $H_d$ and $H_u$, with hypercharges $+1$ and $-1$, respectively. Their corresponding VEVs are denoted by $v \cos \beta$ and $v \sin \beta$ ($v = 174$ GeV), respectively. Neglecting effects of the electroweak scale, these doublets are, by assumption, the only scalar zero modes extant at the GUT scale; this requires a minimal finetuning condition [17,19]. The scalar irreps $10, \bar{126}, 126, 210$ contain each one doublet with the quantum numbers of $H_d$, whereas the $120$ contains two such doublets. The $H_d$ is composed of these doublets [19] with the corresponding amplitudes [25] $\bar{a}_j$ ($j = 1, \ldots, 6$). The analogous coefficients for $H_u$ are denoted by $\alpha_j$. The normalization conditions are

$$\sum_{j=1}^{6} |\alpha_j|^2 = \sum_{j=1}^{6} |\bar{a}_j|^2 = 1.$$  \hspace{1cm} (A.1)

The Dirac mass matrices, taking into account that the $126$ and $210$ have no Yukawa couplings, are given by

$$M_d = v \cos \beta (c_1^d \bar{a}_1 Y_{10} + c_2^d \bar{a}_2 Y_{\bar{126}} + (c_3^d \bar{a}_3 + c_5^d \bar{a}_5) Y_{120}) \quad (a = d, \ell). \hspace{1cm} (A.2)$$

$$M_u = v \sin \beta (c_1^b \bar{a}_1 Y_{10} + c_2^b \bar{a}_2 Y_{\bar{126}} + (c_3^b \bar{a}_3 + c_5^b \bar{a}_5) Y_{120}) \quad (b = u, D). \hspace{1cm} (A.3)$$

with Yukawa coupling matrices $Y_{10}, Y_{\bar{126}}, Y_{120}$ and Clebsch–Gordan coefficients $c_j^{a,b}$ deriving from the $SO(10)$-invariant Yukawa couplings [20,23]. The absolute values of the Clebsch–Gordan coefficients have no physical meaning and some of their phases are convention-depended. With our conventions, the required information reads

$$c_1^d = c_1^u = c_1^D,$$  \hspace{1cm} (A.4)

$$c_2^d = -c_2 = -\frac{1}{3} c_2^D,$$

$$c_3^d = c_3^u = \frac{1}{3} c_3^D,$$

$$c_5^d = c_5^u = c_5^D,$$  \hspace{1cm} (A.5)

Eqs. (A.2) and (A.3) together with this equation lead to the mass matrices (2)–(5). Furthermore, comparing Eqs. (A.2) and (A.3) with Eq. (10), we find

$$H' = v \cos \beta c_1^d \bar{a}_1 |Y_{10}, \quad F' = v \cos \beta c_2^d \bar{a}_2 Y_{\bar{126}}, \quad G' = v \cos \beta (c_3^d \bar{a}_3 + c_5^d \bar{a}_5) Y_{120}. \hspace{1cm} (A.5)$$

Comparison with Eqs. (11)–(14) and using Eq. (A.4) delivers the coefficients

$$r_H = \tan \beta \left| \frac{\alpha_1}{\bar{a}_1} \right|, \quad r_F = \tan \beta \left| \frac{\alpha_2}{\bar{a}_2} \right|, \hspace{1cm} (A.6)$$

$$r_u = \tan \beta \left| \frac{\alpha_6 - \sqrt{3} \alpha_5}{\alpha_6 - \sqrt{3} \alpha_5} \right|, \quad r_\ell = \left| 1 - \frac{2 \bar{a}_6}{\alpha_6 - \sqrt{3} \alpha_5} \right|, \quad r_D = \tan \beta \left| \frac{3 \alpha_6 + \sqrt{3} \alpha_5}{\alpha_6 - \sqrt{3} \alpha_5} \right|. \hspace{1cm} (A.7)$$

Now we want to check the consistency of our numerical solution given by Eq. (17) and Table 2. From $r_D \gg r_u$, it follows that

$$\sqrt{3} \alpha_5 \simeq \alpha_6 \simeq \frac{r_D}{4 \tan \beta} |\bar{a}_6 - \sqrt{3} \bar{a}_5|. \hspace{1cm} (A.8)$$
Furthermore, using $r_\ell \sim 1$, we find the order-of-magnitude relations

$$\tilde{a}_5 \sim \tilde{a}_6 \sim \tan \beta / r_D. \quad (A.9)$$

Then the first of the normalization conditions (A.1) reads approximately

$$\sum_j |a_j|^2 \simeq \frac{1}{\tan^2 \beta} \left( r_H^2 \tilde{a}_1^2 + r_\ell^2 |\tilde{a}|^2 + \frac{1}{12} r_D^2 \tilde{a}_6 - \sqrt{3} \tilde{a}_5^2 \right) + |a_3|^2 + |a_4|^2 \simeq 1. \quad (A.10)$$

This means that $|\tilde{a}_j|^2 \ll 1$ for $j = 1, 2, 5, 6$. Therefore, the second normalization condition is given by

$$\sum_j |\tilde{a}_j|^2 \simeq |\tilde{a}_3|^2 + |\tilde{a}_4|^2 \simeq 1, \quad (A.11)$$

and the brunt of the normalization has to be supplied by the components of $H_D$ in the $126$ and $210$, which do not couple to the fermions. This is a consistency condition for the scenario presented in this Letter.

To translate the condition (16) into the formalism presented here, we note that $|v_d|^2 \ll |v_\ell|^2$ and $\sin \beta \simeq 1$ for $\tan \beta = 10$. Therefore, Eq. (16) effectively checks if the necessary condition $|\tilde{a}_2| \ll 1$ is fulfilled.

Finally, it remains to see if our numerical solution respects the perturbative regime of the Yukawa sector. It suffices to consider the largest elements of the Yukawa couplings

$$Y_d = \frac{1}{\sqrt{v \cos \beta}} M_d, \quad Y_u = \frac{1}{\sqrt{v \sin \beta}} M_u, \quad Y_{\ell} = \frac{1}{\sqrt{v \cos \beta}} M_{\ell}, \quad Y_D = \frac{1}{\sqrt{v \sin \beta}} M_D, \quad (A.12)$$

which reside in $Y_u$ and $Y_D$. The largest entry in $Y_u$ is the 33-element with the main contribution from $r_H h_{33}^\dagger / (v \sin \beta) \simeq 0.59$. The 23-element with $r_D g_{23}^\dagger / (v \sin \beta) \simeq 0.64$ dominates in $Y_D$. These numbers demonstrate that for our numerical solution the Yukawa couplings remain in the perturbative regime.

References

[34] J.A. Nelder, R. Mead, Comp. J. 7 (1965) 306;
Non-supersymmetric charged domain walls

Jan B. Gutowski a, Wafic A. Sabra b,∗

a DAMTP, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK
b Centre for Advanced Mathematical Sciences and Physics Department, American University of Beirut, Lebanon

Received 16 October 2006; accepted 26 October 2006
Available online 9 November 2006

Abstract

We present general non-supersymmetric domain wall solutions with non-trivial scalar and gauge fields for gauged five-dimensional \( N = 2 \) supergravity coupled to Abelian vector multiplets.

© 2006 Elsevier B.V. All rights reserved.

1. Introduction

The recent interest in the study of solutions of gauged supergravity theories in various dimensions has to a large extent been motivated by the conjectured Anti-de Sitter/Conformal Field Theory (AdS/CFT) equivalence [1]. From the CFT perspective, supergravity vacua could correspond to an expansion around non-zero vacuum expectation values of certain operators, or describe a holographic renormalization group flow [2]. It is hoped that this conjectured equivalence can help in gaining some understanding of the non-perturbative structure of gauge theories by studying classical supergravity solutions. Of particular interest are the domain walls of gauged five-dimensional \( N = 2 \) supergravity theories. The theory of ungauged five-dimensional \( N = 2 \) supergravity coupled to Abelian vector supermultiplets can be obtained by compactifying eleven-dimensional supergravity, the low-energy limit of M-theory, on a Calabi–Yau three-fold [3]. Another class of models are those obtained in [4] which are closely related to Jordan algebras. The gauged five-dimensional \( N = 2 \) supergravity theories we consider are those obtained by gauging the \( U(1) \) subgroup of the \( SU(2) \) automorphism group of the superalgebra [4]. The gauging is accomplished by introducing into the Lagrangian of the theory a linear combination of the Abelian vector fields already present in the ungauged theory, i.e. \( A_\mu = V_I A^I_\mu \), with a coupling constant \( \chi \). The coupling of the fermions of the theory to the \( U(1) \) vector field breaks supersymmetry and gauge-invariant terms are added to preserve \( N = 2 \) supersymmetry. In terms of the bosonic action of the theory, we get an additional \( \chi^2 \)-dependent scalar potential \( V \) [4].

Most domain wall solutions constructed so far are configurations preserving some of the supersymmetries (see for example [5]). Explicit supersymmetric domain wall solutions for the theories of [4], where the scalar fields live on symmetric spaces, were given in [6]. These solutions, describing a holographic renormalization group flow, were expressed in terms of Weierstrass elliptic function. Recently a systematic approach has been employed in the classification of general supersymmetric solutions of the gauged five-dimensional supergravity with non-trivial vector multiplets [7,8]. In [8], the requirement that the scalar manifold is a symmetric space was relaxed and the structure of solutions with null Killing vector in both gauged and ungauged supergravity theories was also investigated. In our present work we are interested in finding non-supersymmetric domain wall solutions. We present non-supersymmetric charged domain wall solutions with non-trivial scalars for all gauged five-dimensional \( N = 2 \) supergravity models coupled to vector multiplets. We organize our work as follows. In section two, and in an attempt to make our work self-contained, a brief review of the theories of the \( U(1) \)-gauged supergravity and their equations of motion are given. In section three, the domain wall solutions of [6] are presented as

∗ Corresponding author.
E-mail address: wafic.sabra@gmail.com (W.A. Sabra).
well as general domain wall solutions applicable for a Calabi–Yau compactification [8]. The analysis and the derivation of non-supersymmetric charged domain wall solutions are given in section four. Section five includes the study of the causal structure of the domain walls geometry and we conclude in section six.

2. Gauged five-dimensional \( N = 2 \) supergravity

The bosonic action of the gauged \( D = 5 \) \( N = 2 \) supergravity can be written as [4]

\[
S = \frac{1}{16\pi G} \int \left( R + 2\chi^2 \nabla^2 - G_{IJ} F^I \wedge * F^J - G_{IJ} dX^I \wedge \star dX^J - \frac{1}{6} C_{IJK} F^I \wedge F^J \wedge A^K \right),
\]

(2.1)

where \( I, J, K \) take values \( 1, \ldots, n \) and \( F^I = dA^I \). Here the \( X^I \) represent the scalar fields of the theory, which are constrained via

\[
\frac{1}{6} C_{IJK} X^I X^J X^K = 1,
\]

and may be regarded as being functions of \( n - 1 \) unconstrained scalars \( \phi^a \). In addition, the couplings \( G_{IJ} \) depend on the scalars via

\[
G_{IJ} = \frac{9}{2} X_I X_J - \frac{1}{2} C_{IJK} X^K,
\]

(2.2)

where \( X_I \equiv \frac{1}{6} C_{IJK} X^J X^K \) and therefore one has the following useful relations

\[
G_{IJ} X^J = \frac{3}{2} X_I, \quad G_{IJ} \partial_a X^J = -\frac{3}{2} \partial_a X_I.
\]

(2.3)

The scalar potential of the gauged theory can be written in the form

\[
\mathcal{V} = 9V_I V_J \left( X^I X^J - \frac{1}{2} G_{IJ} \right),
\]

(2.4)

where \( V_I \) are constants [4]. The Einstein equations derived from the action (2.1) are given by

\[
R_{\mu\nu} = G_{IJ} \left( F^I_{\mu\kappa} F^J^{\nu\lambda} - \frac{1}{6} \delta_{\mu\nu} F^I_{\rho\sigma} F^{J\rho\sigma} \right) + \nabla_\mu X^I \nabla_\nu X^J - \frac{2}{3} \chi^2 \nabla g_{\mu\nu}.
\]

(2.5)

The Maxwell equations are

\[
d (G_{IJ} \ast F^J) + \frac{1}{4} C_{IJK} F^J \wedge F^K = 0.
\]

(2.6)

The scalar equations of motion give the following relations [7,8]

\[
d (\ast dX_I) - \left( \frac{1}{6} C_{MNI} - \frac{1}{2} X_I C_{MNI} X^J \right) dX^M \wedge \ast dX^N
+ \left( X_M X^P C_{NPJ} - \frac{1}{6} C_{MNI} - 6 X_J X_M X_N \right)
+ \frac{1}{6} X_I C_{MNI} X^J \right) F^M \wedge \ast F^N
+ \frac{1}{6} X_I C_{MNI} X^J \right) F^M \wedge \ast F^N
+ \frac{1}{6} \chi^2 \left( \frac{1}{2} V_M V_N G^{MF} G^{NF} C_{LPJ} + X_J G^{MN} V_M V_N \right) \text{dvol} = 0.
\]

(2.7)

3. Supersymmetric domain walls

In this section we review the supersymmetric domain wall solutions found in [6] as well as the general supersymmetric solutions with null Killing vector and with vanishing gauge fields presented in [8]. In [6] one starts with the following general ansatz for supersymmetric domain wall solutions:

\[
ds^2 = e^{2A} (-dt^2 + dz^2) + e^{2B} (dr^2 + r^2 d\theta^2 + r^2 d\phi^2),
\]

(3.1)

where \( A \) and \( B \) are functions of \( r \) only. The analysis of the Killing spinor equations (obtained from the vanishing of the gravitini and dilatino supersymmetric variations) gives the following restrictions on the metric and the scalar fields,

\[
\partial_r A - \partial_r B - \frac{1}{r} = 0,
\]

(3.2)

Eq. (3.2) implies

\[
X_I = \frac{1}{r^2} e^{-2B} \left[ -2X^I \int e^{2B} r^2 dr + \Lambda_I \right],
\]

(3.3)

where the \( \Lambda_I \) are integration constants. For symmetric spaces where

\[
C^{IJK} = \delta^{I'} \delta^{J'} \delta^{K'} C_{I'J'K'},
\]

\[
C_{IJK} C_{I'LMM} C_{PQP} \delta^{J'} \delta^{K'} = \frac{4}{3}\delta_{II}(LCPQ),
\]

(3.4)

it was shown that the above equations are completely integrable. Defining the quantity [6]

\[
y(u) = -9a \int e^{3(B+u)} \text{du} + \frac{9}{2} \chi^2 b,
\]

(3.5)

with

\[
a = C^{IJK} V_I V_J V_K, \quad b = C^{IJK} V_I V_J \Lambda_K, \quad c = C^{IJK} V_I \Lambda_J \Lambda_K, \quad d = C^{IJK} \Lambda_I \Lambda_J \Lambda_K,
\]

(3.6)

and where the new radial coordinate \( u \) is given by \( u = \ln \chi r \), we obtain the differential equation

\[
\left( \frac{dy}{du} \right)^2 = 4y^3 - g_2 y - g_3.
\]

(3.7)

where

\[
g_2 = 243 \chi^4 (b^2 - ac), \quad g_3 = \frac{729}{2} \chi^6 (3abc - a^2 d - 2b^3).
\]

(3.8)

The general solution of Eq. (3.7) is given by \( y = \varphi (u + \gamma) \), where \( \varphi (u) \) denotes the Weierstrass elliptic function, and \( \gamma \) is an integration constant.
In the classification of solutions with null Killing vector and vanishing gauge field strengths in gauged supergravity [8], it was found that the metric and the scalar fields can be written in the following form
\[
ds^2 = H^{-1} \left( 2dU \left( dV + \frac{1}{2} F dU \right) - (dx^2)^2 - (dx^3)^2 \right) - H^2 (dx^1)^2,
\]
\[H^{-1} X_I = -2 \chi V_I x^1 + \beta_I (U), \tag{3.9}\]
with \( F \) given by
\[
H \partial_2^2 F + H^4 (\alpha_2^2 + \alpha_3^2) F - 3 \partial_1 H \partial_1 F = \frac{9}{2} H^6 G^{IJ} \partial_U \beta_I \partial_U \beta_J.
\]
Hence we see that solutions for which \( F = 0 \) must have \( \partial_I \beta_I = 0 \) and hence \( H \) and \( X^I \) are also independent of \( U \).

Changing to signature \((-, +, +, +, +)\) and concentrating on solutions with \( F = 0 \), we obtain the following domain wall solutions
\[
ds^2 = H^{-1} (-dt^2 + do^2 + dx^2 + dy^2) + H^2 dz^2, \tag{4.11}\]
\[H^{-1} X_I = -2 \chi V_I z + \beta_I. \tag{4.12}\]
Notice that these solutions are valid for all gauged \( N = 2 \) supergravity theories and in particular for those obtained from a Calabi–Yau compactification. To recover the domain wall solution of [6] described above, one can perform the following change of variable
\[
H^3 \left( \frac{dz}{du} \right)^2 = 1, \quad z = \frac{b}{2 \chi a} - \frac{y}{9 a \chi^3}. \tag{4.13}\]

4. Non-supersymmetric domain walls

In this section we consider non-supersymmetric domain wall solutions which in certain limits give the supersymmetric solutions considered in the previous section. As an ansatz for non-supersymmetric solution we take
\[
ds^2 = \frac{f}{H^5} \left( 4f H^2 - f H'' + f'' H' \right) + \frac{H^2}{f} dz^2, \tag{4.1}\]
where \( f \) and \( H \) are functions of \( z \) only. The non-vanishing the Ricci tensor are given by
\[
R_{tt} = \frac{f}{2 H^3} \left( 4f H^2 - f H'' + f'' H' \right),
\]
\[
R_{xx} = R_{yy} = R_{ww} = \frac{1}{2 H^3} \left( -4f H^2 + f H H'' + f'' H' \right),
\]
\[
R_{zz} = \frac{2 H}{H} - \frac{5 H^2}{H^2} + \frac{2 H' f'}{f} + \frac{f''}{2 f},
\]
where the prime denotes differentiation with respect to the coordinate \( z \). The Einstein equations of motion (2.5) give the following conditions:
\[
G_{IJJ} F_J^I F_J^I = \frac{1}{2 H^3} \left( f'' H^2 - 3 H' f' H \right), \tag{4.3}\]
\[
G_{IJJ} \partial_J X^I \partial_J X^J = \frac{3}{2 H^2} \left( H'' H - 2 H' \right), \tag{4.4}\]
\[
\chi^2 V = \frac{3 f}{4 H^4} \left( 4 H^2 - H H'' \right) + \frac{1}{4 H^3} \left( H f'' - 6 H' f' \right), \tag{4.5}\]
where we allowed the gauge fields to have non-vanishing field strengths \( F_J^I \). Note that the function \( f \) drops out in (4.4) and as a consequence we will not modify the scalars and we will simply use the ansatz as given in the supersymmetric case (3.12). Then it can be easily demonstrated that
\[
V_I X^I = \frac{1}{2 \chi} \frac{H'}{H^2}, \tag{4.6}\]
\[
G^{JJ} V_I V_J = \frac{1}{6 \chi^2} \left( \frac{H''}{H^3} - \frac{H'}{H^4} \right). \tag{4.7}\]

Thus the scalar potential is given by
\[
\mathcal{V} = \frac{3}{4 \chi^2 H^4} \left( 4 H^2 - H H'' \right). \tag{4.8}\]

Upon comparing (4.8) with the expression of \( \mathcal{V} \) in (4.5), the following condition is obtained
\[
6 H H' f' - 3 \left( 1 - f' \right) (H H'' - 4 H^2) - f'' H^2 = 0. \tag{4.9}\]

This can be solved by
\[
f = 1 + (\mu + \alpha z) H^3, \tag{4.10}\]
where \( \mu \) and \( \alpha \) are constants. Going back to the gauge equation of motion (2.6), this gives for our solution
\[
\partial_z \left( \frac{1}{H^2} G_{IJJ} F_J^I \right) = 0, \tag{4.11}\]
from which we obtain
\[
F_J^I = \frac{1}{H^2} G^{JJ} q_J, \tag{4.12}\]
where \( q_J \) are constants representing electric charges. Using (4.3), (4.10) and (4.12), we get
\[
G^{JJ} q_I q_J = \frac{3}{2 H^3} \left[ \alpha H H' + (\mu + \alpha z) (H H'' - H' \right). \tag{4.13}\]

Let us first consider the case with vanishing charges \( q_I = 0 \), and take \( \alpha \neq 0 \). In this case, one solution of (4.13) is given by
\[
H = \frac{c}{(\mu + \alpha z)}. \tag{4.14}\]

Then (4.4) and (3.12) imply that the scalars are constants, with \( X_I = -2 \frac{2 a}{\chi} V_I \).

If however, one takes \( q_I \neq 0 \), with \( \alpha = 0 \) in (4.13) then using (4.7) we obtain the condition
\[
G^{JJ} (q_I q_J - 9 \mu \chi^2 V_I V_J) = 0. \tag{4.15}\]

This can be solved by
\[
q_I = 3 \sqrt{\mu} \chi V_I. \tag{4.16}\]

Finally it remains to check whether the scalar equations of motion are satisfied for our solution. The scalar equations (2.7) for
our solution give

\[ H \partial_\tau \left( f H^{-3} \partial_\tau X_1 \right) - f H^{-2} \left( \frac{1}{6} C_{MNI} \right) \]

\[ - \frac{1}{2} X_I C_{MNI} X^I \partial_\tau X^M \partial_\tau X^N \]

\[ - H^{-1} \left( X_M X^P C_{NP1} - \frac{1}{6} C_{MNI} - 6X_I X_M X_N \right) \]

\[ + \frac{1}{6} X_I C_{MNI} X^I F^M_{iz} F^N_{iz} \]

\[ + 3\chi^2 \left( \frac{1}{2} V_M V_N G^{ML} G^{NP} C_{LP1} + X_I G^{MN} V_M V_N \right) \]

\[ - 2X_I X^M X^N V_M V_N \] \(=0\). \hspace{1cm} (4.17)

To simplify the calculation, we multiply the scalar equations for the supersymmetric case, i.e. multiply

\[ H \partial_\tau (H^{-3} \partial_\tau X_1) - H^{-2} \left( \frac{1}{6} C_{MNI} \right) \]

\[ - \frac{1}{2} X_I C_{MNI} X^I \partial_\tau X^M \partial_\tau X^N \]

\[ + 3\chi^2 \left( \frac{1}{2} V_M V_N G^{ML} G^{NP} C_{LP1} + X_I G^{MN} V_M V_N \right) \]

\[ - 2X_I X^M X^N V_M V_N \] \(=0\). \hspace{1cm} (4.18)

with \( f \) and subtract the resulting equation from (4.17), this gives after using the solution for the gauge fields,

\[ H \partial_\tau X_1 + \chi^2 \left( 4X^K V_I V_K - 4X^K X^L X_I V_L V_K \right) H^3 = 0. \hspace{1cm} (4.19) \]

It can be easily seen that this equation is indeed satisfied for our solution.

To summarize, we have obtained a class of domain wall solutions for all gauged five-dimensional \( N = 2 \) supergravity theories coupled to an arbitrary number of vector multiplets. These solutions are given by

\[ ds^2 = - \left( \frac{1 + \mu H^3}{H} \right) dt^2 + \frac{H^2}{1 + \mu H^3} (dz)^2 \]

\[ + 3H^2 C^{IJK} \sqrt{V_I V_J V_K}, \]

\[ X_1 = H(\beta I \xi + \beta I) \]. \hspace{1cm} (4.20)

In general, the metric is specified only implicitly by (4.20), because \( H \) is not specified explicitly by the equation of \( X_1 \). However, when the scalar manifold is symmetric, we have the relation

\[ \frac{9}{2} C^{IJK} X_I X_J X_K = 1 \] \hspace{1cm} (4.21)

from which we can explicitly solve for \( H \) and find

\[ H = \left( \alpha_0 + \alpha_1 z + \alpha_2 z^2 + \alpha_3 z^3 \right)^{-1/3}. \] \hspace{1cm} (4.22)

where

\[ \alpha_0 = \frac{9}{2} C^{IJK} \beta_I \beta_J \beta_K, \]

\[ \alpha_1 = -27 \chi C^{IJK} V_I \beta_J \beta_K, \]

\[ \alpha_2 = 54 \chi^2 C^{IJK} V_I V_J \beta_K, \]

\[ \alpha_3 = -36 \chi^3 C^{IJK} V_I V_J V_K. \] \hspace{1cm} (4.23)

For the special case of the \( STU \) model solutions, for which the intersection numbers are given by

\[ C_{IJK} = |\epsilon_{IJK}| \] \hspace{1cm} (4.24)

for \( I, J, K = 1, 2, 3 \). In this case, \( H \) factorizes as

\[ H = (\beta_0 z + \lambda_0)^{-1/3} (\beta_1 z + \lambda_1)^{-1/3} (\beta_2 z + \lambda_2)^{-1/3} \] \hspace{1cm} (4.25)

for constants \( \beta_0, \beta_1, \beta_2, \lambda_0, \lambda_1, \lambda_2 \). Note that as \( z \rightarrow -\frac{\lambda_i}{\beta_i} \), then the Ricci scalar diverges as \( (\beta_i z + \lambda_i)^{-4/3} \) if \( \mu = 0 \), and as \( (\beta_i z + \lambda_i)^{-7/3} \) if \( \mu > 0 \).

Hence we observe that both the supersymmetric and non-supersymmetric domain wall solutions contain curvature singularities. However the causal structure of the spacetimes differs considerably between the supersymmetric and non-supersymmetric cases.

5. Causal structure of domain wall spacetime

To proceed, we examine the causal structure of the spacetime geometry given in (4.20) for the solutions with symmetric scalar manifolds. Observe that geodesics on the spacetime with metric (4.20) have the following conserved quantities

\[ E = \frac{1}{H} (1 + \mu H^3) \tau, \] \hspace{1cm} (5.1)

\[ P^i = \frac{1}{H} \dot{\chi}^i \] \hspace{1cm} (5.2)

for \( i = 1, 2, 3 \), where \( \tau = t(\tau), \chi^1, \chi^2, \chi^3 = (x(\tau), y(\tau), w(\tau)), z = z(\tau), \dot{\tau} = \frac{dt}{d\tau} \) and \( \tau \) is an affine parameter. We will restrict our consideration to geodesic motion in the domain of \( z \) for which \( H > 0 \), and we take \( \mu > 0 \). It is convenient to define

\[ P^2 = (P^1)^2 + (P^2)^2 + (P^3)^2. \]

Then null geodesics satisfy

\[ \left( \frac{dz}{d\tau} \right)^2 = \frac{1}{H} (E^2 - (1 + \mu H^3) P^2) \] \hspace{1cm} (5.3)

whereas timelike geodesics satisfy

\[ \left( \frac{dz}{d\tau} \right)^2 = \frac{1}{H} E^2 - \frac{1}{H^2} (1 + \mu H^3) (H P^2 + 1). \] \hspace{1cm} (5.4)

Note that as \( z \rightarrow \infty, H \sim z^{-1} \), and hence (5.4) implies that no timelike geodesic of fixed \( E, P^1 \) can reach \( z = \pm \infty \). In addition, causal geodesics must satisfy

\[ E^2 - P^2 > \frac{E^2}{1 + \mu H^3} - P^2 \geq \frac{H}{1 + \mu H^3} \left( \frac{dz}{d\tau} \right)^2 \geq 0 \]

which implies that \( E^2 > P^2 \).
Null geodesics of the supersymmetric solution satisfy
\[
\left( \frac{dz}{d\tau} \right)^2 = \frac{1}{H} (E^2 - P^2). \tag{5.5}
\]
In the neighborhood of one of the curvature singularities, one can take \( H \sim \alpha z^{-1/3} \) as \( z \to 0 \). It follows that a null geodesic reaches the curvature singularity within finite affine parameter. Also, null geodesics can propagate out to \( z = \infty \), though they do not reach \( z = \infty \) in finite affine parameter.

Timelike geodesics of the supersymmetric solution satisfy
\[
\left( \frac{dz}{d\tau} \right)^2 = \frac{1}{H} \left( E^2 - P^2 \right) - \frac{1}{H^2}. \tag{5.6}
\]
Again, timelike geodesics reach the curvature singularity in finite proper time, but are confined to lie within \( 0 \leq z \leq z_{\text{max}}(E, P^2) \).

Null geodesics of the non-supersymmetric solution with \( P^2 \neq 0 \) cannot reach the singularity. However, null geodesics with \( P^2 = 0 \) satisfy
\[
\left( \frac{dz}{d\tau} \right)^2 = \frac{1}{H} E^2 \tag{5.7}
\]
and reach the singularity in finite affine parameter. In both cases, the null geodesics can propagate out to \( z = \infty \), though they do not reach \( z = \infty \) in finite affine parameter.

Timelike geodesics of the non-supersymmetric solution cannot reach the singularity for any choice of \( P \), and are therefore confined within region \( 0 < z_{\text{min}}(E, P^2) \leq z \leq z_{\text{max}}(E, P^2) \).

6. Conclusion

In this Letter we have constructed non-supersymmetric domain wall solutions of gauged five-dimensional \( N = 2 \) supergravity theories with non-trivial vector multiplets. The causal structure of these solutions was also discussed. These solutions constitute generalizations to a subclass of null solutions with vanishing gauge fields which were considered in [8]. In the supersymmetric limit the scalar fields remain unchanged and the gauge field strengths vanish. The scalar fields structure of these domain wall solutions resembles those for black hole solutions considered in [9] and therefore explicit domain wall solutions for the Calabi–Yau models considered in [9] can be constructed.

It will be of interest to find non-supersymmetric generalizations to the solutions of [7,8] and in particular to the supersymmetric null solutions with non-trivial gauge fields of [8]. We hope to report on this in a future publication.

Acknowledgement

The work of W.S. was supported in part by the National Science Foundation under grant number PHY-0313416.

References

Exact 1/4 BPS loop—Chiral primary correlator

Gordon W. Semenoff¹, Donovan Young *, 2

Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, British Columbia V6T 1Z1, Canada

Received 22 September 2006; accepted 25 October 2006
Available online 3 November 2006
Editor: L. Alvarez-Gaumé

Abstract
Correlation functions of 1/4 BPS Wilson loops with the infinite family of 1/2 BPS chiral primary operators are computed in \( \mathcal{N} = 4 \) super-Yang–Mills theory by summing planar ladder diagrams. Leading loop corrections to the sum are shown to vanish. The correlation functions are also computed in the strong-coupling limit by examining the supergravity dual of the loop–loop correlator. The strong coupling result is found to agree with the extrapolation of the planar ladders. The result is related to known correlators of 1/2 BPS Wilson loops and 1/2 BPS chiral primaries by a simple re-scaling of the coupling constant, similar to an observation of Drukker [N. Drukker, hep-th/0605151] for the case of the 1/4 BPS loop vacuum expectation value.

© 2006 Elsevier B.V. All rights reserved.

Recently, the study of the properties of highly symmetric states has provided considerable insight into the AdS/CFT correspondence. In the case of 1/2 BPS local chiral operators and 1/2 BPS Wilson loops of \( \mathcal{N} = 4 \) supersymmetric Yang–Mills theory, their correspondence with 1/2 BPS gravitons and fundamental string world-sheets has been generalized to large operators where a beautiful picture of giant gravitons [1–3], giant Wilson loops [4–15] and bubbling geometries [16] has emerged. These relate infinite classes of highly symmetric protected operators in Yang–Mills theory to their dual geometries which solve IIB supergravity.

In the case of 1/2 BPS Wilson loops, an essential component of the bubbling loop picture is the ability to compute the loop expectation value and correlators of the loop with chiral primary operators in Yang–Mills theory by summing planar diagrams [11,17–21]. To point, for example, it is this sum, in the form of a matrix model computation, which provides evidence that the giant loops are dual to D3- and D5-branes. The matrix model is thought to coincide with the sum of all Feynman diagrams. This depends on cancellation of loop corrections, which has been demonstrated in leading orders, but has not yet been proven. 3 It apparently holds for the expectation value of the 1/2 BPS Wilson loop and the correlator of the 1/2 BPS Wilson loop with any 1/2 BPS chiral primary operator. In all of these cases, when extrapolated to strong coupling, the sum of planar ladder Feynman diagrams agrees with the supergravity computation using AdS/CFT. This gives an infinite tower of functions which interpolate between weak and strong coupling. In this Letter, we will examine a modest extension of the picture. We will demonstrate similar results for the expectation value and the correlation functions of a 1/4 BPS Wilson loop with 1/2 BPS chiral operators.

The vacuum expectation value of the 1/4 BPS loop was studied by Drukker in Ref. [23]. He observed a number of interesting features of the gauge theory computation. One was that the ladder diagrams had a structure similar to the 1/2 BPS circle loop and they could be summed to obtain an expression very similar to the case of the 1/2 BPS loop. The difference was the replacement of the ’t Hooft coupling \( \lambda \) by \( \lambda \cos^2 \theta_0 \) where \( \theta_0 \) is a parameter of the 1/4 BPS loop. He further showed that, as occurred for the 1/2 BPS loop, the leading corrections from diagrams with internal vertices (those diagrams which are left out of the sum over

* Corresponding author.
E-mail address: dyoung@physics.ubc.ca (D. Young).
1 Work supported in part by the NSERC of Canada.
2 Work supported by NSERC of Canada and a University of British Columbia Graduate Fellowship.
3 There could also be non-perturbative contributions, which are plausibly suppressed in the large \( \mathcal{N} \) limit [22].

0370-2693/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.physletb.2006.10.047
ladders) cancel. He observed that, in the string dual where, following the prescription given in Ref. [24], the expectation value of
the loop is found as the area of an extremal world-sheets bounding the loop, there are two saddle point solutions. He showed that
the strong coupling extrapolation of the sum of diagrams on the gauge theory side carried a vestige of these two saddle points with
some of the expected features of a saddle-point expansion.

In the following, we will study correlators of 1/4 BPS Wilson loops with 1/2 BPS chiral primary operators. We find that these
correlators depend on the $SO(6)$-orientation of the chiral primary. We identify all of the orientations where the Wilson loop and the
chiral primary share some degree of supersymmetry. We find that the ladder diagrams can be summed for correlators of the loop and
these operators and the result is identical to those previously found with the 1/2 BPS Wilson loop [11,20] with a certain rescaling of
the coupling constant. We shall also study the strong coupling limit of the same correlators using the AdS/CFT correspondence.
We identify the supergravity dual of the loop–loop correlation function and compute it in the asymptotic limit that is appropriate to
extracting the contribution of intermediate chiral primary operators. This yields the limit of large $N$ and large 't Hooft coupling $\lambda$.
We find that the results agree with the extrapolation to strong coupling of the Yang–Mills computation.

The Wilson loop operator of $\mathcal{N} = 4$ supersymmetric Yang–Mills theory which is most relevant to the AdS/CFT correspondence is [24]

$$ W[C] = \frac{1}{N} \text{Tr} P \exp \left[ \int_C \left( i A_\mu (x) \dot{x}^\mu (\tau) + | \dot{x}(\tau) | \Theta^I (\tau) \Phi_I (x(\tau)) \right) d\tau \right], $$

where $A_\mu (x)$ are the gauge fields and $\Phi_I (x)$, $I = 1, \ldots, 6$ are the scalar fields of $\mathcal{N} = 4$ supersymmetric Yang–Mills theory. The
curve $C$ is described by $\dot{x}^\mu (\tau)$ and $\Theta^I (\tau)$, with $\sum_I \Theta^I \Theta^I = 1$, describes a loop on the 5-sphere. This loop operator is related to
the holonomy of heavy $W$-bosons in the gauge theory with $SU(N + 1) \to SU(N) \times U(1)$ symmetry breaking. Its string theory
dual is a source for a fundamental open string whose world-sheet ends on the contour $C$ at the boundary of $AdS_5 \times S^5$.

When probed from a distance much larger than the extension of $C$, the Wilson loop operator should look like an assembly of local operators,$^4$

$$ W[C] = \langle 0 | W[C] | 0 \rangle \left( 1 + \sum_{\Delta_i > 0} O_{\Delta_i} (0) L[C]^{\Delta_i} \xi_{\Delta_i} [C] \right), $$

where $L[C] = \int_C | \dot{x}(\tau) | d\tau$ is the length of $C$ and we have assumed that $C$ is near the origin $0$. The operator expansion coefficients
generally depend on the shape and orientation of $C$, as well as the parameters of $\mathcal{N} = 4$ Yang–Mills theory, the coupling constant $g_\text{YM}$ and the number of colors $N$. In the remainder of this Letter, we will consider only the planar 't Hooft large $N$ limit of Yang–
Mills theory where $N \to \infty$ holding $\lambda \equiv g_\text{YM}^2 N$ fixed. In that limit, we can see from (4) below that $\xi_\Delta$ is the ratio of a disc to a
cylinder amplitude and therefore should be of order $\frac{1}{N}$ times a function of $\lambda$.

All operators which can be made from the gauge fields, scalars and their derivatives can appear in the expansion in Eq. (2). We have
classified operators according to their conformal dimensions, $\Delta_\gamma$. In a conformal field theory, the operators of fixed conformal
dimensions can be organized into families which contain a primary operator with smallest $\Delta$ and an infinite tower of descendants.
We will assume that primary operators are normalized so that

$$ \langle 0 | O_\Delta (x) O_{\Delta'} (0) | 0 \rangle = \frac{\delta_{\Delta \Delta'}}{(4\pi^2)^2 x^2)^\Delta}. $$

The operator expansion coefficient $\xi_\Delta$ for a primary operator can be extracted from the asymptotics of the correlator

$$ \frac{\langle 0 | W[C] O_\Delta (x) | 0 \rangle}{\langle 0 | W[C] | 0 \rangle} = \frac{L[C]^\Delta}{(4\pi^2)^\Delta x^2 \xi_\Delta + \cdots}. $$

For example, for the 1/2 BPS circle Wilson loop,

$$ C_{1/2}: \quad x^\mu (\tau) = (R \cos \tau, R \sin \tau, 0, 0), \quad \Theta^I = (1, 0, \ldots) $$

a perturbative expansion of the loop gives

$$ W[C_{1/2}] = \langle 0 | W[C_{1/2}] | 0 \rangle \left( \sum_{k=0}^\infty \frac{(2\pi R)^k}{N k! \lambda^k} \frac{1}{2^k} \text{Tr} (Z(0) + \bar{Z}(0))^k + \cdots \right), $$

$^4$ It is also possible to consider the insertion of supersymmetric operators into the Wilson loop itself. We emphasize that is a different procedure from what we are
discussing here, where correlations of primary operators with the Wilson loop are the objects of most interest. Also, chiral operators of the type that we consider
figure promptly in the discussion of the BMN limit as well as some issues of integrability [25–29].
where $Z = (\Phi_1 + i \Phi_2)$ and the dots indicate quantum corrections as well as operators with derivatives of $Z$, $\hat{Z}$ and containing gauge fields. For the chiral primary operators

$$O_J \equiv \frac{1}{\sqrt{J}} \langle \text{Tr} Z(0) \rangle^J,$$

the weak coupling limit of $\xi_J[C_{1/2}, \lambda]$ is the appropriate coefficient in Eq. (6),

$$\xi_J[C_{1/2}; \lambda \sim 0] = \frac{1}{N} \frac{1}{2^J J!} \sqrt{J \lambda^J}.$$

This expression should receive quantum corrections. The sum of all quantum corrections from planar ladder diagrams was computed in Ref. [20]

$$\xi_J[C_{1/2}; \lambda] = \frac{1}{N} \frac{1}{2^J} \sqrt{\lambda^J} I_J(\sqrt{\lambda}).$$

(9)

where $I_J(x)$ is the $J$th modified Bessel function of the first kind. In the expression (9), as it must, the leading term in a small $\lambda$ expansion agrees with (8). The leading order planar diagrams which are left out of the sum over ladders was also computed in Ref. [20] and were shown to cancel identically. It was then tempting to conjecture that these corrections vanish to all orders.

To support this conjecture, the extrapolation of Eq. (9) to large $\lambda$ can be compared with the result of a computation of the same coefficients using the AdS/CFT correspondence, originally done in Ref. [30],

$$\xi_J[C_{1/2}; \lambda \sim \infty] = \frac{1}{N} \frac{1}{2^J} \sqrt{\lambda^J}.$$

(10)

This coincides with the large $\lambda$ limit of the expression in Eq. (9). The coefficients $\xi_J[C_{1/2}, \lambda]$ in (9), together with the result of Ref. [17]

$$\{W[C_{1/2}] = \frac{2}{\sqrt{\lambda}} I_1(\sqrt{\lambda}).$$

(11)

yield an infinite family of interpolating functions which match both the strong and weak coupling limits computed in string and gauge theory, respectively.

In the present Letter, we will examine the 1/4 BPS loop which has the trajectory

$$C_{1/4}: \quad x^\mu(\tau) = R(\cos \tau, \sin \tau, 0, 0), \quad \Theta^I(\tau) = (\sin \theta_0 \cos \tau, \sin \theta_0 \sin \tau, \cos \theta_0, 0, 0, 0).$$

(12)

The main difference from the 1/2 BPS loop is that $\Theta^I(\tau)$ moves in a circle on an $S^2 \subset S^5$, rather than sitting at a point. Putting $\theta_0$ to zero recovers the 1/2 BPS loop in (5). The special case of this 1/4 BPS loop with $\theta_0 = \pi/2$ was originally discussed by Zarembo [31].

To understand the supersymmetries of the loop with trajectory (12) we recall that the supersymmetry transformation of $\mathcal{N} = 4$ Yang–Mills theory is generated by the spinor

$$\epsilon(x) = \epsilon_0 + \gamma_\mu x^\mu \epsilon_1.$$

(13)

Here, we have to consider both Poincaré supersymmetries, with constant spinor $\epsilon_0$ and conformal supersymmetries, with constant spinor $\epsilon_1$. In order to be supersymmetries of the 1/4 BPS Wilson loop, it is straightforward to see that they have to satisfy the equations [23]

$$\sin \theta_0 (\gamma^1 \Gamma^2 + \gamma^2 \Gamma^1) \epsilon_0 = 0, \quad \sin \theta_0 (\gamma^1 \Gamma^2 + \gamma^2 \Gamma^1) \epsilon_1 = 0,$$

$$\cos \theta_0 \epsilon_0 = R(-i \gamma^1 + \sin \theta_0 \Gamma^5) \Gamma^3 \gamma^2 \epsilon_1.$$

(14)

where the ten-dimensional gamma matrices are $(\gamma^i, \Gamma^I)$ with $i = 1, \ldots, 4$ and $I = 1, \ldots, 6$. Let us count the supersymmetries. Each of the spinors $\epsilon_0$ and $\epsilon_1$ has 16 components. The conditions in (14) are half-rank and reduce the number of each of the spinors by half. Then (15) relates the remaining components of $\epsilon_1$ to those of $\epsilon_0$ in a way which is compatible with (14). The remaining independent components are eight—half of the original 16 components of $\epsilon_0$. This is 1/4 of the original 32 components of $\epsilon_0$ and $\epsilon_1$.

We will consider a chiral operator which has an arbitrary $SO(6)$ orientation, beginning with

$$\text{Tr}(u \cdot \Phi(0))^J,$$

where $u$ is a complex 6-vector, satisfying the constraint that $u^2 = 0$. Being a scalar operator, conformal supersymmetries are automatic. This operator has some Poincaré supersymmetry if there exist some non-zero constant spinors $\epsilon_0$ which solve the
equation
\[ u \cdot \Gamma \epsilon_0 = 0. \] (16)

There are solutions only when \((u \cdot \Gamma)^2 = u^2 = 0\) which, as we have assumed, is the case. Then \(u \cdot \Gamma\) is half-rank and there are exactly eight independent non-zero solutions of Eq. (16).

Now we can ask the question as to whether the eight independent \(\epsilon_0\) which solve (16) have anything in common with the eight solutions of (14) and (15), i.e. are there spinors which solve both of them?

Before we answer this question, let us backtrack to the case of the 1/2 BPS loop geometry (9). There Eq. (14) is absent and the spinors must solve (15) with \(\theta_0 = 0\). This simply relates \(\epsilon_1\) to \(\epsilon_0\), eliminating half of the possible spinors. There are 16 independent solutions of this equation—it is 1/2 BPS. Now, consider a chiral primary operator. Without loss of generality, we can consider the operator \(\text{Tr}(\Phi_1 + i\Phi_2)^J\). It is supersymmetric if \(\epsilon_0\) satisfies the equation
\[ (\Gamma^1 + i\Gamma^2)\epsilon_0 = 0. \]

The matrix \(\Gamma_1 + i\Gamma_2\) has half-rank, so this requirement eliminates half of the supersymmetries generated by \(\epsilon_0\). This leaves eight supersymmetries which commute with both the 1/2 BPS Wilson loop and the 1/2 BPS chiral primary operator. This high degree of residual joint supersymmetry is thought to be responsible for the fact that, apparently, only ladder diagrams contribute to the asymptotic limit of their correlator.

Returning to the 1/4 BPS loop and chiral primary with general orientation, it is easy to see that there is a simultaneous solution of (14)–(16) only when one of the following holds:

- \(u_1 = u_2 = 0\). We can always do an \(SO(6)\) rotation which commutes with the loop operator and sets \((u_4, u_5, u_6) \rightarrow (u_4, 0, 0)\). Then, there will be simultaneous solutions of (14)–(16) only when \(u_3 = iu_4\) or when \(u_3 = -iu_4\). In both of these cases, there are four solutions, corresponding to 1/8 supersymmetry in common between the chiral primary and the Wilson loop. Up to a constant, the chiral primary operator is \(\text{Tr}(\Phi_3 + i\Phi_4)^J\) or the complex conjugate \(\text{Tr}(\Phi_3 - i\Phi_4)^J\).

- \(u_3 = u_4 = 0\). There is a solution when \(u_1 = \pm iu_2\) and there is also 1/8 supersymmetry. The chiral primary is \(\text{Tr}(\Phi_1 + i\Phi_2)^J\) or its complex conjugate. In this case, we show in Appendix C that the coefficient \(\xi_J\) which is extracted from the long range part of the correlator of this operator and the loop vanishes due to R-symmetry. Thus, for all \(J > 0\), the coefficients of \(\text{Tr}(\Phi_1 + i\Phi_2)^J\) or \(\text{Tr}(\Phi_1 - i\Phi_2)^J\) in the operator expansion of the 1/4 BPS loop are zero.

- \(u_1 = \pm iu_2\). There are two non-zero solutions when \(u_3 = iu_4\) or when \(u_3 = -iu_4\). This corresponds to 1/16 supersymmetry. There are essentially four operators,
\[ \text{Tr}(\chi(\Phi_1 + i\Phi_2) + (\Phi_3 + i\Phi_4))^J \]
plus others with substitutions of \(\Phi_1 - i\Phi_2\) or \(\Phi_3 - i\Phi_4\). In this case too, because of R-symmetry the contribution with any non-zero power of \((\Phi_1 \pm i\Phi_2)\) will be zero. The coefficient \(\xi_J[C_{1/4}]\) for these operators is therefore the same as those for the operator \(\text{Tr}(\Phi_3 \pm i\Phi_4)^J\).

Thus we see that the interesting quantity where there is some degree of supersymmetry common to both the loop operator and the primary is
\[ \xi_J[C_{1/4}] = \lim_{|x| \rightarrow \infty} \left( \frac{4\pi^2|x|^2}{2\pi R} \right)^J \frac{1}{\sqrt{J!} \lambda^J} \frac{\langle 0|W[C_{1/4}]\text{Tr}(\Phi_3(x) + i\Phi_4(x))^J|0\rangle}{\langle 0|W[C_{1/4}]|0\rangle}. \] (17)

It is these partially supersymmetric configurations which we expect to have some level of protection from quantum corrections. Indeed, we shall find evidence for this. All other possibilities either vanish, are equivalent to (17) or have no supersymmetry at all. The cases with no supersymmetry at all are apparently not protected.

We will present arguments that the sum of planar ladder diagrams contributing to the correlation function in (17) gives a contribution which differs from the one for the 1/2 BPS loop quoted in Eq. (9) by the simple replacement \(\lambda \rightarrow \lambda \cos^2 \theta_0\), so that the total result is
\[ \xi_J[C_{1/4}] = \frac{1}{N} \frac{1}{2} \sqrt{\lambda \cos^2 \theta_0} \frac{J}{I_J(\sqrt{\lambda \cos^2 \theta_0})} \frac{I_J(\sqrt{\lambda \cos^2 \theta_0})}{I_1(\sqrt{\lambda \cos^2 \theta_0})}. \] (18)

To find this result using Feynman diagrams, we begin with the lowest order diagrams, depicted in Fig. 1. There, each occurrence of the scalar \(\Phi_3\) in the composite operator contracts with a scalar \(\Phi_3\) in the Wilson loop. We consider only the planar diagrams. Each scalar \(\Phi_3\) from the Wilson loop carries a factor of \(\cos \theta_0\), leading to an overall factor of \((\cos \theta_0)^J\). We are taking the convention for Feynman rules where each line in the Feynman diagram results in a factor of \(\lambda^J\) for the diagram in Fig. 1. With this convention, the chiral primary operator has normalization \(\lambda^{-1/2}\) (see (7)). The net result is a factor of \(\lambda^{J/2}\) which combines with
The leading planar contribution to $\langle W [C_{1/4}] \text{Tr}(\Phi_3 + i\Phi_4)^J \rangle$. There are $J$ lines connecting the chiral primary on the left with the circular Wilson loop on the right.

The “rungs” represent the combined gauge field and scalar propagator. For clarity, $J$ has been set to 2.

The one-loop radiative corrections to $\langle W [C_{1/4}] \text{Tr}(\Phi_3 + i\Phi_4)^J \rangle$. Only an adjacent pair of the $J$ scalar lines are shown.
modes, which at large \( \lambda \) are 1/2 BPS supergravitons, the string theory duals of the chiral primary operators. The connection between the graviton propagator and the worldsheet is through a vertex operator which must be identified and the connection point with the vertex operator must be integrated over the worldsheet. The resulting amplitude is proportional to the square of the desired operator expansion coefficient.

To begin, the first step is to identify the minimal surface in \( AdS_5 \times S^5 \) whose boundary is the 1/4 BPS contour \( C_{1/4} \). This was done in Ref. [23]. We will summarize it here in more convenient coordinates. We take the metric of \( AdS_5 \times S^5 \)

\[
ds^2 = \sqrt{\lambda}\left( \frac{dy^2 + dr_1^2 + r_1^2 d\phi_1^2 + dr_2^2 + r_2^2 d\phi_2^2}{y^2} + d\theta^2 + \sin^2 \theta \, d\phi^2 + \cos^2 \theta (dp^2 + \sin^2 \rho \, d\rho^2 + \cos^2 \rho \, d\phi^2) \right).
\]

(19)

The string world-sheet is then embedded as follows,

\[
y = R \tanh \sigma, \quad r_1 = \frac{R}{\cosh \sigma}, \quad \phi_1 = \tau, \quad r_2 = 0, \quad \phi_2 = \text{const},
\]

\[
\sin \theta = \frac{1}{\cosh(\sigma_0 \pm \sigma)}, \quad \phi = \tau, \quad \rho = \frac{\pi}{2}, \quad \hat{\phi} = 0, \quad \hat{\phi} = \text{const},
\]

(20)

where \( \sigma \in [0, \infty) \) and \( \tau \in [0, 2\pi] \) are the world-sheet coordinates. The contour \( C_{1/4} \) is the boundary of the worldsheet at \( \sigma = 0 \), which in turn sits at \( y = 0 \), the boundary of \( AdS_5 \times S^5 \). The parameter \( \cos \theta_0 = \frac{1}{\cosh \sigma_0} \). The choice of \( \pm \) sign in the embedding of \( \theta \) arises because there are two saddle points in the classical action corresponding to wrapping the north or south pole of the \( S^5 \). Of course the sign should be chosen to minimize the classical action, which corresponds to choosing +. The other saddle point is unstable, and the string world-sheet will slip-off the unstable pole.

The supergravity modes that we are interested in are fluctuations of the RR 5-form as well as the spacetime metric. They are by now very well known, and details can be found in Refs. [15,30,32–34]. The fluctuations are

\[
\delta g_{ab} = \left[ -\frac{6J}{5} g_{ab} + \frac{4}{J + 1} D_{(a} D_{b)} \right] s^J(X) Y_J(\Omega), \quad \delta g_{IK} = 2 k g_{IK} s^J(X) Y_J(\Omega),
\]

(21)

where \( \alpha, \beta \) are \( AdS_5 \) and \( I, K \) are \( S^5 \) indices. The symbol \( X \) indicates coordinates on \( AdS_5 \) and \( \Omega \) coordinates on the \( S^5 \). The \( D_{(a} D_{b)} \) represents the traceless symmetric double covariant derivative. The \( Y_J(\Omega) \) are the spherical harmonics on the five-sphere, while \( s^J(X) \) have arbitrary profile and represent a scalar field propagating on \( AdS_5 \) space with mass squared \( = J(J - 4) \), where \( J \) labels the representation of \( SO(6) \) and must be an integer greater than or equal to 2. (This is the representation of \( SO(6) \) which contains the chiral primary operators that we are interested in.)

The supergravity field dual to the operator \( \text{Tr}(u \cdot \Phi)^J \) is obtained by choosing the combination of spherical harmonics with the same quantum numbers and evaluating them on the worldsheet using (20) (see Appendix B) so that,

\[
Y_J(\theta, \phi) = \mathcal{N}_J(u) [u_1 \sin \theta \cos \phi + u_2 \sin \theta \sin \phi + u_3 \cos \theta]^J.
\]

(22)

The worldsheets will be connected by the propagator for the scalar supergravity mode \( s^J(X) \). The asymptotic form of this propagator for large separation \( x \) is

\[
P(X, \tilde{X}) = \langle s^J(X) s^J(\tilde{X}) \rangle \approx A_J \left( \frac{1}{x} \right)^{2J} y^J \tilde{y}^J,
\]

(23)

where \( A_J = 2^J (J + 1)^2 / (16N^2 J) \). The barred quantities are coordinates on the second Wilson loop worldsheet. Then, in the large \( \lambda \) limit, the Wilson loop correlator is

\[
\frac{\langle 0|W[C_{1/4}, x]W^*[C_{1/4}, y]|0\rangle}{|\langle 0|W[C_{1/4}]|0\rangle|^2} = \int \int \bar{\partial}_a X^M \partial^a X^N \delta g_{MN} P(X, \tilde{X}) \bar{\delta} \tilde{g}_{M\tilde{N}} \partial_a X^M \partial^a \tilde{X}^\tilde{N},
\]

(24)
where \( M, N = 1, \ldots, 10 \) and the \( \delta_{g_{MN}} \) are given in (21), except now we have removed the fluctuating parts, \( s^f(X) \) and replaced them by the propagator \( P \). The pullback of the fluctuations (21) to the worldsheet are found in Appendix A. Using them we have,

\[
\frac{\langle 0 \vert W[C_{1/4}, x] W' [C_{1/4}, 0] \vert 0 \rangle}{\langle 0 \vert W[C_{1/4}] \vert 0 \rangle^2} = \frac{\Lambda_J}{\sqrt{2\pi}} \frac{\lambda}{16\pi^2} \left[ 2I \int d\sigma d\tau y^J y^{J-2} Y_J(\theta, \phi) - 2I \int d\sigma d\tau (r_1^2 + r_2^2) y^J y^{J-2} Y_J(\theta, \phi) \right. \\
\left. + 2I \int d\sigma d\tau (\theta^2 + \sin^2\theta) y^J Y_J(\theta, \phi) \right]^2. \tag{25}
\]

Each of the terms inside the square on the right-hand side of the above expression has a common factor of

\[
\int_0^{2\pi} d\tau Y_J(\theta, \phi) = N_J(u) \int_0^{2\pi} d\tau [u^1 \sin\theta \cos\tau + u^2 \sin\theta \sin\tau + u^3 \cos\theta]^J. \tag{26}
\]

From this expression we see that, consistent with our expectations using R-symmetry on the gauge theory side, for the at least 1/16 supersymmetric combination of loop and primary when \( u_2 = \pm iu_1 \), the dependence on \( u_1 \) and \( u_2 \) integrates to zero. If these parameters are chosen more arbitrarily, so that there is no supersymmetry at all, the loop depends on them. In that case the contributions proportional to powers of \( u_1 \) and \( u_2 \) in the final result for the operator expansion coefficients do not follow the rule that they are related to the 1/2 BPS loop ones by the replacement of \( \lambda \) by \( \lambda \cos^2 \theta_0 \). We attribute this to absence of supersymmetry. From here, we will proceed with the supersymmetric case only by putting \( u_1 = u_2 = 0 \) and \( u_3 = 1 \).

We will now compute the integrals in (25) with this assumption. We note that the embedding (20) has some nice properties. For instance \( y^J y^{J-2} = y^J \) and also \( \sin^2\theta = \theta^2 \). Using these, we can express the integrals in (25) as follows,

\[
\frac{2^{-J/2}}{R^J} \int d\sigma y^J y^{J-2} \cos^J \theta = 2^{-J/2} \int_0^\infty d\sigma \left( \frac{\tanh \sigma}{\cosh^J \sigma} \right)^J (\tan^J \sigma \pm \sigma) = 2^{-J/2} \int_0^\infty dz (1 - z^2)^{J/2} \left( \frac{1 \mp \cos \theta_0}{1 \pm \cos \theta_0} \right)^J, \tag{27}
\]

\[
\frac{2^{-J/2}}{R^J} \int d\sigma (r_1^2 + r_2^2) y^J y^{J-2} \cos^J \theta = 2^{-J/2} \int_0^1 dz (1 + z^2)^{J/2} \left( \frac{1 \mp \cos \theta_0}{1 \pm \cos \theta_0} \right)^J, \tag{28}
\]

\[
\frac{2^{-J/2}}{R^J} \int d\sigma (\theta^2 + \sin^2\theta) y^J y^{J-2} \cos^J \theta = -2^{1-J/2} \int_{\mp \cos \theta_0}^1 dz \left( \frac{1 \mp \cos \theta_0}{1 \pm \cos \theta_0} \right)^J. \tag{29}
\]

Putting everything together,

\[
\frac{\langle 0 \vert W[C_{1/4}, x] W' [C_{1/4}, 0] \vert 0 \rangle}{\langle 0 \vert W[C_{1/4}] \vert 0 \rangle^2} = 16J^2 \frac{\Lambda_J}{2\sqrt{\pi}} \left( \frac{R^J}{\lambda} \right)^{2J/4} \left\{ \frac{1}{J+1} \int_0^1 dz \int_{\mp \cos \theta_0}^1 dz \left( \frac{1 \mp \cos \theta_0}{1 \pm \cos \theta_0} \right)^J \right\}^2 \tag{30}
\]

which is just the result for the 1/2 BPS circle [30] with \( \lambda \rightarrow \lambda \cos^2 \theta_0 \). Using the prescription [30] to obtain from the loop-to-loop correlator the overlap with the chiral primary in question, we find \( \xi_{J[C_{1/4}]} = \sqrt{J/\lambda \cos^2 \theta_0} / 2N \). This is identical to the large \( \lambda \) limit of Eq. (18). We have thus confirmed that the sum of planar ladder diagrams agrees with the prediction of AdS/CFT in the strong coupling limit. The emergence of this structure on the supergravity side of the duality is non-trivial. The integrations over the AdS\(_5\) and S\(_5\) portions of the string worldsheet conspire in a complicated way in (30) to give the \( \lambda \rightarrow \cos^2 \theta_0 \lambda \) result.

It is instructive to consider this calculation where both saddle points of the classical action are kept in the path integral, as is discussed in [23]. There it was noted that the semi-classical result for the expectation value of the Wilson loop is a sum of two terms; one proportional to \( \exp(\sqrt{\lambda}^2) \) and the other to \( \exp(-\sqrt{\lambda}^2) \), where \( \lambda' = \cos^2 \theta_0 \lambda \). This was mirrored in the asymptotic expansion [35] of the modified Bessel function of (11),

\[
I_1(\sqrt{\lambda}) = \frac{e^{\sqrt{\lambda}}}{\sqrt{2\pi \sqrt{\lambda}}} \sum_{k=0}^{\infty} \frac{(-1)^k}{2\sqrt{\lambda}^k} \Gamma(3/2 + k) / k! \Gamma(3/2 - k) \pm i e^{-\sqrt{\lambda}} \sqrt{2\pi \sqrt{\lambda}} \sum_{k=0}^{\infty} \frac{1}{2\sqrt{\lambda}^k} \Gamma(3/2 + k) / k! \Gamma(3/2 - k), \tag{31}
\]

where the sign of the \( i \) is ambiguous due to the Stokes’ Phenomenon [36]. The factor of \( i \) was associated with the fluctuation determinant of the three tachyonic modes associated with the worldsheet slipping off the unstable pole of the five-sphere.

Due to the sign structure found in (30) before squaring, the analogous structure for the connected correlator of the primary with the loop is a sum of a term proportional to \( \exp(\sqrt{\lambda}^2) \) and of another proportional to \( (-1)^{J+1} \exp(-\sqrt{\lambda}^2) \). The sum of these
two terms should then be normalized by the expectation value of the Wilson loop. If we employ the asymptotic expansions of the modified Bessel functions in (9), we have

\[
\frac{I_J(\sqrt{\lambda^2})}{I_1(\sqrt{\lambda^2})} = \frac{e^{\sqrt{\lambda^2}} \sum_{k=0}^{\infty} \left( \frac{-1}{2 \sqrt{\lambda^2}} \right)^k \Gamma(J+k+1/2) \Gamma(J+k+1/2) \pm i (-1)^J e^{-\sqrt{\lambda^2}} \sum_{k=0}^{\infty} \left( \frac{1}{2 \sqrt{\lambda^2}} \right)^k \Gamma(J+k+1/2) \Gamma(J+k+1/2)}{e^{\sqrt{\lambda^2}} \sum_{k=0}^{\infty} \left( \frac{-1}{2 \sqrt{\lambda^2}} \right)^k \Gamma(J+3k/2-1) \pm i e^{-\sqrt{\lambda^2}} \sum_{k=0}^{\infty} \left( \frac{1}{2 \sqrt{\lambda^2}} \right)^k \Gamma(J+3k/2-1)}.
\]

This clearly reflects the presence of two saddle points in the functional integrals in both the numerator and denominator.

We also note that the chiral primary has zero overlap with the supersymmetric Wilson loop (i.e. \( W_{\theta} \)). This is expected, since two such Wilson loops should not interact with each other by supersymmetry.

There has been extensive work of late concerning Wilson loops whose \( SU(N) \) representations are of higher rank \([4–8,10–13]\). They have been associated with D-brane solutions analogous to giant gravitons. Explicit solutions are available for the 1/2 BPS loop, and results have been matched to matrix model calculations. It would be very interesting to solve the DBI equations of motion corresponding to the 1/4 BPS loop, and to repeat the calculations done here for that solution, as has been recently done for the 1/2 BPS case \([15]\).

### Appendix A. Metric fluctuations

Given (21) and (19), we must construct the traceless symmetric double covariant derivative,

\[
D_{(\mu} D_{\nu)} = \frac{1}{2} (D_{\mu} D_{\nu} + D_{\nu} D_{\mu}) - \frac{1}{5} g_{\mu\nu} g^{\rho\sigma} D_{\rho\sigma}.
\]

The action of \( D_{\mu} D_{\nu} \) on a scalar field \( \phi \) is,

\[
D_{\mu} D_{\nu} \phi = \partial_{\mu} \partial_{\nu} \phi - \Gamma_{\mu\nu}^{\lambda} \partial_{\lambda} \phi.
\]

The Christoffel symbols for the AdS geometry (19) are,

\[
\Gamma_{\phi_i \phi_i} = -r_i, \quad \Gamma_{\phi_y \phi_i} = r_i^2 y, \quad \Gamma_{\phi_i \phi_y} = \frac{1}{r_i}, \quad \Gamma_{\phi_i \phi_y} = -\frac{1}{y}, \quad \Gamma_{\phi_i \phi_y} = \frac{1}{y}, \quad \Gamma_{r_i r_i} = -\frac{1}{y}, \quad \Gamma_{r_y r_y} = -\frac{1}{y},
\]

where \( i = 1, 2 \). The trace of \( D_{\mu} D_{\nu} \phi \) is given by,

\[
g^{\mu\nu} D_{\mu} D_{\nu} \phi = \sum_{i=1}^{2} \left( y^2 \partial_y^2 + y^2 \partial_i^2 + \frac{y^2}{r_i^2} \partial_{\phi_i}^2 - 3y \partial_y + \frac{y^2}{r_i} \partial_{r_i} \right) \phi.
\]

Because of (23), we only keep those terms of \( D_{(\mu} D_{\nu)} \) which contain derivatives in \( y \). These are,

\[
D_{(x} D_{y)} = \frac{4}{5} \partial_y^2 + \frac{8}{5} \partial_y, \quad D_{(r_i} D_{r_i)} = \frac{1}{r_i^2} D_{(\phi_i} D_{\phi_i)} = \frac{1}{r_i^2} \partial_{\phi_i}^2 - \frac{2}{5} \partial_{y}.
\]

We now note that since the derivatives will be acting on \( y^2 \) from the propagator, we may replace \( \partial_y^2 \to J (J-1)/y^2 \) and \( y^{-1} \partial_y \to J/y^2 \). Therefore the metric fluctuations may be expressed as follows,

\[
\delta g_{yy} = \left[-\frac{6J}{5} + \frac{4}{J+1} \left( \frac{4}{5} J (J-1) + \frac{8}{5} J \right) \right] \frac{L^2}{y^2},
\]

\[
\delta g_{r_ir_i} = \frac{1}{r_i^2} \delta g_{\phi_i \phi_i} = \left[-\frac{6J}{5} - \frac{4}{J+1} \left( \frac{1}{5} J (J-1) + \frac{2}{5} J \right) \right] \frac{L^2}{y^2}.
\]

### Appendix B. Spherical harmonics

The five-sphere is embedded in \( \mathbb{R}^6 \) in the following manner,

\[
x^1 = \sin \theta \cos \phi, \quad x^2 = \sin \theta \sin \phi, \quad x^3 = \cos \theta \sin \rho \cos \phi, \quad x^4 = \cos \theta \sin \rho \sin \phi, \quad x^5 = \cos \theta \cos \rho \cos \phi, \quad x^6 = \cos \theta \cos \rho \sin \phi,
\]

and has the metric

\[
ds^2_5 = d\theta^2 + \sin^2 \theta d\phi^2 + \cos^2 \theta (d\rho^2 + \sin^2 \rho d\tilde{\phi}^2 + \cos^2 \rho d\tilde{\phi}^2).
\]
The embedding (20) takes \( \rho = \pi/2, \hat{\phi} = 0, \) or \( x^4 = x^5 = x^6 = 0. \) Note that \( \rho \in [0, \pi/2] \) while \( \theta \in [0, \pi]. \) A general chiral primary normalized as in (3) may be written as,

\[
\frac{2^{j/2}}{\sqrt{J!}} C^{1 \ldots J} \, \text{Tr} \, \Phi_{1} \cdots \Phi_{J},
\]

where \( C^{1 \ldots J} \) is traceless symmetric and \( C^{1 \ldots J} C^{*1 \ldots J} = 1. \) The corresponding spherical harmonic is given by \( Y_{J}(\theta, \phi) = C^{1 \ldots J} x^{1} \cdots x^{J}. \) A properly normalized (i.e. (3)) operator built on \( \text{Tr}(u \cdot \Phi)^{J} \) will then correspond to

\[
Y_{J}(\theta, \phi) = N_{J}(u)|u_{1} \sin \theta \cos \phi + u_{2} \sin \theta \sin \phi + u_{3} \cos \theta|^{J}
\]

for some normalization \( N_{J}(u). \) If we choose \( u_{1} = u_{2} = 0 \) and \( u_{3} = \pm i u_{4} = 1, \) i.e. the operator \( \text{Tr}(\Phi_{3} \pm i \Phi_{4})^{J} / \sqrt{J!}, \) then \( N_{J}(u) = 2^{-J/2}. \)

### Appendix C. R-symmetry

Let \( O_{J} = \frac{1}{\sqrt{J!}} \text{Tr} (\Phi_{1} + i \Phi_{2})^{J} \) and let \( U \) be a rotation in the \( x^{1}-x^{2} \) plane. Then

\[
[O_{J}(x) W[C_{1/4}]] = [U O_{J}(x) W[C_{1/4}]] = [O_{J}(U x) U W[C_{1/4}] U^{\dagger}].
\]

Examining \( C_{1/4} \) in (12), we see that the spatial rotation acting on \( W[C_{1/4}] \) may be realized by a shift in the contour parameter \( \tau, \) which can in turn by compensated by an R-symmetry rotation \( R \) in the \( \Theta^{1-\Theta^{2}} \) plane, \( U W[C_{1/4}] U^{\dagger} = RW[C_{1/4}] R^{\dagger}. \) Then,

\[
[O_{J}(x) W[C_{1/4}]] = [R O_{J}(U x) R^{\dagger} W[C_{1/4}]].
\]

The operator expansion coefficient depends on the leading asymptotic in large \( x \) which is a function of only the length of \( C_{1/4} \) and \( x^{2}, \)

\[
[O_{J}(x) W[C_{1/4}]] \simeq \left( \frac{2 \pi R}{4 \pi^{2} x^{2}} \right)^{J} \xi_{J} + \cdots.
\]

Performing the \( \Theta^{1-\Theta^{2}} \) plane R-symmetry transformation on \( O_{J} \) multiplies it by a phase \( \exp(i J \phi) \) so that,

\[
[R O_{J}(U x) R^{\dagger} W[C_{1/4}]] \simeq e^{i J \phi} \left( \frac{2 \pi R}{4 \pi^{2} x^{2}} \right)^{J} \xi_{J} + \cdots = e^{i J \phi} \left( \frac{2 \pi R}{4 \pi^{2} x^{2}} \right)^{J} \xi_{J} + \cdots.
\]

Using (C.2) and (C.3), we have \( e^{i J \phi} \xi_{J} = \xi_{J}, \) i.e. \( \xi_{J} = 0. \)

### References

Gauge invariant formulation of massive totally symmetric fermionic fields in (A)dS space

R.R. Metsaev

Department of Theoretical Physics, P.N. Lebedev Physical Institute, Leninsky prospect 53, Moscow 119991, Russia

Received 6 September 2006; received in revised form 1 November 2006; accepted 2 November 2006

Available online 10 November 2006

Editor: N. Glover

Abstract

Massive arbitrary spin totally symmetric free fermionic fields propagating in $d$-dimensional (anti-)de Sitter space–time are investigated. Gauge invariant action and the corresponding gauge transformations for such fields are proposed. The results are formulated in terms of various mass parameters used in the literature as well as the lowest eigenvalues of the energy operator. We apply our results to a study of partial masslessness of fermionic fields in (A)dS, and in the case of $d = 4$ confirm the conjecture made in the earlier literature.

© 2006 Elsevier B.V. All rights reserved.

1. Introduction

Conjectured duality [1] of conformal $\mathcal{N} = 4$ SYM theory and superstring theory in $AdS_5 \times S^5$ Ramond–Ramond background has led to intensive study of field (string) dynamics in AdS space. By now it is clear that in order to understand the conjectured duality better it is necessary to develop powerful approaches to study of field (string) dynamics in AdS space. Light-cone approach is one of the promising approaches which might be helpful to understand AdS/CFT duality better. As is well known, quantization of Green–Schwarz superstrings propagating in flat space is straightforward only in the light-cone gauge. Since, by analogy with flat space, we expect that quantization of the Green–Schwarz AdS superstring propagating with Ramond–Ramond flux [2] will be straightforward only in a light-cone gauge [3] we believe that in the stringy perspective of AdS/CFT correspondence the light-cone approach to field dynamics in AdS is a fruitful direction to go. Light-cone approach to dynamics of massive fields in AdS space was developed in [4,5] and a complete description of massive arbitrary spin bosonic and fermionic fields in $AdS_5$ was obtained in [6].

Unfortunately, this is not enough for a complete study of the AdS/CFT correspondence because in order to apply the light-cone approach to study of superstring in AdS space we need a light-cone formulation of field dynamics in $AdS_5 \times S^5$ Ramond–Ramond background. Practically useful and self-contained way to give a light-cone gauge description is to start with a Lorentz covariant and gauge invariant description of field dynamics in $AdS_5 \times S^5$ Ramond–Ramond background and then to impose the light-cone gauge. Our experience led us to conclusion that the most simple way to develop light-cone approach in $AdS_5 \times S^5$ space is to start with gauge invariant description of fermionic fields. It turns out, however, that gauge invariant description of massive fermionic fields (with fixed but arbitrary spin) even in $AdS_5$ is still not available in the literature. In this Letter we develop Lagrangian Lorentz covariant and gauge invariant formulation\(^1\) for massive totally symmetric arbitrary spin fermionic fields in $(A)dS_d$ space. We believe that our results will be helpful to find a gauge invariant description of arbitrary spin fields in $AdS_5 \times S^5$ case. Our approach allows us to study fermionic fields in $AdS_d$ space and $dS_d$ space on an equal footing. In this Letter we apply our results to study of par-

---

\(^1\) Sometimes, a gauge invariant approach to massive fields is referred to as a Stueckelberg approach.
tial masslessness of fermionic fields in \((A)dS_d\). For \(d = 4\) our results confirm the conjecture made in Ref. [7].

Before proceeding to the main theme of this Letter let us mention briefly the approaches which could be used to discuss gauge invariant action for fields in \((A)dS\). Since the works [8–11] devoted to massless fields in \(AdS\) various descriptions of massive and massless arbitrary spin fields in \((A)dS\) have been developed. In particular, an ambient space formulation was discussed in [12,13] and various BRST formulations were studied in [14–17]. The frame-like formulations of free fields which seems to be the most suitable for formulation of the theory of interacting fields in \((A)dS\) was developed in [18,19]. Other interesting formulations of higher spin theories were also discussed recently in [20–23]. In this Letter we adopt the approach of Ref. [24] devoted to the bosonic fields in \((A)dS\). This approach turns out to be the most useful for our purposes.

2. Gauge invariant action of massive fermionic field in \((A)dS\)

In \(d\)-dimensional \((A)dS_d\) space the massive totally symmetric arbitrary spin fermionic field is labelled by one mass parameter and by one half-integer spin label \(s + \frac{1}{2}\) where \(s > 0\) is an integer number. To discuss Lorentz covariant and gauge invariant formulation of such field we introduce Dirac complex-valued tensor-spinor fields \(\bar{\psi}_s^{\alpha A_1\cdots A_s}\) which define the theory of interacting fields in \((A)dS\) was developed in [18,19]. Other

\[\sum_{s' = 0}^{s} \bigoplus_{\alpha A_1\cdots A_s} \psi_s^{\alpha A_1\cdots A_s}.\]  

\(2.1\)

In order to obtain the gauge invariant description of a massive field in an easy-to-use form, let us introduce a set of the creation and annihilation operators \(\alpha^A, \bar{\alpha}^A, \zeta, \bar{\alpha}^A, \zeta\) defined by the relations

\[\bar{\alpha}^A |0\rangle = 0, \quad \bar{\alpha}^A |\zeta\rangle = 0, \quad \zeta |0\rangle = 1, \quad \zeta |\zeta\rangle = 1.\]  

\(2.2\)

where \(\eta^{AB}\) is the mostly positive flat metric tensor. The oscillators \(\alpha^A, \bar{\alpha}^A\) and \(\zeta, \bar{\zeta}\) transform in the respective vector and scalar representations of the \(so(d - 1, 1)\) Lorentz algebra. The tensor-spinor fields \(2.1\) can be collected into a ket-vector \(\psi\) defined by

\[|\psi\rangle = \sum_{s' = 0}^{s} \langle s'| \psi_{s'}\rangle,\]  

\(2.3\)

\[|\psi_{s'}\rangle = \alpha^{A_1} \cdots \alpha^{A_{s'-1}} \psi^{A_{s'-1} \cdots A_1} (x) |0\rangle.\]  

\(2.4\)

Here and below spinor indices are implicit. The ket-vector \(|\psi_{s'}\rangle\)

\[(a^A \bar{\alpha}^A - s') |\psi_{s'}\rangle = 0, \quad s' = 0, 1, \ldots, s,\]  

\(2.5\)

which tells us that \(|\psi_{s'}\rangle\) is a degree \(s'\) homogeneous polynomial in the oscillator \(a^A\). In terms of the ket-vector \(|\psi\rangle\) the algebraic constraints \(2.5\), \(2.6\) take the form

\[(a^A \bar{\alpha}^A + N_s - s) |\psi\rangle = 0,\]  

\(2.7\)

\[\gamma^A \bar{\alpha}^A |\psi\rangle = 0,\]  

\(2.8\)

Eq. \(2.7\) tells us that \(|\psi\rangle\) is a degree \(s\) homogeneous polynomial in the oscillators \(a^A, \gamma^A\).

Lagrangian for the massive fermionic field in \((A)dS_d\) space we found takes the form

\[L = L_{\text{det}} + L_m,\]  

\(2.10\)

where \(L_{\text{det}}\) stands for a derivative depending part of \(L\), while \(L_m\) stands for a mass part of \(L^A\)

\[i e^{-1} L_{\text{det}} = \langle \psi | \mathcal{L} | \psi \rangle,\]  

\(2.11\)

\[i e^{-1} L_m = \langle \psi | M | \psi \rangle.\]  

\(2.12\)

The standard first-derivative differential operator \(L\) which enters \(L_{\text{det}}\) \(2.11\) is given by

\[L \equiv \bar{\phi} - a D \bar{\alpha} - \bar{\gamma} a \bar{\alpha} D + \gamma a \bar{\phi} \bar{\gamma} \bar{\alpha} + \frac{1}{2} \gamma a \bar{\alpha} a \bar{\gamma} a \bar{\alpha} D^2\]  

\(2.13\)

where we use the notation

\[\gamma a \equiv \gamma^A a^A, \quad \bar{\gamma} a \equiv \gamma^A \bar{\alpha}^A, \quad a^2 \equiv \alpha^A \bar{\alpha}^A, \quad a^2 \equiv \bar{\alpha}^A \bar{\alpha}^A, \]  

\(2.14\)

\[\bar{\phi} \equiv \gamma^A D^A, \quad a D \equiv a^A D^A, \quad \bar{\alpha} D \equiv \bar{\alpha}^A D^A, \quad D_A \equiv e^\mu_A D^\mu,\]  

\(2.15\)

and \(e^\mu_A\) stands for inverse vielbein of \((A)dS_d\) space, while \(D\)

\[D_m \equiv \partial_\mu + \frac{1}{2} \omega^{AB}_\mu M^{AB}.\]  

\(2.16\)

\[\text{Important constraint (2.6) was introduced for the first time in [9] while study of massless fermionic fields in \(AdS\). This constraint implies that the field \(|\psi_{s'}\rangle\) being reducible representation of the Lorentz algebra \(so(d - 1, 1)\) is decomposed into spin \(s' + \frac{1}{2}, s' - \frac{1}{2}, s' - \frac{1}{2}\) \(\text{irreps of the Lorentz algebra. Various Lagrangian formulations in terms of unconstrained fields in flat space and \((A)dS_d\) space may be found e.g. in [26–30].}\]

\[\text{The bra-vector } |\psi\rangle \text{ is defined according the rule } |\psi\rangle = (|\psi\rangle)^\dagger \psi^0.\]
The $\omega_{\mu}^{AB}$ is the Lorentz connection of (A)dS$_d$ space, while a spin operator $M^{AB}$ forms a representation of the Lorentz algebra so(d − 1, 1):

$$M^{AB} = M^b_{AB} + \frac{1}{2} \epsilon^{AB}, \quad M^b_{AB} \equiv \alpha^A \bar{\alpha}^B - \alpha^B \bar{\alpha}^A,$$

$$\gamma^{AB} \equiv \frac{1}{2} (\gamma^A \gamma^B - \gamma^B \gamma^A). \quad (2.17)$$

We note that our derivative depending part of the Lagrangian $\mathcal{L}$ is nothing but a sum of the Lagrangians of Ref. [9] for the tensor-spinor fields (2.1).

We now proceed with discussion of the mass operator $M$ (2.12). The operator $M$ is given by

$$M = \bigg(1 - \gamma^A \gamma^B - \frac{1}{4} \gamma^2 \bar{\gamma}^2\bigg) \bar{m}_1 + \bar{m}_4 \bigg(\gamma \alpha \bar{\zeta} - \frac{1}{2} \alpha^2 \bar{\gamma}^2 \bar{\gamma}\bigg) - \bigg(\zeta \gamma \bar{\alpha} - \frac{1}{2} \gamma \alpha \bar{\zeta} \bar{\gamma}^2\bigg) \bar{m}_4, \quad (2.18)$$

where operators $\bar{m}_1$, $\bar{m}_4$ do not depend on the $\gamma$-matrices and $\alpha$-oscillators, and take the form

$$\bar{m}_1 = \frac{2s + d - 2}{2s + d - 2 - 2N_\zeta}, \quad (2.19)$$

$$\bar{m}_4 = \left(\frac{2s + d - 3 - N_\zeta}{2s + d - 4 - 2N_\zeta} F(\kappa, s, N_\zeta)\right)^{1/2}. \quad (2.20)$$

Function $F(\kappa, s, N_\zeta)$ depends on a mass parameter $\kappa$, spin $s$ and operator $N_\zeta$, and is given by

$$F(\kappa, s, N_\zeta) = \kappa^2 + \theta\left(s^2 - \frac{d - 4}{2} - N_\zeta\right)^2. \quad (2.21)$$

$F$ is restricted to be positive and throughout this Letter, unless otherwise specified, we use the convention\(^5\):

$$\theta = \begin{cases} 
-1 & \text{for AdS space,} \\
0 & \text{for flat space,} \\
+1 & \text{for dS space.} 
\end{cases} \quad (2.22)$$

The mass parameter $\kappa$ is a freedom of our solution, i.e. gauge invariance allows us to find Lagrangian completely by module of mass parameter as it should be for the case of massive fields.

Now we discuss gauge symmetries of the action

$$S = \int d^dx \mathcal{L}. \quad (2.23)$$

To this end we introduce parameters of gauge transformations $\epsilon_1^{A_1} \cdots \epsilon_s^{A_s}$, $s = 0, 1, \ldots, s - 1$ which are $\gamma$-traceless (for $s' > 0$) Dirac complex-valued tensor-spinor spin $s' + \frac{1}{2}$ fields of the so(d − 1, 1) Lorentz algebra, i.e. we start with a collection of the tensor-spinor fields

$$\sum_{s' = 0}^{s-1} \epsilon_1^{A_1} \cdots \epsilon_{s'}^{A_{s'}}, \quad \gamma^A \epsilon^{AA_2 \cdots A_{s'}} = 0, \quad \text{for } s' > 0. \quad (2.24)$$

\(^5\) Thus our Lagrangian gives description of massive fermionic fields in (A)dS space and flat space on an equal footing. Discussion of massive fermionic fields in flat space in framework of BRST approach may be found in [26,31].

As before to simplify our expressions we use the ket-vector of gauge transformations parameter

$$|\epsilon\rangle \equiv \sum_{s' = 0}^{s-1} \epsilon^{s'-1-s'} |\epsilon_{s'}\rangle, \quad (2.25)$$

$$|\epsilon_{s'}\rangle \equiv \epsilon^{A_1} \cdots \epsilon^{A_{s'}} \epsilon^{A_1 \cdots A_{s'}}(x)|0\rangle. \quad (2.26)$$

The ket-vector $|\epsilon\rangle$ satisfies the algebraic constraints

$$\langle \alpha \bar{\alpha} + N_\zeta - s + 1 | \epsilon \rangle = 0, \quad (2.27)$$

$$\gamma \bar{\alpha} |\epsilon\rangle = 0. \quad (2.28)$$

The constraint (2.27) tells us that the ket-vector $|\epsilon\rangle$ is a degree $s - 1$ homogeneous polynomial in the oscillators $\alpha^A$, $\zeta$, while the constraint (2.28) respects the $\gamma$-tracelessness of $|\epsilon\rangle$.

Now the gauge transformations under which the action (2.23) is invariant take the form

$$\delta |\psi\rangle = (\alpha D + \Delta) |\psi\rangle, \quad (2.29)$$

$$\Delta \equiv \xi \Delta_1 + \gamma \alpha \Delta_2 + \alpha^2 \Delta_3 \zeta, \quad (2.30)$$

where operators $\Delta_1$, $\Delta_2$, $\Delta_3$ do not depend on the $\gamma$-matrices and $\alpha$-oscillators, and take the form

$$\Delta_1 = \left(\frac{2s + d - 3 - N_\zeta}{2s + d - 4 - 2N_\zeta} F(\kappa, s, N_\zeta)\right)^{1/2}, \quad (2.31)$$

$$\Delta_2 = \frac{2s + d - 2}{(2s + d - 2 - 2N_\zeta)(2s + d - 4 - 2N_\zeta)} \kappa, \quad (2.32)$$

$$\Delta_3 = -\left(\frac{2s + d - 3 - N_\zeta}{(2s + d - 4 - 2N_\zeta)} F(\kappa, s, N_\zeta)\right)^{1/2}, \quad (2.33)$$

and $F$ is defined in (2.21). Thus we expressed our results in terms of the mass parameter $\kappa$. Since there is no commonly accepted definition of mass in (A)dS we relate our mass parameter $\kappa$ with various mass parameters used in the literature.

One of the most-used definitions of mass, which we denote by $m_D$, is obtained from the following expansion of mass part of the Lagrangian:

$$ie^{-1} \mathcal{L}_m = \langle \psi_s | m_D | \psi_s \rangle + \cdots, \quad (2.34)$$

where dots stand for terms involving $|\psi_s\rangle$, $s' < s$, and for contribution which vanishes while imposing the constraint $\gamma \bar{\alpha} |\psi\rangle$. Comparing (2.34) with (2.18), (2.19) leads then to the identification

$$\kappa = m_D. \quad (2.35)$$

Another definition of mass parameter for fermionic fields in AdS$_d$ [5], denoted by $m$, can be obtained by requiring that the value of $m = 0$ corresponds to the massless fields. For the case of spin $s + \frac{1}{2}$ field in AdS$_d$ the mass parameter $m$ is related with $m_D$ as

$$m_D = m + s + \frac{d - 4}{2} \quad \text{for AdS}_d, \quad (2.36)$$

where $m > 0$ corresponds to massive unitary irreps of the so(d − 1, 2) algebra [5,12]. Below we demonstrate that natural generalization of (2.36) which is valid for both AdS and
dS spaces is given by
\[ m_D = m + \sqrt{-\theta} \left( s + \frac{d-4}{2} \right) \text{ for } (A)dS_d. \] (2.37)

Since sometimes in the case of AdS the formulation in terms of the lowest eigenvalue of energy operator \( E_0 \) is preferable we now express our results in terms of \( E_0 \). To this end we use the relation found in [5]:
\[ m = E_0 - s - d + \frac{5}{2} \text{ for } AdS_d. \] (2.38)

Making use then (2.35), (2.36) we get for the case of AdS\( d \) the desired relations
\[ \kappa = E_0 - \frac{d - 1}{2}, \]
\[ F = \left( E_0 - s - d + \frac{5}{2} + N_\zeta \right) \left( E_0 + s - \frac{3}{2} - N_\zeta \right). \] (2.39) (2.40)

3. Limit of massless fields in \((A)dS_d\)

In previous section we presented the action for the massive field. In limit as the mass parameter \( m \) tends to zero our Lagrangian leads to the Lagrangian for massless field in \((A)dS_d\). Let us discuss the massless limit in detail. To realize limit of massless field in \((A)dS_d\) we take (see (2.35), (2.37))
\[ m_D \to \sqrt{-\theta} \left( s + \frac{d-4}{2} \right) \iff m \to 0. \] (3.1)

We now demonstrate that this limit leads to appearance of the invariant subspace in \( |\psi\rangle \) (2.3) and this invariant subspace, denoted by \( |\psi^{m=0}\rangle \), is given by the leading \( (s' = s) \) term in (2.3):
\[ |\psi^{m=0}\rangle = |\psi_s\rangle. \] (3.2)

All that is required is to demonstrate that in the limit (3.1), the ket-vector \( |\psi^{m=0}\rangle \) satisfies the following requirements: (i) \( |\psi^{m=0}\rangle \) is invariant under action of the mass operator \( M \); (ii) the gauge transformation of the ket-vector \( |\psi^{m=0}\rangle \) becomes the standard gauge transformation of massless field. To this end we note that an action of the mass operator \( M \) on \( |\psi_s\rangle \) and the gauge transformation of \( |\psi_s\rangle \) take the form
\[ M|\psi_s\rangle = \left( 1 - \gamma_\alpha \gamma_\beta - \frac{1}{4} \alpha_\beta \bar{\alpha}_\beta \right) m_1(0)|\psi_s\rangle + \left( \gamma_\alpha - \frac{1}{2} \alpha_\gamma \bar{\alpha}_\gamma \right) m_4(0) |\psi_{s-1}\rangle, \]
\[ \delta|\psi_s\rangle = \left( \alpha D + \gamma_\alpha \bar{\Delta}_2(0) |\epsilon_{s-1}\rangle + \alpha^2 \bar{\Delta}_3(0) |\epsilon_{s-2}\rangle \right), \] (3.3) (3.4)
where \( m_1,4(0) \) (3.3) and \( \bar{\Delta}_{2,3}(n) \) (3.4) stand for \( \bar{m}_{1,4}(2.19), (2.20) \) and \( \bar{\Delta}_{2,3}(2.32), (2.23) \) in which we set \( N_\zeta = 0 \). Taking into account
\[ \lim_{m \to 0} \bar{m}_4(0) = 0, \quad \lim_{m \to 0} \bar{\Delta}_3(0) = 0, \] (3.5)
we see that if \( m = 0 \) then the ket-vector \( |\psi_s\rangle \) is indeed invariant under action of the mass operator \( M \) and a realization of the mass operator on the ket-vector \( |\psi_s\rangle \) takes the form
\[ M|\psi_s\rangle = \sqrt{-\theta} \left( s + \frac{d-4}{2} \right) \left( 1 - \gamma_\alpha \gamma_\beta - \frac{1}{4} \alpha_\beta \bar{\alpha}_\beta \right), \] (3.6)
while the relations (3.4), (3.5) lead to the gauge transformation
\[ \delta|\psi^{m=0}\rangle = \left( \alpha D + \frac{\sqrt{-\theta}}{2} \gamma_\alpha \right) |\epsilon_{s-1}\rangle, \] (3.7)
which is noting but the standard gauge transformation of massless field in \((A)dS_d\) space. Thus the Lagrangian for massless spin \( s + \frac{1}{2} \) fermionic field in \((A)dS_d\) space takes the form:
\[ i e^{-1} \mathcal{L} = \langle \psi^{m=0} | L + M^{m=0} | \psi^{m=0}\rangle, \]
(3.8)
where \( |\psi^{m=0}\rangle \) is given by (3.2), (3.4), while the operators \( L \) and \( M^{m=0} \) are defined by (2.13) and (3.6) respectively. The remaining ket-vectors \( |\psi_{s-1}\rangle, \ldots, |\psi_0\rangle \) (2.3) decouple in the massless limit and they describe spin \( s - \frac{1}{2} \) massive field, i.e. in the massless limit the generic field \( |\psi\rangle \) is decomposed into two decoupling systems—one massive spin \( s + \frac{1}{2} \) field and one massive spin \( s - \frac{1}{2} \) field. Adopting (2.34) for spin \( s - \frac{1}{2} \) field we find mass of the massive spin \( s - \frac{1}{2} \) field:
\[ m_D = \sqrt{-\theta} \left( s + \frac{(d - 4)}{2} \right). \]

4. Partial masslessness of fermionic fields in \((A)dS_d\)

Here we apply our results to study of partial masslessness\footnote{Our Lagrangian (3.8) is a generalization to \( d \)-dimensions of the Lagrangian of Ref. [9] for massless field in \((A)dS_d\). Alternative Lagrangian descriptions of massless fermionic fields in \((A)dS_d\) may be found in [11, 21].} of fermionic fields in \((A)dS_d\). We confirm conjecture of Ref. [7] for \( d = 4 \) and obtain a generalization to the case of arbitrary \( d > 4 \). In this section we assume that the \( \theta \) (2.22) takes the values \( \pm 1 \). We start our discussion of partial masslessness of fermionic fields with simplest case of (see also Ref. [34]).

4.1. Massive spin 5/2 field

Such field is described by ket-vectors \( |\psi_2\rangle, |\psi_1\rangle, |\psi_0\rangle \) (see (2.3)). For spin 5/2 field there is one critical value of \( m_D \) which leads to appearance of partial massless field. For this critical value of \( m_D \) the generic field \( |\psi\rangle \) is decomposed into one partial massless field and one massive spin \( \frac{1}{2} \) field. To demonstrate this we consider the gauge transformations (2.29),
\[ \delta|\psi_2\rangle = \alpha D \epsilon_2(1) \epsilon_1 \epsilon_0 + \alpha^2 \bar{\Delta}_3(0) \epsilon_0, \]
\[ \delta|\psi_1\rangle = \alpha D \epsilon_1 + \bar{\Delta}_1(0) \epsilon_1 + \gamma \alpha \bar{\Delta}_2(1) \epsilon_0, \]
\[ \delta|\psi_0\rangle = \bar{\Delta}_1(1) \epsilon_0, \] (4.1) (4.2) (4.3)
where \( \bar{\Delta}_{1,2,3}(n) \) are given in (2.31)--(2.33) in which we set \( s = 2 \) and argument \( n \) stands for an eigenvalue of the operator \( N_\zeta \). The critical value of \( m_D \) is obtained from the requirement of decoupling of the field \( |\psi_0\rangle \). This requirement amounts to the equation \( \bar{\Delta}_1(1) = 0 \) which leads to the critical value
\[ m^2_{D(0)} = -\theta \left( 1 + \frac{d - 4}{2} \right)^2. \] (4.4)

For this value of \( m_D \) the generic field \( |\psi\rangle \) is decomposed into two decoupling systems—one partial massless field described
by $|\psi_2\rangle$, $|\psi_1\rangle$ and one massive spin $\frac{1}{2}$ field described by $|\psi_0\rangle$.

We proceed with discussion of partial masslessness for

4.2. Massive spin 7/2 field

Spin 7/2 field $|\psi\rangle$ is described by ket-vectors $|\psi_3\rangle$, $|\psi_2\rangle$, $|\psi_1\rangle$, $|\psi_0\rangle$ (see (2.3)). The gauge transformations (2.29) for these ket-vectors take the form

\[
\delta|\psi_3\rangle = aD(\epsilon_2) + \gamma \alpha \tilde{\Delta}(0)|\epsilon_2\rangle + a^2 \tilde{\Delta}(0)|\epsilon_1\rangle,
\]

\[
\delta|\psi_2\rangle = aD(\epsilon_1) + \tilde{\Delta}(0)|\epsilon_2\rangle + \gamma \alpha \tilde{\Delta}(1)|\epsilon_1\rangle
\]

\[
+ 2a^2 \Delta(1)\epsilon_0, \quad \delta|\psi_1\rangle = aD(\epsilon_0) + \tilde{\Delta}(1)|\epsilon_1\rangle + \gamma \alpha \tilde{\Delta}(2)|\epsilon_0\rangle,
\]

\[
\delta|\psi_0\rangle = \tilde{\Delta}(2)|\epsilon_0\rangle,
\]

where expressions for $\tilde{\Delta}_{1,2,3}(n)$ are given in (2.31)–(2.33) in which we set $s = 3$ and argument $n$ stands for an eigenvalue of the operator $N$. For the spin 7/2 field there are two critical values of $m_D$. For each critical value of $m_D$ the generic field $|\psi\rangle$ is decomposed into one partial massless field and one massive field. We consider these critical values in turn.

First critical value of $m_D$ is obtained from the requirement of decoupling of the field $|\psi_1\rangle$ (see (4.8)). This requirement amounts to the equation $\tilde{\Delta}(2) = 0$ which leads to the critical value

\[
m_{D(0)}^2 = -\theta \left( 1 + \frac{d - 4}{2} \right)^2.
\]

For this value of $m_D$ the generic field $|\psi\rangle$ is decomposed into two decoupling systems—one partial massless field described by $|\psi_3\rangle$, $|\psi_2\rangle$, $|\psi_1\rangle$ and one massive spin $\frac{1}{2}$ field $|\psi_0\rangle$.

The second critical value of $m_D$ is obtained from the requirement of decoupling of the fields $|\psi_1\rangle$, $|\psi_0\rangle$ (see (4.6), (4.7)). This requirement amounts to the equations $\tilde{\Delta}(1) = 0$, $\Delta(3) = 0$ which lead to the critical value

\[
m_{D(1)}^2 = -\theta \left( 2 + \frac{d - 4}{2} \right)^2.
\]

For this $m_D$ the generic field $|\psi\rangle$ is decomposed into one partial massless field described by $|\psi_3\rangle$, $|\psi_2\rangle$ and one massive spin $\frac{1}{2}$ field described by $|\psi_1\rangle$, $|\psi_0\rangle$. We finish with partial masslessness for

4.3. Massive arbitrary spin $s + \frac{1}{2}$ field

Such field is described by ket-vectors $|\psi_s\rangle$, $s' = 0, 1, \ldots, s$. Gauge transformations (2.29) for these ket-vectors take the form

\[
\delta|\psi_s\rangle = aD(\epsilon_{s'-1}) + \tilde{\Delta}(s - s'-1)|\epsilon_{s'}\rangle
\]

\[
+ \gamma \alpha \tilde{\Delta}(s - s')|\epsilon_{s'-1}\rangle
\]

\[
+ (s - s') + 1a^2 \Delta(3)\epsilon_0 = 0,
\]

where $\tilde{\Delta}_{1,2,3}(n)$ are given in (2.31)–(2.33) and argument $n$ stands for an eigenvalue of the operator $N$. For values $s' = 0, 1, s$ (4.11) we use the convention $|\epsilon_{-2}\rangle = |\epsilon_{-1}\rangle = |\epsilon_1\rangle = 0$.

For the spin $s + \frac{1}{2}$ field there are $s - 1$ critical values of $m_D$, denoted by $m_{D(n)}$, $n = 0, 1, \ldots, s - 2$ (the case of $n = s - 1$ leads to massless field and was considered in Section 3). For each $m_{D(n)}$ we note that the requirement of decoupling of the fields $|\psi_{n}\rangle, \ldots, |\psi_{0}\rangle$, $n = 0, \ldots, s - 2$ amounts to equations $\tilde{\Delta}(s - n - 1) = 0$, $\Delta(3) = 0$. Solution to these equations

\[
m_{D(n)}^2 = -\theta \left( n + 1 + \frac{d - 4}{2} \right)^2,
\]

is in agreement with conjecture made in Ref. [7] for the case of $d = 4$. Thus we confirmed conjecture of Ref. [7] and obtained $m_{D(n)}$ for $d > 4$. For each $m_{D(n)}$ the generic field $|\psi\rangle$ is decomposed into two decoupling systems—one partial massless field $|\psi_{\text{par}}\rangle$ described by $|\psi_3\rangle, \ldots, |\psi_{n+1}\rangle$, and one massive spin $n + \frac{1}{2}$ field $|\psi_{\text{msv}}\rangle$ described by $|\psi_n\rangle, \ldots, |\psi_0\rangle$. This is to say that by decomposing $|\psi\rangle$ (2.3) into the respective ket-vectors

\[
|\psi_{\text{par}}^{(s)}\rangle \equiv \sum_{s'=n+1}^{s} \xi^{-s'}|\psi_{s'}\rangle,
\]

\[
|\psi_{\text{msv}}^{(s)}\rangle = \sum_{s'=0}^{n} \xi^{-s'}|\psi_{s'}\rangle,
\]

one can make sure that if $m_D = m_{D(n)}$ then the mass part of the Lagrangian (2.12) is factorized

\[
\text{ie}^{-1}L_m = |\psi_{\text{par}}^{(s)}\rangle |\mathcal{M}| |\psi_{\text{par}}^{(s)}\rangle + |\psi_{\text{msv}}^{(s)}\rangle |\mathcal{M}| |\psi_{\text{msv}}^{(s)}\rangle.
\]

For values $m_D = m_{D(n)}$ the gauge transformations (2.29) are also factorized, while $\mathcal{L}_{\text{der}}$ (2.11) is factorized for arbitrary $m_D$.

5. Uniqueness of Lagrangian for massive fermionic field

We now demonstrate that the Lagrangian and gauge transformations are uniquely determined by requiring that the action be gauge invariant. We formulate our statement. Suppose the derivative depending part of the Lagrangian is given by (2.11), while the derivative depending part of gauge transformations (2.29) is governed by $\alpha D$-term. Suppose the gauge field $|\psi\rangle$ (2.3) and the gauge transformations parameter $|\epsilon\rangle$ (2.25) satisfy the respective constraints (2.8), (2.28). Then we state that the mass operator $\mathcal{M}$ given in (2.18) and the operator $\Delta$ which enters gauge transformations (2.29) are uniquely determined

8 We note that the key point is not positivity or even reality of the mass part of action (2.12), but rather stability of the energy and unitarity of the underlying physical representations. For bosons a negative mass term is allowed in AdS (the Breitenlohner–Freedman bound, [35]), while partially massless fermions even have an imaginary mass term in their actions but are still stable and unitary in $dS$. Partially massless fermions are not unitary in AdS (see Refs. [7,34,36]).

9 $m_D$- and $m$-masses of the field $|\psi_{\text{par}}^{(s)}\rangle$ are given by: $m_D = \sqrt{d/n + 1 + (d - 4/2)}$, $m = \sqrt{d/n + 1 - s}$, while for the field $|\psi_{\text{msv}}^{(s)}\rangle$ we get $m_D = \sqrt{d/n + 1 - s}$, $m = \sqrt{d/n + 1 - s}$. 

\[
\]
by the following requirements: (i) the action be gauge invariant; (ii) there are no invariant subspaces in $|\psi\rangle$ under action of gauge transformations. Here we outline prove of this statement.

We start with general form of the mass operator $M$ and the operator $\Delta$:

$$M = m_1 + \gamma am_2 Y\Lambda + a^2 m_3 \alpha^2 + \gamma am_4 + a^2 m_5 \gamma\Lambda - m_7^2 \gamma\Lambda - \gamma am_7 \alpha^2 + a^2 m_6 + m_8 \alpha\Lambda,$$

$$\Delta = \Delta_1 + \gamma a\Delta_2 + a^2 \Delta_3,$$

where $m_1, \ldots, 6$ and $\Delta_1, \Delta_2, 3$ do depend on $\gamma$-matrices and $a$-oscillators, and are given by

$$m_1 = \tilde{m}_1, \quad m_2 = \tilde{m}_2, \quad m_3 = \tilde{m}_3,$$

$$m_4 = \tilde{m}_4 \zeta, \quad m_5 = \zeta \tilde{m}_5,$$

$$m_6 = \tilde{m}_6 \zeta^2, \quad m_7 = \zeta^2 \tilde{m}_6,$$

$$\Delta_1 = \zeta \tilde{\Delta}_1, \quad \Delta_2 = \tilde{\Delta}_2, \quad \Delta_3 = \tilde{\Delta}_3 \zeta.$$

Operators $\tilde{m}_1, \ldots, 6$ and $\tilde{\Delta}_1, 2, 3$ depend only on $N_\tau$ (2.9). $\tilde{m}_1, \ldots, 6$ stand for hermitian conjugate of $\tilde{m}_1, \ldots, 6$. Since $\tilde{m}_1, \ldots, 3$ are hermitian, $\tilde{m}_1, \ldots, 3$ the operators $\tilde{m}_1, \ldots, 3$ are real-valued functions of $N_\tau$ from the very beginning. These properties of $\tilde{m}_1, \ldots, 3$ and $\tilde{\Delta}_1, 2, 3$ and expressions for $M$ (5.1), $\Delta$ (5.2) are obtained by requiring that:

(i) $M$ and $\Delta$ commute with the spin operator of the Lorentz algebra $M^{AB} = (2.17)$ and satisfy the commutators $[\alpha\Lambda + N_\tau, M] = 0, [\alpha\Lambda + \Delta, \Delta] = \Delta$;

(ii) $M$ does not involve terms like $a^2 \gamma \alpha f_1$ and $f_2 a^2 \gamma\Lambda$, where $f_1, 2$ are polynomial in the oscillators (such terms in view of (2.8) do not contribute to $L_m$);

(iii) $\Delta$ does not involve terms like $a^2 \gamma \alpha f_3, \alpha f_4 \tilde{\Lambda}$, where $f_3, 4$ are polynomial in the oscillators (the $f_3$-terms lead to violation of constraint (2.8) for gauge transformed field (2.29), while the $f_4$-terms in view of (2.28) do not contribute to $\delta|\psi\rangle$ (2.29));

(iv) $M$ and hermitian conjugated of $M$ satisfy the relation $M^\dagger = -\gamma^0 M \gamma^0.$

Thus all that is required is to find dependence of the operators $\tilde{m}_1, \ldots, 6$ and $\tilde{\Delta}_1, 2, 3$ on $N_\tau$. We now demonstrate that this dependence can be determined by requiring that the action be gauge invariant. We evaluate the variation of the action (2.10), (2.23) under gauge transformations (2.29), (5.2).

$$\delta S = -i \int d^4x \epsilon e|\psi\rangle (\Phi X_1 + \gamma a\Lambda X_2 + aD \Lambda X_3$$

$$+ a\Lambda \Phi X_4 + a^2 \tilde{\Lambda} DX_5 + a^2 \Phi X_6 + aD\gamma \alpha X_7$$

$$+ a^2 a^2 DX_8 + \tilde{\Lambda} DX_9 + Y(0) + \gamma a (Y(1) + Y(1))$$

$$+ \alpha^2 Y(2))|\epsilon\rangle + h.c.,$$

where we use the notation

$$X_1 = \Delta_1 - m_4^\dagger, \quad X_2 = -\Delta_1 - m_3^\dagger,$$

$$X_3 = -(2s + d - 4 - 2N_\tau) \Delta_2 + m_1,$$

$$X_4 = (2s + d - 4 - 2N_\tau) \Delta_2 + m_2,$$

$$X_5 = \frac{1}{2}(2s + d - 4 - 2N_\tau) \Delta_2 + 2m_3,$$

$$X_6 = -\frac{1}{2}(2s + d - 4 - 2N_\tau) \Delta_3 + m_5,$$

$$X_7 = (2s + d - 4 - 2N_\tau) \Delta_3 + m_4,$$

$$X_8 = m_6, \quad X_9 = m_8^\dagger,$$

$$Y(0) = m_1 \Delta_1 - (2s + d - 2N_\tau) m_4^\dagger \Delta_2,$$

$$Y(1) = m_1 \Delta_1 - (2s + d - 2N_\tau) m_2 \Delta_2 + m_4 \Delta_1$$

$$- 2m_4^\dagger \Delta_3 - (2s + d - 2 - 2N_\tau) m_5^\dagger \Delta_3,$$

$$Y(2) = m_1 \Delta_3 + m_2 \Delta_3 + (2s + d - 4 - 2N_\tau) m_3 \Delta_3$$

$$+ m_4 \Delta_3 + (2s + d - 4 - 2N_\tau) m_5 \Delta_2,$$

$$Y(1)' = \frac{\theta}{4} (2s + d - 3 - 2N_\tau)(2s + d - 4 - 2N_\tau).$$

$$X_1 = 0, \quad a = 1, \ldots, 9,$$

$$Y(0)' = 0,$$

$$Y(1)' + Y(1)' = 0,$$

$$Y(2)' = 0.$$

Solution to Eq. (5.18) is easily found to be

$$m_1 = (2s + d - 4 - 2N_\tau) \Delta_2,$$

$$m_4 = -(2s + d - 4 - 2N_\tau) \Delta_3,$$

$$m_2 = -m_1, \quad m_3 = -\frac{1}{2} m_1, \quad m_5 = -\frac{1}{2} m_4,$$

$$m_4^\dagger = \Delta_1, \quad m_5^\dagger = -\frac{1}{2} \Delta_1, \quad m_6 = m_6^\dagger = 0,$$

i.e. Eqs. (5.18) allow us to express the operators $m_a, a = 1, \ldots, 5$, entirely in terms of the operators $\Delta_2, 3$, which enter the gauge transformations. Moreover, the expressions for $m_4$ (5.22) and $m_4^\dagger$ (5.24) imply the relation

$$\Delta_1^\dagger = -(2s + d - 4 - 2N_\tau) \Delta_3.$$
where $\Delta_{2(0)}$ is dimensionful parameter not depending on $N_\zeta$. Inserting $\Delta_2$ in (5.27) in (5.21) one can make sure that (5.21) is satisfied automatically. That all remains then to solve Eq. (5.20). Making use of (5.22)–(5.25) and (5.27) one can make sure that (5.20) amounts to the equation\footnote{It is easy to demonstrate that making use of field redefinitions, the phase factors of $\Delta_3$ can be normalized to be equal to $-1$. Therefore in (5.28) and below $\Delta_3$ is assumed to be real-valued and negative. Relations (5.5) and Eq. (5.25) imply then that $\Delta_3$ is real-valued and positive.} 

$$Z(N_\zeta) - Z(N_\zeta - 1) \equiv \frac{(2s + d - 3 - 2N_\zeta)(\Delta_{2(0)})^2}{(2s + d - 2 - 2N_\zeta)^2(2s + d - 4 - 2N_\zeta)^2} - \frac{\theta}{4}(2s + d - 3 - 2N_\zeta) = 0, \quad (5.28)$$

where we use the notation

$$Z(N_\zeta) \equiv (2s + d - 4 - 2N_\zeta)(N_\zeta + 1)(\Delta_3)^2. \quad (5.29)$$

The relation (5.29) implies a condition $Z(-1) = 0$. This condition and (5.28) lead to the initial condition

$$Z(0) = \frac{(2s + d - 3)(\Delta_{2(0)})^2}{(2s + d - 2)^2(2s + d - 4)^2} + \frac{\theta}{4}(2s + d - 3). \quad (5.30)$$

Eq. (5.28) and the initial condition (5.30) allow us to find $Z(N_\zeta)$ uniquely and taking into account (5.29) we obtain

$$\tilde{\Delta}_3 = \frac{2s + d - 3 - N_\zeta}{2s + d - 4 - 2N_\zeta} \times \left( \frac{(\Delta_{2(0)})^2}{(2s + d - 2)^2(2s + d - 4 - 2N_\zeta)^2} + \frac{\theta}{4} \right). \quad (5.31)$$

Thus we satisfied all equations imposed on $\mathcal{M}$ and $\Delta$ by the requirement of gauge invariance of the action and the expressions (5.22)–(5.25), (5.27), (5.31) determine $\mathcal{M}$ and $\Delta$ uniquely. In view of the first relation in (5.22) and (5.27) the $\Delta_{2(0)}$ is real-valued and introducing the mass parameter $\kappa$ (which is assumed to be positive) by relation

$$\Delta_{2(0)} = (2s + d - 2)\kappa, \quad (5.32)$$

we arrive at the expressions for $\mathcal{M}$ and $\Delta$ given in Section 2. To summarize, we found the gauge invariant action for the fermionic fields in (A)dS$_d$. All that remains to construct action for fermionic fields in AdS$_5 \times$ S$^5$ Ramond–Ramond background is to add appropriate dependence of S$^5$-coordinates and take into account contribution of Ramond–Ramond background fields.\footnote{Study of some leading contributions of Ramond–Ramond background fields to mass operator of the bosonic fields in AdS$_5 \times$ S$^5$ Ramond–Ramond background may be found in [37]. Precise form of mass operator for bosonic fields is still to understood.} The result will be reported elsewhere.

Acknowledgements

This work was supported by the INTAS project 03-51-6346, by the RFBR Grant No. 05-02-17654, RFBR Grant for Leading

Scientific Schools, Grant No. 1578-2003-2 and Russian Science Support Foundation.

Appendix A. Notation and commutators of oscillators and covariant derivative

We use $2^{[d/2]} \times 2^{[d/2]}$ Dirac gamma matrices $\gamma^A$ in $d$-dimensions, $(\gamma^A, \gamma^B) = 2\eta^{AB}$, $\gamma^A \gamma^B = \gamma^0 \gamma^A \gamma^0$, where $\eta^{AB}$ is mostly positive flat metric tensor and flat vectors indices of the space $(d - 1, 1)$ algebra take the values $A, B = 0, 1, \ldots, d - 1$. To simplify our expressions we drop $\eta_{AB}$ in scalar products, i.e. we use $X^A Y^B \equiv \eta_{AB} X^A Y^B$. Indices $\mu, \nu = 0, 1, \ldots, d - 1$ stand for indices of space–time base manifold.

We use the algebra of commutators for operators that can be constructed out the oscillators $a^A$, $\bar{a}^A$ (2.2) and derivative $D^A$ (2.15), (2.16) (see also Appendix A in Ref. [4]). Starting with

$$\{ \hat{\partial}_A, \hat{\partial}_B \} = \Omega_{ABC} \delta_C, \quad \Omega_{ABC} = -\omega_{ABC} + \omega^{BCA},$$

$$\omega_{ABC} \equiv \epsilon_{\mu \nu} \omega^{A, \mu \nu \alpha}_{\alpha \beta \gamma}, \quad (A.1)$$

where $\omega_{ABC} \equiv \epsilon_{\mu \nu} \partial_{\mu} a^A \partial_{\nu} a^B$, $\omega_{\alpha A, \beta B, \gamma C}$ is a torsion tensor we get the basic commutator

$$[D^A, D^B] = \Omega_{ABC} D^C + \frac{1}{2} R_{ABCD} M^{CD}, \quad (A.2)$$

and $R_{ABCD}$ is a Riemann tensor which for $(A)dS_d$ geometry takes the form

$$R_{ABCD} = \theta (\eta^{AC} \eta^{BD} - \eta^{AD} \eta^{BC}), \quad (A.3)$$

The spin operator $M_{AB}$ is given in (2.17). For flexibility in (A.2) and below we present our relations for a space of arbitrary geometry and for $(A)dS_d$ space. Using (A.2) and the commutators

$$[D^A, a^B] = -\omega_{ABC} a^C, \quad [D^A, \bar{a}^B] = -\omega_{ABC} \bar{a}^C, \quad (A.4)$$

we find straightforwardly

$$[D^A, a^2] = 0, \quad [\bar{a}^2, D^A] = 0, \quad [D^A, \gamma a] = 0, \quad (A.5)$$

$$[\bar{a}^2, \alpha D] = 2 \bar{a} D, \quad [\gamma \bar{a}, \alpha D] = \psi, \quad (A.6)$$

$$[\bar{\psi}, \gamma a] = 2 \alpha D, \quad (A.6)$$

$$\bar{\psi}^2 = D^A D^A + \omega_{AAB} D^B + \frac{1}{4} \gamma^{A, B} R_{ABCD} M^{CD} = D^A D^A + \omega_{AAB} D^B + \frac{\theta}{2} \gamma^{A, B} M^{AB}, \quad (A.7)$$

$$[\bar{a} D, \alpha D] = D^A D^A + \omega_{AAB} D^B - \frac{1}{4} R_{ABCD} M^{AB} M^{CD} = D^A D^A + \omega_{AAB} D^B - \frac{\theta}{2} M^{AB} M^{CD}, \quad (A.8)$$

$$[\bar{\psi}, \alpha D] = \frac{1}{2} \gamma^A a^B R_{ABCD} M^{CD} = \theta (\gamma a (\alpha \bar{a} + \frac{d - 1}{2}) - \alpha^2 \gamma \bar{a}), \quad (A.9)$$
\[ [\bar{\alpha} D, \bar{\phi}] = \frac{1}{2} \gamma A - B R^{ABCD} M_{CD} \]
\[
\theta \left( (\bar{\alpha} \bar{\alpha} + \frac{d - 1}{2}) \gamma \bar{\alpha} - \gamma \alpha \bar{\alpha}^2 \right). \tag{A.10}
\]

References

Spinning gravitating skyrmions

Theodora Ioannidou a, Burkhard Kleihaus b,*, Jutta Kunz b

a Mathematics Division, School of Technology, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece
b Institut für Physik, Universität Oldenburg, Postfach 2503, D-26111 Oldenburg, Germany

Received 7 September 2006; accepted 31 October 2006
Available online 10 November 2006
Editor: N. Glover

© 2006 Published by Elsevier B.V.

Abstract
We investigate self-gravitating rotating solutions in the Einstein–Skyrme theory. These solutions are globally regular and asymptotically flat. We present a new kind of solutions with zero baryon number, which possess neither a flat limit nor a static limit.

1. Introduction

In non-Abelian field theories coupled to gravity particle-like solutions as well as black holes arise [1]. The latter are of importance as counterexamples to the no-hair conjecture. In recent years, in particular, various stationary rotating non-Abelian black holes have been studied [2]. However, the construction of stationary rotating particle-like solutions is still a difficult task. Although existence of such solutions in non-Abelian gauge field theories has been restricted [3–5], they can exist in the topologically trivial sector [6,7]. On the other hand, stationary rotating soliton solutions in flat space have been obtained in the Skyrme model [8] and the $U(1)$ gauged Skyrme model [9] in the nontrivial sector.

The Skyrme model is a nonlinear chiral field theory in which baryons and nuclei are described in terms of solitons (so-called skyrmions). Due to the long-standing difficulties in finding a satisfactory theoretical model for the interaction of baryons, much effort has been devoted to the study of classical and quantized interactions of skyrmions. However, the quantization of the Skyrme model is not only difficult since it is a non-renormalizable field theory; but also spinning skyrmions must be considered which means that the solutions must consist of massive pions. This follows from the fact that the skyrmion can only spin at a frequency up to the pion mass before it begins to radiate pions [10]. Recently in [8], it was shown numerically that a good description of protons and neutrons can be achieved with spinning skyrmions, provided the pion mass is chosen twice the experimental value.

The rotating skyrmion solutions of [8] are expected to persist, when the coupling to gravity is turned on gradually, analogous to the static skyrmion solutions [11]. In the static limit, a branch of gravitating skyrmions emerges from the flat space skyrmion, when the coupling to gravity is increased from zero [11]. This branch terminates at a maximal value of the coupling parameter, when the coupling to gravity becomes too large for solutions to persist. A second branch of solutions exists which merges with the first one at the maximal value of the coupling parameter and extends back to zero. The solutions on the second branch possess a larger mass and they are unstable [12]. In the limit of vanishing coupling the solutions shrink to zero size and their mass diverges. As shown in [13], in this limit the skyrmion solutions approach the lowest mass Bartnik–McKinnon (BM) solution of the $SU(2)$ Einstein–Yang–Mills theory [14].

In this Letter we investigate the stationary rotating generalization of the static gravitating skyrmions. We show that in contrast to the static case, additional branches of solutions arise, which are not related to the flat space skyrmion or to the BM solution. Most interestingly, we find a new kind of solution with zero baryon number, which exists for arbitrary (finite) coupling.
parameter, but does not possess a flat space limit. In particular, Section 2 presents the Einstein–Skyrme Lagrangian and the ansatz for the Skyrme field and the metric, which lead to stationary rotating skyrmions. In Section 3 the numerical solutions are discussed, while the conclusions are given in Section 4.

2. Einstein–Skyrme theory

The $SU(2)$ Einstein–Skyrme Lagrangian reads

$$\mathcal{L} = \frac{R}{16\pi G} + \frac{\kappa^2}{4} \text{Tr}(K_\mu K^\mu) + \frac{1}{32\pi^2} \text{Tr}([K_\mu, K_\nu] [K^\mu, K^\nu]) + \frac{m_\pi^2}{2} \text{Tr}\left(\frac{U + U^\dagger}{2} - 1\right), \quad (1)$$

and its action is given by

$$S = \int \mathcal{L} \sqrt{-g} d^4x. \quad (2)$$

Here $R$ is the curvature scalar, $G$ is the Newton constant, $\kappa$ and $e$ are the Skyrme model coupling constants, $m_\pi$ is the pion mass, and $g$ corresponds to the determinant of the metric. The $SU(2)$ Skyrme field $U$ enters via $K_\mu = \partial_\mu U U^{-1}$.

Variation of (2) with respect to the metric $g^\mu\nu$ leads to the Einstein equations

$$G^\mu\nu = R^\mu\nu - \frac{1}{2} g^\mu\nu R = 8\pi G T^\mu\nu, \quad (3)$$

where the stress–energy tensor is given by

$$T^\mu\nu = -\frac{\kappa^2}{2} \text{Tr}(K_\mu K_\nu - \frac{1}{2} g_{\mu\nu} K^a K^a) - \frac{1}{8\pi^2} \text{Tr}\left(g^{a\beta}[K_\mu, K_\alpha][K_\nu, K_\beta]\right) - \frac{1}{4} g_{\mu\nu} [K_\alpha, K_\beta] \left[K^\alpha, K^\beta\right] + g_{\mu\nu} m_\pi^2 \text{Tr}\left(\frac{U + U^\dagger}{2} - 1\right). \quad (4)$$

For stationary rotating solutions, two commuting Killing vector fields are imposed on the space–time: $\xi = \partial_t$ and $\eta = \partial_\phi$, in a system of adapted coordinates $(t, r, \theta, \phi)$. In these coordinates the metric can be expressed in Lewis–Papapetrou form

$$ds^2 = -f dt^2 + \frac{m}{f} (dr^2 + r^2 d\theta^2) + lr^2 \sin^2 \theta (d\phi - \frac{\omega}{r} dt)^2, \quad (5)$$

where $f$, $m$, $l$ and $\omega$ are functions of $r$ and $\theta$ only.

Then, the total mass and angular momentum are defined by

$$M = \frac{1}{4\pi} \int_\Sigma R_{\mu\nu} k^\mu \xi^\nu dV, \quad (6)$$

$$J = -\frac{1}{8\pi G} \int_\Sigma R_{\mu\nu} k^\mu \eta^\nu dV, \quad (6)$$

respectively. Here $\Sigma$ denotes an asymptotically flat hyper-surface, $dV$ is the natural volume element on $\Sigma$, $k^\mu$ is normal to $\Sigma$ and $k_\mu k^\mu = -1$.

In order for finite energy configurations to exist the Skyrme field must tend to a constant matrix at spatial infinity: $U \to I$ as $r \to \infty$. This effectively compactifies the three-dimensional space into $S^1$ and implies that the Skyrme fields can be considered as maps from $S^1$ into $SU(2)$. As the third homotopy class of $SU(N)$ is $Z$, every field configuration is characterized by a topologically invariant integer $B$, which can be obtained as

$$B = \int_\Sigma B^\mu k_\mu dV, \quad (7)$$

where $B^\mu$ is the topological current

$$B^\mu = \frac{1}{\sqrt{-g}} \frac{1}{24\pi^2} \varepsilon^{\mu\nu\alpha\beta} \text{Tr}(K_\alpha K_\beta K_\rho). \quad (8)$$

This winding number classifies the solitonic sectors in the model and may be identified with the baryon number of the field configuration.

For spinning skyrmions the ansatz is of the form $\bar{1}$:

$$U = n_1 I + in_2 (\tau_x \cos(\omega_s t) + \tau_y \sin(\omega_s t)), \quad (9)$$

where $\tau_x$, $\tau_y$, $\tau_z$ are the Pauli matrices; $n_i$ are functions of $r$ and $\theta$ only, satisfying the constraint $C := (1 - \sum n_i^2) = 0$, and the constant $\omega_s$ corresponds to the spinning frequency of the skyrmions.

When deriving the partial differential equations (PDEs) for the skyrmion functions $n_i$, we have to take into account the constraint $C = 0$. This can be achieved by adding the constraint multiplied by some constant, say $c_0$, to the Lagrangian and deriving the variational equations:

$$E_i = \frac{\partial}{\partial (\partial_i n_i)} \left( \frac{\partial \mathcal{L}}{\partial (\partial_i n_i)} \right) - \frac{\partial \mathcal{L}}{\partial n_i} + 2c_0 n_i \sqrt{-g} = 0. \quad (10)$$

Then, the constant $c_0$ can be obtained from the linear superposition $\sum n_i E_i = 0$, and substituted back in the PDEs of (10).

3. Numerical solutions

3.1. Parameters and boundary conditions

Introducing the dimensionless radial coordinate $x = \kappa er$, the gravitational coupling parameter $a^2 = 4\pi G \kappa^2$, the spinning frequency $\hat{\omega}_s = \omega_s / \kappa e$, and the pion mass $\bar{m}_\pi = m_\pi / \sqrt{\kappa / e}$, action (2) becomes

$$S = \frac{\kappa}{e} \int \left[\frac{R}{4a^2} + \frac{1}{4} \text{Tr}(K_\mu K^\mu) + \frac{1}{32} \text{Tr}([K_\mu, K_\nu] [K^\mu, K^\nu]) + \frac{\bar{m}_\pi^2}{2} \text{Tr}\left(\frac{U + U^\dagger}{2} - 1\right)\right] \sqrt{-g} d^4x. \quad (11)$$

$\bar{1}$ Strictly speaking, the ansatz is neither stationary nor axially symmetric, since it depends explicitly on time and the azimuthal angle. However, the stress–energy tensor does possess the corresponding symmetries. See also Ref. [4].
while the Einstein equations read: \( G_{\mu \nu} = 2\alpha^2 T_{\mu \nu} \). We also introduce the dimensionless mass \( M = M/[4\pi \kappa/e] \) and angular momentum \( J = J/[4\pi/e^2] \). That way, the solutions depend only on the parameters \( \alpha, \bar{\omega}_s, \) and \( \tilde{m}_\alpha \). For convenience we will rename \( \bar{\omega}_s \rightarrow \omega_s \).

At the origin, the boundary conditions are

\[
\begin{align*}
n_1(0) &= -1, & n_2(0) &= n_3(0) = 0, \\
\partial_x f|_0 &= 0, & \partial_x m|_0 &= 0, & \omega(0) &= 0, \\
\end{align*}
\]  

while for large \( x \), since the asymptotic value of the Skyrme field is the unit matrix and of the metric is the Minkowski metric, we get

\[
\begin{align*}
n_1(\infty) &\rightarrow 1, & n_2(\infty)|_{l=2,3} &\rightarrow 0, & f(\infty) &\rightarrow 1, \\
l(\infty) &\rightarrow 1, & m(\infty) &\rightarrow 1, & \omega(\infty) &\rightarrow 0. \\
\end{align*}
\]  

On the \( z \)-axis (\( \theta = 0 \)) the boundary conditions follow from regularity

\[
\begin{align*}
\partial_\theta n_1|_{\theta=0} &= 0, & n_2(\theta = 0) &= 0, & \partial_\theta n_3|_{\theta=0} &= 0, \\
\partial_\theta f|_{\theta=0} &= 0, & \partial_\theta l|_{\theta=0} &= 0, & \partial_\theta m|_{\theta=0} &= 0, \\
\partial_\theta \omega|_{\theta=0} &= 0, \\
\end{align*}
\]  

while in the \( xy \)-plane (\( \theta = \pi/2 \)) from reflection symmetry

\[
\begin{align*}
\partial_\theta n_1|_{\theta=\pi/2} &= 0, & n_2(\theta = \pi/2) &= 0, & n_3(\theta = \pi/2) &= 0, \\
\partial_\theta f|_{\theta=\pi/2} &= 0, & \partial_\theta l|_{\theta=\pi/2} &= 0, \\
\partial_\theta m|_{\theta=\pi/2} &= 0, & \partial_\theta \omega|_{\theta=\pi/2} &= 0. \\
\end{align*}
\]  

In what follows we will encounter two special cases, the Bartnik–McKinnon solution and the nontrivial solutions in the vacuum sector. The first, is obtained after rescaling \( x = \alpha \bar{x} \) and taking the limit of vanishing \( \alpha \). In this limit, the solutions are equivalent to the Bartnik–McKinnon one with lowest mass:

\[
\begin{align*}
n_1 &= -w(\bar{x}), & n_2 &= \sqrt{1 - n_1^2} \sin \theta, \\
n_3 &= \sqrt{1 - n_1^2} \cos \theta, & l &= m, & \omega &= 0, \\
\end{align*}
\]  

where the gauge potential of the \( SU(2) \) Einstein–Yang–Mills theory is parametrized from \( w(\bar{x}) \) via the relation \( A^a_i = (1 - w(\bar{x}))\epsilon_{iab} \bar{x}_j/(2\bar{x}^2) \).

The second is obtained by setting

\[
\begin{align*}
n_1 &= \cos(h), & n_2 &= \sin(h), & n_3 &= 0, \\
\end{align*}
\]  

where the function \( h \) depends on \( x \) and \( \theta \). Regularity and finite energy of the solutions require that \( h \) vanishes on the \( z \)-axis and at infinity. We will refer to these solutions as ‘pion cloud’.

### 3.2. Numerical results

The solutions are constructed using the software package CADSOL [15] based on the Newton–Raphson algorithm. In order to map the infinite range of the radial variable \( x \) to the finite interval \([0, 1]\) we introduce the compactified radial variable \( \bar{x} = x/(1 + x) \). Typical grids contain 70 × 50 points. The estimated relative errors are approximately \( \approx 0.1\% \), except close to \( \alpha_{\text{max}} \) where they become as large as \( 1\% \).

In particular, gravitating skyrmions are constructed and their dependence on the coupling parameter \( \alpha \) and the spinning frequency \( \omega_s \) are studied for fixed pion mass: \( \tilde{m}_\alpha = 1 \). This sets a limit to the range of the spinning frequency \( \omega_s \leq 1 \). We start the discussion by a qualitative description of the dependence of the solutions on the gravitational parameter \( \alpha \) for fixed spinning frequency \( \omega_s \). Different branches of solutions exist which are characterized by their limit as \( \alpha \) tends to zero. First, there are branches of solutions which tend to the flat space skyrmions and to the scaled BM solution which we call (for obvious reasons) ‘skyrmion’ and ‘BM’ branches, respectively.

Second, branches of solutions exist which form a ‘pion cloud’ for large \( x \) and the way different branches merge depends on the value of \( \omega_s \) relative to critical values \( \omega_{s}^\pm \approx 0.9607 \). So, for \( \omega_s < \omega_{s}^\pm \) the ‘skyrmion’ branches merge with the ‘BM’ ones; however for \( \omega_s > \omega_{s}^\pm \), the ‘skyrmion’ and the ‘BM’ branches merge with the ‘cloudy skyrmion’ and the ‘cloudy BM’ ones, respectively. Moreover, the ‘cloudy skyrmion’ branches merge with the ‘cloudy BM’ branches only when \( \omega_s < \omega_{s}^\pm \).

Next a quantitative description in terms of the dimensionless mass \( M \) and the value of the function \( l_0 = l(0) \) is presented. Since with vanishing \( \alpha \) the mass diverges on the ‘BM’, the ‘cloudy skyrmions’ and the ‘cloudy BM’ branches we also consider the scaled masses \( M \alpha \) and \( M \alpha^3 \).

Fig. 1(a) presents the mass \( M \) for the ‘skyrmion’ branches (solid) which merge either with the ‘BM’ branches (dashed) or the ‘cloudy skyrmion’ branches (dotted) when \( \omega_s < \omega_{s}^\pm \) and \( \omega_s > \omega_{s}^\pm \), respectively. Note that, as \( \alpha \) increases the mass decreases on the ‘skyrmion’ branches, for small \( \alpha \); but diverges on the ‘BM’ and the ‘cloudy skyrmion’ branches as \( \alpha \) tends to zero. Also Fig. 1(a) shows the mass of the ‘BM’ branches merging with the ‘cloudy BM’ branches (dash-dotted) when \( \omega_s > \omega_{s}^\pm \), and the mass of the ‘cloudy skyrmion’ branches merging with the ‘cloudy BM’ branches when \( \omega_s < \omega_{s}^\pm \).

The scaled mass \( M \alpha \) is plotted in Fig. 1(b). Note that, as \( \alpha \) decreases along the ‘BM’ branches the scaled mass \( M \alpha \) tends to a finite value which is equal to the mass of the Bartnik–McKinnon solution. Also, Fig. 1(b) shows the scaled mass \( M \alpha^3 \) of the ‘BM’ branches as it merges with the ‘cloudy BM’ branches, and of the ‘cloudy skyrmion’ branches as it merges with the ‘cloudy BM’ branches. Clearly \( M \alpha \) diverges on the ‘cloudy skyrmion’ and the ‘cloudy BM’ branches for vanishing \( \alpha \).

Fig. 1(c) reveals that in the limit \( \alpha \rightarrow 0 \) the scaled mass \( M \alpha^3 \) of the ‘cloudy skyrmion’ and the ‘cloudy BM’ branch tends to a unique value which depends only on \( \omega_s \).
Fig. 1. The dimensionless mass $M$ (a), the scaled masses $M\alpha$ (b) and $M\alpha^3$ (c), and the value of $l$ at the origin (d) as function of $\alpha^2$ for several values of $\omega_s$.

Next we study the quantity $l_0$ of Fig. 1(d) in order to have a better understanding. Following a solution along a ‘skyrmion’ branch which merges with a ‘BM’ branch, $l_0$ decreases monotonically first along the ‘skyrmion’ branch as $\alpha$ increases and then along the ‘BM’ branch as $\alpha$ decreases, to take the value of the Bartnik–McKinnon solution as $\alpha \to 0$. In contrast, when a ‘skyrmion’ branch merges with a ‘cloudy skyrmion’ branch, $l_0$ reaches a minimum on the ‘skyrmion’ branch and increases on the ‘cloudy skyrmion’ branch as $\alpha$ decreases. On the other hand, when the ‘BM’ branch merges with a ‘cloudy BM’ branch, $l_0$ increases with increasing $\alpha$ along the ‘BM’ branch until it reaches a maximum and decreases with decreasing $\alpha$ along the ‘cloudy BM’ branch. Finally, when a ‘cloudy skyrmion’ branch merges with a ‘cloudy BM’ branch, $l_0$ decreases monotonically if one follows the solutions first on the ‘cloudy skyrmion’ branch with decreasing $\alpha$ and then on the ‘cloudy BM’ branch with increasing $\alpha$.

Also, while we observe that the scaled mass $M\alpha^3$ of the ‘cloudy skyrmion’ and the ‘cloudy BM’ branches tends to the same value as $\alpha$ tends to zero—this is not true for the quantity $l_0$. Thus, we conclude that the solutions approach different limits, though with the same (scaled) mass.

In the following we compare the mass and the angular momentum. We here restrict to $\omega_s = 0.95 < \omega_{s}^{ct}$, as examples. In particular, Fig. 2(a) shows that when $\omega_s = 0.95 < \omega_{s}^{ct}$ the angular momentum decreases monotonically on the ‘skyrmion’ branch as $\alpha$ increases and on the ‘BM’ branch as $\alpha$ decreases, while it tends to zero on the ‘BM’ branch as $\alpha \to 0$. In contrast, for $\omega_s = 0.9608 > \omega_{s}^{ct}$ the angular momentum increases monotonically on the ‘skyrmion’ branch as $\alpha$ increases and on the ‘cloudy skyrmion’ branch as $\alpha$ decreases, while in the limit $\alpha \to 0$ the angular momentum diverges like $\alpha^3$. Fig. 2(b) presents the scaled mass $M\alpha^3$ and angular momentum $J\alpha^2$ for the ‘BM’ and ‘cloudy BM’ branches when $\omega_s = 0.9608$ and for the ‘cloudy skyrmion’ and ‘cloudy BM’ branches when $\omega_s = 0.95$. As $\alpha \to 0$ the (scaled) angular momentum tends to zero on the ‘BM’ branch, but takes finite values on the ‘cloudy’ branches.

The new interesting feature of the rotating gravitating skyrmions is the formation of the ‘pion cloud’. This can be demonstrated by plotting the skyrmion functions $n_2$ and $n_3$ when $\alpha = 0.1$ for $\omega_s = 0.95 < \omega_{s}^{ct}$ and $\omega_s = 0.9608 > \omega_{s}^{ct}$ as presented in Figs. 3 and 4 (respectively).

In particular, Figs. 3(a) and (b) show that on the ‘skyrmion’ branch the solution is close to the flat space skyrmion whereas on the ‘BM’ branch it is close to the (scaled) Bartnik–McKinnon solution. In contrast, the functions $n_2$ of the solutions on the ‘cloudy skyrmion’ and ‘cloudy BM’ branch (pre-
Fig. 2. The dimensionless mass $M$ and angular momentum $J$ (a), and the scaled mass $M\alpha^3$ and angular momentum $J\alpha^3$ (b) as function of $\alpha^2$ for $\omega_s = 0.95$ and $\omega_t = 0.9608$.

Fig. 3. The functions $n_2$ (left) and $n_3$ (right) are plotted for $\theta = 0, \pi/4, \pi/2$ for the four solutions with parameter values $\omega_s = 0.95$ and $\alpha = 0.1$. Presented in Fig. 3(c)) almost coincide at large $x$ where the 'pion cloud' forms. However, for small $x$ the functions $n_2$ are similar in shape to the corresponding ones of Fig. 3(a). Finally, by comparing the functions $n_3$ of the ‘skyrmion’ and ‘BM’ solutions (of Fig. 3(b)) with their ‘cloudy’ counterparts (of Fig. 3(d)) we observe that they are similar in shape for all $x$—which means that the ‘pion cloud’ does not reflect itself in the function $n_3$.

Fig. 4 presents the solutions on connected branches when $\omega_s > \omega_t^2$. Note that, the function $n_2$ of both the ‘skyrmion’ and the ‘cloudy skyrmion’ branch (plotted in Fig. 4(a)) almost coincide for small values of $x$ and differ for larger $x$, where the ‘pion cloud’ is apparent. In contrast, the functions $n_3$ are almost identical for both solutions as shown in Fig. 4(b). Similar observations hold for the solutions of the ‘BM’ branch and the ‘cloudy BM’ branch plotted in Fig. 4(c)–(d).

On the ‘cloudy BM’ branch the scaled BM solution in the core separates from the surrounding ‘pion cloud’ as $\alpha$ decreases. Therefore, in what follows, we show that pure ‘pion cloud’ solutions can be constructed numerically by extracting the data of the ‘pion cloud’. The ansatz for the ‘pion cloud’ so-
Fig. 4. Analogous as Fig. 3 for $\omega_s = 0.9608$.

For these solutions $n_3 = 0$ so the chiral matrix can be regarded as a map from $S^3 \to S^2$ and thus, the 'pion cloud' solutions have zero baryon number. In addition, due to the boundary conditions of the profile function $h(x, \theta)$, the 'pion cloud' solutions can be deformed continuously to the vacuum.

Fig. 5 presents the scaled mass $M\alpha^3$ of the 'pion cloud' solutions and of the skyrmions on the 'cloudy' branches as function of $\alpha^2$ for fixed $\omega_s = 0.95$. Note that, for small $\alpha$ the masses of the 'cloudy skyrmion' and 'cloudy BM' solutions coincide with the mass of the 'pion cloud' solution. However, whereas the branches of the skyrmion solutions exist only up to a maximal value of $\alpha$, the 'pion cloud' solutions exist for arbitrarily large $\alpha$.

Indeed, as $\alpha$ increases the magnitude of $h$ decreases linearly like $1/\alpha$. So by expanding the Lagrangian up to quadratic order in $h$ yields

\[ L_h = \frac{R}{2\alpha^2} - \frac{1}{2} \left[ \partial_\mu h \partial^\mu h + \frac{1}{x^2} \partial_\mu h \partial^\theta h + \frac{f}{lx^2 \sin^2 \theta} h^2 \right] - \frac{1}{f} \left( \omega_3 - \frac{\omega}{x} \right)^2 h^2 + \hat{m}_3^2 h^2, \]

which is equivalent to the Lagrangian of the rotating boson star

\[ L_{BS} = \frac{R}{2\alpha^2} - \frac{1}{4} g^{\mu\nu} (\partial_\mu \Phi^* \partial_\nu \Phi + \partial_\mu \Phi \partial_\nu \Phi^*) - V(|\Phi|), \]

for $\Phi = h(x, \theta) e^{i(\varphi - \omega_s t)}$ and $V(|\Phi|) = |\Phi|^2 m_3^2 / 2$. The limit $\alpha \to \infty$ of rotating boson stars has been studied in Ref. [16] (though with different notation) where it was shown that the
field equations become independent of the coupling parameter $\alpha$ after re-scaling $h = \hat{h}/\alpha$. Moreover, it was argued in [16] that several branches of solutions exist in certain ranges of $\omega_s$ which suggests that several branches of ‘pion cloud’ solutions might (also) exist, at least for large values of $\alpha$.

In the limit $\alpha \to 0$ the scaled mass $M\alpha^3$ takes finite values. By introducing the scaled radial coordinate $\xi = \alpha x$, the Skyrme field equations reduce to the constraint

$$\sin^2(h)[\cos(h)\omega_s^2 - f m^2] = 0, \quad (22)$$

as $\alpha \to 0$. This implies that the Skyrme field function is either $h = 0$ or $\cos(h) = f m^2/\omega_s^2$. Note that none of these solutions can hold globally; the former one yields the trivial solution and the latter is not consistent with the asymptotic boundary conditions $h \to 0$ and $f \to 1$ for $\omega_s^2 < m^2$. However, both can hold locally. Indeed, we find the solution

$$\cos(h) = f m^2 / \omega_s^2, \quad [\xi, \theta] \in D = (0, \xi_0] \times (0, \pi),$$

$$h = 0, \quad \text{elsewhere,}$$

where $\xi_0$ depends on $\omega_s$. Although this solution for $h$ is not continuous at the origin and on the $z$-axis, the metric functions are continuous globally. Moreover, in the domain $D$ the function $h$ depends only on the radial coordinate $\xi$. Consequently, the metric is spherically symmetric and, for $\xi \geq \xi_0$, given by the Schwarzschild solution.

Thus, in the limit $\alpha \to 0$ the ‘pion cloud’ becomes confined to the finite domain $D$. Outside of $D$ the metric is the vacuum solution determined by the connecting conditions at $\xi_0$. To see the physical picture, however, we have to return to unscaled coordinates. Then the domain $D$ extends over the whole space, except the $z$-axis and infinity. Consequently, the ‘pion cloud’ occupies an increasing volume in space as $\alpha$ decreases.

The profile function $h$ is plotted in Fig. 6(a) as function of the cylindrical coordinates $\rho = x \sin \theta, z = x \cos \theta$ for $\omega_s = 0.95$ and $\alpha = 0.07$ and in Fig. 6(b) as function of the scaled coordinates $\rho' = \xi \sin \theta, z' = \xi \cos \theta$ for such parameter values that the constraint is almost satisfied (i.e. $\omega_s = 0.95$ and $\alpha = 10^{-4}$).

Finally we state that the ‘pion cloud’ solutions do not exist for arbitrarily small $\omega_s$. In fact, we observed that the coefficient of the second order derivative term in the Skyrme field equation

$$\left[ f^2 \sin^2(h) - \left( \left( \omega_s - \frac{\omega}{r} \right)^2 \sin^2(h) - f \right) r^2 \sin^2 \theta \right]$$

$$\times \left( \partial_{\rho'}^2 h + \frac{1}{r^2} \partial_{\theta'}^2 h \right) + \cdots = 0 \quad (23)$$

develops a zero at some point on the $\theta = \pi/2$ axis, when $\omega_s$ decreases to a critical value, and no solution exists for $\omega_s$ below the critical value. Therefore, the ‘pion cloud’ solutions do not possess a static limit.

4. Conclusions

We studied stationary rotating solutions of the Einstein–Skyrme theory. These solutions are asymptotically flat, globally regular and axially symmetric. Branches of stationary rotating skyrmions emerge from corresponding branches of static skyrmions, when the rotational frequency is increased form zero. If the rotational parameter $\omega_r$ is smaller than a critical value, the rotating skyrmions on these branches behave similar to their static counterparts, when the coupling to gravity is varied. However, additional branches of solutions exist, which do not have a flat space limit. These branches are characterized by the formation of a ‘pion cloud’ for small values of the coupling $\alpha$.

In addition to the rotating skyrmions with baryon number one, we found new solutions in the topologically trivial sector. These ‘pion cloud’ solutions are also asymptotically flat and globally regular, but possess neither a flat limit nor a static limit. In contrast to the skyrmions the ‘pion cloud’ solutions exist for arbitrary coupling, i.e. $0 < \alpha < \infty$.

Axially symmetric rotating skyrmions with higher baryon number $B > 1$ should easily be obtained by the replacement $\varphi \rightarrow B\varphi$ in the ansatz Eq. (9).

Einstein–Skyrme theory also possesses black hole solutions. So far the skyrmion black holes have been studied in the static limit only. Families of black hole solutions emerge from the globally static regular skyrmions, when the horizon radius is increased from zero. Similarly, we expect several families of stationary rotating skyrmion black holes to emerge from the globally regular rotating skyrmions on the different branches obtained here [17]. Moreover, one may speculate that also sta-
tionary rotating black holes emerge from the ‘pion cloud’ solutions.

Acknowledgements

B.K. gratefully acknowledges support by the DFG under contract KU612/9-1 and T.I. thanks Oldenburg University for its hospitality.

References

    H. Luckock, Black hole skyrmions, in: H.J. de Vega, N. Sanchez (Eds.),
    Proceedings of the 1986 Paris–Meudon Colloquium, World Scientific,
    Singapore, 1987;
    M. Schauder, R. Weiß, W. Schönauer, The CADSOL Program Package,
    Universität Karlsruhe, Interner Bericht Nr. 46/92 (1992).
Calogero models and nonlocal conformal transformations

Anton Galajinsky a, Olaf Lechtenfeld b,*, Kirill Polovnikov a

a Laboratory of Mathematical Physics, Tomsk Polytechnic University, 634050 Tomsk, Lenin Ave. 30, Russian Federation
b Institut für Theoretische Physik, Leibniz Universität Hannover, Appelstrasse 2, D-30167 Hannover, Germany

Received 25 July 2006; accepted 27 October 2006
Available online 9 November 2006
Editor: L. Alvarez-Gaumé

Abstract

We propose a universal method of relating the Calogero model to a set of decoupled particles on the real line, which can be uniformly applied to both the conformal and nonconformal versions as well as to supersymmetric extensions. For conformal models the simplification is achieved at the price of a nonlocal realization of the full conformal symmetry in the Hilbert space of the resulting free theory. As an application, we construct two different \( N = 2 \) superconformal extensions.

© 2006 Elsevier B.V. All rights reserved.

PACS: 04.60.Ds; 11.30.Pb; 12.60.Jv

Keywords: Calogero model; Nonlocal conformal transformations; Supersymmetry

1. Introduction

The range of physical and mathematical applications of the Calogero model is impressive. Being originally formulated as an exactly solvable multi-particle quantum mechanics in one dimension [1], it played an important role in the study of matrix models [2,3], fractional statistics [4], classical and quantum integrable systems [5], the quantum Hall effect [6], superstring theory on the \( \text{AdS}_2 \) background [7], the WDVV equation [8] and BPS operators in \( N = 4 \) SYM theory [9] (for a recent review see [10]).

If one is concerned with only the pairwise interaction \( g^2 \sum_{i<j} (x_i - x_j) \) and disregards the harmonic potential \( \omega^2 \sum_i (x_i^2) \), the Calogero model exhibits conformal symmetry [11]. This property and the fact that the isometry group of \( \text{AdS}_2 \) space is \( \text{SO}(1, 2) \) led the authors of [12] to conjecture that an \( N = 4 \) superconformal extension of the Calogero model might provide a microscopic description of the extreme Reissner–Nordström black hole in the near horizon limit, which corresponds to \( \text{AdS}_2 \times S^2 \) geometry. Unfortunately, a consistent \( N = 4 \) superconformal generalization of the Calogero model has not yet been constructed (for previous attempts see [8,13–15]). The latter problem partially motivated the present investigation.

It has been known since the original work of Calogero [1] that in the presence of harmonic forces the energy eigenvalues of the problem differ from those of decoupled oscillators only by a constant. An explicit but nonunitary similarity transformation connecting their Hamiltonians has been constructed in [16] (see also [17] for a supersymmetric extension).

When the harmonic potential is switched off one expects a similar relation between identical particles interacting via the inverse-square potential and free particles in one dimension to hold. A unitary transformation that maps the Hamiltonian of the Calogero model to that of free particles was constructed in [18]. However, the full conformal symmetry, which characterizes the case at hand, was not taken into account. Note also that the transformation considered in [18] cannot be obtained from that examined in [16] by taking the limit \( \omega \to 0 \). This indicates that the two approaches are essentially different.

* Corresponding author.

E-mail addresses: galajin@mph.phtd.tpu.edu.ru (A. Galajinsky), lechtenf@itp.uni-hannover.de (O. Lechtenfeld), kir@mph.phtd.tpu.edu.ru (K. Polovnikov).

0370-2693/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.physletb.2006.10.062
The purpose of this Letter is to propose a universal method of relating the Calogero model to decoupled particles, which can be uniformly applied to both the conformal and nonconformal versions as well as to supersymmetric extensions. Our approach is different from [18] in that it makes use of all conformal generators when constructing the transformation. In other words, we study the behaviour of the Calogero model under specific (unitary) transformations generated by the conformal algebra so(1, 2). As shown below, although the Hamiltonian $H$ of the Calogero model can indeed be mapped to the free Hamiltonian $H_0$, the generator $K$ of special conformal transformations gets modified and keeps track of the original potential $H_{\text{int}} = H - H_0$ via a nonlocal contribution,

$$K = \frac{1}{2} x^i x^i \rightarrow \tilde{K} = K + \alpha^2 e^{iB} H_{\text{int}} e^{-iB} \quad \text{with} \quad H_{\text{int}} = \sum_{i<j} \frac{g^2}{(x^i - x^j)^2}. \quad (1)$$

Here $\alpha$ is a constant, and the explicit form of the operator $B$ is given below. A similar relation holds for an $N = 2$ superconformal extension of the Calogero model, for which also the superconformal generators are modified appropriately. Thus, after applying a unitary transformation one arrives at free particles in one dimension with the (super)conformal group being realized in a nonstandard (nonlocal) way. Although quantum states look particularly simple in this framework, the action of the full conformal group in the Hilbert space proves to be rather involved.

The organization of the Letter is as follows. In Section 2 we use general properties of the so(1, 2) algebra and construct a novel unitary transformation which maps the conformal Calogero model to a set of free particles on the real line. In Section 3 the method is applied to the nonconformal Calogero model which features an external harmonic potential for each particle. A map to a set of decoupled harmonic oscillators is constructed and shown to be much simpler than the one proposed in [16]. We then proceed to explore supersymmetric generalizations in Section 4. The $N = 2$ superconformal extension of the Calogero model built in [19] is related to a set of free $N = 2$ superparticles, with the SU(1, 1|1) symmetry group being realized in a nonstandard fashion. We argue that the $N = 2$ superconformal extension is not unique. Furthermore, our transformation may pave the way to constructing $N > 2$ superconformal extensions of the Calogero model from a set of free superparticles. We conclude by discussing possible further developments in Section 5.

2. From the Calogero model to free particles

Our starting point is the so(1, 2) algebra realized in the quantized $n$-particle Calogero model via the Weyl-ordered generators

$$H = \frac{1}{2} p_i p_i + \sum_{i<j} \frac{g^2}{(x^i - x^j)^2}, \quad D = -\frac{1}{4} (x^i p_i + p_i x^i), \quad K = \frac{1}{2} x^i x^i, \quad (2)$$

which satisfy

$$[H, D] = iH, \quad [H, K] = 2iD, \quad [D, K] = iK. \quad (3)$$

Here, $g$ is a dimensionless coupling constant ($[x] = [t^{1/2}]$), and the index $i$ labels $n$ identical particles (of unit mass) on the real line mutually interacting via the inverse-square potential. Putting $g = 0$ yields a free-particle representation of so(1, 2), whose generators we denote by $H_0, D$ and $K$.

Each generic Lie-algebra element

$$A = \alpha H + \beta K + \gamma D, \quad (4)$$

where the real constants $\alpha$ and $\beta^{-1}$ have the dimension of length and $\gamma$ is dimensionless, determines a unitary transformation

$$(H, D, K) \quad \rightarrow \quad (H', D', K') = (e^{iA} H e^{-iA}, e^{iA} D e^{-iA}, e^{iA} K e^{-iA}) \quad (5)$$

which is an automorphism of the algebra. It is instructive to use the Baker–Campbell–Hausdorff formula

$$T' = e^{iA} T e^{-iA} = \sum_{n=0}^{\infty} \frac{i^n}{n!} T_n', \quad \text{where} \quad T_0' = T \quad \text{and} \quad T_n' = [A, [A, \ldots [A, T], \ldots]], \quad (6)$$

and calculate the first three terms of the transformed Hamiltonian,

$$H'_0 = H, \quad iH'_1 = 2\beta D + \gamma H, \quad \frac{i^2}{2!} H'_2 = \left( \frac{\gamma^2}{2} - \alpha \beta \right) H + \beta^2 K + \beta \gamma D. \quad (7)$$

Apparently, the particular choice

$$\gamma = \pm 2 \sqrt{\alpha \beta} \quad \text{for} \quad \alpha \beta > 0 \quad (8)$$
produces $\frac{i^2}{2} H'_n = \beta A$ and $H'_{n>2} = 0$, terminating the series in (6) at the third step. In what follows, we always adopt this choice. The condition (8) also terminates the series for the transformed dilatation and special conformal generators, so together we have

$$H' = (1 + \gamma + \alpha \beta) H + \beta (2 + \gamma) D + \beta^2 K = \kappa^2 H + 2\beta \kappa D + \beta^2 K,$$

$$D' = -\alpha \left( 1 + \frac{\gamma}{2} \right) H + \left( 1 - \frac{\gamma}{2} \right) D + \beta \left( 1 - \frac{\gamma}{2} \right) K = -\alpha \kappa H + (1 - 2\alpha \beta) D + \beta \kappa K,$$

$$K' = \alpha^2 H + \alpha (\gamma - 2) D + \left( 1 - \frac{\gamma}{2} \right)^2 K = \alpha^2 H - 2\alpha \kappa D + \kappa^2 K,$$

where we abbreviated

$$\kappa_{\pm} := 1 \pm \sqrt{\alpha \beta} \quad \text{for} \ \alpha \beta > 0.$$  

An important simplification occurs for

$$\alpha \beta = 1 \quad \rightarrow \quad \kappa_+ = 2, \quad \kappa_- = 0$$

and the lower sign choice, $\gamma = -2$, namely

$$H' = \beta^2 K, \quad D' = -D + 2\beta K, \quad K' = 4K - 4\alpha D + \alpha^2 H.$$  

Note that $H$ is mapped to the free-field generator $K \equiv K_0$. For the upper sign choice one gets $K' = \alpha^2 H$ instead.

In our consideration it is only the structure of the conformal algebra which matters. So, by changing the operator $A$ in (4) for

$$B = \lambda H_0 + \sigma K + \delta D,$$

analogous relations hold for a system of free particles with the generators $H_0$, $D$ and $K$. This observation suggests (in this respect see also [18]) that one can compose the transformations generated by $A$ and by $B$ to map

$$H \mapsto K \equiv K_0 \mapsto H_0 \quad \text{via} \quad \alpha \beta = 1, \quad \gamma = -2, \quad \lambda \sigma = 1, \quad \delta = +2.$$  

The second map,

$$(H_0, D, K) \quad \mapsto \quad \left( H''_0, D'', K'' \right) = \left( e^{iB} H_0 e^{-iB}, e^{iB} D e^{-iB}, e^{iB} K e^{-iB} \right),$$

reads

$$K'' = \lambda^2 H_0, \quad D'' = -D - 2\lambda H_0, \quad H''_0 = 4H_0 + 4\sigma D + \sigma^2 K.$$  

A successive application of the two transformations then produces

$$H \mapsto \tilde{H} = H_0, \quad D \mapsto \tilde{D} = D, \quad K \mapsto \tilde{K} = K + \alpha^2 H''_\text{int},$$

provided we impose the further relations

$$\beta \lambda = -1 \quad \Rightarrow \quad \alpha \sigma = -1 \quad \text{and} \quad \alpha + \lambda = 0.$$  

Thus, with the help of the unitary operator $e^{iB} e^{-iA}$ one can transform the Hamiltonian of the Calogero model into that describing a system of free particles.

A few comments are in order. Firstly, a similar transformation of $H$ to $H_0$ has been discussed in [18]. However, the authors of [18] employed (4) with $\gamma = 0$, whence their Baker–Campbell–Hausdorff series did not terminate. As was demonstrated above, our generic choices for $A$ and $B$ allow for a drastic simplification. Secondly, not the entire $\text{so}(1, 2)$ algebra was studied in [18]. According to our analysis, the operator of special conformal transformations gets modified. In fact, it effectively “hides” the interaction potential, which disappears for the Hamiltonian but gives a nonlocal contribution $\alpha^2 H''_\text{int} = 2\gamma B \left( \sum_{i<j} \frac{r^2}{(\xi'_i - \xi'_j)^2} \right) e^{-iB}$ to $K$.

Thirdly, consistency requires the operator $e^{iB} e^{-iA}$ to be independent of the remaining free parameter $\alpha$, as the latter is not fixed by the formalism and has a dimension of length. In order to check this, let us differentiate $e^{iB} e^{-iA}$ with respect to $\alpha$ and demonstrate that

$$\frac{d}{d\alpha} (e^{iB} e^{-iA}) = 0$$

for our special Lie-algebra elements

$$A = \alpha H + \frac{1}{\alpha} K - 2D \quad \text{and} \quad B = -\alpha H_0 - \frac{1}{\alpha} K + 2D.$$  

(20)
Taking into account also the commutation relations (3), which are valid for both \(H\) and \(H_0\), one can easily verify the relations
\[
\left[ \frac{dB}{d\alpha}, A^n \right] = -2in\frac{1}{\alpha}A^n \quad \Rightarrow \quad \frac{de^{ib}}{d\alpha} = i \left( \frac{dB}{d\alpha} + \frac{1}{\alpha}B \right) e^{ib} = 2i \left( \frac{1}{\alpha}D - H_0 \right) e^{ib},
\]
\[
\left[ \frac{dA}{d\alpha}, A^n \right] = 2in\frac{1}{\alpha}A^n \quad \Rightarrow \quad \frac{de^{IA}}{d\alpha} = i \left( \frac{dA}{d\alpha} - \frac{1}{\alpha}A \right) e^{IA} = 2i \left( \frac{1}{\alpha}D - \frac{1}{\alpha^2}K \right) e^{IA}.
\]
Together with (16) they lead to the desired result (19).

To summarize, the quantum mechanical Hamiltonian of the Calogero model can be transformed into a free Hamiltonian by applying an appropriate unitary transformation. Knowing its explicit form, the stationary states of the former model can be immediately constructed from those of the latter. This is in agreement with the claim of [4] that the quantum Calogero model hiddenly describes free particles in one dimension. It should be remembered, however, that the price paid for this change of variables is a nonlocal realization of the full conformal algebra in the Hilbert space.

3. Adding the harmonic potential

Let us now add an external harmonic potential to the model. The analysis of the previous section makes it clear that our technique can still be applied. Such a treatment of the Calogero model in the presence of a harmonic force should be much less intricate than the computation of [16], whose similarity transformation to decoupled harmonic oscillators explicitly involves the correlated ground state of the Calogero model.

Consider then the Hamiltonian
\[
H_1 = \frac{1}{2} p_i p_i + \sum_{i<j} \frac{g^2}{(x^i - x^j)^2} + \frac{\omega^2}{2} x^i x^i = H + \omega^2 K.
\]
Application of the first transformation with \(A\) as in (9) for the lower sign choice in (8) yields
\[
H'_1 = (\kappa_- + \alpha^2 \omega^2)H + (2\beta \kappa_- - 2\alpha \kappa_+ \omega^2)D + (\beta^2 + \kappa_+^2 \omega^2)K.
\]
It is clear that the first term on the r.h.s. can no longer vanish for a real value of \(\kappa_\pm = 1 \pm \sqrt{\alpha \beta}\). Hence, we must allow \(\alpha\) and/or \(\beta\) to become complex in
\[
\kappa_- = i\alpha \omega \quad \Rightarrow \quad \alpha \beta = (1 - i\alpha \omega)^2,
\]
where \(\alpha\) remains arbitrary. This means that, as in [16], an ultimate similarity transformation is realized by a nonunitary operator. With the above relations replacing (11), the transformation specializes to
\[
H'' = 2i\omega D + \left( \frac{1}{\alpha^2} - 4i\frac{\omega}{\alpha^2} - 2\omega^2 \right)K,
\]
which indeed reduces to (12) for \(\omega \to 0\).

The same recipe works for the \(B\) transformation, which is again found from \(A\) by replacing \(H \to H_0\) and changing the overall sign,
\[
A = \alpha H + \frac{1}{\alpha} (1 - i\alpha \omega)^2 K - 2(1 - i\alpha \omega)D, \quad B = -\alpha H_0 - \frac{1}{\alpha} (1 - i\alpha \omega)^2 K + 2(1 - i\alpha \omega)D.
\]
It is straightforward to write down the second transformation and verify that
\[
\tilde{H}_1 \equiv e^{ib} e^{IA} H_1 e^{-iA} e^{-ib} = H_0 + \omega^2 K,
\]
which proves that we have indeed mapped the nonconformal Calogero model to decoupled harmonic oscillators, via a simple explicit albeit nonunitary similarity transformation. Clearly, the limit \(\omega \to 0\) connects with the results of the previous section.

Finally, like in the previous case one can establish the independence of the transformation on the parameter \(\alpha\). Thus, the formalism developed in the preceding section is universal and can be applied to both the conformal and nonconformal Calogero models.

4. Superconformal extensions

The unitary transformation constructed above has many interesting applications. In particular, it allows one to address the issue of superconformal extensions of the Calogero model. Below we treat in detail the \(N = 2\) case. In our setting, this amounts to adding fermionic coordinates to the free model and to properly modifying the nonlocal generator \(\tilde{K}\) such as to close the superconformal algebra. The inverse unitary transformation with the standard form (20) for \(A\) and \(B\) then maps the set of free superparticles back to the desired superconformal Calogero model with the standard representation of \(K\).
Apart from the so(1, 2) generators, the \( N = 2 \) superconformal algebra contains two supersymmetry generators \( Q \) and \( \tilde{Q} \) which are hermitian conjugates of each other, two superconformal generators \( S \) and \( \tilde{S} \) also related by Hermitian conjugation, and a \( u(1) \) generator \( J \). Altogether there are four bosonic and four fermionic operators, which obey the nonvanishing commutation relations (suppressing Hermitian conjugates)

\[
[H, D] = iH, \quad [K, D] = -iK, \quad [Q, D] = \frac{i}{2}Q, \quad [S, D] = -\frac{i}{2}S,
\]

\[
[Q, J] = -\frac{1}{2}Q, \quad [S, J] = -\frac{1}{2}S, \quad [H, K] = 2iD, \quad [Q, K] = -iS,
\]

\[
(Q, \tilde{Q}) = 2H, \quad (S, \tilde{S}) = 2K, \quad (Q, \tilde{S}) = -2D - 2iJ + iC, \quad [S, H] = iQ.
\]  

(28)

Here, \( C \) is a real constant which stands for a central charge. For the realization of this algebra we need to add to the coordinates \( x^i \) the same number \( \psi_i, \bar{\psi}_i, \) subject to the standard anticommutation relations

\[
{\psi_i, \bar{\psi}_j} = \delta^{ij} \quad \text{and} \quad {\psi^i, \psi^j} = 0 = {\bar{\psi}^i, \bar{\psi}^j} \quad \text{with} \quad (\psi^i)^\dagger = \bar{\psi}^i.
\]  

(29)

The algebra (28) suggests that \( \tilde{H} = H_0 = \frac{1}{2}p^i p^i \) is accompanied by

\[
\tilde{Q} = Q_0 = \psi^i p^i \quad \text{and} \quad \tilde{\psi} = \tilde{Q}_0 = \bar{\psi}^i p^i
\]  

(30)

and the dilatation and \( u(1) \) generators

\[
\tilde{D} = D = -\frac{1}{4}(x^i p_i + p_i x^i) \quad \text{and} \quad \tilde{J} = J = \frac{1}{4}(\psi^i \bar{\psi}^j - \bar{\psi}^j \psi^i).
\]  

(31)

The remaining (conformal) generators \( \tilde{K}, \tilde{S} \) and \( \tilde{\bar{S}} \) are nonlocal but acquire the standard form in the interacting model,

\[
K = \frac{1}{2} x^i x^i \quad \text{and} \quad S = \psi^i x^i, \quad \tilde{S} = \bar{\psi}^i x^i.
\]  

(32)

The goal is to construct the interacting-model Hamiltonian \( H \) and supercharges \( Q \) and \( \tilde{Q} \) by working our way back from the free model with the help of the algebra (28). To this end, we begin with the special conformal generator and parametrize as before

\[
\tilde{K} = K + a^2 e^{iB} H_{\text{int}} e^{-iB} \quad \text{but with} \quad H_{\text{int}} = \sum_{i<j} \frac{g^2}{(x^i - x^j)^2} + V,
\]  

(33)

allowing for a new contribution \( V \) due to the fermions. The algebra commutators (28) then consistently fix the form of the superconformal generator \( \tilde{S} \),

\[
[\tilde{Q}, \tilde{K}] = -i\tilde{S} \quad \Rightarrow \quad \tilde{S} = -i a e^{iB} [S, V] e^{-iB}.
\]  

(34)

Hermitian conjugation produces \( \tilde{S} \). Other structure relations of the superconformal algebra (28) yield the following restrictions on the form of \( V \):

\[
[K, V] = 0, \quad [D, V] = -iV, \quad [J, V] = 0, \quad [Q, H_{\text{int}}] + i[H_0 + V, [S, V]] = 0, \quad \{ S, \tilde{S}, V \} = C, \quad [S, V], [\tilde{S}, V] + i[Q, [S, V]] + i[\tilde{Q}, [S, V]] + 2H_{\text{int}} = 0,
\]  

(35)

plus their Hermitian conjugates.

Let us define an \( N = 2 \) Calogero model by finding a solution to Eqs. (35). The first line in (35) implies that the potential \( V \) is a homogeneous function of the \( x^i \) of degree \(-2\). Being \( u(1) \) neutral, it involves an equal number of \( \psi^i \) and \( \bar{\psi}^i \). Thus, it is natural to take the simplest ansatz

\[
V = V_{ij}(x) \psi^i \bar{\psi}^j = \frac{1}{2} V_{ii}(x) + \frac{1}{2} V_{jj}(x) [\psi^i, \bar{\psi}^j]
\]  

(36)

with unknown functions \( V_{ij}(x) \). Substituting this form into the remaining (anti)commutators in (35) one obtains a system of partial differential equations,

\[
-2V_{ij} = \partial_i (V_{jp} x^p) + \partial_j (V_{ip} x^p), \quad \partial_i V_{ij} + \partial_j (V_{ip} x^p) = 0, \quad \partial_p V_{ij} = \partial_i V_{pj},
\]

\[
\partial_i (V_{ip} x^p) + (V_{ip} x^p) (V_{ij} x^j) - 2 \sum_{i<j} \frac{g^2}{(x^i - x^j)^2} = 0, \quad V_{ij} x^i x^j = C.
\]  

(37)
The first equation implies that $V_{ij} = V_{ji}$. Then the second restriction gives the condition
\[ \partial_i (V_j p^i) - \partial_j (V_i p^i) = 0 \implies V_i p^i = \partial_i \Phi \] (38)
with some scalar function $\Phi$. The remaining equations in (37) imply that
\[ V_{ij} = -\partial_i \partial_j \Phi \] (39)
and constrain $\Phi$ to obey the partial differential equations
\[ \partial_i \partial_j \Phi + (\partial_i \Phi) (\partial_j \Phi) = 2 \sum_{i<j} \frac{\delta^2}{(x^i - x^j)^2} \quad \text{and} \quad x^i \partial_i \Phi = C. \] (40)

Any solution $\Phi$ to these equations will give rise to an $N = 2$ superconformal extension of the Calogero model.

The general solution to (40) can be put in the form
\[ \Phi = \mu \sum_{i<j} \ln|x^i - x^j| + v \ln \sqrt{x^2 + \Lambda \left( \frac{x^i}{x^j} \right)} \] (41)
where $\mu$ and $v$ are dimensionless constants, $x^2 \equiv x^i x^i$, and $\Lambda$ is a general function of coordinate ratios. Putting for simplicity $\Lambda \equiv 0$ and inserting (41) into (40), we find the conditions
\[ \mu (\mu - 1) = g^2 > -\frac{1}{4} \quad \text{and} \quad v (v + n(n-1)\mu + n - 2) = 0, \] (42)
which give four solutions for the pair $\mu(n, g)$ and $v(n, g)$. The central charge is fixed at
\[ C(n, g) = \frac{n(n-1)}{2} \mu + v. \] (43)

Differentiating twice as in (39) and inserting in (36) yields
\[ V = \sum_{i<j} \frac{\mu}{(x^i - x^j)^2} - \frac{n-2}{2} \frac{v}{x^2} + \frac{1}{2} \sum_{i \neq j} \frac{\mu}{(x^i - x^j)^2} \left[ \psi^i, \bar{\psi}^j - \bar{\psi}^i \right] - \frac{1}{2} \sum_{i,j} \frac{v}{x^2} \delta^{ij} - 2 x^i x^j \left[ \psi^i, \bar{\psi}^j \right] \] (44)
and, hence, with (42) the interaction Hamiltonian
\[ H_{int} = \sum_{i<j} \frac{\mu^2}{(x^i - x^j)^2} + \frac{1}{2} \sum_{i \neq j} \frac{\mu}{(x^i - x^j)^2} \left[ \psi^i, \bar{\psi}^j - \bar{\psi}^i \right] - \frac{n-2}{2} \frac{v}{x^2} - \frac{1}{2} \sum_{i,j} \frac{v}{x^2} \delta^{ij} - 2 x^i x^j \left[ \psi^i, \bar{\psi}^j \right] \] (45)
but also $\tilde{K}$ and $\tilde{S}$. The original Calogero coupling $g^2$ has been replaced by $\mu^2$, of which $v$ is a function via (42). By the very construction, this $H = H_0 + H_{int}$ along with $D$ and $K$ from (33) furnish a representation of so(1, 2). Therefore, they can be used to construct the inverse transformation $e^{-iA} e^{-iB}$ and hence the supercharge, which for $v = 0$ reads
\[ Q = e^{-iA} e^{-iB} (\psi^i p^i) e^{iB} e^{iA} = \psi^i p^i + i[V, S] = \psi^i \left( p^i + i \sum_{k(\neq i)} \frac{\mu}{x^i - x^k} \right). \] (46)

It may be checked that the same transformation maps $\tilde{S}$ of (34) back to $S$ as it should.

Beautifully enough, with $v = 0$ we have reproduced precisely the $N = 2$ superextension constructed by Freedman and Mende [19] in the framework of supersymmetric quantum mechanics. For the other solution to (42), $v = 2 - n - n(n-1)\mu$, we have apparently found an alternate superextension (see also [20]).

5. Concluding remarks

In this Letter we have constructed a simple unitary transformation relating the conformal Calogero model to a system of free particles on the real line. The simplification was achieved at a price of a highly nontrivial and, in particular, nonlocal realization of the full conformal symmetry in the resulting free theory. The transformation was shown to be universal and applicable to the nonconformal Calogero model as well as to $N = 2$ supersymmetric extensions. In the latter case we reconstructed not only the model of Freedman and Mende but found a second variant.

Turning to possible further developments, first to mind comes the $N = 4$ superconfomral extension of the Calogero model, which seems crucial for testing a conjecture of Gibbons and Townsend [12]. The construction realized for the su(1, 1|1) superalgebra in Section 4 can literally be generalized to the su(1, 1|2) superalgebra. This project is under way. Another interesting point is to employ our transformation for deriving the propagator of the Calogero model starting from the free propagator. Finally, it may be worthwhile to generalize the analysis of Section 3 to the case of a harmonic pair potential.
Acknowledgements

A.G. thanks the Institute for Theoretical Physics at Hannover University for the hospitality extended to him during the course of this work. He is grateful to the DAAD for financial support (A/06/27302). The research was supported by RF Presidential grants MD-8970.2006.2, NS-4489.2006.2, INTAS grant 03-51-6346, DFG grant 436 RUS 113/669/0-3, RFBR grant 06-02-16346 and RFBR-DFG grant 06-02-04012.

References

Constraints on gravitational scaling dimensions from non-local effective field equations

H.W. Hamber *,1, R.M. Williams 2

Theory Division, CERN, CH-1211 Geneva 23, Switzerland

Received 4 August 2006; received in revised form 30 October 2006; accepted 30 October 2006

Available online 7 November 2006

Abstract

Quantum corrections to the classical field equations, induced by a scale dependent gravitational constant, are analyzed in the case of the static isotropic metric. The requirement of general covariance for the resulting non-local effective field equations puts severe restrictions on the nature of the solutions that can be obtained. In general the existence of vacuum solutions to the effective field equations restricts the value of the gravitational scaling exponent $\nu^{-1}$ to be a positive integer greater than one. We give further arguments suggesting that in fact only for $\nu^{-1} = \frac{3}{2}$ consistent solutions seem to exist in four dimensions.

Over the last few years evidence has been increasing to suggest that quantum gravitation, even though plagued by uncontrollable divergences in standard weak coupling perturbation theory [1], might actually make sense, and lead to testable predictions at the non-perturbative level. These new results in general arise from the non-trivial scaling properties of the gravitational coupling constants in the vicinity of a non-trivial ultraviolet fixed point in four dimensions. As is often the case in physics, the best arguments do not come from often incomplete and partial results in a single model, but more appropriately from the level of consistency that various, often quite unrelated, field theoretic approaches provide.

The main aspect we wish to investigate in this Letter is the nature of the specific predictions about the running of Newton’s constant $G$, as they apply to the standard static isotropic metric. Our starting point will be the solution of the non-relativistic Poisson equation, which for a localized point source can be investigated for various values of the gravitational scaling exponent $\nu$. But a more appropriate setting will be a relativistic, generally covariant framework, wherein the effects of the leading quantum correction can be studied systematically, and for which we will show that the existence of vacuum solutions severely restricts the possible values for the exponent $\nu$. Specifically, we will show that no consistent solution to the effective non-local field equations can be found unless $\nu^{-1}$ is an integer greater than one. To check the overall consistency of the results, a different approach to the solution of the covariant effective field equations for the static isotropic metric will be pursued, in terms of an effective vacuum density and pressure. In this case one finds that unless the exponent $\nu$ is equal to $1/3$, a consistent solution cannot be obtained.

The starting point for our discussion is the form of the running gravitational coupling in the vicinity of the ultraviolet fixed point at $G_c$, as obtained from the lattice theory of gravity, and given in [3]

$$G(k^2) = G_c \left[ 1 + a_0 \left( \frac{m^2}{k^2} \right)^{\frac{1}{2\nu}} + O \left( (m^2/k^2)^{\frac{1}{2\nu}} \right) \right]$$

(1)

with $m = 1/\xi$, $a_0 > 0$ and $\nu \approx 1/3$. Usually the quantity $G_c$ in the above expression is identified with the laboratory scale value, $\sqrt{G_c} \sim \sqrt{G_{\text{phys}}} \sim 1.6 \times 10^{-33}$ cm, the reason being that the scale $\xi$ can be very large, roughly of the same order as the
scaled cosmological constant $\lambda$. Quantum corrections on the r.h.s. are therefore quite small as long as $k^2 \gg m^2$, which in real space corresponds to the "short distance" regime $r \ll \xi$.\(^3\)

For more details the reader is referred to the recent papers [3–5], and further references therein.

For $k^2 \to 0$ the quantum correction proportional to $a_0$ diverges, and the spurious infrared divergence needs to be regulated. A natural infrared regulator exists in the form of $m = 1/\xi$, and therefore a properly infrared regulated version of the previous expression is

$$G(k^2) \simeq G_c \left[ 1 + a_0 \left( \frac{m^2}{k^2 + m^2} \right)^{\frac{1}{2}} + \cdots \right]$$

(2)

with $m = 1/\xi$ the (tiny) infrared cutoff. Thus the gravitational coupling approaches the finite value $G_{\text{irr}} = (1 + a_0 + \cdots) G_c$, independent of $m = 1/\xi$, at very large distances $r \gg \xi$. The procedure for removing the spurious infrared divergence of Eq. (1) at small $k^2$ completely parallels the situation in non-Abelian gauge theories, where similar spurious infrared divergences appear [11,12]. There too the parameter $m = \xi^{-1}$, related to the non-perturbative gluon condensate, acts as a natural infrared regulator. A less elegant, but equivalent, procedure would consist in cutting off momentum integrals at $k_{\text{min}} = m$, but we shall not pursue such an approach here.\(^4\)

In this work we will be concerned with the static limit, where the non-relativistic Newtonian potential can be defined as

$$\phi(r) = (-M) \int \frac{d^3k}{(2\pi)^3} e^{ik \cdot x} G(k^2) \frac{4\pi}{k^2}.$$

(3)

The static potential $\phi(r)$ can be obtained from Eq. (2) directly by Fourier transform, or equivalently from the solution of Poisson’s equation with a point source at the origin. In the limit of weak fields, the relativistic field equations give for the $\phi$ field (with $g_{00}(x) \simeq -(1 + 2\phi(x))$)

$$\Delta \phi(x) = 4\pi G\rho(x)$$

(4)

and for a point source at the origin the first term on the r.h.s. is just $4\pi Mg_0^2\delta(0)(x)$. The solution for $\phi(r)$, obtained by Fourier transforming back to real space Eq. (3), gives in the large $r$ limit

$$\phi(r) \sim \frac{MG}{r} \left[ 1 + a_0 \left( 1 - c_1(mr)^{\frac{1}{2}} e^{-mr} \right) \right]$$

(5)

with $c_1 = 1/(2^{\frac{3}{2}}\Gamma(\frac{3}{2}))$. The part in $G(k^2)$ proportional to $a_0$ can equivalently be represented as a source term $\rho_m$ in Poisson’s equation, the latter determined from the inverse Fourier transform of the correction term in Eq. (2),

$$a_0 M \left( \frac{m^2}{k^2 + m^2} \right)^{\frac{1}{2}}.$$

(6)

One finds

$$\rho_m(r) = \frac{1}{8\pi} c_q a_0 M m^3 (mr)^{-\frac{3}{2}} K_{\frac{5}{12}}(3\frac{\xi}{\xi}) (mr)$$

(7)

with $c_q \equiv 2^{\frac{5}{6}} (\frac{5}{3})^{\frac{1}{2}}/\sqrt{\Gamma(\frac{1}{3})}$. Note that the vacuum polarization density $\rho_m(r)$ has the normalization property

$$4\pi \int_0^\infty r^2 dr \rho_m(r) = a_0 M$$

(8)

and that $\rho_m(r)$ diverges at small $r$ for $\nu \geq 1/3$. In the small $r$ limit and for general $\nu > 1/2$, one then finds from Poisson’s equation, using the expansion of the modified Bessel function $K_\nu(x)$ for small arguments,

$$\phi(r) \sim -\frac{MG}{r} + a_0MGc_2m^\frac{1}{2} r^{\frac{1}{2} - 1} + \cdots$$

(9)

with $c_2 = \nu[\sec(\frac{\pi}{2\nu})]/\Gamma(\frac{1}{2})$.

Solutions to Poisson’s equation with a running $G$ provide some useful insights into the structure of quantum corrections, but a complete analysis requires a study of the full relativistic field equations, which will be discussed next. A set of effective field equations incorporating the running of $G$ is obtained from the replacement [2]

$$G \to G(\Box) = G \left[ 1 + a_0 \left( \frac{m^2}{\Box} \right)^{\frac{1}{2}} + \cdots \right] \equiv G(1 + A(\Box))$$

(10)

with the d’Alembertian $\Box$ expressing the running of $G$ as in either Eqs. (1) or (2). The non-local effective field equations then read

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R + \lambda g_{\mu\nu} = 8\pi G(1 + A(\Box)) T_{\mu\nu}$$

(11)

with $A(\Box)$ given by Eq. (10), and $\lambda \simeq 1/\xi^2$. The use of the d’Alembertian $\Box$ to describe the running of couplings in gauge theories and quantum gravity was discussed in some detail, for example, in [13]. The corresponding trace equation is

$$R - 4\lambda = -8\pi G(1 + A(\Box)) T.$$

(12)

Being manifestly covariant, these expressions at least satisfy some of the requirements for a set of consistent field equations incorporating the running of $G$. The d’Alembertian $\Box$ operator is defined here through the appropriate combination of covariant derivatives

$$\Box \equiv g^{\mu\nu} \nabla_\mu \nabla_\nu$$

(13)

and its explicit form depends on the specific tensor nature of the object it is acting on. In general the operator $A(\Box)$ has to be defined by a suitable analytic continuation from positive integer powers, which is usually done by computing $\Box^n$ for positive integer $n$, and then analytically continuing to $n \to -1/2\nu$.

\(^3\) The result of Eq. (1) is in fact quite similar what one finds for gravity in 2 + $r$ dimensions [6–8], if one allows for a different value of exponent $\nu$ as one transitions from two to four dimensions, $G(k^2) \simeq G_c(1 + (m^2/k^2)^{(4-2)/2} + \cdots)$. See also the recent results discussed in [9,10].

\(^4\) In addition, the above expression of the running of $G$ only applies to the scaling regime for which $k \ll \xi^{-1}$, corresponding to distances much larger than the Planck length. At distances comparable to the Planck length string corrections, higher derivative terms and conformal anomaly contributions should be considered as well [3].
us set for now the cosmological constant $\lambda = 0$, since its contribution can always be added at a later stage. As long as one is interested in static isotropic solutions, one can take for the metric the most general form

$$ds^2 = -B(r) dt^2 + A(r) dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\phi^2).$$  \hfill (14)

For the energy momentum tensor we will take the perfect fluid form

$$T_{\mu \nu} = \text{diag}[B(r) \rho(r), A(r) p(r), r^2 p(r), r^2 \sin^2 \theta p(r)].$$  \hfill (15)

and a point source as the origin is simply represented as

$$T_{\mu \nu}(r) = \text{diag}[B(r) \rho(r), 0, 0, 0]$$  \hfill (16)

with the source proportional to a $3-d$ delta function.

Consider first the trace equation

$$R = -8\pi G(1 + A(\Box)) T = +8\pi G (1 + A(\Box)) \rho,$$  \hfill (17)

where we have used the fact that the point source at the origin is described just by the density term. One then computes the repeated action of the invariant d’Alembertian on $T$,

$$\Box (-8\pi GT) = \Box (8\pi G \rho) \equiv 16G\pi \rho \frac{A'}{RA}
- \frac{4G\pi A' \rho'}{AB}
+ \frac{4G\pi B' \rho'}{A}
+ \frac{8G\pi \rho''}{A}.$$  \hfill (18)

In view of the rapidly escalating complexity of the problem, it seems sensible to expand around the Schwarzschild solution, and set

$$A(r) = 1 - 2MG \frac{\sigma(r)}{r}, \quad B(r) = 1 - 2MG \frac{\theta(r)}{r},$$  \hfill (19)

where the correction to the standard solution are parametrized here by the two functions $\sigma(r)$ and $\theta(r)$, both assumed to be “small”, i.e. proportional to $a_0$ as in Eq. (10), with $a_0$ considered a small parameter. To simplify the problem even further, we will assume that for $2MG \ll r \ll \xi$ (the “physical” regime) one can set

$$\sigma(r) = -a_0 MG \xi \alpha, \quad \theta(r) = -a_0 MG \xi \beta.$$  \hfill (20)

This assumption is in part justified by the form of the non-relativistic correction of Eq. (9). Then for $\alpha = \beta$ (the equations seem impossible to satisfy if $\alpha$ and $\beta$ are different) one obtains for the scalar curvature

$$R = 0 + \alpha(2c_\sigma + (\alpha - 1)c_\theta) a_0 MG \xi^{\alpha - 3} + O(a_0^2).$$  \hfill (21)

A first result can be obtained in the following way. Since in the ordinary Einstein case one has for a perfect fluid $R = -8\pi GT = +8\pi G(\rho - 3p)$, and since $\rho_{\Box}(r) \sim r^{3-\alpha}$ from Eq. (7) in the same regime, one concludes that a solution is given by

$$\alpha = \frac{1}{\nu},$$  \hfill (22)

which is also consistent with the Poisson equation result of Eq. (9).

The next step up would be the consideration of the action of $\Box$ on the point source, as it appears in the full effective field equations of Eq. (11), with again $T_{\mu \nu}$ described by Eq. (16). One perhaps surprising fact is the generation of an effective pressure term by the action of $\Box$, suggesting that both terms should arise in the correct description of vacuum polarization effects,

$$(\Box T_{\mu \nu})_{rr} = \frac{\rho B''}{2AB} + \frac{2Bp' B' - BA' \rho'}{2A^2} + \frac{B' \rho'}{2A} + \frac{Bp''}{A},$$  \hfill (23)

and $(\Box T_{\mu \nu})_{\theta \theta} = (\Box T_{\mu \nu})_{\varphi \varphi} = 0$. A similar effect, namely the generation of an effective vacuum pressure term in the field equations by the action of $\Box$, was seen already in the case of the Robertson–Walker [2].

To check the overall consistency of the approach, consider next the set of effective field equations that are obtained when the operator $(1 + A(\Box))$ appearing in Eqs. (11) and (12) is moved over to the gravitational side. Since the r.h.s. of the field equations then vanishes for $r \neq 0$, one has apparently reduced the problem to one of finding vacuum solutions of a modified, non-local field equation. Let us first look at the relatively simple trace equation. If we denote by $\delta R$ the lowest order variation (that is, of order $a_0$) in the scalar curvature over the ordinary vacuum solution $R = 0$, then one has

$$\frac{1}{8\pi GA(\Box)} \delta R = 0.$$  \hfill (24)

On a generic scalar function $F(r)$ one has the following action of the covariant d’Alembertian $\Box$:

$$\Box F(r) = \frac{A' F'}{2A^2} + \frac{B' F'}{2AB} + \frac{2F'}{2A} + \frac{F''}{A}.$$  \hfill (25)

Assuming a power law correction, as in Eq. (20), with $\alpha = \beta$, as in Eq. (21), one then finds

$$\Box R \to a_0 MG a^{-3} (2c_\sigma + c_\theta (\alpha - 1) \alpha (\alpha - 2) (\alpha - 3),$$

$$\Box R \to a_0 MG a^{-3} (2c_\sigma + c_\theta (\alpha - 1) \alpha (\alpha - 2) (\alpha - 3) \alpha (\alpha - 4) (\alpha - 5)$$

and so on, and for general $n \to +\frac{1}{2}$

$$\Box^n R \to a_0 MG (2c_\sigma + c_\theta (\alpha - 1) \alpha (\alpha - 2) (\alpha - 3) \alpha (\alpha - 4) \alpha (\alpha - 5)$$

Therefore the only possible power solution for $r \gg MG$ is $\alpha = 0, 2, \ldots, \frac{1}{2}$, with $c_\sigma$ and $c_\theta$ unconstrained to this order.

Next we examine the full effective field equations (as opposed to just their trace part) as in Eq. (11) with $\lambda = 0$. If one denotes by $\delta T_{\mu \nu} \equiv \delta(R_{\mu \nu} - \frac{1}{2} g_{\mu \nu} R)$ the lowest order variation (that is, of order $a_0$) in the Einstein tensor over the ordinary vacuum solution $G_{\mu \nu} = 0$, then one has

$$\frac{1}{8\pi GA(\Box)} \delta \left( R_{\mu \nu} - \frac{1}{2} g_{\mu \nu} R \right) = 0.$$  \hfill (28)
again for $r \neq 0$. Here the covariant d’Alembertian operator $\Box$ acts on a second rank tensor and would thus seem to require the calculation of as many as 1920 terms, of which many fortunately vanish by symmetry. In the static isotropic case the components of the Einstein tensor are given by

\[
G_{tt} = \frac{A'B}{rA^2} - \frac{B}{r^2A} + \frac{A}{r^2},
\]

\[
G_{rr} = -\frac{A}{r^2} + \frac{B'}{rB} + \frac{1}{r^2},
\]

\[
G_{\theta\theta} = -\frac{B''r^2}{4AB^2} + \frac{A'B' + A''r^2}{2AB} - \frac{A'r^2}{2A^2} + \frac{B'r}{2AB},
\]

\[
G_{\phi\phi} = \sin^2 \theta G_{\theta\theta}.
\]

After acting with $\Box$ on this expression one finds a rather complicated result. Here we will list only ($\Box G$)$_{tt}$:

\[
\frac{6BA^3}{rA^5} + \frac{2BA^2}{r^2A^4} - \frac{4B'A}{r^3A^3} - \frac{2BA'}{r^3A^3} + \frac{6BA''A'}{rA^4} + \frac{B''A'}{rA^3} + \frac{6B}{r^4A} - \frac{4B'A}{r^4A} - \frac{2BA}{r^3A^3} + \frac{B''A}{r^2A^2} + \frac{B'A'}{r^2A^3} + \frac{BA}{rA^3} + \frac{B''}{r^2A} - \frac{B'A}{r^2A^2} + \frac{BA}{rA^3}.
\]

If one again assumes that the corrections are given by a power, as in Eq. (20), with $\alpha = 2$, then one has to lowest order

\[
G_{tt} = a_0MGc_\alpha \alpha^{-3},
\]

\[
G_{rr} = -a_0MG(c_\alpha + c_\alpha (\alpha - 1))\alpha^{-3},
\]

\[
G_{\theta\theta} = -\frac{1}{2}a_0MG(c_\alpha + c_\alpha (\alpha - 1))(\alpha - 1)\alpha^{-1}
\]

with the $\phi\phi$ component again proportional to the $\theta\theta$ component. Applying $\Box$ on the above Einstein tensor one then gets

\[
(\Box G)_{tt} = a_0MGc_\alpha \alpha^{-5},
\]

\[
(\Box G)_{rr} = -a_0MG(c_\alpha + c_\alpha (\alpha - 1))\alpha^{-5},
\]

\[
(\Box G)_{\theta\theta} = -\frac{1}{2}a_0MG(c_\alpha + c_\alpha (\alpha - 1))(\alpha - 1)\alpha^{-3} r^2\alpha^{-3}
\]

(with the $\phi\phi$ component proportional to the $\theta\theta$ component), and so on. One then has for general $n \rightarrow +\frac{1}{2}$

\[
(\Box^n G)_{tt} \rightarrow a_0MGc_\alpha \Gamma(2 + \frac{1}{2} - \alpha)\Gamma(\alpha - 1)\Gamma(-\alpha) \alpha^{-3} - \frac{1}{2},
\]

\[
(\Box^n G)_{rr} \rightarrow -a_0MG(c_\alpha + c_\alpha (\alpha - 1)) \times \frac{\Gamma(2 + \frac{1}{2} - \alpha)\Gamma(\alpha - 1)\Gamma(-\alpha)}{(\alpha - 1)(\alpha - 1)(\alpha - 1)\Gamma(-\alpha)} \alpha^{-3} - \frac{1}{2},
\]

\[
(\Box^n G)_{\theta\theta} \rightarrow -\frac{1}{2}a_0MG(c_\alpha + c_\alpha (\alpha - 1)) \times \frac{(\alpha - 1)(\alpha - 1)(\alpha - 1)\Gamma(-\alpha)}{(\alpha - 1)(\alpha - 1)(\alpha - 1)\Gamma(-\alpha)} \alpha^{-3} - \frac{1}{2}.
\]

Inspection of the above results reveals a common factor $1/\Gamma(-\alpha)$, which would allow only integer powers $\alpha = 0, 1, 2, \ldots$ but the additional factor of $1/(\alpha - 1)$ excludes $\alpha = 1$ from being a solution. Even for $\alpha$ close to $1/\nu$ (as expected on the basis of the non-relativistic expression of Eq. (9), as well as from Eq. (22)) $\nu \sim 1/\alpha - \epsilon$ only integer values $\alpha = 2, 3, 4, \ldots$ are allowed. In general the problem of finding a complete general solution to the effective field equations by this method lies in the difficulty of computing arbitrarily high powers of $\Box$ on general functions such as $\sigma(r)$ and $\theta(r)$, which eventually involve a large number of derivatives. Assuming for these functions a power law dependence on $r$ simplifies the problem considerably, but also restricts the kind of solutions that one is likely to find. More specifically, if the solution involves (say for small $r$, but still with $r \gg 2MG$) a term of the type $r^n lnmr$, as in Eqs. (9), (48) and (51) for $\nu \rightarrow 1/3$, then this method will have to be dealt with very carefully. This is presumably the reason why in some of the $\Gamma$-function coefficients encountered here one finds a power solution (in fact $\alpha = 3$) for $\nu$ close to a third, but one gets indeterminate expression if one sets exactly $\alpha = 1/\nu = 3$.

The earlier discussion of the non-relativistic case suggests that the quantum correction due to the running of $G$ can be approximately described by Poisson’s equation, with a source term related to a vacuum energy density $\rho_m(r)$, distributed around the static source of strength $M$ in accordance with the result of Eqs. (7) and (8). These expressions, in turn, were obtained by Fourier transforming back to real space the original result for $G(k^2)$ of Eq. (2). Furthermore, in the preceding discussion of the relativistic case it was found (as in [2] for the Robertson–Walker metric case) that a manifestly covariant implementation of the running of $G$, via the $G(\Box)$ given in Eq. (10), will induce a non-vanishing effective pressure term in the field equations. This result can be seen clearly, in the case of the static isotropic metric, for example from the result of Eq. (23). We will therefore now consider a relativistic perfect fluid, with energy–momentum tensor, which in the static isotropic case reduces to Eq. (15). The $tt$, $rr$ and $\theta\theta$ components of the field equations then read

\[ -\lambda B + \frac{A'B}{rA^2} - \frac{B}{r^2A} + \frac{B}{r^2} = 8\pi GB\rho, \]

\[ \lambda A - \frac{A}{r^2} + \frac{B'}{rB} + \frac{1}{r^2} = 8\pi GA\rho, \]

\[ \lambda r^2 - \frac{B''r^2}{4AB^2} - \frac{A'B'r^2}{4A^2B} - \frac{B''r^2}{2AB} - \frac{B'r}{2AB} = 8\pi Gr^2p. \]

(34)

with the $\phi\phi$ component equal to $\sin^2 \theta$ times the $\theta\theta$ component. Energy conservation $\nabla^\mu T_{\mu\nu} = 0$ implies

\[ (p + \rho) \frac{B'}{2B} + p' = 0 \]

(35)

and forces a definite relationship between $B(r)$, $\rho(r)$ and $p(r)$. The three field equations and the energy conservation equation are, as usual, not independent, because of the Bianchi identity.

It seems reasonable to attempt to solve the above equations (usually considered in the context of relativistic stellar structure) with the density given by the $\rho_m(r)$ of Eq. (7). This of course raises the question of how the relativistic pressure $p(r)$ should be chosen, an issue that the non-relativistic calculation did not have to address. We will argue below that covariant energy conservation completely determines the pressure in the
static case, leading to consistent equations and solutions (note that in particular it would not be consistent to take \( p(r) = 0 \).

Since the function \( B(r) \) drops out of the \( tt \) field equation, the latter can be integrated immediately, giving

\[
A(r)^{-1} = 1 + \frac{c_1}{r} - \frac{\lambda}{3} r^2 - \frac{8 \pi G}{r} \int_0^r dx x^2 \rho(x). \tag{36}
\]

It also seems natural here to identify \( c_1 = -2MG \), of course corresponds to the correct solution for \( a_0 = 0 \) (\( p = \rho = 0 \)). Next, the \( rr \) field equation can be solved for \( B(r) \),

\[
B(r) = \exp \left\{ c_2 - \int_{r_0}^r dy \frac{1 + A(y) (\lambda y^2 - 8 \pi G y^2 p(y) - 1)}{y} \right\} \tag{37}
\]

with the constant \( c_2 \) again determined by the requirement that the above expression for \( B(r) \) reduce to the standard Schwarzschild solution for \( a_0 = 0 \) (\( p = \rho = 0 \)), giving \( c_2 = \ln(1 - 2MG/r_0 - \lambda/r_0^2/3) \). The last task left therefore is the determination of the pressure \( p(r) \). Using the \( rr \) field equation, \( B''(r)/B(r) \) can be expressed in term of \( A(r) \) in the energy conservation equation. Inserting then the explicit expression for \( A(r) \), from Eq. (36), one obtains

\[
p''(r) + \left( 8 \pi G r^3 p(r) + 2MG - \frac{2}{3} \frac{\lambda}{3} r^3 + 8 \pi G \int_0^r dx x^2 \rho(x) \right) \times (p(r) + \rho(r)) \\
\times \left[ 2 \pi \left( r - 2MG - \frac{\lambda}{3} r^3 - 8 \pi G \int_0^r dx x^2 \rho(x) \right) \right]^{-1} = 0 \tag{38}
\]

which is usually referred to as the equation of hydrostatic equilibrium. From now on we will focus only the case \( \lambda = 0 \). The last equation, a non-linear differential equation for \( p(r) \), can be solved to give the desired solution \( p(r) \), which then, by Eq. (37), determines the remaining function \( B(r) \). In our case though it will be sufficient to solve the above equation for small \( a_0 \), where \( a_0 \) (see Eqs. (2) and (7)) is the dimensionless parameter which, when set to zero, makes the solution revert back to the classical one. It will also be convenient to pull out of \( A(r) \) and \( B(r) \) the Schwarzschild solution part, by introducing the small corrections \( \sigma(r) \) and \( \theta(r) \), as defined in Eq. (19), both of which are expected to be proportional to the parameter \( a_0 \). One then has

\[
\theta(r) = \exp \left\{ c_2 + \int_{r_0}^r dy \frac{1 + 8 \pi G y^2 p(y)}{y - 2MG - 8 \pi G \int_0^r dx x^2 \rho(x)} \right\} + 2MG - r. \tag{39}
\]

Again, the integration constant \( c_2 \) needs to be chosen here so that the normal Schwarzschild solution is recovered for \( p = \rho = 0 \). To order \( a_0 \) the resulting equation for \( p(r) \), from Eq. (38), is

\[
\frac{MG(p(r) + \rho(r))}{r(r - 2MG)} + p'(r) \simeq 0. \tag{40}
\]

Note that in regions where \( p(r) \) is slowly varying, \( p'(r) \simeq 0 \), one has \( p \simeq -\rho \), i.e. the fluid contribution is acting like a cosmological constant term with \( \sigma(r) \sim \theta(r) \sim -\rho/(3r^3) \). The last differential equation can then be solved for \( p(r) \),

\[
p_m(r) = \frac{1}{\sqrt{1 - 2MG/r}} \left( c_3 - \int_{r_0}^r dz \frac{MGp(z)}{z^2 \sqrt{1 - 2MG/r}} \right), \tag{41}
\]

where the constant of integration has to be chosen so that when \( \rho(r) = 0 \) (no quantum correction) one has \( p(r) = 0 \) as well. Because of the singularity in the integrand at \( r = 2MG \), we will take the lower limit in the integral to be \( r_0 = 2MG + \epsilon \), with \( \epsilon \to 0 \). To proceed further, one needs the explicit form for \( p_m(r) \), which was given in Eqs. (7) and (8). The required integrands involve for general \( \nu \) the modified Bessel function \( K_\nu(x) \), and can be therefore a bit complicated. Here we will limit our investigation to the small \( \nu \) (\( m \ll 1 \)) behavior of the modified Bessel function \( K_\nu(x) \), and can be therefore a bit complicated. Here we will limit our investigation to the small \( \nu \) (\( m \ll 1 \)) behavior. From now on we will focus only the case \( \nu = 1/3 \), with the constant

\[
A_0 = \frac{|\sec(\pi/2\nu)|}{4\pi Gamma(\frac{1}{\nu} - 1)}, \tag{42}
\]

determined from the small \( x \) behavior of the modified Bessel function \( K_\nu(x) \). For \( \nu < 1/3 \) \( p_m(r) \sim \text{const} \times a_0 Mm^3 \), independent of \( r \). For \( \nu = 1/3 \) the expression for \( p_m(r) \) is given later in Eq. (46). Therefore in this limit, with \( \frac{1}{\nu} < \nu < 1 \), one has

\[
A^{-1}(r) = 1 - \frac{2MG}{r} - 2a_0MGc_p m^{\frac{1}{\nu} - 1} + \ldots \tag{44}
\]

with the constant \( c_p = \nu|\sec(\pi/2\nu)|/\Gamma(\frac{1}{\nu} - 1) \). For \( \nu = 1/3 \) the last contribution is indistinguishable from a cosmological constant term \( -\frac{1}{3} r^2 \), except for the fact that the coefficient here is quite different, being proportional to \( a_0 Mm^3 \). To determine the pressure, we suppose that it as well has a power dependence on \( r \) in the regime under consideration. \( p_m(r) = c_p A_0 r^\nu \), where \( c_p \) is a numerical constant, and then substitute \( p_m(r) \) into the pressure equation (40). This gives, past the horizon \( r \gg 2MG \),

\[
(2\nu - 1)c_p MG r^{\nu - 1} - c_p \nu r^\nu - MG r^{1/\nu - 4} \simeq 0 \tag{45}
\]

giving the same power \( \nu = 1/3 - 3 \) as for \( \rho(r) \), \( c_p = -1 \) and surprisingly also \( \nu = 0 \), implying that in this regime only \( \nu = 1/3 \) gives a consistent solution. Again, the resulting correction is quite similar to what one would expect from a cosmological term, with an effective \( \lambda_m/3 \simeq 8\pi \nu a_0 Mm^3 \).

The case \( \nu = 1/3 \) requires special treatment, and one needs to go back to the expression for \( p_m(r) \) for \( \nu = 1/3 \),

\[
\rho_m(r) = \frac{1}{2\pi^2} a_0 Mm^3 K_0(mr). \tag{46}
\]
For small $r$ one then has

$$\rho_m(r) = -\frac{a_0}{2\pi^2} M^3 m^3 \left( \ln \frac{mr}{2} + \gamma \right) + \cdots$$

(47)

and consequently from Eq. (36),

$$A^{-1}(r) = 1 - \frac{2MG}{r} + \frac{4a_0GMm^3}{3\pi r^2} \ln(mr) + \cdots$$

(48)

From Eq. (40) one then obtains an expression for the pressure $p_m(r)$, and one finds

$$p_m(r) = \frac{a_0Mm^3}{2\pi^2} \log(mr)$$

$$- \frac{a_0Mm^3}{2\pi^2} \sqrt{1 - \frac{2MG}{r} - MG}$$

$$+ \frac{a_0M^2}{\pi^2} + \frac{a_0M^2c_3}{2\pi^2} \frac{1}{\sqrt{1 - \frac{2MG}{r}}}$$

(49)

where $c_3$ is again an integration constant. Here we will be content with the $r \gg 2MG$ limit of the above expression, which we shall write therefore as

$$p_m(r) = \frac{a_0}{2\pi^2} M^3 m^3 \ln(mr) + \cdots$$

(50)

After performing the required $r$ integral in Eq. (39), and evaluating the resulting expression in the limit $r \gg 2MG$, one obtains an expression for $\theta(r)$, and from it

$$B(r) = 1 - \frac{2MG}{r} + \frac{4a_0GMm^3}{3\pi r^2} \ln(mr) + \cdots$$

(51)

The expressions for $A(r)$ and $B(r)$ are, for $r \gg 2MG$, consistent with a gradual slow increase in $G$ in accordance with the formula

$$G \to G(r) = G \left( 1 + \frac{a_0}{3\pi} m^3 \ln \frac{1}{m^2r^2} + \cdots \right)$$

(52)

and therefore consistent as well with the original result of Eqs. (1) or (2), namely that the classical laboratory value of $G$ is obtained for $r \ll \xi$. In fact it is quite reassuring that the renormalization properties of $G(r)$ as inferred from $A(r)$ are the same as what one finds from $B(r)$. For large $r$ one has instead, from Eq. (7) for $\rho_m(r)$,

$$\rho_m(r) \sim A_0 r^{\frac{3}{2} - 2} e^{-mr}$$

(53)

with $A_0 = 1/\sqrt{128\pi c_4 a_0 M^4 m^{1+\frac{1}{d}}}$. In the same limit, the integration constants is chosen so that the solution for $A(r)$ and $B(r)$ at large $r$ corresponds to a mass $M' = (1 + a_0) M$ (see the expression for the integrated density in Eq. (8)), or equivalently

$$\sigma(r) \sim \theta(r) \sim -2a_0 MG.$$

(54)

One then recovers a result similar to the non-relativistic expression of Eq. (5), with $G(r)$ approaching the constant value $G_\infty = (1 + a_0) G$, up to an exponentially small correction in $mr$ at large $r$.

In conclusion, it appears that a solution to the relativistic static isotropic problem of the running gravitational constant can be found, provided that the exponent $\nu$ in either Eq. (2) or Eq. (11) is close to one third. This last result seems to be linked with the fact that the running coupling term acts in some way like a local cosmological constant term, for which the $r$ dependence of the vacuum solution for small $r$ is fixed by the nature of the Schwarzschild solution with a cosmological constant term. Furthermore, in $d \geq 4$ dimensions the Schwarzschild solution to Einstein gravity with a cosmological term is given by [14]

$$A^{-1}(r) = B(r)$$

$$= 1 - \frac{8MG\pi^2}{(d-2)(d-1)} r^{3-d} - \frac{2\lambda}{(d-2)(d-1)} r^2$$

(55)

which would suggest, in analogy with the results for $d = 4$ given previously, that in $d \geq 4$ dimensions only $\nu = 1/(d-1)$ is possible, if the correction again behaves locally like a cosmological constant term. This last result would also be in agreement with the exact value $\nu = 0$ found at $d = \infty$ [5], as well as with approximate renormalization group studies [10].

To summarize, the starting point for our discussion of the renormalization group running of $G$ is Eq. (1), valid at short distances $k \gg m$, or its improved infrared regulated version of Eq. (2). While a solution to the non-relativistic Poisson equation can be given for various values of the exponent $\nu$, the scale dependence of $G$ can also be consistently embedded in a relativistic covariant framework using the d’Alembertian $\Box$ operator. This then leads to a set of non-local effective field equations, whose consequences can be worked out for the static isotropic metric, at least in a regime where $2MG \ll r \ll \xi$, and under the assumption of a power law correction. We have found that the structure of the leading quantum correction is such that it severely restricts the possible values for the exponent $\nu$, in the sense that no consistent solution to the effective non-local field equations, incorporating the running of $G$, can be found unless $\nu^{-1}$ is an integer. A somewhat different approach to the solution of the static isotropic metric was pursued in terms of an effective vacuum density of Eq. (7), and a vacuum pressure chosen so as to satisfy a covariant energy conservation for the vacuum polarization contribution. The main result there is the derivation, from the relativistic field equations, of an expression for the metric coefficients $A(r)$ and $B(r)$, given for $2MG \ll r \ll \xi$ in Eqs. (48), (51) and (52). From the nature of the solution for $A(r)$ and $B(r)$ one finds again that unless the exponent $\nu$ is close to $1/3$, a consistent solution to the field equations cannot be found.

In related papers, the authors of Refs. [15,16] have also investigated quantum corrections to the Schwarzschild metric arising from scale dependent gravitational couplings. In contrast to the results presented here, these approaches are not based on a manifestly covariant set of effective field equations incorporating the renormalization group running of $G$, and provide therefore no useful constraint on the gravitational scaling dimensions, arising in our work from the imposition of general covariance.
Acknowledgements

The authors wish to thank Luis Álvarez Gaumé, Gabriele Veneziano and the Theory Division at CERN for their warm hospitality. The work of Ruth M. Williams was supported in part by the UK Particle Physics and Astronomy Research Council.

References

Determining $F_\pi$ from spectral sum rules

Magdalena Luz 1

Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark

Received 24 July 2006; received in revised form 13 October 2006; accepted 2 November 2006

Available online 10 November 2006

Editor: G.F. Giudice

Abstract

We derive spectral sum rules for a system with two quarks coupled to an imaginary isospin chemical potential in the $\epsilon$ regime. The sum rules show an explicit dependence on the pion decay constant which should make it possible to measure $F_\pi$ from the eigenvalue spectrum of this particular Dirac operator.

© 2006 Elsevier B.V. All rights reserved.

1. The determination of low energy constants of QCD such as the pion decay constant $F_\pi$ remains an important problem. In particular the computation of such quantities from lattice calculations is notoriously difficult due to the exceeding computational challenge posed by simulations of small quark masses. In addition when approaching the chiral limit on the lattice, finite size effects inevitably become more and more significant. Hence, a method which takes finite size scaling explicitly into account seems to be highly profitable. Such a technique is provided by the so-called $\epsilon$-regime of QCD [1]. This regime applies in a region where the Compton wave length of the pion $1/m_\pi$ is larger than the one-dimensional size $L$ of the physical volume $V = L^4$, while still being much smaller than the typical hadronic scale $\Lambda_{QCD}$, i.e., $1/m_\pi > L \gg 1/\Lambda_{QCD}$. The lowest order effective partition function of the $\epsilon$-regime is known analytically. The fact that it depends explicitly on the infinite volume chiral condensate $\Sigma = \langle \bar{\psi} \psi \rangle$ can be exploited to determine this constant with high accuracy: back in 1992 Leutwyler and Smilga derived a set of spectral sum rules for the Dirac operator by restricting the partition function to sectors with fixed topological charge $\nu$ [2]. In this way the chiral condensate is linked to the spectrum of the Dirac operator in finite volume which can be determined from lattice simulations [3,4].

However, with the standard Dirac operator other low energy constants such as the pion decay constant $F_\pi$ appear only in higher order corrections [5]. As a consequence, the computation of $F_\pi$ was believed to be much more demanding [4,6,7]. It is therefore still more widespread to extrapolate the pion decay constant down to the chiral limit from simulations in the $p$ regime of chiral perturbation theory [8,9].

Recently a new approach has been proposed [10–12] which avoids these difficulties: if the quarks are coupled to an external source, dependency on $F_\pi$ appears already at the lowest order in the partition function [5,13]. The authors of Refs. [10,11] have used this fact to derive correlation functions of the eigenvalue densities which are sensitive to $F_\pi$ for both quenched and unquenched chiral perturbation theory. Here, we will use the same partition function to deduce a set of sum rules in a way analogous to Ref. [2]. These rules depend then likewise on $F_\pi$ and make it possible to determine this important quantity from a lattice simulations in finite volume.

2. We consider a system with two quark flavors $u$ and $d$ which are coupled to an external source $\mu_{iso}$. The source can be interpreted as an imaginary isospin chemical potential which couples differently to the two flavors or as twisted boundary conditions
by the following argument [2]: the rhs of Eq. (6) is actually proportional to the quark condensate $\langle \bar{u}u \rangle$. Here we use a non-mass-degenerate quark pair for purely technical reasons: as we will see below this gives us the only handle to distinguish between the two sets of eigenvalues. Thus keeping the masses distinct allows us to extract sum rules for only one set of eigenvalues, $\lambda_+$ or $\lambda_-$ accordingly.

For both sets of eigenvalues the non-zero modes come in positive and negative pairs, $\pm \lambda_+\mu$ and $\pm \lambda_-\mu$, respectively. Hence in a sector of fixed topology, the partition function can be written as

$$Z_\lambda(m_u, m_d) = \left( m_u^2 m_d^2 \prod_n \left( \lambda_+^2 + m_u^2 \right) \prod_n \left( \lambda_-^2 + m_d^2 \right) \right) \langle \cdots \rangle,$$

where $\langle \cdots \rangle$ denotes the gauge average over all configurations with topological charge $\nu$ and the products are restricted to strictly positive values of $\lambda$.

Our derivation of the sum rules deviates from the procedure originally used by Leutwyler and Smilga in Ref. [2], we will follow the somewhat easier approach presented in Ref. [15] instead. Massive spectral sum rules for $\lambda_+$ (or $\lambda_-$) can be derived from formula (3) by taking logarithmic derivatives with respect to the masses $m_\pm$ (or $m_d$). A first order sum rule for instance is given by

$$\frac{1}{2m_u} \left( \frac{\partial}{\partial m_u} \ln Z_\lambda(m_u, m_d) - \frac{\nu}{m_u} \right) = \left( \sum_n \frac{1}{\lambda_+^2 + m_u^2} \right) \langle \cdots \rangle.$$

As in the products of Eq. (3) the sum on the rhs runs over the positive eigenvalues only.

On the other hand in the $\epsilon$-expansion of chiral perturbation theory the partition function can be shown [11,13] to be

$$Z_{\epsilon}(x_u, x_d) = e^{-2VF_\pi^2 \mu_\pi^2} \int_0^1 ds e^{2VF_\pi^2 \mu_\pi^2 s^2} I_{\nu}(s x_d) I_{\nu}(s x_u),$$

where we introduced the scaling variables of the $\epsilon$-regime $x_i = V \Sigma m_i$, $i = u, d$ and $I_{\nu}$ are modified Bessel functions. The dependence on $F_\pi$ is through the product $VF_\pi^2 \mu_\pi$ only. In particular a sign change in $\mu_\pi$ leaves the partition function invariant. The two flavors are indeed only distinguishable through their masses. Upon inserting this explicit formula for $Z_{\epsilon}$ into Eq. (4), we obtain the first order sum rule

$$\left( \sum_n \frac{1}{\lambda_+^2 + m_u^2} \right) \langle \cdots \rangle = \frac{V^2 \Sigma^2}{2x_u} \int_0^1 ds e^{2VF_\pi^2 \mu_\pi^2 s^2} I_{\nu}(s x_d) I_{\nu+1}(s x_u).$$

Eq. (6) has to be taken with some care. The lhs needs to be properly regularized. That this expression is UV finite can be seen by the following argument [2]; the lhs of Eq. (6) is actually proportional to the quark condensate $\langle \bar{u}u \rangle$.

$$\langle \bar{u}u \rangle = - \lim_{V \to \infty} \frac{2m_u}{V} \sum_n \left( \sum_n \frac{1}{\lambda_+^2 + m_u^2} \right) \langle \cdots \rangle = -2m_u \int_0^\infty d\lambda_+ \frac{\rho(\lambda_+)}{\lambda_+^2 + m_u^2},$$

as the eigenvalues become dense in the infinite volume limit. Eq. (7) is ultraviolet divergent, because the density scales proportional to $\lambda_+^3$, for large $\lambda_+$. It can however be regularized by the introduction of counterterms of order $m_u$ and $m_u^3$ and a proper tuning of their coefficients. This argument is related to the fact that the partition function of QCD can be made finite by the introduction of a cosmological constant. As a consequence, higher order sum rules can be shown to be finite by the same argument. Let us introduce a cutoff $A$ such that $m_u \ll A$ and add the counterterms

$$\langle \bar{u}u \rangle = -2m_u \int_0^A d\lambda_+ \frac{\rho(\lambda_+)}{m_u^2 + \lambda_+^2} - 2m_u \int_A^\infty d\lambda_+ \frac{\rho(\lambda_+)}{m_u^2 + \lambda_+^2} + c_1 m_u + c_2 m_u^3.$$

[12,14]. This gives rise to two independent eigenvalue equations

$$D_+ + m_u \psi_+ = (\Phi[A] + i \mu_\pi \gamma_0 + m_u) \psi_+, \quad (1)$$

$$D_- + m_d \psi_- = (\Phi[A] - i \mu_\pi \gamma_0 + m_d) \psi_-, \quad (2)$$

where $A$ denotes the gauge field. The advantage of an imaginary isospin chemical potential is twofold. Firstly, an isospin chemical potential preserves the positivity of the fermion matrix [14]. The system can thus be simulated on the lattice with the usual Monte Carlo techniques without running into sign problems. Secondly, if it is chosen to be imaginary the massless operators $D_+$ and $D_-$ are anti-Hermitian and the eigenvalues $\lambda_\pm$ lie on the real axis.

In the case of degenerate masses $m_\mu = m_d$ the two flavors can be converted into each other by the transformation $\mu_\pi \to -\mu_\pi$. Here we use a non-mass-degenerate quark pair for purely technical reasons: as we will see below this gives us the only handle to distinguish between the two sets of eigenvalues. Thus keeping the masses distinct allows us to extract sum rules for only one set of eigenvalues, $\lambda_+$ or $\lambda_-$ accordingly.
The divergent contributions can be subtracted from the high momentum part of the integral

\[ 2m_u \int_{A}^{\infty} d\lambda_+ \frac{\rho(\lambda_+)}{m_u^2 + \lambda_+^2} + c_1 m_u + c_2 m_u^3 = 2m_u^5 \int_{A}^{\infty} d\lambda_+ \frac{\rho(\lambda_+)}{\lambda_+^4(m_u^2 + \lambda_+^2)} + \gamma_1 m_u + \gamma_2 m_u^3. \]  
\( (9) \)

All these terms disappear, as the mass is taken to zero and we are left with the low momenta contribution in Eq. (8). In lattice simulations we have of course a finite mass and a finite volume, but as we keep the scaling variable \( \xi = V \Sigma m_i \) constant and simulate at a fixed cutoff the correction terms become irrelevant as we go to larger lattices.

Formula (6) simplifies a good deal if we take both quark masses to zero. The Bessel functions disappear up to remnants powers of the parameter \( s \)

\[ \left( \sum_n \frac{1}{\lambda_+ n} \right)_v = \frac{V^2 \Sigma^2}{4(\nu + 1)} \int_0^1 ds e^{2V F_\pi^2 \mu_{iso}^2 s^2} \frac{s^{2v+3}}{(s+x_u) (s+x_d)}. \]
\( (10) \)

For general \( v \) this expression can only be evaluated in terms of incomplete and ordinary \( \Gamma \)-functions

\[ \left( \sum_n \frac{1}{\lambda_+ n} \right)_v = -\frac{V \Sigma^2}{8(\nu + 1) F_\pi^2 \mu_{iso}^2} \Gamma(\nu + 2) - \Gamma(\nu + 2, -2V F_\pi^2 \mu_{iso}^2) = -\frac{V \Sigma^2}{8(\nu + 1) F_\pi^2 \mu_{iso}^2} \int_0^0 \frac{s^{2v+3}}{(s+x_u) (s+x_d) s^{2v+1}} \frac{d e^{-s}}{s}. \]
\( (11) \)

However, for a given topological charge it is straightforward to calculate the parameter integrals. The rule reduces then to a rather simple expression, where \( F_\pi \) appears only in polynomials and exponentials and which can easily be fitted to lattice data. Let us illustrate this for the case of vanishing topological charge, there we simply have

\[ \left( \sum_n \frac{1}{\lambda_+ n} \right)_v = V^2 \Sigma^2 \frac{1 + e^{2V F_\pi^2 \mu_{iso}^2} (2V F_\pi^2 \mu_{iso}^2 - 1)}{e^{2V F_\pi^2 \mu_{iso}^2} - 1}. \]
\( (12) \)

Eqs. (10) and (11) are the direct equivalents of the first order sum rule given in Ref. [2], but evaluated for a system with two quarks coupled the chemical potential \( \mu_{iso} \). Indeed, in the limit where \( \mu_{iso} \) vanishes Eq. (10) becomes

\[ \left( \sum_n \frac{1}{\lambda_+ n} \right)_v = \frac{V^2 \Sigma^2}{4(\nu + 1)} \int_0^1 ds s^{2v+3} \frac{e^{2V F_\pi^2 \mu_{iso}^2 s^2}}{e^{2V F_\pi^2 \mu_{iso}^2} - 1} = \frac{V^2 \Sigma^2}{4(\nu + 2)}. \]
\( (13) \)

which is precisely the Leutwyler–Smilga result for \( N_f = 2 \) flavors.

Due to the symmetry of the partition function under an exchange of the quark masses, the corresponding sum rule for \( \lambda_- \) can be obtained from Eq. (6) by simply substituting \( s_u \) for \( s_d \). If we had treated the quarks as mass degenerate from the beginning we would have arrived at the sum of those two sum rules, which is just twice Eq. (6).

**Higher order sum rules**

We can derive two different types of second order sum rules by either taking the second derivative with respect to \( m_u, \frac{\partial^2}{\partial m_u \partial m_d} \ln Z_v \), or a mixed derivative \( \frac{\partial^2}{\partial m_u \partial m_d} \ln Z_v \). Let us start with the second possibility. Applied to Eq. (3) the mixed derivative yields the subtracted correlation of the two sets of eigenvalues

\[ \frac{1}{4m_u m_d} \left[ \frac{\partial^2}{\partial m_u \partial m_d} \ln Z_v(m_u, m_d) \right] = \left( \sum_n \frac{1}{\lambda_+ - n_m + m_u^2} \right)_v \left( \sum_n \frac{1}{\lambda_- - n_m + m_d^2} \right)_v - \left( \sum_n \frac{1}{\lambda_+ - n_m + m_d^2} \right)_v \left( \sum_n \frac{1}{\lambda_- - n_m + m_u^2} \right)_v. \]
\( (14) \)

Since we have calculated the disconnected parts already, the only new contribution from Eq. (14) is a sum rule for

\[ \left( \sum_n \frac{1}{\lambda_+ - n_m + m_u^2} \right)_v \left( \sum_n \frac{1}{\lambda_- - n_m + m_d^2} \right)_v = \frac{V^4 \Sigma^4}{4x_u x_d} \int_0^1 ds e^{2V F_\pi^2 \mu_{iso}^2 s^2} 3 I_{v+1}(s x_u) I_v(s x_d). \]
\( (15) \)

In the limit of vanishing quark masses we can again expand the Bessel functions and obtain a much simpler expression for the massless mixed second order sum rule

\[ \left( \sum_n \frac{1}{\lambda_+ - n_m} \right)_v \left( \sum_n \frac{1}{\lambda_- - n_m} \right)_v = \frac{V^4 \Sigma^4}{16(\nu + 1)^2} \int_0^1 ds e^{2V F_\pi^2 \mu_{iso}^2 s^2} 2^{2v+5} = \frac{V^2 \Sigma^4}{64 F_\pi^2 \mu_{iso}^2 (\nu + 1)^2} \Gamma(\nu + 3) - \Gamma(\nu + 3, -2V F_\pi^2 \mu_{iso}^2). \]
\( (16) \)
Apart from the coefficients, the difference to the massless first order rule is given by the higher power of the parameter $\nu$ in the first line of Eq. (16). These powers are indeed a distinctive feature for any massless sum rule of a given order (third order sum rules for instance carry a power of $s^{2(2n+1)}$). Concerning the evaluation of the integral we can make exactly the same remarks as for the first order rule, they are easily calculated in a fixed topological sector. For completeness we give again the result at $v = 0$

$$
\left(\left\{\sum_{n} \frac{1}{\lambda_{n}^{2}}\right\}\right)_{v} = \left(\sum_{n} \frac{1}{\lambda_{n}^{2}}\right)_{0} = \frac{V^{2} \Sigma^{4}}{32 F_{\pi}^{2} \mu_{iso}^{4}} \left(1 - 2 V F_{\pi}^{2} \mu_{iso}^{2} (1 + V F_{\pi}^{2} \mu_{iso}^{2}) e^{2 V F_{\pi}^{2} \mu_{iso}^{2} - 1}\right).
$$

(17)

Again, in the limit $\mu_{iso} \to 0$ where $\lambda_{+}$ and $\lambda_{-}$ become degenerate the formula reproduces the result of Ref. [2]. In addition, this second order sum rule can be compared to the results of Ref. [11]. There the mixed two point spectral correlation function

$$
\rho^{(2)}(\lambda_{1}, \lambda_{2}, m_{u}, m_{d}, i \mu_{iso}) = \left(\sum_{n} \delta(\lambda_{1} - \lambda_{+}) \sum_{n} \delta(\lambda_{2} - \lambda_{-})\right) - \left(\sum_{n} \delta(\lambda_{1} - \lambda_{-}) \sum_{n} \delta(\lambda_{2} - \lambda_{+})\right)
$$

(18)

is derived with the replica method. By integrating out $\lambda_{1}$ and $\lambda_{2}$ in

$$
\int d\lambda_{1} \int d\lambda_{2} \rho^{(2)}(\lambda_{1}, \lambda_{2}, m_{u}, m_{d}, i \mu_{iso}) = \frac{V^{4} \Sigma^{4} \sum_{n}^{1}}{4 x^{4}} \int_{0}^{1} \frac{J_{v}^{1}(2)}{s^{2 v}} I_{v}(s_{xd}) I_{v}(s_{xd}) I_{v}(s_{xd}) I_{v}(s_{xd})
$$

(19)

we should reproduce our subtracted mixed sum rule of Eq. (14). Indeed, we do find exact agreement in the massless limit.

With the first choice above $\frac{\partial^{2} Z_{v}}{\partial m_{u}^{2}}$ in $Z_{v}$, we have means to extract the correlation between eigenvalues of one series $\lambda_{+}$ (or equivalently $\lambda_{-}$). More precisely, from taking the second derivative with respect to $m_{u}$ and subtracting the zero mode contributions and the known first order terms, we can deduce a formula for

$$
\left(\left\{\sum_{n} \frac{1}{\lambda_{n}^{2} + m_{u}^{2}}\right\} - \sum_{n} \frac{1}{\lambda_{n}^{2} + m_{u}^{2}}\right)_{v} = \frac{V^{4} \Sigma^{4}}{4 x^{4}} \int_{0}^{1} \frac{J_{v}^{1}(2)}{s^{2 v}} I_{v}(s_{xd}) I_{v}(s_{xd}) I_{v}(s_{xd}) I_{v}(s_{xd})
$$

(20)

Again the formula simplifies drastically in the massless limit where we get

$$
\left(\left\{\sum_{n} \frac{1}{\lambda_{n}^{2} + m_{u}^{2}}\right\} - \sum_{n} \frac{1}{\lambda_{n}^{2} + m_{u}^{2}}\right)_{v} = \frac{V^{4} \Sigma^{4}}{16(\nu + 1)(\nu + 2)} \int_{0}^{1} \frac{J_{v}^{1}(2)}{s^{2 v + 5}} I_{v}(s_{xd}) I_{v}(s_{xd}) I_{v}(s_{xd}) I_{v}(s_{xd})
$$

(21)

Let us remark here that at finite $\mu_{iso}$ we cannot extract a separate sum rule for the fourth order term $(\sum_{n} 1/\lambda_{n}^{4})_{v}$ from the partition function in Eq. (5). Since the $u$ and the $d$ quark obey two different eigenvalue equations, each of them behaves effectively like a one flavor system. In order to isolate the fourth order term above we would need another flavor with the same eigenvalues $\lambda_{+}$ but different mass $\tilde{m}_{u}$. At vanishing isospin chemical potential however, where $\lambda_{+}$ and $\lambda_{-}$ become degenerate a fourth order term can be calculated from the difference of the two second order sum rules described here and it again coincides with the result of Ref. [2].

It is in principle possible to continue this series and to calculate sum rules of any order by taking higher and higher derivatives, and subtracting the known lower order terms. The computation however becomes tedious and its usefulness doubtful.

3. In this Letter we have derived a set of spectral sum rules for a system of two quarks coupled to an imaginary isospin chemical potential from the finite volume partition function. This can be seen as an extension of the pioneering work of Leutwyler and Smilga in Ref. [2]. The sum rules derived here inherit the $F_{\nu}$ dependence from the partition function and provide means to determine this important constant from the spectrum of the Dirac operator introduced here.

Acknowledgements

The author thanks P.H. Damgaard for useful discussions and a careful reading of the manuscript. Part of this work was accomplished during the Doctoral Training Programme 2006 at ECT* in Trento. Financial support from ECT* is gratefully acknowledged.

References

Constraints on minimal SUSY models with warm dark matter neutralinos

Graciela Gelmini, Carlos E. Yaguna *

Department of Physics and Astronomy, UCLA, 475 Portola Plaza, Los Angeles, CA 90095, USA

Received 19 July 2006; received in revised form 22 September 2006; accepted 28 October 2006

Available online 14 November 2006

Editor: T. Yanagida

Abstract

If the energy density of the Universe before nucleosynthesis is dominated by a scalar field $\phi$ that decays and reheats the plasma to a low reheating temperature $T_{RH}$, neutralinos may be warm dark matter particles. We study this possibility and derive the conditions on the production mechanism and on the supersymmetric spectrum for which it is viable. Large values of the $\mu$ parameter and of the slepton masses are characteristic features of these models. We compute the expected direct detection cross sections and point out that split-SUSY provides a natural framework for neutralino warm dark matter.

© 2006 Elsevier B.V. All rights reserved.
PACS: 95.35.+d; 98.80.Cq; 12.60.Jv; 14.80.Ly

Thermally produced neutralinos are typical and well-motivated cold dark matter candidates (see e.g. [1]). They are produced by scatterings in the thermal bath, then reach equilibrium and finally decouple when non-relativistic. Non-thermally produced neutralinos, on the contrary, are usually produced as relativistic final states in the decay of heavy particles and may never reach chemical or kinetic equilibrium. As a result, they are not necessarily cold. Indeed, if they manage to keep most of their initial energy they might behave as warm dark matter suppressing the evolution of small-scale structures in the Universe.

Although consistent with the observations of the large scale structure of the Universe and the cosmic microwave background radiation anisotropies, cold dark matter models seem to have problems at galactic scales. They not only tend to form cuspy structures in the halo density profile [2] but also predict a large overabundance of small halos near galaxies such as our own [3]. Warm dark matter, with its larger free-streaming scale, may solve these problems while maintaining the celebrated success of cold dark matter models at large scales [4].

Neutralinos as warm dark matter candidates were initially discussed in Refs. [5,6]. In this Letter we will extend, in several ways, the analysis presented in those references. Ref. [5] concentrated on the cosmological aspects of this possibility, while we study a particular particle model. Reheating temperatures smaller than 5 MeV or 2 MeV as those found in Ref. [6] are hardly compatible with the standard cosmological scenario, which require $T_{RH} > 4$ MeV [7]. We, instead take $T_{RH} \approx 10$ MeV as a characteristic value and find the conditions on the initial energy and on the supersymmetric spectrum for which the neutralino is a warm dark matter particle. In doing so, we ignore, following Ref. [8], the naturalness argument giving up the idea that SUSY stabilizes the weak scale, notion that is the basis for split-SUSY models. We also discuss the non-thermal production of neutralinos, as well as the implications for direct dark matter searches of neutralino warm dark matter.

We concentrate on non-standard cosmological models (see for example Refs. [9–11]) in which the late decay of a scalar field $\phi$ reheats the Universe to a low reheating temperature $T_{RH}$, smaller than the standard neutralino freeze out temperature. Such scalar fields are common in superstring models where they appear as moduli fields. These fields get mass at the low energy supersymmetry breaking scale, typically of the order of $10^2–10^3$ TeV. The decay of $\phi$ into radiation increases the entropy, diluting the neutralino number density. The decay of $\phi$ into supersymmetric particles, which eventually decay into neutralinos, increases the neutralino number density. We denote

* Corresponding author.
E-mail address: cyaguna@gmail.com (C.E. Yaguna).
by $b$ the net number of neutralinos produced on average per $\phi$ decay. The number $b$ is highly model-dependent, so is the $\phi$ field mass $m_\phi$. They are determined by the physics of the hidden sector, by the mechanism of supersymmetry breaking, and in superstring-inspired models by the compactification mechanism [9–16].

The coupling of the $\phi$ to the gravitino arises from the term $e^{K/2} \bar{\psi}_\nu \sigma^{\mu \nu} \psi_\nu$, where $K$ is the Kähler potential. Under the assumption of $m_\phi \gg m_{3/2}$ the $\phi$ decay through mixing with the field responsible for supersymmetry breaking was found to be important [12], unless a symmetry of $\phi$ is preserved at the vacuum. If $m_\phi$ is larger than twice the gravitino mass $m_{3/2}$, the decay mode $\phi \rightarrow \psi_{3/2} \psi_{3/2}$ of the moduli field into two gravitinos is present with branching ratio of order 0.01 (see Ref. [13], which correct previous claims [14] that this branching would be chirally suppressed by a factor $(m_{3/2}/m_\phi)^2$. Even if this decay is kinematically forbidden, the decay of $\phi$ into its supersymmetric partner and a gravitino may happen as long as $m_\phi > m_{3/2}$ [15]. Gravitinos must then decay rapidly not to disrupt nucleosynthesis (so $m_{3/2} \gtrsim 100$ TeV), and they produce comparable amounts of normal particles and their supersymmetric partners. If $m_\phi \gg m_{3/2}$, the gravitino decays during the radiation dominated epoch after the decay of the $\phi$ field (here we do not address this case and we focus on neutralino production during $\phi$ decay). When $m_\phi$ and $m_{3/2}$ are of the same order of magnitude, we can consider the gravitino decay as part of the $\phi$ decay, since they happen almost simultaneously. In this case, depending on how important the direct decay of $\phi$ into supersymmetric particles other than the $\psi_{3/2}$ is, $b$ can typically be 0.01–0.001, but not smaller.

If instead $m_\phi < m_{3/2}$ more possibilities open up. The yield per $\phi$ decay $b$ can still be of order one but it can also be much smaller. Supergravity models with chiral superfields $\Phi_I$ are specified in terms of the Kähler potential $K(\Phi_I, \bar{\Phi}_I)$, the superpotential $W(\Phi_I)$, and the gauge kinetic function $f_{\alpha \beta}(\Phi_I)$. Specific relations between the $\phi$ mass $m_\phi$, the gravitino mass $m_{3/2}$, and the gaugino mass $m_{1/2}$ arise as a consequence of the relations $m_{3/2} = (e^{K/2} W)$, $m_{1/2} = (F^{-1} \partial_\Phi \ln f)$, and $m_\phi = (\phi^2 V/\partial \phi^2)$. With appropriate choices of $K$, $W$, and $f$, the hierarchy $m_{3/2} \gtrsim m_\phi \gg m_{1/2}$ may be achieved. Here $V$ is the scalar potential and $F^I$ is the $F$-term of the chiral superfield $\Phi_I$.

One finds that $b \simeq O(1)$, for example, when the main $\phi$ decay mode is through a coupling of the type $h_\phi \phi^2$ with a chiral matter supermultiplet $\psi$ in the superpotential $W$. This leads to comparable decay rates of $\phi$ into the scalar and fermionic components of $\psi$ (which are supersymmetric partners). On the other hand, it is possible that the $\phi$ field decays mostly into Higgs fields, or gauge fields ($W$'s, $Z$’s, photons, gluinos). In this case $b$ can be very small $10^{-2}$, $10^{-4}$, $10^{-6}$, etc. [10,11]. For example, the coupling of $\phi$ to the gauge bosons arises from the term $\partial f_{\alpha \beta}(\Phi_I) F^\alpha_{\mu \nu} F^{\mu \nu \beta}$ and with non-minimal kinetic terms $f_{\alpha \beta}$ may contain $\phi$. The $\phi$ decay width into gauge bosons is then $\Gamma_{\phi} \sim \lambda_8 m_\phi^2 / M_P^2$ with $\lambda_8 \equiv \partial \ln f$, while that into gauginos is $\Gamma_{\phi} \sim \lambda_8^2 m_\phi / M_P^2$ with $\lambda_8 \equiv m_{1/2} \partial \ln (F\phi / \partial \phi^2)$. Thus in principle the gaugino coupling may be suppressed relative to the coupling to gauge bosons.

Here we consider $b$ and $m_\phi$ as free parameters.

To account for the dark matter of the Universe, the neutralino relic density must be in agreement with the observed dark matter density. In low $T_{RH}$ cosmological models essentially all neutralinos can have the dark matter density provided the right combination of the following two parameters can be achieved in the high energy theory: the reheating temperature $T_{RH}$ and the ratio of the number of neutralinos produced per $\phi$ decay over the $\phi$ field mass, i.e. $b/m_\phi$ [16,17]. We will find later the values of $T_{RH}$ and $b/m_\phi$ for which the neutralinos we are interested in have the right dark matter density.

A crucial quantity that distinguishes warm from cold dark matter is the free-streaming length at matter-radiation equality $\lambda_{FS}$, which depends on the parameter $r_X = a(t) p_{X}(t)/\rho_X$ (see for example Refs. [5,6] and references therein). This parameter would be the present characteristic speed of neutralinos of mass $m_\chi$, if their momentum $p_X$ only redshifted from neutralino production onwards ($a(t)$ is the scale factor with $a_0 = 1$). During the cosmic evolution $r_X$ is constant. Structures smaller than $\lambda_{FS}$ are damped because neutralinos can freely flow out of them. $N$-body simulations have shown that to explain the lack of substructure in the local group, $\lambda_{FS}$ should be of order 0.1 Mpc [4], thus $r_X \simeq 10^{-7}$.

The parameter $r_X$ also determines (for neutralinos which are relativistic at production [18]) the neutralino phase-space density $Q$. This is defined as $Q = \rho / (\langle v^2 \rangle^{3/2}$ where $\rho$ is the neutralino energy density and $\langle v^2 \rangle$ is the mean square value of the particle velocity. In the absence of dissipation, the coarse-grained phase-space density (the quantity that can actually be observed) can only decrease from its primordial value. The observation of dwarf-spheroidal galaxies places a lower bound on $Q$ which translates into an upper bound on $r_X$ of about $2.5 \times 10^{-7}$ [5]. In what follows we impose the condition $r_X = 10^{-7}$ to our model.

Neutralinos are produced in $\phi$ decays, thus their average initial energy is $E_I = m_\phi / N$, where $N$ is a number which depends on the production spectrum. We expect $N$ to be of order one and require $E_I \gg m_\phi$ so that neutralinos are relativistic at production. Thus, $p_X \simeq E_I$ at the moment of $\phi$ decay. Assuming an instantaneous $\phi$ decay at $T_{RH}$, with no subsequent entropy production, the scale factor at decay is $a = T_0 / T_{RH}$, where $T_0$ is the present photon temperature, and the parameter $r_X$ in our model is $r_X = (T_0 E_I) / (T_{RH} m_\chi)$, i.e.

$$r_X \simeq 10^{-7} \left( \frac{2.3}{N} \right) \left( \frac{m_\phi}{10^3 \text{ TeV}} \right) \left( \frac{10 \text{ MeV}}{T_{RH}} \right) \left( \frac{100 \text{ GeV}}{m_\chi} \right).$$

(1)

Thus, the condition $r_X = 10^{-7}$ fixes $m_\phi$ in terms of the reheating temperature and the neutralino mass

$$m_\phi = 10^3 \text{ TeV} \left( \frac{N}{2.3} \right) \left( \frac{m_\chi}{100 \text{ GeV}} \right) \left( \frac{T_{RH}}{10 \text{ MeV}} \right).$$

(2)

In the estimation of $r_X$ we have assumed that the neutralinos do not lose their energy in scattering processes with the thermal bath. To ensure this condition we will simply require that the interactions of neutralinos with the particles present in the plasma, $e^\pm, \nu, \gamma$, are out of equilibrium. Neutralino interactions are determined by the neutralino composition in terms of
gauge eigenstates. The lightest neutralino ($\chi$) is a linear superposition of bino ($\tilde{B}$), wino ($\tilde{W}$) and higgsino ($\tilde{H}$) states,

$$\chi = N_{11} \tilde{B} + N_{12} \tilde{W} + N_{13} \tilde{H} + N_{14} \tilde{H}^\ast.$$

According to the dominant term in Eq. (3), $\chi$ is classified as bino-like, wino-like, or higgsino-like.

Wino and higgsino-like neutralinos are always accompanied by a chargino state (\tilde{\chi}^\pm) with a mass difference $\Delta m = m_{\chi^\pm} - m_\chi \approx \sin^2(\theta_W) (M_Z/m_\chi)^2 m_\chi$ for higgsinos and much smaller for winos. When $m_\chi \Delta m < E_\nu T$ relativistic neutralinos scatter inelastically (e.g. $\chi e^- \rightarrow \chi^0 e^-$) besides scattering elastically (e.g. $\chi \nu \rightarrow \chi \nu$) on electrons and neutrinos. Since it is not possible to simultaneously suppress the elastic and the inelastic scattering rates, wino and higgsino-like neutralinos easily lose their initial energy. In Ref. [6] was indeed pointed-out that a wino-like neutralino cannot be warm dark matter and that a higgsino-like neutralino would only be viable if $T_{\text{RH}} < 2$ MeV were allowed. Bino-like neutralinos, on the other hand, couple to electrons and neutrinos mainly through slepton exchange and typically have small interaction cross sections. In general, there are no charginos close in mass to bino-like neutralinos and inelastic scatterings can be suppressed. We will therefore focus on bino-like neutralinos in the following.

In the approximation $m_\ell \gtrsim 6T E_\nu$, which is guaranteed by the lower bound found below, the cross section for elastic bino–electron scattering mediated by a slepton of mass $m_\ell$ is estimated to be

$$\sigma v \sim \frac{g^4}{16\pi} \frac{E_\nu E_\ell}{m_\ell^4}.$$  \hspace{1cm} (4)

Here $g$ is the weak hypercharge coupling constant, $E_\ell \sim 3T$ is the energy of the incoming thermal-bath neutrino or electron, and $E_\nu$ is the energy of the outgoing lepton. We require that $\chi-\ell$ interactions are out-of-equilibrium, i.e. that the interaction rate is smaller than the expansion rate of the Universe, $\Gamma \leq \sigma v < H$, where $n$ is the thermal neutrino or electron number density and $H$ is the Hubble parameter. Since the interaction rate goes as $T^5$ and $H$ as $T^2$, if the interactions are out-of-equilibrium at $T_{\text{RH}}$ they will also be out of equilibrium at any later temperature $T < T_{\text{RH}}$. Replacing $E_\ell$ in Eq. (4) by its maximum value $E_\ell = m_\phi / N$, we obtain that neutralino interactions are out of equilibrium at $T_{\text{RH}}$ (and subsequently if

$$0.1 \left( \frac{m_\phi}{N \times 10^3 \text{ TeV}} \right) \left( \frac{T_{\text{RH}}}{10 \text{ MeV}} \right)^2 \left( \frac{40 \text{ TeV}}{m_\ell} \right)^4 < 1.$$  \hspace{1cm} (5)

Sleptons therefore must be heavy.

Due to the structure of the neutralino mass matrix, bino-like neutralinos always have a small higgsino component. As a result, they also couple to fermions through $Z$ exchange. If sleptons are heavy, the $Z$-exchange diagram could give the dominant contribution to the scattering rate. In this case the neutralino cross section is

$$\sigma v \sim \frac{g^4}{32\pi} \tan^2(\theta_W) \cos^2 \theta_W \left( \frac{2E_\ell}{\mu^4} \right) E_\ell,$$

where $\theta_W$ is the weak mixing angle and $\tan \beta = v_2/v_1$ is the ratio between the vacuum expectation values of the two Higgs doublets. Eq (6) holds for $m_\ell^2 \lesssim 6T E_\ell$ and $m_\ell^2 \lesssim 6T E_\nu$. The latter condition, after using Eq. (2), becomes $T_{\text{RH}}/10$ MeV $\gtrsim 1.3(m_\chi/100 \text{ GeV})$. The condition $\Gamma < H$ implies for the interaction in Eq. (6) (again replacing $E_\ell$ by its maximum possible value, i.e. $E_\ell$)

$$0.1 \left( \frac{m_\phi}{N \times 10^3 \text{ TeV}} \right) \left( \frac{T_{\text{RH}}}{10 \text{ MeV}} \right)^2 \left( \frac{8 \text{ TeV}}{\mu} \right)^4 < 1.$$  \hspace{1cm} (7)

If $m_\chi^2 \gtrsim 6T E_\ell$ (i.e. $T_{\text{RH}}/10$ MeV $\gtrsim 1.3(m_\chi/100 \text{ GeV})$, the right-hand side of Eq. (6) and the left-hand side of Eq. (7) must be multiplied by the factor $m_\chi^4/3T_{\text{RH}}^2 E_\ell$, which yields

$$0.1 \left( \frac{m_\chi}{100 \text{ GeV}} \right)^2 \left( \frac{T_{\text{RH}}}{10 \text{ MeV}} \right) \left( \frac{6 \text{ TeV}}{\mu} \right)^4 < 1.$$  \hspace{1cm} (8)

In any event, a large $\mu$ parameter value is required.

Replacing Eq. (2) into Eqs. (5) and (7) and directly from Eq. (8) we obtain bounds of the slepton masses and the parameter $\mu$ which depend only on the neutralino mass and the reheating temperature,

$$m_\ell > 18 \text{ TeV} \left( \frac{m_\chi}{100 \text{ GeV}} \right)^{1/4} \left( \frac{T_{\text{RH}}}{10 \text{ MeV}} \right)^{3/4},$$

$$\mu > 4 \text{ TeV} \left( \frac{m_\chi}{100 \text{ GeV}} \right)^{1/4} \left( \frac{T_{\text{RH}}}{10 \text{ MeV}} \right)^{3/4},$$

from Eq. (7) or

$$\mu > 3 \text{ TeV} \left( \frac{m_\chi}{100 \text{ GeV}} \right)^{1/2} \left( \frac{T_{\text{RH}}}{10 \text{ MeV}} \right)^{1/4}$$

when Eq. (8) holds instead.

Using the cosmological lower bound $T_{\text{RH}} > 4$ MeV we obtain $m_\ell > 9$ TeV ($m_\chi/100$ GeV)$^{1/4}$ and $\mu > 2.0$ TeV ($m_\chi/100$ GeV)$^{1/4}$ or $\mu > 2$ TeV ($m_\chi/100$ GeV)$^{1/2}$.

Let us now return to the density requirement. In general there are four different ways in which the neutralino relic density $\Omega_{\chi h^2}$ depends on $T_{\text{RH}}$, in the low reheating cosmologies we are considering. The four cases have thermal or non-thermal dominant neutralino production, with or without chemical equilibrium [16,17]. The case of interest here is that of a very low reheating temperature and high standard neutralino relic density, corresponding to small neutralino cross sections for which the production is non-thermal without chemical equilibrium [16]. In this case neutralinos are produced in the decay of the $\phi$ field into supersymmetric particles, with subsequent fast decay of all other supersymmetric particles into neutralinos, and the production is not compensated by annihilation. This implies

$$\Omega_{\chi h^2} \lesssim 2 \times 10^3 b \left( \frac{10^3 \text{ TeV}}{m_\phi} \right) \left( \frac{m_\chi}{10^2 \text{ GeV}} \right) \left( \frac{T_{\text{RH}}}{10 \text{ MeV}} \right).$$  \hspace{1cm} (12)

This equation fixes the value of $b/m_\phi$ required for neutralinos to have the dark matter density, $\Omega_{\text{DM}} h^2 = 0.11,^1$

$$b \left( \frac{m_\phi}{10^3 \text{ TeV}} \right) = 5 \times 10^{-7} \left( \frac{m_\chi}{100 \text{ GeV}} \right) \left( \frac{10 \text{ MeV}}{T_{\text{RH}}} \right).$$  \hspace{1cm} (13)

$^1$ $\Omega_{\text{DM}} h^2 = 0.109^{+0.003}_{-0.006}$ was obtained for a  $\Lambda$CDM model with scale-invariant primordial perturbation spectrum through a global fit of cosmic microwave background, supernovae, and large scale structure data [20].
The prospects for direct detection of bino-like neutralinos depend on the neutralino–nucleon cross section, which in turn is determined by the diagrams mediated by the heavy neutral Higgs boson (H) and squarks. We show in Fig. 2 the spin-independent neutralino–proton cross section for different values of the squark mass. The models solve a number of phenomenological problems associated with ordinary supersymmetry: they alleviate proton decay, increase the mass of the Higgs boson, and minimize the flavor and CP problems. Thus, a split-SUSY model with a bino-like LSP and \( \mu \gtrsim 10 \) TeV would be a perfectly viable framework for neutralino warm dark matter.

In standard Cosmological models, a bino-like neutralino is allowed in split-SUSY models only if it is almost degenerate in mass with other neutralino and chargino states \((M_1 \simeq M_2 \text{ or } M_1 \simeq \mu)\) and coannihilation effects determine its relic density [22]. These constraints, however, do not hold in our model. By assuming a non-standard cosmology which allows for non-thermal production of neutralinos we have effectively enlarged the viable parameter space of split-SUSY models.

In split-SUSY models all scalar superpartners are usually heavy, but our conditions require only that the sleptons be heavy, not necessarily the squarks. A model with light squarks would certainly be appealing from a phenomenological point of view. The most general low energy spectrum of masses below a TeV of a model compatible with neutralino warm dark matter would consist of: the lightest neutralino, the next-to-lightest neutralino, the lightest chargino, the three Higgs bosons, the gluino and the squarks. Such low energy spectrum is sufficiently rich to give significant signals at the LHC, particularly for light gluinos and squarks. Besides, the absence of sleptons in the low energy spectrum would be a crucial test for models with neutralino warm dark matter.
patible with neutralino warm dark matter, each with a different value of $m_A$, the pseudoscalar Higgs boson mass. The mass of heavy boson $H$ is close to $m_A$. The lines are not continued to smaller squark masses because they enter into a region incompatible with the Higgs mass bound [23] or $b \to s \gamma$ [24]. The figure shows that the expected neutralino-nucleon scattering cross sections in neutralino warm dark matter models are small, between $10^{-46}$ cm$^2$ and $10^{-49}$ cm$^2$. Hence, the expected spin-independent neutralino-proton cross sections are below the reach of the largest dark matter detectors envisioned at present, i.e. Zeplin IV-MAX and Xenon 1 Ton [25].

In this Letter we studied non-thermally produced neutralinos that are viable warm dark matter candidates, in low reheating temperature cosmological scenarios. We have shown that models compatible with neutralino warm dark matter satisfy a number of non-trivial requirements, some of them testable: the reheating temperature is low, the mass of the decaying scalar field which reheats the Universe and the number of neutralinos it produces per decay are constrained, the neutralino is bino-like, slepton masses and the $\mu$ parameter are both large and the neutralino interacts too weakly to be observed in direct dark matter experiments in the near future. To validate the idea of neutralino warm dark matter, therefore, cosmological observations as well as accelerator and dark matter searches will be essential.

Acknowledgements

This work was supported in part by the US Department of Energy Grant DE-FG03-91ER40662, Task C and NASA grants NAG5-13399 and ATP03-0000-0057. We thank P. Gondolo for helpful suggestions.

References

Measurements of branching fractions for inclusive $\bar{K}^0/K^0$ and $K^*(892)^\mp$ decays of neutral and charged $D$ mesons

BES Collaboration

M. Ablikim\textsuperscript{a}, J.Z. Bai\textsuperscript{a}, Y. Ban\textsuperscript{i}, J.G. Bian\textsuperscript{a}, H.F. Chen\textsuperscript{p}, H.S. Chen\textsuperscript{a}, H.X. Chen\textsuperscript{a}, J.C. Chen\textsuperscript{a}, Jin Chen\textsuperscript{a}, Y.B. Chen\textsuperscript{a}, S.P. Chi\textsuperscript{b}, Y.P. Chu\textsuperscript{a}, X.Z. Cui\textsuperscript{a}, Y.S. Dai\textsuperscript{i}, L.Y. Diao\textsuperscript{i}, Z.Y. Deng\textsuperscript{a}, Q.F. Dong\textsuperscript{o}, S.X. Du\textsuperscript{a}, J. Fang\textsuperscript{a}, S.S. Fang\textsuperscript{b}, C.D. Fu\textsuperscript{a}, C.S. Gao\textsuperscript{a}, Y.N. Gao\textsuperscript{o}, S.D. Gu\textsuperscript{a}, Y.T. Gu\textsuperscript{d}, Y.N. Guo\textsuperscript{a}, Y.Q. Guo\textsuperscript{a}, K.L. He\textsuperscript{a}, M. He\textsuperscript{m}, Y.K. Heng\textsuperscript{a}, H.M. Hu\textsuperscript{a}, T. Hu\textsuperscript{a}, G.S. Huang\textsuperscript{a,1}, X.T. Huang\textsuperscript{m}, X.B. Ji\textsuperscript{a}, X.S. Jiang\textsuperscript{a}, X.Y. Jiang\textsuperscript{c}, J.B. Jiao\textsuperscript{m}, D.P. Jin\textsuperscript{a}, S. Jin\textsuperscript{a}, Yi Jin\textsuperscript{b}, Y.F. Lai\textsuperscript{a}, G. Li\textsuperscript{b}, H.B. Li\textsuperscript{a}, H.H. Li\textsuperscript{a}, J. Li\textsuperscript{i}, R.Y. Li\textsuperscript{a}, S.M. Li\textsuperscript{a}, W.D. Li\textsuperscript{a}, W.G. Li\textsuperscript{a}, X.L. Li\textsuperscript{a}, X.N. Li\textsuperscript{a}, X.Q. Li\textsuperscript{k}, Y.L. Li\textsuperscript{d}, Y.F. Liang\textsuperscript{n}, H.B. Liao\textsuperscript{a}, B.J. Liu\textsuperscript{a}, C.X. Liu\textsuperscript{a}, F. Liu\textsuperscript{f}, Fang Liu\textsuperscript{a}, H.H. Liu\textsuperscript{a,*}, H.M. Liu\textsuperscript{a}, J. Liu\textsuperscript{i}, J.B. Liu\textsuperscript{a}, J.P. Liu\textsuperscript{q}, Q. Liu\textsuperscript{a}, R.G. Liu\textsuperscript{a}, Z.A. Liu\textsuperscript{a}, Y.C. Lou\textsuperscript{c}, F. Lu\textsuperscript{a}, G.R. Lu\textsuperscript{c}, J.G. Lu\textsuperscript{a}, C.L. Luo\textsuperscript{j}, F.C. Ma\textsuperscript{i}, H.L. Ma\textsuperscript{a}, L.L. Ma\textsuperscript{a}, Q.M. Ma\textsuperscript{a}, X.B. Ma\textsuperscript{a}, Z.P. Mao\textsuperscript{a}, X.H. Mo\textsuperscript{a}, J. Nie\textsuperscript{a}, H.P. Peng\textsuperscript{p,2}, R.G. Ping\textsuperscript{a}, N.D. Qi\textsuperscript{a}, H. Qin\textsuperscript{a}, J.F. Qiu\textsuperscript{a}, Z.Y. Ren\textsuperscript{a}, G. Rong\textsuperscript{a}, L.Y. Shan\textsuperscript{a}, L. Shang\textsuperscript{a}, C.P. Shen\textsuperscript{a}, D.L. Shen\textsuperscript{a}, X.Y. Shen\textsuperscript{a}, H.Y. Sheng\textsuperscript{a}, H.S. Sun\textsuperscript{a}, J.F. Sun\textsuperscript{a}, S.S. Sun\textsuperscript{a}, Y.Z. Sun\textsuperscript{a}, Z.J. Sun\textsuperscript{a}, Z.Q. Tan\textsuperscript{d}, X. Tang\textsuperscript{a}, G.L. Tong\textsuperscript{a}, D.Y. Wang\textsuperscript{a}, L. Wang\textsuperscript{a}, L.L. Wang\textsuperscript{a}, L.S. Wang\textsuperscript{a}, M. Wang\textsuperscript{a}, P. Wang\textsuperscript{a}, P.L. Wang\textsuperscript{a}, W.F. Wang\textsuperscript{a,3}, Y.F. Wang\textsuperscript{a}, Z. Wang\textsuperscript{a}, Z.Y. Wang\textsuperscript{a}, Zhe Wang\textsuperscript{a}, Zheng Wang\textsuperscript{b}, C.L. Wei\textsuperscript{a}, D.H. Wei\textsuperscript{a}, N. Wu\textsuperscript{a}, X.M. Xia\textsuperscript{a}, X.X. Xie\textsuperscript{a}, G.F. Xu\textsuperscript{a}, X.P. Xu\textsuperscript{f}, Y. Xu\textsuperscript{k}, M.L. Yan\textsuperscript{p}, H.X. Yang\textsuperscript{a}, Y.X. Yang\textsuperscript{c}, M.H. Ye\textsuperscript{b}, Y.X. Ye\textsuperscript{p}, Z.Y. Yi\textsuperscript{a}, G.W. Yu\textsuperscript{a}, C.Z. Yuan\textsuperscript{a}, J.M. Yuan\textsuperscript{a}, Y. Yuan\textsuperscript{a}, S.L. Zhang\textsuperscript{a}, Y. Zeng\textsuperscript{g}, Yu Zeng\textsuperscript{a}, B.X. Zhang\textsuperscript{a}, B.Y. Zhang\textsuperscript{a}, C.C. Zhang\textsuperscript{a}, D.H. Zhang\textsuperscript{a}, H.Q. Zhang\textsuperscript{a}, H.Y. Zhang\textsuperscript{a}, J.W. Zhang\textsuperscript{a}, J.Y. Zhang\textsuperscript{a}, S.H. Zhang\textsuperscript{a}, X.M. Zhang\textsuperscript{a}, X.Y. Zhang\textsuperscript{m}, Yiyun Zhang\textsuperscript{n}, Z.P. Zhang\textsuperscript{p}, D.X. Zhao\textsuperscript{a}, J.W. Zhao\textsuperscript{a}, M.G. Zhao\textsuperscript{a}, PP. Zhao\textsuperscript{a}, W.R. Zhao\textsuperscript{a}, Z.G. Zhao\textsuperscript{a,4}, H.Q. Zheng\textsuperscript{l}, J.P. Zheng\textsuperscript{a}, Z.P. Zheng\textsuperscript{a}, L. Zhou\textsuperscript{a}, N.F. Zhou\textsuperscript{a,4}, K.J. Zhu\textsuperscript{a}, Q.M. Zhu\textsuperscript{a}, Y.C. Zhu\textsuperscript{a}, Y.S. Zhu\textsuperscript{a}, Yingchun Zhu\textsuperscript{a,2}, Z.A. Zhu\textsuperscript{a}, B.A. Zhuang\textsuperscript{a}, X.A. Zhuang\textsuperscript{a}, B.S. Zou\textsuperscript{a}

\textsuperscript{a} Institute of High Energy Physics, Beijing 100049, People’s Republic of China
\textsuperscript{b} China Center for Advanced Science and Technology (CCAST), Beijing 100080, People’s Republic of China
\textsuperscript{c} Guangxi Normal University, Guilin 541004, People’s Republic of China
\textsuperscript{d} Guangxi University, Nanning 530004, People’s Republic of China
\textsuperscript{e} Henan Normal University, Xinxiang 453002, People’s Republic of China
\textsuperscript{f} Huazhong Normal University, Wuhan 430079, People’s Republic of China
\textsuperscript{g} Hunan University, Changsha 410082, People’s Republic of China
\textsuperscript{h} Jinan University, Jinan 250022, People’s Republic of China
\textsuperscript{i} Liaoning University, Shenyang 110036, People’s Republic of China
\textsuperscript{j} Nanjing Normal University, Nanjing 210097, People’s Republic of China
\textsuperscript{k} Nankai University, Tianjin 300071, People’s Republic of China
\textsuperscript{l} Peking University, Beijing 100871, People’s Republic of China
\textsuperscript{m} Shandong University, Jinan 250100, People’s Republic of China
\textsuperscript{n} Sichuan University, Chengdu 610064, People’s Republic of China
\textsuperscript{o} Tsinghua University, Beijing 100084, People’s Republic of China
\textsuperscript{p} University of Science and Technology of China, Hefei 230026, People’s Republic of China
\textsuperscript{q} Wuhan University, Wuhan 430072, People’s Republic of China

0370-2693/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.physletb.2006.10.057
Abstract

Using the data sample of about 33 pb\(^{-1}\) collected at and around 3.773 GeV with the BES-II detector at the BEPC collider, we have studied inclusive \(\bar{K}^0/K^0\) and \(K^{*\pm}(892)\) decays of \(D^0\) and \(D^+\) mesons. The branching fractions for the inclusive \(\bar{K}^0/K^0\) and \(K^{*\pm}(892)\) decays are measured to be \(BF(D^0 \rightarrow \bar{K}^0/K^0 X) = (47.6 \pm 4.8 \pm 3.0)\%\), \(BF(D^+ \rightarrow K^{*\pm}/K^0 X) = (60.5 \pm 5.5 \pm 3.3)\%\), \(BF(D^0 \rightarrow K^{*-} X) = (15.3 \pm 8.3 \pm 1.9)\%\) and \(BF(D^+ \rightarrow K^{*-} X) = (5.7 \pm 5.2 \pm 0.7)\%\). The upper limits of the branching fractions for the inclusive \(K^{*\pm}(892)\) decays are set to be \(BF(D^0 \rightarrow K^{*+} X) < 3.6\%\) and \(BF(D^+ \rightarrow K^{*+} X) < 20.3\%\) at 90\% confidence level.

© 2006 Elsevier B.V. All rights reserved.

1. Introduction

Measurements of the branching fractions for inclusive \(\bar{K}^0/K^0\) and \(K^{*\pm}\) decays of \(D\) mesons are important in understanding of the \(D\) decay mechanisms. Comparing the measured inclusive branching fraction with the sum of those for the exclusive decays [1] provides some information about the decay modes which have not been observed yet. In addition, measurements of the branching fractions for the inclusive \(K^{*-}\) and \(K^{*+}\) decays of \(D\) mesons can also help us to study the relative strength of the Cabibbo-favored and Cabibbo-suppressed decays. Up to now, these branching fractions have not been measured yet.

This Letter reports measurements of the branching fractions for the inclusive decays \(D \rightarrow K^0/K^0 X\) (\(X\) = any particles) and \(D \rightarrow K^{*\pm} X\). The branching fractions are obtained based on analyses of the data sample of integrated luminosity of 33 pb\(^{-1}\) collected with the BES-II detector at and around 3.773 GeV. Throughout the Letter, charge conjugation is implied.

2. BES-II detector

The BES-II is a conventional cylindrical magnetic detector [2] operated at the Beijing Electron–Positron Collider (BEPC) [3]. A 12-layer Vertex Chamber (VC) surrounding the beryllium beam pipe provides input to the event trigger, as well as coordinate information. A forty-layer main drift chamber (MDC) located just outside the VC yields precise measurements of charged particle trajectories with a solid angle coverage of 85\% of 4\(\pi\); it also provides ionization energy loss \((dE/dx)\) measurements which are used for particle identification. Momentum resolution of 1.7\%\(\sqrt{1+p^2}\) (\(p\) in GeV/c) and \(dE/dx\) resolution of 8.5\% for Bhabha scattering electrons are obtained for the data taken at \(\sqrt{s} = 3.773\) GeV. An array of 48 scintillation counters surrounding the MDC measures the time of flight (TOF) of charged particles with a resolution of about 180 ps for electrons. Outside the TOF, a 12 radiation length, lead-gas barrel shower counter (BSC), operating in limited streamer mode, measures the energies of electrons and photons over 80\% of the total solid angle with an energy resolution of \(\sigma_E/E = 0.22/\sqrt{E}\) (\(E\) in GeV) and spatial resolutions of \(\sigma_\theta = 7.9\) mrad and \(\sigma_\phi = 2.3\) cm for electrons. A solenoidal magnet outside the BSC provides a 0.4 T magnetic field in the central tracking region of the detector. Three double-layer muon counters instrument the magnet flux return and serve to identify muons with momentum greater than 500 MeV/c. They cover 68\% of the total solid angle.

3. Data analysis

Around the center-of-mass energy of 3.773 GeV, \(\psi(3770)\) is produced in the annihilation of \(e^+e^-\). In decays to \(D\bar{D}\) pairs (\(D^0\bar{D}^0\) or \(D^+D^-\)) with a large branching fraction of about (85 \pm 6)\% [3]. These provide us a unique method to directly measure the branching fractions for \(D\) meson decays. In the analyses we first reconstruct a \(D\) meson of the \(D\bar{D}\) pair (this is called a singly tagged \(\bar{D}\)), then select the inclusive decays \(D \rightarrow \bar{K}^0/K^0 X\) or \(D \rightarrow K^{*\pm} X\) on the recoil side of the singly tagged \(\bar{D}\), and measure the absolute branching fractions for these decays.

3.1. Events selection

To select the candidate events for the decays, it is first required that at least two charged tracks be well reconstructed in the MDC with good helix fits. In order to ensure the well-measured 3-momentum vectors and the reliability of the charged-particle identification, the polar angle \(\theta\) of each charged track must satisfy \(|\cos\theta| < 0.85\). It is then required that each charged track, except for those from \(K^0_S\), originate from the interaction region defined by \(|V_x| + |V_y| < 2.0\) cm and \(|V_z| < 20.0\) cm, where \(V_x\), \(V_y\), and \(V_z\) are the closest approach of the charged track in the \(x\), \(y\), and \(z\) directions.

---

* Corresponding author.
E-mail address: liuhb@mail.ihep.ac.cn (H.H. Liu).
1 Current address: Purdue University, West Lafayette, IN 47907, USA.
2 Current address: DESY, D-22607 Hamburg, Germany.
3 Current address: Laboratoire de l’Accélérateur Linéaire, F-91898 Orsay, France.
4 Current address: University of Michigan, Ann Arbor, MI 48109, USA.
Pions and kaons are identified using the $dE/dx$ and TOF measurements, with which the combined confidence levels ($CL_x$ or $CL_K$) for a pion or kaon hypotheses are calculated. A pion candidate is required to have $CL_x > 0.001$ and a kaon candidate is required to satisfy $CL_K > CL_x$.

Neutral pions are reconstructed through the decay $\pi^0 \to \gamma\gamma$. For the $\gamma$ from $\pi^0$ decay, the energy deposited in the BSC is required to be greater than 70 MeV; the electromagnetic shower is required to start in the first 5 readout layers; and the angle between the $\gamma$ and the nearest charged track is required to be greater than $22^\circ$ [4,5].

3.2. Singly tagged $\bar{D}^0$ and $D^-$ samples

The singly tagged $\bar{D}^0$ and $D^-$ samples used in the analyses have been selected in the previous works [4,5], where the $\bar{D}^0$ mesons are reconstructed in four hadronic decay modes $K^+\pi^-, K^+\pi^-\pi^+, K^0\pi^+\pi^-\pi^0$ and $K^+\pi^-\pi^0\pi^0$ ($Kn\pi$, $n = 1, 2, 3$), and the $D^-$ mesons are reconstructed in nine hadronic decay modes $K^+\pi^-\pi^-, K^0\pi^-, K^0\bar{K}^-, K^-K^-, K^0\pi^-\pi^+, K^0\pi^-\pi^-\pi^0, K^+\pi^+\pi^-\pi^-\pi^0$ and $\pi^+\pi^-\pi^-$ ($mKn\pi$, $m = 0, 1, 2; n = 0, 1, 2, 3, 4$). These give the total numbers 7584 ± 198(stat.) ± 341(sys.) singly tagged $\bar{D}^0$ mesons [4] and 5321 ± 149(stat.) ± 160(sys.) singly tagged $D^-$ mesons [5].

3.3. Candidates for $D \to \bar{K}^0/K^0 X$ and $D \to K^{\ast+}(K^{\ast+})X$

Candidates for the inclusive decays $D \to \bar{K}^0/K^0 X$ and $D \to K^{\ast+}(K^{\ast+})X$ are selected from the survival tracks on the recoil side of the singly tagged $\bar{D}$. Neutral kaons are reconstructed through the decay $K_S^0 \to \pi^+\pi^-$. We require that $\pi^+\pi^-$ must originate from a secondary vertex which is displaced from the event primary vertex by 7 mm at least. $K_S^{\ast-}(K^{\ast+})$ mesons are reconstructed through the decay $K^{\ast-}(K^{\ast+}) \to K_S^{\ast-}(K_S^{\ast+})$.

In each invariant masses spectrum for the $mKn\pi$ combinations, the region within a $\pm 3\sigma_{MD_{\bar{D}}}$ window around the fitted $\bar{D}$ mass $M_{\bar{D}_i}$ is defined as the singly tagged $\bar{D}$ signal region, where $\sigma_{MD_{\bar{D}_i}}$ is the standard deviation of the mass spectrum for the $i$th tag mode. The region outside a $\pm 4\sigma_{MD_{\bar{D}_i}}$ window around the fitted $\bar{D}$ mass is taken as the $\bar{D}$ sideband region. In order to estimate the number of background events in the $\bar{D}$ signal regions, the number of the $\bar{D}$ sideband events is normalized by the ratio of the area of the fitted background in the $\bar{D}$ signal region to that of the $\bar{D}$ sideband.

Figs. 1 and 2 show the distributions of the $\pi^+\pi^-$ invariant masses for the events observed on the recoil side of the $\bar{D}^0$ and $D^-$ tags for studying $D \to \bar{K}^0/K^0 X$. In each figure, (a) is the mass spectrum for the events with the $mKn\pi$ invariant masses in the $\bar{D}$ signal regions, and (b) is the normalized mass spectrum for the $\bar{D}$ sideband events. Fitting each mass spectrum with a Gaussian function for $K_S^0$ signal and a polynomial to describe the background shape, the numbers of $K_S^0$ mesons are obtained. These numbers are summarized in Table 1, where $N$ and $N_b$ are the numbers of $K_S^0$ mesons observed from the events with the $mKn\pi$ invariant masses in the $\bar{D}$ signal region and $D$ sideband regions, respectively. Subtracting $N_b$ from $N$, we obtain the number $n$ of the signal events for $D \to \bar{K}^0/K^0 X$.

Figs. 3–6 show the distributions of invariant masses for the $K_S^0\pi^-$ or $K_S^0\pi^+$ combinations observed on the recoil side of $D^0$ or $D^-$ tags respectively for studying the decays $D \to K^{\ast-}(K^{\ast+})X$. In each figure, (a) is the mass spectrum for the events in which the $mKn\pi$ invariant masses are in the $\bar{D}$ signal regions, and (b) is the normalized mass spectrum for the $\bar{D}$ sideband events. The histograms are for the events with the $\pi^+\pi^-$ invariant masses in the $K_S^0$ signal region (within a $\pm 3\sigma_{MK_S^0}$ window around the fitted $K_S^0$ mass), and the shadows are the normalized background estimated by $K_S^0$ sideband (outside a $\pm 4\sigma_{MK_S^0}$ window around the fitted $K_S^0$ mass). Fitting each mass spectrum with a Gaussian function for $K^{\ast-}(K^{\ast+})$ signal and a polynomial to describe the background shape, we obtain the numbers of $K^{\ast-}(K^{\ast+})$ mesons. In the fit, the mass and width
The detection efficiencies for the inclusive $\bar{K}^0$/$K^0$ and $K^+/(K^+)$ decays of $D$ mesons are estimated by Monte Carlo simulation. The Monte Carlo events are generated as $e^+e^- \rightarrow D\bar{D}$, where $D$ decays into the singly tagged $\bar{D}$ modes and $D$ decays into $\bar{K}^0/K^0X$ or $K^-(K^+)X$. The particle trajectories are simulated with the GEANT3 based Monte Carlo simulation package of the BES-II detector [6]. Weighting the efficiencies by the branching fractions of $D$ decays quoted from PDG [1]

### 4. Results

#### 4.1. Monte Carlo efficiency

The detection efficiencies for the inclusive $\bar{K}^0$/$K^0$ and $K^+/(K^+)$ decays of $D$ mesons are estimated by Monte Carlo simulation. The Monte Carlo events are generated as $e^+e^- \rightarrow D\bar{D}$, where $D$ decays into the singly tagged $\bar{D}$ modes and $D$ decays into $\bar{K}^0/K^0X$ or $K^-(K^+)X$. The particle trajectories are simulated with the GEANT3 based Monte Carlo simulation package of the BES-II detector [6]. Weighting the efficiencies by the branching fractions of $D$ decays quoted from PDG [1]
and the numbers of the singly tagged $\bar{D}$ mesons, we obtain the averaged efficiencies to be $(6.94 \pm 0.06)\%$ for $D^0 \rightarrow \bar{K}^0/K^0 X$, $(7.57 \pm 0.06)\%$ for $D^+ \rightarrow \bar{K}^0/K^0 X$, $(2.56 \pm 0.04)\%$ for $D^0 \rightarrow K^+\!(K^+\!)X$ and $(2.36 \pm 0.06)\%$ for $D^+ \rightarrow K^{*-}(K^{*-})X$.

### 4.2. Branching fractions

The branching fractions for the inclusive decays $D \rightarrow \bar{K}^0/K^0 X$ and $D \rightarrow K^{*-}X$ are determined by dividing the numbers of the signal events by the numbers of the singly tagged $\bar{D}$ mesons and the detection efficiencies. The branching fractions for the inclusive decays are

$$BF(D^0 \rightarrow \bar{K}^0/K^0 X) = (47.6 \pm 4.8 \pm 3.0)\%,$$

$$BF(D^+ \rightarrow \bar{K}^0/K^0 X) = (60.5 \pm 5.5 \pm 3.3)\%,$$

$$BF(D^0 \rightarrow K^{*-}X) = (15.3 \pm 8.3 \pm 1.9)\%$$

and

$$BF(D^+ \rightarrow K^{*-}X) = (5.7 \pm 5.2 \pm 0.7)\%,$$

where the first error is statistical and the second systematic.

The upper limits of the branching fractions for the inclusive decays $D^0 \rightarrow K^{+}X$ and $D^+ \rightarrow K^{*+}X$, which include the systematic errors, are set to be

$$BF(D^0 \rightarrow K^{+}X) < 3.6\%$$

and

$$BF(D^+ \rightarrow K^{*+}X) < 20.3\%$$

at 90% confidence level.

In the measurement of the branching fractions for $D^0 \rightarrow K^{*-}X$ and $D^0 \rightarrow K^{+}X$, we use three singly tagged $\bar{D}^0$ modes ($K^+\pi^-, K^+\pi^-\pi^-\pi^-\pi^+$ and $K^+\pi^-\pi^0$). These give us $7033 \pm 193(\text{stat.}) \pm 316(\text{sys.})$ singly tagged $\bar{D}^0$ mesons. In the measured branching fractions, the systematic error arises from the uncertainties in particle identification ($\sim 0.5\%$ per track), in tracking ($\sim 2.0\%$ per track), in the number of the singly tagged $\bar{D}$ mesons ($\sim 4.5\%$ for $\bar{D}^0$ and $\sim 3.0\%$ for $D^-$) [4,5], in $K^0_S$ selection ($\sim 1.1\%$) [5], in background parameterization ($1.3\% \sim 9.3\%$) and in Monte Carlo statistics ($0.8\% \sim 2.5\%$). Adding these uncertainties in quadrature yields the total systematic error to be $6.4\%$ for $D^0 \rightarrow \bar{K}^0/K^0 X$, $5.4\%$ for $D^+ \rightarrow \bar{K}^0/K^0 X$, $12.2\%$ for $D^0 \rightarrow K^{*-}(K^{*-})X$ and $11.9\%$ for $D^+ \rightarrow K^{*-}(K^{*-})X$.

The upper limits of the branching fractions for the inclusive $K^{*-}$ decays are determined to be $BF(D^0 \rightarrow K^{*-}X) = (15.3 \pm 8.3 \pm 1.9)\%$ and $BF(D^+ \rightarrow K^{*-}X) = (5.7 \pm 5.2 \pm 0.7)\%$. These are measured for the first time. The upper limits of the branching fractions for the inclusive $K^{*+}$ decays are set to be $BF(D^0 \rightarrow K^{*+}X) < 3.6\%$ and $BF(D^+ \rightarrow K^{*+}X) < 20.3\%$ at 90% confidence level.

### Acknowledgements

The BES Collaboration thanks the staff of BEPC and computing center for their hard efforts. This work is supported in part by the National Natural Science Foundation of China under contracts Nos. 10491300, 10225524, 10225525, 10425523, the Chinese Academy of Sciences under contract No. KJ 95T-03, the 100 Talents Program of CAS under Contract Nos. U-11, U-24, U-25, and the Knowledge Innovation Project of CAS under Contract Nos. U-602, U-34 (IHEP), and the National Natural Science Foundation of China under Contract No. 1025522 (Tsinghua University).

### References

Threshold hyperon production in proton–proton collisions at COSY-11


Abstract

The $\Sigma^+$ hyperon production was measured at the COSY-11 spectrometer via the $pp \rightarrow nK^+\Sigma^+$ reaction at excess energies of $Q = 13$ MeV and $Q = 60$ MeV. These measurements continue systematic hyperon production studies via the $pp \rightarrow pK^+\Lambda/\Sigma^0$ reactions where a strong decrease of the cross section ratio close-to-threshold was observed. In order to verify models developed for the description of the $\Lambda$ and $\Sigma^0$ production we have performed the measurement on the $\Sigma^+$ hyperon and found unexpectedly that the total cross section is by more than one order of magnitude larger than predicted by all anticipated models. After the reconstruction of the kaon and neutron four momenta, the $\Sigma^+$ is identified via the missing mass technique. Details of the method and the measurement will be given and discussed in view of theoretical models.

© 2006 Elsevier B.V. All rights reserved.
PACS: 13.60.Hb; 13.75.-n; 25.40.Ve; 28.20.-n

Keywords: Strangeness; Kaon; Near threshold hyperon production; Sigma production; COSY-11

1. Introduction

The study of the hyperon production in hadron induced multi particle exit channels like $pp \rightarrow NKY$ includes several aspects. The nucleon–hyperon interaction can be extracted by analyzing the $NY$ subsystem in the appropriate kinematical region. Closely related to that is the issue of the reaction mechanisms of the hyperon production which have to be clarified for an unambiguous interpretation of the data. If the hyperon production is due to the excitation and a subsequent decay of intermediate nucleon resonances it allows to extract information about the structure of the relevant resonances.

The $pp \rightarrow pK^+\Lambda$ excitation function close-to-threshold shows a clear deviation from the pure phase space distribution and a proton–hyperon final state interaction (FSI) has to be included to describe the data [1–4]. In the $pp \rightarrow pK^+\Sigma^0$ channel the $pY$ FSI seems to be negligible and the pure phase space calculations follow reasonably well the data points. The cross section ratio $\sigma(pp \rightarrow pK^+\Lambda)/\sigma(pp \rightarrow pK^+\Sigma^0)$ below excess energies of $Q \sim 20$ MeV is in the order of 28 [2,3] in contrast to the value of about 2.5 determined for excess energies higher than $Q = 300$ MeV [5] (see Fig. 1). This value is in good agreement with the $\Lambda/\Sigma^0$ isospin relation. The question arises if this drastic cross section increase close-to-threshold is

* Corresponding author.
E-mail address: trozek@us.edu.pl (T. Rożek).

0370-2693/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.physletb.2006.07.066
a mere effect of the \( N\Lambda \) FSI or whether it is partly due to the reaction mechanisms in the \( NY \) channels. To explain the increase of the \( \Lambda/\Sigma^0 \) cross section ratio in the close-to-threshold region, different scenarios were proposed.

Calculations of the strangeness production by solely \( \pi \) and \( K \) exchange added incoherently have been performed in Ref. [6]. The \( \pi N \rightarrow Y K \) and \( KN \rightarrow Y K \) scattering amplitudes for pion and kaon exchange, respectively, were taken from the existing data in the higher energy region [5]. Since this incoherent \( \pi/K \) exchange model describes the \( pp \rightarrow pK^+\Lambda \) cross section over the whole energy range, but overestimates the close-to-threshold region in case of the \( pp \rightarrow pK^+\Sigma^0 \) channel, the predicted \( \Lambda/\Sigma^0 \) ratio is too low for \( Q \leq 20 \text{ MeV} \) (see curve (1) in Fig. 1).

A better description of the strong rise of the ratio towards lower \( Q \)-values is achieved by the resonance model (curve (2) in Fig. 1) [6–9]. In this model the nonresonant direct contributions like \( \pi \) or \( K \) exchange were not included, but the \( \pi \), \( \eta \), and \( \rho \) meson exchange with the excitation of the intermediate baryonic resonances \( N(1650) \), \( N(1710) \), \( N(1720) \), and \( \Delta(1920) \) are taken into account. In this resonance model the close-to-threshold region of the \( \Lambda/\Sigma^0 \) cross section ratio seems to be better reproduced than the higher energy values i.e. \( (Q \geq 10 \text{ MeV}) \). It should be stressed that in these calculations the parameters were fixed on the basis of higher energy data, before the close-to-threshold \( \Lambda \) and \( \Sigma^0 \) data were available.

Other calculations by Shyam [10] (based on the effective Lagrangian model) include for the strangeness production also meson exchange (\( \pi, \rho, \sigma \) and \( \omega \)) together with the excitation of resonances. The coupling constant was determined by fitting data of the \( \pi^+p \rightarrow \Sigma^+K^+, \pi^-p \rightarrow \Sigma^0K^0 \) and \( \pi^-p \rightarrow \Sigma^-K^+ \) reaction channels. The coherent sum of resonant states and meson exchange processes describes the experimental data for the \( pp \rightarrow pK^+\Lambda \) and \( pp \rightarrow pK^+\Sigma^0 \) channels very well. The effective Lagrangian model is depicted by the curve (3) in Fig. 1.

The Jülich theory group has performed calculations including \( \pi \) and \( K \) exchange [11,12]. In their approach the interaction between the hyperons (\( \Lambda, \Sigma \)) and the nucleon is described by a microscopic \((\Lambda N - \Sigma N)\) coupled channel model [13] with a coherent superposition of the production amplitudes. The \( \Lambda \) production is dominated by the \( K \) exchange and therefore the contribution due to an interference between \( \pi \) and \( K \) exchange is negligible in this hyperon channel. On the other hand the \( \pi \) and \( K \) exchanges give a comparable contribution to the cross section in the case of \( \Sigma^0 \) production. An interference between \( \pi \) and \( K \) exchange amplitudes act very differently on the two channels. Within the Jülich meson exchange model the large cross section ratio can be described by a destructive interference of the \( \pi \) and \( K \) exchange amplitudes only. For excess energies above 20 MeV the model is not valid any more but qualitatively the cross section ratio given by the model stays at a nearly constant level.

Although the various descriptions of the cross section ratio differ even in the dominant reaction mechanism, all reproduce more or less the trend of an increase of the \( \Lambda/\Sigma^0 \) cross section ratio in the threshold region (see Fig. 1). The present data are not sufficient to definitely exclude possible explanations and therefore an unambiguous identification of the dominant reaction mechanism is impossible. To clarify this point further data are needed. Especially the other isospin channels should allow to extract information about the production mechanisms. Recently, besides the \( \Lambda \) and \( \Sigma^0 \) production the reaction channel \( pp \rightarrow nK^+\Sigma^+ \) became accessible at the COSY-11 detection system after the installation of a neutron detector. The measurement of the \( \Sigma^+ \) hyperon production via this reaction was performed at two beam momenta, \( p_{beam} = 2.6 \text{ GeV}/c \) and \( p_{beam} = 2.74 \text{ GeV}/c \), corresponding to excess energies of 13 MeV and 60 MeV, respectively.

2. Experiment

COSY-11 is an internal magnetic spectrometer at the COoler SYnchrotron and storage ring COSY [14] in Jülich. The interaction between a proton in the beam and a proton from the \( \text{H}_2 \) cluster target [15] may lead to the production of the \( \Sigma^+ \) hyperon in the \( pp \rightarrow nK^+\Sigma^+ \) reaction. The charged reaction products are separated from the circulating beam in the magnetic field of one of the regular COSY dipoles [16]. The \( \Sigma^+ \) hyperon is identified via the missing mass technique by detecting the \( K^+ \) and the neutron. The momentum vector of the \( K^+ \) meson can be established by tracking back the \( K^+ \) trajectory reconstructed in the drift chambers (DC1 and DC2 in Fig. 2) through the known magnetic field back to the target point. Together with the velocity measurement in the two scintillators S8 and S1, the kaon is identified via its invariant mass.

Assuming a hit in the neutron detector being due to a neutron, the four momentum vector of the neutron is given by the measured velocity, the direction of the neutron (given by the first hitted module) and the known mass. The background from charged particles hitting the neutron detector is discriminated by veto scintillators.

In Fig. 3 the experimental distributions of the squared missing mass \((m^2)\) of the \( pp \rightarrow nK^+X \) system for the two beam momenta are shown. For the higher momentum, an enhancement around the squared \( \Sigma^+ \) mass is clearly seen on a large
background (Fig. 3b), but for the lower beam momentum (Fig. 3a) a $\Sigma^+$ peak is not directly visible.

In order to determine the number of $\Sigma^+$ events in the higher energy data set, a fit has been done with a polynomial function superposed by the expected missing mass distribution of the $nK^+\Sigma^+$ system for the $pp \rightarrow nK^+\Sigma^+$ reaction obtained from the simulation studies. In Fig. 4a the experimental missing mass spectrum of the $pp \rightarrow nK^+\Sigma^+$ system is compared with the fitted polynomial function. The expected distribution from MC studies with $X = \Sigma^+$ is depicted in the figure as well. Fig. 4b shows the result of the subtraction of the fitted polynomial from the experimental missing mass distribution together with the MC distribution.

In order to understand the background distribution, 22 reaction channels (mostly multi-pion reactions but also $pp \rightarrow pK^+\Lambda$ ($\Sigma^0$, $\Lambda\gamma$)) were simulated and their contributions to the missing mass distribution were determined. These studies showed that the reactions $pp \rightarrow pK^+\Lambda$ and $pp \rightarrow pK^+\Delta\gamma$ ($\gamma'$s) are the dominant background channels in the $\Sigma^+$ region. The Monte Carlo code includes the realistic geometry and physics processes like energy loss and straggling which occasionally cause the misidentification of the particle type.

All background channels result in a rather smooth distribution of the missing mass spectrum as can be inferred from calculations from MC studies and by comparing the two experimental distributions (see Fig. 3).

For the lower energy data set a $\Sigma^+$ peak is not obviously visible via the missing mass distribution. Therefore, a simple polynomial background fitting cannot be used. To determine the number of $\Sigma^+$ events it was assumed that the background shape for this data set is the same as that at the higher energy. This assumption is justified since there is no new open channel for the higher energy.

At the COSY-11 experiment the shape of the missing mass distribution is mainly determined by the acceptance of the detection system and is dependent on the excess energy of an individual event. From the analysis of $\eta$ and $\eta'$ production studies at COSY-11 it was verified that the background shape resulting here mainly from multi pion production is in very good agreement with the expectations from Monte Carlo studies taking into account the detector characteristics and is comparable at different beam momenta. In addition Monte Carlo data of the reaction channels which contribute dominant to the background in the $\Sigma^+$ production were compared in view of the background shape by adjusting the kinematical limits. Within error bars their shapes were identical. Therefore it is justified to assume that the background shape is the same for both beam momenta. For a detailed discussion on the background shape at COSY-11 we refer to [17]. The background shape from the experimental missing mass distribution for the higher energy data set is the same as that at the higher energy. This assumption was made, that the kaon peak in the experimental distribution has the same position and width as in the simulated distribution.

3. Results

3.1. Total cross section

For the lower energy data set, even after applying all cuts, there was no clear enhancement around the kaon mass in the invariant mass distribution, and therefore the assumption was made, that the kaon peak in the experimental distribution has the same position and width as in the simulated distribution.
As a cross check, event samples with cuts on different regions but still within the kaon range in the experimental invariant mass distribution were taken and the corresponding missing mass region were applied. Events from these regions, namely: $\mu_{\exp} \pm 0.25 \sigma_{\exp}$, $\mu_{\exp} \pm 0.5 \sigma_{\exp}$ and $\mu_{\exp} \pm 1.0 \sigma_{\exp}$ were taken and the corresponding missing mass distributions were generated. Next the number of $\Sigma^+$ events for each of these distributions was determined. The results are listed in the middle column of Table 1 and in the last column the number of the $\Sigma^+$ events corresponding to the full Gaussian distribution.

In order to calculate the cross section for the $pp \to nK^+\Sigma^+$ reaction the number of $\Sigma^+$ hyperon events and the detection efficiency of the COSY-11 apparatus for the two excess energies were determined. The luminosity was determined by a simultaneous measurement of proton–proton elastic scattering.

In Table 2 the total cross sections for both beam momenta are given. The systematical errors are due to: (i) errors of the detection efficiency determination which is 8.5% for the lower and 3.5% for the higher energy data set (including the inaccuracy of the effective detector position and of the beam momentum determination), (ii) uncertainty in the form of the background, and (iii) error of the luminosity calculation which is 3% for both data sets and includes the uncertainty due to the normalization procedure and the error of the solid angle determination. For the data at 2.74 GeV/c the uncertainty in the background form was estimated by comparing the polynomial fit with a background form resulting from an adjusted sum of known background reaction channels generated in Monte Carlo studies. The difference is about 18%. For the data at 2.6 GeV/c the region for the adjustment of the background used for the subtraction was varied resulting in an error of about 20%. The values in Table 2 include also a change in the detection efficiency resulting from the inclusion of higher partial waves. Close to the reaction threshold higher than 5 partial wave contributions are...
not expected, however, if the excess energy $Q$ for the studied channel go beyond a few MeV range, higher partial waves can contribute to the production mechanism. Since in the case of the $\Sigma^+$ hyperon production this contribution is unknown, its effect is assumed on the basis of the $pp \rightarrow pK^+\Lambda$ channel studied at the TOF experiment at COSY [18–20]. Higher partial waves in a strength given in [18,19] result in a decrease of the detection efficiency by 30% for the lower and by 7.7% for the higher energy data set. The sum of the total systematical errors equals to 60% for the lower and 34% for the higher energy data set.

3.2. Comparison with model predictions

Among the models described in the introduction only two give predictions for the $pp \rightarrow nK^+\Sigma^+$ reaction, namely the Jülich meson exchange model [11,12] and the resonance model [6,9]. Calculations of the $\Sigma^+$ production within the Jülich meson exchange model predict a total cross section of $\sigma = 0.23 \mu b$ at $Q = 13$ MeV for the destructive interference (which was necessary to describe the high $\Lambda/\Sigma^0$ cross section ratio at threshold). This is about a factor of 20 below the experimental value of 4.56 $\mu b$ given in Table 2. A constructive interference would result in a cross section even a factor of 53 too low.

For the resonance model the predictions for the $pp \rightarrow nK^+\Sigma^+$ channel for the close-to-threshold region deviate even more from the data. In Fig. 7 the model predictions and the available data for the $pp \rightarrow nK^+\Sigma^+$ (a), $pp \rightarrow pK^0\Sigma^0$ (b) and $pp \rightarrow pK^+\Lambda$ (c) channels are shown. The data points presented by triangles, dots and squares in the close-to-threshold region were measured by the COSY-11 Collaboration [1–3]. The data point in the $pp \rightarrow pK^+\Lambda$ channel indicated by the arrow was determined in parallel by selecting the $K^+\pi$ exit channel which was included in the triggered events from the $\Sigma^+$ production data at 2.74 GeV/c. The high energy data for the given reactions were taken from [5] and [21]. The model calculations for each channel are given by the solid line [9,22].

The data point for the $pp \rightarrow nK^+\Sigma^+$ channel at $Q = 13$ MeV is underestimated in the total cross section calculated using the resonance model [6,9] by about a factor of 500 and for $Q = 60$ MeV by about a factor of 50. For the $pp \rightarrow pK^+\Sigma^0$ channel, this model calculation describes the existing data set and in the case of the $pp \rightarrow pK^+\Lambda$ channel the underestimation of the cross section in the close-to-threshold region is about a factor of 16 being 30 times smaller than for the $\Sigma^+$ production. At high excess energies, the $\Sigma^+$ data points are by a factor of 3–4 below the model calculations. Previous COSY-11 hyperon production studies conclude, that final state interactions (FSI) plays an important role in the close-to-threshold $\Lambda$ production [1–3,23]. In the resonance model the FSI is not included [6,9] and therefore the deviation of the model calculations from the data points in close-to-threshold region is expected if a strong FSI is present. This effect is clearly seen for the $pp \rightarrow pK^+\Lambda$ and barely observed for the $pp \rightarrow pK^+\Sigma^0$ reaction channel.

In the investigation of the hyperon production in COSY-11 it was observed [1–3] that a pure 3-body phase space (PS) dependent cross section expressed as [25]:

$$\sigma = K \cdot Q^2,$$

(1)

where $K$ is a normalization factor and $Q$ the excess energy cannot describe the $pp \rightarrow pK^+\Lambda$ data, and therefore a modification is needed which takes into account the proton–hyperon FSI. In order to describe the close-to-threshold region, the parametrisation of the excitation function including the FSI proposed by Fäldt–Wilkin [4] was used. It is expressed by:

$$\sigma = C \cdot \frac{Q^2}{(1 + \sqrt{1 + Q/\varepsilon})^2},$$

(2)

where $C$ and $\varepsilon$ are parameters related to the FSI strength.

In Fig. 8 the cross sections for different production channels for the hyperon $\Lambda$, $\Sigma^0$ and $\Sigma^+$ are compared to predictions of the 3-body phase space (PS, dotted line) and the 3-body phase space calculations modified by the $pY$ FSI (PS + FSI, solid line), following equation (2) with $\varepsilon$ and $C$ as free parameters.

Fig. 7. Comparison of the experimental total cross section with the resonance model [6,9] predictions for various $pp \rightarrow nK^+\Sigma^+$ reactions. Full triangles in (a) are data obtained in this work. Data in the close-to-threshold region (presented as full symbols in (b) and (c)) are taken from Refs. [1–3] and data from the high excess energy region (open symbols) from Refs. [5,21]. In (c) the data point indicated by the arrow was determined from our data as a cross check of the luminosity calculation.

Fig. 8. The $pp \rightarrow nK^+\Sigma^+$, $pp \rightarrow pK^+\Lambda$ and $pp \rightarrow pK^+\Sigma^0$ cross sections as a function of the excess energy $Q$. Experimental data are from Refs. [1–3, 24] and from this work. The errors for the $pp \rightarrow nK^+\Sigma^+$ reactions represent a sum of statistical and systematical uncertainties given in Table 2. The lines show the calculations corresponding to 3-body phase space with (solid line) and without (dashed line) final state interaction.
These parameters are related to the scattering length $a$ and the effective range $r$ of the $pY$ potential [4].

For the $pp \rightarrow nK^+\Sigma^+$ data the resulting $\varepsilon$ and $C$ parameters are of similar values as for the $pp \rightarrow pK^+\Lambda$ channel. It seems that in the case of the $\Sigma^+$ production via the $pp \rightarrow nK^+\Sigma^+$ reaction a rather strong $n\Sigma^+$ FSI is present, however, within the error bars also the curve obtained without FSI describes the two data points. Therefore for an unambiguous conclusion about the $p-\Sigma^+$ FSI more data are needed to disentangle the reaction mechanisms and especially the role of nuclear resonances.

4. Conclusions and perspectives

The total cross section of the $pp \rightarrow nK^+\Sigma^+$ reaction was determined at the COSY-11 detection system for excess energies of $Q = 13$ and 60 MeV. However, the values established are by more than an order of magnitude larger than the expectations of any currently available model predictions.

It should be noticed that the unexpected large total $\Sigma^+$ production cross section is somehow in line with an observation by Tan [26] who concluded that when assuming charge symmetry in $\Sigma^+n$ and $\Sigma^0p$ scattering, the contribution from the $\Sigma^0$ diagram is less than one seventh of the one from the $\Sigma^+$ channel. Further, recently [27] for the case of the $\phi$ production it was suggested that a strong enhancement of the reaction amplitude towards threshold might be due to the presence of a crypto exotic baryon with hidden strangeness. Though this observation is not one-to-one conferrable to other isospin channels, in the $\Sigma^0p$ system no corresponding structure was observed, it might give a hint for some exotic mechanisms. Certainly the present results do not prove such a reaction process but might indicate the appearance of an interesting phenomena. In any case, present theoretical predictions of the cross sections strongly underestimate the experimental data. The adjustment of the excitation function expected from a phase space distribution including $N-Y$ FSI to the data results in parameters comparable to the $p-\Lambda$ system which may indicate a strong $n-\Sigma^+$ interaction but due to the large systematic uncertainties the data are also consistent with a pure phase space distribution without $p-\Sigma^+$ FSI.

Further studies of the $\Sigma^+$ production are necessary to clarify the picture. On the experimental side additional data points should be added for which an improved event selectivity is favorable to reduce the large uncertainties introduced by the background subtraction. A 4$\pi$ detection system for neutral and charged particles which will be soon available with WASA at COSY could be used [28]. On the theoretical side an improved model has to be developed which consistently reproduces the hyperon cross section data close-to-threshold.

Note added in proof

During the evaluation of the article we have been made aware of the predictions of the excitation function of the total cross section for the $pp \rightarrow nK^+\Sigma^+$ reaction which is closer to the data in comparison with the models discussed, yet still underpredict the determined total cross sections by more than an order of magnitude [29].

Acknowledgements

This work has been supported by the European Community—Access to Research Infrastructure action of the Improving Human Potential Programme, by the FFE grants (Nos. 41266606 and 41266654) from the Research Center Jülich, by the DAAD Exchange Programme (PPP-Polen), and by the Polish State Committee for Scientific Research (grant No. PB1060/P03/2004/26).

References

$g$ factors of $^{31,32,33}\text{Al}$: Indication for intruder configurations in the $^{33}\text{Al}$ ground state

P. Himpe $^a$, G. Neyens $^{a,*}$, D.L. Balabanski $^b$, G. Bélier $^c$, D. Borremans $^a$, J.M. Daugas $^c$, F. de Oliveira Santos $^d$, M. De Rydt $^a$, K. Flanagan $^a$, G. Georgiev $^{d,2}$, M. Kowalska $^e$, S. Mallion $^a$, I. Matea $^{d,3}$, P. Morel $^c$, Yu.E. Penionzhkevich $^f$, N.A. Smirnova $^g$, C. Stodel $^d$, K. Turzó $^a$, N. Vermeulen $^a$, D. Yordanova $^a$

$^a$ Instituut voor Kern- en Stralingsfysica, K.U. Leuven, Celestijnenlaan 200 D, B-3001 Leuven, Belgium
$^b$ Dipartimento di Fisica, Università di Camerino, I-62032, Camerino, Italy
$^c$ CEA/DIF/DPTA/PN, BP 12, F-91680 Bruyères le Chatel, France
$^d$ GANIL, BP 55027, F-14076 Caen cedex 5, France
$^e$ Institut für Physik, Universität Mainz, D-55099 Mainz, Germany
$^f$ Joint Institute for Nuclear Research, FLNR, Dubna 141980, Moscow Region, Russia
$^g$ Vakgroep Subatomaire en Stralingsfysica, Universiteit Gent, Proeftuinstraat 86, B-9000 Gent, Belgium

Received 24 May 2006; received in revised form 10 August 2006; accepted 20 October 2006
Available online 30 October 2006
Editor: V. Metag

Abstract

The $g$ factors of $^{31,32,33}\text{Al}$ have been measured using the $\beta$-nuclear magnetic resonance ($\beta$-NMR) technique on spin-polarized beams produced in the fragmentation of a $^{36}\text{S}$ (77.5 MeV/u) beam on a $^9\text{Be}$ target. Nearly pure beams of $\text{Al}$ ($Z = 13$) isotopes were selected with the high-resolution fragment separator LISE at GANIL. An asymmetry as high as 6% has been observed in the $\beta$-NMR curve for $^{32}\text{Al}$ implanted in a Si single crystal. The magnetic moment of the $N = 20$ nucleus $^{33}\text{Al}$ is obtained for the first time: $\mu({}^{33}\text{Al}, I^\pi = 5/2^+ \!+ \!1) = 4.088(5)\mu_N$, while those of $^{31,32}\text{Al}$ are obtained with improved accuracy: $\mu({}^{31}\text{Al}, I^\pi = 5/2^+ \!+ \!1) = 3.830(5)\mu_N$ and $\mu({}^{32}\text{Al}, I^\pi = 1^+ \!+ \!1) = 1.9516(22)\mu_N$. Comparison of the results to shell-model calculations in the sd and the sdpf shell-model spaces leads to the conclusion that $^{33}\text{Al}$ must contain some contribution from 2p–2h intruder configurations in its ground-state wave function. This indicates a gradual transition from the normal sd shell Si ($Z = 14$) isotopes to the intruder Mg ($Z = 12$) isotopes.

© 2006 Elsevier B.V. All rights reserved.

The gyromagnetic factor ($g$ factor) is a very sensitive probe to the wave function of a nuclear state. Systematic precise measurements of $g$ factors in a chain of isotopes/isotones allow to probe small changes in the nuclear structure with changing isospin. Nuclei near the $\beta$-stability line with protons and neutrons filling the sd shell-model orbits ($8 \leq N, Z \leq 20$) are well described using the USD shell-model interaction developed by Brown and Wildenthal [1]. They show that the experimental $g$ factors in this region are well reproduced using the free nucleon $g$ factors. However, the properties of nuclei at the border of this region, with magic neutron number $N = 20$ and a mid-shell number of protons ($Z = 11, 12$), are found to deviate strongly from the predictions by the sd shell model. Nuclear masses and binding energies, excitation energies and $\beta$-decay properties, electromagnetic moments, all confirm that the ground states of the neutron rich Na and Mg isotopes with about twenty neutrons...
[2–7] are dominated by neutron excitations from the sd- to the pf-orbit (called ‘intruder’ states). The shell-model interaction has consequently been altered and the model space extended, in order to reproduce the unusual properties of these exotic nuclei. The earliest modification (Warburton et al., 1990 [8]) explains the special behavior of nuclei belonging to the so-called ‘island of inversion’, a region of nuclei with $Z = 10, 11, 12$ and $N = 20, 21, 22$. This name refers to the fact that the ground state wave function is dominated by particle-hole excitations of neutrons across the reduced $N = 20$ shell gap, which has experimentally been confirmed for the nuclei in dark-grey in Fig. 1. Later experiments have proven that also isotopes with $N = 18$ [6] or $N = 19$ [7] belong to this island of inversion. New interactions were developed (e.g. Caurier et al., 1998 [9]; Otsuka et al., 1999 [10]) as more experimental evidence became available. For the nuclei at the borders of this island, the different models predict different ground state properties, and thus experiments are needed to further refine the model parametrizations.

The Al isotopes with $Z = 13$ protons are located at the border of the island of inversion. Isotopes with just one proton more (Si) are known to have a normal shell structure at low energy [11,12], while the ground states of isotopes with one proton less ($^{32}$Mg ($N = 20$) and $^{31}$Mg ($N = 19$)) are dominated by particle-hole (intruder) configurations [3,7,13]. Thus the transition from the ‘normal’ shell model region into the ‘island of inversion’ is expected to happen in the Al isotopes. Up to $^{31}$Al ($N = 18$) the low-energy spectrum of these isotopes [14–16] is found to be in very good agreement with the USD shell-model. In $^{32}$Al evidence for an intruder configuration around 1 MeV is found [17], while the ground state is suggested to be a normal sd-state based on a recent $g$ factor measurement [14]. Also in $^{33}$Al a low-lying state around 700 keV is suggested to be an intruder state [18], while no evidence for mixing with intruder configurations in the ground state is found from a $\beta$-decay study [19].

This Letter reports on the first precise measurements of the ground state $g$ factors of the neutron rich Al isotopes from $^{31}$Al up to $^{33}$Al ($N = 20$). The obtained precision is a result of significant improvements in the experimental method and careful studies on the systematic errors. Such data can provide information on small admixtures of $2p-2h$ intruder configurations in the wave function of exotic nuclei, as demonstrated recently for the neutron-rich Na isotopes [20].

Neutron rich Al isotopes have been produced in the fragmentation of a $^{36}$S$^{16+}$ beam (77.5 MeV/μ) on a $^{9}$Be target (~1 mm). Secondary beams of $^{31,32,33}$Al with purities of respectively 95–85–97% were selected using the high-resolution fragment separator LISE at GANIL [21,22]. A $^{9}$Be wedge degrader of about 1 mm thickness (198 mg/cm$^2$) was used to obtain such high beam purities. Fragments were identified by standard energy loss versus time of flight measurements using Si pin diodes (500 μm thickness). Spin polarization was obtained by deflecting the primary beam by $2(1)^{\circ}$ with respect to the spectrometer entrance where the fragmentation target was mounted. In order to obtain the highest polarization, fragments are selected in the wing of their longitudinal momentum distribution [23]. The polarized beam rates that were implanted in our crystal varied from 15000/s for $^{31}$Al ($t_{1/2} = 644$ ms), 5000/s for $^{32}$Al ($t_{1/2} = 33$ ms) to 1500/s for $^{33}$Al ($t_{1/2} = 42$ ms [19]).

A detailed description of the experimental set-up and methodology can be found in [15,24]. A new movable crystal holder was installed in the vacuum chamber, on which several crystals can be mounted. This allowed easy identification of the best implantation host for maintaining the reaction induced polarization. Both metallic and ionic crystals have been investigated, all having a cubic lattice structure. A sufficiently high magnetic field is applied to maintain the spin orientation after implantation. The magnetic field is measured with a Hall probe at 7 cm from the crystal, the read-out being integrated in the data acquisition system. To calibrate the field at the position of the crystal a field mapping was performed before and after each experiment. The asymmetry in the $\beta$-decay of the polarized nuclei is observed in two scintillator telescopes. Scattering or noise events are excluded by requiring a coincidence between signals in a thin and a thick scintillator. The ratio of coincident counts in the Up and Down telescopes, $R = N_{up}/N_{down}$, is proportional to the amount of polarization in the implanted ensemble:

$$R = \frac{N_{up}}{N_{down}} = R_0 \left(1 + A_1 B_1^0 Q_1^{\exp} \right) / \left(1 - A_1 B_1^0 Q_1^{\exp} \right) \approx R_0 \left(1 + 2A_1 B_1^0 Q_1 \right). \tag{1}$$

$R_0$ is the experimental asymmetry (e.g. due to different efficiency of both telescopes), $A_1$ is the asymmetry parameter of the $\beta$-decay, $B_1^0$ is the first order orientation tensor related to the spin-polarization and $Q_1$ quantifies the experimental asymmetry losses (e.g. due to relaxation, scattering of $\beta$-particles, experimental geometry, ...). A coil, which is part of a series LRC circuit, is placed around the implantation crystal to induce a radio-frequent (rf) field with frequency $\nu_{RF}$. When this frequency matches the Larmor frequency $\nu_L = g\mu_N B_0/h$ the polarization in the ensemble is resonantly destroyed [25]. From this resonance frequency and the applied magnetic field, the $g$ factor can be derived.

Before starting a $g$ factor measurement, the experimental conditions to detect maximum polarization are optimized. All experiments have been done at room temperature. First the amount of polarization maintained in different crystals is investigated. Then the reaction-induced polarization is investigated as a function of the longitudinal momentum of the selected frag-
ment beam. The procedure for this is described in [26]. The polarization of the Al isotopes is maintained best in a Si single crystal (asymmetries up to 6% were observed) while this was up to 2 and 3 times less in MgO and NaCl.

Once polarization is established, it can be resonantly destroyed by the rf-field. Each rf-frequency was modulated continuously around a fixed value over a modulation range: $\nu_{RF} \pm \Delta \nu_{RF}$. When the applied frequency range covers the Larmor frequency, the polarization is destroyed and this is reflected as a change in the $\beta$-decay asymmetry. The resonance condition can be searched by changing the magnetic field $B_0$ or the rf-frequency. The normalized asymmetry $R_N$ is plotted, such that one can directly deduce the amount of polarization destroyed at resonance:

$$R_N = \frac{R - R_{\text{base}}}{R} \approx -2A_1 \sqrt{\frac{3I}{I+1}} P Q_1^{\text{exp}}.$$ (2)

The baseline $R_{\text{base}}$ of the resonance curves is defined by measuring the $\beta$-decay asymmetry without rf-field applied.

Since a series LRC circuit with a narrow power peak (FWHM $\sim 100$ kHz) is used to generate a high-power rf-signal, only a small $g$ factor range can be scanned as a function of frequency (Fig. 2(b)). For isotopes with unknown ground state structure, we first search in a large $g$ factor range by changing the static magnetic field $B_0$. The rf-frequency is fixed to a certain value and modulated over a broad range. Like that an extended $g$ factor region can be scanned with a few values of $B_0$, as demonstrated in Fig. 2(a). All data were taken in a sweep mode in order to average out experimental asymmetry fluctuations during the measurement. The static field is changed every two minutes and during the field change (about 3 s) no data are collected. For every field value data are collected with and without rf-field to determine $R_{\text{base}}$. The rf-frequency is modified every ten seconds in a frequency scan because the rf-frequency changes instantaneously. $R_{\text{base}}$ is determined after each full sweep in this case.

The observed NMR data are fit with a theoretical curve that is obtained by numerically solving the time evolution equation with a Hamiltonian that describes the time-dependent interaction of the nuclei with the static magnetic field and the modulated rf-field. An inhomogeneous line broadening of the NMR resonance is taken into account by convoluting the resulting resonance with a Gaussian line shape. More details are given in [26]. The statistical error on the $g$ factor is deduced by fitting the data with the calculated function, using a multi-parameter $\chi^2$-minimization procedure with the $g$ factor (determining the resonance position) and the amount of destroyed polarization (determining the amplitude) as parameters. The rf-field strength and the Gaussian line width (determining the line shape and width) are determined by making a consistent fit of many resonance curves taken for different rf-conditions [26]. This was done for the Al isotope with the highest NMR amplitude, namely $^{32}$Al. The systematic uncertainty on our data is investigated by measuring NMR resonances for $^{32}$Al in Si during four different experiments. The error due to the magnetic field calibration, which is made for every experiment independently, is derived from the scattering on these values. The standard deviation is 0.11% as shown in Fig. 3(a), and this will be taken as our systematic error. This systematic error is almost an order of magnitude larger than most statistical errors. As final error the square root of the sum of the squared statistical and system-
atic uncertainties is used. This total error is shown as well on the data points of Fig. 3(a), demonstrating that with this error a normal statistical ensemble is obtained. In Fig. 3(b) the results of five measurements on $^{32}$Al, obtained during the same experiment, are presented. The weighted mean value of these five measurements gives $g = 1.9516(4)$, and including the systematic error we obtain $g(^{32}$Al) = 1.9516(22). This value is in agreement with the earlier reported value $g = 1.959(9)$ [14].

As the $g$ factor of $^{31}$Al was known from [15], a field scan with a small frequency modulation $\nu_{RF} = (1100 \pm 10)$ kHz was performed (Fig. 4(a)). The deduced value $g(^{31}$Al) = 1.532(2), which includes also a 0.11% systematic error, is in a good agreement with the earlier published value $g = 1.517(20)$ [15]. The measurement on $^{33}$Al was performed during the same experiment: First a field scan was measured with the same rf-frequency and a large modulation $\nu_{RF} = (1100 \pm 50)$ kHz from which we deduced $g = 1.68(6)$. Then a fine scan is made (Fig. 4(b)), resulting in $g(^{33}$Al) = 1.635(2).

The experimental $g$ factors and level schemes of $^{30}$–$^{33}$Al are compared to the results from large scale shell model calculations performed with the Antoine code [27]. Two interactions have been used: The USD interaction [1] with protons and neutrons restricted to the sd-shell and the sdpf–sm interaction [4] with different model spaces for the neutrons. All $g$ factors are calculated using free nucleon values for $g_s$ and $g_l$. The experimental level schemes and $g$ factors of $^{30}$Al and $^{31}$Al (Fig. 5) are in very good agreement with the calculated ones. A less good agreement is observed for $^{32}$Al: The first two excited states observed by Robinson et al. [28] are inverted with respect to the calculations, and the suggested 1p–1h intruder state observed by Fornal et al. [17] is reproduced only by the sdpf–sm calculation if the neutron model space includes the $f_{7/2}p_{3/2}$ orbits. Thus the first three excited states in $^{32}$Al do not support a normal sd shell-model picture. In Table 1 the $g$ factors of all odd–even and odd–odd Al isotopes are compared to the calculated values and a good general agreement is observed (less than 2% deviation for most isotopes). For $^{32}$Al the 6% higher experimental value cannot be attributed to a missing intruder contribution in the ground state wave function. Such a 2p–2h contribution would significantly lower the $g$ factor, as a pure 2p–2h configuration has $g(\nu \nu, 2p–2h, 1^+) = 0.34$. Thus the deviation is entirely due to the composition of the wave function in the sd-shell.

In trying to understand the possible origin of this deviation, we performed calculations in different restricted model spaces for the protons (Fig. 6). The major contribution in the $^{32}$Al wave function comes from the $(\pi d_{5/2} \nu d_{3/2})_1^+$ configuration, for which the Schmidt value is $g(\pi d_{5/2} \nu d_{3/2}, 1^+) = 2.78$ using the additivity rule [29]. All other sd-shell configurations have a Schmidt value below 1, thus lowering the calculated $g$ factor with respect to this Schmidt value as illustrated in Fig. 6. No-

---

[Fig. 4](#) NMR scans as a function of the magnetic field for $^{31}$Al (a) and $^{33}$Al (b) implanted in a Si single crystal. The applied frequency is 1100 kHz.

[Fig. 5](#) Comparison between calculated and experimental level schemes and $g$ factors for $^{30,31,32}$Al.
notice that the ground state g factor and the energies of the excited states are sensitive to different components in the wave function: The g factor is hardly modified by including the $\pi s_{1/2}$ orbit while the energy of the $4^+$ state is modified drastically. On the other hand, by including also the $\pi d_{3/2}$ orbital the level scheme remains similar, while it strongly reduces the g factor. This illustrates that g factors and excitation energies should both be considered when fitting the parameters for a new interaction. Less contribution from the $\pi d_{3/2}$ orbital would lead to a better agreement for both the g factor and the energies of the first excited states in $^{32}$Al. This suggests that the $Z = 16$ shell gap might play a role in explaining the disagreement for both the $^{32}$Al g factor and its energy level scheme.

For the odd Al isotopes the experimental g factors all deviate less than 2% from the calculated values (Table 1), except for the semi-magic $^{34}$Al ($N = 20$) which deviates almost 4%. Furthermore, one expects the lowest excited state around 3 MeV (a nearly pure $\pi s_{1/2}$ state) if no excitations of neutrons are allowed across the $N = 20$ shell gap (Fig. 7). Experimentally, the first excited state is observed as low as 730 keV [18], and it was suggested that this is a 2p–2h $5/2^+$ intruder state. With the sdpf–sm interaction a pure 2p–2h intruder state appears indeed around 1.5 MeV in a reduced $f_{7/2}p_{3/2}$ space and around 650 keV if the full pf space is used (right part of Fig. 7). The g factor of these intruder states is lower than that of the sd shell state, and the experimental value is found to be in-between both. Thus a configuration where some amount of intruder configurations is mixed with the normal sd shell configuration could explain the observed ground state g factor. With the present sdpf–sm interaction, the calculated g factor reduces (from 1.702 → 1.689 → 1.680) if two neutrons are allowed in the reduced or full pf space, due to the fact that a small amount of intruder components occurs (up to 10% in these calculations). This suggests that intruder configurations do play an important role in the low-energy structure of $^{33}$Al and that in the present interaction this amount is too low. Note that the half-life and branching ratio for both $\beta$ and $\beta$-delayed neutron decay of $^{33}$Al has been well-described by the sd shell-model [19], illustrating again that different observables are sensitive to different parts of the wave function.

Our experimental value suggests that the amount of intruder mixing is at least 25%. This is supported by calculations performed by Utsuno et al. using the Monte Carlo Shell Model in the $sd$–$p_{3/2}f_{3/2}$ model space with the SDFP-M interaction [10]. They predict that the ground state of $^{33}$Al has 50% of intruder mixture with a g factor $g(^{33}\text{Al})_{\text{MCSM}} = 1.55$ [31] (using free nucleon values). Later, the interaction was modified in its monopole part as to reproduce the experimental g factors and quadrupole moments of the Na isotopes (Fig. 8) and using effective g factors [20]. The value for $^{33}$Al calculated with this modified interaction and effective g factors [32] is in very good agreement with our experimental number, as shown on (Fig. 8).
In conclusion, the $g$ factors of exotic Al-isotopes, produced via projectile-fragmentation reactions, were measured with the NMR technique, after optimizing all NMR conditions. This has lead to a significant amount of polarization (up to 3.5%), allowing precision studies for the $g$ factors up to the $N = 20$ nucleus $^{33}$Al. The results are compared with large scale shell model calculations. The nuclear properties of $^{30}$Al and $^{31}$Al are in very good agreement with the theoretical predictions in the sd-shell model space, placing these isotopes clearly outside the island of inversion. No evidence for the presence of intruder configurations in the ground state of $^{32}$Al was found, but the experimental level scheme and the ground state $g$ factor might be better reproduced with an enhanced $Z = 16$ shell gap. For $^{33}$Al, the first indication is observed for a non-negligible (at least 25%) contribution of intruder configurations into the ground state wave function. The sdpf–sm interaction does not predict enough intruder mixing in order to explain the measured $g$ factor.

**Acknowledgements**

This work has been supported by the European Community FP6—Structuring the ERA—Integrated Infrastructure Initiative contract EURONS No. RI3-CT-2004-506065. P.H. is grateful to the IWT-Vlaanderen (Institute for the Promotion of Innovation through Science and Technology in Flanders) for providing a scholarship. This work was supported also by INTAS 00-463, the IUAP project No. p5-07 of OSCT Belgium and by the FWO-Vlaanderen. D.L.B. acknowledges support through grant BgNSF VUF06/05.

**References**

Effective field theory calculation of nd radiative capture at thermal energies

H. Sadeghi \(^a,\) S. Bayegan \(^b\), Harald W. Grießhammer\(^c\)

\(^a\) Department of Physics, University of Arak, PO Box 38156-879, Arak, Iran
\(^b\) Department of Physics, University of Tehran, PO Box 14395-547, Tehran, Iran
\(^c\) Center for Nuclear Studies, Department of Physics, The George Washington University, Washington, DC 20052, USA

Received 2 May 2006; received in revised form 26 September 2006; accepted 9 October 2006
Available online 27 October 2006
Editor: W. Haxton

Abstract

The cross section for thermal neutron capture by the deuteron is calculated with pionless effective field theory (EFT). No new three-nucleon forces are needed up to next-to-next-to leading order in order to achieve cut-off independent results, besides those fixed by the triton binding energy and nd scattering length in the triton channel. The cross section is accurately determined to be \(\sigma_{\text{tot}} = [0.503 \pm 0.003]\) mb. At zero energies, the magnetic \(M_1\) transition gives the dominant contribution and is calculated up to next-to-next-to leading order (N\(^2\)LO). Close agreement between the available experimental data and the calculated cross section is reached. We demonstrate convergence and cutoff independence order by order in the low-energy expansion.

\(\text{©} 2006\) Elsevier B.V. All rights reserved.
PACS: 27.10.+h; 21.45.+v; 25.40.Lw; 11.80.Jy; 21.30.Fe

Keywords: Effective field theory; Three-body system; Three-body force; Faddeev equation; Radiative capture

1. Introduction

The study of the three-body nuclear system involving neutron radiative capture by deuteron has been investigated in theoretical and experimental works over the past years. The experimental result of this process has most accurately been measured by Jurney et al. [1]. The value of 0.508 \(\pm 0.015\) (mb) for the cross-section was resulted for 2200 m/s neutrons.

Rapid progress has been made in the theoretical study of the nd \(\rightarrow^3\text{H}\gamma\) reaction such as the p–d and n–d radiative capture. At such energies a magnetic dipole \((M_1)\) transition is almost entirely participated. These reactions were studied in plane wave (Born) approximation by Friar et al. [2]. In these investigations the authors employed their configuration-space Faddeev calculations of the helium wave function, with inclusion of three-body forces and pion exchange currents. More recently a rather detailed investigation of such processes has been performed by Viviani et al. [3,19]. In their calculations the quite accurate three-nucleon bound- and continuum-states were obtained in the variational pair-correlated hyperspherical method from a realistic Hamiltonian model with two- and three-nucleon interactions.

They obtained in Ref. [3] the cross-section from Argonne \(v_{14}\) two-nucleon and Urbana VIII three-nucleon interactions (AV14/UVIII), also from Argonne \(v_{18}\) two-nucleon and Urbana IX three-nucleon interactions (AV18/UIX) and including \(\Delta\) admixtures. Cross-section values were found 0.600 (mb) and 0.578 (mb) which overestimate the experimental value by 18% and 14% value, respectively, see Table 2. It should be noted, however, that the explicit inclusion of \(\Delta\)-isobar degrees of freedom in the nuclear wave function are found to be in significantly better agreement with experiment than those obtained from perturbation theory, \(\Delta\)PT, estimates. This shows that their results for this very-low energy observable are sensitive to details of the short-range part of interaction. A recent calculation using manifestly gauge-invariant currents reduced the spread [19], but the result including three-body currents,
0.558 mb, still over-predicts the cross-section by 10%. Model-dependent currents associated with the Δ(1232) were identified as source of the discrepancy. Thus, the question remains how such details of short-range physics can so severely influence a very-long-range reaction with maximal energies of less than 10 MeV.

During the last few years, nuclear effective field theory (EFT) has been applied to two-, three-, and four-nucleon systems, see e.g. [4–10]. The pionless effective field theory would be an ideal tool to calculate low-energy cross sections in a model-independent way and to possibly reduce the theoretical errors by a systematic, model-independent calculation with an a priori estimate of the theoretical uncertainties. An example of a precise calculation is the reaction np → γd, which is relevant to big-bang nucleosynthesis (BBN). The cross section for this process was computed to 1% error for center of mass energies $E \lesssim 1$ MeV [11–13].

We have suggested a method for computation of neutron–deuteron radiative capture for extremely low energy ($20 \leq E \leq 200$ keV) with pionless EFT [15], with where this formalism, we can estimate errors in a perturbative expansion up to N2LO within a few percent of the ENDF values [16].

The purpose of the present Letter is to study of the cross section for radiative capture of neutrons by deuterons nd → γ3H at zero energies with pionless EFT. At these energies, the magnetic $M_1$ transition gives the dominant contribution. The $M_1$ amplitude is calculated up to next-to-next-to-leading order (N2LO) with insertion of three-body force. Results show less than 1% deviation from the available experimental data at zero energy (0.0253 eV).

This Letter is organized as follows. In the next section, a brief description of the formalism and its input for total cross section of the neutron–deuteron radiative capture will be presented. We discuss the theoretical errors, tabulation of the calculated cross section in comparison with the other theoretical approaches and the newly available experimental data [1] in Section 3. Finally, summary and conclusions follow in Section 4.

2. Neutron–deuteron scattering in triton channel and radiative capture

The $^2S_{1/2}$ channel to which $^3$He and $^3$H belong is qualitatively different from the other three-nucleon channels because all three nucleons can occupy the same points in space. Consequently, $^2S_{1/2}$ describes the preferred mode for nd → $^3$Hy and pd → $^3$Hey. The three-nucleon Lagrangian is well-known and will not be repeated here, see e.g. [14,18] for details.

The derivation of the integral equation describing neutron–deuteron scattering has also been discussed before, see e.g. [7, 18]. We present here only the results. The integral equation is solved numerically by imposing a cut-off $\Lambda$. In that case, a unique solution exists in the $^2S_{1/2}$ channel for each $\Lambda$ and vanishing three-body force, but no unique limit as $\Lambda \rightarrow \infty$. As long-distance phenomena must however be insensitive to details of the short-distance physics (and in particular of the regulator chosen), Bedaque et al. [6,7,14,18] showed that the system must be stabilized by a three-body force

$$\mathcal{H}(E; A) = \frac{2}{\Lambda^2} \sum_{n=0}^{\infty} H_{2n}(\Lambda) \left( \frac{ME + \gamma_r^2}{\Lambda^2} \right)^n$$

$$= \frac{2H_0(\Lambda)}{\Lambda^2} + \frac{2H_2(\Lambda)}{\Lambda^4} (ME + \gamma_r^2) + \cdots,$$

which absorbs all dependence on the cut-off as $\Lambda \rightarrow \infty$. It is analytical in $E$ and can be obtained from a three-body Lagrangian, employing a three-nucleon auxiliary field analogous to the treatment of the two-nucleon channels [14]. Contrary to the terms without derivatives, there are different, inequivalent three-body force terms with two derivatives, but only one of them, $H_2$, is enhanced over its naive dimensional estimate, mandating its inclusion in N2LO [14,20]. Neutron–deuteron scattering amplitude including the new term generated by the two-derivative three-body force is shown schematically in Fig. 1. We refer to Ref. [18] for details of the notation. Two amplitudes get mixed: $t_s$ describes the $d_t + N \rightarrow d_s + N$ process, and $t_t$ describes the $d_t + N \rightarrow d_t + N$ process, where $d_t$ ($d_s$) is an auxiliary field of two nucleons in a relative singlet-S (triplet-S) wave:

$$t_s(p, k) = \frac{1}{4} \left[ 3K(p, k) + 2\mathcal{H}(E, A) \right]$$

$$+ \frac{1}{2\pi} \int_0^\Lambda dq \ q^2 \left[ D_s(q) [K(p, q) + 2\mathcal{H}(E, A)] t_s(q) \right]$$

$$+ D_t(q) \left[ 3K(p, q) + 2\mathcal{H}(E, A) \right] t_t(q),$$

$$t_t(p, k) = \frac{1}{4} \left[ K(p, k) + 2\mathcal{H}(E, A) \right]$$

$$+ \frac{1}{2\pi} \int_0^\Lambda dq \ q^2 \left[ D_t(q) [K(p, q) + 2\mathcal{H}(E, A)] t_t(q) \right]$$

$$+ D_s(q) \left[ 3K(p, q) + 2\mathcal{H}(E, A) \right] t_s(q),$$

where $D_s,t(q) = D_s,t(E - \frac{q^2}{2\pi}, q)$ are the propagators of deuteron and $K$ the projector of the exchanged nucleon, projected into the S-wave. For the spin-triplet S-wave channel, one determines the two-nucleon interaction up to N2LO by the deuteron binding momentum $\gamma_r = 45.7025$ MeV and effective range $\rho_z = 1.764$ fm. Because there is no real bound state in the spin singlet channel of the two-nucleon system, its free parameters are better determined by the scattering length $a_s = 1/\gamma_s = -23.714$ fm and the effective range $r_s = 2.73$ fm at zero momentum.

The neutron–deuteron $J = 1/2$ phase shifts $\delta$ are determined by the on-shell amplitude $t_i(k, k)$, multiplied with the wave function renormalisation

$$T(k) = Z t_i(k, k) = \frac{3\pi}{M} \frac{1}{k\cot \delta - ik}.$$

At thermal energies, the reaction proceeds through S-wave capture predominantly via a magnetic dipole transition, $M_1^{J=1/2}$, where $L = 0$, $S = 1/2$, $3/2$ and $I = 1$. To obtain the spin structure, which corresponds to a definite value of $J$ for the entrance channel, it is necessary to build special linear combinations of
products $\tilde{D}N$ and $\tilde{\sigma} \times \tilde{D}N$, with $J^P = \frac{1}{2}^+$ or $J^P = \frac{3}{2}^+$ and $\tilde{D}$ the deuteron spin-one field, see [15] for details:

$$\hat{\phi}_{1/2} = (i\tilde{D} + \tilde{\sigma} \times \tilde{D})N \quad \text{and} \quad (2i\tilde{D} - \tilde{\sigma} \times \tilde{D})N.$$ 

For both possible magnetic dipole transitions with $J^P = \frac{1}{2}^+$ (amplitude $g_1$) and $J^P = \frac{3}{2}^+$ (amplitude $g_3$) we can write:

$$g_1: \quad i^+(i\tilde{D} \cdot \tilde{e}^a \times \tilde{k} + \tilde{\sigma} \times \tilde{D} \cdot \tilde{e}^a \times \tilde{k})N,$$

$$g_3: \quad i^+(i\tilde{D} \cdot \tilde{e}^a \times \tilde{k} + \tilde{\sigma} \times \tilde{D} \cdot \tilde{e}^a \times \tilde{k})N.$$ 

The contribution of the electric transition $E_{L,S,J}^{i}$ for energies of less than 60 keV to the total cross section is indeed very small. Therefore, the electric quadrupole transition $E_2^{(0/5/2)(3/2)}$ from the initial quartet state will not be considered at thermal energies. The $M_1$ amplitude receives contributions from the magnetic moments of the nucleon and dibaryon operators coupling to the magnetic field, which are described by the Lagrange density

$$L_B = \frac{e}{2M_N} N^3 (k_0 + k_1 \tau^3) \sigma_B \cdot B + e \frac{L_1}{M_N \sqrt{\rho^{(S_0)}}} \rho^{(S_1)} d_i \sigma^\dagger d_i B_j + \text{h.c.},$$

where $k_0 = 1/2(k_0 + k_1) = 0.4399$ and $k_1 = 1/2(k_0 - k_1) = 2.35294$ are the isoscalar and isovector nucleon magnetic moment in nuclear magnetons, respectively. The coefficient $L_1$ is fixed at its leading non-vanishing order by the total cross section [11].

The radiative capture cross section $\text{nd} \rightarrow \text{^3H}$ at very low energy is given by

$$\sigma = \frac{2}{9} \frac{\alpha}{v_{\text{rel}}} \frac{\rho^3}{4M_N^3} \sum_{L,S,J} \left| \tilde{\chi}_{LSJ} \right|^2,$$

where

$$\tilde{\chi}_{LSJ} = \frac{\sqrt{6\pi}}{p_{\mu N}} \frac{\sqrt{4\pi}}{\rho_{\mu N}} \chi_{LSJ},$$

with $\chi$ standing for either $E$ or $M$ and $\mu_N$ is nuclear magneton and $p$ is momentum of the incident neutron in the center of mass.

We now turn to the Faddeev integral equation to be used in the $M_1$ calculation. We solve the Faddeev equation for nd-scattering and also for the triton bound state to some order (e.g. LO), then we take these Faddeev amplitudes and sandwich the photon-interactions with nucleons between them when the photon kernel is expanded to the same order. This process will be done separately for NLO and N$^2$LO. Finally the wave function renormalization in each order will be done.

The diagrams in Fig. 2 represent contributions of electromagnetic interaction with nucleon, deuteron, four-nucleon-magnetic-photon operator described by a coupling between the $^3S_1$-dibaryon and $^1S_0$-dibaryon and a magnetic photon. As mentioned in the introduction, in another paper [15], we have presented the detailed schematics of these diagrams in neutron–deuteron radiative capture for $(20 \leq E \leq 200 \text{ keV})$ up to N$^2$LO.

The last diagram in Fig. 2 with insertion of a photon to the N$^2$LO three-nucleon force $H_2$ vertex is not $M_1$ and we know that $M_1$ contribution is the dominant contribution at very low energy and especially for zero energy. Its contribution should therefore be very tiny. Because the leading three-nucleon force $H_0$ has no derivatives, it is not affected by the minimal substitution $p \rightarrow p - eA$. But the parameter $H_2$ is the strength of the three-nucleon interaction with two derivatives. Naturally for the energy range near zero momentum, insertion of photon to $H_2$ vertices for momentum $p \sim 0.025$ eV and $M_1$ transition, could be neglected. $H_2$ is necessary in neutron–deuteron scattering to improve cut-off independence but is defined such that it does not contribute at zero momentum. Contributions of a photon coupling to $H_2$ are however indeed negligible at zero energy.

3. Neutron–deuteron radiative capture results at zero energy

We numerically solved the Faddeev integral equation up to N$^2$LO. We used $\hbar c = 197.327 \text{ MeV fm}$, a nucleon mass of $M = 938.918 \text{ MeV}$, for the $NN$ triplet channel a deuteron binding energy (momentum) of $B = 2.225 \text{ MeV}$ ($\gamma_B = 45.7066 \text{ MeV}$), a residue of $Z_d = 1.690(3)$, for the $NN$ singlet channel an $^1S_0$ scattering length of $a_{1S0} = -23.714 \text{ fm}$ and effective range $r_s = 2.73 \text{ fm}$. $L_1 \sim -4.5 \text{ fm}$ by fixing at its leading non-vanishing order by the thermal cross section.

As in Ref. [20], we can determine which three-body forces are required at any given order, and how they depend on the cut-
off. Low-energy observables must be insensitive to the cut-off, namely to any details of short-distance physics in the region above the break-down scale of the pion-less EFT, set approximately by the pion-mass. It was found in Ref. [20] that no additional three-nucleon forces are necessary to render a renormalisable amplitude at N^2LO in this process, besides those needed already in nucleon–deuteron scattering: H_0 and H_2.

At N^2LO, where we saw that H_2 is required, we checked this by varying the cut-off between 150 and 500 MeV. This is a reasonable estimate of the errors of our calculation due to higher-order effects. As seen in Fig. 3, in the thermal energy range the cutoff variation is very small and decreases steadily as we increase the order of the calculation and it is of the order of (k/Λ)^n, (γ/Λ)^n, where n is the order of the calculation and Λ = 150 MeV is the smallest cutoff used (see Table 1 and Fig. 3). Also, errors due to cutoff variation are decreasing when the order of calculation is increased up to N^2LO.

We determined the two-nucleon parameters from the deuteron binding energy, triplet effective range (defined by an expansion around the deuteron pole, not at zero momentum), the singlet scattering length, effective range (defined by expanding at zero momentum), the singlet binding energy, triplet effective range (defined by an expansion around the deuteron pole, not at zero momentum), and two body capture process (obtained with comparison between experimental data and theoretical results for np → dγ process at zero energy [12]). We fix the three-body parameters as follows: because we defined H_2 such that it does not contribute at zero momentum scattering, one can first determine H_0 from the 2S1/2 scattering length a_3 = (0.65 ± 0.04) fm [17]. At LO and NLO, this is the only three-body force. At N^2LO, H_2 is required. It is determined by the triton binding energy B_3 = 8.48 MeV. Finally, we solve by insertion of the potential at a given order in the integral equation and iteration of kernel.

The cross section for neutron–deuteron radiative capture as function of the center-of-mass kinetic energy E in MeV. The short dashed, long dashed and solid lines correspond to the contribution of M_1 capture cross section up to LO, NLO and N^2LO, respectively. Single point shows experimental results for this cross section at 0.025 eV [1].

Table 2 shows a comparison between different theoretical results for neutron radiative capture by deuteron at zero energy (0.0253 eV). Last row shows our EFT result. The last line quotes deviation between data [1] and theory, if it is larger than the theoretical or experimental uncertainty.

**Table 1**

Results for the relative cutoff variation of the cross section up to N^2LO is shown between Λ = 150 MeV and Λ = 500 MeV

<table>
<thead>
<tr>
<th>E (10^{-8} MeV)</th>
<th>LO</th>
<th>NLO</th>
<th>N^2LO</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0006</td>
<td>0.00005</td>
<td>0.0000002</td>
</tr>
<tr>
<td>2</td>
<td>0.0010</td>
<td>0.00040</td>
<td>0.0000005</td>
</tr>
<tr>
<td>2.65</td>
<td>0.0012</td>
<td>0.00060</td>
<td>0.0000090</td>
</tr>
<tr>
<td>3</td>
<td>0.0014</td>
<td>0.00084</td>
<td>0.0000150</td>
</tr>
<tr>
<td>10</td>
<td>0.0020</td>
<td>0.00131</td>
<td>0.0000600</td>
</tr>
</tbody>
</table>

**Table 2**

Comparison between different theoretical results for neutron radiative capture by deuteron at zero energy (0.0253 eV). Last row shows our EFT result. The last line quotes deviation between data [1] and theory, if it is larger than the theoretical or experimental uncertainty.

<table>
<thead>
<tr>
<th>Theory</th>
<th>σ (nb)</th>
<th>Deviation from exp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AV14/VIII(IA + MI + MD) [3]</td>
<td>0.509</td>
<td></td>
</tr>
<tr>
<td>AV18/IX(IA + MI + MD) [3]</td>
<td>0.489</td>
<td></td>
</tr>
<tr>
<td>AV14/VIII(IA + MI + MD + 4πρτ) [3]</td>
<td>0.658</td>
<td>29%</td>
</tr>
<tr>
<td>AV18/IX(IA + MI + MD + 4πρτ) [3]</td>
<td>0.631</td>
<td>24%</td>
</tr>
<tr>
<td>AV14/VIII(IA + MI + MD + Δ) [3]</td>
<td>0.600</td>
<td>18%</td>
</tr>
<tr>
<td>AV18/IX(IA + MI + MD + Δ) [3]</td>
<td>0.578</td>
<td>14%</td>
</tr>
<tr>
<td>AV18/IX(gauge inv) [19]</td>
<td>0.523</td>
<td></td>
</tr>
<tr>
<td>AV18/IX(gauge inv + 3N-current) [19]</td>
<td>0.556</td>
<td></td>
</tr>
<tr>
<td>EFT(LO)</td>
<td>0.485</td>
<td>5%</td>
</tr>
<tr>
<td>EFT(NLO)</td>
<td>0.496</td>
<td>2.3%</td>
</tr>
<tr>
<td>EFT(N^2LO)</td>
<td>0.503 ± 0.003</td>
<td></td>
</tr>
<tr>
<td>Experiment [1]</td>
<td>0.508 ± 0.015</td>
<td></td>
</tr>
</tbody>
</table>
above, this is not the case: There are no new three-nucleon forces besides those already fixed in nd scattering at the same order. The contribution from the photon coupling to a three-nucleon force is negligible in our calculation. As our result is model-independent and universal, any model with the same input must—within the accuracy of our calculation—lead to the same result. Our inputs are the first two terms of the effective-range expansion in the singlet- and triplet-S wave of NN scattering, the proton and neutron magnetic moments, the triton binding energy and nd scattering length in the doublet-S-wave, and finally the thermal cross section of the reaction np → dγ (determining $L_1$). More work is needed to understand why the potential-model calculations [3,19] have the same input but do not seem to reproduce the same result.

Addressing convergence of the EFT calculation, we notice that the contributions which are characterised as higher-order in the power-counting are indeed small: The LO result is 0.485 mb, with NLO adding 0.011 mb, and N2LO another 0.007 mb. Cut-off dependence is negligible. The typical size of the expansion parameter in the pion-less EFT is about $\gamma_t/m_\pi \approx 1/3$. We therefore estimate the uncertainty from leaving out corrections at N3LO and higher as about 1/3 of the N2LO corrector or 0.003 mb.

4. Conclusion

The cross section for radiative capture of neutrons by deuterons nd → γ 3H at zero energies was calculated with pionless effective field theory, the unique, model independent and systematic low-energy version of QCD for processes involving momenta below the pion mass. We applied pionless EFT to find numerical results for the $M_1$ contributions. Incident thermal neutron energies have been considered for this capture process. At these energies our calculation is dominated by only $S$-wave state and magnetic transition $M_1$ contribution. The $M_1$ amplitude is calculated up to next-to-next to leading order N2LO. Three-nucleon forces are needed up to N3LO for cut-off independent results. The triton binding energy and nd scattering length in the triton channel have been used to fix them. Hence the cross-section is in total determined as $\sigma_{tot} = [0.485(LO) + 0.011(NLO) + 0.007(N^{2}LO)] = [0.503 \pm 0.003]$ mb. It converges order by order in low energy expansion. It is also cut-off independent at this order. We notice that our calculation has a systematic error which is now smaller than the experimental error-bar.

Acknowledgement

We thank L. Marcucci and A. Kievsky for enlightening discussions which spurred this work.

References

  nddc.bnl.gov.
Single particle spectra based on modern effective interactions

C. Barbieri\textsuperscript{a,\textemdash b}

\textsuperscript{a} Gesellschaft für Schwerionenforschung, Planckstr. 1, 64291 Darmstadt, Germany\textsuperscript{1}
\textsuperscript{b} TRIUMF, 4004 Wesbrook Mall, Vancouver, BC, V6T 2A3, Canada

Received 19 May 2006; received in revised form 23 October 2006; accepted 25 October 2006
Available online 7 November 2006
Editor: J.-P. Blaizot

Abstract

The self-consistent Green’s function method is applied to \(^{16}\text{O}\) using a G-matrix and \(V_{\text{UCOM}}\) as effective interactions, both derived from the Argonne \(v_{18}\) potential. The present calculations are performed in a larger model space than previously possible. The experimental single particle spectra obtained with the G-matrix are essentially independent of the oscillator length of the basis. The results shows that \(V_{\text{UCOM}}\) better reproduces spin–orbit splittings but tends to overestimate the gap at the Fermi energy.

© 2006 Elsevier B.V. All rights reserved.
PACS: 24.10.Cn; 21.60.-n; 21.30.Fe; 21.60.Jz
Keywords: Spectral function; Effective interactions

A fundamental problem in nuclear physics is how to obtain descriptions of finite nuclei starting from a microscopic nuclear Hamiltonian. Much progress has been achieved for few body systems. The Green’s function Monte Carlo [1] technique is able to give exact results up to \(A = 12\), while the no-core shell model [2] has been applied to even larger nuclei. A wide range of exact methods is also available for very light systems [3]. In general, it has been found that both two- and three-nucleon (2N and 3N) forces are required to reproduce the experimental observations. Other recent attempts to push the limits of ab initio methods into the medium mass region have focused on the nucleus of \(^{16}\text{O}\) and its neighbor isotopes [4,5]. These works computed separation energies and spin orbit splittings of the orbits near the Fermi level. Coupled cluster theory appears to produce converged results for these nuclei [6]. These achievements have been possible by computing the contributions of long-range correlations (LRC) directly within very large models spaces where, however, one still needs to employ a proper effective interaction that accounts for the excluded degrees of freedom. In particular the effects due to short-range correlations (SRC) can be separated efficiently by such partitioning procedure, since they are characterized by high momenta degrees freedom [7].

Several ab initio methods employ similar partitioning techniques. Typically, two classes of microscopic approaches are possible to derive an effective interaction from a realistic nucleon–nucleon force [8]. Bloch–Horowitz theory makes use of the Feshbach projection formalism to devise an energy dependent interaction [9,10]. This gives solutions for every eigenstate with nonzero projection onto the model space, however, the energy dependence severely complicates the calculations. The G-matrix interaction [11], obtained by solving the Bethe–Goldstone equation, is also energy dependent. Alternatively, one can employ a proper unitary transformation to map a finite set of solutions of the initial Hamiltonian into states belonging to a numerically tractable space. In this case, one has the advantage to work with an energy independent interaction. Examples of such approaches are the Lee–Suzuki method [12] and the unitary correlator operator method (UCOM) [13–15]. The UCOM formalism is such that one can apply the inverse transformation to reinsert SRC into the nuclear wave function. A discussion of the similarities and differences between Lee–Suzuki and Bloch–Horowitz is given in Ref. [8]. Differently, one can derive a low momentum force, indicated as \(V_{\text{low-k}}\)
by using the renormalization group or the Lee–Suzuki method [8]. It should be noted that both $V_{\text{low-}}$ and $V_{\text{UCOM}}$ are phase shift equivalent at low energy and can be regarded as bare realistic interactions in this regime. The above methods, in principle, generate effective many-nucleon forces in addition to the 2N interactions and the intrinsic 3N ones. In practice, however, in calculating medium and large nuclei one wish to avoid as much as possible these complications, possibly by choosing interactions and model spaces that require weak overall 3N terms. It is therefore important to investigate how truncating to a 2N Hamiltonian affects the results for the different approaches outlined above.

In Ref. [17] we proposed to employ a set of Faddeev equations within the self-consistent Green’s function (SCGF) approach [7] to obtain a microscopic description of LRC. This allows to couple simultaneously quasiparticles (qp) and quasiholes (qh) to both particle–hole (ph) and particle–particle/hole–hole (pp/hh) collective excitations. The latter are eventually also expressed in terms of dressed qp and qh modes. Such formalism was later applied to $^{16}$O to investigate mechanisms that could possibly quench the spectroscopic factors of mean field orbits [18]. These calculations were already performed in a no-core fashion. However, the model space employed was still somewhat limited and phenomenological corrections were applied to tune the values of specific single particle (sp) energies (doing this allows studying correlations by artificially suppressing the couplings among selected excitation modes). Note that here and in the following we use the terms sp energies and sp spectra to refer to the poles of the one-body Green’s function (defined below Eq. (1)). These represent the excitation energies of the $A \pm 1$ neighbor nuclei, which are observable quantities. In this Letter the calculations of Ref. [18] are repeated by avoiding any phenomenology and employing a large model space. We discuss the results of 2N interactions belonging to the two types discussed above, namely a standard G-matrix and $V_{\text{UCOM}}$.

We consider the calculation of the sp Green’s function

$$g_{\alpha\beta}(\omega) = \sum_{\eta} \frac{\langle \Psi_0^\dagger | c^\dagger_{\alpha} c_{\beta} | \Psi_0 \rangle}{\omega - \epsilon_{\alpha}^+ - i\eta} + \sum_{\eta} \frac{\langle \Psi_0^\dagger | c^\dagger_{\beta} c_{\alpha} | \Psi_0 \rangle}{\omega - \epsilon_{\beta}^+ - i\eta},$$

from which both the one-hole and one-particle spectral functions, for the removal and addition of a nucleon, can be extracted. In Eq. (1), $\lambda_{\alpha}^\eta = \langle \Psi_0^\dagger | c^\dagger_{\alpha} c_{\eta} | \Psi_0 \rangle$ ($\gamma_{\alpha}^k = \langle \Psi_0^\dagger | c^\dagger_{\alpha} c_{\eta} | \Psi_0 \rangle$) are the spectroscopic amplitudes for the excited states of a system with $A+1$ ($A-1$) particles and the poles $\epsilon_{\alpha}^+ = E_{\alpha}^{A+1} - E_0^0$ ($\epsilon_{\alpha}^- = E_0^0 - E_{\alpha}^{A-1}$) correspond to the excitation energies with respect to the $A$-body ground state. The one-body Green’s function can be computed by solving the Dyson equation [19,20],

$$g_{\alpha\beta}(\omega) = g_{\alpha\beta}^0(\omega) + \sum_{\gamma\delta} g_{\alpha\gamma}(\omega) \Sigma_{\gamma\delta}^*(\omega) g_{\delta\beta}(\omega),$$

where the irreducible self-energy $\Sigma_{\gamma\delta}^*(\omega)$ acts as an effective, energy-dependent, potential that governs the single particle behavior of the system. The self-energy is expanded in a Faddeev series as in Fig. 1. This couples the exact propagator $g_{\alpha\beta}(\omega)$ (which is itself a solution of Eq. (2)) to other phonons in the system [17]. The relevant information regarding pp and ph/hh collective excitations is included in the polarization and the two-particle propagators. Respectively,

$$\Pi_{\alpha\beta,\gamma\delta}(\omega) = \sum_{n \neq 0} \frac{\langle \Psi_0^A | c_{\gamma}^\dagger c_{\delta} | \Psi_n^A \rangle \langle \Psi_n^A | c_{\alpha}^\dagger c_{\beta} | \Psi_0^A \rangle}{\omega - (E_n^A - E_0^A) + i\eta}$$

$$- \sum_{n \neq 0} \frac{\langle \Psi_0^A | c_{\gamma}^\dagger c_{\delta} | \Psi_n^A \rangle \langle \Psi_n^A | c_{\alpha}^\dagger c_{\beta} | \Psi_0^A \rangle}{\omega - (E_n^A - E_0^A) - i\eta},$$

and

$$g_{\alpha\beta}^H(\omega) = \sum_{n} \frac{\langle \Psi_0^A | c_{\alpha}^\dagger c_{\beta} | \Psi_n^{A+2} \rangle \langle \Psi_n^{A+2} | c_{\gamma}^\dagger c_{\delta} | \Psi_0^A \rangle}{\omega - (E_n^{A+2} - E_0^A) + i\eta}$$

$$- \sum_{k} \frac{\langle \Psi_0^A | c_{\alpha}^\dagger c_{\beta} | \Psi_k^{A-2} \rangle \langle \Psi_k^{A-2} | c_{\gamma}^\dagger c_{\delta} | \Psi_0^A \rangle}{\omega - (E_k^{A-2} - E_0^A) - i\eta},$$

which describe the one-body response and the propagation of two-particles/two-holes. In this work, $\Pi(\omega)$ and $g_H^H(\omega)$ are obtained by solving the dressed RPA (DRPA) equations [21,22], which account for the redistribution of strength in the sp spectral function. Since this information is carried by the correlated propagator $g_{\alpha\beta}(\omega)$, Eq. (2), the SCGF formalism requires an iterative solution. It can be proven that full self-consistency guarantees to satisfy the conservation of the number of particles and other basic quantities [23].

The coupled cluster studies of Refs. [6,24] found that eight major harmonic oscillator shells can be sufficient to obtain converging results for $^{16}$O with G-matrix interactions. At the same time, the experience with the calculations of Ref. [18] suggests that high partial waves do not contribute sensibly. In this work, all the orbits of the first eight shells with orbital angular momentum $\ell \leq 4$ were included. Inside this model space a G-matrix and the $V_{\text{UCOM}}$ potential were employed as effective interactions. The former was computed using the CENS library routines [11,25]. For the latter, the UCOM matrix-elements code [26] was employed with the constraint $I_0 = 0.09$ fm$^3$. This choice of the UCOM correlator reproduces, in perturbation theory, the binding energies of several nuclei up to $^{208}$Pb [27]. In both cases the Argonne $v_{18}$ potential [28] was used as starting interaction. However, we chose to neglect the Coulomb and the other charge independence breaking terms in the present work. The Hartree–Fock (HF) equations (Brueckner–Hartree–Fock (BHF) for the G-matrix) were first solved for the unperturbed propagator $g_{\alpha\beta}^{\text{BHF}}(\omega)$, which was employed in the first calculation. After that, the (dressed) solution $g_{\alpha\beta}(\omega)$ was used to generate $\Pi(\omega)$ and $g_H^H(\omega)$ in DRPA and then to solve
the Faddeev equations for an improved self-energy. At each iteration the two most important fragments close to the Fermi level of each partial wave were retained, both in the quasiparticle and the quasihole domains. The remaining strength was collected in few effective poles that correspond to the (B)HF particle and the quasihole domains. The remaining strength was also considered in Ref. [4]. There, the effective interaction was derived in the unitary-model-operator approach and an explicit diagonalization was performed. The resulting sp ener-

Fig. 2. Single particle spectrum obtained with the G-matrix as a function of the oscillator length. The dashed lines refer to values of $b_{HO}$ for which the solutions are sensible to the number of sp fragments collected, at each iteration, into effective poles of $g_{ap}(\omega)$.

Fig. 3. Self-consistent single particle spectrum obtained with $V_{UCOM}$ as a function of the oscillator length.

Table 1. Spin–orbit splittings (in MeV). The experimental values refer to the spectra of $^{17}$O/$^{15}$O [29]

<table>
<thead>
<tr>
<th>$b_{HO}$ [fm]</th>
<th>1.6</th>
<th>1.7</th>
<th>1.8</th>
<th>1.9</th>
<th>2.0</th>
<th>2.1</th>
<th>exp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta E_{p_{1/2}-p_{3/2}}$</td>
<td>–</td>
<td>–</td>
<td>3.1</td>
<td>3.1</td>
<td>3.2</td>
<td>6.176</td>
<td></td>
</tr>
<tr>
<td>$\Delta E_{d_{3/2}-d_{5/2}}$</td>
<td>–</td>
<td>–</td>
<td>3.5</td>
<td>3.6</td>
<td>3.5</td>
<td>3.4</td>
<td>5.084</td>
</tr>
<tr>
<td>$V_{UCOM}$</td>
<td>4.7</td>
<td>4.4</td>
<td>4.1</td>
<td>4.5</td>
<td>4.4</td>
<td>4.1</td>
<td>6.176</td>
</tr>
<tr>
<td>$\Delta E_{p_{1/2}-p_{3/2}}$</td>
<td>4.9</td>
<td>4.4</td>
<td>4.1</td>
<td>3.9</td>
<td>3.6</td>
<td>3.3</td>
<td>5.084</td>
</tr>
<tr>
<td>$\Delta E_{d_{3/2}-d_{5/2}}$</td>
<td>4.9</td>
<td>4.4</td>
<td>4.1</td>
<td>3.9</td>
<td>3.6</td>
<td>3.3</td>
<td>5.084</td>
</tr>
</tbody>
</table>

are not converged with respect to the number of iterated poles are shown, for completeness, by dashed lines in Fig. 2. However, they will not be considered any further in the following. No similar complications were encountered for $V_{UCOM}$ which generates, for each shell, at most one main fragment and a smaller satellite peak near the Fermi energy. The relevant poles of Eq. (1) were therefore iterated exactly. Fig. 3 shows that the spin orbit splittings for this interaction are approximately constant, although the sp energies are not yet independent of the oscillator length. This can be understood considering that these spin orbit partners correspond to particularly simple and similar configurations (one particle or one hole on top of the correlated ground state). Conversely, separation energies are linked to the total binding energy of neighbor isotopes. Larger model spaces will probably be required for a full convergence with $V_{UCOM}$.

The splittings obtained from both interactions are reported in Table 1. The $0\rho$ results obtained with the G-matrix show little dependence of the oscillator length. These are in line with previous Green’s function calculations [30] and account for about half of the experimental value. Better solutions are obtained with the present choice of the UCOM correlator. For the $0\delta$ orbits the results for the two interactions are more similar to each other but not totally independent of the oscillator length.

Long-range correlations at the level of 2p1h and 2h1p were also considered in Ref. [4].
energies showed a somewhat stronger dependence on the oscillator length than the one found in this work. The present formalism explicitly employs a basis of 2qp1qh/2qh1qp configurations. By using fully dressed quasiparticle and quasihole states additional ph excitation are included, in principle up to promoting all the nucleons above the Fermi level. The effects of self-consistency are thus twofold: additional excitations are implicitly included beyond the bare 2p1h/2h1p level and (as discussed above) these contributions are selected in such a way to preserve basic conservation laws [7,23]. The importance of these can be judged by comparing the second and third columns of Figs. 4 and 5. The spectrum obtained in Fig. 2 is nearly convergent, suggesting that the all order summation employed here and the proper accounting of the fragmentation of the sp strength allow to select most of the relevant configurations. Coupled cluster calculations are also available for $^{17}$O with an analogous $v_{18}$/G-matrix and also including 2p1h/2h1p cluster operators [6]. These authors find convergence with respect to the model space and report splittings of the 0p and 0d orbits larger than those of Table 1 by about 1.5 and 0.5 MeV, respectively. We note that the LRC studied in this work are in the form of couplings to small amplitude excitations of the core—which can be described at the DRPA level. More complex collective modes are also present [31] and should be included for full solution of the many-body problem. For example, the phenomenological studies of Ref. [18] suggest further contributions to the $p_{3/2}$ quasihole wave function coming from couplings to the first excited $0^+$ state in oxygen. Testing this conjecture would first require being able to reproduce the correct excitation energy of this level—since it can couple effectively only when it is low enough in energy. To our knowledge this is still a challenge for the available ab initio methods.

Figs. 4 and 5 show the effects of LRC on the sp spectrum for $^{17}$O with $h_{9/2} = 1.9$ fm, and compare to the experimental values for the addition/removal of a neutron. For both interactions the coupling to collective phonons reduces the splitting of the 0p orbits, with respect to the HF approximation. Including the effects of fragmentation tends instead to compress the sd shell and to lower the whole spectrum. The self-consistent results for the energy gap between particle and hole states, $\Delta E_F = \epsilon_{d_{5/2}} - \epsilon_{p_{1/2}}$, are 13.0 MeV with the G-matrix and 15.4 MeV with $V_{UCOM}$. Both of them exceed the experimental value of 11.5 MeV. However, the differences $\epsilon_{d_{5/2}} - \epsilon_{p_{1/2}} = 12.2$ MeV and $\epsilon_{d_{5/2}} - \epsilon_{p_{1/2}} = 16.5$ MeV obtained with the G-matrix are close to the experiment (12.4 and 16.6 MeV, respectively), suggesting that the shortcomings of this interaction lie mainly in the poor description of the spin–orbit splittings. The mean square radii obtained are $r_{rms} = 2.63$ fm (G-matrix) and $r_{rms} = 2.45$ fm ($V_{UCOM}$).

Using the renormalization group with different momentum cutoffs it was shown that it is possible to shift the binding energies for $A = 3, 4$ systems along the Tjon line [32]. The same result has been obtained in the UCOM approach by modifying the correlator in the tensor-isoscalar channel [33]. Usually, tuning the binding energies to the experimental values increases the nonlocality of the interaction and leads to improved spin orbit splittings, as seen in Table 1. On the other hand, our $V_{UCOM}$ result for $\Delta E_F$—with 2N forces—overestimates the experiment. This behavior is seen already at the HF level for soft interactions like $V_{UCOM}$ and $V_{low-k}$ [27,34] and it is only slightly modified by the LRC considered here. We note that $V_{UCOM}$ is obtained by applying the UCOM correlator operator to the nuclear Hamiltonian and then truncating to a two-nucleon interaction. Hence, it is not expected to generate the same results of the original interaction (Argonne $v_{18}$ in this work). In both cases (G-matrix and $V_{UCOM}$), three-body forces appear necessary in order to reproduce the whole spectrum of observations. We note, however, that the UCOM method offers some advantages to reduce the contributions needed from many-body forces since it allows to treat SRC in different channels separately [15].

In conclusion, SCGF calculations have been performed for the first time in a large model space, including up to eight oscillator shells. Long-range correlations in the form of coupling sp to ph and pp/hh DRPA modes were investigated for $^{16}$O. A comparison was made between the results of a G-matrix and the $V_{UCOM}$ interactions, both derived from same realistic potential (Argonne $v_{18}$). The spectra of adjacent nuclei were found to be nearly convergent for the G-matrix, while they depend only weakly on the oscillator length for $V_{UCOM}$. In general it was found that the LRC effects considered here, tend to compress the spectra of $A \pm 1$ nuclei but do not affect sensibly the gap between quasiparticle and quasihole energies at the Fermi level.
Acknowledgements

The author is in debt with B.K. Jennings for suggesting this work, as well as with H. Feldmeier and R. Roth for several useful discussions. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC). Part of the calculations were performed at the TRIUMF/UBC node of WestGrid Canada.

References

[34] N. Paar, P. Papakonstantinou, H. Hergert, R. Roth, nucl-th/0506076.
Saturation physics at HERA and RHIC: 
An unified description

V.P. Gonçalves a, *, M.S. Kugeratski b, M.V.T. Machado c, F.S. Navarra b

a High and Medium Energy Group (GAME), Instituto de Física e Matemática, Universidade Federal de Pelotas Caixa Postal 354, CEP 96010-900, Pelotas, RS, Brazil
b Instituto de Física, Universidade de São Paulo, C.P. 66318, 05315-970 São Paulo, SP, Brazil
c High Energy Physics Phenomenology Group, GFPAE, IF-UFRGS Caixa Postal 15051, CEP 91501-970 Porto Alegre, RS, Brazil

Received 4 August 2006; received in revised form 17 October 2006; accepted 26 October 2006

Abstract

One of the frontiers of QCD which are intensely investigated in high energy experiments is the high energy (small $x$) regime, where we expect to observe the non-linear behavior of the theory. In this regime, the growth of the parton distribution should saturate, forming a color glass condensate (CGC). In fact, signals of parton saturation have already been observed both in $ep$ deep inelastic scattering at HERA and in deuteron-gold collisions at RHIC. Currently, a global description of the existing experimental data is possible considering different phenomenological saturation models for the two processes within the CGC formalism. In this Letter we analyze the universality of these dipole cross section parameterizations and verify that they are not able to describe the HERA and RHIC data simultaneously. We analyze possible improvements in the parameterizations and propose a new parameterization for the forward dipole amplitude which allows us to describe quite well the small-$x$ $ep$ HERA data on $F_2$ structure function as well as the $dAu$ RHIC data on charged hadron spectra. It is an important signature of the universality of the saturation physics.

© 2006 Elsevier B.V. All rights reserved.

In the past few years much theoretical effort has been devoted towards the understanding of the high energy limit of the strong interaction theory (for recent reviews see, e.g. [1–3]). In the high energy limit, perturbative quantum chromodynamics (pQCD) predicts that the small-$x$ gluons in a hadron wavefunction should form a color glass condensate (CGC), which is described by an infinite hierarchy of the coupled evolution equations for the correlators of Wilson lines [4–7]. In the absence of correlations, the first equation in the Balitsky–JIMWLK hierarchy decouples and is then equivalent to the equation derived independently by Kovchegov within the dipole formalism [8]. The color class condensate is characterized by the limitation on the maximum phase-space parton density that can be reached in the hadron wave function (parton saturation), with the transition being specified by a typical scale, which is energy dependent and is called saturation scale $Q_s$ [$Q_s^2 \propto A^{d} x^{-\lambda}$]. Moreover, in the CGC formalism the dipole-target forward scattering amplitude $N$ for a given impact parameter $b$, which is directly related with the two-point function of Wilson lines, encodes all the information about the hadronic scattering, and thus about the non-linear and quantum effects in the hadron wave function. The function $N$ can be obtained by solving the Balitsky–JIMWLK evolution equation in the rapidity $Y \equiv \ln(1/x)$ [9]. Its main properties are: (a) for the interaction of a small dipole ($r \ll 1/Q_s$), $N(r) \approx r^2$, implying that this system is weakly interacting; (b) for a large dipole ($r \gg 1/Q_s$), the system is strongly absorbed and therefore $N(r) \approx 1$. This property is associated to the large density of saturated gluons in the hadron wave function. Another remarkable feature of CGC formalism is that the dense, saturated system of partons to be formed in hadronic wave functions at high energy has universal properties, the same for all hadrons or nuclei.

In the CGC formalism the description of the observables is directly related to the behavior of $N$. For instance, the $F_2$...
structure function is probed in $ep(A)$ process and is given by
\[
F_{F_x}(x, Q^2) = (Q^2/4\pi^2a_{em})(\sigma_{T}^{\gamma^{*}p(A)} + \sigma_{L}^{\gamma^{*}p(A)}), \tag{10}
\]
where $a_{em}$ is the electromagnetic correction factor. $\sigma_{T}$ and $\sigma_{L}$ represent the total and longitudinal cross sections, respectively.

In Eq. (2), the parton distribution functions are integrated over the rapidity $y$ and transverse momentum $p_T$ of the produced hadron. These quantities evolve according to the BFKL equation.

In the context of the CGC formalism, the forward dipole amplitude $F_{F_x}(x, Q^2)$ can be expressed as
\[
F_{F_x}(x, Q^2) = \sum_{f} \int d^2\mathbf{r} \left| \Psi_{L,T}^{f}(x, r, Q^2) \right|^2 \sigma_{\text{dip}}(x, r)
\]
\[
= \sum_{f} \int d^2\mathbf{r} \left| \Psi_{L,T}^{f}(x, r, Q^2) \right|^2 \times 2 \int d^2\mathbf{b} N_{F}(x, r, b),
\]
with $N_{F}$ being the fundamental representation of the forward dipole amplitude, $r$ defining the relative transverse separation of the pair (dipole) and $z=(1-z)$ the longitudinal momentum fraction of the quark (antiquark). The photon wave functions $\Psi_{L,T}$ are determined from light cone perturbation theory (see e.g. Ref. [11]). It is useful to assume that the impact parameter dependence of $N_{F}$ can be factorized as $N_{F}(x, r, b) = N_{F}(x, r)S(b)$, so that $\sigma_{\text{dip}}(x, r) = \sigma_{0}N_{F}(x, r)$, with $\sigma_{0}$ being a free parameter related to the non-perturbative QCD physics. Similarly, the single-inclusive hadron production in hadron--hadron processes is described in the CGC formalism by [12]
\[
\frac{d\sigma_{pp(A)} \rightarrow X F}{dx F d^2p_T d^2b} = \frac{1}{(2\pi)^2} \int_{x_F}^{1} dx_F \frac{f_{q/p}(x_F, Q^2_F)N_{F}(x_F, p_T, b)}{x_F} \times D_{h/q}(x_F, p_T, Q^2_F) + f_{g/p}(x_F, Q^2_F)N_{A}(x_F, p_T, b)
\]
\[
\times D_{h/g}(x_F, p_T, Q^2_F),
\]
where $p_T$ and $x_F$ are the transverse momentum and the Feynman-$x$ of the produced hadron, respectively. The variable $x_F$ denotes the momentum fraction of a projectile parton and $b$ is the impact parameter. Moreover, $f_{q/p}(x_F, Q^2_F)$ is the projectile parton distribution functions and $D(z, Q^2_F)$ the parton fragmentation functions into hadrons. These quantities evolve according to the DGLAP [13] evolution equations and respect the momentum sum-rule. In Eq. (2), $N_{F}(k, b)$ and $N_{A}(k, b)$ are the fundamental and adjoint representations of the forward dipole amplitude in momentum space. The amplitudes $N_{F}(k, b)$ and $N_{A,F}(r, b)$ are directly related by a Fourier transform.

The search of signatures for the parton saturation effects has been an active subject of research in the last years. In particular, it has been observed that the HERA data at small $x$ and low $Q^2$ can be successfully described with the help of saturation models [14–19]. Moreover, experimental results for the total [20], diffractive [21] and inclusive charm cross sections [22,23] present the property of geometric scaling. On the other hand, the observed [24] suppression of high $p_T$ hadron yields at forward rapidities in $dAu$ collisions at RHIC had its behavior anticipated on the basis of CGC ideas [25]. A current shortcoming of these analyzes comes from the non-existence of an exact solution of the non-linear equation in the full kinematic range, which implies the construction of phenomenological models satisfying the asymptotic behavior which is under theoretical control. Several models for the forward dipole cross section have been used in the literature in order to fit the HERA and RHIC data. In particular, the phenomenological models from Refs. [14–17] have been proposed in order to describe the HERA data, while those from Refs. [12,26] have been able to describe the $dAu$ RHIC data. An important aspect should be emphasized at this point. Although at HERA it is possible to probe values of $x$ two orders of magnitude smaller than at RHIC, the saturation scales for these two scenarios are very similar due to the nuclear medium (see Fig. 1 in Ref. [27]). Consequently, one can expect to be possible to cross relate these experiments in this respect and gain a clear understanding of the CGC in high energy experiments. There are several similarities among the phenomenological models proposed in Refs. [12,14–17,26]. In particular, in these models the function $N_{F}$ has been modeled in terms of a simple Glauber-like formula
\[
N_{F}(x, r) = 1 - \exp \left[ -\frac{1}{4}(r^2 Q_{F}^2(x))^\gamma(x, r^2) \right],
\]
where $\gamma$ is the anomalous dimension of the target gluon distribution. The main difference comes from the predicted behavior for the anomalous dimension, which determines the transition from the non-linear to the extended geometric scaling regimes, as well as from the extended geometric scaling to the DGLAP regime. A detailed comparison has been presented in Ref. [27]. As the models from Refs. [14,15,17] have been exhaustively discussed in the literature, in this Letter we only present a brief review of the models proposed in Refs. [12,26]. In the KKT model [26] the expression for the quark dipole-target forward-scattering amplitude is given by [26]:
\[
N_{F}(r, x) = 1 - \exp \left[ -\frac{1}{4}(r^2 Q_{F}^2)^\gamma(Y, r^2) \right],
\]
where $Q_{F}^2 = \frac{C_F}{N_c} Q_T^2$ and the anomalous dimension $\gamma(Y, r^2)$ is
\[
\gamma(Y, r^2) = \frac{1}{2} \left( 1 + \frac{\xi(Y, r^2)}{\xi(Y, r^2) + \sqrt{2\xi(Y, r^2) + 7\xi(3)c} + \gamma(3)c} \right),
\]
with $c$ a free parameter (which was fixed in [26] to $c=4$) and $\xi(Y, r^2) = \ln[1/(r^2 Q_{F}^2)]/(\Lambda^2/(2\pi)(Y-Y_0).$}

The authors assume that the saturation scale can be expressed by $Q_{F}^2(Y) = \Lambda^2 A^{1/3}(1/Y)^\gamma$. The form of the anomalous dimension is inspired by the analytical solutions to the BFKL equation [28]. Namely, in the limit $r \rightarrow 0$ with $Y$ fixed we recover the anomalous dimension in the double logarithmic approximation $\gamma \approx 1 - \sqrt{1/(2\xi)}$. In another limit of large $Y$ with $r$ fixed, Eq. (5) reduces to the expression of the anomalous dimension near the saddle point in the leading logarithmic approximation $\gamma \approx \frac{1}{2} + \frac{k_T^2}{4\xi(3)c}$. Therefore, Eq. (5) mimics the onset of the geometric scaling region [17,29]. In the calculations of Ref. [26] it is assumed that a characteristic value of $r$ is $r \approx 1/(2k_T)$ where $k_T$ is the transverse momentum of the valence quark and $\gamma$ was approximated by $\gamma(Y, r^2) \approx \gamma(Y, 1/(4k_T^2))$. In the above expressions the parameter $\Lambda = 0.6$ GeV and
\[ \gamma(Y, r^2) = \gamma_s + \Delta \gamma(Y, r^2) \]  \hspace{1cm} (7)

where

\[ \Delta \gamma(Y, r^2) = (1 - \gamma_s) \frac{|\log \frac{1}{r^2 Q_T^2}|}{\lambda Y} + |\log \frac{1}{r^2 Q_T^2}| + d\sqrt{Y}, \]  \hspace{1cm} (8)

with \( Q_T = Q_s(Y) \) a typical hard scale in the process, \( \lambda = 0.3 \) and \( d = 1.2 \). Moreover, \( \gamma_s = 0.63 \) is the anomalous dimension for BFKL evolution with saturation boundary condition. Similarly to the KKT model this model is able to describe the RHIC data.

As already discussed in Ref. [27], based on the universality of the hadronic wave function predicted by the CGC formalism, we might expect that the KKT and DHJ parameterizations would also describe the HERA data on proton structure functions in the kinematical region where the saturation effects should be present (small \( x \) and low \( Q^2 \)). However, this expectation fails when the KKT model is applied, as verified in [27]. Here we extend the analysis of the DHJ model without any modification of the parameters fitted at RHIC, only assuming \( \lambda = 1 \) and adjusting the non-perturbative parameter \( \alpha_0 \), which defines the normalization, in order to describe the \( F_2 \) experimental data at \( Q^2 = 10 \text{ GeV}^2 \). In Figs. 1 and 2 we present the predictions of the DHJ model for the proton structure function and compare with the ZEUS data [30]. We can see that this parameterization fails for both small and large values of \( Q^2 \). Consequently, the current parameterizations of the forward dipole cross section which are constrained at RHIC are not able to describe the HERA data. An open question is if minimal modifications in these parameterizations allow to describe both sets of data. Following Ref. [31] we consider a modification of the KKT model assuming that the saturation momentum scale is given as in the GBW model, \( Y_0 = 4.6, c = 0.2 \) and that the typical scale in the computation of \( \xi(Y, r^2) \) is the photon virtuality. Its predictions (KKTm lines) are presented in Figs. 1 and 2. It is observed that these modifications imply a quite good description of the HERA data. Similarly, as the \( Q^2 \) evolution of the \( F_2 \) data is not well described by the DHJ model it is possible to improve this model by the modification of the anomalous dimension. Here we propose to modify the DHJ model assuming now that \( Q_T = Q_0 = 1.0 \text{ GeV} \), i.e. that the typical scale is energy independent. It is important to emphasize that this modification preserves the main properties of the anomalous dimension proposed in [12]. Basically, we still have that the anomalous dimension increases logarithmically with \( p_T \) from \( \gamma = \gamma_s \) to its asymptotic value \( \gamma \approx 1 \), while decreasing with \( Y \) as \( \Delta \gamma \approx 1/Y \) at very large rapidity. As shown in Figs. 1 and 2, with this modification our predictions (GKMN lines) agree with the experimental data.

The question which follows is whether the RHIC data are still well reproduced after these modifications. Following Ref. [12] we have calculated the single inclusive hadron production cross section in \( dAu \) collisions at different rapidities. We have used the CTEQ5L quark and gluon distributions [32] and the LO KKP quark–hadron fragmentation functions [33]. Our results are presented in Fig. 3 and compared with the BRAHMS data [24]. The KKTm and GKMN predictions are represented by long-dashed and solid curves respectively. As in Ref. [12] we need a \( K \)-factor in our calculations, since it has been performed at leading order in \( \alpha_s \). Although the normal-
The proton structure function at different values of the photon virtualities. Data from ZEUS.

Comparison of theory and BRAHMS data for minimum-bias $dAu$ collisions at RHIC energy.

Before presenting a summary of our main results, let us briefly discuss the basic properties of the resulting GKMN model (a more detailed analysis will be presented elsewhere). In Fig. 4(a) we present the forward dipole cross section as a function of the scaling variable $r Q_s$ for distinct parameterizations. As it can be seen the DHJ, KKTm and GKMN models have a similar behavior. The difference among the models can be demonstrated studying the $Q^2$ behavior of the effective anomalous dimension, defined by $\gamma_{\text{eff}} = \frac{d \ln N(r Q_s, Y)}{d \ln (r^2 Q^2_s / 4)}$ (see similar analyzes in Ref. [31]). In Fig. 4(b) is shown $\gamma_{\text{eff}}$ as a function of the virtuality $Q^2$, using the average dipole size as $r = 2/Q$. While the GBW model presents a fast convergence to the DGLAP anomalous dimension at large $Q^2$, the IIM parameterization has a mild growth with virtuality, converging to $\gamma \approx 0.85$ at large $Q^2$. The KKTm and IIM parameterizations are similar at large $Q^2$, but differ at small virtualities, with the KKTm one predicting a smaller value. On the other hand, the predictions of the DHJ and GKMN parameterizations are similar at small $Q^2$ and differ at large virtualities. In particular, we have a strict difference between these models in the intermedi-
Fig. 4. (a) Forward dipole cross section as a function of the scaling variable $rQ_s$. (b) The $Q^2$ behavior of the effective anomalous dimension at $x = 3 \times 10^{-4}$.

A wide range of virtualities, which can explain why the DHJ model does not describe the $Q^2$ evolution of the $F_2$ structure function.

As a summary, in this letter we have analyzed current parameterizations for the dipole scattering amplitude which are able to describe separately the $ep$ HERA and $dAu$ RHIC data. We have shown that an unified description using these parameterizations is not possible. We have proposed a modification in the DHJ parameterization for the dipole scattering amplitude, based on saturation physics, which allows to describe simultaneously the $ep$ HERA and $dAu$ RHIC data. This result has been obtained adjusting the normalizations of the dipole cross section and single inclusive hadron cross section and assuming an energy independent typical scale, keeping all other original parameters. A global least $\chi^2$ fit of data would change slightly the values of our parameters. This would be a fine tuning which is beyond the scope of this work. We rather prefer to keep the level of fitting accuracy of [12] and emphasize the strategy to reconcile two different sets of data. Apart from this fine tuning, a more detailed theoretical study of the proposed anomalous dimension is necessary. We postpone these improvements for a future publication. Finally, our results demonstrate that an unified description of the experimental data which probes the high energy regime of QCD is possible. This is an important signature of the universality of the saturation physics.

Acknowledgements

V.P.G. would like to thanks M.A. Betemps for informative and helpful discussions. This work was partially financed by the Brazilian funding agencies CNPq, FAPESP and FAPERGS.

References

Non-thermal leptogenesis and baryon asymmetry in different neutrino mass models

G. Panotopoulos

Department of Physics and Institute of Plasma Physics, University of Crete, 71003 Heraklion, Crete, Greece
ASC, Physics Department LMU, Theresienstr. 37, 80333 Munich, Germany

Received 21 September 2006; accepted 28 October 2006
Available online 7 November 2006
Editor: T. Yanagida

Abstract

In the present work we study non-thermal leptogenesis and baryon asymmetry in the universe in different neutrino mass models discussed recently. For each model we obtain a formula relating the reheating temperature after inflation to the inflaton mass. It is shown that all but four cases are excluded and that in the cases which survive the inflaton mass and the reheating temperature after inflation are bounded from below and from above.

© 2006 Elsevier B.V. All rights reserved.

1. Introduction

The Standard Model (SM) of particle physics (for a review on the subject see e.g. [1]) is a very successful theoretical framework for all low-energy phenomena. However, it is widely considered to be a low-energy limit of some underliner fundamental theory. Perhaps the most direct evidence for physics beyond the SM is the recent discovery that neutrinos have small but finite masses [2–4]. A simple and natural way to explain the tiny neutrino masses is via the seesaw mechanism [5]. According to that, the existence of super-heavy right-handed neutrinos is postulated and the smallness of the masses of the usual SM neutrinos is due to the largeness of the masses of the new neutrinos. Solar, atmospheric, reactor and accelerator neutrino experiments (for a summary of three-flavour neutrino oscillation parameters see e.g. [6]) seem to indicate neutrino masses in the sub-eV range \((0.001 < m_\nu < 0.1 \text{ eV})\), which implies that heavy right-handed neutrinos weigh \(\sim 10^{10} - 10^{15} \text{ GeV} \) [7].

On the other hand, the baryon asymmetry in the universe (BAU) is one of the most challenging problems for modern cosmology. Both Big-Bang nucleosynthesis [8] and CMB data (for example from WMAP [9]) show that in the universe one baryon corresponds approximately to one billion photons. This very small number should be computable in the framework of the theory of the elementary particles and their interactions we know today. Nowadays, the most popular way to obtain the BAU is through leptogenesis (for an incomplete list see e.g. [10] and for a review see [11]). Initially a lepton asymmetry is generated through the out-of-equilibrium decays of right-handed neutrinos and then the lepton asymmetry is partially converted to baryon asymmetry through the non-perturbative “sphaleron” effects [12]. In general leptogenesis can be thermal or non-thermal. Thermal leptogenesis usually requires very high reheating temperature after inflation [13]. This can be problematic because of the gravitino constraint. In supersymmetric models (for reviews in supersymmetry see e.g. [14] and for supersymmetry in cosmology see e.g. [15]) with spontaneous supersymmetry breaking the superpartner of the graviton, the gravitino, gets a mass depending on how the supersymmetry is broken. In gravity mediated supersymmetry breaking the gravitino mass is in the range \(m_{3/2} = 100 \text{ GeV–1 TeV} \) and the gravitino (if not the lightest supersymmetric particle) is unstable with a lifetime larger than Nucleosynthesis time \(t_N \sim 1 \text{ s} \) and dangerous for cosmology. This gravitino problem [16] can be avoided provided that the reheating temperature after inflation is bounded from above in a certain way, namely \(T_R \lesssim (10^6 - 10^7) \text{ GeV} \) [17].
Therefore one can see that heavy right-handed neutrinos can have important implications both for particle physics and cosmology. Various neutrino mass models [18,19] have been proposed and their predictions on neutrino masses and mixings have been studied thoroughly. The requirement for the right baryon asymmetry in the universe as well as for the right phenomenology for light neutrino masses and mixings puts severe constraints on right-handed neutrinos. Recently six concrete neutrino mass models were discussed and a comparison of numerical predictions on baryon asymmetry for these models was presented [20]. Two of the models were almost consistent with the observed BAU, while the rest of them predicted either a small ($\eta \lesssim 10^{-19}$) or a large ($\eta \geq 10^{-6}$) baryon asymmetry. The analysis was performed in the framework of thermal leptogenesis. The aim of the present work is to study the same models in the framework of non-thermal leptogenesis and derive the constraints on the inflaton mass and the reheating temperature after inflation.

Our Letter is organized as follows. After this introduction we review the six neutrino mass models and lepton asymmetry in Section 2 and we discuss non-thermal leptogenesis for these models in Section 3. Our results are presented in Section 4 and we conclude in Section 5.

2. Review of the different neutrino mass models and of lepton asymmetry

Here we give a brief review of the six neutrino mass models [19] discussed recently in [20]. The interested reader can find more details in [19,20]. In particular, all the information about the models are collected in Appendix A of [20]. There is one normal hierarchical model (NHT3), two inverted hierarchical models (InvT2A, InvT2B) and three degenerate models (DegT1A, DegT1B, DegT1C). According to seesaw mechanism, the light left-handed neutrino mass matrix $m_\nu$, the heavy right-handed neutrino mass matrix $M_R$ and the Dirac neutrino mass matrix $m_D$ are related as follows

$$m_\nu = m_D M_R^{-1} m_D^T.$$  

(1)

where $M_R^{-1}$ is the inverse of $M_R$ and $m_D^T$ is the transpose of $m_D$. The predicted values of the neutrino mass-squared differences and mixing parameters are shown in Table 1.

In thermal leptogenesis for the SM case the BAU $\eta \equiv n_B/n_\gamma = 6.1 \times 10^{-10}$ is computed by the formula [20]

$$\eta = 0.0216 \kappa \epsilon,$$

(2)

where $\kappa$ is the dilution factor and $\epsilon$ is the $CP$ asymmetry. The dilution factor is not needed for our discussion in non-thermal leptogenesis scenario. However we remark in passing that it is determined by numerical integration of Boltzmann equations and that it can be estimated by analytical expressions given in [21].

On the other hand, the $CP$ asymmetry is a basic quantity for our presentation and we shall now discuss its relation to the neutrino mass matrices. The lepton asymmetry in the universe is generated by $CP$ violating out-of-equilibrium decay of the heavy neutrinos $N \to lH^+$ and $N \to lH$. The $CP$ asymmetry $\epsilon$ is defined as

$$\epsilon = \frac{\Gamma - \bar{\Gamma}}{\Gamma + \bar{\Gamma}}$$  

(3)

with $\Gamma = \Gamma(N \to lH^*)$ and $\bar{\Gamma} = \Gamma(N \to \bar{l}H)$ the decay rates. It is given by the interference between tree-level and one-loop decay amplitudes and it is found to be [22]

$$\epsilon = \frac{1}{8\pi(Y_\nu Y_{\nu}^*)_{11}} \sum_{j=2,3} \text{Im}[(Y_{\nu} Y_{\nu}^*)_{1j}^2] \times (f(M_{Rj}^2/M_{R1}^2) + 2g(M_{Rj}^2/M_{R1}^2)), $$  

(4)

where the two functions $f(x)$, $g(x)$ have the form

$$f(x) = \sqrt{x} [1 - (1 + x) \ln(1 + 1/x)],$$  

(5)

$$g(x) = \frac{\sqrt{x}}{2(1-x)}.$$  

(6)

Both functions behave like $\sim -1/(2\sqrt{x})$ for $x \gg 1$. In this approximation the asymmetry $\epsilon$ takes the form

$$\epsilon = -\frac{3}{8\pi(Y_\nu Y_{\nu}^*)_{11}} \sum_{j=2,3} \text{Im}[(Y_{\nu} Y_{\nu}^*)_{1j}^2] \frac{M_{Rj}}{M_{R1}}.$$  

(7)

Finally, using the seesaw formula $\epsilon$ becomes [23]

$$\epsilon = \frac{3M_{R1} m_\nu \delta_{\text{eff}}}{16\pi \bar{v}^2}. $$  

(8)

where $\bar{v}$ is the Higgs vev and $\delta_{\text{eff}}$ is the $CP$-violating phase. However, in the quasi-degenerate spectrum $M_{R1} \simeq M_{R2} < M_{R3}$ the $CP$ asymmetry is enhanced by a factor given by [24]

$$R = \frac{M_{R1}}{2(M_{R2} - M_{R1})}.$$  

(9)

The three right-handed neutrino masses for each model are shown in Table 2 while the $CP$ asymmetry and baryon asymmetry are shown in Table 3. The Dirac neutrino mass matrix

<table>
<thead>
<tr>
<th>Type</th>
<th>$\Delta m_{21}^2$ [10^{-5} eV^2]</th>
<th>$\Delta m_{31}^2$ [10^{-3} eV^2]</th>
<th>$\tan^2 \theta_{12}$</th>
<th>$\sin^2 2\theta_{23}$</th>
<th>$\sin \theta_{13}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DegT1A</td>
<td>8.80</td>
<td>2.78</td>
<td>0.98</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>DegT1B</td>
<td>7.91</td>
<td>2.50</td>
<td>0.27</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>DegT1C</td>
<td>7.91</td>
<td>2.50</td>
<td>0.24</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>InvT2A</td>
<td>9.36</td>
<td>2.50</td>
<td>0.98</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>InvT2B</td>
<td>9.30</td>
<td>2.50</td>
<td>0.98</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>NHT3</td>
<td>9.04</td>
<td>3.01</td>
<td>0.55</td>
<td>0.98</td>
<td>0.074</td>
</tr>
</tbody>
</table>

Table 1: Predicted values of the solar and atmospheric neutrino mass-squared differences and three mixing parameters (from [20])
$m_D$ can be either the charged lepton mass matrix $m_l$ (case (i)) or the up-quark mass matrix $m_u$ (case (ii)). We see that NHT3 and DegT1A models are almost consistent with the observed BAU, while the rest of the models lead either to very small baryon asymmetry, $\eta \leq 10^{-19}$ (DegT1B, DegT1C, InvT2A), or to large baryon asymmetry, $\eta \geq 10^{-6}$ (InvT2B).

### 3. Non-thermal leptogenesis

In the non-thermal leptogenesis scenario [25] the heavy neutrinos are produced through the direct non-thermal decay of the inflaton. We start by introducing three heavy right-handed neutrinos (one for each family) $N_i$, $i = 1, 2, 3$, with masses $M_{R_1}, M_{R_2}, M_{R_3}$. They interact with the inflaton through Yukawa couplings with $\lambda_i$ the coupling constants for this type of interaction. We assume that after the slow-roll phase of inflation the inflaton decays predominantly into the heavy neutrinos. With a Yukawa coupling between the inflaton and the heavy neutrinos, the inflaton decay rate $\Gamma_\phi$ is given by

$$\Gamma_\phi \equiv \Gamma (\phi \rightarrow N_i N_i) = \frac{1}{4\pi} |\lambda_i|^2 M_\phi,$$

where $M_\phi$ is the inflaton mass. The reheating temperature after inflation $T_R$ (defined by $H(T_R) = \Gamma_\phi$, with $H$ the Hubble parameter) is given by

$$T_R = \left( \frac{45}{4\pi^2 g_s} \right)^{1/4} \left( \Gamma_\phi M_{pl} \right)^{1/2},$$

where $M_{pl}$ is Planck mass and $g_s$ is the effective number of relativistic degrees of freedom at the reheating temperature. For the reheating temperatures that we shall consider all the particles are relativistic and for MSSM $g_s = 915/4 = 228.75$, while for SM $g_s = 427/4 = 106.75$.

Any lepton asymmetry $Y_L \equiv n_L/s$ produced before the electroweak phase transition is partially converted into a baryon asymmetry $Y_B \equiv n_B/s$ via sphaleron effects [12]. The result-
ing $Y_B$ is

$$Y_B = CY_L$$

with the fraction $C$ computed to be $C = -8/15$ in the MSSM and $C = -28/79$ in the SM [26]. The lepton asymmetry, in turn, is generated by the $CP$-violating out-of-equilibrium decays of the heavy neutrino

$$N_1 \rightarrow l H^+, \quad N_1 \rightarrow \bar{\nu} H.$$

In the framework of non-thermal leptogenesis the lepton asymmetry can be obtained by a simple formula [11]

$$Y_L = \frac{3}{2} BR (\phi \rightarrow N_1 N_1) T_R M_1 \epsilon,$$

where $BR$ is the branching ratio for the decay of the inflaton to the lightest heavy right-handed neutrino. Following the fourth paper in [25] we shall consider that $M_1 \geq 100 T_R$, because in that case the neutrino $N_1$ is always out of thermal equilibrium. The decay $\phi \rightarrow N_1 N_1$ is kinematically allowed provided that

$$M_1 \geq 2 M_1.$$  

We will assume that $BR \approx 1$, that is the inflaton decays practically only to the lightest of the right-handed neutrinos. This is possible even if the inflaton is heavy enough to decay to all right-handed neutrinos as long as $|\lambda_1|^2 \gg |\lambda_2|^2, |\lambda_3|^2$. Combining the above formulae we obtain

$$Y_B = CY_L = C \frac{T_R}{2 M_1} \epsilon,$$

or

$$T_R = \left( \frac{2Y_B}{3C \epsilon} \right) M_1.$$

From the WMAP data [9] we know that

$$\eta_B \equiv \frac{n_B}{n_\gamma} = 6.1 \times 10^{-10}.$$
If we recall that the entropy density for relativistic degrees of freedom is \( s = \frac{2}{3} n_{\gamma} T^3 \) and that the number density for photons is \( n_{\gamma} = \frac{2c^3}{\pi^2} T^3 \), one easily obtains for today that \( s = 7.04 n_{\gamma} \). Thus for \( Y_B \) we have

\[
Y_B = 8.7 \times 10^{-11}.
\]  

Finally we recall that \( M_I > 2 M_1 \) and \( M_1 \geq 100 T_R \).

### 4. Results

Now we can present our results. We shall begin with the SM case first and we shall use for the fraction \( C \) the SM value, namely \( C = -28/79 \). For each neutrino model (12 cases in total) the \( C \) \( P \) asymmetry \( \epsilon \) as well as the right-handed neutrino mass \( M_1 \) are known. Therefore we have (i) a formula relating the reheating temperature to the inflaton mass, (ii) a lower bound for the inflaton mass \( M_I > 2 M_1 \), and (iii) an upper bound for the reheating temperature \( T_R \leq 0.01 M_1 \). Furthermore, using the relationship between \( T_R \) and \( M_I \) we are able to convert the upper limit for \( T_R \) to a corresponding upper limit for \( M_I \) and also the lower limit for \( M_I \) to a corresponding lower limit for \( T_R \). So both \( T_R \) and \( M_I \) are bounded both from above and from below. Let \( T_R^{\text{min}} \) and \( T_R^{\text{max}} \) be the lower and higher value for the reheating temperature respectively. Then \( T_R^{\text{min}} < T_R < T_R^{\text{max}} \) and obviously it is required that \( T_R^{\text{max}} > T_R^{\text{min}} \), which is not satisfied for all cases. In fact most of the cases are excluded. The only cases for which the constraint is satisfied are:

- **DegT1A**, case (i), for which:
  \[
  8.56 \times 10^0 < M_I \leq 5.49 \times 10^{11} \text{ GeV},
  \]  
  \[
  6.67 \times 10^5 < T_R \leq 4.28 \times 10^7 \text{ GeV};
  \]
- **NHT3**, case (i), for which:
  \[
  1.3 \times 10^{11} < M_I \leq 2.35 \times 10^{12} \text{ GeV},
  \]  
  \[
  3.6 \times 10^7 < T_R \leq 6.51 \times 10^8 \text{ GeV};
  \]
- **InvT2B**, case (i), for which:
  \[
  1.13 \times 10^{11} < M_I \leq 5.09 \times 10^{16} \text{ GeV},
  \]  
  \[
  1.25 \times 10^3 < T_R \leq 5.65 \times 10^6 \text{ GeV};
  \]
- **InvT2B**, case (ii), for which:
  \[
  9.2 \times 10^8 < M_I \leq 4.55 \times 10^{12} \text{ GeV},
  \]  
  \[
  9.29 \times 10^2 < T_R \leq 4.6 \times 10^6 \text{ GeV}.
  \]

One can see from the results presented above that inflationary models in which \( M_I \sim 10^{13} \text{ GeV} \), like e.g. chaotic [27] or natural [28] inflation, are compatible only with one neutrino model (InvT2B, case (i)). Furthermore, for a concrete inflationary model with a given inflaton mass our results allow us to know what the reheating temperature must be and also what the inflaton decay rate \( \Gamma_\phi \) is and what the inflaton Yukawa coupling \( |\lambda_1| \) is. For example, in chaotic or natural inflation we obtain

\[
M_I \sim 10^{13} \text{ GeV},
\]

\[
T_R \sim 10^5 \text{ GeV},
\]

\[
\Gamma_\phi \sim 10^{-8} \text{ GeV},
\]

\[
|\lambda_1| \sim 10^{-10}.
\]

On the other hand, if some day it turns out that for example model NHT3 is the correct one for neutrino masses, then we have a prediction for the inflaton mass, \( M_I \sim (10^{11}-10^{12}) \text{ GeV} \). In that case all inflationary models that predict a different inflaton mass are ruled-out.

At this point we should add a comment regarding the gravitino constraint. In supersymmetric models one has to address the gravitino problem. Adding supersymmetry the expression for the baryon asymmetry will change slightly by a numerical factor of order one. Therefore one could use the results obtained so far for the non-supersymmetric case. If we require that \( T_R \leq (10^6-10^7) \text{ GeV} \) then we see that the models InvT2B and DegT1A are already compatible with the gravitino constraint, the model NHT3 is marginally compatible (for the lower values for \( T_R \)) with the gravitino constraint and finally the model InvT2B can be made compatible with the gravitino constraint lowering the upper bound for \( T_R \)

\[
1.25 \times 10^3 < T_R \leq (10^6-10^7) \text{ GeV}.
\]

### 5. Conclusions

In the present work we have studied non-thermal leptogenesis in six neutrino mass models proposed earlier and discussed recently in the literature. For each model we have obtained a formula relating the inflaton mass \( M_I \) to the reheating temperature after inflation \( T_R \). In fact according to this formula \( T_R \) is proportional to \( M_I \). Hence, the bigger the inflaton mass the bigger the reheating temperature. In a concrete inflationary model (chaotic [27], natural [28], supersymmetric hybrid [29], etc.) with a given mass for the inflaton, the right baryon asymmetry implies a certain reheating temperature after inflation. This in turn implies a certain decay rate for the inflaton field and a certain value for the inflaton Yukawa coupling. Furthermore, kinematical reasons and the requirement for non-thermal leptogenesis lead to a lower and an upper bound both for \( M_I \) and \( T_R \). Our results show that in most of the neutrino models under study the lower bound is not compatible with the upper bound and therefore only four cases survive. If we also take into account the gravitino constraint \( T_R \leq (10^6-10^7) \text{ GeV} \), then in one of these cases the reheating temperature is even more constrained.

### Acknowledgements

We would like to thank T.N. Tomaras for a critical reading of the manuscript. This work was supported by the EU grant MRTN-CT-2004-512194.

### References


References

Hidden sector baryogenesis

Bhaskar Dutta *, Jason Kumar

Department of Physics, Texas A&M University, College Station, TX 77843-4242, USA

Received 7 September 2006; accepted 22 September 2006
Available online 17 November 2006
Editor: M. Cvetič

Abstract

We introduce a novel mechanism for baryogenesis, in which mixed anomalies between the hidden sector and $U(1)_B$ drive the baryon asymmetry. We demonstrate that this mechanism occurs quite naturally in intersecting-brane constructions of the Standard Model, and show that it solves some of the theoretical difficulties faced in matching baryogenesis to experimental bounds. We illustrate with a specific example model. We also discuss the possible signals at the LHC.

© 2006 Elsevier B.V. All rights reserved.

1. Introduction

One of the great puzzles facing theoretical physics is the question of why we observe so little anti-matter, as compared to matter [1]. The generation of the asymmetry is called baryogenesis, and Sakharov demonstrated [2] that the three conditions required for it to occur are

- Violation of baryon number, $B$;
- $C$ and $CP$ violation;
- Departure from thermal equilibrium.

Several ideas have been proposed to satisfy these conditions, the most prominent of which are GUT baryogenesis, Affleck–Dine baryogenesis [3], baryogenesis via leptogenesis [4] and electroweak baryogenesis [5]. However all of these models have various theoretical or experimental constraints which make the fit to data problematic.

A parallel thread in model-building has been the construction, within the context of string theory, of intersecting brane models (IBMs) [6–8] which can perhaps provide a description of real-world low energy physics. These stringy constructions have provided new insights into the types of beyond-the-Standard-Model physics one might expect to find at colliders.

One of the ubiquitous features of these IBMs is the existence of hidden sectors, arising from the gauge theory living on extra branes (above and beyond the visible sector branes needed to generate the Standard Model). It is rather generic in this context for the hidden sector groups to have mixed anomalies with Standard Model $U(1)$’s. We suggest here a mechanism in which mixed anomalies between baryon number, $U(1)_B$, and hidden sector gauge groups can drive baryogenesis.

This provides a new mechanism for baryogenesis which not only provides a unique phenomenological signature, but also seems to appear rather generically in a large class of stringy constructions.

2. Motivation

In intersecting brane models, one obtains a Standard Model gauge theory as the low-energy limit of the theory of open strings which begin and end on a set of intersecting D-branes (the “visible sector”). In such models, however, one generically can find additional D-branes, and the gauge theory living on those branes provides a hidden sector. The strings which stretch between those hidden sector branes, and between hidden and visible sector branes, form exotic matter.1

---

1 Since there can be exotic matter charged under both the hidden and Standard Model gauge groups, our hidden sector is more precisely a pseudo-hidden sector.
In a construction with $N = 1$ supersymmetry (which arises in orientifold models), the strings stretching between branes yield degrees of freedom which are arranged into chiral multiplets. The net number of bifundamental chiral multiplets\footnote{Other representations arise when the effects of the orientifold are accounted for, but we will not need them here.} stretched between two branes \cite{9} is given by the topological intersection number $I_{ab}$:

$$I_{ab} \text{ multiplicities} \rightarrow (\square, \square_b).$$

(1)

Non-trivial topological intersections are somewhat generic in many constructions. One of the best understood IBMs arises from toroidal orientifolds (for example, $T^6/Z_2 \times Z_2(\times \Omega R)$) of Type IIA string theory, on which D6-branes are wrapped. The branes in this example wrap 3-cycles on the compact dimensions, which can be represented by three coprime ordered pairs of wrapping numbers: $(n_1, m_1)(n_2, m_2)(n_3, m_3)$, where $(n_i, m_i)$ are the wrapping numbers on the $a$ and $b$ cycles of the $i$th torus. The topological intersection between branes $a$ and $b$ is then

$$I_{ab} = \prod_{i=1}^{3}(n_i^a m_i^b - m_i^a n_i^b).$$

(2)

It is clear that $I_{ab} = 0$ only if D6-branes $a$ and $b$ have the same wrapping numbers on at least one torus. For a generic choice of wrapping numbers, this will not be the case, and $I_{ab} \neq 0$.

For a more general manifold, an orthogonal basis of 3-forms may be written as $\alpha_i, \beta^j$, where $f_\Sigma \alpha_i \wedge \beta^j = \delta_i^j$. Without loss of generality, assume brane $a$ wraps the 3-cycle dual to $\alpha_1$. The form dual to the cycle wrapped by brane $b$ is

$$\gamma = \sum_i d\alpha_i + b_i \beta^i.$$  

(3)

In this case, $I_{ab} = 0$ only if $b_1 = 0$. For a generic choice of $a_i, b_i$, this will not be the case, and again $I_{ab} \neq 0$.

It is necessary to ensure that the gauge theory has canceled anomalies. The cancellation of cubic anomalies is automatically ensured by the RR-tadpole constraints (i.e., the constraint that all space-filling charges cancel). There can also be mixed anomalies, however, which are canceled by a generalized Green–Schwarz mechanism. If a symmetry is broken by such an anomaly, then the associated gauge boson will receive mass through the Steckelberg mechanism, and the symmetry will appear to be an anomalous global symmetry at low-energies. If two branes $a$ and $b$ have non-trivial intersection, then there will be chiral fermions transforming under the groups $G_a$ and $G_b$, where $G_{a,b}$ are the gauge groups living on branes $a$ and $b$ respectively. It is clear from the field theory analysis that there will thus be a $U(1)_a - G^B_b$ mixed anomaly (where $U(1)_a$ is the diagonal $U(1)$ subgroup of $G_a$) given by

$$\partial_{\mu} j_{a}^{\mu} = \frac{I_{ab}}{32\pi^2} \text{Tr} F_b \wedge F_b.$$  

(4)

In a large class of intersecting brane world models, $SU(3)_{\text{qcd}}$ arises as a subgroup of a $U(3)$ gauge group living on a stack of 3 parallel D-branes (in certain cases where there is an orientifold plane, there will actually be 6 parallel D-branes in this stack). In such cases, the charge under the diagonal $U(1)_B$ is baryon number. As we have seen, $U(1)_B$ will generically have mixed anomalies with other gauge groups (both visible and hidden sector), provided that the $U(3)_{\text{qcd}}$ stack of branes and the other stack have non-trivial intersection.

We will consider the case where there is an anomaly between the hidden sector and $U(1)_B$. As a result, the divergence of the baryon current will be given by

$$\partial_\mu j_\mu^B \sim \text{Tr} F \wedge F,$$  

(5)

where $F$ is the field strength of the hidden sector gauge theory. Instantons or spherelons in the hidden sector will then violate baryon number, providing a source for the baryogenesis.

3. Baryogenesis driven by the hidden sector

Having motivated this mechanism from intersecting brane world constructions, we will develop this idea from the point of view of the low-energy effective field theory. In fact, motivation aside, this mechanism can appear just as readily in non-stringy constructions, and it will be easier to find specific models in low-energy effective field theory. We refer to Ref. \cite{10} for other work on baryogenesis in related contexts.

We will consider a theory with $N = 1$ SUSY and Standard Model gauge group and matter content, as well as a non-trivial hidden sector including hidden group $G$. We will need four features:

- The cancellation of all cubic anomalies (in IBMs, this is ensured by the RR-tadpole constraints);
- A non-vanishing $U(1)_B - G^2$ mixed anomaly;
- Vanishing $U(1)_Y$ mixed anomalies;
- A Yukawa coupling which permits exotic baryons to decay to SM baryons.

All multiplets charged under the fundamental of $SU(3)_{\text{qcd}}$ have charge $\frac{1}{3}$ under $U(1)_B$. In an intersecting brane model this will arise naturally, as $U(1)_B = \frac{1}{3} U(1)_{\text{diag}}$ is a gauged subgroup of $U(3)_{\text{qcd}}$. In a more general field theory model, $U(1)_B$ arises simply as a global symmetry.

The vanishing of $U(1)_Y$ mixed anomalies is easy to arrange in intersecting brane models. In that case, $U(1)_Y$ arises as a linear combination of $U(1)$’s, and in many constructions it is easy to arrange for the existence of such a non-anomalous symmetry. In such constructions, the vanishing of the hypercharge anomaly naturally leads to the existence of the appropriate Yukawa coupling. For example, one might arrange for the $U(1)_Y - G^2$ anomaly to vanish by ensuring a non-trivial intersection between the $G$ branes and a $U(1)_B$ brane. But as we will see, this permits a Yukawa coupling which allows exotic quarks to decay to right-handed quarks, plus an exotic scalar. From the effective field theory point of view, we merely need to choose our exotic with matter with appropriate hypercharge couplings to ensure vanishing anomalies and the appropriate Yukawa couplings.
3.1. A specific model

We will now look at a specific model with a hidden gauge group $G$ contained in a larger hidden sector. In our model, we have $U(1)_{T} = \frac{1}{2}(U(1)_{B} - U(1)_{L} + U(1)_{T})$, where $U(1)_{C}$ is the diagonal $U(1)$ subgroup of $G$. We have 2 chiral multiplets $q_{i}$ transforming in the bifundamental of $(U(3)_{B}, G)$ and with hypercharge $Q_{Y} = \frac{2}{3}$; four multiplets $\lambda_{j}$ transforming in the fundamental of $G$ with charge $Q_{T} = -1$ and hypercharge $Q_{Y} = -1$; one chiral multiplet $\eta$ transforming in the fundamental of $G$ with charge $Q_{L} = 1$ and hypercharge $Q_{Y} = 0$; and one chiral multiplet $\xi$ transforming in the anti-fundamental of $G$ with charge $Q_{L} = 1$ and hypercharge $Q_{Y} = 0$. The charges for this specific model are described in Table 1. This could arise in a brane model (assuming we label the branes as follows: $a = U(3)_{B}, b = U(1)_{T}, c = U(1)_{L}$ and $g = G$) with intersection numbers $I_{ab} = 2, I_{bc} = 4, I_{gb} = 1$, and $I_{g} = 1$, where $b'$ is the orientifold image of the $b$ brane.

We see that all of this matter is charged against the (anti-)fundamental of $G$, and the net hypercharge of this matter content is zero. As a result, we induce no $U(1)_{Y} - G^{2}$ mixed anomaly. Note however, that $U(1)_{B}$ and $U(1)_{B-L}$ have mixed anomalies with $G$. We assume that the rest of the hidden sector cancels the RR-tadpoles. This ensures the cancellation of all cubic anomalies. Assuming that there are no symmetric or anti-symmetric representations of $SU(3)_{qcd}$, it is easy to arrange by a judicious choice of the QCD branes, this also ensures that there are no net chiral exotics.

The divergence of the baryon current will contain a hidden-sector contribution given by

$$\rho_{\mu} f_{B}^{\mu} \propto \frac{1}{32\pi^{2}} (g_{G}^{2} \text{Tr} F_{G} \wedge F_{G} + \ldots).$$

At low energies $U(1)_{B}$ will appear to be an anomalous global symmetry. We will assume that $G$ and other hidden sector gauge groups break at some scale (breaking or confinement will be necessary in order to avoid exotic massless fermions which are charged under the Standard Model). The breaking of $G$ can involve complicated hidden sector dynamics such as brane recombination [8,11] in which other hidden sector groups simultaneously break. For a phenomenologically viable model, however, the $U(1)$ hypercharge must remain unbroken.

3.2. Phase transitions and baryogenesis

Having discussed the basic setup, one can now address the way baryogenesis actually occurs. As the universe expands and cools, we assume that there is a phase transition (such as the spontaneous symmetry breaking of $G$) at some temperature $T_{C}$.

If this $G$ phase transition is strongly first-order, then it will result in the nucleation of expanding bubbles of a broken symmetry vacuum.

At the bubble walls, there will be a departure from equilibrium. During this process, CP can generally be violated in the $G$-sector. $G$-sphalerons correspond to transitions from one vacuum of the hidden sector theory to another, and the mixed $U(1)_{B} - G^{2}$ anomaly implies that these transitions are accompanied by a discrete violation of baryon number [12]. All of the Sakharov conditions are thus satisfied, and a baryon asymmetry can be produced during the phase transition. $G$-sphalerons will be unsuppressed above the phase transition, but will generally be suppressed at low temperatures [13]. Thus, to avoid washout of the produced asymmetry by $G$-sphalerons after the phase transition, one must demand the usual condition [14]

$$\frac{v(T_{c})}{T_{c}} \geq 1,$$

where $v(T_{c}) = \langle \phi \rangle$ is the order parameter of the first order phase transition. This mechanism is reminiscent of electroweak baryogenesis [15], but does not suffer from the tunings required to fit electroweak baryogenesis into the parameter space allowed by LEP-II data and EDM bounds [16–18].

It is interesting to note that the amount of chiral matter charged under $G$ and $U(1)_{L}$ need not be the same as the amount charged under $G$ and $U(1)_{Y}$. As a result, there may be a $U(1)_{B-L} - G^{2}$ anomaly (indeed, $U(1)_{L}$ may have no anomaly). In IBMs, one expects a $U(1)_{B-L} - G^{2}$ anomaly unless $U(1)_{L}$ lives on a lepton brane which is parallel to the QCD branes, as in a Pati–Salam model. On thus expects that these $G$ sphalerons can violate both $B$ and $B-L$. As a result, even if $G$ breaks at a scale significantly larger than TeV, electroweak sphalerons will not wash out the baryon asymmetry. This naturally avoids one of the difficulties of GUT baryogenesis.

Of course, one can choose models where the $G$-sphalerons does preserve $B-L$. This will occur if the $U(1)_{L} - G^{2}$ anomaly has the same magnitude as the $U(1)_{B} - G^{2}$ anomaly, as is the case in Pati–Salam constructions of the SM sector (where $U(1)_{B-L}$ is a non-diagonal subgroup of $U(4)$). In this case, the hidden sector drives baryogenesis only if the scale of $G$ breaking is approximately at or below the electroweak scale. This would be natural in a scenario where supersymmetry breaking is communicated to both the $G$ and SM sectors by gravity/moduli.

In our specific example, the exotic particles generated by the $G$ sphalerons will be the exotic baryons schematically represented by $q_{i} q_{i} q_{i}$, as well as $\lambda_{i}, \eta$, and $\xi$. We will use a tilde to represent the scalar of the appropriate chiral multiplet, while the fermion will be represented without a tilde where no confusion is caused. In order to provide realistic baryogenesis, there must be a process whereby the $qqq$ baryons decay to Standard Model baryons and the $\lambda$ fermions decay to Standard Model particles. Generically, there will be Yukawa coupling terms of the form $^{3}$

<table>
<thead>
<tr>
<th>Particle</th>
<th>$Q_{B}$</th>
<th>$Q_{G}$</th>
<th>$Q_{T}$</th>
<th>$Q_{L}$</th>
<th>$Q_{Y}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q_{i}$</td>
<td>$\frac{1}{3}$</td>
<td>$-1$</td>
<td>0</td>
<td>0</td>
<td>$\frac{2}{3}$</td>
</tr>
<tr>
<td>$\lambda_{j}$</td>
<td>0</td>
<td>1</td>
<td>$-1$</td>
<td>0</td>
<td>$-1$</td>
</tr>
<tr>
<td>$\eta$</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>$-1$</td>
</tr>
<tr>
<td>$\xi$</td>
<td>0</td>
<td>$-1$</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

3 The subscripts on $w^{c}$ and $e^{c}$ denote flavors of right-handed up-type quarks and electron-type leptons.
\[ W_{\text{yuk.}} = c_{ij} q_i u_j^c \eta + d_{jm} \lambda_j e_m^c \xi + \cdots \] (8)

which allow an exotic \( q_i \) quark to decay to \( u^c \) and \( \bar{\eta} \) and allow \( \lambda_i \) to decay to \( e^c \) and \( \bar{\xi} \); we assume that these decays are kinematically allowed (if this is not the case, then we would instead find exotic baryons which do not decay to SM baryons). These decays conserve \( R \)-parity if we assign the following charges: \[ Q_\eta = Q_\lambda = -1, \quad Q_\bar{\eta} = Q_{\bar{\xi}} = 1. \] Indeed, we must be sure that \( \lambda_i \) can decay to only charged SM and neutral exotic bosons before nucleosynthesis, in order to avoid \( \text{Li}_6 \) production bounds \[19\].

The fields \( \eta \) and \( \xi \) can play the role of dark matter particles. These fields can get Majorana masses once the \( U(1)_{\text{T}SR} \times U(1)_{\text{L}} \) symmetry is broken (leaving only \( U(1)_Y \)). The fermionic parts of the fields (assuming they are lighter) can take part in constituting the dark matter of the universe. The annihilation of these new particles can happen via a \( t \)-channel exchange of exotic quarks.

But in a more general scenario where \( U(1)_L \) is unbroken, \( \xi \) cannot obtain a Majorana mass. As \( \xi \) is produced by the same \( G \)-sphalerons, one expects the \( \xi \) number density to be related to the baryon number density (the precise ratio depends on the specifics of a model). In a simple scenario the mass of \( \xi \) could be \( 10 m_{\text{proton}} \) which would provide a nice mechanism for relating the baryon and dark matter densities, along the lines of \[20\].

As we have not specified the precise nature of the hidden sector, it is not clear whether baryogenesis is dominated by local or non-local processes. If non-local baryogenesis dominates, and if \( G \)-sphalerons do not violate \( L \), then one might face a variety of effects which suppress baryogenesis \[21\].

### 3.3. Different transitions

In many known intersecting braneworld models, the hidden sector gauge groups are known to confine (the \( \beta \)-function for the \( USp \) groups are negative) \[7\], rather than break at low energies. It is interesting to consider how this impacts baryogenesis. The role of the first-order transition in the Sakharov conditions is to drive the system away from thermal equilibrium. From that point of view, a first-order confining phase transition will do just as well as a symmetry breaking transition. The fundamental question is the suppression of \( G \)-sphalerons after the transition. In a higgsing transition, it is clear that \( G \)-sphalerons will be suppressed below the transition, and thus would be unable to wash out the baryon asymmetry. But after a confining transition, it is not entirely clear if sphaleron-like processes are suppressed. This is analogous to the question of whether or not strong sphalerons are suppressed at temperatures below \( \Lambda_{\text{QCD}} \). It is of course difficult to make any concrete calculations, due to the inherent difficulties in computing in a strongly coupled gauge theory near confinement. But we expect that there should be a mass-gap on the confining side of the phase transition. Thus, we expect that there will be an upper limit on the size of instantons after confinement. As such, the energy barrier which the sphaleron-like process must cross should have a non-zero minimum size, which in turn implies Boltzmann suppression at low temperatures. So although it is not clear, it seems quite plausible that a first-order confining transition in the \( G \) sector can also produce a departure from equilibrium and seed baryogenesis, while shutting off sphalerons to prevent washout.

### 4. Signatures at the LHC

It is of prime interest to determine the signatures for this type of hidden sector baryogenesis (HSB) at LHC. As mentioned, HSB can occur even if \( G \) breaks at a relatively high-scale, provided that \( B-L \) is also broken. In this case, however, there will not necessarily be any clear signature visible at LHC. However, if supersymmetry breaking is mediated to both the visible and hidden sectors by gravity, then one might expect \( G \) to in any case break at a scale \( \sim \) TeV. One might expect that \( G \)-sphaleron processes can then be accessed at LHC. Unfortunately, this is likely not the case. As shown in \[22\] in the context of electroweak theory, sphalerons can be accessed efficiently at high temperature, but not in high-energy scattering. On the other hand, if \( G \) sector particles have masses set by the TeV scale, then they can be produced directly at LHC.

At the LHC, the exotic quark (\( q \)) can be pair-produced and the production process in this case would be \( gg \to q \bar{q} \) via a \( t \)-channel exchange of the exotic quark. The exotic quark \( q \) would then decay into \( q_{\text{SM}} \) and missing energy (\( \eta \)). The \( q_{\text{SM}} \) could be one of the up type quarks. The Yukawa couplings between the exotic quarks and the SM quarks are controlled by the structure of intersections between the \( G \)-branes and SM branes. In general, the signal will be multiple jets + leptons (arising from the decay of top quarks) + missing energy. The exotic quark can also be singly produced via \( qg \to q_{\text{exotic}} \eta \). The exotic quark then decays into a SM quark and \( \eta \). The jet \( E_T \) depends on the mass difference between the exotic quark and \( \eta \). If the \( E_T \) is large the signal becomes more easily accessible. So the final state can have a high \( E_T \) jet plus missing energy. We can also have leptonic signals once the \( \lambda \) 's are produced (via \( Z \) interaction) which will then decay into lepton plus missing energy (\( \xi \)). So the signal is similar as in \( R \)-parity conserving SUSY scenarios.

Interestingly, if the exotic quarks do not have the same hypercharge as Standard Model quarks, then the scalars \( \lambda \) and \( \eta \) would have fractional charge. This would provide a unique signature of new physics. But due to the difficulty in decaying fractionally charged particles into SM particles, such models would be tightly bound by cosmology data and direct tests, requiring the fractionally charged particles to recombine or annihilate almost entirely.

### 5. Conclusions

We have discussed a novel mechanism for baryogenesis which avoids many of the tight constraints arising from electroweak and GUT baryogenesis. This model utilizes a first-order transition in a hidden sector which has a mixed anomaly with \( U(1)_B \) to drive baryogenesis. As such, this mechanism naturally provides a way to break \( B-L \), allowing the hidden sector group to break at any scale without washout from electroweak sphalerons (a major concern for GUT baryogenesis). Furthermore, this mechanism does not face the same challenge...
as electroweak baryogenesis in fitting the precision data from LEP-II and other experiments.

Perhaps most notable, however, is that this mechanism seems natural in IBM’s. Hidden sectors appear generically, and as electroweak baryogenesis in fitting the precision data from SU(3) must break, and if the breaking is a first-order transition then one would expect baryogenesis. This mechanism can be expected to occur quite naturally, regardless of any other sources of baryon asymmetry. If the exotic baryons produced at the hidden sector transition can decay to SM baryons, then HSB can provide a substantial component of the asymmetry. If not, then it will provide exotic baryons which become a challenge in reconciling the IBM with observation. The signal of this scenario at the LHC will be consistent with multiple jets plus missing energy and jets plus leptons plus missing energy.

This is a fascinating example of how string theoretic input can provide intuition for low-energy phenomenology and cosmology. It will be interesting to see how this mechanism works in specific models, particularly those IBMs for which flux vacua can be counted. It would be quite interesting to determine, for example, brane models with large amounts of flux vacua which exhibit HSB. An analysis of the open string hidden-sector landscape [23] would be quite useful for this purpose.

Acknowledgements

We gratefully acknowledge Z. Chacko, M. Cvetič, J. Harvey, D. Morrissey, A. Rajaraman, S. Thomas, J. Wells and especially C. Wagner for very helpful discussions. J.K. would like to thank the Aspen Center for Physics, where part of this work was completed, for their hospitality. This work is supported in part by NSF grant PHY-0314712.

References

Coset models and D-branes in group manifolds

Domenico Orlando

Centre de Physique Théorique de l’École Polytechnique, 91128 Palaiseau, France
Laboratoire de Physique Théorique de l’École Normale Supérieure, 24 rue Lhomond, 75231 Paris Cedex 05, France

Received 26 September 2006; accepted 25 October 2006
Available online 7 November 2006

Editor: L. Alvarez-Gaumé

Abstract

We conjecture the existence of a duality between heterotic closed strings on homogeneous spaces and symmetry-preserving D-branes on group manifolds, based on the observation about the coincidence of the low-energy field description for the two theories. For the closed string side we also give an explicit proof of a no-renormalization theorem as a consequence of a hidden symmetry and infer that the same property should hold true for the higher order terms of the DBI action.

© 2006 Elsevier B.V. All rights reserved.

One of the main technical advantages provided by the study of models on group manifolds is that the geometrical analysis can be recast in Lie algebraic terms. At the same time the underlying conformal symmetry makes it possible to explicitly study the integrability properties that, in general, allow for extremely nice behaviours under renormalization. Wess–Zumino–Witten models can be used as starting points for many interesting models: The main challenge in this case consists in partially removing the symmetry while retaining as many algebraic and integrability properties as possible.

In this Letter we aim at pointing out an analogy (or, as we will say, a duality) between two—in principle disconnected—constructions based on WZW models: Closed string (heterotic) backgrounds obtained via asymmetric deformations and symmetry-preserving D-branes on group manifolds. As we will show, in fact, the low-energy field contents for both theories are the same, although they minimize different effective actions (SUGRA for the former and DBI for the latter). For one of the sides of the duality (the closed string one) we will also show a no-renormalization theorem stating that the effect of higher-order terms can be resummed to a shift in the radii of the manifold. A similar behaviour can also be conjectured from the D-brane side, and this would be consistent with some remark in literature about the coincidence between the DBI and CFT results concerning mass spectra, ... up to the said shift [1,2].

Let us start with the open-string side of this duality, by reminding some known facts about the geometric description of D-branes in WZW models on compact groups, pointing out in particular the low-energy field configuration. Natural boundary conditions on WZW models are those in which the gluing between left- and right-moving currents can be expressed in terms of automorphisms \( \omega \) of the current algebra. The corresponding world-volumes are then given by (twisted) conjugacy classes on the group [3]:

\[
\mathcal{C}^\omega(g) = \{ h g \omega(h^{-1}) \mid h \in G \}.
\]

As it was pointed out in [1], one can use Weyl’s theory of conjugacy classes so to give a geometric description of \( \mathcal{C}^\omega(g) \). For a given automorphism \( \omega \) we can always find an \( \omega \)-invariant maximal torus \( T \subset G \) (such as \( \omega(T) = T \)). Let \( T^\omega \subset T \) be the set of elements \( t \in T \) invariant under \( \omega \) (\( T^\omega = \{ t \in T \mid \omega(t) = t \} \)) and \( T^\omega_0 \subset T^\omega \) the connected component to the unity. When \( \omega \) is inner \( T = T^\omega = T^\omega_0 \) while in general (i.e. if we allow \( \omega \) to be outer) \( \dim(T^\omega_0) \leq \operatorname{rank} G \).

Let \( \omega \) be inner. Define a map:

\[
q : G/T \times T \to G, \quad ([g], t) \mapsto q([g], t) = g t g^{-1}.
\]
One can show that this map is surjective, so that each element in \( G \) is conjugated to some element in \( T \). This implies in particular that the conjugacy classes are characterized by elements in \( T \), or, in other words, fixing \( t \in T \) (so to take care of the action of the Weyl group), we find that the (regular) conjugacy classes \( C_{t}^{o}(g) \) are isomorphic to the homogeneous space \( G/T \). A similar result holds for twisted classes, but in this case

\[
C_{tw}^{o}(g) \simeq G/T_{0}^{o}.
\]

(3)

The description of the D-brane is completed by the \( U(1) \) gauge field that lives on it. The possible \( U(1) \) fluxes are elements in \( H^{2}(G/T_{0}^{o}, \mathbb{R}) \) and one can show that

\[
H^{2}(G/T_{0}^{o}) \simeq \mathbb{Z}^{\dim T_{0}^{o}}.
\]

(4)

Summarizing we find that the gauge content of the low energy theory is given by:

- the metric on \( G/T_{0}^{o} \) (in particular \( G/T \) for untwisted branes),
- the pull-back of the Kalb–Ramond field on \( G/T_{0}^{o} \),
- \( \dim T_{0}^{o} \) independent \( U(1) \) fluxes (rank \( G \) for untwisted branes).

These fields extremize the DBI action

\[
S = \int dx \sqrt{\det(g + B + 2\pi F)}
\]

(5)

and according to some coincidence with known exact CFT results there are reasons to believe that the fields only receive a normalization shift when computed at all loops.

Let us now move to the other—closed string—side of the advertised duality. A good candidate for a deformation of a WZW model that reduces the symmetry, at the same time preserving the integrability and renormalization properties, is obtained via the introduction of a truly marginal operator written as the product of a holomorphic and an antiholomorphic current

\[
\mathcal{O} = \sum_{ij} c_{ij} J^{i} \bar{J}^{j}.
\]

(6)

As it was shown in [4], a necessary and sufficient condition for this marginal operator to be integrable is that the left and right currents both belong to Abelian groups. If we consider the heterotic super-WZW model, a possible choice consists in taking the left currents in the Cartan torus and the right currents from the heterotic gauge sector [5,6]:

\[
\mathcal{O} = \sum_{a=1}^{N} H_{a} J^{a} \bar{J}^{a},
\]

(7)

where \( J^{a} \in H \subset T \), \( T \) being the maximal torus in \( G \).

Using a construction bearing many resemblances to a Kaluza–Klein reduction it is straightforward to show that the background fields corresponding to this kind of deformation consist in a metric, a Kalb–Ramond field and a \( U(1)^{N} \) gauge field. Their explicit expressions are simply given in terms of Maurer–Cartan one-forms on \( G \) as follows:

\[
g = \frac{k}{2} \delta_{ab} J^{a} \otimes J^{b} - k \delta_{ab} H_{a} J^{a} \otimes \bar{J}^{b},
\]

(8a)

\[
H_{[3]} = dB - \frac{1}{k_{g}} A^{a} \wedge dA^{a} = \frac{k}{2} f_{abc} J^{a} \wedge J^{b} \wedge J^{c} - k H_{a}^{2} f_{abc} J^{a} \wedge J^{b} \wedge J^{c},
\]

(8b)

\[
A^{a} = H_{a} \sqrt{\frac{2k}{k_{g}}} \bar{J}^{a} \quad \text{(no summation over } a \text{ implied)},
\]

(8c)

where \( \bar{J}^{a} \) are the currents that have been selected for the deformation operator. In this way we get an \( N \)-dimensional space of exact models. Here we will concentrate on a special point in this space, namely the one that corresponds to \( \{ H_{a} = 1/\sqrt{2}, \forall a = 1, 2, \ldots, N \} \). This point is remarkable for it corresponds to a decompactification limit where \( N \) dimensions decouple and we’re left with the homogeneous \( G/H \) space times \( N \) non-compact dimensions.\(^3\) More precisely, when \( H \) coincides with the maximal torus \( T \), the background fields read:

\[
G = \frac{k}{2} \sum_{\mu} J^{\mu} \otimes J^{\mu},
\]

(9a)

\[
H_{[3]} = dB = \frac{1}{2} f_{\mu \nu \rho} J^{\mu} \wedge J^{\nu} \wedge J^{\rho},
\]

(9b)

\[
F^{a} = - \sqrt{\frac{k}{2k_{g}}} H_{a} \frac{f^{a}_{\mu \nu}}{f_{\mu \nu} \wedge J^{\mu} \wedge J^{\nu}}
\]

(9c)

(no summation over \( a \)). Geometrically:

- \( g \) is the metric on \( G/T \) obtained as the restriction of the Cartan–Killing metric on \( G \);
- \( H_{[3]} \) is the pullback of the usual Kalb–Ramond field present in the wzw model on the group \( G \);
- \( F^{a} \) are rank(\( G \)) independent \( U(1) \) gauge fluxes that satisfy some quantization conditions and hence naturally live in \( H^{2}(G/T, \mathbb{Z}) \).

Having chosen a truly marginal operator for the deformation we know that this model is conformal. This implies in particular that the background fields solve the usual \( \beta \) equations that stem from the variation of the effective SUGRA action:

\[
S = \int dx \sqrt{\bar{g}} \left( R - \frac{1}{12} H_{\mu \nu \rho} H^{\mu \nu \rho} - \frac{k}{8} F^{a}_{\mu \nu} F^{a}_{\mu \nu} + \frac{\delta c}{3} \right),
\]

(10)

In example if we consider \( G = SU(2) \), then \( T = U(1) \) and the decompactification limit \( H \to 1/\sqrt{2} \) we get the exact \( S^{2} = SU(2)/U(1) \) background supported by a \( U(1) \) magnetic monopole field (see e.g. [7,8]).

Our conjecture stems precisely from this: The gauge field above exactly match the ones we found before for symmetry-preserving D-branes. Moreover both sides of the duality are derived from wzw models that enjoy a no-renormalization property which would make this correspondence true at all orders. In this spirit we now pass to prove that a similar theorem holds

\(^3\) One can see the initial group manifold \( G \) as a principal fibration of \( H \) over a \( G/H \) basis: The deformation changes the radii of the fiber and eventually trivializes in correspondence of this special point.
for closed heterotic strings on coset models inferring that the duality, when proven, would give a direct way to deduce the same feature for the D-brane action.

In studying symmetrically deformed WZW models, i.e., those where the deformation operator is written as the product of two currents belonging to the same sector $O = \lambda J$, one finds that the Lagrangian formulation only corresponds to a small-deformation approximation. For this reason different techniques have been developed so to read the background fields at every order in $\lambda$ [9–13] but, still, the results are in general only valid at first order in $\lambda$ and have to be modified so to take into account the effect of instanton corrections. In this section we want to show that this is not the case for asymmetrically deformed models, for which the background fields in Eqs. (8) are exact at all orders in $h_a$ and for which the effect of renormalization only amounts to the usual (for WZW models) shift in the level of the algebra $k \to k + c G$ where $c G$ is the dual Coxeter number.

Consider in example the most simple SU(2) case. In terms of Euler angles the deformed Lagrangian is written as:

$$S = S_{SU(2)}(\alpha, \beta, \gamma) + \delta S = \frac{k}{4\pi} \int d^2z \delta \phi \bar{\delta} \phi + \delta \phi \bar{\delta} \phi + \bar{\delta} \phi \delta \phi + 2 \cos \beta \bar{\delta} \alpha \bar{\delta} \gamma + 2 \cos \beta \delta \alpha \delta \gamma + \frac{\sqrt{k} H}{2\pi} \int d^2z (\delta \phi + \cos \beta \bar{\delta} \alpha) \bar{I}.$$  

(11)

If we bosonize the right-moving current as $\bar{I} = \bar{\delta} \phi$ and add a standard $U(1)$ term to the action, we get:

$$S = S_{SU(2)}(\alpha, \beta, \gamma) + \delta S(\alpha, \beta, \gamma, \phi) + \frac{k}{4\pi} \int d^2z \delta \phi \bar{\delta} \phi = S_{SU(2)}(\alpha, \beta, \gamma) + \frac{k}{4\pi} \int d^2z \delta \phi \bar{\delta} \phi + \frac{k}{4\pi} \int d^2z \delta \phi \bar{\delta} \phi$$  

(12)

and in particular at the decoupling limit $H \to 1/\sqrt{2}$, corresponding to the $S^2$ geometry, the action is just given by $S = S_{SU(2)}(\alpha, \beta, \gamma + 2 \sqrt{k} H \phi)$. This implies that our (deformed) model inherits all the integrability and renormalization properties of the standard SU(2) WZW model. In other words the three-dimensional model with metric and Kalb–Ramond field with $SU(2) \times U(1)$ symmetry and a $U(1)$ gauge field is uplifted to an exact model on the SU(2) group manifold (at least locally): The integrability properties are then a consequence of this hidden SU(2) symmetry that is manifest in higher dimensions.

The generalization of this particular construction to higher groups is easily obtained if one remarks that the Euler parametrization for the $g \in SU(2)$ group representative is written as:

$$g = e^{i \gamma t_1} e^{i \beta t_2} e^{i \phi t_3},$$  

(13)

where $t_1 = \sigma_1 / 2$ are the generators of $su(2)$ ($\sigma_1$ being the usual Pauli matrices). As stated above, the limit deformation corresponds to the gauging of the left action of an Abelian subgroup $T \subset SU(2)$. In particular here we chose $T = \{ h \mid h = e^{i \phi t_3} \}$, hence it is natural to find (up to the normalization) that:

$$h(\phi) g(\alpha, \beta, \gamma) = g(\alpha, \beta, \gamma + \phi).$$  

(14)

The only thing that one needs to do in order to generalize this result to a general group $G$ consists in finding a parametrization of $g \in G$ such as the chosen Abelian subgroup appears as a left factor. In example if in SU(3) we want to gauge the $U(1)^2$ Abelian subgroup generated by $(\lambda_3, \lambda_8)$ (Gell–Mann matrices), we can choose the following parametrization for $g \in SU(3)$ [14]:

$$g = e^{i \lambda_3 \phi} e^{i \lambda_5 \phi} e^{i \lambda_2 \beta} e^{i \lambda_3 \delta} e^{i \lambda_5 \delta} e^{i \lambda_2 \beta} e^{i \lambda_3 \alpha}.$$  

(15)

The deep reason that lies behind this property (differentiating symmetric and asymmetric deformations) is the fact that not only the currents used for the deformation are preserved (as it happens in both cases), but here their very expression is just modified by a constant factor. In fact, if we write the deformed metric as in Eq. (8a) and call $\tilde{K}^\mu$ the Killing vector corresponding to the chosen isometry (that does not change along the deformation), we see that the corresponding $\tilde{J}_\mu^{(H)}$ current is given by:

$$\tilde{J}_\mu^{(H)} = \tilde{K}^\mu \gamma^{(H)} \tilde{g}^{\mu \nu} = (1 - 2 H^2) \tilde{J}_\mu^{(0)}.$$  

(16)

The most important consequence (from our point of view) of this integrability property is that the SUGRA action in Eq. (10) is exact and the only effect of renormalization is the $k \to k + c G$ shift.

It is very tempting to extend this no-renormalization theorem to the D-brane side. Of course this would require an actual proof of the duality we conjecture. Nevertheless we think that this kind of approach might prove (at least for these highly symmetric systems) more fruitful than adding higher loop corrections to the DBI action, which on the other hand remains an interesting directions of study by itself.

Is this duality just a coincidence, due to the underlying Lie algebraic structures that both sides share, or is it a sign of the presence of some deeper connection? Different aspects of the profound meaning of the DBI effective action are still poorly understood and it is possible that this approach—pointing to one more link to conformal field theory—might help shedding some new light.

Acknowledgements

I would like to thank C. Bachas, C. Kounnas and S. Ribault for illuminating discussions and especially thank M. Petropoulos for encouragement during different stages of this work and for the careful reading of the manuscript.

References

Supersymmetric renormalization prescription in $\mathcal{N} = 4$ super-Yang–Mills theory

Laurent Baulieu $^{a,b}$, Guillaume Bossard $^b$,*

$^a$ Theoretical Division CERN, CH-1211 Geneva 23, Switzerland

$^b$ LPTHE, CNRS, Universités Pierre et Marie Curie et Denis Diderot, 4 place Jussieu, F-75252 Paris Cedex 05, France

Received 12 October 2006; accepted 25 October 2006

Available online 3 November 2006

Editor: L. Alvarez-Gaumé

Abstract

Using the shadow dependent decoupled Slavnov–Taylor identities associated to gauge invariance and supersymmetry, we discuss the renormalization of the $\mathcal{N} = 4$ super-Yang–Mills theory and of its coupling to gauge-invariant operators. We specify the method for the determination of non-supersymmetric counterterms that are needed to maintain supersymmetry.

© 2006 Elsevier B.V. All rights reserved.

1. Introduction

Non-linear aspects of supersymmetry and the non-existence of a supersymmetry-preserving regulator make the renormalization of supersymmetric theories a subtle task. Whichever is the choice of regularization, we expect non-supersymmetric counterterms for maintaining supersymmetry at the renormalized level. A very effective regularization of UV divergences of super-Yang–Mills theories, called-dimensional reduction, was introduced quite early by Siegel [1]. Whether this regularization holds true at all orders in perturbation theory was questioned in [2]. With suitable improvements, its compatibility with the quantum action principle was shown in [3]. In fact, this regularization cannot preserve supersymmetry beyond 3-loop order [4], which implies the introduction of non-supersymmetric counterterms for 4-loop computations. As another complication, the renormalizable Lorentz covariant gauge conditions (Landau–Feynman-type gauges) break supersymmetry. This breaking of a global symmetry is analogous to that of the Lorentz invariance by axial or Coulomb gauges for the ordinary Yang–Mills theories, but it is more intricate, because supersymmetry is realized non-linearly. This question was addressed by Dixon [5,6], who completed the ordinary BRST symmetry transformations for gauge invariance by adding supersymmetry transformations, whose supersymmetry parameter is a commuting constant spinor. The “enlarged BRST symmetry” determines a Slavnov–Taylor identity. It was shown, in a series of papers by Stöckinger et al., that this process allows the determination, order by order in perturbation theory, of non-invariant counterterms [7] that restore supersymmetry covariance of Green functions in the $\mathcal{N} = 1$ models [8]. The unusual feature that occurs is that Feynman rules depend on the parameter of supersymmetry, but it is advocated that observables do not depend on it. This method has a conceptual backlash. To define the “enlarged BRST symmetry”, Dixon changed the transformation law of the Faddeev–Popov ghost (to achieve nilpotency of the “enlarged BRST transformations”). But then, the BRST equation of the Faddeev–Popov ghost loses its geometrical meaning. Moreover, observables are not defined as they should be, from the cohomology of the BRST differential, since, in this case, they would be reduced to supersymmetry scalars. They must be introduced as gauge-invariant functionals of physical fields, which are well defined classically, but are sources of confusion at the quantum level, because of their possible mixing with non-gauge-invariant operators. In fact, the previous methods are sufficient to define certain rules for practical perturbative computations,
but the way they are obtained lacks the important feature of relying on a well-funded algebraic construction. The latter must be independent of the renormalization scheme and clearly separates gauge invariance from supersymmetry.

In recent papers, we indicated the possibility of disentangling these two invariances, for defining the quantum theory, with independent Slavnov–Taylor identities [9]. We introduced new fields, which we called shadows, not to confuse them with the usual Faddeev–Popov ghosts. The advantage of doing so is as follows. The obtained pair of differential operators allows us to define the two Slavnov–Taylor operators that characterize the gauge-fixed BRST-invariant supersymmetric quantum field theory, while the Faddeev–Popov ghost keeps the same geometrical interpretation as in the ordinary Yang–Mills theory. Observables are defined by the cohomology of the BRST differential Slavnov–Taylor operator and their supersymmetry covariance is controlled at the quantum level by the other Slavnov–Taylor operator for supersymmetry. This will allow for an unambiguous perturbative renormalization of supersymmetric gauge theories.

The shadow fields are assembled into BRST doublets, and they do not affect the physical sector. The quantum field theory has an internal bigrading, the ordinary ghost number and the new shadow number. The commuting supersymmetry parameter is understood as an ordinary gauge parameter for the quantum field theory. The prize one has to pay for having shadows is that they generate a perturbative theory with more Feynman diagrams. If we consider physical composite operators that mix through renormalization with BRST-exact operators, we have in principle to consider the whole set of fields in order to compute the supersymmetry-restoring non-invariant counterterms. For certain “simple” Green functions, which cannot mix with BRST-exact composite operators, there exist gauges in which some of the additional fields can be integrated out, in a way that justifies the work of Stöckinger et al. in the \( \mathcal{N} = 1 \) theories. By doing this elimination, we lose the geometrical meaning, but we may gain in computational simplicity.

In the conformal phase of the \( \mathcal{N} = 4 \) super-Yang–Mills theory, the observables are usually defined as correlation functions of gauge-invariant operators. The aim of this Letter is to discuss their quantum definition and the methodology that is needed to non-ambiguously compute non-invariant counterterms and maintain supersymmetry. In fact, our results apply to the renormalization of all supersymmetric theories.

2. Shadow fields and supersymmetry Slavnov–Taylor identities

2.1. Action and symmetries

The physical fields of the \( \mathcal{N} = 4 \) super-Yang–Mills theory in \( 3 + 1 \) dimensions are the gauge field \( A_\mu \) the \( SU(4) \)-Majorana spinor \( \lambda \), and the six scalar fields \( \phi^i \) in the vector representation of \( SO(6) \sim SU(4) \). They are all in the adjoint representation of a compact gauge group that we will suppose simple. The classical action is uniquely determined by \( Spin(3,1) \times SU(4) \), supersymmetry and gauge invariance. It reads

\[
S \equiv \int d^4x \text{Tr} \left( -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \frac{1}{2} D_\mu \phi^i D^\mu \phi_i + i \frac{1}{2} (\bar{\lambda} \gamma_\mu \lambda - \frac{1}{2} (\bar{\lambda} \lambda) - \frac{1}{4} (\phi^i, \phi^j) [\phi_i, \phi_j] \right) \tag{1}
\]

with \( \phi \equiv \phi^i \tau_i \) and the supersymmetry transformations \( \delta^{\text{Susy}} \)

\[
\delta^{\text{Susy}} A_\mu = i(\bar{\epsilon} \gamma_\mu \lambda), \quad \delta^{\text{Susy}} \phi^i = -(\bar{\epsilon} \tau^i \lambda), \quad \delta^{\text{Susy}} \lambda = \left( \mathcal{F} + i \bar{\psi} \phi + \frac{1}{2} [\phi, \phi^j] \right) \epsilon. \tag{2}
\]

Out of \( \delta^{\text{Susy}} \), we can build an operator \( Q \) that also acts on the Faddeev–Popov field \( \Omega \) and new shadow fields \( c, \mu \) [9]. \( Q \) acts on all the physical fields as \( Q = \delta^{\text{Susy}} - \delta^\text{gauge}(\epsilon) \), and we have

\[
Qc = (\bar{\epsilon} \phi - i A^c_\epsilon) - c^2, \quad Q\Omega = -\mu - [c, \Omega], \quad Q\mu = -[(\bar{\epsilon} \phi) \Omega] + i (\bar{\epsilon} \gamma_\mu \epsilon) D_\mu \Omega - [c, \mu]. \tag{3}
\]

The BRST operator \( s \) is nothing but a gauge transformation of parameter \( \Omega \) on all physical fields, and we have

\[
s\Omega = -\Omega^2, \quad sc = \mu, \quad s\mu = 0. \tag{4}
\]

To define a BRST-exact supersymmetric gauge-fixing, we introduce the trivial quartet \( \bar{\mu}, \bar{c}, \bar{\Omega}, b \), with

\[
s\bar{\mu} = \bar{c}, \quad sc = 0, \quad s\bar{\Omega} = b, \quad sb = 0.
\]

\[
Q\bar{\mu} = \bar{\Omega}, \quad Q\bar{c} = -b, \quad Q\bar{\Omega} = -i (\bar{\epsilon} \gamma_\mu \epsilon) \partial_\mu \bar{\mu}, \quad Qb = i (\bar{\epsilon} \gamma_\mu \epsilon) \partial_\mu \bar{c}. \tag{5}
\]

\[\footnote{We do not exclude the possibility of also reducing the set of fields in the general case, including observables that mix with BRST-exact operators through renormalization, but further investigations are needed in order to establish this statement.}

\[\footnote{The parameter \( \epsilon \) is a commuting spinor, so that \( \delta^{\text{Susy}} \approx \delta^\text{gauge}(\bar{\epsilon} \phi - i A^c_\epsilon) - i (\bar{\epsilon} \gamma_\mu \epsilon) \partial_\mu \), where \( \approx \) stands for the equality modulo equations of motion.} \]
On all fields, we have $s^2 = 0$, $Q^2 \approx -i(\bar{\epsilon}\gamma^\mu\epsilon)\partial_\mu$, $\{s, Q\} = 0$. We have the following renormalizable supersymmetric $sQ$-exact gauge-fixing actions:

$$-sQ \int d^4x \mathrm{Tr}\left(\tilde{\mu}\partial^\mu A_\mu + \frac{\alpha}{2}\tilde{\mu}\bar{\epsilon}\right).$$

(6)

By introducing sources associated to the non-linear $s$, $Q$ and $sQ$ transformations of fields, we get the following $\epsilon$-dependent action, which initiates a BRST-invariant supersymmetric perturbation theory:

$$\Sigma \equiv \frac{1}{g^2} S - \int d^4x \mathrm{Tr}\left(b\partial^\mu A_\mu + \frac{\alpha}{2}b^2 - \tilde{\epsilon}\partial^\mu(D_\mu c + i(\bar{\epsilon}\gamma^\mu\lambda)) - \frac{i\alpha}{2}(\bar{\epsilon}\gamma^\mu\epsilon)c\tilde{\partial}_\mu\tilde{c}
+ \tilde{\Omega}\partial^\mu D_\mu - \tilde{\mu}\partial^\mu(D_\mu + [D_\mu, c]) - i(\bar{\epsilon}\gamma^\mu[\Omega, \lambda])\right)
+ \int d^4x \mathrm{Tr}\left(A_\mu^+(D^\mu + \tilde{\lambda}^+(\Omega, \lambda) - \phi_i^+(\Omega, \phi_i) + A_\mu(Q)A^\mu - \tilde{\lambda}^+(Q)\lambda + \phi_i^+(Q)\phi^i
+ A_\mu^+(sQ\mu - \tilde{\lambda}^+(Q)\lambda + \phi_i^+(Q)\phi^i + \Omega^+(\Omega)\Omega - \Omega^+(Q)\Omega - \Omega^+(Q)\Omega
- c^+(Q)c + \mu(Q)\mu + \frac{g^2}{2}(\tilde{\lambda}^+(Q) - [\tilde{\lambda}^+(Q), \Omega])\right).$$

(7)

Because of the $s$ and $Q$ invariances, the action is invariant under both Slavnov–Taylor identities defined in [9], which are associated respectively to gauge and supersymmetry invariance, $S_{(s)}(\Sigma) = S_{(Q)}(\Sigma) = 0$. The supersymmetry Slavnov–Taylor operator is:

$$S_{(Q)}(\mathcal{F}) \equiv \int d^4x \mathrm{Tr}\left(\delta R_{A^\mu} \delta L_{A^\mu_{(Q)}} + \delta R_{\lambda} \delta L_{\lambda_{(Q)}} + \delta R_{\phi_i} \delta L_{\phi_i_{(Q)}} + \delta R_{\Omega} \delta L_{\Omega_{(Q)}} + \frac{\delta R_{\bar{\epsilon}}}{\delta \bar{\epsilon} c_{(Q)}} + \frac{\delta R_{\Phi_{(Q)}}}{\delta \Phi_{(Q)}} + \frac{\delta R_{\Lambda_{(Q)}}}{\delta \Lambda_{(Q)}} - \frac{\delta R_{\lambda}}{\delta \bar{\epsilon} c_{(Q)}} - \frac{\delta R_{\phi_i}}{\delta \Phi_{(Q)}} - \frac{\delta R_{\Omega}}{\delta \Lambda_{(Q)}} \right)
- i(\bar{\epsilon}\gamma^\mu\epsilon)\left(-\partial_\mu A_\mu^+(Q)\delta L_{A^\mu} + \partial_\mu \tilde{\lambda}^+(Q)\delta L_{\lambda} - \partial_\mu \phi_i^+(Q)\delta L_{\phi_i} + \partial_\mu \Omega^+(Q)\delta L_{\Omega} - \partial_\mu \lambda^+(Q)\delta L_{\lambda}
+ A_\mu^+(Q)\partial_\mu A^\nu + \tilde{\lambda}^+(Q)\partial_\mu \lambda + \phi_i^+(Q)\partial_\mu \phi^i + \Omega^+(Q)\partial_\mu \Omega + c^+(Q)\partial_\mu c + \mu(Q)\partial_\mu \mu\right).$$

(8)

2.2. Observables

The observables of the $N = 4$ super-Yang–Mills theory in the conformal phase are Green functions of local operators in the cohomology of the BRST linearized Slavnov–Taylor operator $S_{(s)}(\Sigma)$. From this definition, these Green functions are independent of the gauge parameters of the action, including $\epsilon$. Classically, they are represented by gauge-invariant polynomials of the physical fields [9,10]. We introduce classical sources $\alpha$ for all these operators. We must generalize the supersymmetry Slavnov–Taylor identity for the extended local action that depends on these sources. Since the supersymmetry algebra does not close off-shell, other sources $\nu$, coupled to unphysical $S_{(s)}(\Sigma)$-exact operators, must also be introduced. We define the following field and source combinations $\varphi^s$:

$$A^s_\mu \equiv A^+(Q) - \partial_\mu \tilde{c} - [A^+(Q), \partial_\mu \bar{\epsilon}, \Omega].$$
$$c^s \equiv c^+(Q) - [\mu(Q), \Omega].$$
$$\phi_i^s \equiv \phi_i^+(Q) - [\phi_i^+(Q), \Omega].$$
$$\lambda^s \equiv \lambda^+(Q) - [\lambda^+(Q), \Omega].$$

(9)

They verify $S_{(Q)}(\Sigma)\varphi^s = -[\Sigma, \varphi^s]$. The collection of local operators coupled to the $\nu$’s is made of all possible gauge-invariant (i.e. $S_{(s)}(\Sigma)$-invariant) polynomials in the physical fields and the $\varphi^s$’s. These operators have ghost number zero, and their shadow number is negative, in contrast with the physical gauge-invariant operators, which have shadow number zero.

3 Note that power counting forbids a gluino dependence for the argument of the $sQ$-exact term, and that $Q$ is nilpotent on all the functionals that do not depend on the gluinos. $\alpha$ is the usual interpolating Landau–Feynman gauge parameter.

4 $M$ is the $32 \times 32$ matrix $M = \frac{1}{2}(\bar{\epsilon}\gamma^\mu\epsilon)\gamma_{\mu} + \frac{1}{2}(\bar{\epsilon}\gamma^\nu\epsilon)\gamma_{\nu} + (\bar{\epsilon}\gamma^\nu\bar{\epsilon} - (\bar{\epsilon}\gamma^\nu\gamma^\mu\epsilon)\gamma_{\mu} + (\bar{\epsilon}\gamma^\nu\gamma^\mu\bar{\epsilon})\gamma_{\mu}$). It occurs because $Q^2$ is a pure derivative only modulo equations of motion. The dimension of $A_\mu$, $\lambda$, $\phi_i$, $\Omega$, $\tilde{\lambda}^+(\Omega)$, $\beta$, $\bar{\epsilon}$, $c$, and $\tilde{c}$ are respectively 1, $\frac{1}{2}$, 1, 0, 2, 2, 2, 2, 2, and 2. Their ghost and shadow numbers are respectively $(0, 0)$, $(0, 0)$, $(0, 0)$, $(1, 0)$, $(-1, 0)$, $(0, 0)$, $(1, 1)$, $(0, 0)$, $(1, 1)$, $(0, 0)$, and $(1, 0)$. They are all non-negative.

5 The linearized Slavnov–Taylor operator $S_{(Q)}(\Sigma)$ [9] verifies $S^2_{(Q)}(\Sigma) = -i(\bar{\epsilon}\gamma^\mu\epsilon)\partial_\mu$, which solves in practice the fact that $Q^2$ is a pure derivative only modulo equations of motion.
The relevant action is thus \( \Sigma[u, v] \equiv \Sigma + \mathcal{Y}[u, v] \), with

\[
\mathcal{Y}[u, v] \equiv \int d^4x \left( u_{ij} \frac{1}{2} \text{Tr} \phi^i \phi^j + u_{i}^{\alpha} \text{Tr} \phi^i \lambda_\alpha + u_{ijk} \frac{1}{3} \text{Tr} \phi^i \phi^j \phi^k \right.
\]

\[
+ K_{ij}^\mu \text{Tr} \left( i \phi^{ij} D_{\mu} \phi^j \right) + \frac{1}{2} \gamma_{ij} \tau^j \phi^i \lambda + K_{ij}^{\mu\nu} \text{Tr} \left( F_{\mu\nu} \phi^i - \frac{1}{2} \bar{\lambda} \gamma_{\mu\nu} \tau^j \phi^i \right) + K_{ij}^{\mu} \frac{1}{2} \text{Tr} \bar{\lambda} \gamma_{\mu} \tau^j \phi^i \lambda
\]

\[
+ C_{ijk} \text{Tr} \left( \frac{1}{3} \phi^{ijk} \phi^k \right) + \frac{1}{2} \bar{\lambda} \gamma_{\mu} \tau^j \lambda + C_{i j}^{\mu\nu} \text{Tr} \left( i \phi^{ij} D_{\mu} \phi^j \right) - \frac{1}{4} \frac{1}{\lambda} \gamma_{\mu\nu} \tau^j \phi^i \lambda
\]

\[
+ C_{i j}^{\mu} \frac{1}{2} \text{Tr} \bar{\lambda} \gamma_{\mu} \tau^j \lambda + C_{ij}^{\mu\nu} \text{Tr} \left( F_{\mu\nu} \phi^i - \frac{1}{2} \bar{\lambda} \gamma_{\mu\nu} \tau^j \phi^i \right) + C_{i j}^{\mu} \frac{1}{2} \text{Tr} \bar{\lambda} \gamma_{\mu} \tau^j \phi^i \lambda
\]

\[
+ \frac{1}{4} \lambda^a \phi^{ij} + i \nu^{\alpha\beta} \text{Tr} \phi^i \lambda_\alpha \lambda_\beta + i \nu^{\alpha} \phi^i \phi^j \phi^k + i \nu^{\alpha} \text{Tr} \phi^i \phi^j \phi^k + i \nu^{\alpha} \text{Tr} \phi^i \phi^j \phi^k + \cdots\right).
\]

Here, the \( \cdots \) stand for all other analogous operators.

The Slavnov–Taylor operator \( S_{(Q)} \) can be generalized into a new one, \( S_{(Q)}^{\text{ext}} \), by addition of terms that are linear in the functional derivatives with respect to the sources \( u \) and \( v \), in such a way that

\[
S_{(Q)}^{\text{ext}}(\Sigma[u, v]) = S_{(Q)}(\Sigma) + S_{(Q)}^{\text{ext}} \mathcal{Y} + \int d^4x \text{Tr} \left( \frac{\delta R \mathcal{Y}}{A_{\mu}} \frac{\delta L_{\gamma}}{\delta A_{\mu}^a} + \frac{\delta R \mathcal{Y}}{\delta \lambda} \frac{\delta L_{\gamma}}{\delta \phi^i} \frac{\delta L_{\gamma}}{\delta \bar{\phi}^i} \right) = 0.
\]

Indeed, if we were to compute \( S_{(Q)}(\Sigma[u, v]) \) without taking into account the transformations of the sources \( u \) and \( v \), the breaking of the Slavnov–Taylor identity would be a local functional linear in the set of gauge-invariant local polynomials in the physical fields, \( A_{\mu}^a, c^a, \phi^i, \) and \( \lambda^\lambda \).

Eq. (11) defines the transformations \( S_{(Q)}^{\text{ext}}(\Sigma) \) of the sources \( u \) and \( v \). Simplest examples for the transformation laws of the \( u \)’s are for instance

\[
S_{(Q)}^{\text{ext}}(\Sigma) u_{ij} = -i \left[ \gamma_{\mu} \tau^{ij} \gamma_{\epsilon} \gamma_{\epsilon} \right] \partial_{\mu} u_{ij}^{\alpha} + \partial_{\mu} \partial_{\alpha} u_{ij}^{\alpha} + 2 u_{ij}^{\alpha} u_{\alpha}^{\beta} v_{ij}^{\beta} - i \partial_{\mu} \left( u_{ij}^{\alpha} v_{ij}^{\alpha} + u_{ij}^{\beta} v_{ij}^{\beta} \right).
\]

\[
S_{(Q)}^{\text{ext}}(\Sigma) u_{i}^{\alpha} = -i \left[ \gamma_{\mu} \tau^{i} \gamma_{\epsilon} \gamma_{\epsilon} \right] \partial_{\mu} u_{i}^{\alpha} + 2 i \left[ \gamma_{\mu} \gamma_{\epsilon} \gamma_{\epsilon} \right] \partial_{\mu} u_{i}^{\alpha} + i \left[ \gamma_{\mu} \gamma_{\epsilon} \gamma_{\epsilon} \right] \partial_{\mu} v_{i}^{\alpha}.
\]

\[
- u_{ij}^{\alpha} v_{ij}^{\alpha} + u_{ij}^{\beta} v_{ij}^{\beta} + u_{i}^{\alpha} v_{i}^{\beta} + i \partial_{\mu} \left( u_{ij}^{\alpha} v_{ij}^{\beta} - u_{ij}^{\beta} v_{ij}^{\alpha} \right).
\]

These transformations are quite complicated in their most general expression. However, for many practical computations of non-supersymmetric local counterterms, we can consider them at \( v = 0 \). We define \( Q u \equiv (S_{(Q)}^{\text{ext}}(\Sigma) u \big|_{v=0}) \). By using \( \delta \Sigma^{\text{Susy}} \mathcal{Y}[u] + \mathcal{Y}[Q u] = 0 \) we can in fact conveniently compute \( Qu \). Notice that \( Q \) is not nilpotent on the sources, but we have the result that \( \mathcal{Y}[Q^2 u] \) is a linear functional of the equation of motion of the fermion \( \lambda \).

In [9,11], we showed the absence of anomaly and the stability of the \( N = 4 \) action \( \Sigma \) under renormalization. Thus, the complete theory involving shadows and ghosts can be renormalized, in any given regularization scheme, so that supersymmetry and gauge invariance are preserved at any given finite order. It is a straightforward and precisely defined process to compute all observables, provided that a complete set of sources has been introduced. This lengthy process cannot be avoided because there exists no regulator that preserves both gauge invariance and supersymmetry. We must keep in mind that renormalization generally mixes physical observables with BRST-exact operators, and a careful analysis must be done [12].

### 3. Enforcement of supersymmetry

We now turn to the problem of determining non-invariant counterterms, which are necessary to ensure supersymmetry at the quantum level. We will use the notation of [7] for the 1PI correlation functions, an example of which is

\[
\left( A_{\mu}^{a}(p) \bar{\lambda}^{a}(k) \phi_{i}^{Q}(x) \right) = \int d^4x \int d^4y \ e^{i(p + k) \cdot y} \frac{\delta^L \delta^L \delta^L \Gamma}{\delta \phi^{a}(x) \delta \phi^{a}(0) \delta \lambda^{a}(y) \delta A^{a}(x)}.
\]

All fields and sources are set equal to zero after the differentiation. Latin letters \( a, b, \ldots \) label the index of the gauge Lie algebra. In this section we focus on the case of observables that do not mix with non-gauge-invariant operators. The following subsection explains how computations are simplified in this case.
3.1. Loop cancellations

We can first eliminate by Gaussian integration the Faddeev–Popov ghosts \( \Omega, \tilde{\Omega} \) against the shadows \( \tilde{\mu}, \mu \) for computing some observables, in our class of linear gauges.

The antighost Ward identities of [9] determine the following dependence of the 1PI generating functional \( \Gamma \) in the fields \( \tilde{\mu}, \tilde{\Omega}, \bar{c} \) and \( b \)

\[
\Gamma[\ldots, \tilde{\mu}, \bar{c}, \tilde{\Omega}, b, A_\mu^{(s)}, A_\mu^{(Q)}, A_{\mu,\nu}^{(Q)}] = \Gamma[\ldots, 0, 0, 0, A_\mu^{(s)} - \partial_\mu \tilde{\Omega}, A_\mu^{(Q)} + \partial_\mu \bar{c}, A_{\mu,\nu}^{(Q)} + \partial_\mu \tilde{\mu}] \\
- \int d^4x \text{Tr} \left( b \partial^\mu A_\mu + \frac{\alpha}{2} b^2 - \frac{i\alpha}{2} (\bar{c} \gamma^\mu \epsilon) c \partial_\mu \bar{c} \right),
\]

where the \( \ldots \) stand for the dependence on all other fields and sources. Consider the generating functional of 1PI correlation functions of the subset of fields \( \varphi_{\text{sub}} \) made of the physical fields, the shadow \( c \), \( \bar{c} \) and the sources associated to the \( Q \) variations of these fields. The pair of \( Q \)-doubles \( \Omega, \mu \) and \( \tilde{\mu}, \tilde{\Omega} \) only appear in the Feynman diagrams through propagators and interactions defined by the following part of the action

\[
\int d^4x \text{Tr} (\partial^\mu \tilde{\Omega} D_\mu \Omega - \partial^\mu \tilde{\mu} D_\mu \mu).
\]

Feynman rules show that the fields \( \Omega, \tilde{\Omega}, \mu \) and \( \tilde{\mu} \) exactly compensate in closed loops of opposite contributions at least at the regularized level. The following Ward identities imply that this property is maintained after renormalization

\[
(\tilde{\mu}^{(a)}(p)\mu^b) + (\tilde{\Omega}^{(a)}(p)\Omega^{(O)}(p)\mu^b) = (\bar{c} \gamma^\mu \epsilon) c (\partial^\mu \tilde{\Omega}^{(a)}(p)\Omega^{(O)}(p)\mu^b) = 0.
\]

\( \Omega^{(O)} \) is the source of the operator \( \mu + [\Omega, c] \) and the term linear in \( \mu \) of the insertion of \( [\Omega, c] \) in \( \Gamma \) must be zero. It follows that \( (\Omega^{(O)}(p)\mu^b) = \delta^b_c \), at any given finite order of perturbation theory. The only superficially divergent 1PI Green functions depending on \( \Omega, \tilde{\Omega}, \mu \) and \( \tilde{\mu} \) that must be considered are \( (\tilde{\mu}^{(a)}(p)\mu^b) = -(\tilde{\Omega}^{(a)}(p)\Omega^{(O)}(p)\mu^b) \) and \( (\mu^{(a)}(p)A^b_\mu(k)\partial^\mu \tilde{\Omega}^{(O)}(p + k)\mu^c) = -(\mu^{(a)}(p)A^b_\mu(k)\Omega^{(O)}(p + k)\mu^c) \).

We can thus integrate out these fields in all correlation functions of the fields \( \varphi_{\text{sub}} \).

After this elimination, the supersymmetry Slavnov–Taylor identity is sufficient to constrain the 1PI Green functions of the fields \( \varphi_{\text{sub}} \) to the same values as they would have in the complete procedure without the ab initio elimination of \( \tilde{\mu}, \tilde{\Omega} \) and \( \bar{c} \) and \( b \), and the following simplified super-symmetry Slavnov–Taylor identity

\[
\Sigma^{(Q)}(\mathcal{F}) \equiv \frac{1}{g^2} S + Q \int d^4x \text{Tr} \left( \bar{c} \partial^\mu A_\mu + \frac{\alpha}{2} \bar{c} b \right) \\
+ \int d^4x \text{Tr} \left( A_\mu^{(Q)} Q A^\mu - \bar{c} (\partial^\mu \tilde{\Omega}^{(Q)} \Omega + \phi^{(Q)} Q \phi - c^{(Q)} Q c + \frac{g^2}{2} \bar{c} \lambda^{(Q)} M \lambda^{(Q)} \right).
\]

The ambiguities of the quantum theory are fixed by the antighost Ward identities for \( \bar{c} \) and \( b \), and the following simplified super-symmetry Slavnov–Taylor identity

\[
S_{(Q)}(\mathcal{F}) \equiv \int d^4x \text{Tr} \left( \frac{\delta R \mathcal{F}}{\delta A_\mu^{(Q)}} \frac{\delta L \mathcal{F}}{\delta A_\mu^{(Q)}} + \frac{\delta R \mathcal{F}}{\delta \lambda^{(Q)}} \frac{\delta L \mathcal{F}}{\delta \lambda^{(Q)}} + \frac{\delta R \mathcal{F}}{\delta \phi^{(Q)}} \frac{\delta L \mathcal{F}}{\delta \phi^{(Q)}} + \frac{\delta R \mathcal{F}}{\delta c^{(Q)}} \frac{\delta L \mathcal{F}}{\delta c^{(Q)}} \\
- i(\bar{c} \gamma^\mu \epsilon) \left( A_\mu^{(Q)} \partial_\mu \lambda + \bar{c} \lambda^{(Q)} \partial_\mu \lambda + \phi^{(Q)} \partial_\mu \phi - c^{(Q)} \partial_\mu c \right) - \frac{b}{\delta c^{(Q)}} + i(\bar{c} \gamma^\mu \epsilon) \partial_\mu \bar{c} \partial_\mu \mathcal{F} \right).
\]

This identity is analogous to that in [5,6,8]. However, we now understand that \( c \) is not the Faddeev–Popov ghost, and that observables must be defined in the enlarged theory.

This simplified process with less fields can be applied also for computing 1PI correlation functions with insertions of certain physical composite operators (we call them “simple” operators), as long as these operators do not mix through renormalization with BRST-exact operators (which would imply computing insertions of operators depending on other fields than the \( \varphi_{\text{sub}} \)). At the tree level, these “simple” operators are all the gauge-invariant polynomials in the physical fields that are in representations of \( \text{Spin}(3,1) \times SU(4) \) in which there exist no BRST-exact operators of the same canonical dimensions that depend on the antighost \( \tilde{\mu}, \tilde{\Omega} \) and \( \bar{c} \) only through their derivatives. Examples of “simple” operators are the local operators of canonical dimension \( [O] < 4 \) and the BPS primary operators.
3.2. Renormalization of the action

We assume that the “restricted” theory has been renormalized at a given order of perturbation theory, say $n$, by using the best available regularization, namely-dimensional reduction, and renormalization conditions such that the supersymmetry Slavnov–Taylor identity and the so-called antighost Ward identities are satisfied. Within this scheme, finite gauge-invariant, but not supersymmetric, counterterms must occur after a certain order of perturbation theory. At a given order $n$, the action is thus of the following form

$$
\Sigma^{n} = \frac{1}{g^{2}} \int d^{4}x \text{Tr} \left( -\frac{1}{4} F_{\mu \nu}^{a} F^{a}_{\mu \nu} - D_{\mu}^{a} \phi^{\prime a} D^{\mu} \phi^{\prime a} + i \frac{1}{2} (\bar{\lambda}^{a} \phi^{\prime b} \lambda^{b}) 
- \frac{g^{1}}{2} (\lambda^{a} [\phi^{\prime a}, \lambda^{b}]) - \frac{g^{2}}{4} [\phi^{\prime a}, \phi^{\prime b}] (\bar{\phi}^{\prime a}, \bar{\phi}^{\prime b}) + h_{1} \phi^{\prime a} \phi^{\prime b} \lambda^{a} \lambda^{b} + h_{2} \phi^{\prime a} \phi^{\prime b} \lambda^{a} \lambda^{b} ) 
+ \int d^{4}x \left( g_{3} \text{Tr} \phi^{\prime a} \phi^{\prime b} \phi^{\prime c} \phi^{\prime d} + h_{3} \text{Tr} \phi^{\prime a} \phi^{\prime b} \phi^{\prime c} \phi^{\prime d} \right) 
+ \int d^{4}x \left( -b^{2} \phi^{\prime a} A_{\mu}^{a} - \frac{a}{2} b^{2} + c^{2} \partial^{\mu} \left( D_{\mu}^{a} c^{a} + i y_{1} (\bar{c} \gamma_{\mu} \lambda^{a}) \right) \right) + y_{2} i \alpha^{b} \left( \bar{c} \gamma^{\mu} \epsilon \right) c^{a} \partial_{\mu} \right) 
+ \int d^{4}x \left( A^{(Q)}_{a \mu} (x_{1} (\bar{c} \gamma^{\mu} \lambda^{a}) + D^{\mu} c^{a} ) - \phi^{(Q) a} \lambda^{a} \right) \right) 
+ \int d^{4}x \left( -x^{b} \phi^{\prime a} + i x_{4} \phi^{\prime a} + \frac{x_{5}}{2} \left[ \phi^{\prime a}, \phi^{\prime b} \right] + h_{4} \phi^{\prime a} \lambda^{a} \right) \right) 
+ c^{(Q) a} \left( -x_{6} (\bar{c} \phi^{\prime a} \epsilon ) + i x_{7} (\bar{c} \phi \epsilon ) (c^{a} + c^{\dagger a} ) \right) + \frac{g^{2}}{2} \langle \lambda^{(n \rightarrow n)} N \lambda^{(Q) a} \rangle. 
\right) 
\right)
$$

(19)

The conformal property of $\mathcal{N} = 4$ implies that the coupling constant is not renormalized. In this expression, the index $b$ on top of a field $\phi$ indicates its multiplicative renormalization by an infinite factor $\sqrt{Z_{\phi}}$, which is a Taylor series of order $n$ in the coupling constant $g$. The sources $\phi^{(Q) a}$ are renormalized by the inverse factor $1/\sqrt{Z_{\phi}}$ as a result of the BRST Slavnov–Taylor identities. The parameters $g_{1}, h_{1}, x_{1}, y_{1}$ and the $32 \times 32$ symmetric matrix $N$ (quadratic in $c^{a}$) are finite power series in $g^{2}$ of order $n$, which have been fine-tuned to enforce supersymmetry. In the simplest case of the $SU(2)$ gauge group, the parameters $h_{1}$ are redundant and can be set to zero. This action permits us to perturbatively compute the renormalized 1PI generating functional $\Gamma^{n+1}_{\text{corr}}$ of the $\phi^{\text{sub}}$ at order $n$, such that the Slavnov–Taylor identity of supersymmetry is verified at this order. To obtain the action (19) at the following order $n+1$, we then use the minimal subtraction scheme with dimensional reduction, which defines the infinite factors $Z_{\phi}$ at order $n+1$. They yield as an intermediary result the “minimally” renormalized 1PI generating functional $\Gamma^{n+1}_{\text{min}} = \sum_{p=0}^{n} \Gamma_{(p)} + \Gamma^{n+1}_{(n+1)}$. The supersymmetry Slavnov–Taylor identity is possibly broken at order $n+1$, as follows.\footnote{\begin{enumerate}
\item $N$ can be parametrized by five parameters as follows
\begin{equation}
N = a_{1} (\bar{c} \gamma^{\mu} \epsilon ) y_{1} + a_{2} (\bar{c} \gamma^{e} \epsilon ) t_{1} + 6a_{3} (\bar{c} \gamma^{\mu} \gamma^{e} \gamma^{\mu} \gamma^{e} ) y_{1} t_{1} + 2a_{4} (\bar{c} \gamma^{y_{5}} \gamma^{e} \gamma^{\mu} \gamma^{e} ) y_{5} t_{1} + a_{5} (\bar{c} \gamma^{y_{5}} \gamma^{e} ) y_{5} t_{1}.
\end{equation}
\item $\langle \mathcal{F}, G \rangle$ is the antibracket
\begin{equation}
\int d^{4}x \text{Tr} \left( \delta \mathcal{F} \delta \mathcal{G} + \frac{\delta \mathcal{F}}{\delta A^{a}_{\mu}} \frac{\delta \mathcal{G}}{\delta A^{a}_{\mu}} + \frac{\delta \mathcal{F}}{\delta \lambda} \frac{\delta \mathcal{G}}{\delta \lambda} + \frac{\delta \mathcal{F}}{\delta \phi^{a}} \frac{\delta \mathcal{G}}{\delta \phi^{a}} + \frac{\delta \mathcal{F}}{\delta \phi^{\prime a}} \frac{\delta \mathcal{G}}{\delta \phi^{\prime a}} + \frac{\delta \mathcal{F}}{\delta c^{a}} \frac{\delta \mathcal{G}}{\delta c^{a}} + \frac{\delta \mathcal{F}}{\delta c} \frac{\delta \mathcal{G}}{\delta c} + \frac{\delta \mathcal{F}}{\delta A^{(Q) a}} \frac{\delta \mathcal{G}}{\delta A^{(Q) a}} + \frac{\delta \mathcal{F}}{\delta \lambda^{(Q) a}} \frac{\delta \mathcal{G}}{\delta \lambda^{(Q) a}} + \frac{\delta \mathcal{F}}{\delta \phi^{\prime (Q) a}} \frac{\delta \mathcal{G}}{\delta \phi^{\prime (Q) a}} \right).
\end{equation}
\end{enumerate}}

\begin{equation}
S_{(Q)}(\Gamma^{n+1}_{(n+1)}) = \frac{1}{2} \sum_{p=1}^{n} (\Gamma_{(p)} + \Gamma_{(n+1)-p}) + \text{S}_{(Q)} \Sigma^{n+1}_{(n+1)} + \mathcal{O}(g^{2n+4}).
\end{equation}

(20)

Any given term in the right-hand side of Eq. (20) may be non-local, but the sum of these terms is a local functional of fields and sources, as is warranted by the quantum action principle. There is no supersymmetry anomaly \cite{6,9} and the consistency relation $S_{(Q)}(\Gamma^{(n+1)}_{(n+1)}) = 0$ implies the existence of the local functional $\Sigma^{n+1}_{(n+1)}$ such that $S_{(Q)}(\Gamma^{(n+1)_{(n+1)}} + \Sigma^{n+1}_{(n+1)}) = \mathcal{O}(g^{2n+4})$. Thus the component of order $n+1$ of the parameters $g_{1}, h_{1}, x_{1}, y_{1}$ and the matrix $N$ can be modified in such a way that the resulting 1PI generating functional $\Gamma^{n+1}_{(n+1)}$ satisfies the supersymmetry Slavnov–Taylor identity at order $n+1$.

The fine-tuning at order $n+1$ will be achieved if a large enough number of relations between 1PI Green functions are satisfied. They are obtained by suitable differentiations of the supersymmetry Slavnov–Taylor identity. The number of ambiguities removed by the Slavnov–Taylor identity is finite and corresponds to that of parameters of the action. Thus the relations between the 1PI Green functions only have to be implemented on their renormalization conditions. These relations must be expanded on Lorentz and gauge group invariant tensors.
The antighost Ward identities fix the ambiguities on the Green functions that contain the antishadow $\bar{c}$ and the $b$ field. The identities
\begin{equation}
(\bar{c}^a(p)\lambda^b_{\alpha c})^\mu = -i\mu^\mu [A^{(Q)\mu}_a(p)\lambda^b_{\alpha c}]^\mu, \quad (\bar{c}^a(p)\phi^b_{\alpha} = \alpha(\hat{e}\phi)\delta^{ab} - i\mu^\mu [A^{(Q)\mu}_a(p)\phi^b_{\alpha}] + i\mu^\mu [A^{(Q)ib}_a(-p)\phi^b_{\alpha}]
\end{equation}
permit to compute the value of $y_1$ in function of $x_1$, and $y_2$ at the $n + 1$ order.

We first use the components of the Slavnov–Taylor identity that expresses the closure of the supersymmetry algebra at the quantum level
\begin{equation}
[A^{(Q)\mu}_a(p)\lambda^c_{\alpha b}][A^{(Q)\mu}_b(p)\lambda^a_{\alpha c}] + [A^{(Q)\mu}_a(p)\phi^b_{\alpha c}][A^{(Q)\mu}_c(p)\phi^a_{\alpha b}] + [\bar{c}^a(p)\eta_{\mu \nu}] [A^{(Q)\sigma}_a(p)\lambda^b_{\beta c}][A^{(Q)\sigma}_b(p)\lambda^c_{\beta a}] = 0,
\end{equation}
\begin{equation}
[A^{(Q)\mu}_a(p)\lambda^c_{\alpha b}][A^{(Q)\mu}_c(p)\lambda^a_{\alpha b}] + [A^{(Q)\mu}_a(p)\phi^b_{\alpha c}][A^{(Q)\mu}_c(p)\phi^a_{\alpha b}] + [\bar{c}^a(p)\eta_{\mu \nu}] [A^{(Q)\sigma}_a(p)\lambda^b_{\beta c}][A^{(Q)\sigma}_b(p)\lambda^c_{\beta a}] = 0,
\end{equation}
\begin{equation}
[A^{(Q)\mu}_a(p)\lambda^c_{\alpha b}][A^{(Q)\mu}_c(p)\phi^a_{\alpha b}] + [A^{(Q)\mu}_a(p)\phi^b_{\alpha c}][A^{(Q)\mu}_c(p)\phi^a_{\alpha c}] + [\bar{c}^a(p)\eta_{\mu \nu}] [A^{(Q)\sigma}_a(p)\lambda^b_{\beta c}][A^{(Q)\sigma}_b(p)\phi^c_{\beta a}] = 0.
\end{equation}

These identities imply that the quantities $x_i$ are functions of only two independent parameters. In turn, both parameters are determined from the following Slavnov–Taylor identities, which express the supersymmetry covariance of physical Green functions
\begin{equation}
[A^{(Q)\mu}_a(p)A^{(Q)\nu}_b(p)]\lambda^c_{\alpha d} + [\bar{c}^a(p)\delta^{(2)}_{ab}] [A^{(Q)\alpha d}_a(p)\lambda^c_{\beta d}] = 0,
\end{equation}
\begin{equation}
[\bar{c}^a(p)\phi^b_{\alpha c}][A^{(Q)\mu}_a(p)\phi_{\alpha c}] + [\bar{c}^a(p)\phi^b_{\alpha c}][A^{(Q)\mu}_a(p)\phi_{\alpha c}] + [\bar{c}^a(p)\phi^b_{\alpha c}][A^{(Q)\mu}_a(p)\phi_{\alpha c}] = 0.
\end{equation}

It remains to determine the matrix $N$, which is related to the terms quadratic in the sources. This can be done using the identity
\begin{equation}
[A^{(Q)\mu}_a(p)\lambda^b_{\alpha c}][A^{(Q)\mu}_b(p)\lambda^c_{\alpha b}] + [A^{(Q)\mu}_a(p)\phi_{\alpha c}][A^{(Q)\mu}_c(p)\phi_{\alpha b}] + [\bar{c}^a(p)\phi^b_{\alpha c}][\bar{c}^a(p)\phi^b_{\alpha c}] = 0.
\end{equation}

3.3. Renormalization of local observables

We must also renormalize the part of the action that is linear in the components of the local observables. Consider a set of local operators that mix together by renormalization. Suppose that each one of these operators can be considered as the element of an irreducible supersymmetry multiplet. Then, all the other components of the supersymmetry multiplets will mix by renormalization with the same matrix of anomalous dimensions. As for the ordinary Green functions, non-supersymmetric counterterms must be perturbatively computed for enforcing the Ward identities. The method of the preceding section can be generalized. We decompose each source into irreducible representations of Spin$(3,1) \times SU(4)$ and write the most general gauge-invariant functional linear in the sources.

\begin{equation}
\mathcal{Y}^{u, v} = \int d^4x \left( Z^K u_i \frac{1}{2} Tr \phi^{ij} \phi^j + Z^K u_{ij} \frac{1}{2} Tr \left( \phi^{ij} \phi^{ij} - \frac{1}{6} \delta^{ij} \phi^b \phi^b \right) + Z^K u_i r^l Tr \phi^{b c} \right)
\end{equation}
\begin{equation}
+ Z^K u_i u_j Tr \left( \phi^{ij} \phi^j - \frac{2}{1} r^l \phi^l \phi^l \right) + Z^K \left[ u_{ij} Tr \phi^{ij} \left[ \phi^{ij} , \phi^{ij} \right] + Z^K u_{ij} Tr \left( \frac{1}{3} \phi^{ij} \phi^{ij} \phi^k + \frac{1}{8} z_{ij} \phi^{ij} \phi^k \phi^k \right) \right)
\end{equation}
\begin{equation}
+ Z^K C u_{ij} Tr \phi^{ij} \left[ \phi^{ij} , \phi^{ij} \right] + Z^K C u_{ij} Tr \left( \frac{1}{3} \phi^{ij} \phi^{ij} \phi^k + \frac{1}{8} z_{ij} \phi^{ij} \phi^k \phi^k \right) + \ldots \right).
\end{equation}

There is an ambiguity corresponding to each one of the renormalization factors $Z^i$, to be fixed by the supersymmetry Slavnov–Taylor identity. At a given order, we first perturbatively compute the infinite part of the renormalization factors. Then the finite part of the renormalization factors must be adjusted, as for ordinary Green functions.

Consider as the simplest cases the Konishi operator $\mathcal{O}^{KL} \equiv \frac{1}{2} Tr \phi^L \phi^L$, and the $\frac{1}{2}$ BPS operator $\mathcal{O}^{KL} \equiv \frac{1}{2} Tr (\phi^L \phi^L - 1/6 \delta^{ij} \phi^k \phi^k)$. The renormalization factors of the first two components of the associated supermultiplets are related because of the Ward identity
\begin{equation}
[\bar{e} \phi^L \phi^L] [u_{ij} (p) \phi^a_i (k) \phi^b_l] + [u_{ij} (p) \phi^a_i (k) \bar{e} \phi^{b c}] [\lambda^{(Q)\mu}_a (p + k) \phi^b_l] + [u_{ij} (p) \phi^a_i (k) \bar{e} \phi^{b c}] [\lambda^{(Q)\mu}_a (p) \phi^b_l] = 0.
\end{equation}
A less simple example, for which there could be non-supersymmetric counterterms with a mixing-matrix, is for the cubic operator \( \text{Tr} \phi^i \phi^j \phi^k \) of the Konishi multiplet. The global symmetries and power counting allow this operator to mix with \( \text{Tr}(\frac{1}{8} \sigma^i \phi^j \phi^k) \) belonging to the \( \frac{1}{2} \) BPS multiplet associated to \( O^4 \). The identity

\[
3\mathbf{e}^{t k j} \mathbf{e}^{i} \left[ u_{k i m}(-p_1 - p_2 - p_3)\phi_{1}^{a_1} (p_1)\phi_{2}^{a_2} (p_2)\phi_{3}^{a_3} \right] + 3\mathbf{e}^{t j k} \mathbf{e}^{i} \left[ c_{k i m}(-p_1 - p_2 - p_3)\phi_{1}^{a_1} (p_1)\phi_{2}^{a_2} (p_2)\phi_{3}^{a_3} \right] + \sum_{r \in \mathcal{E}_3} [u_{r j}(-p_1 - p_2 - p_3)\phi_{1}^{a_{1+}} (p_1+r)\alpha^{rb} ||^{(Q)} (p_1+r)\phi_{3}^{a_{1+}} (p_1+r)\phi_{3}^{a_{1+}} = 0
\]

and the component in \( [\tau^{mnp}]_{lq} \) of the following one

\[
3\mathbf{e}^{t k j} \mathbf{e}^{i} \left[ u_{k i j} (p)\lambda^{a_2} (k)\lambda^{a_2} (k) \right] + 3\mathbf{e}^{t j k} \mathbf{e}^{i} \left[ c_{j i r} (p)\lambda^{a_2} (k)\lambda^{a_2} (k) \right] + \left[ \left[ u^{r j} (p)\lambda^{a_2} (k)\phi^{ij}_{(Q)} (p + k)\lambda^{a_2} (p + k) \right] - \left[ u^{r j} (p)\lambda^{a_2} (p - k)\phi^{ij}_{(Q)} (p - k)\lambda^{a_2} (p - k) \right] \right] [\tau^{mnp}]_{lq} = 0
\]

permits us to determine perturbatively the renormalization factors of these operators in function of those of \( O^K \) and \( O^C \).

In fact, the renormalization factors of all the other components of the supermultiplet containing \( O^K \) and \( O^C \) can be perturbatively computed as a function of those of the operators \( O^K \) and \( O^C \).

### 3.4. Contact terms

After computing the renormalization of one insertion of “simple” physical operators in all Green functions of fields \( \phi_{sub} \), we may want to compute their multicorrelators. The renormalization of these correlation functions possibly involves the addition of contact terms. Such counterterms cannot be generated in the minimal scheme prescription. However, dimensional reduction breaks supersymmetry, and we expect that finite contact-counterterms must be added to the action, for restoring supersymmetry. To compute these possible counterterms, we write the more general \( \text{Spin}(3, 1) \times SU(4) \)-invariant action that depends only on the sources \( u \), in a polynomial way

\[
\mathbb{Z}[u] = \frac{1}{2} \int d^4 x \left( z_1 u^{ij} u_{ij} + z_2 u^{ij} u_{ij} - i z_3 u_{ij} u^{ij} - i z_4 u_{ij} u^{ij} + z_5 u^{(ijk)} \partial^i u_{(ijk)} + z_6 u^{(ij)} \partial^i u_{(ij)} + z_7 u^{(ij)} \partial^i u_{(ij)} + z_8 u^{(ij)} \partial^i u_{(ij)} + \cdots \right).
\]

The values of the renormalization factors \( z_l \) can then be computed, by imposing the supersymmetry Slavnov–Taylor identity, order by order in perturbation theory. As before, it is sufficient to enforce some identities between relevant correlation functions. The simplest identity

\[
[\delta \tau^{mj} \mathbf{e}^{i} \left[ u^{(jk)} (p)u^{(j)} \right] + \left[ \delta \mathbf{e}^{i} \right]^{(k)} (p)u^{(k)} = 0
\]

constrains \( z_3 \) and \( z_4 \) as functions of \( z_1 \) and \( z_2 \), and so on. In practice, we have to define renormalization conditions for each one of the classes of superficially divergent correlation functions that are not related by the supersymmetry Slavnov–Taylor identity. Within a given class, the renormalization conditions of all correlation functions are related by supersymmetry Slavnov–Taylor identities. The non-invariant contact-counterterms can then be perturbatively computed by perturbatively enforcing these renormalization conditions.

### Acknowledgements

We thank Luis Alvarez-Gaumé and Raymond Stora for good discussions on the subject.

### References

Mixing angles and non-degenerate coupled systems of particles

Q. Duret, B. Machet *

Laboratoire de Physique Théorique et Hautes Énergies †, Unité Mixte de Recherche UMR 7589, Université Pierre et Marie Curie—Paris 6,
CNRS, Université Denis Diderot—Paris 7, France

Received 11 August 2006; accepted 12 September 2006
Available online 28 September 2006
Editor: L. Alvarez-Gaumé

Abstract

Defining, in the framework of quantum field theory, their mass eigenstates through their matricial propagator, we show why the mixing matrices of non-degenerate coupled systems should not be parametrized as unitary. This is how, for leptonic binary systems, two-angles solutions with discrete values $\pi/4 \text{ mod } \pi/2$ and $\pi/2 \text{ mod } \pi$ (in addition to the trivial case $0 \text{ mod } \pi$) arise when weak leptonic currents of mass eigenstates approximately satisfy the two properties of universality and vanishing of their non-diagonal neutral components. Charged weak currents are also discussed, which leads to a few remarks concerning oscillations. We argue that quarks, which cannot be defined on shell because of the confinement property, are instead more naturally endowed with unitary Cabibbo-like mixing matrices, involving a single unconstrained mixing angle. The similarity between neutrinos and neutral kaons is outlined, together with the role of the symmetry by exchange of families.

© 2006 Elsevier B.V. All rights reserved.

PACS: 11.10.-z; 12.15.Ff; 14.60.Pq

1. Introduction

The observed large mixing angles in the neutrino sector have already long been a matter of surprise and questioning [1,2]. Symmetries [3] have been invoked, which should be approximate since the mixing is only close to maximal, various “textures” of mass matrices have been proposed [4,5] which are not fully satisfying either, and furthermore unstable by unitary transformations on flavour eigenstates [6].

In the while, the $CP$ violating parameters $\epsilon_L$ and $\epsilon_S$ of the physical neutral kaons $K_L$ and $K_S$ have been shown [7,8], using the propagator formalism of quantum field theory, not to be rigorously identical. This amounts to a tiny lack of unitarity in the mixing matrices linking “in” (or “out”) mass eigenstates to the orthonormal basis of flavour eigenstates. Phrased in another way, mixing matrices of physical kaons cannot be parametrized with a single mixing angle. It was also shown that systems of this type should not be described by a single constant mass matrix and that their mass eigenstates cannot be correctly determined by a bi-unitary transformation.

In this Letter we establish a link between these two peculiarities and show they are among general properties of non-degenerate coupled systems of particles.

2. General framework

2.1. Flavour and mass eigenstates

Since their couplings to the Higgs boson are not flavor-diagonal, massive fermions in the standard model form coupled systems (like neutral kaons). The usual approach to such systems makes use of a mass matrix. It was however shown in [7,8] that it is an inadequate procedure: indeed, a (constant) mass matrix can only be introduced as a linear approximation to the inverse propagator in the vicinity of each of its poles, such that, in the case under study, as many mass matrices as there are poles should be considered. This is why we stick below to basic principles of quantum field theory, which state in particular
that the physical masses of particles, bosons or fermions, can only be the poles of their full propagator. The corresponding eigenstates—the propagating states—are the mass/spin eigenstates of the Lorentz–Poincaré group.

Two bases play a fundamental role, in particular in the electroweak physics of leptons: the $2^{nf}$ flavour eigenstates $|^f e, \pm f, \nu_e, f, \nu_\mu, f, \ldots\rangle$ which, by convention, couple to weak vector bosons, and the $2^{nf}$ propagating eigenstates $|^m e, \pm m, \nu_e, m, \nu_\mu, m, \ldots\rangle$, which are also the mass eigenstates. At the classical level only left-handed flavour fermions weakly couple, but the situation changes when quantum corrections are included.

The physical masses $z_i = m_i^2$ satisfy by definition the gauge invariant pole equation (the variable $z$ is used for $q^2$)

$$\det \Delta(z) = 0, \quad \text{for } z = z_i, \quad (1)$$

where $\Delta(z)$ is the full (renormalized) $2nf \times 2nf$ matrix propagator in momentum space. The solutions of (1) are independent of the renormalization procedure. The propagating (mass) eigenstates $\psi_m^i$ are the corresponding eigenvectors with vanishing eigenvalues

$$\Delta^{-1}(z = z_i) \psi_m^i = 0. \quad (2)$$

It is also convenient to introduce $\Delta^{-1}(z) = L^{(2)}(z)$ as the renormalized quadratic Lagrangian operator. (1) then reads

$$\det L^{(2)}(z) = 0, \quad (3)$$

and the mass eigenstates satisfy the equation (equivalent to (2))

$$L^{(2)}(z = z_i) \psi_m^i = 0. \quad (4)$$

The situation is accordingly that of a $z$ dependent $2nf \times 2nf$ matrix $L^{(2)}(z)$, the $2nf$ eigenvalues $\lambda_j(z), j = 1, \ldots, 2nf$ of which are supposed non-degenerate and satisfy, by definition of the poles $z_i$ of $\Delta(z), \lambda_i(z_i) = 0$. At any $z$ it has $2nf$ eigenvectors $\psi^j(z), j = 1, \ldots, 2nf$. When $z \rightarrow z_i$, $\psi^i(z) \rightarrow \psi_m^i$ and $\psi^j(z), j \neq i \rightarrow \omega_i^j$. So, among the $2nf$ eigenvectors of $L^{(2)}(z_i)$ lies the mass eigenstate $\psi_m^i$ corresponding to the vanishing eigenvalue and $2nf - 1$ other eigenstates $\omega_i^j$, that we call spurious [7,8], and which correspond to non-vanishing eigenvalues $\lambda_j(z_i), j \neq i$. They just represent off-mass-shell states. The case of two flavors ($nf = 1$) is depicted on Fig. 1.

Mixing matrices link flavour ($\Psi_f$) to mass ($\Psi_m$) eigenstates: simplifying to $nf = 2$ (4 flavors)

$$\Psi_f = J\Psi_m, \quad \Psi_f = \begin{pmatrix} \nu_e, f \\ \nu_\mu, f \\ e^- f \\ \mu^- f \end{pmatrix}, \quad \Psi_m = \begin{pmatrix} \nu_e, m \\ \nu_\mu, m \\ e^- m \\ \mu^- m \end{pmatrix}, \quad J = \begin{pmatrix} K_\nu \\ K_\ell \end{pmatrix}, \quad (5)$$

where we have split every $2nf \times 2nf$ matrix into four $nf \times nf$ sub-blocks. The entries of $\Psi_m$ are the $\psi_m^i$’s of (2) and (4).

The (renormalized) quadratic Lagrangian density is $L^{(2)}(\chi) = \bar{\Psi}_f L^{(2)}(\chi) \Psi_f$.

Fermions are usually considered as bi-spinors (of Dirac or Majorana types) built from two Weyl spinors of different chiralities which are also orthogonal to each other. Two sets of different mixing matrices therefore generally occur, respectively for left and right spinors. $K_\nu$ and $K_\ell$ in (5) must then be also attributed a subscript $L$ or $R$ depending on which chirality is considered. In order not to overload the notations, these subscripts will be understood in the following.
2.2. Mixing matrices for mass-split on-shell fermions are not unitary

As already mentioned in [7,8], the connection between flavour eigenstates and non-degenerate mass eigenstates is not a unitary transformation. Indeed:

in flavour space, \( L^{(2)}(z) \) being, at each \( z \), a hermitian \( 2n_f \times 2n_f \) operator (matrix), its \( 2n_f \) eigenstates form an orthonormal basis \( \Psi(z) \) (because it is in particular normal, left and right eigenstates coincide). At \( z = z_i = m_i^2 \), \( \langle \phi_i^m | \phi_i^m \rangle = 1 \) and \( \langle \phi_i^m | \phi_j^m, j \neq i \rangle = 0 \). Thus, at the \( 2n_f \) values \( z = z_i \), \( 2n_f \) different orthonormal bases (of \( 2n_f \) eigenstates) occur. Since two non-degenerate mass eigenstates \( \phi_i^m \) and \( \phi_m^k \) belong to two different orthonormal bases, they are in general not orthogonal:

\[
\langle \phi_i^m | \phi_m^k \rangle \neq 0, \quad i \neq k.
\]

This being true in the neutral and charged sectors, both \( K_v \) and \( K_{\ell} \), which connect the flavour basis to a non-orthonormal one, have no reasons to be unitary:

\[
K_v^T K_v \neq 1, \quad K_{\ell}^T K_{\ell} \neq 1, \quad \text{q.e.d.}
\]

The non-unitarity of mixing matrices does not however jeopardize the unitarity of the theory.\textsuperscript{4} It simply states that, at a given \( q^2 \), all physical states cannot be simultaneously on-shell when they are non-degenerate.

It may happen, for example to describe unstable particles (like neutral kaons), that one is led to introduce an (effective) Hamiltonian, or Lagrangian, which is non-hermitian, and even non-normal. Then, at each \( z \), the set of eigenstates \( \psi^i(z) \) do not form any more an orthonormal basis. Spurious states still accompanied the mass eigenstate at \( z = z_i \). Different mass eigenstates, corresponding to different \( z_i \)'s, have no reason either in this case to form an orthonormal basis, as explicitly checked in [7].

The simplest case of two flavours (\( n_f = 1 \)) is depicted on Fig. 1 which represents either the neutral kaon system, or, in the cases of two lepton families, the neutrino sector or the charged lepton sector. The \( z \)-independent flavour basis, \( (\psi_1, \psi_2) \) for example (\( K^0, \bar{K}^0 \)) for neutral kaons) has been represented by the two horizontal lower lines.

The two eigenstates of the hermitian (normal) renormalized propagator \( \psi^1(z) \) and \( \psi^2(z) \) form a \( z \)-dependent orthonormal basis. When \( z \) varies, this builds up an infinite set of orthonormal bases which is depicted by the two (parallel) curved lines. At a given \( z \), the orthonormal basis \( (\psi^1(z), \psi^2(z)) \) is connected to the orthonormal flavour basis by a unitary mixing matrix with angle \( \theta(z) \). At \( z = z_1 \), \( \varphi^1_{m1} \) and \( \omega^1_{i1} \) form an orthonormal basis, and so do, at \( z = z_2 \), \( \varphi^2_{m1} \) and \( \omega^2_{i1} \). They are respectively related to the basis of flavour eigenstates by two different unitary matrices, with respective angles \( \theta_1 \) and \( \theta_2 \). \( \varphi^1_{m1} \) and \( \omega^1_{i1} \) do not form in general an orthonormal basis, and it intuitively appears on the picture that the mixing matrix connecting them to the flavour basis cannot be parametrized with a single angle (both angles \( \theta_1 \) and \( \theta_2 \) obviously play a role).

\[
\text{2.3. The case of quarks}
\]

(7) applies to states for which the full propagator has poles, corresponding to physical (“on-shell”) propagating states which can be identified with particles. In contrast, quarks are never produced on shell; the poles of their full propagator are ill-defined and so are accordingly their “physical” masses and mass splittings. The only unambiguous orthonormal basis which then occurs in \( L^{(2)} \) (supposed to be hermitian) is the \( z \)-dependent basis \( \psi^i(z) \). At each \( z \) are associated two unitary, \( z \)-dependent mixing matrices \( K_v(z) \) and \( K_{\ell}(z) \). Their unitary product \( K(z) = K_v^i(z)K_{\ell}(z) \) we propose to consider as the equivalent of the renormalized unitary Cabibbo–Kobayashi–Maskawa (CKM) matrix of the standard approach, in which the complex mass matrices \( M_u \) and \( M_d \) generated by the couplings of quarks to the Higgs boson are diagonalized by bi-unitary transformations.

3. Leptonic weak currents

3.1. Fermion coupling to weak gauge bosons

In the flavour basis \( \Psi_f \), the weak Lagrangian reads

\[
\mathcal{L}_{\text{weak}} = \bar{\psi} f \gamma^\mu \frac{1 - \gamma^5}{2} \left[ W^+_\mu T^+ + W^-_\mu T^- + W^{3}_\mu T^3 \right] \psi_f,
\]

\[
T^+ = \left( \begin{array}{cc} 1 & 0 \\ 0 & -1 \end{array} \right), \quad T^- = \left( \begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array} \right), \quad T^3 = \left( \begin{array}{cc} 1 & 0 \\ 0 & -1 \end{array} \right).
\]

\[
(8)
\]

\( T^\pm, T^3 \) form a representation of the \( SU(2) \) group of weak interactions: \( [T^+, T^-] = T^3 \), etc. In the orthonormal basis \( \Psi(z) \) one finds another \( SU(2) \) representation \( [\hat{T}^+, \hat{T}^-] = \hat{T}^3 \):

\[
\mathcal{L}_{\text{weak}} = \bar{\psi}(z) y^\mu \frac{1 - \gamma^5}{2} \left[ W^+_\mu \hat{T}^+ + W^-_\mu \hat{T}^- + W^{3}_\mu \hat{T}^3(z) \right] \psi(z),
\]

\[
\hat{T}^+ = \left( K_v^\dagger(z) K_{\ell}(z) \right), \quad \hat{T}^- = \left( K_v^\dagger(z) K_v(z) \right),
\]

\[
\hat{T}^3 = \left( \begin{array}{cc} 1 & 0 \\ 0 & -1 \end{array} \right).
\]

\[
(9)
\]

In the non-orthonormal basis \( \Psi_m \) of mass eigenstates, \( 8 \) becomes

\[
\mathcal{L}_{\text{weak}} = \bar{\psi}_m y^\mu \frac{1 - \gamma^5}{2} \left[ W^+_\mu \Sigma^+ + W^-_\mu \Sigma^- + W^{3}_\mu \Sigma^3 \right] \psi_m,
\]

\[
\Sigma^i = J^i T^i J,
\]

\[
\Sigma^+ = \left( \begin{array}{cc} K_v^\dagger K_{\ell} & 0 \\ 0 & K_v^\dagger K_v \end{array} \right), \quad \Sigma^- = \left( \begin{array}{cc} K_v^\dagger K_v & 0 \\ 0 & -K_v^\dagger K_{\ell} \end{array} \right),
\]

\[
\Sigma^3 = \left( \begin{array}{cc} K_v^\dagger K_{\ell} & 0 \\ 0 & -K_v^\dagger K_v \end{array} \right).
\]

\[
(10)
\]

and, because of \( 7 \), the \( SU(2) \) commutation relations are not systematically satisfied by the \( \Sigma \)'s (and \( \Psi_m \) does not simply decompose into two \( SU(2) \) doublets).

\textsuperscript{4} This has been explicitly checked in the case of neutral kaons.
$K_v$ is related to $K_v(z_1)$ and $K_v(z_2)$ by

$$K_v = \frac{1}{K_v[22](z_1)K_v[11](z_2) - K_v[12](z_1)K_v[21](z_2)}$$

$$\times \left( \begin{array}{cc} D_v(z_1)K_v[11](z_2) & D_v(z_2)K_v[12](z_1) \\ D_v(z_1)K_v[21](z_2) & D_v(z_2)K_v[22](z_1) \end{array} \right),$$

$D_v(z) = \det K_v(z), \quad z_1 = m^2_{\nu_{\text{ren}}}, \quad z_2 = m^2_{\nu_{\text{ren}}}.$ (11)

The following exact relations

$$\langle \psi_m^1 | \psi_1 \rangle = K_{[11]}(z_1), \quad \langle \psi_m^2 | \psi_1 \rangle = K_{[12]}(z_2),$$

$$\langle \psi_m^1 | \psi_2 \rangle = K_{[21]}(z_1), \quad \langle \psi_m^2 | \psi_2 \rangle = K_{[22]}(z_2),$$

(12) also hold, which become compatible with the approximate formula

$$K_v \simeq \left( \begin{array}{cc} K_v[11](z_1) & K_v[12](z_2) \\ K_v[21](z_1) & K_v[22](z_2) \end{array} \right)$$

(13)

when one neglects the scalar products $\langle \psi_m^1 | \psi_2 \rangle$ supposed to be small. When $z_2 \to z_1$, (12) and (13) give back a unitary mixing matrix $K_v(z_1) = \lim_{z_2 \to z_1} K_v$. This shows the role of the non-degeneracy in the non-unitarity of $K_v$.

3.2. Weak currents

It is remarkable, but often unnoticed that, in the quark sector of the standard model, with unitary $K_u$ and $K_d$, the built-in characteristic of the weak Lagrangian that the couplings of flavour fermions to gauge bosons are symmetric by family exchange of families, translates into a similar property for mass states, without any constraint on the (then unique) mixing angle. The same holds concerning the absence of flavour changing neutral currents.

The situation becomes different for mass-split physical states. Considering the case of two families, when a single mixing angle is not enough to describe the system, the symmetry by exchange of families for mass states (that we call universality for mass states) and the absence of “mass changing neutral currents” (MCNCs) are no longer automatically achieved. They instead require well defined relations between mixing angles.

We demonstrate below that, within each $2 \times 2$ subspace of given electric charge, they constrain the two mixing angles to be either $0, \pi/2 \text{ mod } \pi,$ or $\pi/4 \text{ mod } \pi/2$ ($0 \text{ mod } \pi$ is the trivial case). In the first cases the family indices of flavour and mass states are either identical or crossed. The last case corresponds to the “maximal mixing” (approximately) observed for neutrinos and for neutral kaons. Mass eigenstates are then symmetric or antisymmetric by family exchange (for example so are the PC eigenstates $K^0 \pm K^0$). Maximal mixing accordingly realizes universality in both spaces of flavour and mass eigenstates.

3.2.1. Neutral weak currents of mass eigenstates

We deal with weak currents of mass eigenstates and investigate the property that MCNCs are very small and that their diagonal counterparts are quasi-universal. Allowing a lack of unitarity (7), we parametrize, with transparent notations (preserving a unit norm for all states and discarding irrelevant global phases)

$$K_v = \left( \begin{array}{cc} e^{i\alpha} c_1 & e^{i\beta} s_1 \\ -e^{i\beta} s_2 & e^{i\gamma} c_2 \end{array} \right),$$

$$K_\ell = \left( \begin{array}{cc} e^{i\theta} c_3 & e^{i\epsilon} s_3 \\ -e^{i\epsilon} s_4 & e^{i\phi} c_4 \end{array} \right).$$

(14)

Note that $[K_v, K^\dagger_v] \neq 0$: $K_v$ and $K_\ell$ are not normal. (14) entails

$$K_v \dagger K_v = \left( \begin{array}{cc} c_1^2 + s_2^2 & c_1 s_1 e^{i(\beta-\alpha)} - c_2 s_2 e^{i(\gamma-\beta)} \\ c_1 s_1 e^{-i(\beta-\alpha)} - c_2 s_2 e^{-i(\gamma-\beta)} & c_2^2 + s_1^2 \end{array} \right),$$

$$K_\ell \dagger K_\ell = \left( \begin{array}{cc} c_3^2 + s_4^2 & c_3 s_3 e^{(\epsilon-\zeta)} - c_4 s_4 e^{(\phi-\zeta)} \\ c_3 s_3 e^{-i(\epsilon-\zeta)} - c_4 s_4 e^{-i(\phi-\zeta)} & c_4^2 + s_3^2 \end{array} \right).$$

The two requests of (quasi) universality and absence of MCNCs, equivalent to $K_v \dagger K_v \approx 1 \approx K_\ell \dagger K_\ell$, translate into, respectively, the identity of diagonal elements and the vanishing of non-diagonal elements. The condition $K_v \dagger K_v = 1$, in the simple case where the phases $\alpha, \beta, \gamma, \delta$ are vanishing, can be visualized on Fig. 2, which is drawn in the orthonormal flavour basis:

- the two unit vectors $(c_1, s_1)$ and $(-s_2, c_2)$ are the two dashed vectors.
the mass eigenstates proportional to \((c_2, -s_1)\) and \((s_2, c_1)\) are the two dash-dotted vectors;
• all vectors are uniquely determined by \(\theta_1\) and \(\theta_2\); the condition under scrutiny is that of finding these two angles such that the vectors \((c_1, -s_2)\) and \((s_1, c_2)\), drawn with continuous lines (or, equivalently, the mass eigenstates) become orthonormal.

The discussions for neutrinos and for charged leptons being similar, we proceed with the former.

• Quasi-universality is satisfied for \(c_1^2 + s_2^2 = c_2^2 + s_1^2 \Leftrightarrow c_1^2 = c_2^2\) which requires \(\theta_2 \approx \pm \theta_1 + k\pi\) (a).
• The quasi-absence of MCNCs requires
  1. either: \(c_1 s_1 = c_2 s_2\) \((f_1)\) and \(e^{i(\alpha - \delta)} = e^{i(\beta - \gamma)}\) \((f_2)\),
  2. or: \(c_1 s_1 = -c_2 s_2\) \((g_1)\) and \(e^{i(\alpha - \delta)} = -e^{i(\beta - \gamma)}\) \((g_2)\).

\((f_1)\) requires either \(\theta_2 \approx \theta_1 + n\pi\) \((b)\), or \(\theta_2 \approx -\theta_1 + \pi/2 + n\pi\) \((c)\), while \((g_1)\) requires either \(\theta_2 = -\theta_1 + n\pi\) \((d)\) or \(\theta_2 = \theta_1 + \pi/2 + n\pi\) \((e)\). The different cases to consider are accordingly \((a) \cup (b) \cup (f_2)\), \((a) \cup (c) \cup (f_2)\), \((a) \cup (d) \cup (g_2)\) and \((a) \cup (e) \cup (g_2)\).

The solutions of \((a) \cup (b) \cup (f_2)\) and \((a) \cup (d) \cup (g_2)\) are:

\[
(a) \cup (b) \cup (f_2): \quad \theta_2 = \theta_1 + k\pi \quad \text{or} \\
\begin{cases} 
\theta_1 = (k - n) \frac{\pi}{2}, & \theta_2 = (k + n) \frac{\pi}{2} = -\theta_1 + k\pi \\
\end{cases},
\]

\[
(a) \cup (d) \cup (g_2): \quad \theta_2 = -\theta_1 + k\pi \quad \text{or} \\
\begin{cases} 
\theta_1 = (n - k) \frac{\pi}{2}, & \theta_2 = (n + k) \frac{\pi}{2} = \theta_1 + k\pi \\
\end{cases}.
\]

The solutions of \((a) \cup (c) \cup (f_2)\) and \((a) \cup (e) \cup (g_2)\) are:

\[
(a) \cup (c) \cup (f_2): \quad \begin{cases} 
\theta_1 = \frac{\pi}{4} + (n - k) \frac{\pi}{2}, & \theta_2 = \frac{\pi}{4} + (n + k) \frac{\pi}{2} = \theta_1 + k\pi \\
\end{cases},
\]

\[
(a) \cup (e) \cup (g_2): \quad \begin{cases} 
\theta_1 = -\frac{\pi}{4} + (k - n) \frac{\pi}{2}, & \theta_2 = \frac{\pi}{4} + (k + n) \frac{\pi}{2} = -\theta_1 + k\pi \\
\end{cases}.
\]

There exist accordingly two sets of solutions:

• Cabibbo-like solutions \(\theta_2 = \pm \theta_1 + k\pi\) for which the equations for universality and for the absence of MCNCs coincide: the \((\theta_1, \theta_2)\) surfaces defined by \(c_1^2 = c_2^2\) and \(c_1 s_1 = c_2 s_2\) intersect along a line, which yields a one parameter solution (with a single, unconstrained, mixing angle).

• Cases for which the two equations are independent: the two surfaces (one of which can be checked to always present a saddle point in the vicinity of the intersections) intersect at discrete points.

The set of all solutions is depicted on Fig. 3 (in which the conditions \((f_2)\) or \((g_2)\) are supposed to be realized). It is made of the entire thick continuous and dashed lines + all black and white dots. The thick continuous and dashed lines correspond respectively to \(\theta_2 = \theta_1 + k\pi\) and \(\theta_2 = -\theta_1 + k\pi\) (conditions \((a)\), \((b)\) or \((d)\)). They represent the Cabibbo-like situations. They cross (white dots) at the discrete values \(\pi/2 + k\pi\) or \(k\pi\) of \(\theta_1\) and \(\theta_2\). The thin dotted and continuous lines correspond respectively to \(\theta_2 = \pi/2 + \theta_1 + k\pi\) (conditions \((c)\), \((b)\) or \((d)\)). They represent the Cabibbo-like situations. They cross (white dots) at the discrete values \(\pi/2 + k\pi\) or \(k\pi\) of \(\theta_1\) and \(\theta_2\). Their intersections (black dots) with the thick continuous and dashed lines (on which in particular condition \((a)\) holds) provide the maximal mixing solutions \(\pm \pi/4 + n\pi/2\).

Note that \(\theta_{1,2} \approx 0, \pi\) are allowed for physical non-degenerate particles. The case \(\theta = 0\) means that mass and flavour eigenstates are exactly aligned (which is usually assumed for charged leptons).

Though discrete solutions are also located on thick continuous and dashed lines, they should not be mixed up with
Cabibbo-like solutions: the former are one-parameter solutions, while the latter depend on two parameters.

Both types, when exact, can be shrunk, by rephasing the fermions, to a single mixing angle which is unconstrained for Cabibbo-like cases and has fixed values for others. We give below a Cabibbo-like example.\footnote{Since they lie on the trajectories of Cabibbo-like solutions, this is also a general property of all exact discrete solutions.} For $\theta_2 = \theta_1 + \pi$ and $e^{i(\delta - \omega)} = e^{i(\theta - \beta)} = e^{i\xi}$ (i.e., for conditions (b) $\cup$ (f_2)), or $\theta_2 = -\theta_1$ and $e^{i(\delta - \omega)} = -e^{i(\theta - \beta)} = e^{i\xi}$ (i.e., for conditions (d) $\cup$ (g_2)), one has

$$
\begin{pmatrix}
    e^{-i\nu \psi_{em}} \\
    e^{-i\nu \psi_{num}}
\end{pmatrix} = \begin{pmatrix}
    \cos(-\theta_1) & \sin(-\theta_1) \\
    -\sin(-\theta_1) & \cos(-\theta_1)
\end{pmatrix} \begin{pmatrix}
    \nu_{ef} \\
    -e^{i\xi} \nu_{uf}
\end{pmatrix}.
$$

(17)

So doing, for all exact solutions, $K_v$ becomes unitary.

However, exact solutions are purely academic since, for example, the absence of MCNCs is expected to be only approximate.\footnote{Note that (b) $\cup$ (f_2) or (c) $\cup$ (g_2), which entail the exact absence of MCNC, is enough to have a single mixing angle, since they also entail exact universality (a).} As for exact universality (a), it is by itself not enough to have a unique mixing angle since, in particular, $(f_2)$ or $(g_2)$ may not be satisfied. Moreover, (a) may be only approximately realized. So, in the vicinity of the solutions above, a single mixing angle is not enough to describe the system.

Another characteristic of the discrete solutions is their low sensitivity to small translations in the $(\theta_2, \theta_1)$ plane. If one varies, for example, $\theta_2$, by $\epsilon$ close to a specific point, the l.h.s.’s of the universality condition, $(c_2^4 + s_2^4) - (c_1^2 + s_1^2) = 0$, and of the condition for the absence of MCNCs, $c_{2\,2} \pm i s_{1\,1} = 0$, vary respectively by $-4\epsilon c_2 s_2$ and $\epsilon (c_2^2 - s_2^2)$. Hence:

- at the discrete values $m\pi/2 + n\pi$, the universality condition is satisfied at $O(\epsilon^2)$ while the MCNC condition is only satisfied at $O(\epsilon)$. Referring to Fig. 2, this means that, if one varies by $\epsilon$ the angle (which is then a right angle) between the two dashed vectors, the (right) angle between the continuous vectors (and the one between the mass eigenstates) also vary by $\epsilon$, while their (unit) lengths are only altered at $O(\epsilon^2)$;
- at the “maximal mixing” values, the reverse holds: the absence of MCNCs is specially enforced; by the same variation as above, it is now the angle between the continuous vectors (and the one between the mass eigenstates) which only varies by $\epsilon^2$, while their lengths are altered at $O(\epsilon)$;
- outside the set of discrete solutions, in particular for Cabibbo-like solutions, both variations are instead $O(\epsilon)$.

The absence of MCNCs is thus specially enforced at maximal mixing, while universality is at angles $m\pi/2 + n\pi$.

As seen on Fig. 1, Cabibbo-like systems, characterized by a single unconstrained mixing angle, can only be:

- degenerate particles $z_1 = z_2$ (in which case $\omega_1^2 = \omega_2^2$ and $\omega_1^2 = \omega_2^2$ such that $(\varphi_m^2, \varphi_m^2)$ form an orthonormal basis);
- “off-shell” systems $(\psi^1(z), \psi^2(z))$ evaluated at a common scale $z = q^2$, like quarks, for which the mixing angle is $\theta(z)$;
- very special systems like the ones satisfying Eqs. (81), (82) of [7].

Physical non-degenerate mesonic systems like $K^0 - \bar{K}^0$ correspond to the other category (non-Cabibbo-like): when $CP$ (or exact family symmetry) holds, mixing angles are identical and maximum, but when $CP$ is broken, two angles occur, as shown in [7], which are only close to maximum. Such systems are expected to lie inside the small (2-dimensional) areas in the vicinity of the discrete solutions (the extended dots of Fig. 3), and not inside 1-dimensional deformations of exact Cabibbo-like systems, that stay on the thick continuous or dashed lines.

3.2.2. Charged weak currents of mass eigenstates. Short comments on oscillations

Charged weak currents are coupled through $K_{\nu} K_{\ell}$, the so-called PMNS matrix [9]. Since charged leptons are non-degenerate coupled fermions too, we expect, like previously obtained for neutrinos, the occurrence of a discrete set of mixing angles $\pi/4 \mod \pi/2$ and $\pi/2 \mod \pi$. $K_{\nu}$, like $K_{\ell}$, lies accordingly close to one of the “academic” unitary matrices evoked above, such that $K_{\nu}^\dagger K_{\ell}$ should also be close to a unitary matrix with a mixing angle in the same set of discrete values.\footnote{After convenient rephasing of the fermions (see (17)), both $K_{\ell}$ and $K_{\nu}$ become close to unitary matrices with respective mixing angles $\theta_2$ and $\theta_4$: the PMNS matrix is then close to a unitary matrix with angle $(\theta_2 - \theta_4)$.} Several cases arise, the relevance of which with respect to oscillations we would like to briefly discuss:

- if one among $K_{\nu}$ and $K_{\ell}$ is close to “maximal” and the other close to a multiple of $\pi/2$, the PMNS matrix is close to “maximal”;
- if both $K_{\ell}$ and $K_{\nu}$ are close to “maximal” with respective mixing angles $(2k + 1)\pi/4$ and $(2n + 1)\pi/4$, the PMNS matrix is close to a matrix with mixing angle $(k - n)\pi/2$; this includes the diagonal unit matrix (up to an irrelevant sign) and the antidiagonal unit matrix;
- if both mixing angles of $K_{\ell}$ and $K_{\nu}$ are close to a multiple of $\pi/2$, the same result holds.

Let us first stress that, while neutrino oscillations are determined by $K_{\nu}$ alone, the detection of neutrinos on earth always goes through their coupling to charged leptons, which involves the PMNS matrix. We will consider two configurations for the latter which both seem able to reproduce the observed solar electron neutrino deficit on earth.

“Measuring” a PMNS matrix close to maximal for two generations favors the first possibility. One among the two sets $(\nu_e, \nu_\mu)$ and $(e^-, \mu^-)$ of leptons has then a maximal mixing, while the mixing angle of the second is a multiple of $\pi/2$ (in which case only simple mass-flavour alignment or nearly perfect “crossing” can occur). The following picture may then be conceived. Let us suppose that the flux of neutrinos stays unperturbed during its travel from the center of the sun to the surface of the earth, where it is detected. This can for example happen if the Mikheyev–Smirnov–Wolfenstein (MSW) effect [11]
does not operate inside the sun and if, then, vacuum oscillations do not modify the neutrino spectrum. Its detection through the charged currents (and, so, through the maximal PMNS matrix) introduces a coefficient \( \approx \pm 1/\sqrt{2} \), which yields a factor 1/2 in the square of the corresponding amplitude. A 1/2 “deficit” occurs though, in reality, no oscillation took place.

Mass-flavour alignment for one fermion species, which is one of the two alternatives leading to this first possibility, rules out the corresponding oscillations. It is natural to assume this property for charged leptons (as usually done), since such oscillations cannot anyhow be observed as soon as one measures their energy with a precision much higher than their mass-splitting [10]. The emerging picture may appear coherent, though the asymmetry arising between the two species of leptons raises questions concerning the role of the electric charge. Another possible weakness of this point of view lies in the importance acquired by the measuring process through which, furthermore, the determination of the PMNS matrix cannot be truly asserted.

Now, we would like to point out that the following scenario, with a PMNS matrix \( \approx \text{diag}(1, 1) \), is possible as well. This belongs to the second possibility in the original list, and accordingly provides a symmetric treatment of neutral and charged leptons, which both have maximal mixing. In this case, we are led instead to consider that neutrinos do oscillate in their travel from the core of the sun to the earth. So, with respect to what is expected from solar models, a modified flux of \( \nu_{\text{em}} \) reaches the earth, which can for example be altered by a factor \( \approx \pm 1/\sqrt{2} \). These neutrinos then diagonally couple, in the detector, to charged leptons with the coefficient 1 occurring now in the PMNS matrix, such that a global factor 1/2 again occurs in the (amplitude) \( ^2 \). It can rightly be interpreted as “neutrino oscillations”.

This treatment avoids the slightly opportunistic eviction of electron-muon oscillations that occurred before, and which is not mandatory: it indeed undoubtedly leads to potential such oscillations, which can appear problematic, but can be argued away, as already explained, according to [10]. Charged currents now differ from those in the quark sector by a stronger suppression of their off-diagonal components as compared with the ones obtained from the CKM matrix.

The two scenarios just described, which differ, seem nevertheless to lead to the same conclusion, i.e. the observed depletion of electronic solar neutrinos on earth. Discriminating between them (and others?) needs a more careful investigation which lies beyond the scope of this work.

As for the third possibility, it can easily be shown never to lead to any neutrino deficit. Thus, the maximal character of \( K^0 \), which is common to the first two cases, appears as the essential ingredient for the occurrence of this phenomenon.

4. Neutral mesons

Neutral kaons are composite states, and any Lagrangian that limits to their description can only be effective. A similar propagator formalism can nevertheless be applied [7,8].

The second order electroweak transitions, which couple \( K^0 \) to \( K^0 \), are family changing transitions in which \( d \) and \( s \) quarks get swapped. When \( CP \) is conserved, the \( K^0_1 \) and \( K^0_2 \) mass eigenstates, respectively symmetric and antisymmetric with respect to \( d \leftrightarrow s \) family exchange, correspond to exact maximal mixing. They form an orthonormal basis and no spurious state occurs. \( CP \) violation alters this situation: the \( CP \) violating parameters \( \epsilon_L \) and \( \epsilon_S \) for \( K_L \) and \( K_S \) mass eigenstates slightly differ due to their mass splitting and the Hamiltonian is no longer normal. Inside each in or out space mass eigenstates no longer form orthonormal basis while in and out mass eigenstates, which differ, form a bi-orthogonal basis.

The striking similarity between the latter and neutrinos suggests that the symmetry by exchange of families (universality) plays an important role in the nature of physical states.

Composite states (mesons) are however more complex that fundamental particles. Indeed, while the underlying electroweak theory for quarks does satisfy the criteria of universality and absence of flavour/mass changing neutral currents, the corresponding two types of conditions are not directly available in an effective theory for neutral kaons alone. Whether or not they could be implemented in a larger frame of an effective theory for all scalar and pseudoscalar mesons, in which a general mass matrix in flavour space should be diagonalized (see for example [12]), is a forthcoming matter of investigation.

5. Conclusion

In this short Letter, we have proposed an enlargement of the mixing scheme between mass and flavour eigenstates, which incorporates the peculiarity of both neutrinos and neutral kaons that their mixing angles are close to maximal. In continuation of [7,8] we have shown that, in quantum field theory, the mixing matrices of on-shell coupled mass-split fermions should not be parametrized as unitary. The physics of two massive neutrinos is then not that of a single mixing angle, but of two. A new family of discrete mixing angles then springs out, among which lies the quasi-maximal mixing observed for neutrinos and neutral kaons. When two different mixing angles are concerned, the naïve \( \theta_2 \rightarrow \theta_1 \) (“Cabibbo”) limit does not exist, which explains how discrete solutions can easily be overlooked.

The role of family exchange symmetry has been emphasized. The generalization of this simple exercise to more than two flavours will be the subject of a subsequent work. Other aspects of coupled fermionic systems will also appear in [13].

Acknowledgements

Comments and suggestions by O. Babelon, A. Djouadi, O. Lychkovskiy, V.A. Novikov, L.B. Okun, M.I. Vysotsky and J.B. Zuber are gratefully acknowledged.

References

   (1967) 1717;
   L. Wolfenstein, Phys. Rev. D 20 (1979) 2634;
   Phys. 42 (1985) 913;
   field theory, in preparation.
The asymptotic behavior of Casimir force in the presence of compactified universal extra dimensions

Hongbo Cheng

Department of Physics, East China University of Science and Technology, Shanghai 200237, China

Received 6 September 2006; received in revised form 19 October 2006; accepted 28 October 2006
Available online 3 November 2006

Abstract

The Casimir effect for parallel plates in the presence of compactified universal extra dimensions within the frame of Kaluza–Klein theory is analyzed. Having regularized and discussed the expressions of Casimir force in the limit, we show that the nature of Casimir force is repulsive if the distance between the plates is large enough and the higher-dimensional spacetime is, the greater the value of repulsive Casimir force between plates is. The repulsive nature of the force is not consistent with the experimental phenomena.

© 2006 Elsevier B.V. All rights reserved.

PACS: 11.10.Kk; 04.62.+v

Unifying the interactions in nature needs a powerful ingredient like the model of higher-dimensional spacetime. Nearly 80 years ago the idea that our universe has more than four dimensions was put forward by Kaluza and Klein [1,2]. In this theory named Kaluza–Klein theory, one extra dimension in our Universe was introduced to be compactified in order to unify gravity and classical electrodynamics. Recently the quantum gravity such as string theories or brane-world scenario is developed to reconcile the quantum mechanics and gravity with the help of introducing seven extra spatial dimensions. In Randall–Sundrum model the matter fields may be localized on a four-dimensional brane considered as our real universe, and only gravitons can propagate in the extra space transverse to the brane [3,4]. In some approaches larger extra dimensions were also invoked for providing a breakthrough of hierarchy problem [5–7]. The order of the compactification scale of the extra dimensions has not been confirmed and are also of considerable interest recently. In a word, studies of higher-dimensional spacetime have therefore been pursued vigorously and extensively and more achievements have been made.

The Casimir effect as a fundamental aspect of quantum field theory in confined geometries and the physical manifestation of zero-point energy has received great attention and has been extensively studied in a wide variety of topics [8–18]. The topics include the influence from the effect on the stability of radion in the Randall–Sundrum model, the cosmological aspects like the cosmological constant and the primordial cosmic inflation [19,20]. The effect was also explored in the context of string theory [21–24]. The precision of the measurement has been greatly improved practically [25–28], leading the Casimir effect to be a remarkable observable and trustworthy consequence of the existence of quantum fluctuations. The experimental results clearly show that the attractive Casimir force between the parallel plates vanishes when the plates move apart from each other to the very distant place. In particular it must be pointed out that no repulsive force appears. Therefore the Casimir effect can become a powerful tool for the study and development of a large class of topics on the model of Universe with more than four dimensions.

Exploring the possible existence or size of extra dimensions by means of Casimir effect attracts more attentions of the physical community. The electromagnetic Casimir effect for parallel plates in higher-dimensional spacetime without compactified universal extra dimensions has been studied and some important results were obtained [29,30]. In that case there are divergences in the Casimir energy at the boundaries unless a careful subtraction was performed. Here we focus on the topic...
in the frame of Kaluza–Klein theory. In this approach research on the Casimir effect in five-dimensional spacetimes is just the first step of generalization to investigate the higher-dimensional spacetimes. Having examined the Casimir effect for the rectangular cavity in the presence of a compactified universal extra dimension, we show analytically that the extra-dimension corrections to the standard Casimir effect are very manifest [31]. The Casimir effect for parallel plates in the spacetime with one extra compactified dimension was discussed. Only when the plates gap is very small, the size of the additional dimension satisfying $L \lesssim 10$ nm was obtained by comparison to experimental data [32]. We also scrutinized the same problem and show rigorously that there must appear repulsive Casimir force between the parallel plates within the experimental reach when the plates distance is large enough in the spacetime with one compactified additional dimension [33]. Therefore the results obtained from the Kaluza–Klein theory including only one compactified spatial extra dimension are not consistent with the experimental results mentioned above, which means that the model that the spacetime with only one extra dimension cannot be realistic.

As mentioned above a lot of models such as the string theories motivate the models with more than five dimensions, suggesting it is necessary to continue exploring the Casimir effect in the presence of more compactified universal extra dimensions in detail in order to know whether the models of spacetime with more than one additional dimensions are realistic. This problem, to our knowledge, has not been discussed. For simplicity and comparison to the measurement the system consisting of two parallel plates is always chosen. The purpose of this Letter is to reexamine the Casimir effect for parallel plates in the universe with $d$ compactified spatial dimensions carefully. We regularize the total energy to obtain the Casimir energy, and then Casimir force. In particular we focus on the asymptotic behaviour of the Casimir force between plates for their large enough gap and the dependence of dimensionality of the spacetime in order to compare our results with the measuring evidence listed above directly. Finally the conclusions are emphasized.

In the Kaluza–Klein theory we start to consider the scalar field in the system consisting of two parallel plates in the spacetime with $d$ extra compactified dimensions. Along the extra dimensions the wave vectors of the field have the form $k_i = \frac{n_i \pi}{L}$, $i = 1, 2, \ldots, d$, respectively, $n_i$ an integer. Here we choose that the extra dimensions possess the same radius as $L$. At the plates the fields satisfy the Dirichlet condition, leading the wave vector in the directions restricted by the plates to be $k_n = \frac{n \pi}{R}$, $n$ a positive integer and $R$ the separation of the plates. Under these conditions, the zero-point fluctuations of the fields can give rise to observable Casimir forces.

In the case of $d$ additional compactified dimensions we find the frequency of the vacuum fluctuations to be

$$\omega_{[n]} = \sqrt{k^2 + \frac{n^2 \pi^2}{R^2} + \sum_{i=1}^{d} \frac{n_i^2}{L^2}},$$

(1)

where

$$k^2 = k_1^2 + k_2^2,$$

(2)

$k_1$ and $k_2$ are the wave vectors in directions of the unbound space coordinates parallel to the plates surface. Here $[n_i]$ represents a short notation of $n_1, n_2, \ldots, n_d$, $n_i$ a nonnegative integer. Following Refs. [9–16], therefore the total energy density of the fields in the interior of system reads,

$$\varepsilon = \int \frac{d^2k}{(2\pi)^2} \sum_{n=1}^{\infty} \sum_{[n]} \frac{1}{\sqrt{[n]}} \rho_{[n]} n$$

$$= \frac{\pi}{2} \frac{\Gamma(-\frac{3}{2})}{\Gamma(-\frac{1}{2})} \sum_{l=0}^{d-1} \left( \frac{d}{l} \right) E_{d-l+1} \left( \frac{\pi^2}{R^2}, \frac{1}{L^2}, \frac{1}{L^2}, \ldots, \frac{1}{L^2}, \frac{3}{2} \right) + \frac{3}{2} \frac{\Gamma(-\frac{3}{2}) \zeta(-3)}{2 \Gamma(-\frac{1}{2})}$$

(3)

in terms of the Epstein zeta function $E_p(a_1, a_2, \ldots, a_p; s)$ defined as

$$E_p(a_1, a_2, \ldots, a_p; s) = \sum_{[n_j]} \left( \sum_{j=1}^{p} a_j n_j^s \right),$$

(4)

where $[n_j]$ stands for a short notation of $n_1, n_2, \ldots, n_p, n_j$ a positive integer. We regularize Eq. (3) by means of the following result,

$$\Gamma\left(-\frac{3}{2} \right) E_{d-l+1} \left( \frac{\pi^2}{R^2}, \frac{1}{L^2}, \frac{1}{L^2}, \ldots, \frac{1}{L^2}, \frac{3}{2} \right) = \frac{1}{2} \Gamma\left(-\frac{3}{2} \right) E_{d-l} \left( 1, 1, \ldots, 1; \frac{3}{2} \right) \frac{1}{L^3}$$

$$+ \frac{1}{L} \sum_{k=0}^{\infty} \frac{16-k}{k!} \left( \frac{R}{L} \right)^{k-\frac{1}{2}} \prod_{j=1}^{k} \left[ 16 - (2j-1)^2 \right]$$

$$\times \sum_{n_1, n_2, \ldots, n_d} \left( n_1^2 + n_2^2 + \ldots + n_d^2 + 1 \right)^{-2k+3}$$

$$\times \exp \left( -\frac{2R}{L} n_1 (n_2^2 + n_3^2 + \ldots + n_d^2 + 1) \right)$$

$$+ \frac{\Gamma(-2)}{2 \sqrt{\pi}} E_{d-l} \left( 1, 1, \ldots, 1; -2 \right) \frac{R}{L^3}$$

(5)

to obtain the Casimir energy density of parallel plates in the spacetime with $d$ extra compactified spatial dimensions. It is certainly fundamental to investigate the Casimir force for the same system in the same background in order to compare our results with the experimental phenomenon. The Casimir force is given by the derivative of the Casimir energy with respect to the plate distance. Here we focus on the property that the plate distance $R$ approaches to the infinity, then the expression for the Casimir force in the limiting case is defined as

$$f = \frac{\partial \varepsilon}{\partial R},$$

(6)

where

$$\mu = \frac{R}{L},$$

(7)
Eqs. (3) and (5) are substituted into Eq. (6) and let \( \mu \to \infty \), then the last term of Eq. (3) and the first two terms of Eq. (5) vanish, and we obtain the expression of the asymptotic behavior of the Casimir force in the limiting case of extremely large plates separation,

\[
f = \frac{\Gamma(-2) \left( \sum_{i=0}^{d-1} \binom{d}{i} E_{d-l}(1,1, \ldots, 1; -2) \right)}{8 \Gamma(l)} \left( \frac{1}{L^4} \right). \tag{8}
\]

Here the dimensionality of the high-dimensional spacetime is \( D = 4 + d \), \( d \) the number of extra compactified dimensions. The burden and surprisingly difficult calculation is performed to regularize (8) while the formula as follow is applied repeatedly,

\[
\Gamma(-2)E_{d-l}(1,1, \ldots, 1; -2)
\]

\[= -\frac{1}{2} \Gamma(-2)E_{d-l-1}(1,1, \ldots, 1; -2)
\]

\[+ \frac{\sqrt{\pi}}{2} \Gamma\left( -\frac{5}{2} \right) E_{d-l-1}\left( 1,1, \ldots, 1; -\frac{5}{2} \right)
\]

\[+ \pi^{-\frac{1}{2}} \sum_{k=0}^{\infty} \frac{(16\pi)^{-k}}{k!} \prod_{j=1}^{k} [25 - (2j - 1)^2]
\]

\[\times \sum_{n_1, n_2, \ldots, n_{d-l}=1}^{\infty} \frac{1}{n_1^{k-3} (n_1^2 + n_2^2 + \cdots + n_{d-l}^2)^{-\frac{k-2}{2}}}
\]

\[\times \exp\left[ -2\pi n_1 (n_1^2 + n_2^2 + \cdots + n_{d-l}^2) \right] \tag{9}
\]

then the asymptotic value of Casimir force for parallel plates in the spacetime with \( d \) extra dimensions is obtained and depicted in Fig. 1. In the spacetimes with different dimensionality, the asymptotic value of Casimir force for parallel plates in the limiting of extremely large plates distance is definite and non-negative. When the dimensionality is four, the value of Casimir force vanishes in the limiting case. We find that the asymptotic values are positive in the background with more than four dimensions, which means that there must exist repulsive Casimir force as the plates gap is large enough in the presence of additional compactified dimensions no matter how long the size of extra dimensions is. It is interesting that the asymptotic values of Casimir force depend on the dimensionality of spacetime, the higher dimensionality, the greater asymptotic value. We should not neglect that the repulsive Casimir force between parallel plates is excluded in the practice. We must point out that the experiment is always performed on electromagnetic fields that may obey more complicated boundary conditions than the case of scalar field we consider here, and the asymptotic value of Casimir force for different kinds of fields satisfying different boundary conditions will certainly be different, but all of the asymptotic values keep positive, which means that the repulsive Casimir force must appear as the plates are sufficiently far away from each other.

In conclusion, the model of higher-dimensional spacetime described by standard Kaluza–Klein theory cannot be realistic. Having discussed the Casimir force between parallel plates in the frame of Kaluza–Klein approach in detail, we discover that the values of Casimir force always remain positive as the plates move apart from each other to farther enough, which means that there must exist the repulsive Casimir force when the separation is sufficiently large. The experimental evidence confirm that no repulsive Casimir force appear in this case. Although we are limited here by only the asymptotic case for simplicity and comparison, it is enough for us to declare that the results obtained from the standard Kaluza–Klein theory including extra compactified dimensions disagree with the experimental evidence inevitably, leading a reliable negative verdict on the theory. It must be pointed out that in the higher-dimensional spacetime in which all of the extra dimensions are not compactified the electromagnetic Casimir energy density for parallel plates is divergent at the plates, however, a careful subtraction of the divergent self-energy of the plates is needed to restore consistency with the force [29,30]. Our conclusions above are drawn in the case that all of extra dimensions are compactified. It is interesting that for the same system consisting of two parallel plates the different interesting results are obtained in frames of different theories. The higher-dimensional spacetime theories including standard Kaluza–Klein theory needs to be developed further and related topics also need further research.

**Acknowledgements**

The author thank Professor I. Brevik and Professor E. Elizalde for helpful discussions. This work is supported by the Shanghai Municipal Science and Technology Commission No. 04dz05905.

**References**


Constraints on a variable dark energy model with recent observations

Puxun Wu\textsuperscript{a,b}, Hongwei Yu\textsuperscript{a,\ast}

\textsuperscript{a} Department of Physics and Institute of Physics, Hunan Normal University, Changsha, Hunan 410081, China
\textsuperscript{b} School of Sciences and Institute of Physics, Central South University of Forestry and Technology, Changsha, Hunan 410004, China

Received 25 September 2006; accepted 7 November 2006

Editor: M. Cveti\v{c}

Abstract

We place, by the maximum likelihood method, constraints on a variable dark energy model with the equation of state $w = w_0/[1 + b \ln(1 + z)]^2$ using some recent observational data, including the new Sne Ia data from the SNLS, the size of baryonic acoustic oscillation peak from SDSS and the CMB data from WMAP3. We find that the SNLS data favor models with $w_0$ around $-1$, in contrast to the Gold data set which favors a more negative $w_0$. By combining these three databases, we obtain that $\Omega_m = 0.27^{+0.036}_{-0.038}$, $w_0 = -1.11^{+0.21}_{-0.30}$ and $b = 0.31^{+0.71}_{-0.31}$ with $\chi^2 = 110.4$ at the 95% confidence level. Our result suggests that a varying dark energy model and a crossing of the $w = -1$ line are favored, and the present value of the equation of state of dark energy is very likely less than $-1$.

\copyright 2006 Elsevier B.V. All rights reserved.

PACS: 98.80.-k; 98.80.Es

1. Introduction

Since the Hubble diagram of Type Ia Supernovae (Sne Ia)\textsuperscript{[1]} first indicated that the universe is undergoing an accelerating expansion, many works have been done, trying to explain this phenomenon. A large number of models for the cosmic acceleration are proposed by assuming the existence of an energy component with negative pressure in the universe, named dark energy, which at late times dominates the total energy density of the universe and drives its acceleration of expansion. The simplest candidate of dark energy is the cosmological constant $\lambda$\textsuperscript{[2]}, with the equation of state $w = p/\rho = -1$. However, the cosmological constant has to be extremely fine-tuned in order to induce the observed cosmic expansion, and this has led many authors to use scalar fields, such as quintessence\textsuperscript{[3]}, phantom\textsuperscript{[4]}, quintom\textsuperscript{[5]} and Chaplygin gas\textsuperscript{[6]}, as alternative models for dark energy. On the other hand there are also some other models where the observed cosmic acceleration is not driven by dark energy, such as modified gravity\textsuperscript{[7]}, theories with compactified extra dimensions\textsuperscript{[8]}, DGP model\textsuperscript{[9]} and Cardassian cosmology\textsuperscript{[10]}, etc.

In order to determine the recent expansion history of our universe, one can also use a different approach, that is, to assume an arbitrary parametrization for the equation of state $w(z)$ for dark energy, where $z$ is the redshift. The parametrization may not be motivated by any particular fundamental physical theory and is thus “model-independent”. It however needs to be designed to give a good fit to the observational data. The simplest parametrization is $w = \text{const}$. Some other proposals, including $w(z) = w_0 + w_1 z$\textsuperscript{[11]}, $w(z) = w_0 + w_1 z/(1 + z)$\textsuperscript{[12]}, etc., are also made. Recently Wetterich\textsuperscript{[13]} proposed an interesting phenomenological parametrization for a variable dark energy, in which the effective equation of state is expressed as:

$$w(z) = \frac{w_0}{[1 + b \ln(1 + z)]^2},$$

where $w_0$ represents the present value of the equation of state and $b$ is a positive constant characterizing the change of $w(z)$ with redshift. Apparently with this $w(z)$ there are three model parameters ($\Omega_m, w_0, b$) to be determined by the observations, where $\Omega_m = \rho_m/\rho_c$ represents the present matter density parameter and $\rho_c = 3H_0^2/8\pi G$ is the present critical density of our
universe. The advantage of this parametrization over other parameterizations of a time varying equation of state, $w(z)$ is that it covers the whole available redshift range while other parameterizations proposed before cover only a restricted range of redshift [11,12,14]. Later, using Cosmic Microwave Background, Large Scale Structure, and SNe Ia data, Doran, Karwan and Wettermich [15] discussed the $w_0$ and the dark energy fraction at very high redshift $\Omega_d^0$ in this model, and found that the 95% confidence level $w_0 < -0.8$ and $\Omega_d^0 < 0.03$, where $\Omega_d^0$ and $w_0$ are related to $b$ by $b = -3w_0(\ln \frac{1 - \Omega_d^0}{\Omega_d^0} + \ln \frac{1 - \Omega_d^0}{\Omega_d^0} - 1)$ with $\Omega_d^0$ being the present energy density of dark energy. Movahed and Rahvar [16] have used the Gold Sna Ia data [18], the position of first acoustic peak of the Cosmic Microwave Background radiation (CMB) and the size of baryonic acoustic oscillations peak to constrain this model and obtained at the 2σ confidence level $\Omega_m = 0.27^{+0.04}_{-0.03}$, $w_0 = -1.45^{+0.65}_{-2.17}$ and $b = 1.35^{+6.30}_{-1.35}$.

Recently Astier et al. [19] released the data of high redshift supernovae from the Supernova Legacy Survey (SNLS). In this survey the systematic uncertainties and systematic errors are reduced. It is worth noting that the SNLS data set is a better agreement with the WMAP data compared to the Gold Sna Ia set [17]. Thus, in this Letter we will reexamine this variable dark energy by using the 115 new SNLS SNe Ia data, the size of baryonic acoustic oscillations peak detected in the large-scale correlation function of luminous red galaxies from Sloan Digital Sky Survey (SDSS) [20] and the CMB data obtained from the three-year WMAP result [21]. We obtain at a 95% confidence level $\Omega_m = 0.27^{+0.036}_{-0.038}$, $w_0 = -1.14^{+0.21}_{-0.36}$ and $b = 0.31^{+0.71}_{-0.31}$.

2. The basic equation

By using Eq. (1) and the equation of energy conservation, it is easy to obtain the evolution of density of dark energy [13,16]

$$\rho_d = \rho_{d0}(1 + z)^{3(1 + \bar{w}(z))},$$  

(2)

where $\rho_{d0}$ denotes the present density of dark energy and $\bar{w}(z) = \frac{w_0}{1 + \Omega_m(1 + z)}$. Since WMAP observations strongly indicate that the geometry of our universe is spatially flat [22], we will ignore the term containing curvature factor in Friedman equation. If the radiation components in universe are further ignored, we find for the Hubble parameter

$$H^2(z; \Omega_m, w_0) = H_0^2[\Omega_m(1 + z)^3 + (1 - \Omega_m)(1 + z)^{3(1 + \bar{w}(z))}],$$  

(3)

where $H_0$ is the present Hubble constant. Meanwhile one can show that for a flat universe the luminosity distance, $d_L$, can be expressed as

$$d_L(z; H_0, \Omega_m, w_0) = \frac{c}{H_0}(1 + z) \int_0^z \frac{dz'}{E(z', \Omega_m, w_0)}.$$  

(4)

Here $c$ is the velocity of light and $E(z, \Omega_m, w_0) = H(z; \Omega_m, w_0)/H_0$.

3. Constraints from SNe Ia, SDSS and CMB data

The 115 new SNe Ia data includes 44 previously published nearby SNe Ia and 71 distant SNe Ia released recently by the Supernova Legacy Survey (SNLS) [19] which is a planned five year survey of SNe Ia with $z < 1$. Constraints from SNe Ia can be obtained by fitting the distance modulus $\mu(z)$

$$\mu(z) = 5 \log_{10}[D_L(z)] + \mathcal{M}.$$  

(5)

Here $D_L = H_0d_L$ and $\mathcal{M} = M - 5 \log_{10}(H_0)$, $M$ being the absolute magnitude of the object.

Recently Eisenstein et al. [20] successfully found the size of baryonic acoustic oscillation peak using a large spectroscopic sample of luminous red galaxy from the SDSS and obtained a parameter $A$, which is independent of dark energy models and for a flat universe can be expressed as

$$A = \sqrt{\Omega_m} \int_{z_1}^{z_f} \frac{dz}{E(z)},$$  

(6)

while the Gold set favors a more negative $A$.

For the CMB data, the shift parameter $R$ can be used to constrain the dark energy models and it can be expressed as [23]

$$R = \sqrt{\Omega_m} \int_0^{z_r} \frac{dz}{E(z)},$$  

(7)

for a flat universe, where $z_r = 1089$. From the three-year WMAP result [22], the shift parameter is constrained to be $R = 1.70 \pm 0.03$ [21].

In order to place limits on model parameters ($\Omega_m, w_0, b$) with the observation data, we make use of the maximum likelihood method, that is, the best fit values for these parameters can be determined by minimizing

$$\chi^2 = \sum_i \left[ \frac{[\mu_{\text{obs}}(z_i) - \mu(z_i)]^2}{\sigma^2} \right] + \frac{(A - 0.469)^2}{0.017^2} + \frac{(R - 1.70)^2}{0.03^2}. $$  

(8)

For the SNLS SNe Ia data set, at a 95.4% confidence level we obtain $\Omega_m = 0.27^{+0.25}_{-0.27}$, $w_0 = -1.03^{+0.46}_{-1.42}$ and $b = 0.0^{+4.3}_{-3.29}$. These are different from the result obtained in Ref. [16] using 157 Gold Sna Ia data, where $\Omega_m = 0.01^{+0.51}_{-0.01}$, $w_0 = -1.96^{+0.75}_{3.29}$ and $b = 6.00^{+7.35}_{6.00}$. Apparently the SNLS data set favors models with $w_0$ around $-1$ while the Gold set favors a more negative $w_0$. Meanwhile the best fit value of $b( = 0)$ for SNLS data indicates that a non-varying dark energy model is favored. However, if combining the SNLS SNe Ia, SDSS and CMB, we find that at a 95% confidence level $\Omega_m = 0.27^{+0.036}_{-0.038}$, $w_0 = -1.11^{+0.21}_{-0.30}$ and $b = 0.31^{+0.71}_{-0.31}$ with $\chi^2 = 110.4$. The best fit values show that a variable dark energy model is favored since $b$ is nonzero and it is very likely the present value of the equation of state is less than $-1$. In Fig. 1 the Hubble diagram for 115 SNLS SNe Ia data set is shown with the
Fig. 1. The Hubble diagram for 115 SNLS Sne Ia data with the best fit parameters $(\Omega_m, w_0, b) = (0.27, -1.11, 0.31)$ obtained from the combination of SNLS, SDSS and CMB databases.

Fig. 2. The $1\sigma$, $2\sigma$ and $3\sigma$ confidence contours for $\Omega_m$ and $w_0$ by fixing $b$ at its best fit value 0.31 from the combination of SNLS, SDSS and CMB databases.

Fig. 3. The $1\sigma$, $2\sigma$ and $3\sigma$ confidence contours for $\Omega_m$ and $b$ with $w_0$ at its best fit value $-1.11$ from the combination of Gold, SNLS, SDSS and CMB databases.

Fig. 4. The behavior of $w(z)$. The solid line plots $w(z)$ by using the best fit parameters $(w_0 = -1.11, b = 0.31)$ obtained from the SNLS + SDSS + CMB data and the dotted lines are for $1\sigma$ errors. The solid line shows clearly a varying equation of state parameter.

In this Letter we have placed constraints on a parameterized dark energy model [13] using the new SNLS Sne Ia data sets, the size of baryonic acoustic oscillation peak from SDSS and the shift parameter from the CMB observation. It is found

$$(\Omega_m, w_0, b) = (0.27, -1.11, 0.31).$$

The contour plots of $\Omega_m$ and $w_0$ by fixing $b$ at its best fit value 0.31 are shown in Fig. 2. The contour plots of $\Omega_m$ and $b$ by fixing $w_0$ at its best fit value $-1.11$ are shown in Fig. 3. In Fig. 4 we give the evolutionary curves $w(z)$ vs $z$ with $1\sigma$ error bar based on SNLS Sne Ia + SDSS + CMB data. This figure shows graphically that SNLS Sne Ia + SDSS + CMB data favor a varying dark en-

energy model. Comparing with the results obtained in Ref. [16], we find that in our results at the 95% confidence level stronger constraints on $w_0$ and $b$ are obtained. Meanwhile it is easy to see that we also obtained a stronger constraint on $w_0$ than that obtained in Ref. [15].

4. Conclusion

In this Letter we have placed constraints on a parameterized dark energy model [13] using the new SNLS Sne Ia data sets, the size of baryonic acoustic oscillation peak from SDSS and the shift parameter from the CMB observation. It is found...
318


that a non-varying phantom dark energy model is favored if the
SNLS Sne Ia data set is used in contrast to a varying dark energy when the Gold Sne Ia data is utilized [16], and different
from the case of the Gold set, the model with w0 around −1
is favored by SNLS data at a 95.4% confidence level. Combing three databases (SNLS Sne Ia, SDSS and CMB), we obtain
a constraint on the model parameters (Ωm , w0 , b), which suggests that a varying dark energy model and a crossing of the
w = −1 line are favored [24], and the present value of the equation of state of dark energy is very likely less than −1.
Acknowledgements
We would like to thank Z.H. Zhu for his helpful discussions.
This work was supported in part by the National Natural Science Foundation of China under Grants No. 10375023 and No.
10575035, and the Program for NCET under Grant No. 040784.
References
P.J.E. Peebles, B. Ratra, Rev. Mod. Phys. 75 (2003) 595;
R.R. Caldwell, M. Kamionkowski, N.N. Weinberg, Phys. Rev. Lett. 91
(2003) 071301;
V. Faraoni, Class. Quantum Grav. 22 (2005) 3235;
(2005) 193;
R. Cai, A. Wang, JCAP 0503 (2005) 002;
S.K. Srivastava, hep-th/0605010;
177;

265;
M.C. Bento, O. Bertolami, A.A. Sen, Phys. Rev. D 67 (2003) 063003;
G. Dvali, G. Gabadadze, M. Kolanovic, F. Nitti, Phys. Rev. D 64 (2001)
084004;
024031;
M. Fairbairn, A. Goobar, astro-ph/0511029;
123003.
501;
123520;
[24] For other analyses of observational data that also indicate a crossing of the
w = −1 line, see, for example:
U. Alam, V. Sahni, A.A. Starobinsky, JCAP 0406 (2004) 008;
R. Lazkoz, S. Nesseris, L. Perivolaropoulos, JCAP 0511 (2005) 010;


Transverse single spin asymmetries in inclusive deep-inelastic scattering

A. Metz, M. Schlegel *, K. Goeke

Institut für Theoretische Physik II, Ruhr-Universität Bochum, D-44780 Bochum, Germany

Received 3 November 2006; received in revised form 8 November 2006; accepted 8 November 2006
Available online 16 November 2006
Editor: N. Glover

Abstract

In inclusive deep-inelastic lepton–hadron scattering multi-photon exchange between the leptonic and the hadronic part of the process causes single spin asymmetries. The asymmetries exist for a polarized target as well as a polarized incoming or outgoing lepton, if the polarization vector has a component transverse with respect to the reaction plane. The spin dependent parts of the single polarized cross sections are suppressed like $\alpha_{em} m_{pol}/Q$—where $m_{pol}$ denotes the mass of the polarized particle—compared to the leading terms of the cross section for unpolarized or double-polarized deep-inelastic scattering. Both the target and the beam spin asymmetry are evaluated in the parton model. In the calculation only quark–quark correlators are included. While this approximation turns out to be justified for the lepton spin asymmetries, it is not sufficient for the target asymmetry.

© 2006 Elsevier B.V. All rights reserved.

During the last decades an enormous amount of information on the partonic structure of the nucleon has been extracted from inclusive deep-inelastic lepton–nucleon scattering (DIS, $l(k) + N(P) \rightarrow l(k') + X(P_X)$). It is well known that the cross section for this process is fully described by four independent structure functions, provided that one only considers the electromagnetic interaction. For instance the unpolarized cross section of inclusive DIS is given by

$$k' \frac{d \sigma_{unp}}{d^3 k'} = \frac{4 \alpha_{em}^2}{Q^4} \left( x y F_1(x, Q^2) + \frac{1 - y}{y} F_2(x, Q^2) \right),$$

and contains the two structure functions $F_1$ and $F_2$. In Eq. (1) we make use of the standard DIS variables

$$Q^2 = 2 k \cdot k', \quad x = \frac{Q^2}{2 P \cdot (k - k')}, \quad y = \frac{P \cdot (k - k')}{P \cdot k}.$$  

Neglecting the nucleon mass $M$ the variables in (2) are related by means of $y = Q^2/(xs)$ with $s = 2 P \cdot k$ being the squared cm-energy of the reaction. Two additional structure functions, often denoted by $g_1$ and $g_2$, appear in double-polarized DIS (longitudinal lepton polarization, and longitudinal or transverse target polarization) (see, e.g., Refs. [1,2]).

As long as the typically used one-photon exchange approximation is considered any single spin asymmetry (SSA) is strictly forbidden in inclusive DIS due to parity and time reversal invariance [3]. However, this is no longer true if multi-photon exchange is taken into account. In fact, in inclusive DIS (transverse) SSAs exist if one goes beyond the one-photon exchange approximation [3]. The SSAs arise from a specific correlation between a polarization vector $S$ of a particle as well as the 4-momenta of the nucleon and of the leptons,

$$\varepsilon_{\mu\nu\rho\sigma} S^\mu P^\nu k^\rho k'^\sigma,$$

where

* Corresponding author.

E-mail address: marc.schlegel@rub.de (M. Schlegel).

0370-2693/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.physletb.2006.11.009
where $\varepsilon^{\nu\nu'\rho\sigma}$ is the totally antisymmetric Levi-Civita tensor. One can readily convince oneself that only those components of $S$ contribute to the correlation (3) which are transverse with respect to the reaction plane. (In the target rest frame, e.g., the reaction plane is given by the leptonic plane.) The vector $S$ can represent the polarization of the nucleon but also the polarization of the incoming or outgoing lepton. Since the SSAs turns out to be proportional to the mass of the polarized particle the target SSA should be the most attractive candidate from the experimental point of view. The expression in (3) is a so-called artificial time-reversal or correlation. Artificial time-reversal and ordinary time-reversal differ in the sense that in the former case the initial and final state of a reaction are not interchanged. (For a recent discussion on this issue we refer the reader to [4].) In order to generate a correlation of the type (3) a non-zero phase (imaginary part) on the level of the amplitude of the process is required. Such a phase can be provided by multi-photon exchange between the leptonic and the hadronic part of the reaction. Therefore, there is no argument which forbids the existence of the correlation (3) in inclusive DIS. On the other hand, the SSAs are proportional to the electromagnetic fine structure constant $\alpha $ ($\approx 1/137$) which may lead to relatively small effects. Indeed, the results of early measurements of the transverse target SSA at the Cambridge Electron Accelerator [5] and at the Stanford Linear Accelerator [6] were compatible with zero within the error bars. However, present experiments with their higher precision should be able to observe such effects.

We also note that a lot of work has been devoted to transverse SSAs in processes like one-hadron inclusive production in hadron–hadron collisions, semi-inclusive DIS, and the Drell–Yan process (see, e.g., Refs. [7–19]). In particular over the past 4–5 years this field of research has been considerably growing. On the other hand, for decades no measurement/analysis of a transverse SSA in inclusive DIS has been performed.

In addition we mention that in elastic lepton scattering off the nucleon transverse SSAs were already discussed long ago [20]. (Note also Ref. [21] where the transverse SSA for elastic scattering of two point-like spin–1/2 particles was computed.) On the theoretical side, renewed interest for this observable emerged recently [22–31] because measurements became feasible and non-zero results were observed [32,33]. Since the elastic scattering is the limit of inclusive DIS for $x \to 1$, one certainly can also expect non-vanishing asymmetries in inclusive DIS for arbitrary values of $x$.

In this Letter we compute the transverse SSAs for a polarized incoming lepton and for a polarized nucleon target in inclusive DIS by considering two-photon exchange between the leptonic and the hadronic part of the reaction. The calculation is performed in the framework of the parton model. (An early phenomenological calculation of the transverse target SSA only considered the excitation of the nucleon to the $\Delta$ in the framework of the parton model. (An early phenomenological calculation of the transverse target SSA only considered the DIS by considering two-photon exchange between the leptonic and the hadronic part of the reaction. The calculation is performed

Thus it is worthwhile to mention that two-photon exchange may also be at the origin of the observed large discrepancy between the outcome of two extraction methods—Rosenbluth separation and polarization transfer—for the electric form factor of the proton [35–38]. Moreover, our work here is related to Refs. [23,39,40] in which the two-photon exchange contribution to elastic electron scattering off the nucleon was treated in the parton model.

We start by recalling some elements of the collinear parton model. This approach essentially relies on two ingredients/approximations:

1. A fast moving hadron looks like a bunch of partons moving in the same direction. If, e.g., the nucleon has a large light-cone plus-momentum $p^+ = (P^0 + P^3)/\sqrt{2}$ a parton inside the nucleon has a large plus-momentum $p^+$ as well. The light-cone minus-momentum $p^-$ and the transverse momentum $p_T$ of a given parton are small compared to $p^+$ and are ignored.

2. In the case of a hard process, like inclusive DIS off the nucleon at large $Q^2$, the reaction is computed in the impulse approximation, i.e., one considers the reaction rate for the corresponding process with free partons and sums incoherently over the contributions from the different partons. For DIS this means in particular that the virtual photon interacts with a single free quark, while the remaining partons inside the nucleon merely act as spectators of the reaction.

On the basis of the collinear parton model the structure functions in Eq. (1) are given by

$$F_2(x) = 2xF_1(x) = \sum_q e_q^2 f_q^{u}(x),$$  

(4)

where $f_q^{u}$ represents the ordinary unpolarized distribution of a quark with flavor $q$ in a nucleon. The summation in (4) is running both over quarks and antiquarks, and $e_q$ denotes the quark charge in units of the elementary charge. The plus-momentum of the quark, which is struck by the virtual photon, is specified by means of the relation $p^+ = xP^+$. The field-theoretical definition of the quark distribution reads (see, e.g., Ref. [41])

$$f_1(x) = \int \frac{d\xi}{4\pi} e^{ip\xi} \langle P, S | \bar{\psi}(0) \gamma^+ \gamma^5 \psi(\xi) | P, S \rangle |_{\xi^+ = \xi_T = 0}. $$  

(5)
Here we have used the light-cone gauge in which the Wilson-line, connecting the two quark fields in (5) and ensuring color gauge invariance of the operator, disappears. We also mention that the scale dependence of parton distributions is neglected throughout this work since it is irrelevant for the main point of the discussion.

Now we turn our attention to the calculation of the transverse SSAs in the parton model by discussing in a first step the case of a polarized incoming lepton. To this end we consider the two-photon exchange diagram in Fig. 1 together with its Hermitian conjugate. The so-called crossed box graph, where the lower vertices of the two photons on the lhs of the cut in Fig. 1 are interchanged, does not contribute to the SSA since it cannot provide an imaginary part. The second ingredient of the parton model implies that only such diagrams are taken into account in which both photons couple to the same quark.

The diagram in Fig. 1 provides the following contribution to the squared matrix element of the process,

$$\frac{e^6}{Q^2 i} \int \frac{d^4 l}{(2\pi)^4} \frac{1}{[(l-k)^2 - \lambda^2 + i\epsilon][(l-k')^2 - \lambda^2 + i\epsilon][l^2 + i\epsilon]} L_{\mu\nu\rho} 4\pi W^{\mu\nu\rho},$$

with

$$L^{\mu\nu\rho} = \frac{1}{2} \text{Tr}((\not{k} + m)\gamma_5 \not{y}((\not{k} + m)\gamma^\nu(f + m)\gamma^\rho)),
4\pi W^{\mu\nu\rho} = \sum_q \frac{e_q^3}{Q^2} \frac{f_1^q(x) \text{Tr}(\not{p} \gamma^\mu(\not{p} + \not{k} - \not{k}')\gamma^\nu(\not{p} + \not{k} - f)\gamma^\rho)}{(p + k - l)^2 + i\epsilon}.$$

Note that in Eq. (6) only the term showing up for a transversely polarized incoming lepton is listed. In order to obtain a non-zero asymmetry one has to work with a finite lepton mass $m$. When performing the calculation we ignore a term proportional to $m^3$ in the lepton tensor $L^{\mu\nu\rho}$ and also the mass in the denominator of the lepton propagator in the loop. Both effects are suppressed for large $Q^2$. The quark is treated as massless particle. On the other hand, to avoid a potential IR divergence, a mass $\lambda$ is assigned to the photon.

It turns out that in the collinear parton model only the imaginary part of the loop-integral in (6) survives as soon as one adds the contribution coming from the Hermitian conjugate diagram. This imaginary part can be conveniently evaluated by means of the Cutkosky rules. Here we avoid giving details of the calculation and just quote our final result for the spin dependent part of the single polarized cross section,

$$k^0 \frac{d\sigma_{L,\text{pol}}}{d^3 \vec{k}} = 4\frac{e_m^3}{Q^8} m x y \varepsilon_{\mu\nu\rho\sigma} S^\mu p^\nu k^\rho k'^\sigma \sum_q e_q^3 x f^q_1(x).$$

At this point several comments are in order. The result in Eq. (7) is the leading term in the Bjorken limit ($Q^2 \rightarrow \infty$, $x$ fixed). Corrections to this formula are suppressed at least by a factor $M/Q$. The sign of the spin dependent part of the polarized cross section depends on the charge of the lepton which enters to the third power. The result in (7) holds for a negatively charged lepton. (It is interesting to note that in one of the early measurements of the target SSA [6] there is evidence for the expected sign change when switching from an electron to a positron beam.) We have taken the convention $\varepsilon^{0123} = 1$ for the Levi-Civita tensor. The spin dependent part of the single polarized cross section behaves like $e_m m/Q$ relative to the unpolarized cross section given in Eq. (1) (and relative to the dominant term of the double polarized DIS cross section). In this context note that the correlation (3) showing up in Eq. (7) is given by

$$\varepsilon_{\mu\nu\rho\sigma} S^\mu p^\nu k^\rho k'^\sigma \propto \frac{Q^3}{x y} \sqrt{1 - y},$$

in the Bjorken limit.
We emphasize that the expression in Eq. (7) is IR finite. Terms proportional to \( \ln(Q^2/\lambda^2) \) appearing at intermediate steps of the calculation cancel in the final result. In related studies of transverse SSAs in semi-inclusive processes a comparable cancellation of IR divergent terms has been observed (see, e.g., Refs. [18,42]). Because of its IR finiteness the parton model result (7) probably constitutes a reliable estimate of the leading term (in the Bjorken limit) of the lepton beam SSA. Nevertheless, a word of caution has to be added. At present we have no rigorous proof that other diagrams, not included in the parton model approximation, cannot provide a leading (and separately IR finite) contribution to the transverse beam SSA. In fact, one might in particular question the second ingredient of the parton model according to which both photons only couple to the same quark. When performing the integration upon the loop-momentum \( l \) also photons with an arbitrary long wavelength contribute. Photons with a long wavelength, however, interact with the entire nucleon rather than just a single parton. On the other hand, it is possible that such effects caused by soft photon emission in general cancel when computing the lepton SSA.

In connection with the second lepton SSA (polarized lepton in the final state) it is sufficient to mention that also this observable is IR finite in the collinear parton model. The calculation is basically a copy of the one for the beam SSA. Because the asymmetries are proportional to the mass of the lepton they become quite small for electron scattering. Corresponding measurements of the transverse lepton beam SSA in elastic electron–nucleon scattering show effects of \( \mathcal{O}(10^{-6}–10^{-5}) \) [32,33]. However, in comparison much larger asymmetries can be expected for polarized muon scattering off the nucleon.

Now we proceed in order to discuss in a second step the transverse target SSA. As we will see below, from a theoretical point of view this observable is more challenging than the lepton spin asymmetries. The main reason for this difference is the twist-3 nature of the target asymmetry, whereas the lepton asymmetries, though suppressed like 1/\( Q \), are given by the twist-2 parton density \( f_1 \).

We start again by using the collinear parton model. Also for the target SSA the diagram in Fig. 1 together with its Hermitian conjugate is considered. The calculation proceeds along the lines of the lepton asymmetry, but here we entirely neglect the lepton mass. The result for the spin dependent part of the single polarized cross section is now given by

\[
k^0 \frac{d\sigma_{\text{pol}}}{d^3k'} = \frac{4\alpha^2_{\text{em}} M x^2 y}{Q^2} \bar{\epsilon}_{\mu \nu \rho \sigma} S^\mu P^\nu k'^\rho k'^\sigma \left(1 - y\right)^2 \ln\frac{Q^2}{\lambda^2} + y(2 - y) \ln y + y(1 - y) \sum_q e_q^2 x g_q^T(x).
\]

Like in the case of the lepton SSA we have just kept the leading term in the Bjorken limit. As already mentioned the target SSA is a twist-3 effect which is reflected by the presence of the twist-3 quark distribution \( g_T \) defined through

\[
S^i g_T(x) = \frac{p^+}{M} \int \frac{d^3\tilde{k}}{4\pi} e^{ip^\perp \tilde{k}} \langle P, S | \bar{\psi}(0) \gamma^i \gamma_5 \psi(\tilde{x}) | P, S \rangle |_{\tilde{x}^+ = \tilde{y}^+ = 0},
\]

with \( i \) denoting a transverse index. If one would keep a quark mass then also a term proportional to the transversity distribution of the quark would appear in (9).

The crucial difference between the result in (9) and the lepton asymmetry in (7) is the uncancelled IR divergence as the photon mass \( \lambda \to 0 \). One has to conclude that the collinear parton model is not suitable for describing this observable. As mentioned above it is possible that also diagrams where both photons couple to different quarks in the nucleon have to be taken into account in order to arrive at an IR finite result. In addition, it is known for a long time that even in the one-photon exchange approximation the collinear parton model have to be included (see, e.g., Ref. [43]): First, the transverse momentum of the struck quark cannot be neglected; second, also quark–gluon–quark correlators have to be taken into account. In this note we limit ourselves to the first correction, and leave a detailed investigation of effects due to quark–gluon–quark correlations inside the nucleon for future work.

If one considers \( p_T \)-dependent terms in the quark–quark correlator, which appear in combination with the imaginary part of the electron–quark correlator in Fig. 1, one arrives at the following result for the single polarized cross section,

\[
k^0 \frac{d\sigma_{\text{pol}}}{d^3k'} = \frac{4\alpha^2_{\text{em}} M x^2 y}{Q^2} \bar{\epsilon}_{\mu \nu \rho \sigma} S^\mu P^\nu k'^\rho k'^\sigma \int d^2 \tilde{p}_T \ H(\tilde{p}_T^2) \sum_q e_q^2 \left( x g_q^T(x, \tilde{p}_T^2) - \frac{\tilde{p}_T^2}{2M^2} S^q_{\text{IT}}(x, \tilde{p}_T^2) \right).
\]

In comparison to the result (9) the main new ingredient in (11) is the unintegrated (\( p_T \)-dependent) parton density \( g_{1T} \) which is given by (see, e.g., Ref. [44]),

\[
\tilde{p}_T \cdot \tilde{S}_T = \int \frac{d^3\tilde{x}}{2(2\pi)^3} e^{(p^+ \tilde{x}^- - \tilde{p}_T \cdot \tilde{S}_T)} \langle P, S_T | \bar{\psi}(0) \gamma^+ \gamma_5 \psi(\tilde{x}) | P, S_T \rangle |_{\tilde{x}^+ = 0}.
\]

We avoid here discussing the subtle issues of a proper gauge invariant definition of \( p_T \)-dependent parton densities and just refer to the literature [45–51]. Obviously, the \( g_{1T} \)-contribution in Eq. (11) is not suppressed compared to the \( g_T \)-term which is already present in the collinear approach. It is worthwhile to note that, in contrast to the target SSA, \( p_T \)-dependent effects are suppressed in the case of the lepton asymmetries. The result in (11) is IR divergent as well, where the divergent terms are contained in the function \( H \). In particular the \( p_T \)-dependent term in \( H \) is also associated with an IR divergence. Provided that this divergence cancels after the inclusion of quark–gluon–quark correlators, one can perform the \( p_T \)-integral in (11) and might arrive at a description of the target SSA in terms of ordinary integrated correlators.
Concerning the inclusion of quark–gluon–quark correlators we limit ourselves here to a short qualitative discussion. Without detailed algebra one finds that two such correlators can contribute to the target SSA. Symbolically these objects can be written as

\[ \langle P, S | \bar{\psi} \gamma^+ A_T^g \psi | P, S \rangle, \quad \langle P, S | \bar{\psi} \gamma^+ p_S A_T^g \psi | P, S \rangle, \]

with \( A_T^g \) denoting the transverse components of the gluon field. It is possible that upon inclusion of such contributions an IR finite result for the target SSA can be obtained. Here one has to keep in mind that the QCD equations of motion relate quark–gluon–quark correlators of the type given in (13) to quark–quark correlators. Actually, it is quite interesting and promising that a certain linear combination of the matrix elements in (13) is connected to the particular combination of \( g_T \) and \( g_{1T} \) in Eq. (11).

To summarize, we have investigated transverse SSAs in inclusive DIS off the nucleon which, in general, can be induced by multi-photon exchange between the leptonic and the hadronic part of the reaction. Such SSAs exist for a polarized target as well as a polarized incoming or outgoing lepton. We have computed the asymmetry for a transversely polarized lepton beam and for a transversely polarized nucleon target in the framework of the parton model. So far we have only considered contributions from the quark–quark correlator. In this approach the beam spin asymmetry turns out to be proportional to the twist-2 unpolarized quark density inside the nucleon. In contrast, the target SSA is a genuine twist-3 observable which is reflected by the appearance of the twist-3 parton density \( g_T \). We have also studied the influence of the transverse motion of the struck quark. While such effects are suppressed for the lepton asymmetries, one finds a leading contribution in the case of the target SSA. Due to an uncancelled IR divergence our present result for the target SSA apparently is incomplete. However, it is likely that the full leading contribution (in the Bjorken limit) can be obtained if quark–gluon–quark correlators are also taken into consideration.

The transverse spin asymmetries are of \( \mathcal{O}(a_{em}) \) and may therefore be small. Moreover, they are suppressed like \( m_{pol}/Q \) with \( m_{pol} \) denoting the mass of the polarized particle. At least in the case of the nucleon target SSA this suppression is not severe as long as \( Q \) is in the region of a few GeV. Although probably difficult, we think it is definitely worthwhile to experimentally explore such transverse SSAs. Currently, measurements could be performed at CERN, DESY, and at Jefferson Lab.

Acknowledgements

We would like to thank A. Afanasev and C. Weiss for discussions on the transverse target asymmetry in DIS and, in particular, for drawing our attention to the importance of quark–gluon–quark correlators. A conversation with X. Jiang was very valuable and helped us to find earlier work on single spin asymmetries in DIS. We are also grateful to J.C. Collins and O. Nachtmann for discussions. This research is part of the EU Integrated Infrastructure Initiative Hadronphysics Project under contract number RII3-CT-2004-506078. The work is partially supported by the Verbundforschung “Hadronen und Kerne” of the BMBF.

References

Threshold resummation of Drell–Yan rapidity distributions

Paolo Bolzoni

Dipartimento di Fisica, Università di Milano and INFN, Sezione di Milano, Via Celoria 16, I-20133 Milano, Italy

Received 8 September 2006; received in revised form 27 October 2006; accepted 31 October 2006
Available online 17 November 2006
Editor: G.F. Giudice

Abstract

We present a derivation of the threshold resummation formula for the Drell–Yan rapidity distribution. Our argument is valid for all values of rapidity and to all orders in perturbative QCD and can be applied to all Drell–Yan processes in a universal way, i.e. both for the production of a virtual photon $\gamma^*$ and the production of a vector boson $W^\pm$, $Z^0$. We show that for the fixed-target experiment E866/NuSea used in current parton fits, the NLL resummation corrections are comparable to NLO fixed-order corrections and are crucial to obtain agreement with the data.

In perturbative QCD, it is well known that, when one approaches to the boundary of the phase space, the cross section receives logarithmically-enhanced contributions at all orders. These large terms have been resummed a long time ago for the classes of inclusive hadronic processes of the type of deep-inelastic and Drell–Yan \cite{1,2} to next-to-leading-logarithmic order (NLL). More recently, the next-to-next-to-leading-logarithmic (NNLL) accuracy has been reached \cite{3}.

Threshold resummation of inclusive processes can affect significantly cross sections and the extraction of parton densities \cite{4,5}. For the case of small transverse momentum distributions in Drell–Yan processes, it has been shown that resummation is necessary to reproduce the correct behavior of the cross section \cite{6}.

The differential rapidity Drell–Yan cross section is used for the extraction of the ratio $\bar{d}/\bar{u}$ of parton densities. The accurate knowledge of these functions is needed to study Higgs boson production and the asymmetry $W^\pm$. The resummation of Drell–Yan rapidity distributions was first considered in 1992 \cite{7}. At that time, it was suggested a resummation formula for the case of zero rapidity \cite{8}. Very recently, thanks to the analysis of the full NLO calculation of the Drell–Yan rapidity distribution, it has been shown \cite{8}, that the result given in \cite{7} is valid at NLL for all rapidities.

In this Letter, we will give a simple proof of an all-order resummation formula valid for all values of rapidity. To do this, we will use the technique of the double Fourier–Mellin moments developed in \cite{9}. In particular, we will show that the resummation can be reduced to that of the rapidity-integrated process, which is given in terms of a dimensionless universal function for both DY and $W^\pm$ and $Z^0$ production, and has been largely studied \cite{1,2} even to all logarithmic orders \cite{10}. Finally, we implement numerically the resummation formula and give predictions of the full rapidity-dependent NLL Drell–Yan cross section for the case of the fixed-target E866/NuSea experiment.

We find that resummation at the NLL level is necessary and that its agreement with the experimental data is better than the NNLO calculation \cite{11}.

We consider the general Drell–Yan process in which the collisions of two hadrons ($H_1$ and $H_2$) produce a virtual photon $\gamma^*$ (or an on-shell vector boson $V$) and any collection of hadrons ($X$):

$$H_1(P_1) + H_2(P_2) \rightarrow \gamma^*(V)(Q) + X(K). \quad (1)$$

In particular, we are interested in the differential cross section $\frac{d\sigma}{dQ^2}(x, Q^2, Y)$, where $Q^2$ is the invariant mass of the photon or of the vector boson, $x$ is defined as usual as the fraction of invariant mass that the hadrons transfer to the photon (or to the vector boson) and $Y$ is the rapidity of $\gamma^*(V)$ in the hadronic

---

E-mail address: paolo.bolzoni@mi.infn.it (P. Bolzoni).
center-of-mass:
\[ x \equiv \frac{Q^2}{S}, \quad S = (P_1 + P_2)^2, \quad Y \equiv \frac{1}{2} \ln \left( \frac{E + p_z}{E - p_z} \right). \tag{2} \]
where \( E \) and \( p_z \) are the energy and the momentum along the collisional axis of \( \gamma^*(V) \), respectively. At the partonic level, a parton 1 (2) in the hadron \( H_1 (H_2) \) carries a longitudinal momentum \( p_1 = x_1 P_1 \) (\( p_2 = x_2 P_2 \)). Thus, the rapidity in the partonic center-of-mass (\( y \)) is obtained performing a boost of \( Y \) between the two frames:
\[ y = Y - \frac{1}{2} \ln \left( \frac{x_1}{x_2} \right). \tag{3} \]
In order to understand the kinematic configurations in terms of rapidity, it is convenient to define a new variable \( u \),
\[ u \equiv \frac{Q \cdot p_1}{Q \cdot p_2} = e^{-2y} = \frac{x_1}{x_2} e^{-2y}, \tag{4} \]
which can assume all the values in the closed interval,
\[ z \leq u \leq \frac{1}{z}, \tag{5} \]
with
\[ z = \frac{Q^2}{2p_1 \cdot p_2} = \left( \frac{Q^2}{(p_1 + p_2)^2} \right) = \frac{x}{x_1 x_2}. \tag{6} \]
The upper and lower bounds in Eq. (5) are reached when the extra radiation is emitted collinear to the incoming parton 1 and 2, respectively. Eqs. (3), (4) allow us to rewrite the relation in Eq. (5) as a relation for the upper and lower bounds of the partonic center-of-mass rapidity:
\[ \frac{1}{2} \ln z \leq y \leq \frac{1}{2} \ln \frac{1}{z}. \tag{7} \]
Substituting Eqs. (4), (6) into the two conditions \( u \geq z \) and \( u \leq 1/z \), we obtain the lower bound for \( x_1 \) and \( x_2 \):
\[ x_1 \geq \sqrt{e^y} x_1^0, \quad x_2 \geq \sqrt{e^{-y}} x_2^0 \tag{8} \]
and the obvious requirement \( x_{1(2)}^0 \leq 1 \) implies that the hadronic rapidity has a lower and an upper bound:
\[ \frac{1}{2} \ln x \leq Y \leq \frac{1}{2} \ln \frac{x}{x}. \tag{9} \]
The variable \( z \) in Eq. (6) can be viewed as the fraction of invariant mass that the incoming partons transfer to \( \gamma^*(V) \) and, hence, the threshold limit is reached when \( z \) approaches to 1.
According to the standard factorization of collinear singularities of perturbative QCD, the expression for the hadronic differential cross section in rapidity has the form,
\[ \frac{d\sigma}{dQ^2 dY} = \sum_{i,j} \int_{x_j^0}^{1} \int_{x_i^0}^{1} dx_1 \int_{x_j^0}^{x_i^0} dx_2 \ F_i^{H_1}(x_1, \mu^2) F_j^{H_2}(x_2, \mu^2) \times \frac{d\hat{\sigma}_{ij}}{dQ^2 dy}(x_1, x_2, \frac{Q^2}{\mu^2}, \alpha_s(\mu^2), y), \tag{10} \]
where \( y \) depends on \( Y, x_1 \) and \( x_2 \) according to Eq. (3). The sum runs over all possible partonic subprocesses, \( F_i^{(1)}, F_j^{(2)} \) are respectively the parton densities of the hadron \( H_1 \) and \( H_2, \mu \) is the factorization scale (chosen equal to renormalization scale for simplicity) and \( d\hat{\sigma}_{ij}/(dQ^2 dy) \) is the partonic cross section. In the threshold limit the gluon–quark channels are suppressed by powers of \((1 - z)^2 \) \cite{10} and, so, in order to study resummation, we will consider only the quark–anti-quark contributions of the sum in Eq. (10). These last terms are related to the same dimensionless coefficient function \( C(z, Q^2/\mu^2, \alpha_s(\mu^2), y) \) through the relations
\[ x_1 x_2 \frac{d\hat{\sigma}_{qq}^{*}}{dQ^2 dy}(x_1, x_2, \frac{Q^2}{\mu^2}, \alpha_s(\mu^2), y) = \frac{4\pi \alpha_s}{9 Q^2 S} C(z, \frac{Q^2}{\mu^2}, \alpha_s(\mu^2), y) \tag{11} \]
for the virtual photon vertex and
\[ x_1 x_2 \frac{d\hat{\sigma}_{Vq'}^{*}}{dQ^2 dy}(x_1, x_2, \frac{Q^2}{\mu^2}, \alpha_s(\mu^2), y) = \frac{\pi G_F Q^2}{3S} \delta(Q^2 - M_V^2) C(z, \frac{Q^2}{\mu^2}, \alpha_s(\mu^2), y), \tag{12} \]
for the real vector boson vertex. Here \( G_F \) is the Fermi constant, \( M_V \) is the mass of the produced vector boson. The coefficients \( c_{q\bar{q}'} \) are given by:
\[ c_{q\bar{q}'} = \frac{Q^2}{Z_q} \delta_{q\bar{q}'} \quad \text{for } \gamma^*, \tag{13} \]
\[ c_{q\bar{q}'} = |V_{qq'}|^2 \quad \text{for } W^\pm, \tag{14} \]
\[ c_{q\bar{q}'} = 4[(g_q^2)^2 + (g_{q'}^2)^2] \delta_{q\bar{q}'} \quad \text{for } Z^0, \tag{15} \]
where \( Q^2_q \) is the square charge of the quark \( q \), \( V_{qq'} \) are the CKM mixing factors for the quark flavors \( q, q' \) and
\[ g_V^q = \frac{1}{2} - \frac{4}{3} \sin^2 \theta_W, \tag{16} \]
\[ g_W^q = \frac{1}{2} \quad \text{for an up-type quark}, \tag{16} \]
\[ g_Z^q = -\frac{1}{2} + \frac{2}{3} \sin^2 \theta_W, \tag{17} \]
\[ g_W^q = -\frac{1}{2} \quad \text{for a down-type quark}, \tag{17} \]
with \( \theta_W \) the Weinberg weak mixing angle. Thus, we are left with a dimensionless cross section of the form:
\[ \sigma(x, Q^2, Y) = \int x_1 \int x_2 \ F_i^{H_1}(x_1, \mu^2) F_j^{H_2}(x_2, \mu^2) \times C(z, \frac{Q^2}{\mu^2}, \alpha_s(\mu^2), y), \tag{18} \]
where \( F_i \) and \( F_2 \) are quark or anti-quark parton densities in the hadron \( H_1 \) and \( H_2 \), respectively. This shows the universality of resummation in Drell–Yan processes in the sense that only the quantity defined in Eq. (18) has to be resummed.
We shall now show that the resummed expression of Eq. (18) is obtained by simply replacing the coefficient function \( C(z, Q^2/\mu^2, \alpha_s(\mu^2), y) \) with its integral over \( y \), resummed to the
desired logarithmic accuracy. To show this, we recall that resummation is usually performed in the space of the variable $N$, which is the Mellin conjugate of $x$, since Mellin transformation turns convolution products into ordinary products. In the case of the rapidity distribution, however, this is not sufficient. In fact, we see that the Mellin transform with respect to $x$,

$$\sigma (N, Q^2, Y) \equiv \int_0^1 dx \, x^{N-1} \sigma (x, Q^2, Y),$$  

(19)

does not diagonalize the double integral in Eq. (18), because the partonic center-of-mass rapidity $y$ depends on $x_1$ and $x_2$ through Eq. (3). The ordinary product in Mellin space can be recovered performing the Mellin transform with respect to $x$ of the Fourier transform with respect to $Y$. Using Eqs. (9), (7) and the fact that the coefficient function must be symmetric in $y$, we find

$$\sigma (N, Q^2, M) \equiv \int_0^1 dx \, x^{N-1} \int dY \, e^{iMY} \sigma (x, Q^2, Y)$$  

(20)

$$= F_i^{H_i} (N + i M/2, \mu^2) F_{iM}^{H_i} (N - i M/2, \mu^2)$$

$$\times C \left( N, \frac{Q^2}{\mu^2}, \alpha_s (\mu^2), M \right),$$  

(21)

where

$$F_i^{H_i} (N \pm i M/2, \mu^2) = \int_0^1 dx \, x^{N-1-\pm i M/2} F_i^{H_i} (x, \mu^2),$$  

(22)

$$C \left( N, \frac{Q^2}{\mu^2}, \alpha_s (\mu^2), M \right)$$

$$= 2 \int_0^{\frac{1}{\sqrt{\mu}}} dz \, z^{N-1} \int_0^1 dy \, \cos (M y) C \left( z, \frac{Q^2}{\mu^2}, \alpha_s (\mu^2), y \right).$$  

(23)

The dependence on $M$, the Fourier conjugate of the rapidity $y$, originates from the parton densities, that depend on $N \pm i M/2$, and from the factor of $\cos (M y)$ in the integrand of Eq. (23). This last dependence, however, is irrelevant in the large-$N$ limit. Indeed, one can expand $\cos (M y)$ in powers of $y$,

$$\cos (M y) = 1 - \frac{M^2 y^2}{2} + O (M^4 y^4),$$  

(24)

and observe that the first term of this expansion leads to a convergent integral (the rapidity-integrated cross section), while the following terms are suppressed by powers of $(1-z)$, since the upper integration bound is

$$\ln \frac{1}{\sqrt{\mu}} = \frac{1}{2} (1-z) + O ((1-z)^2).$$  

(25)

Hence, up to terms suppressed by factors $1/N$, Eq. (23) is equal to the Mellin transform of the rapidity-integrated Drell–Yan coefficient function that we call $C_i (N, Q^2/\mu^2, \alpha_s (\mu^2))$. This completes our proof. We get

$$\sigma^{\text{res}} (N, Q^2, M) = F_i^{H_i} (N + i M/2, \mu^2) F_{iM}^{H_i} (N - i M/2, \mu^2)$$

$$\times C_i^{\text{res}} \left( N, \frac{Q^2}{\mu^2}, \alpha_s (\mu^2) \right).$$  

(26)

This is the main theoretical result of our Letter: It shows that, near threshold, the Mellin–Fourier transform of the coefficient function does not depend on the Fourier moments and that this is valid to all orders of QCD perturbation theory. Furthermore this result remains valid for all values of hadronic center-of-mass rapidity, because we have introduced a suitable integral transform over rapidity. The resummed rapidity-integrated Drell–Yan coefficient function to NLL is well known [1,2] and, using the notation of [10], it is given in a compact form (in the $\overline{\text{MS}}$ scheme) by

$$C_i^{\text{res}} \left( N, \frac{Q^2}{\mu^2}, \alpha_s (\mu^2) \right)$$

$$= \exp \left\{ - \int \frac{dn}{n} \left[ \int \frac{dk^2}{k^2} \left( A_1 \alpha_s \left( \frac{k^2}{n} \right) + A_2 \alpha_s^2 \left( \frac{k^2}{n} \right) \right) + B_1 \alpha_s \left( \frac{Q^2}{n} \right) \right] \right\},$$  

(27)

where

$$A_1 = \frac{C_F}{\pi}, \quad A_2 = \frac{C_F}{2 \pi} \left[ C_A \left( 67 - \frac{\pi^2}{6} \right) - \frac{5}{9} N_f \right],$$  

(28)

$$B_1 = - \frac{\gamma_E A_1}{2 \pi},$$  

(29)

with $C_F = 4/3$, $C_A = 3$, $N_f$ the number of flavors and with the Euler gamma $\gamma_E = 0.5772\ldots$. The use of only the first coefficient $A_1$ allows us to resum all the LL contributions $\alpha_s^k \log^{k+1} (N)$ while using the use of all the three coefficients in Eq. (28) enable us to add also the NLL terms $\alpha_s^k \log^k (N)$.

A NLL expression of the rapidity distribution is obtained by taking the inverse Mellin and Fourier transform of $\sigma^{\text{res}} (N, Q^2, M)$. This procedure requires the use of some specific prescription [12–14] in order to overcome the problem of the Landau singularity in $\alpha_s (Q^2/N^2)$. Here, we adopt the “Minimal Prescription” proposed in [12], which is simply obtained choosing the integration contour of the inverse Mellin transform in such a way that all the poles of the integrand are to the left, except the Landau pole. Furthermore, in order to improve numerical convergence and to avoid the singularities of the parton densities of Eq. (26) which are computed out of the real axis, we perform the $N$-integral along a path $\Gamma$ given by:

$$\Gamma = \Gamma_1 + \Gamma_2 + \Gamma_3,$$  

(29)

$$\Gamma_1 (t) = C_{MP} - t \frac{M}{2} + t (1+i), \quad t \in (-\infty, 0),$$  

(30)

$$\Gamma_2 (s) = C_{MP} + is \frac{M}{2}, \quad s \in (-1, 1),$$  

(31)

$$\Gamma_3 (t) = C_{MP} + i \frac{M}{2} - t (1-i), \quad t \in (0, +\infty),$$  

(32)

where $C_{MP}$ is a positive number below the Landau pole of $\alpha_s (Q^2/N^2)$. Performing the changes of variable $M = -\ln m$
and \( t = - \ln s \), the double inverse transform over the curve \( \Gamma \) becomes:

\[
\sigma^\text{res}(x, Q^2, Y) = \frac{1}{\pi} \int \frac{dm}{m} \cos(-Y \ln m) \sigma^\text{res}(x, Q^2, -\log m),
\]

where \( \sigma^\text{res}(x, Q^2, M) \) is given by

\[
\sigma^\text{res}(x, Q^2, M) = \frac{1}{\pi} \int \frac{ds}{s} \left[ x^{-C_{MF P} - \ln s + i(M/2 + 1)} \sigma^\text{res} \right.
\]

\[
\times \left( C_{MF P} + \ln s - i(M/2 + 1), Q^2, M \right)(1 - i)
\]

\[
+ \frac{sM}{2} x^{-C_{MF P} - i s M/2} \sigma^\text{res} \left( C_{MF P} + i s M/2, Q^2, M \right). \quad (34)
\]

Eqs. (33), (34) are the expressions that we use to evaluate numerically the resummed dimensionless cross section in the variables \( x \) and \( Y \). The explicit expression of Eq. (27) is easily obtained performing the integrals and is given by

\[
C^\text{res}_I \left( N, \frac{Q^2}{\mu^2}, \alpha_s(\mu^2) \right) = \exp \left[ \ln N g_1(\lambda) + g_2(\lambda) \right], \quad (35)
\]

where

\[
g_1(\lambda) = \frac{A_1}{\beta_0 \lambda} \left[ 2 \lambda + (1 - 2 \lambda) \log(1 - 2 \lambda) \right], \quad (36)
\]

\[
g_2(\lambda) = - \frac{2 A_1 Y_E}{\beta_0} \log(1 - 2 \lambda)
\]

\[
+ \frac{A_1}{\beta_0^3} \left[ 2 \lambda + \log(1 - 2 \lambda) + \frac{1}{2} \log^2(1 - 2 \lambda) \right]
\]

\[
- \frac{A_2}{\beta_0^3} \left[ 2 \lambda + \log(1 - 2 \lambda) \right] + \log \left( \frac{Q^2}{\mu^2} \right) \frac{A_1}{\beta_0} \log(1 - 2 \lambda)
\]

\[
\lambda = \beta_0 \alpha_s(\mu^2) \ln N, \quad \beta_0 = \frac{1}{4\pi} \left( 11 - \frac{2}{3} N_f \right), \quad (37)
\]

and

\[
\beta_1 = \frac{1}{16\pi^2} \left( 102 - \frac{38}{3} N_f \right). \quad (38)
\]

Here, we choose the factorization scale equal to the renormalization scale for simplicity. To study the dependence on the renormalization scale one has simply to express \( \alpha_s(\mu^2) \) in terms of it. Furthermore, we need the analytic continuations to the whole complex plane of the Mellin-transformed parton densities that appear in Eq. (26). In order to overcome this problem, we have to evolve up a partonic fit taken at a certain scale solving the DGLAP evolution equations in Mellin space [15]. The LO and NLO expressions of the splitting functions are reported in [16] and their analytic continuations are given in [17] and [18].

Finally, we want to obtain a NLO determination of the cross section improved with NLL resummation. In order to do this, we must keep the resummed dimensionless part of the cross section Eq. (33), multiply it by the correct dimensionless prefactors looking Eqs. (11)–(17), add the full NLO cross section and subtract the double-counted logarithmic enhanced contributions. This matching has to be done in the \( x \) and \( Y \) spaces, because we are not able to calculate the Mellin–Fourier moments of the full NLO cross section analytically. Thus, we have

\[
\frac{d\sigma}{dQ^2 dY} = \frac{d\sigma^\text{FO}}{dQ^2 dY} + \frac{d\sigma^\text{res}}{dQ^2 dY} \left[ \alpha_s(\mu^2) - \alpha_s \right] \frac{d\sigma^\text{res}}{dQ^2 dY} \bigg|_{\alpha_s=0}.
\]

The first term is the full NLO cross section reported in [8,19–21], which includes even the quark–gluon channel. The third and the fourth terms in Eq. (39) are obtained in the same way as the second one, but with the substitutions

\[
C^\text{res}_I \left( N, \frac{Q^2}{\mu^2}, \alpha_s(\mu^2) \right) \rightarrow 1, \quad (40)
\]

\[
C^\text{res}_I \left( N, \frac{Q^2}{\mu^2}, \alpha_s(\mu^2) \right) \rightarrow \alpha_s(\mu^2) 2 A_1 \left\{ \ln^2 N + \ln N \left[ 2 Y_E - \log \left( \frac{Q^2}{\mu^2} \right) \right] \right\}, \quad (41)
\]

respectively. The terms that appear in Eq. (41) are exactly the \( O(\alpha_s) \) logarithmic enhanced contributions in the \( \overline{\text{MS}} \) scheme.

We note that the final expression Eq. (39) is relevant even when the variable \( x \) is not large. In fact, the cross section can get the dominant contributions from the integral in Eq. (18) for values of \( z \) (Eq. (6)) that are near the threshold even when \( x \) is not close to one, because of the strong suppression of parton densities \( \Gamma(x_i, \mu^2) \) when \( x_i \) are large.

To show the importance of this resummation, we have calculated the Drell–Yan rapidity distribution for proton–proton collisions at the Fermilab fixed-target experiment E866/NuSea [22]. The center-of-mass energy has been fixed at \( \sqrt{s} = 38.76 \text{ GeV} \) and the invariant mass of the virtual photon \( \gamma^* \) has been chosen to be \( Q^2 = 64 \text{ GeV}^2 \) in analogy with [11]. Clearly the contribution of the virtual \( Z^0 \) can be neglected, because its mass is much bigger than \( Q^2 \). In this case \( x = 0.04260 \) and the upper and lower bound of the hadronic rapidity \( Y \) Eq. (9) are given by \( \pm 1.57795 \). We have evolved up the MRST 2001 parton distributions (taken at \( \mu^2 = 1 \text{ GeV}^2 \)) as in [11], where the NNLO calculation is performed. However, results obtained using more modern parton sets should not be very different. The LO parton set is given in [23] with \( \alpha_s^{\text{LO}}(m_Z) = 0.130 \) and the NLO set is given in [24] with \( \alpha_s^{\text{NLO}}(m_Z) = 0.119 \). The evolution of parton densities at the scale \( \mu^2 \) has been performed in the variable flavor number scheme. The quarks has been considered massless and, at the scale of the transition of the flavor number \( (N_f \rightarrow N_f + 1) \), the new flavor is generated dynamically. The resummation formula Eq. (26) together with Eqs. (35)–(38) has been used with the number of flavors \( N_f = 4 \).

In Fig. 1, we plot the rapidity-dependence of the cross section at LO, NLO and LO improved with LL resummation. The effect of LL resummation is small compared to the effect of the full NLO correction. We see that, at leading order, the impact of the resummation is negligible in comparison to the NLO
fixed-order correction. This means that the NLL resummation is necessary.

The LO, the NLO and its NLL improvement cross sections are shown in Fig. 2. The effect of the NLL resummation in the central rapidity region is almost as large as the NLO correction, but it reduces the cross section instead of enhancing it for not large values of rapidity. Going from the LO result to the NLO with NLL resummation, we note a reduction of the dependence on the factorization scale i.e. a reduction of the theoretical error. It is interesting to observe that logarithmically enhanced and constant terms account for more than 80% of the NLO contribution for all relevant rapidities. Therefore, they have the same sign. Nevertheless a suppression arises due to the shift in the complex plane of the dominant contribution of the resummed exponent. This suppression starts at order $O(\alpha_s^3)$.

In Fig. 3, we report the experimental data of [22] converted to the $Y$ variable, together with our NLO and NLL resummed predictions. The agreement with data is good and a great improvement for not large rapidity is obtained with respect to the NLO calculation. We note also that the NLL resummation gives better result than the NNLO calculation performed in [11]. The NNLO prediction has a worse agreement with data than the NLO one for not large values of rapidity. This result suggests that, for the case of rapidity distributions, NLL resummation is more important than high-fixed-order calculation and that it can be so even at higher center-of-mass energies.

To summarize, we have proved a resummation formula for the Drell–Yan rapidity distributions to all logarithmic accuracy and valid for all values of rapidity. Isolating a universal dimensionless coefficient function, which is exactly that ones of the Drell–Yan rapidity-integrated, we have shown a general procedure to obtain resummed results to NLL for the rapidity distributions of a virtual photon $\gamma^*$ or of a real vector boson $W^\pm, Z^0$. Furthermore, we have outlined a general method to calculate numerical predictions and analyzed the impact of resummation for the fixed-target experiment E866/NuSea. This shows that NLL resummation has an important effects on predictions of differential rapidity cross sections giving an agreement with data that is better than NNLO full calculations.

**Acknowledgements**

I want to thank S. Forte and G. Ridolfi for suggesting this problem; A. Vicini for useful discussions about numerical calculations; G. Cerati, B.M. Mognetti, S. Montesano and A. Zellio for their concern about my programming difficulties.

**References**

Gluonic phase versus LOFF phase in two-flavor quark matter

O. Kiriyama a,∗, D.H. Rischke a, b, I.A. Shovkovy b, c, 1

a Institut für Theoretische Physik, J.W. Goethe-Universität, D-60438 Frankfurt am Main, Germany
b Frankfurt Institute for Advanced Studies, J.W. Goethe-Universität, D-60438 Frankfurt am Main, Germany
c Department of Physics, Western Illinois University, Macomb, IL 61455, USA

Received 6 June 2006; received in revised form 26 September 2006; accepted 30 October 2006

Abstract

We study the gluonic phase in a two-flavor color superconductor as a function of the ratio of the gap over the chemical potential mismatch, \( \Delta/\delta\mu \). We find that the gluonic phase resolves the chromomagnetic instability encountered in a two-flavor color superconductor for \( \Delta/\delta\mu < \sqrt{2} \). We also calculate approximately the free energies of the gluonic phase and the single plane-wave LOFF phase and show that the former is favored over the latter for a wide range of coupling strengths.

© 2006 Elsevier B.V. All rights reserved.

PACS: 12.38.-t; 11.30.Qc; 26.60.+c

It is widely accepted that sufficiently cold and dense quark matter is a color superconductor [1]. The most likely and, probably, the only place where color superconductivity can exist in the universe is the interior of compact stars. Thus, studies of phases of quark matter under conditions realized in the bulk of compact stars (i.e., color and electric charge neutrality, and \( \beta \)-equilibrium) have recently attracted a great deal of interest. The density regime of relevance for compact stars is up to a few times the normal nuclear density \( \rho_0 \simeq 0.16 \text{ fm}^{-3} \). In this “moderate” density regime, the studies of QCD-motivated effective theories are most useful and have revealed a rich phase structure [2–4].

One of the most striking features of neutral and \( \beta \)-equilibrated color-superconducting phases is unconventional cross-flavor Cooper pairing of quarks with the possibility of gapless superconductivity, e.g., in the form of the gapless 2SC (g2SC) phase [5] or the gapless color-flavor-locked (gCFL) phase [6]. It was, however, quickly realized that the 2SC/g2SC phases suffer from a chromomagnetic instability, indicated by imaginary Meissner screening masses of some gluons [7]. In the 2SC phase, these instabilities occur when the ratio of the gap over the mismatch of the chemical potential, \( \Delta/\delta\mu \), decreases below a value \( \sqrt{2} \). Similar instabilities were found also in the gCFL phase [8].

Resolving the chromomagnetic instability and clarifying the nature of the true ground state of dense quark matter are the most pressing tasks in the study of color superconductors. It was proposed that the chromomagnetic instability in two-flavor quark matter can be removed by the formation of a single plane-wave LOFF state [9–12] (first studied by Larkin and Ovchinnikov [13], and Fulde and Ferrell [14] in the context of solid state physics, and by Alford et al. [15] for cold, dense quark matter), or a gluonic phase with vector condensation in the ground state [16]. (For a recent discussion of this issue, see also Refs. [17,18].) Alternatives include a mixed phase [19] and, in the case of three-flavor quark matter, also phases with spontaneously induced meson supercurrents [20]. While the neutral LOFF state is free from the chromomagnetic instability in the weak-coupling regime [10], this is, in fact, not the case for somewhat larger values of the coupling [21]. At the same time, the gluonic phase can resolve the instability there. So far, how-

* Corresponding author.

E-mail addresses: kiriyama@th.physik.uni-frankfurt.de (O. Kiriyama), drischke@th.physik.uni-frankfurt.de (D.H. Rischke), shovkovy@th.physik.uni-frankfurt.de (I.A. Shovkovy).

1 On leave from Bogolyubov Institute for Theoretical Physics, 03143 Kiev, Ukraine.

0370-2693/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.physletb.2006.10.067
ever, the gluonic phase phase has been studied only around the critical point \( \Delta / \delta \mu = \sqrt{2} \) [16].

The gluonic phase and the LOFF phases are currently viewed as the most likely candidates for resolving the chromomagnetic instability and, thus, for the true ground state of two-flavor color-superconducting quark matter. (Here we exclude the possibility of phase separation [19] which may have limitations of its own and deserves a separate in-depth study.) In order to see which of the two proposed phases is actually preferred, one first has to extend the analysis of Ref. [16] to a computation of the free energy away from the critical point \( \Delta / \delta \mu = \sqrt{2} \), and then compare the results to the free energy of the single-plane LOFF state.\(^2\) This is done in the present work. We qualitatively confirm the results of Ref. [18] and extend them by (approximately) including the neutrality condition and explicitly comparing the free energy of the gluonic phase to that of the single plane-wave LOFF phase as a function of the coupling strength.

In order to study various phases of two-flavor quark matter, we use a gauged Nambu–Jona-Lasinio (NJL) model with massless up and down quarks:

\[
\mathcal{L} = \bar{\psi} (i \partial - i \mu \gamma^0) \psi + G_D (\bar{\psi} i \gamma_5 \varepsilon^{eb} C \bar{\psi}^T) (\psi^T C i \gamma_5 \varepsilon^{eb} \psi) - \frac{1}{4} F_{\mu\nu}^a F^{a \mu\nu}, \tag{1}
\]

where the quark field \( \psi \) carries flavor (\( i, j = 1, \ldots, N_f \) with \( N_f = 2 \)) and color (\( \alpha, \beta = 1, \ldots, N_c \) with \( N_c = 3 \)) indices, \( C \) is the charge conjugation matrix, \( (\varepsilon)^k = \delta^{ik} \) and \( (\varepsilon^{eb})^{\rho\beta} = \delta^{\rho\beta} \) are the antisymmetric tensors in flavor and color spaces, respectively. The covariant derivative and the field strength tensor are defined as

\[
D_{\mu} = \partial_{\mu} - i g A_{\mu}^a T^a, \tag{2a}
\]

\[
F_{\mu\nu}^a = \partial_{\mu} A_{\nu}^a - \partial_{\nu} A_{\mu}^a + g f^{abc} A_{\mu}^b A_{\nu}^c, \tag{2b}
\]

To evaluate loop diagrams we use a three-momentum cutoff \( \Lambda \). Hence, the model has two phenomenological model parameters, the cutoff \( \Lambda \) and the diquark coupling \( G_D \). We use \( \Lambda = 653.3 \) MeV throughout this Letter, but we consider \( G_D \) as a free parameter. Henceforth, in order to specify the diquark coupling we use \( \Lambda_0 \) which is the value of the 2SC gap at \( \delta \mu = 0 \) (see below).

In \( \beta \)-equilibrated neutral 2SC/g2SC matter, the elements of the diagonal matrix of quark chemical potentials \( \mu \) are given by

\[
\mu_{uv} = \mu_{ag} = \tilde{\mu} - \delta \mu, \tag{3a}
\]

\[
\mu_{dt} = \mu_{dg} = \tilde{\mu} + \delta \mu, \tag{3b}
\]

\[
\mu_{ub} = \tilde{\mu} - \delta \mu - \mu_8, \tag{3c}
\]

\[
\mu_{db} = \tilde{\mu} + \delta \mu - \mu_8, \tag{3d}
\]

with

\[
\tilde{\mu} = \mu - \frac{\delta \mu}{3} + \frac{\mu_8}{3}, \quad \delta \mu = \frac{\mu_e}{2}. \tag{4}
\]

In a gauge theory, the self-consistent solution of the Yang–Mills equations requires background gauge fields [23]. These can be viewed as electric- and color-chemical potentials which ensure electric and color-charge neutrality of the system. Note that a generalization of this holds true even in the case of inhomogeneous phases. Then, of course, the corresponding fields would not be constant in space. Instead, they would have a constant central value contribution and, on top of it, a coordinate-dependent modulation describing color-electric fields induced by the inhomogeneities.\(^3\) The constant contribution would take care of the global neutrality, while the modulation describes the local field needed to prevent the local flow of currents.

On the other hand, in NJL-type models without dynamic gauge fields, one has to ensure electric and color-charge neutrality by introducing appropriate chemical potentials by hand [24]. In the case of the 2SC/g2SC phases, we only require an electron chemical potential \( \mu_e \), and a color-chemical potential \( \mu_8 \) which ensures that the color-charge density \( n_8 \) is zero. In principle, in other phases like the gluonic phase one has to check that no other color-charge density is non-vanishing and necessitates the introduction of a respective color-chemical potential. Indeed, the gluonic phase introduced in Ref. [16] requires a non-vanishing temporal component of the gluon field of the third color, \( (\Phi^a) \). In our gauged NJL model, this is equivalent to a non-vanishing color-chemical potential \( \mu_3 \) besides \( \mu_8 \). In this first exploratory study, however, we use the fact that both \( \mu_3 \) and \( \mu_8 \) are known to be numerically small and we simply neglect them.

In Nambu–Gor’kov space, the inverse full quark propagator \( S^{-1}(p) \) is written as

\[
S^{-1}(p) = \begin{pmatrix} (S^+_0)^{-1} & \Phi^- \\ \Phi^+ & (S^-_0)^{-1} \end{pmatrix} \tag{5}
\]

with

\[
(S^+_0)^{-1} = \gamma^\mu p_\mu + (\tilde{\mu} - \delta \mu T^3)\gamma^0 + g \gamma^a A^a T^a, \tag{6a}
\]

\[
(S^-_0)^{-1} = \gamma^\mu p_\mu - (\tilde{\mu} - \delta \mu T^3)\gamma^0 - g \gamma^a A^a T^a T^a, \tag{6b}
\]

and

\[
\Phi^- = -ie\bar{\varepsilon}^b\gamma_5 \Delta, \quad \Phi^+ = -ie\bar{\varepsilon}^b\gamma_5 \Delta. \tag{7}
\]

Here \( T^3 = \text{diag}(1, -1) \) is a matrix in flavor space. Following the usual convention, we choose the diquark condensate to point in the third (blue) direction in color space.

In the one-loop approximation, the free energy of two-flavor quark matter at \( T = 0 \) is given by

\[
V_k = \frac{\Delta^2}{4 G_D} - \frac{1}{2} \int \frac{d^4 p}{(2\pi)^4} \ln \det S^{-1}(p), \tag{8}
\]

\(^{2}\) Note that while the comparison with a multi-plane wave LOFF states would be more desirable, the corresponding free energy cannot be easily estimated within a microscopic approach. For current state-of-the-art calculations using an effective theory see Ref. [22].

\(^{3}\) In the special case of a mixed phase, e.g., a color-electric field is generated around the boundary layer between the two phases and prevents the generation of a color-electric current across this layer.
where “Det” stands for the determinant in Dirac, flavor, color, and Nambu–Gor’kov space. Unlike the free energy in Ref. [16], Eq. (8) does not have a quartic term in $A^0_μ$. This is because we neglected the color-chemical potential $μ_5$ and, in addition, we take into account only one dynamic gluonic field (see below).

In the gluonic phase [16], the chromomagnetic instability at $Δ/δμ < √2$ triggers a non-vanishing vacuum expectation value of the spatial component of

$$K_μ = \frac{1}{\sqrt{2}} \left( A^4_μ - i A^5_μ \right).$$

(For a simpler version of such a phenomenon, see also Ref. [25]). Using the SO(3)$_{rot}$ rotational symmetry and the SU(2)$_c$ color symmetry, one can choose $B \equiv g \langle A_5^0 \rangle \neq 0$ without loss of generality. Consequently, the non-zero vacuum expectation value of $B$ breaks SO(3)$_{rot}$, leaving only SO(2)$_{rot}$ [16]. Furthermore, a non-vanishing $B$ together with $Δ$ and $μ_c$ breaks the original symmetry of QCD down to

$$U(1)_Ω \otimes U(1)_{t^3_L} \otimes U(1)_{t^3_R} \otimes \text{SO}(2)_{rot},$$

(10)

where $U(1)_{t^3}$ is a subgroup of the SU(2)$_{L/R}$ chiral symmetry and the charge $Ω$ is given by

$$Ω = Q_f \otimes \mathbb{1}_c - \mathbb{1}_f \otimes T^3 - \frac{1}{\sqrt{3}} \mathbb{1}_f \otimes T^8,$$

(11)

with $Q_f = \text{diag}(\frac{2}{3}, -\frac{1}{3})$ being the flavor matrix of the electric charges of quarks.

The reduced symmetry of the ground state with $B \neq 0$ allows for additional condensates, $C = g \langle A_1^0 \rangle \neq 0$ and $D = g \langle A_3^0 \rangle \neq 0$. In fact, as discussed in Ref. [16], such condensates are required by the equation of motion. As discussed above, the gluonic field $D$ is nothing but a color chemical potential $μ_3$. The field $C$, on the other hand, induces electric superconductivity in the ground state and, therefore, is physically more interesting. However, including all three gluonic fields makes the analysis quite involved. In this work, we retain only the $B$ field that is directly connected to the Meissner masses of gluons $4–7$ and, thus, is the most relevant field for the chromomagnetic instability.

It is straightforward to show that the mass of the $B$ field at $B = 0$ (i.e., in the 2SC/g2SC phases) coincides with the Meissner screening masses of gluons of adjoint color $4–7$ calculated in the hard-dense-loop (HDL) approximation [7],

$$M_B^2 = \frac{\partial^2 V}{\partial B^2} |_{B=0} = \frac{\mu_c^2}{6\pi^2} \left[ 1 - \frac{2\Delta^2}{\Delta^2} + \frac{2\delta\mu\sqrt{\delta\mu^2 - \Delta^2}}{\Delta^2} \theta(\delta\mu - \Delta) \right].$$

(12)

In order to derive this expression we neglected terms of order $O(\mu_c^2/\Delta^2)$ and $O(\Delta^2/\mu_c^2)$.

In order to effect the condensate field $B$ on the free energy of the 2SC/g2SC phases, we calculate the difference of the thermodynamic potentials in a dense medium and in vacuum at the same value of $B$,

$$Ω_g ≡ V_g(Δ, B, δμ, μ) - V_g(0, B, 0, 0).$$

(13)

In a gauge theory, this subtraction in the one-loop free energy also takes care of the renormalization of the gauge coupling constant. As a result, the cutoff dependence of the free energy can be completely removed in this approximation.

Let us look at the free energy $Ω_g$ in detail. Fig. 1 shows the free energy $Ω_g$ (measured with respect to the normal phase at $B = 0$) as a function of $Δ/δμ$. The results are plotted for $\tilde{μ} = 500$ MeV and $δμ = 80$ MeV, with the diquark coupling chosen so that $Δ_0 = 132$ MeV. Here we do not restrict $Δ/δμ$ to its physical value, determined by the stationary point of $Ω_g(Δ)$, but treat it as a free parameter.

Several important features of the free energy as a function of the $B$ field are evident from Fig. 1. When $Δ/δμ > √2$, one can see that the free energy monotonically increases with $B$. (Strictly speaking, since our model reproduces the HDL result only up to terms of order $O(\mu_c^2/\Delta^2)$ and $O(\Delta^2/\mu_c^2)$, the actual critical point is somewhat lower than $√2$.) In other words, $Ω_g(B)$ has a global minimum at $B = 0$, and the 2SC phase is stable against gluon condensation in this regime. This is also clear from a different representation of the results, shown in Fig. 2.

When $Δ/δμ < √2$, on the other hand, we observe the onset of the chromomagnetic instability. For small $B$, the free energy first decreases with increasing $B$ and then grows at larger $B$. This can be seen clearly in Fig. 2. The behavior of $Ω_g$ agrees well with Eq. (12) at small $B$. In this regime, the 2SC/g2SC phase is no longer the ground state. It is unstable with respect to the formation of a non-zero $B$ condensate, i.e., the so-called gluonic phase. The corresponding ground state is determined by the minimum of $Ω_g(B)$. The Meissner masses squared, which are given by the curvature of the free energy at the minimum, are non-negative in this state. (It is interesting to note that, although the free energy in the normal phase, $Δ/δμ = 0$, cf.
where the 2SC and $g_{2SC}$ energy is vanishing, see Eq. (12). Let us note that the results of the 2SC gluonic phase resolves the chromomagnetic instability of the Ref. [18].

Fig. 2 are in qualitative agreement with those shown in Fig. 2 of Ref. [18].

From the results for the free energy it is clear that the gluonic phase resolves the chromomagnetic instability of the 2SC/g2SC phases. A neutral LOFF state is another candidate for the solution to the instability: it has been shown that such a state is free from the chromomagnetic instability, however, only in the weak-coupling regime [10] (for strong coupling, this is not the case [21]). In order to determine the energetically most favored state, it is necessary to compare the free energies of the 2SC/g2SC phases, the neutral LOFF state and the gluonic phase.

To this end we use the following approximation derived in Ref. [21] for the 2SC/g2SC phases and the neutral LOFF state:

$$\Omega = \Omega_{2SC} + \Omega_{g2SC/LOFF},$$

where the 2SC and g2SC/LOFF parts of the free energy are given by

$$\Omega_{2SC} = \frac{\mu_{e}^{2}}{12\pi^{2}} - \frac{\mu_{ab}^{2}}{12\pi^{2}} - \frac{\mu_{db}^{2}}{12\pi^{2}} - \frac{\mu_{fd}^{2}}{3\pi^{2}} - \frac{\Delta^{2}}{4G_{D}} - \frac{\Delta^{2}}{\pi^{2}} \ln \frac{4\Lambda^{2} - \mu^{2}}{\Lambda^{2} - \mu^{2}} - \frac{\Delta^{2}}{\pi^{2}} \left( \Lambda^{2} - 2\mu^{2} \right),$$

$$\Omega_{g2SC/LOFF} = \frac{2\mu_{e}^{2}q^{2}}{\pi^{2}} + \frac{\mu_{f}^{2}}{\pi^{2}} \left( \frac{(q + \delta\mu)^{3}}{q} \right) \frac{1}{2} \left[ (1 - x_{1}^{2}) \ln \frac{1 + x_{1}}{1 - x_{1}} - x_{1} + \frac{2}{3} x_{1}^{3} \right] + (q \rightarrow -q),$$

with the dimensionless parameter $x_{1}$ being

$$x_{1} = \theta \left( 1 - \frac{\Delta^{2}}{(\delta\mu + q)^{2}} \right) \left[ 1 - \frac{\Delta^{2}}{(\delta\mu + q)^{2}} \right]$$

and

$$q = |\vec{q}|, \quad \vec{q} = \frac{\delta}{2\sqrt{3}}(\Lambda^{8}).$$

Note that the wave vector of the diquark condensate $\vec{q}$ is equivalent to a gauge field condensate $(\Lambda^{8})$ in the case of single plane-wave LOFF pairing. In Eq. (15), the 2SC/g2SC part of the free energy is obtained by taking the $q \rightarrow 0$ limit. Also a non-zero color chemical potential $\mu_{B}$ has been neglected there.

The free energy of a given phase can be computed by solving the gap equations, e.g., $\partial\Omega/\partial\delta\Delta = 0$ and $\partial\Omega/\partial q = 0$, and the neutrality condition $\partial\Omega/\partial\delta\mu = 0$. To simplify the calculations in the gluonic phase, we evaluate the free energy approximately as follows: (i) we obtain $\Delta^{*}$ and $\delta\mu^{*}$ in the 2SC/g2SC phase by solving the coupled set of equations $\partial\Omega/\partial\Delta = 0$ and $\partial\Omega/\partial\delta\mu = 0$; (ii) by using these solutions, we calculate $\Omega_{g}(B, \Delta^{*}, \delta\mu^{*})$ which is an approximate value for the free energy in the gluonic phase. For the densities of interests, $B$ is at most of the order of 100 MeV, whereas $q$ is of the order of tens of MeV. We performed a preliminary test of the quality of our approximation by varying $B$ from 0 to 300 MeV and computing the value of $\delta\mu$ necessary to ensure electric neutrality. We found that this value changes at most by 10%.

We illustrate the comparison of the free energies of all three phases in Fig. 3 (cf. Fig. 2 in Ref. [21]). We take $\mu = 400$ MeV and choose the normal phase as a reference point for the free energy. In Ref. [21], it has been demonstrated that the neutral LOFF state is more stable than the 2SC/g2SC phases in the whole LOFF window 63 MeV $< \Delta_{0} < 137$ MeV, which includes the entire g2SC window 92 MeV $< \Delta_{0} < 130$ MeV. However, whereas the Meissner masses squared of gluons 4–7 in the weakly coupled neutral LOFF state are positive [9,10], they remain negative in the intermediate and the strongly coupled regimes (at all values of $\Delta_{0}$ above 81 MeV) [21]. In contrast, the gluonic phase removes the instability and is energetically favored over the 2SC/g2SC phases in the whole window in which the instability takes place (at all values of $\Delta/\delta\mu$ below $\sqrt{2}$). The gluonic phase and the neutral LOFF state coexist...
in the region $92 \text{ MeV} < \Delta_0 < 137 \text{ MeV}$, but, as our results indicate, the gluonic phase is more stable than the neutral LOFF state in a wide region $\Delta_0 > 103 \text{ MeV}$, which is close to the edge of the g2SC window with the normal phase. We argue therefore that the instability in the 2SC/g2SC phase is resolved by the formation of the gluonic phase.

In summary, we explored the gluonic phase away from the critical point $\Delta/\delta\mu = \sqrt{2}$. We demonstrated that the energetically favored state of a neutral two-flavor color superconductor is not the 2SC/g2SC phase in the intermediate- and strong-coupling regimes but the gluonic phase in which the dynamic gluonic field $B = g(A_0^c)$ acquires a vacuum expectation value. In particular, the whole g2SC phase is replaced by the gluonic phase which is chromomagnetically stable.

We also compared the free energies of the 2SC/g2SC phase, the neutral LOFF state, and the gluonic phase. We found that the gluonic phase is energetically favored in the intermediate- and strong-coupling regimes. The encouraging results of this analysis should be further improved in the future by (i) taking into account the most general ansatz for the gauge-field configuration in the gluonic phase, (ii) by calculating the free energy in a self-consistent manner. We already performed a preliminary investigation including the effect of the gluon field $D$, responsible for enforcing Gauss’s law, and found that, for $\Delta_0 \leq 130 \text{ MeV}$, the additional cost in the free energy is of order $0.01 \text{ MeV/fm}^3$, and thus negligible. For $\Delta_0 \gtrsim 130 \text{ MeV}$, however, the effect of the $D$ field could be an order of magnitude larger.

It is appropriate to mention that the instability related to the 8th gluon was not studied in the present work. In this sense, the neutral LOFF state is appealing, because the Meissner mass of the 8th gluon is automatically zero in this state. It should also be mentioned that in the strongly coupled LOFF state [9,11] the longitudinal Meissner mass squared of the 8th gluon is negative. Although this instability was not addressed in this work, it is unlikely, however, that the LOFF state is energetically favored in the strong-coupling regime.

Note added

It has recently been demonstrated that, in the three-flavor case, realistic crystal structures are more robust than a single plane-wave LOFF state [22]. In the two-flavor case, Bowers and Rajagopal [26] already indicated that a LOFF state with multiple plane waves would have a lower free energy than that with a single plane wave. The result shown in Fig. 3 would be altered by the inclusion of crystal structures with more plane waves. This is therefore an important project that needs to be addressed in future work.

Acknowledgements

I.A.S. acknowledges discussions with V.A. Miransky. This work was supported in part by the Virtual Institute of the Helmholtz Association under grant No. VH-VI-041, by the Gesellschaft für Schwerionenforschung (GSI), and by the Deutsche Forschungsgemeinschaft (DFG).

References

[48] A. Gerhold, A. Rebhan, Phys. Rev. D 72 (2005) 065020(R);
Neutrino masses and CDM in a non-supersymmetric model

Jisuke Kubo, Daijiro Suematsu *

Institute for Theoretical Physics, Kanazawa University, Kanazawa 920-1192, Japan

Received 3 October 2006; received in revised form 24 October 2006; accepted 4 November 2006

Available online 13 November 2006

Editor: T. Yanagida

Abstract

We propose a model for neutrino mass generation based on both the tree-level seesaw mechanism with a single right-handed neutrino and one-loop radiative effects in a non-supersymmetric framework. The generated mass matrix is composed of two parts which have the same texture and produce neutrino mass eigenvalues and mixing suitable for the explanation of neutrino oscillations. The model has a good CDM candidate which contributes to the radiative neutrino mass generation. The stability of the CDM candidate is ensured by $Z_2$ which is the residual symmetry of a spontaneously broken $U(1)'$. We discuss the values of $U_{e3}$ and also estimate the masses of the relevant fields to realize an appropriate abundance of the CDM.

© 2006 Elsevier B.V. All rights reserved.

1. Introduction

Recent experimental and observational results on neutrino masses [1] and cold dark matter (CDM) [2] suggest that the standard model (SM) should be extended by introducing some neutral fields. A well studied candidate for the extension is the minimal supersymmetric SM (MSSM). Although the MSSM contains a good CDM candidate as the lightest superparticle (LSP) as long as the $R$-parity is conserved, the parameter regions preferable for the explanation of the WMAP data are found to be strictly restricted in certain types of the MSSM [3]. Confronting these situations, it seems to be interesting to consider models in which we can explain these new features from the same origin in a non-supersymmetric extension of the SM: a certain symmetry related to the smallness of neutrino masses can guarantee the stability of a CDM candidate. Backgrounds that forth coming collider experiments like LHC may find signatures of such extended models make this kind of trials worthy enough at present stage. Several recent works have been done along this line [4].

In this Letter we follow this line to propose an extension of a previously considered model by introducing a local $U(1)'$ symmetry at TeV regions. As in the radiative mass generation models [5], we introduce an additional SU(2) doublet $\eta^T \equiv (\eta^+, \eta^0)$ to the ordinary Higgs doublet $H^T \equiv (H^+, H^0)$. We also introduce a singlet $\phi$ whose vacuum expectation value breaks $U(1)'$ symmetry spontaneously down to $Z_2$ which is responsible for the stability of the CDM candidate. This extension seems to remedy defects in the previous models that certain fine tunings are required for both the generation of small neutrino masses and the reconciliation between the CDM abundance and the constraints from lepton flavor violating processes. Based on such a model we calculate the value of the element $U_{e3}$ of the MNS matrix and masses of the relevant fields which produce an appropriate abundance of the CDM.

2. A model

We consider a model with a similar symmetry to the model in [4,6]. We extend it by introducing a singlet Higgs scalar $\phi$. This extension makes the model able to contain an additional $U(1)'$ symmetry. In this Letter we assume that this symmetry is leptophobic and then leptons do not have its charge for simplicity.1 The $U(1)'$ charge for the ingredients of the model is shown

---

1 We need to introduce some additional fermions to cancel gauge anomaly. Since such extensions will be done without changing the following results, we do not go further into this problem here.
in Table 1, in which fermions are assumed to be left-handed. Note that we need only two right-handed neutrinos $N_1$ and $N_2$ to generate appropriate neutrino masses and mixings in a minimal case. Then the invariant Lagrangian relevant to the neutrino masses can be expressed as

$$\mathcal{L}_m = \sum_{\alpha=e,\mu,\tau} (h_{a1} L_a H N_1 + h_{a2} L_a \eta N_2) + \frac{1}{2} M_6 \eta^2 + \frac{1}{2} \lambda \phi \eta^2 + \text{h.c.},$$

where we assume that Yukawa couplings for charged leptons are diagonal. The most general invariant scalar potential up to dimension five may also be written as

$$V = \frac{1}{2} \lambda_1 (H^\dagger H)^2 + \frac{1}{2} \lambda_2 (\eta^\dagger \eta)^2 + \frac{1}{2} \lambda_3 (\phi^\dagger \phi)^2 + \frac{1}{2} \lambda_6 (\phi^\dagger \phi) (\eta^\dagger \eta)^2 + \frac{1}{2} m_\eta^2 \eta^\dagger \eta + \frac{1}{2} m_\phi^2 \phi^\dagger \phi + \text{h.c.}. \quad (1)$$

We add a non-renormalizable $\lambda_6$ term and a bare mass term for $N_1$. The scalar potential (2) without the $\lambda_6$ term has an accidental U(1) symmetry, which forbids the one-loop contribution of the $\eta$ exchange diagram to neutrino masses. This symmetry is explicitly broken by the Yukawa interactions (1), so that terms like the $\lambda_6$ term, i.e. $(\phi^\dagger \phi) (\eta^\dagger \eta)^2$, can be generated in high orders in perturbation theory in general. All of them contribute to radiative neutrino masses.\footnote{It turns out that one-loop corrections generating the $\lambda_6$ term, i.e. $(\phi^\dagger \phi)(\eta^\dagger \eta)H^2$, vanish if the condition (5) discussed later is satisfied.} Here we do not ask the origin of the $\lambda_6$ term. They might be supposed to be effective terms generated through some dynamics at an intermediate scale $M_\ast$. We can check that there are no other dimension five operators invariant under the above mentioned symmetry in the scalar potential.

As the model discussed in\cite{4}, $H$ plays the role of the ordinary doublet Higgs scalar in the SM but $\eta$ is assumed to obtain no vacuum expectation value (VEV). A singlet scalar $\phi$ is assumed to obtain a VEV, which breaks U(1)$'$ down to Z$_2$ (see Table 1). This VEV also gives the mass for $N_2$ through $M_{N_2}^2 = \lambda_6 (\phi)$ and also yields an effective coupling for the $\lambda_6$ term as $\lambda_6 (\phi)/M_\ast$. It can be small enough as long as $\langle \phi \rangle \ll M_\ast$ is satisfied. Thus, the masses of the real and imaginary parts of $\eta^0$ are found to be almost degenerate. They are expressed as $M_{\eta^0}^2 \simeq m_\eta^2 + (\lambda_4 + \lambda_5) (H^0)^2 + \lambda_6 (\phi)^2$. In the model discussed in\cite{4}, the coupling constant of the term corresponding to this $\lambda_6$ term is required to be extremely small to generate appropriate neutrino masses. This point is automatically improved by introducing the new U(1)$'$ symmetry.

### 3. Masses and mixings of neutrinos

We find that there are two origins for the neutrino masses under these settings for the model. One is the ordinary seesaw mass induced by a right-handed neutrino $N_1$\cite{7} and another is one-loop radiative mass mediated by the exchange of $\eta^0$ and $N_2$\cite{5,6}. These effects generate a mass matrix for three light neutrinos. It is expressed by

$$M_\nu = \frac{v^2}{M_\ast} \left[ \mu^{(1)} + \frac{\lambda_6}{8\pi^2}\lambda \left( \frac{M_{N_2}^2}{M_{\eta^0}^2} \right) \mu^{(2)} \right],$$

$$I(x) = \frac{x}{1-x} \left( 1 + \frac{\ln x}{1-x} \right), \quad (3)$$

where $v = \langle H^0 \rangle$ and $\mu^{(a)}$ is defined by

$$\mu^{(a)} = \left( \begin{array}{ccc} h_{\alpha a}^{\tau \alpha} & h_{\alpha a}^{\mu \alpha} & h_{\alpha a}^{\tau \alpha} \\ h_{\alpha a}^{\mu \alpha} & h_{\alpha a}^{\mu \alpha} & h_{\alpha a}^{\mu \alpha} \\ h_{\alpha a}^{\tau \alpha} & h_{\alpha a}^{\tau \alpha} & h_{\alpha a}^{\tau \alpha} \end{array} \right) \quad (a = 1, 2). \quad (4)$$

Although both terms of $M_\nu$ may be characterized by different mass scales, the texture of both terms is the same as found in Eq. (4). This type of the texture for neutrino mass matrix has been studied in\cite{7,8}. We neglect CP phases in the following discussion.

Now we study eigenvalues and the mixing matrix for the neutrino mass matrix (3). We consider to diagonalize $M_\nu$ by using an orthogonal matrix $U$ in such a way as $UT \nu U = \text{diag}(m_1, m_2, m_3)$. If Yukawa couplings satisfy a condition

$$h_{e1} h_{\tau 2} + h_{\mu 2} h_{\mu 1} + h_{\tau 1} h_{\tau 2} \propto [\mu^{(1)}, \mu^{(2)}] = 0, \quad (5)$$

$\mu^{(1)}$ and $\mu^{(2)}$ can be simultaneously diagonalized. Since $U$ can be analytically found in such cases, we confine ourselves to these interesting ones. We define a matrix $\tilde{U}$ as

$$\tilde{U} = \left( \begin{array}{ccc} 1 & 0 & 0 \\ 0 & \cos \theta_2 & \sin \theta_2 \\ 0 & -\sin \theta_2 & \cos \theta_2 \end{array} \right) \left( \begin{array}{ccc} \cos \theta_3 & 0 & \sin \theta_3 \\ 0 & 1 & 0 \\ -\sin \theta_1 & 0 & \cos \theta_3 \end{array} \right). \quad (6)$$

The first term of $M_\nu$ can be diagonalized by this $\tilde{U}$ if the following conditions is satisfied:

$$\tan \theta_2 = \frac{h_{\mu 1}}{h_{\tau 1}}, \quad \tan \theta_3 = \frac{h_{e 1}}{\sqrt{h_{\mu 1}^2 + h_{\tau 1}^2}}. \quad (7)$$

Then the mass eigenvalues for the first term of $M_\nu$ are obtained by using the following eigenvalues of $\mu^{(1)}$:

$$\mu^{(1)} = \text{diag}(0, 0, h_{e 1}^2 + h_{\mu 1}^2 + h_{\tau 1}^2). \quad (8)$$

We consider diagonalization of $\mu^{(2)}$ next. At first, it should be noted that $\mu^{(2)}$ is transformed by the same $\tilde{U}$. However, if the condition (5), which can be written as

$$h_{e 2} \sin \theta_3 + (h_{\mu 2} \sin \theta_2 + h_{e 1} \cos \theta_2) \cos \theta_3 = 0 \quad (9)$$
is satisfied, $\mu^{(2)}$ can be diagonalized by applying an orthogonal transformation $\hat{U}U_3$ supplemented by an additional one given by

$$U_3 = \begin{pmatrix} \cos \theta_1 & \sin \theta_1 & 0 \\ -\sin \theta_1 & \cos \theta_1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{10}$$

This additional transformation by $U_3$ does not affect the diagonalization of $\mu^{(1)}$. Consequently, both terms of $M_\nu$ can be simultaneously diagonalized by setting

$$\tan \theta_1 = -\frac{\tan \tilde{\theta}_2 \tan \theta_2 + 1}{(\tan \tilde{\theta}_2 - \tan \theta_2) \sin \tilde{\theta}_2}, \tag{11}$$

where we define $\tilde{\theta}_2$ as $\tan \tilde{\theta}_2 = h_{\mu 2}/h_{\tau 2}$. Finally, we obtain non-zero mass eigenvalues of the light neutrinos as

$$m_2 = AB \frac{\tan^2 \theta_1 + 1}{\tan^2 \tilde{\theta}_2 + 1} (\tan \tilde{\theta}_2 - \tan \theta_2)^2,$$

$$m_3 = A \left(\frac{\tan^2 \theta_2 + 1}{\tan^2 \tilde{\theta}_2 + 1}\right) (\tan^2 \tilde{\theta}_2 + 1), \tag{12}$$

where $A = 2h_{\tau 4}v^2/M_\tau$ and $B = (\lambda_6/16\pi^2 \lambda)(h_{\tau 2}/h_{\tau 1})^2 \times I(M_{N2}/M_{\tilde{\nu}_\mu})$.

Here we fix $\tan \theta_2 = 1$ which is supported by the data of the atmospheric neutrino and K2K experiment. CHOOZ experiments give the constraint on $\theta_3$ such as $|\sin \theta_3| < 0.22$ [9]. If we use these conditions, the mixing matrix $U = \hat{U}U_3$ can be approximately written as

$$U = \begin{pmatrix} \cos \theta_1 & \sin \theta_1 & \sin \theta_3 \\ -\sin \theta_1 & \cos \theta_1 & \cos \theta_3 \\ \sin \theta_1 & \cos \theta_1 & -\cos \theta_3 \end{pmatrix}. \tag{13}$$

Only two mass eigenvalues $m_2$ are non-zero and then we impose that squared mass differences required by the neutrino oscillation data satisfy $m_3 = \sqrt{\Delta m^2_{\text{atm}}}$ and $m_2 = \sqrt{\Delta m^2_{\text{sol}}}$. Although there is a possibility that two non-zero eigenvalues have almost degenerate values such as $\sqrt{\Delta m^2_{\text{atm}}}$ and their squared difference is given by $\Delta m^2_{\text{sol}}$, we do not consider it since $\mu^{(1)}$ and $\mu^{(2)}$ have independent origins. We suppose $\theta_1 = \theta_{\text{sol}}$, where $\theta_{\text{sol}}$ is a mixing angle relevant to the solar neutrino. Then we can determine $\theta_3$ through Eq. (11) by using $\tan \theta_{\text{sol}}, \sqrt{\Delta m^2_{\text{atm}}}$, $\sqrt{\Delta m^2_{\text{sol}}}$ and $B$. If we use neutrino oscillation data for these, we can find allowed regions of $\theta_3$ as a function of $B$. This is shown in Fig. 1, where we have used values of the measured neutrino oscillation parameters [10]

$$\Delta m^2_{\text{sol}} = 8.0^{+0.6}_{-0.4} \times 10^{-5} \text{ eV}^2,$$

$$\Delta m^2_{\text{atm}} = (1.9–3.6) \times 10^{-3} \text{ eV}^2,$$

$\tan^2 \theta_{\text{sol}} = 0.45^{+0.09}_{-0.07}. \tag{14}$

This figure shows that $B$ is restricted in narrow regions such as $0.03 < B < 0.1$.

As an example, let us assume $M_{\tilde{\nu}_\mu}/M_{N2} = 0.3–0.7$ and then $I(M_{N2_2}/M_{\tilde{\nu}_\mu}) = 0.1–1.3$. In such cases $h_{\tau 2}/h_{\tau 1} \simeq 10(\lambda/\lambda_6)^{1/2}$ should be satisfied. If we obtain more constraints on the relevant coupling constants, we may restrict the value of $U_{\nu 3}$ much more. Although $U_{\nu 3}$ takes a non-zero value for $0.03 < B < 0.05$ and $0.08 < B < 0.1$, $U_{\nu 3} = 0$ is also allowed for $0.03 < B < 0.08$. The condition for the coupling constants can be easily satisfied even if we assume that coupling constants are $O(1)$. Therefore, the model needs no fine tuning to be consistent with all the present experimental data for neutrino oscillations. The effective mass $m_{ee}$ for the neutrinoless double beta decay takes the values in the range $|m_{ee}| \lesssim 6.3 \times 10^{-3}$ eV.

### 4. Relic abundance of a CDM candidate

The lightest field with an odd $Z_2$ charge can be stable since an even charge is assigned to each SM content. If both the mass and the annihilation cross section of such a field have appropriate values, it can be a good CDM candidate as long as it is neutral. As found from Table 1, such candidates are $N_2$ and $\eta^0$. Since they have a new U(1)' gauge interaction, their annihilation to quarks is considered to be dominantly mediated by this interaction. If their annihilation is mediated only by the exchange of $\eta^0$ or $N_2$ through Yukawa couplings as in the model discussed in [4], we cannot simultaneously explain, without fine tuning of coupling constants, both the observed value of the CDM abundance and the constraints coming from lepton flavor violating processes such as $\mu \to e\gamma$. Since $U(1)'$ is supposed to be a generation independent gauge symmetry, we can easily escape this problem by assuming that the Yukawa couplings $h_{a2}$ are small enough or both $\eta^0$ and $N_2$ are heavy enough. In the following study we consider the case that $N_2$ is lighter than $\eta^0$.

As seen in the last part of the previous section, this case is consistent with the present experimental bounds for $U_{\nu 3}$ without fine tuning.

Now we estimate the relic abundance of $N_2$ and compare it with the CDM abundance obtained from the WMAP data. We

---

3 A role of U(1)' in annihilation of the CDM in supersymmetric models has been studied in [13].
suppose that possible annihilation processes $N_2N_2 \rightarrow f \bar{f}$ are dominantly mediated by the U(1)' gauge field. If it is expanded by relative velocity $v$ between annihilating $N_2$'s as $\sigma v = a + bv^2$, the coefficients $a$ and $b$ are expressed as

$$a = \sum_f c_f \frac{g^4}{2\pi} Q^2_{fA} q^2 \frac{m^2_N \beta}{(s - M^2_N)^2},$$

$$b = \sum_f c_f \frac{g^4}{6\pi} (Q^2_{fV} + Q^2_{fA}) q^2 \frac{M^2_N \beta}{(s - M^2_N)^2},$$

where $\beta = \sqrt{1 - m^2/\sqrt{M^2_N}}$ and $c_f = 3$ for quarks, $s$ is the center of mass energy of collisions and $q$ is the U(1)' charge of $N_2$ given in Table 1. The charge of the final state fermion $f$ is defined as

$$Q_{fV} = Q_{fR} + Q_{fL}, \quad Q_{fA} = Q_{fR} - Q_{fL}. \tag{15}$$

Using these quantities, the present relic abundance of $N_2$ can be estimated as [11],

$$\Omega_{N_2} h^2 \bigg|_{0} = \frac{M_{N_2} r_{N_2}}{\rho_c / h^2} \simeq \frac{8.76 \times 10^{-11} g_*^{-1/2} x_F}{(a + 3b/x_F)} \text{GeV}^2, \tag{16}$$

where $g_*$ enumerates the degrees of freedom of relativistic fields at the freeze-out temperature $T_F$ of $N_2$. $T_F$ is determined through the equation for a dimensionless parameter $x_F = M_{N_2} / T_F$

$$x_F = \ln \frac{0.0955 m_{pl} M_{N_2}(a + 6b/x_F)}{(g_* x_F)^{1/2}}, \tag{17}$$

where $m_{pl}$ is the Planck mass. If we fix the U(1)' charge of fields and its coupling constant $g'$, we can estimate the present $N_2$ abundance using these formulas. Assuming a GUT relation $g' = \sqrt{5/3} g_R$ and $q = 0.6$ as an example, we calculate $\Omega_{N_2} h^2$. The results are given in Fig. 2.

In the left figure of Fig. 2 we plot favorable regions in the $(M_{Z'}, M_{N_2})$ plane, where $\Omega_{N_2} h^2$ takes values in the range 0.0945–0.1285, which is required by the WMAP data. $\Omega_{N_2} h^2$ has a valley in the parameter region of Fig. 2, and therefore the allowed regions appear as two narrow bands, each sandwiched by a solid line and a dashed line. Since $M_{N_2}$ and $M_{Z'}$ are induced through the mediation of $\eta'$ and written as

$$M_{N_2} = \lambda \langle \phi \rangle, \quad M_{Z'} = 2\sqrt{2} g' q \langle \phi \rangle, \tag{18}$$

$M_{N_2}$ is determined by $M_{Z'}$. We plot this $M_{N_2}$ values by green dotted lines for $\lambda = 0.2$ and 0.7. The lower bounds of $M_{Z'}$ come from constraints for $ZZ'$ mixing and direct search of $Z'$. $H$ is assumed to have no U(1)' charge and then its VEV induces no $ZZ'$ mixing. Moreover, since it is leptophobic, the constraints on $M_{Z'}$ obtained from its hadronic decay is rather weak. Thus, the lower bounds of $M_{Z'}$ may be $M_{Z'} \gtrsim 450$ GeV in the present model [14]. Taking account of this, Fig. 1 shows that this model can well explain the CDM abundance. Since $\lambda$ is included in the definition of $B$, values of $\beta_1$ may be constrained by the mass of the CDM if we can obtain more informations on $\lambda_6, h_{\tau 2}/h_{\tau 1}$ and $M_{Z'}$.

Here we briefly discuss the relation to lepton flavor violating processes such as $\mu \rightarrow e\gamma$. As in the model of [15], $\mu \rightarrow e\gamma$ is induced through the mediation of $\eta'$ and $N_2$. Its branching ratio can be given by

$$B(\mu \rightarrow e\gamma) = \frac{3\alpha}{64\pi (G_F M^2_{\eta'})^2} \left| h_{\mu 2} h_{e 2} F_2 \left( \frac{M^2_{N_2}}{M^2_{\eta'}} \right) \right|^2,$$

$$F_2(x) = \frac{1}{6(1 - x)^6} (1 - 6x + 3x^2 + 2x^3 - 6x^2 \ln x). \tag{19}$$

Taking account that $1/12 < F_2(x) < 1/6$ is satisfied in case of $M_{N_2} < M_{\eta'}$ and imposing the present experimental upper bound $B(\mu \rightarrow e\gamma) \lesssim 1.2 \times 10^{-11}$, we find that $M_{\eta'}$ should satis-
\( M_{\rho} \gtrsim (360-500) \left( \frac{h_{\tau 2}}{0.1} \right) \) GeV. \tag{21}

Here we use the results of the previous section. Constraints coming from \( \mu \rightarrow e\gamma \) and the CDM abundance can be consistent for reasonable values of \( h_{\tau 2} \). Since \( N_2 \) annihilation due to an \( h^0 \) exchange is ineffective for these values of couplings and masses \([4]\), the results of the \( N_2 \) abundance given above is not affected by this process.

Finally, it may be useful to refer to the cases of general \( U(1)' \). In these cases a crucial condition for the mass of the \( U(1)' \) gauge field comes from the constraint for \( ZZ' \) mixing. A mass matrix for neutral gauge bosons can be expressed as

\[
\begin{pmatrix}
\frac{1}{2}(g_1^2 + g_2^2)v^2 & -g'\sqrt{g_1^2 + g_2^2}q Hv^2 \\
-g'\sqrt{g_1^2 + g_2^2}q Hv^2 & 2g^{'2}q_3^2(4\phi^2 + v^2)
\end{pmatrix},
\tag{22}
\]

where \( q_H \) and \( q_\phi \) stand for the \( U(1)' \) charge of \( H \) and \( \phi \). Since a \( ZZ' \) mixing angle \( \theta \) is known to be strongly suppressed \([12]\), the magnitude of \( \langle \phi \rangle \) should satisfy

\[
\langle \phi \rangle \gtrsim \frac{v}{2} \left( \frac{g_1^2 + g_2^2}{2 (2g^{'2}q_3^2)} \right)^{1/4}.
\tag{23}
\]

This condition gives a lower bound on both \( M_{Z'} \) and \( M_{N_2} \). In the right panel of Fig. 2 we plot this bound in cases of \( |\theta| = 10^{-2}, 5 \times 10^{-3}, 10^{-3} \), which are drawn by vertical dash-dotted lines. We also plot the values of \( M_{N_2} \) for \( \lambda = 0.2, 0.3, 0.6 \) and 0.9. They are drawn by green dotted lines. Although \( |\theta| \) should be less than \( 10^{-3} \), we may suppose larger values of \( |\theta| \) in Eq. (23) by extending the model without changing the results in the previous section. In fact, if the model has two Higgs doublets \( H_u \) and \( H_d \) which couple to up- and down-sectors respectively, off-diagonal elements of Eq. (22) is proportional to \( g'q_{H_u}(H_u)^2 - q_{H_d}(H_d)^2 \) where \( q_{H_{u,d}} \) expresses the \( U(1)' \) charge. Cancellation between these two contributions can make the \( ZZ' \) mixing smaller for the same value of \( \langle \phi \rangle \). In such cases we can apply this effect by using larger \( |\theta| \) values in Eq. (23). In this figure \( \theta \) values larger than \( 10^{-3} \) should be understood based on this reasoning.

On the other hand, the introduction of additional Higgs doublets may require us to take account of new final states for the \( N_2 \) annihilation induced by the \( Z' \) exchange. If \( N_2 \) is heavier than \( W_\pm \), the final states should include gauge bosons and Higgs scalars such as \( W^+ W^- \), \( H_1^0 H_0^0 \), \( W^\pm H_\mp \), \( H^+ H^- \) and \( Z H_0^0 \) where \( H_1^0 \) is a mass eigenstate of the neutral Higgs. Since the annihilation to \( W^+ W^- \) is suppressed by the \( ZZ' \) mixing in the present model, important modes are expected to be \( H_1^0 H_0^0 \) and they may give the same order of contributions as the annihilation to \( f \bar{f} \) \([11]\). In order to take such effects into account without practicing tedious estimation of such processes, we show in the right figure of Fig. 2 an additional \( \Omega_{N_2}h^2 \) contour which is obtained by using \( 5 \times (\sigma v)_{f \bar{f}} \) for cross section. It is drawn by blue lines. An original contour for the cross section \( (\sigma v)_{f \bar{f}} \) is drawn by red lines.\(^4\) Since main parts of the cross section into these final states are expected to have the similar dependence on \( M_{Z'} \) and \( M_{N_2} \), this is considered to give good references for these cases. This figure also suggests that this kind of models can explain the CDM abundance even under the constraint for \( Z' \) physics.

5. Summary

We have studied neutrino masses and CDM abundance in a non-supersymmetric, but \( U(1)' \) symmetric model which is obtained from the SM by adding certain neutral fields. Neutrino masses are generated through both the seesaw mechanism with a single right-handed neutrino and the one-loop radiative effects. They induce the same texture which can realize favorable mass eigenvalues and mixing angles. One of the introduced neutral fields is stable due to an unbroken \( Z_2 \) symmetry which is the residual symmetry of the spontaneously broken \( U(1)' \). Thus it can be a good CDM candidate. Since it has the \( U(1)' \) gauge interaction, the annihilation is dominantly mediated through this interaction. If this \( U(1)' \) symmetry is broken at a suitable scale, the present relic abundance of right-handed neutrinos can explain the WMAP result for the CDM abundance. This model suggests that two of the biggest questions in the SM, that is, neutrino masses and the CDM may be explained on the common basis of an extension of the SM. An interesting feature of the model is that the value of the third mixing angle \( \theta_3 \) may be related to the mass of the CDM. The model may be examined through the search of the \( Z' \) and the additional Higgs doublet \( \eta \) at LHC.

Acknowledgements

This work is partially supported by a Grant-in-Aid for Scientific Research (C) from Japan Society for Promotion of Science (Nos. 17540246 and 18540257). We would like to thank the referee of Physics Letters B for pointing out that the originally proposed \( Z_2 \) is redundant because \( Z_2 \) is nothing but the residual symmetry of \( U(1)' \).

References


H. Baer, C. Balazs, JCAP 0305 (2003) 006;  
S. Profumo, C.E. Yaguna, hep-ph/0407036;  
E.A. Baltz, P. Gondolo, hep-ph/0407039;  

E. Ma, hep-ph/0607142;  


Majorana neutrinos and lepton-number-violating signals in top-quark and $W$-boson rare decays

Shaouly Bar-Shalom a,*, Nilendra G. Deshpande b, Gad Eilam a, Jing Jiang b, Amarjit Soni c

a Physics Department, Technion—Institute of Technology, Haifa 32000, Israel
b Institute for Theoretical Science, University of Oregon, OR 97403, USA
c Theory Group, Brookhaven National Laboratory, Upton, NY 11973, USA

Received 2 September 2006; received in revised form 26 October 2006; accepted 30 October 2006
Available online 13 November 2006

Abstract

We discuss rare lepton-number-violating top-quark and $W$-boson four-body decays to final states containing a same-charge lepton pair, of the same or of different flavors: $t \rightarrow b W^{-} \ell_{i}^{+} \ell_{j}^{-}$ and $W^{+} \rightarrow J J^{+}$, where $i \neq j$ or $i = j$ and $J J^{+}$ stands for two light jets originating from a $u d$ or a $c \bar{c}$ pair. These $\Delta L = 2$ decays are forbidden in the Standard Model and may be mediated by exchanges of Majorana neutrinos. We adopt a model independent approach for the Majorana neutrinos mixing pattern and calculate the branching ratios (BR) for these decays. We find, for example, that for $\mathcal{O}(1)$ mixings between heavy and light Majorana neutrinos (not likely but not ruled out) and if at least one of the heavy Majorana neutrinos has a mass of $\lesssim 100$ GeV, then the BR’s for these decays are: $\text{BR}(t \rightarrow b \ell_{i}^{+} \ell_{j}^{-} W^{-}) \sim 10^{-6}$ and $\text{BR}(W^{+} \rightarrow \ell_{i}^{+} \ell_{j}^{+} J J^{+}) \sim 10^{-7}$ if $m_{N} \sim 100$ GeV and $\text{BR}(t \rightarrow b \ell_{i}^{+} \ell_{j}^{-} W^{-}) \sim 10^{-6}$ and $\text{BR}(W^{+} \rightarrow \ell_{i}^{+} \ell_{j}^{+} J J^{+}) \sim 10^{-10}$ for $m_{N} \sim 100$ GeV and $\text{BR}(t \rightarrow b \ell_{i}^{+} \ell_{j}^{-} W^{-}) \sim 10^{-6}$ and $\text{BR}(W^{+} \rightarrow \ell_{i}^{+} \ell_{j}^{+} J J^{+}) \sim 10^{-10}$ for $m_{N} \sim 100$ GeV and $\text{BR}(t \rightarrow b \ell_{i}^{+} \ell_{j}^{-} W^{-}) \sim 10^{-6}$ and $\text{BR}(W^{+} \rightarrow \ell_{i}^{+} \ell_{j}^{+} J J^{+}) \sim 10^{-10}$ for $m_{N} \sim 100$ GeV and $\text{BR}(t \rightarrow b \ell_{i}^{+} \ell_{j}^{-} W^{-}) \sim 10^{-6}$ and $\text{BR}(W^{+} \rightarrow \ell_{i}^{+} \ell_{j}^{+} J J^{+}) \sim 10^{-10}$ for $m_{N} \sim 100$ GeV and $\text{BR}(t \rightarrow b \ell_{i}^{+} \ell_{j}^{-} W^{-}) \sim 10^{-6}$ and $\text{BR}(W^{+} \rightarrow \ell_{i}^{+} \ell_{j}^{+} J J^{+}) \sim 10^{-10}$ for $m_{N} \sim 100$ GeV and $\text{BR}(t \rightarrow b \ell_{i}^{+} \ell_{j}^{-} W^{-}) \sim 10^{-6}$ and $\text{BR}(W^{+} \rightarrow \ell_{i}^{+} \ell_{j}^{+} J J^{+}) \sim 10^{-10}$ for $m_{N} \sim 100$ GeV and $\text{BR}(t \rightarrow b \ell_{i}^{+} \ell_{j}^{-} W^{-}) \sim 10^{-6}$ and $\text{BR}(W^{+} \rightarrow \ell_{i}^{+} \ell_{j}^{+} J J^{+}) \sim 10^{-10}$ for $m_{N} \sim 100$ GeV.

© 2006 Elsevier B.V. All rights reserved.

The recent discovery of neutrino oscillations which indicates mixing between massive neutrinos [1], was a major turning point in modern particle physics, since it stands as the first direct evidence for physics beyond the Standard Model (SM). Thus, it is now clear that the SM has to be expanded to include massive neutrinos that mix. Since there is still no understanding of the nature of these massive neutrinos, i.e., Majorana or Dirac-like, the extension of the SM can basically go either way. In particular, a simple way to consistently include sub-eV massive Majorana neutrinos in the SM is to add superheavy right-handed neutrinos with GUT-scale masses and to rely on the seesaw mechanism [2], which yields the desired light neutrinos mass scale: $m_{\nu} \sim M_{EW}^{2}/M_{GUT} \sim 10^{-2}$ eV, $M_{EW}$ being the electroweak (EW) scale. The seesaw mechanism, therefore, links neutrino masses with new physics at the GUT-scale, which is well motivated theoretically. On the other hand, a simple way to include massive Dirac neutrinos within the SM is to add Higgs-neutrinos Yukawa couplings which are more than 8 orders of magnitude smaller than the Higgs-electron one. Consequently, the Yukawa couplings of fermions (in the SM) unnaturally span over more than 13 orders of magnitude. Thus, within these simple extensions to the SM, the Majorana neutrinos seem to be favored from the theoretical point of view.

The fact that a Majorana mass term violates lepton number by two units, i.e. $\Delta L = \pm 2$, has dramatic phenomenological signatures that can be used to distinguish Majorana neutrinos from Dirac neutrinos within many extensions of the SM. The most extensively studied process is neutrinoless double beta decay $(A, Z) \rightarrow (A, Z + 2) + e^{-} + e^{-}$ [3]. Also interesting are the $\Delta L = 2$ lepton-number-violating (LNV) processes in various high-energy collisions such as: $e^{-} \gamma$ [4,5], $pp$ and $p \bar{p}$ [5–9],

* Corresponding author.
E-mail addresses: shaouly@physics.technion.ac.il (S. Bar-Shalom), desh@uoregon.edu (N.G. Deshpande), eilam@physics.technion.ac.il (G. Eilam), jing@uoregon.edu (J. Jiang), soni@bnl.gov (A. Soni).

0370-2693/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.physletb.2006.10.060
In what follows we will, therefore, focus on the effect of the neutrino sector, see e.g., [4] (in what follows $n$ means $n_{\alpha}$). In the simplest scenario which relies on the classic seesaw mechanism [2], the couplings of heavy neutrinos ($N$) to SM particles, e.g., $B_{iN}$ in (3), are highly suppressed by $\sqrt{m_{\nu}/m_{N}}$, where $m_{\nu}$ is the mass of the light neutrinos typically of the order of the solar or atmospheric neutrino masses. However, the possibility of non-seesaw realizations or internal symmetries in the neutrino sector, that may decouple the heavy-to-light neutrino mixing from the neutrino masses, cannot be excluded [13]. This motivates us to adopt a purely phenomenological approach by assuming no a-priori relation between the mixing angles in $B_{i\alpha}$ and the neutrino masses. Within such a model independent approach, the elements in $B_{i\alpha}$ need only be bounded by existing model independent experimental constraints. For example, the 95% CL mass limits from LEP are $m_{N} \gtrsim 80–90$ GeV, depending on whether it couples to an electron, muon or a tau [14]. For such heavy Majorana neutrinos and assuming the dominance of only one heavy neutrino $N$ (see discussion below), the limits on its couplings to the charged leptons can be expressed in terms of the products $\Omega_{\ell\ell'} \equiv B_{iN}B_{j\ell'N}$ (see [4] and references therein).

In particular, the limits on its flavor-diagonal couplings come from precision electroweak data, and at 90% CL are [15]:

$$\Omega_{ee} \lesssim 0.012, \quad \Omega_{\mu\mu} \lesssim 0.0096, \quad \Omega_{\tau\tau} \lesssim 0.016,$$

while the limits on its flavor-changing couplings come from limits on rare flavor-violating lepton decays such as $\mu \to e\gamma$, $\mu, \tau \to e\gamma\gamma$ [4]:

$$|\Omega_{e\mu}| \lesssim 0.0001, \quad |\Omega_{e\tau}| \lesssim 0.02, \quad |\Omega_{\mu\tau}| \lesssim 0.02.$$ 

Using (3), the kernel amplitude is given by (including both $t$ and $u$-channel diagrams):

$$iM(W^{-}W^{+} \to \ell_{i}^{+}\ell_{j}^{+}) = W_{-\mu}W_{-\nu}^{2} B_{i\alpha} B_{j\alpha} m_{\nu} n_{\mu} \gamma_{\nu} (1 + \gamma_{5})$$

$$\times \left[ \frac{\gamma_{\mu} \gamma_{\nu}}{p_{n(\mu)}^{2} - m_{n}^{2} + i m_{n} \Gamma_{n}^{\nu}} - \frac{\gamma_{\nu} \gamma_{\mu}}{p_{n(\mu)}^{2} - m_{n}^{2} + i m_{n} \Gamma_{n}^{\nu}} \right] \Gamma_{1},$$

where $p_{n(\mu)} \neq p_{\ell} - p_{\nu}$, $p_{\nu} = p_{W} - p_{\ell}$, $m_{n}$ and $\Gamma_{n}$ are the Majorana neutrino $t$- and $u$-channel 4-momenta, mass and total width, respectively. From (6) it is evident that for $m_{n}^{2} \gg m_{W}^{2}$ (recall that for the top-quark and $W$-boson decays the momentum transfer scale is of $O(m_{W})$), the amplitude is proportional to $B_{i\alpha} B_{j\alpha} m_{\nu} / m_{n}^{2}$. On the other hand, for the sub-eV light neutrinos, i.e., $m_{\nu}$ of order of the solar and atmospheric mass scales, the kernel amplitude is proportional to $B_{i\alpha} B_{j\alpha} m_{\nu} / m_{W}^{2}$. Thus, these decays are far dominated by the exchanges of the heavy Majorana neutrinos, if their masses are of $O(m_{W})$. In what follows, we will, therefore, focus on the effect of the heavy Majorana neutrinos, with a further simplifying assumption that the kernel amplitude is dominated by an exchange of only one heavy Majorana neutrino, $N$, which maximizes the quantity $B_{iN} B_{j\alpha} n_{\nu} / m_{N}$, either because it is much lighter than...
the other heavy neutrinos or because its mixing with the left-handed neutrinos is much larger, i.e., $B_{1iN} \ll B_{ijN}$ for $N' \neq N$.

For the dominant decays of $N$ we take $N \to \ell_{k}^{\pm}W^{\mp}, \nu_{k}Z, \nu_{k}H$, where $\nu_{k}, k = 1–3$, are the three light sub-eV neutrinos. The partial widths for these decay channels are given by (see e.g., [4]):

$$
\sum_{k} \Gamma(N \to \ell_{k}^{\pm}W^{\mp}) \approx C(m_{N}^{2} + 2m_{W}^{2})(m_{N}^{2} - m_{W}^{2})^{2},
$$

$$
\sum_{k} \Gamma(N \to \nu_{k}Z) \approx C(m_{N}^{2} + 2m_{Z}^{2})(m_{N}^{2} - m_{Z}^{2})^{2},
$$

$$
\sum_{k} \Gamma(N \to \nu_{k}H) \approx Cm_{N}^{2}(m_{N}^{2} - m_{H}^{2})^{2},
$$

(7)

where

$$
C \equiv \frac{g^{2}}{64\pi m_{W}^{2}m_{N}^{2}} \sum_{k} |B_{kN}|^{2}.
$$

(8)

We note that our results depend very weakly on $m_{H}$. Nonetheless, for definiteness, we will set $m_{H} = 120$ GeV throughout our analysis. Also, the widths for the partial decays $N \to \nu_{k}Z$ and $N \to \nu_{k}H$ depend on the neutral couplings $C_{\nu N}$ which appear in the interaction terms of $Z$ and $H$ with a pair of Majorana neutrinos. In Eq. (7) we have used the approximate relation between the charged and neutral couplings of $N$ to the gauge-bosons: $\sum_{k} |B_{kN}|^{2} \approx \sum_{k} |C_{\nu N}|^{2}$, see e.g., [4].

Let us now define a generic “reduced” amplitude squared:

$$
\bar{\mu}_{nn'}^{2} \equiv \frac{1}{\text{pol}} \sum_{\text{pol}} \mu_{n}^{\dagger} \mu_{n'},
$$

(9)

where pol is the number of polarization states of the decaying particle (pol = 2 and pol = 3 for the top-quark and the $W$-boson decays, respectively), $\mu_{n}$ is the top-quark or $W$-boson decay amplitude for an exchange of a Majorana neutrino $n$ (see Fig. 1), and $n, n' = 1–6$ are indices of the six Majorana neutrino states.

Then, using (9) and summing over all intermediate Majorana neutrino states, we obtain the total amplitude squared:

$$
|\bar{\mathcal{M}}|^{2} \equiv \sum_{n=1}^{6} \mu_{nn'}^{2} + \sum_{n<n'} 2 \text{Re}(\mu_{nn'}^{2}).
$$

(10)

and the decay width for either the top-quark in (1) or the $W$-boson in (2):

$$
\Gamma = \frac{(1 - \frac{\alpha}{2})}{2M(2\pi)^{3/2}} \int \frac{d^{3}p_{k}}{2E_{k}} \delta^{\mu}(P - \sum_{k=1}^{4} p_{k}) |\bar{\mathcal{M}}|^{2},
$$

(11)

where $M$ and $P$ are the mass and 4-momentum of the decaying particle, $i, j$ are flavor indices of the lepton pair in the final state of both decays and $p_{k}$ are the momenta of the final state particles. The reduced amplitude squared for the top-quark ($\bar{\mu}(t)^{2}_{nn'}$) and for the $W$-boson decays ($\bar{\mu}(W)^{2}_{nn'}$) are given by (neglecting the masses of the final state fermions)

$$
\bar{\mu}(t)^{2}_{nn'} = 8A_{nn'}^{ij} |p_{i} \cdot p_{j}| \left[ 1 + \frac{m_{t}^{2}}{m_{W}^{2}} \right] \frac{1}{2m_{W}^{2}} \left[ \frac{1}{2}p_{t} \cdot p_{W} \right] p_{b} \left( \frac{(p_{W} \cdot p_{W}^{*})^{2}}{m_{W}^{2}} - \frac{4}{m_{W}^{2}} \right)
$$

$$
+ \frac{m_{t}^{2}}{2m_{W}^{2}} \left[ p_{b} \cdot p_{W}^{*} - \frac{1}{m_{W}^{2}} p_{b} \cdot p_{W} \cdot p_{W} \cdot p_{W}^{*} \right],
$$

$$
\bar{\mu}(W)^{2}_{nn'} = \frac{16}{3} A_{nn'}^{ij} |p_{i} \cdot p_{j}| \left( p_{f} \cdot p_{f}^{*} \right)
$$

$$
+ \frac{2}{m_{W}^{2}} p_{f} \cdot p_{f} \cdot p_{W} \cdot p_{W}^{*},
$$

(12)

where

$$
A_{nn'}^{ij} = \left( \frac{g}{\sqrt{2}} \right)^{6} |\Pi_{W}^{*}|^{2} B_{ii} B_{jj} B_{iN} B_{jN} m_{i} m_{j} \Pi_{N} \Pi_{N}^{*}.
$$

(13)

and $\Pi_{x} = (p_{x} - m_{x} + i\Gamma_{x})^{-1}$, where $p_{x}, m_{x}$ and $\Gamma_{x}$ are the 4-momentum, mass and width of the particle $x$, respectively. Also, $p_{W}^{*}$ which appears in (12) and (13) is the 4-momentum of the virtually exchanged $W$-boson in both the top-quark and $W$-boson decays (see Fig. 1).

Note that, within our assumption of a single-$N$ dominated amplitude, we get:

$$
|\bar{\mathcal{M}}|^{2} = \bar{\mu}_{NN}^{2},
$$

$$
A_{NN}^{ij} = \left( \frac{g}{\sqrt{2}} \right)^{6} |\Pi_{W}^{*}|^{2} |\Pi_{N}|^{2} m_{N}^{2} |B_{iN}|^{2} |B_{jN}|^{2}.
$$

(14)

We will first consider the case $m_{N} > m_{W}$ and then discuss the implications of a “light” $m_{N}, m_{N} < m_{W}$, on the top and $W$ decays under investigation. In Fig. 2 we plot the BR’s for both the top-quark and the $W$-boson decays, scaled by the neutrino mixing parameters, i.e., setting $B_{1iN} = B_{ijN} = 1$, as a function of the Majorana neutrino mass, $m_{N}$, in the mass range $m_{N} > m_{W}$. We see that, for both decays, a sizable and experimentally accessible BR can arise only for $m_{N}$ values around 100 GeV, for which we obtain:

$$
\text{BR}(t \to bW^{-}e_{i}^{+}e_{j}^{+}) |B_{iN}|^{2} |B_{jN}|^{2} \approx 10^{-4},
$$

(15)

$$
\text{BR}(W^{\pm} \to J_{F}^{\pm}e_{i}^{\mp}e_{j}^{\mp}) |B_{iN}|^{2} |B_{jN}|^{2} \approx 10^{-7}.
$$

(16)

To obtain more realistic BR’s we can use the bounds on the neutrino mixing couplings in (4) and (5). For the $W$-boson decay (this decay is essentially insensitive to the heavy neutrino width), the largest BR subject to the constraints in (4) and (5) is of order of $10^{-10}$. This is too small to be observed at the large hadron collider (LHC), where about $10^{9}$–$10^{10}$ inclusive on-shell $W$’s are expected to be produced through $pp \to W + X$, at an integrated luminosity of $O(100)$ fb$^{-1}$ [16].

However, as was shown in [9], for on-shell production of $N$ via $ud \to W^{\pm} \to \ell N$, the sensitivity to the heavy Majorana neutrino can be significantly enhanced. Indeed, in this case, the s-channel $W^{\ast}$ “decays”, as in (2), to $W^{\ast} \to J_{F}^{\ast}e_{i}^{\pm}e_{j}^{\pm}$, by first
where, for $m_N < m_t$, the BR$(t \rightarrow b W^+ \ell_i^+ \ell_j^+)$ can be approximated by:

$$\text{BR}(t \rightarrow b W^- \ell_i^- \ell_j^+) \approx \text{BR}(t \rightarrow b \ell_i^+ N) \times \text{BR}(N \rightarrow W^- \ell_i^-) \times \delta(i \leftrightarrow j) \quad (i \neq j), \quad (17)$$

where, for $m_N \sim 100$ GeV we obtain:

$$\frac{\text{BR}(t \rightarrow b \ell_i^+ N)}{|B_{iN}|^2} \sim 10^{-4}, \quad (18)$$

and

$$\text{BR}(N \rightarrow W^- \ell_j^+) \sim 0.5 \times \frac{|B_{jN}|^2}{|B_{iN}|^2 + |B_{jN}|^2}. \quad (19)$$

We recall that the cross-section for $t \bar{t}$ production at the LHC is $\sim 850$ pb [16], yielding about $10^8 \, t \bar{t}$ pairs at an integrated luminosity of $\mathcal{O}(100)$ fb$^{-1}$. Thus, a BR$(t \rightarrow b W^- \ell_i^- \ell_j^-) \sim 10^{-6}$ that can arise in most $\ell_i^- \ell_j^+$ channels (see Table 1), should be accessible at the LHC. In particular, the flavor conserving channels $t \rightarrow b W^- e^+ e^+$ and $t \rightarrow b W^- \mu^+ \mu^+$ are expected to be more effective, since the channels involving the $\tau$-lepton will suffer from a low $\tau$ detection efficiency.

Let us now consider the case of a lighter $N$ with a mass $m_N < m_W$. Such a “light” Majorana neutrino is not excluded by LEP data if its couplings/mixings with the SU(2) leptonic doublets are small enough [17]. For example, $N$ can have a mass in the range $5 \text{ GeV} \lesssim m_N \lesssim 50 \text{ GeV}$ if $|B_{iN}|^2 \sim 10^{-4}$. In this mass range the 5-body cascade top decay $t \rightarrow b W^+ \rightarrow b \ell_i^+ N \rightarrow b \ell_i^+ \ell_j^+ J J'$ (recall that $J J'$ stands for a pair of light jets originating from a $u \bar{d}$ or $c \bar{s}$ pair) has a much larger width than the 4-body decay $t \rightarrow b \ell_i^+ \ell_j^- W^-$, since the intermediate $N$ cannot decay to an on-shell $W$. Thus, for $m_N < m_W$ both $t \rightarrow b \ell_i^+ \ell_j^+ J J'$ and $W^+ \rightarrow \ell_i^+ \ell_j^+ J J'$ originate from the cascade decay $W^+ \rightarrow \ell_i^+ N$ followed by $N \rightarrow \ell_j^+ J J'$ and, for BR$(t \rightarrow b W^+) \sim 1$, they have equal branching ratios since:

$$\text{BR}(t \rightarrow b \ell_i^+ \ell_j^+ J J') \sim \text{BR}(t \rightarrow b W^+) \times \text{BR}(W^+ \rightarrow \ell_i^+ N) \times \text{BR}(N \rightarrow \ell_j^+ J J') + (i \leftrightarrow j) \quad (i \neq j).$$

$$\text{BR}(W^+ \rightarrow \ell_i^+ \ell_j^+ J J') \sim \text{BR}(W^+ \rightarrow \ell_i^+ N) \times \text{BR}(N \rightarrow \ell_j^+ J J') + (i \leftrightarrow j) \quad (i \neq j). \quad (20)$$

where the partial width for $W^+ \rightarrow \ell_i^+ N$ is:

$$\Gamma(W^+ \rightarrow \ell_i^+ N) = \frac{g^2}{96\pi} |B_{iN}|^2 m_W \left(2 - \frac{3m_N^2}{m_W^2} + \frac{m_N^2}{m_W^2}\right). \quad (21)$$

![Figure 2](image-url)
where $\Gamma_{N}(Z)\equiv BR(t\rightarrow b\ell^{-}_{i}\ell^{+}_{j}J^{\mp})/|B_{i}|^{2}$ and $BR(W^{+}\rightarrow \ell^{+}_{i}\ell^{+}_{j}J^{\mp})/|B_{i}|^{2}$, as a function of $m_{N}$.

Also, in the mass range $10\text{ GeV} \lesssim m_{N} \lesssim m_{W}$, the BR for $N\rightarrow \ell^{+}_{j}J^{\mp}$ is:

$$BR(N\rightarrow \ell^{+}_{j}J^{\mp}) \approx \frac{\Gamma(N\rightarrow \ell^{+}_{j}d\bar{u}) + \Gamma(N\rightarrow \ell^{+}_{j}s\bar{c})}{\Gamma(N(Z) + \Gamma_{N}(H) + \Gamma_{N}(W))} \approx \frac{1}{4}. \quad (22)$$

where

$$\Gamma_{N}(Z,H) = \sum_{f} \Gamma(N\rightarrow v_{j}Z^{*}(H^{*}) \rightarrow v_{j}f\bar{f});$$

$$f = u,d,c,s,b,e,\mu,\tau,\nu_{e},\nu_{\mu},\nu_{\tau},$$

$$\Gamma_{N}(W) = \sum_{(f\bar{f})} \Gamma(N\rightarrow \ell^{+}_{j}W^{*\mp} \rightarrow \ell^{+}_{j}(f\bar{f})\bar{f});$$

$$(f\bar{f})^- = (d\bar{u}), (s\bar{c}), (e\nu_{x}), (\mu\nu_{x}), (\tau\nu_{x}). \quad (23)$$

Thus, combining the scaled BRs $BR(W^{+}\rightarrow \ell N)/|B_{i}|^{2}$ calculated from (21) with the BR of the 3-body $N$-decay $BR(N\rightarrow \ell J^{\mp}J^{\mp})$ given in (22), we plot in Fig. 3 the scaled BR’s for the top and W decays in (20). We see that for e.g., $5\text{ GeV} \lesssim m_{N} \lesssim 50\text{ GeV}$ with $|B_{i}|^{2} \sim 10^{-4}$, not excluded by LEP [17], we obtain $BR(t\rightarrow b\ell^{+}_{i}\ell^{+}_{j}J^{\mp}) \sim BR(W^{+}\rightarrow \ell^{+}_{i}\ell^{+}_{j}J^{\mp}) \approx 10^{-6}$. For the $W$-decay, this rather large BR will be well within the reach of the LHC, which as mentioned above, is expected to produce $10^{9}–10^{10}$ inclusive on-shell W’s through $pp\rightarrow W + X$.

Before summarizing let us add a few comments:

- The heavy Majorana neutrino induced $\Delta L = 2$ branching ratios considered here i.e., $t\rightarrow b\ell^{+}_{i}\ell^{+}_{j}J^{\mp}$ (or $t\rightarrow b\ell^{+}_{i}\ell^{+}_{j}J^{\mp}$ if $m_{N} < m_{W}$) and $W^{+}\rightarrow \ell^{+}_{i}\ell^{+}_{j}J^{\mp}$ are of course forbidden in the SM, thus a sighting of each constitutes a spectacular signal of lepton flavor violation (as well as LNV). Take for example the above top decay which produces 2 same-charge leptons, possibly of a different flavor, in addition to a wrong charge $W^{-}$ (recall that in the SM its dominant decay mode is $t\rightarrow bW^{+}$). Of course, these decays are not “stand-alones” since the $t$-quarks and $W$-bosons are created and decay in a specific accelerator and measured by a specific detector. Therefore, the background is highly accelerator and detector dependent—a detailed discussion of which is beyond the scope of this Letter. As mentioned above, this top decay is unique since it has both a pair of same-charge leptons and a “wrong” charge $W^{-}$, unlike the positively charged $W$ produced in the dominant $t\rightarrow bW^{+}$ decay. The observation of this $\Delta L = 2$ top quark decay would therefore be a clear signal for LNV.

- A Majorana exchange is not necessarily the only mechanism leading to $\Delta L = 2$ processes. One can envisage, for instance, a situation in which another type of new physics contributes together with the heavy Majorana exchange. Viable examples are R-parity violating supersymmetry [18], or leptoquark exchanges [19]. In cases like these it is in principle possible to obtain destructive interference between the different mechanisms, thus evading the limits in (4) and (5), leaving the Majorana exchange significant for at least the top-quark decay considered here. Therefore, the rather sizable branching ratios in (15) obtained for $O(1)$ mixing angles cannot be excluded.

- There are some discussions about a super-LHC (SLHC) [20] in which the luminosity of the LHC would increase by about factor of 10. There is also some mention [20] of an energy upgrade from $\sqrt{s} = 14\text{ TeV}$ to $25–28\text{ TeV}$, which may require a new machine. Such an upgrade in both luminosity and energy would yield more than an order of magnitude increase in the number of $tt$ pairs and $W$’s produced, making the $O(10^{-6})$ BR of the top-quark decay in question easily accessible to this machine.

To summarize: we have discussed the $\Delta L = 2$ decays of the top-quark and of the $W$-boson, where both are mediated by a heavy Majorana neutrino $N$. Our main results appear in Figs. 2 and 3 and in Table 1 and are significant for both the top-quark case if $m_{N} < 100\text{ GeV}$ and the $W$-boson case if $m_{N} < m_{W}$.

Acknowledgements

S.B.S. thanks the hospitality of the theory group in Brookhaven National Laboratory where part of this study was performed. The work of S.B.S. and of A.S. was supported in part by US DOE under Grants Nos. DE-FAG02-94ER40817 (USA) and DE-AC02-98CH10886 (BNL). The work of N.G.D and of J.J. was supported in part by US DOE under Grants Nos. DE-FAG02-94ER40817 (USA) and DE-AC02-98CH10886 (BNL). The work of S.B.S. thanks the hospitality of the theory group in Brookhaven National Laboratory where part of this study was performed. The work of S.B.S. and of A.S. was supported in part by US DOE under Grants Nos. DE-FAG02-94ER40817 (USA) and DE-AC02-98CH10886 (BNL). The work of N.G.D and of J.J. was supported in part by US DOE under Grant No. DE-FAG02-96ER40969. G.E. would like to thank Vernon Barger, Tao Han and Tom Rizzo for helpful discussions. The research of G.E. was supported in part by the Israel Science Foundation and by the Fund for Promotion of Research at the Technion.

References


T. Yanagida, in: O. Sawada, A. Sugamoto (Eds.), Proceedings of the Workshop on the Unified Theory and Baryon Number in the Universe, KEK, Tsukuba, Japan, 1979, p. 95, KEK report 79-18;
[3] For recent reviews see e.g.: A.S. Barabash, JINST 1 (2006) P07002, hep-ex/0602037;
For recent limits on sub-eV neutrinos from neutrinoless double-beta see e.g., A. Atre, V. Barger, T. Han, Phys. Rev. D 71 (2005) 113014.
Stability analysis of the Randall–Sundrum braneworld in presence of bulk scalar

Debaprasad Maity a, Soumitra SenGupta a,∗, Sourav Sur b

a Department of Theoretical Physics, Indian Association for the Cultivation of Science, Calcutta 700 032, India
b Institute of Physics, Bhubaneswar 751 005, India

Received 26 July 2006; received in revised form 14 October 2006; accepted 2 November 2006

Abstract

The stability problem of Randall–Sundrum braneworld is readdressed in the light of stabilizing bulk scalar fields. It is shown that in such scenario the instability in general persists because of back-reaction even when an arbitrary potential is introduced for a canonical scalar field in the bulk. It is further shown that a bulk scalar field can indeed stabilize the braneworld when it has a tachyon-like action. The full back-reacted metric in such model is derived and a proper resolution of the hierarchy problem (for which the Randall–Sundrum scenario was originally proposed) is found to exist by suitable adjustments of the parameters of the scalar potential.

© 2006 Elsevier B.V. All rights reserved.

Standard model for strong, weak and electromagnetic interactions based on the gauge group $SU(3) \times SU(2) \times U(1)$ has been extremely successful in explaining physical phenomena upto TeV scale. Such a model, however, encounters the well known fine tuning problem in connection with Higgs mass related to the gauge hierarchy problem which refers to the vast disparity between the weak and Planck scale. By invoking supersymmetry one can resolve this problem at the expense of incorporating a large number of (hitherto unseen) superpartners in the theory. In an alternative approach, theories with extra spatial dimension(s) have attracted a lot of attention because of the new geometric approach to solve the same problem which is distinct from ordinary Kaluza–Klein theory. In such models the standard model fields are localized on a $(3+1)$-dimensional brane [1–7] but gravity can propagate in the bulk spacetime. One of the most significant model in this context was proposed by Randall and Sundrum (RS) [4] where the mass hierarchy emerged naturally in an exponentially warped geometry along the extra dimension. The model contains two $(3+1)$-dimensional branes sitting at the two orbifold fixed points where the single extra dimension in a $(4+1)$-dimensional bulk has been compactified on a $S_1/Z_2$ orbifold. There are two parameters namely the bulk cosmological constant $\Lambda$ and the brane separation $r_c$. The geometry and the Higgs mass is warped exponentially by a dimensionless parameter $kr_c$, where $k = \sqrt{-\Lambda/24M^3}$, $M$ being the 5D Planck mass. For the desired warping one should have $kr_c \sim 11$. If $k \sim$ Planck mass, then $r_c$ should have a stable value near Planck length. To stabilize the value of $r_c$, Goldberger and Wise (GW) [8] proposed a simple mechanism by introducing a minimally coupled massive scalar field in the bulk. Later several other works have been done in this direction [9–13]. In the original work of GW the effect of back-reaction of the scalar field on the background metric was neglected. Such back-reaction was later included in subsequent works and a modified solution for the metric was found [14,15]. However none of these works addressed the modulus stability issue following Goldberger–Wise (GW) mechanism where an effective potential for the modulus $r_c$ is determined by integrating out the scalar field. In particular [15] found the value of the brane separation from the boundary values of the scalar field and assumed that this value corresponds to the stable value if one follows the GW procedure. In this Letter we carefully re-examine the stability issue in the back-reacted RS model when the scalar field is introduced in the bulk. We show

* Corresponding author.

E-mail addresses: tpdm@iacs.res.in (D. Maity), tpssg@iacs.res.in (S. SenGupta), sourav@iopb.res.in (S. Sur).

0370-2693/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.physletb.2006.11.014
on a very general ground that stabilization in general is not possible with a minimally coupled scalar field in the bulk if the scalar back-reaction on the metric is taken into consideration. Following GW we subsequently show that if the dependence of scalar field action on the extra coordinate is tachyon-like then we can stabilize the brane world and the stabilized value of $r_c$ can produce the desired hierarchy from Planck scale to TeV scale.

Let us first consider the following bulk action [8,14–16]

$$S = \int d^5x \sqrt{-G} \left[ -M^3 R + \frac{1}{2} \partial_\mu \phi \partial^\mu \phi - V(\phi) \right] - \int d^4x \sqrt{-g_c} \delta(y - y_a) \lambda_a(\phi),$$  

(1)

where $\phi$ is a bulk scalar field with a potential $V(\phi)$ and the index $a$ runs over the brane locations. The corresponding brane potentials are denoted by $\lambda_a$.

Taking the line element in the form

$$ds^2 = e^{-2A(y)} \eta_{\mu\nu} dx^\mu dx^\nu - dy^2,$$  

(2)

where $\{y\}$ is the extra compact coordinate with radius $r_c$ such that $dy^2 = r_c^2 d\theta^2$, $\{\theta\}$ is the angular coordinate, the field equations are given as

$$\phi'' - 4A' \phi' = \frac{\partial V}{\partial \phi} + \sum_a \frac{\partial \lambda_a}{\partial \phi} \delta(y - y_a),$$  

(3a)

$$4A'^2 - A'' = -\frac{V}{3M^3} - \frac{1}{6M^3} \sum_a \lambda_a \delta(y - y_a),$$  

(3b)

$$4A'^2 - 4A'' = -\frac{V}{3M^3} - \frac{1}{2M^3} \sum_i \phi_i'^2 - \frac{2}{3M^3} \sum_a \lambda_a \delta(y - y_a),$$  

(3c)

where prime $\{\prime\}$ denotes partial differentiation with respect to the extra spatial coordinate $y$.

The boundary conditions are

$$[\phi']_a = \frac{\partial \lambda_a}{\partial \phi}, \quad [A']_a = \frac{1}{6M^3} \lambda_a(\phi).$$  

(4)

Denoting $C = 1/24M^3$, it follows

$$A'^2 = 2C (\phi'^2 - 2V), \quad A'' = 4C \phi'^2,$$  

(5a)

$$\phi'' = 4A' \phi' + \frac{\partial V}{\partial \phi}.$$  

(5b)

In order to obtain analytic closed form solutions we resort to the particular class of effective potentials as in Refs. [14,17]:

$$V(\phi) = \frac{1}{8} \left[ \frac{\partial W}{\partial \phi} \right]^2 - 2CW^2,$$  

(6)

where $W(\phi)$ is a superpotential.

Eqs. (5a) and (5b) lead to the first-order equations:

$$\phi' = \frac{1}{2} \frac{\partial W}{\partial \phi}, \quad A' = 2CW.$$  

(7)

Following GW [8], we calculate the effective potential $V_{\text{eff}}$ for $r_c$ on the 3-brane as by using Eqs. (5a) and (5b) for bulk potential as follows:

$$V_{\text{eff}}(r_c) = 2r_c \int_0^\pi d\theta e^{-4A(r_c, \theta)} \left[ \frac{1}{2} \frac{\partial A \phi \partial A \phi - V(\phi)}{\partial \phi} - e^{-4A(0)} \lambda_0(\phi^0) - e^{-4A(r_c, \pi)} \lambda_\pi(\phi^\pi) \right]$$  

$$= 2r_c \int_0^\pi d\theta e^{-4A(r_c, \theta)} \left[ -\phi'^2 + \frac{1}{2C} A'^2 \right] - \lambda_0(\phi_0) - e^{-4A(r_c, \pi)} \lambda_\pi(\phi^\pi)$$  

$$= 2r_c \int_0^\pi d\theta e^{-4A(r_c, \theta)} \left[ -\frac{1}{4C} A'^2 + \frac{1}{2C} A'^2 \right] - \lambda_0(\phi_0) - e^{-4A(r_c, \pi)} \lambda_\pi(\phi^\pi),$$  

(8)
where the Planck brane is located at \( \theta = 0 \) and the visible brane is at \( \theta = \pi \), and we have used the notations: \( \phi^0 \equiv \phi(\theta = 0) \), \( \phi^\pi \equiv \phi_i(\theta = \pi) \). Using the boundary conditions (4) which now take the form

\[
\left[ W(\phi) \right]_0 = 2\lambda_0(\phi), \quad \left[ \frac{\partial W}{\partial \phi} \right]_0 = 2\frac{\partial \lambda_0}{\partial \phi},
\]

(9a)

\[
\left[ W(\phi) \right]_{r_c \pi} = 2\lambda_{\pi}(\phi), \quad \left[ \frac{\partial W}{\partial \phi} \right]_{r_c \pi} = 2\frac{\partial \lambda_{\pi}}{\partial \phi}
\]

(9b)

the brane potentials \( \lambda_0(\phi) \) and \( \lambda_{\pi}(\phi) \) are given by [14]

\[
\lambda_0(\phi) = W(\phi^0) + \left[ \frac{\partial W}{\partial \phi^0} (\phi - \phi^0) \right] + \gamma_0^2 (\phi - \phi^0)^2,
\]

(10a)

\[
\lambda_{\pi}(\phi) = -W(\phi^\pi) - \left[ \frac{\partial W}{\partial \phi^\pi} (\phi - \phi^\pi) \right] + \gamma_{\pi}^2 (\phi - \phi^\pi)^2,
\]

(10b)

where the constants \( \gamma_0 \) and \( \gamma_{\pi} \) are parameters of various potentials.

Now using the boundary condition Eq. (4), we explicitly calculate the four-dimensional effective potential for \( r_c \) as:

\[
V_{\text{eff}} = r_c \int_0^\pi d\theta e^{-4A(r_c \theta)} \left[ -\frac{1}{4C} A'^2 + \frac{1}{2} A' \right] - \lambda_0(\phi_c) + e^{-4A(r_c \pi)} \lambda_{\pi}(\phi_{\pi})
\]

\[
= -\frac{1}{2C} e^{-4r_c \theta} A'[0] - \frac{r_c}{C} \int_0^\pi d\theta \left[ e^{-4A(r_c \theta)} A' \right] + \frac{1}{2C} e^{-4r_c \theta} A'[\pi] - \frac{1}{2C} e^{-4r_c \theta} A'[0]
\]

\[
= -4Cr_c \int_0^\pi d\theta e^{-4A(r_c \theta)} W^2.
\]

(11)

From the above equation we immediately obtain,

\[
\frac{1}{\pi} \left[ \frac{\partial V_{\text{eff}}}{\partial r_c} \right] = -4Ce^{-4A(r_c \pi)} W(r_c \pi)^2.
\]

(12)

Now, in order to have extremum for the above effective potential at some value of \( r_c \) (= \( r_s \), say)

\[
\frac{1}{\pi} \left[ \frac{\partial V_{\text{eff}}}{\partial r_c} \right]_{r_c=r_s} = -4Ce^{-4A(r_s)} W(r_s)^2 = 0,
\]

(13)

which implies \( W(r_s) = 0 \) as well as \( \partial^2 V_{\text{eff}} / \partial r_c^2 = 0 \) at \( r_c = r_s \). On the other hand, the third derivative is given by

\[
\left[ \frac{\partial^3 V_{\text{eff}}}{\partial r_c^3} \right]_{r_c=r_s} = -e^{-4A(r_s)} \left[ \frac{\partial W(r_s)}{\partial r_c} \right]^2
\]

(14)

which is clearly non-vanishing in accord with the above boundary conditions. Therefore, no extremum exists and we only have a point of inflexion at the value of \( r_c = r_s \) for which \( W \) vanishes.

Let us now resort to a more general solution ansatz for the metric function \( A(y) \) and single scalar field \( \phi \) so that the standard RS [4] solution is obtained in the limit where the scalar \( \phi(y) \) becomes trivial:

\[
A(y) = ky + f(\phi_1(y)), \quad \phi = \phi_1(y, a_1, a_2)
\]

(15)

where \( a_1, a_2 \) are the initial parameters of the theory and constant \( k \sim 5D \) Planck scale as in the RS picture.

The field Eqs. (5a) and (5b) yield

\[
\phi_1 = \frac{1}{4C} \frac{\partial f}{\partial y^2}, \quad \frac{\partial^2 f}{\partial y^2} = k + \frac{\partial f}{\partial y} \frac{\partial f}{\partial y^2}.
\]

(16a)

\[
V = \frac{1}{8C} \frac{\partial^2 f}{\partial y^2} - \frac{1}{2C} \left( k + \frac{\partial f}{\partial y} \right)^2.
\]

(16b)

Clearly in the limit \( f \to 0, V \to -12M^2k^2 = \Lambda \) we get back the standard \( AdS_5 \) bulk geometry of RS model.

From Eq. (4), the boundary values of the brane potentials and their first derivatives are obtained as

\[
[\lambda_0]_0 = 2C \left[ k + \frac{\partial f}{\partial y} \right], \quad \left[ \frac{\partial \lambda_0}{\partial y} \right]_0 = 2C \left[ \frac{\partial^2 f}{\partial y^2} \right]_0.
\]

(17a)
\[
[\lambda_{\pi}]_{r_c,\pi} = -2C \left[ k + \frac{\partial f}{\partial y} \right]_{r_c,\pi}, \quad \left[ \frac{\partial \lambda_{\pi}}{\partial y} \right]_{r_c,\pi} = -2C \left[ \frac{\partial^2 f}{\partial y^2} \right]_{r_c,\pi}. \tag{17b}
\]

At a stable point \( r_c = r_s \), the effective 4D potential (8) has a vanishing first derivative
\[
\left[ \frac{\partial V_{\text{eff}}}{\partial r_c} \right]_{r_c = r_s} = -\frac{e^{-4A(r_s)}}{C} \left[ k + \frac{\partial f}{\partial y} \right]_{r_c = r_s} = 0 \tag{18}
\]
which at once implies that the second derivative \( \partial^2 V_{\text{eff}}/\partial r_c^2 \) also vanishes at the stable point \( r_s \). The third derivative, however, is given at \( r_c = r_s \) as
\[
\left[ \frac{\partial^3 V_{\text{eff}}}{\partial r_c^3} \right]_{r_c = r_s} = -2 \frac{e^{-4A(r_s)}}{C} \left[ \frac{\partial^2 f}{\partial y^2} \right]_{r_c = r_s} \tag{19}
\]
which is, of course, non-vanishing in consistence with Eq. (16a).

Considering, for example
\[
f = a_1 \exp(-2a_2 y) \tag{20}
\]
we find from Eqs. (16a) and (16b)
\[
\phi_{\pm} = \sqrt{a_1} e^{-a_2 y}, \quad V = \frac{1}{8} \left[ \frac{\partial W}{\partial \phi} \right]^2 - 2CW^2, \tag{21}
\]
where \( W = 12kM^3 - a_2 \phi^2 \). Clearly, one gets back the same results as discussed in [14].

The analysis so far brings out the inherent difficulty in stabilizing the RS two-brane model using canonical bulk scalar with arbitrary potential.

We now consider a tachyon-like scalar in the bulk [18,19]. The action is given as
\[
S = -\int d^5 x \sqrt{-G} \left[ R + V \sqrt{1 - \partial_A \phi \partial_A \phi} \right] - \int d^4 x \sqrt{-g_{\ell}} \Gamma_\ell(\phi), \tag{22}
\]
where \( V \) is the potential of the field.

With a similar metric ansatz the field equations take the form
\[
A'^2 = -\frac{1}{12} \frac{V}{\sqrt{1 + \phi^2}}, \quad A'' = \frac{1}{6} \frac{V \phi^2}{\sqrt{1 + \phi^2}}, \tag{23a}
\]
\[
\frac{\phi''}{(1 + \phi^2)} - 4A' \phi' - \frac{1}{V} \frac{\partial V}{\partial \phi} = 0 \tag{23b}
\]
and the corresponding boundary conditions are
\[
[A']_i = \frac{1}{6} \lambda_i(\phi), \quad \left[ \frac{V \phi'}{(1 + \phi^2)^{3/2}} \right]_i = \frac{\partial \lambda_i(\phi)}{\partial \phi}, \tag{24}
\]
where \( [i] \) corresponds to the brane locations at orbifold fixed points along the extra direction.

Proceeding along the same line as in the previous section, we obtain from Eqs. (23a) and (23b)
\[
\frac{\phi''}{2A'^2} = -\frac{A''}{2A'^2}, \quad V = -12A' \sqrt{A'^2 - A''^2}/2. \tag{25}
\]

Let us now consider the simple solution ansatz as
\[
A = ky + a_1(y - a_2)^2 \tag{26}
\]
where \( k, a_1, a_2 \) are the parameters of the theory. We get
\[
\phi_{\pm}^2 = -\frac{a_1}{[y + 2a_1(y - a_2)]^2}, \quad V = -12[k + 2a_1(y - a_2)] \sqrt{\frac{1}{k + 2a_1(y - a_2)}}. \tag{27a}
\]

These equations imply that \( a_1 = 0 \) gives \( V = -12k \) and \( \phi = \text{constant} \), indicating pure RS background. We are, however, interested in \( a_1 \neq 0 \), in which case a real solution for \( \phi \) can be obtained only when \( a_1 \) is negative:
\[
\phi_{\pm} = \pm(4a_1)^{-1/2} \ln[k + 2a_1(y - a_2)]. \tag{28}
\]
After a long but straightforward calculation following GW [8] and along same line of Eqs. (8), (11), we get

\[
\frac{1}{\pi} \frac{\partial V}{\partial r_c} = -72e^{-4A} \left[ \frac{-A''^2 + 4A^4 - A''A'^2}{12A^2 - 6A''} \right]_{r_c},
\]

(29)

and the second derivative turns out to be

\[
\frac{1}{\pi^2} \frac{\partial^2 V}{\partial r_c^2} \propto A'A''.
\]

(30)

So by doing the straightforward calculation, the minimum stable point is at

\[
\pi r_s = -\sqrt{\frac{1}{2|a_1|(|\sqrt{5} + 1| + \frac{k}{2|a_1|} + a_2}}.
\]

(31)

The corresponding metric function, upon setting $|a_1| = m^2$, is given by

\[
A = \frac{k^2}{4m^2} + ka_2 - 0.154.
\]

(32)

It is now clear that one can easily get the acceptable hierarchy by appropriately choosing the values of $k/m$ and $ka_2$ with the stabilized radius $r_s$.

Our work clearly reveals an inherent instability in Randall–Sundrum two brane model. It is shown that just by introducing a canonical scalar field in the bulk, stabilization of the braneworld in general cannot be achieved even with an arbitrary potential if full back-reaction of the scalar field on the metric is considered. However, we also show that in such models the two-brane separation can indeed be stabilized if the bulk scalar field has a tachyon-like action with respect to the bulk coordinate. We finally show that the full back-reacted metric in such case can also yield the desired warping from Planck scale to TeV scale and thus resolves the fine tuning problem of the Higgs mass in a stable braneworld scenario.

Acknowledgement

D.M. acknowledges Council of Scientific and Industrial Research, Govt. of India for providing financial support.

References


S. Forste, hep-th/0110055.

E.E. Boos, et al., hep-th/0511185;

P.C. Ferreira, P.V. Moniz, hep-th/0601086;


C. Csaki, hep-ph/0404096;


A. Sen, JHEP 0207 (2002) 065;

Long-range $SL(2)$ Baxter equation in $\mathcal{N} = 4$ super-Yang–Mills theory

A.V. Belitsky

Department of Physics, Arizona State University, Tempe, AZ 85287-1504, USA

Received 19 September 2006; received in revised form 9 November 2006; accepted 9 November 2006
Available online 15 November 2006

Abstract

Relying on a few lowest order perturbative calculations of anomalous dimensions of gauge invariant operators built from holomorphic scalar fields and an arbitrary number of covariant derivatives in maximally supersymmetric gauge theory, we propose an all-loop generalization of the Baxter equation which determines their spectrum. The equation does not take into account wrapping effects and is thus asymptotic in character. We develop an asymptotic expansion of the deformed Baxter equation for large values of the conformal spin and derive an integral equation for the cusp anomalous dimension.

© 2006 Elsevier B.V. All rights reserved.

1. Introduction

Four-dimensional non-Abelian gauge theories were found to possess integrable structures. The latter imply the existence of hidden symmetries of the dilatation operator whose eigenvalues determine anomalous dimensions of gauge invariant composite operators of elementary fields in underlying models. Integrability was revealed in one-loop anomalous dimensions of twist-$L$ maximal-helicity Wilson operators in QCD by identifying the former with eigenenergies of the $L$-site XXX Heisenberg spin chain [1]. The magnet turns out to be noncompact, for the spin operators acting on its sites transform in the infinite-dimensional representation of the collinear subgroup $SL(2, \mathbb{R})$ of the conformal group $SO(4, 2)$. Since the one-loop phenomenon is spawned by gluons, invariably present in Yang–Mills theories—supersymmetric or not—they all necessarily exhibit the same, universal integrable structures. The differences arise merely due to distinct particle contents of the models: while only holomorphic sectors are integrable in QCD and its nearest supersymmetric $\mathcal{N} = 1, 2$ siblings [2], the maximal supersymmetry of the $\mathcal{N} = 4$ super-Yang–Mills theory extends integrability to all operators [3,5,2]. Recent perturbative studies build up a growing amount of direct evidence that integrability persists in certain closed compact [4,5] and noncompact [6–10] subsectors of gauge theories even in higher orders of perturbation theory. Thus, while ruled out for gauge theories with $\mathcal{N} < 4$ supercharges, it is plausible that the maximally supersymmetric Yang–Mills theory is completely integrable. An additional confirmation for this conjecture comes from studies of multi-loop multi-leg scattering amplitudes which display intriguing iterative structures [11,12]. These arguments suggest that the spectrum of all-loop anomalous dimensions in $\mathcal{N} = 4$ SYM theory is determined by a putative long-range integrable spin chain with the dilatation operator being its Hamiltonian.

In this note we probe the underlying integrable long-range magnet by proposing its multi-loop perturbative structure within the framework of the Baxter $Q$-operator [13]. This approach is based on the existence of an operator $Q(u)$ depending on a spectral parameter $u$ and acting on the Hilbert space of the magnet. For different values of $u$ it forms a family of mutually commuting operators, simultaneously commuting with the spin-chain Hamiltonian as well. Although in the present circumstances, the formalism is equivalent to the Bethe Ansatz approach, it possesses certain advantages. First, the eigenvalue $Q(u)$ of the Baxter operator $Q(u)$ determines the single-particle wave function of the chain in the representation of separated variables [14]. Second, the equation for the $Q$-operator—known as the Baxter equation—is polynomial, to be contrasted with a set of coupled transcendental Bethe equa-
Currenty we restrict our consideration to the closed \[15,16\] noncompact SL(2) sector \[1,15\] of the gauge theory which is spanned by single-trace maximal R-charge Wilson operators built from the holomorphic scalar fields \(X = \phi_1 + i \phi_2\) and covariant derivatives,

\[
\mathcal{O}_{n_1 n_2 \ldots n_L}(0) = \text{tr}[(i D_+)^{n_1} X(0)(i D_+)^{n_2} X(0) \cdots (i D_+)^{n_L} X(0)].
\]

(1.1)

Here \(D_+ = D_0 n^\mu\) is projected on the light cone with a null vector \(n^\mu, n^2 = 0\), in order to factor out the maximal Lorentz-spin component from the operator in question. These Wilson operators mix with each other under renormalization group evolution and acquire anomalous dimensions at all orders of perturbative series in coupling constant\(^1\)

\[
\gamma(g) = \sum_{n=1}^\infty g^{2n} \gamma^{(n)}.\]

(1.2)

We find it convenient to use the expansion parameter \(g\) related to the \(\ 't\) Hooft coupling constant \(\lambda\) via

\[
g = \sqrt{2\lambda} = \frac{g_{\text{YM}} \sqrt{N_c}}{2\pi}.
\]

(1.3)

The anomalous dimension \(\gamma(g)\) depends on parameters characteristic of the operator: its twist \(L\), determined by the number of \(X\)-fields, and its Lorentz spin \(N = n_1 + n_2 + \cdots + n_L\). Within the method of the Baxter \(Q\)-operator, the eigenspectrum of one-loop anomalous dimensions \(\gamma^{(0)}\) and the corresponding quasimomentum \(\theta^{(0)}\) are determined by the leading order Baxter function \(Q^{(0)}(u)\)

\[
\gamma^{(0)} = \frac{i}{2} \Big[ \ln Q^{(0)} \left( \frac{i}{2} \right) \Big] - \frac{i}{2} \Big[ \ln Q^{(0)} \left( -\frac{i}{2} \right) \Big],
\]

\[
\theta^{(0)} = \ln Q^{(0)} \left( \frac{i}{2} \right) - \ln Q^{(0)} \left( -\frac{i}{2} \right).
\]

(1.4)

Since the Baxter function \(Q^{(0)}(u)\) is related to the eigenfunction of the mixing matrix, it corresponds to a multiplicatively renormalizable Wilson operator and thus has to be polynomial in \(u\) of order \(N\), \(Q^{(0)}(u) = (u - u_1^{(0)})(u - u_2^{(0)}) \cdots (u - u_N^{(0)})\). The zeros of this polynomial are determined by the Bethe roots \(u_k^{(0)}\) which take only real values for the noncompact SL(2, \(\mathbb{R}\)) spin chain \[17\]. The function \(Q^{(0)}(u)\) obeys the finite-difference Baxter equation \[13\]

\[
u^L u^L\frac{Q^{(0)}(u+i)}{Q^{(0)(u)}} + \nu^L u^L\frac{Q^{(0)}(u-i)}{Q^{(0)(u)}} = \tau^{(0)}(u) Q^{(0)}(u),
\]

(1.5)

where the spectral parameter in the dressing factors \(\nu\) is shifted by the conformal spin \(s = \frac{1}{2}\) of the scalar field \(X\), \(u_{\pm} = u \pm \frac{i}{2}\) and \(\tau^{(0)}(u)\) is an order-\(L\) polynomial in \(u\) depending on the integrals of motion.

\[\text{1 Their complete two-loop planar mixing matrix has been recently computed in Ref. [10].}\]

2. Three-loop Baxter equation

Explicit perturbative calculations \[6,7,9\] of two-loop corrections to the anomalous dimensions of the scalar operators (1.1) exhibit double degeneracy of energy levels with zero quasi-momentum. This hints at the existence of nontrivial odd-parity conserved charges and thus persistence of integrability at higher orders of perturbation theory.

Beyond one loop, the formalism of the Baxter operator gets modified accordingly. The Bethe roots acquire corrections in coupling constant to all orders of perturbation theory,

\[
u_n(g) = \sum_{k=0}^{\infty} g^{2k} \nu_n(i),
\]

(2.1)

and obey deformed Bethe Ansatz equations \[18\]. The reality of Bethe roots \(u_k(g)\) have to be preserved to all orders since the eigenvalue \(Q(u)\) of \(Q(u)\) is a wave function of the chain with the number of its nodes on the real \(u\)-axis coinciding with the spin \(N\) of the operator. The polynomial

\[
Q(u) = \prod_{n=1}^N (u - u_n(g)),
\]

(2.2)

fulfills these properties and is real \(Q^*(u) = Q(u^*)\) for \(u^* = u\). In Ref. \[10\] we found from available two- \[19,20,6,7\] and three-loop \[21–23\] diagrammatic calculations of anomalous dimensions that the Baxter equation possesses the form

\[
x_u^L e^{\nu_{+}(x_u)} Q(u+i) + x_u^L e^{-\nu_{-}(x_u)} Q(u-i) = \tau(u) Q(u),
\]

(2.3)

with the dressing factors depending on the renormalized spectral parameter \[24\]

\[
x[u] = \frac{1}{2} (u + \sqrt{u^2 - g^2}), \quad x_{\pm} = x[u_{\pm}].
\]

(2.4)

The multi-loop transfer matrix\(^2\)

\[
t(u) = 2 u^L + q_1(g) u^{L-1} + q_2(g) u^{L-2} + \cdots + q_L(g)
\]

(2.5)

acquires the “missing” term \(\sim u^{L-1}\) at \(g^2\)-order, i.e., \(q_1(g) \sim O(g^2)\), while the rest of the charges start from \(O(g^6)\), \(q_k(g) = q_k^{(0)} + O(g^2)\). The additional dressing factors \(\sigma\) obey the complex conjugation condition \((\sigma_{+} (x))^{*} = \sigma_{-} (x^*)\) for \(\forall m u = 0\) and encode the renormalization of the noncompact charges \(q_k^{(0)}\) at higher orders. An analysis yielded the following result to three-loop order \[10\]

\[
\sigma_{+}(x) = -\frac{g^2}{2x} \left[ \ln Q \left( \pm \frac{i}{2} \right) \right] - \frac{g^4}{16x^2} \left[ \ln Q \left( \pm \frac{i}{2} \right) \right]'' + \mathcal{O}(g^6).
\]

(2.6)

\[\text{2 Note that with this transfer matrix the resulting Baxter equation breaks down already at order } O(g^{2n}) \text{ in coupling constant with } n = L. \text{ It turns out that one can correctly incorporate order } n = L \text{ corrections by replacing the leading term in } t(u) \text{ with the following combination } 2u^L \rightarrow x_u^L + x_u^L - (\frac{L}{2})^2 - (-\frac{L}{2})^2. \]

This also allows to set \(q_1(x) = 0\).

While the anomalous dimension is expressed order-by-order in coupling constant \( g \) in terms of the solution to Eq. (2.3) as [10]

\[
\gamma(g) = i \left[ \frac{g^2}{2} \left[ \ln Q(u) \right]' + \frac{g^4}{16} \left[ \ln Q(u) \right]'' + \frac{g^6}{384} \left[ \ln Q(u) \right]^{(5)} + O(g^8) \right] \bigg| \substack{u = i/2 \\ \theta = 0}.
\]  

(2.7)

The anomalous dimensions found using these equations reproduced exactly available perturbative predictions. One can demonstrate that the condition of the pole-free transfer matrix at Bethe roots \( u_n(g), t(u_n) = 0 \) immediately produces the three-loop Bethe Ansatz of Ref. [25].

3. Multi-loop conjecture

The above representation (2.6) of the dressing factors \( \sigma_{\pm} \) can be brought to a very suggestive form. Namely, a quick inspection allows one to rewrite these terms as an expansion in terms of the Chebyshev polynomials of the second kind \( U_k \),

\[
\sigma_{\pm}(x) = \frac{2g^2}{\pi} \int_{-1}^{1} dt \sqrt{1 - t^2} \left( \ln Q \left( \pm \frac{i}{2} - gt \right) \right)''
\]

\[
\quad \times \sum_{n=0}^{n_{\text{max}}} \left( - \frac{g}{2x} \right)^n + \frac{U_n(t)}{n+1},
\]  

(3.1)

with \( n_{\text{max}} = 2 \), valid to \( O(g^3) \) in the approximation of Eq. (2.6). Having this representation at our disposal, we may naturally extend the first few terms of the available perturbative series to all orders in coupling \( g \), by sending \( n_{\text{max}} \to \infty \). Using the summation theorem for Chebyshev polynomials, one can sum the infinite series up into the function \(-\text{arccot} \left( \frac{1+2x/g}{\sqrt{1-t^2}} \right)\) and, upon a variable transformation, write \( \sigma_{\pm} \) in the form (with \( \bar{z} = 1 - z \))

\[
\sigma_{\pm}(x) = \frac{g^2}{2\pi x} \int_{0}^{1} dz \int_{-1}^{1} dt \sqrt{1 - t^2}
\]

\[
\times \left[ \ln Q \left( \pm \frac{i}{2} - g \sqrt{\bar{z}} t + \frac{\bar{z}}{4x} \frac{g^2}{4} \right) \right]''
\]

\[
=i \theta_{\pm} - \frac{1}{\pi} \int_{-1}^{1} dt \ln Q \left( \pm \frac{i}{2} - gt \right) \sqrt{u^2 - g^2} \frac{u^2 - g^2}{u + gt}.
\]  

(3.2)

Here we integrated by parts in the second line in order to separate the components \( \theta_{\pm} \) of the spin-chain quasimomentum \( \theta = \theta_+ - \theta_- \).

\[
i \theta_{\pm} = \frac{1}{\pi} \int_{-1}^{1} dt \ln Q \left( \pm \frac{i}{2} - gt \right) \sqrt{1 - t^2}.
\]  

(3.3)

Notice that \( \theta \) reduces to the one-loop expression (1.4) upon setting \( g = 0 \). While the condition \( t(u_n) = 0 \) yields the all-order Bethe Ansatz equations suggested in Ref. [18].

The conjectured multi-loop Baxter equation (2.3) with (3.2) and the known pattern of renormalization of the conformal spin in field theories can be used to determine the all-loop analytic expression for the anomalous dimensions in terms of the Baxter function.

To this end, recall that the conformal spin of Wilson operators \( J^{(0)} = N + \frac{1}{2} L \) defining the quadratic Casimir \( Q^{(2)} = -J^{(0)}(J^{(0)} - 1) - \frac{1}{2} L \) gets additive renormalization by the anomalous dimensions \( \gamma(g) \) of composite Wilson operators at higher orders in coupling, \( J^{(0)} \to J = N + \frac{1}{2} L + \frac{1}{2} \gamma(g) \).

This conclusion arises from considerations of conformal Ward identities for Green functions with conformal operator insertion [26,7]. Then a short inspection of the Baxter equation (2.3) with the dressing factors \( \sigma_{\pm} \) in the form (3.1) demonstrates that the first term in the series of \( \sigma_{\pm}(x) = i \gamma_{\pm}(g)/x + \cdots \) induces the shift of the conformal spin,

\[
J^{(0)} = N + \frac{1}{2} L \to J = N + \frac{1}{2} L + \frac{1}{2} \gamma(g).
\]  

(3.4)

Consequently, we may naturally identify the addendum with the anomalous dimensions of a multiplicatively renormalizable composite operators, \( \gamma(g) = \gamma_+(g) - \gamma_-(g) \). Making use of the explicit form of the dressing factors \( \sigma_\pm \), we find the integral representation of \( \gamma(g) \) in terms of the solution to the Baxter equation,

\[
\gamma(g) = \frac{g^2}{\pi} \int_{-1}^{1} dt \sqrt{1 - t^2} \left[ \ln Q \left( \frac{i}{2} - gt \right) \right]' - \ln Q \left( \frac{i}{2} - gt \right)'.
\]  

(3.5)

The Taylor expansion shows that the lowest three orders in \( g^2 \) coincide with Eq. (2.7).

The Baxter equation (2.3) can be solved analytically order-by-order in coupling constant for specific values of \( L \) and \( N \), e.g., for \( L = 4, N = 2 \) eigenvalue with zero quasimomentum reads,\(^3\)

\[
\gamma(g) = \frac{5 \pm \sqrt{3}}{2} g^2 - \frac{17 \pm 5 \sqrt{5}}{8} g^4 + \frac{585 \pm 207 \sqrt{5}}{160} g^6 - \frac{5185 \pm 2039 \sqrt{5}}{640} g^8 + O(g^{10}).
\]  

(3.6)

However, it has a limited range of applicability being asymptotic in character; it allows to find the anomalous dimensions up to order \( O(g^{2n}) \) only for operators of length \( L \geq n \). This restriction arises from the breaking of its polynomiality above a boundary value of \( n \), i.e., for \( L \leq n \). Analogous limitations apply to the Bethe Ansatz equations of Ref. [24]. A generic dependence of \( \gamma(g) \) on the parameters \( L \) and \( N \) is not known however and below we will develop an asymptotic scheme to find it in the large spin limit.

\(^3\) This anomalous dimension, when related to Berenstein–Maldacena–Nastase operators [27], agrees with previous one-, two- and five-loop analyses of Refs. [28], [4] and [24], respectively.
4. Asymptotic expansion

The large-$N$ behavior of anomalous dimension is of special interest in its own right since it governs the Sudakov asymptotics of scattering amplitudes [29,30], and in light of gauge/string duality, for it can be compared (at strong coupling) to energies of quasiclassical strings [31–35]. Recall at first that the anomalous dimensions of twist-$L$ operators occupy a band of width $L - 2$, with the upper and lower boundaries scaling like [1,15]

\[ \gamma_{\text{lower}}(g) = 2 \Gamma_{\text{cusp}}(g) \ln N, \]
\[ \gamma_{\text{upper}}(g) = L \Gamma_{\text{cusp}}(g) \ln N, \]  

(4.1)

and the coefficient $\Gamma_{\text{cusp}}(g)$ being the cusp anomalous dimension [36,37], known to one- [36], two- [37,19,20,15] and three-loop orders [22,23,12]. The minimal anomalous dimension $\gamma_{\text{lower}}(g)$ of high-twist operators develops the asymptotic behavior identical to the one of twist-two operators [1,34,9]. Since the single-logarithmic regime is realized for $L \ll N$ with $L, N \to \infty$ [34], this allows one to evade the limitation of the asymptotic character of the Baxter equation and to derive an all-loop equation for the cusp anomaly $\Gamma_{\text{cusp}}$.

Notice that although we have to solve the problem with large quantum numbers, we cannot apply traditional WKB expansion for $Q(u)$ (see, e.g., Ref. [11]) since the latter is valid for the spectral parameter which scales as $u \sim N^1$ while the energy is determined by the Baxter function $Q(u)$ evaluated at the argument $u = \pm \frac{1}{2} - gt$ which behaves as $u \sim N^0$. Therefore, we have to resort to other techniques. To this end, we will use in the following the approach developed in Refs. [38,34] for one-loop anomalous dimensions and which, as we will see momentarily, is easily generalizable beyond leading order of perturbation theory.

4.1. One-loop Baxter equation

Let us briefly review the formalism of Refs. [38,34] applied to the one-loop Baxter equation (1.5). Though we are interested only in the lowest energy curve, at the beginning we will be general enough to discuss subleading trajectories as well in order to point out approximations which have to be imposed to separate the lowest anomalous dimension only. In the regime in question, the conserved charges are large $q_k^{(0)} \sim N^k$ and, therefore, the transfer matrix is large $|r^{(0)}(u)| \gg 1$. Introducing a new function

\[ \phi^{(0)}(u) = \frac{Q^{(0)}(u + i)}{Q^{(0)}(u)}, \]

(4.2)

we can rewrite the Baxter equation in the form

\[ u^L \phi^{(0)}(u) + \frac{u^L}{\phi^{(0)}(u) - i} = r^{(0)}(u). \]

(4.3)

The solution to it is based upon different scaling behavior of the right- and left-hand sides with $N$. For the spectral parameter $u \sim N^0$, the solution is given by an infinite fraction. Keeping the leading terms only we come to two difference equations

\[ u^L Q_+^{(0)}(u + i) = r^{(0)}(u) Q_+^{(0)}(u), \]
\[ u^L Q_0^{(0)}(u - i) = r^{(0)}(u) Q_-^{(0)}(u). \]

(4.4)

The additive corrections to their right-hand sides go as $O(1/q_0^{(0)})$, where $q_0^{(0)}$ is a conserved charge which scales with the maximal power of $N$. For cyclically symmetric states $\theta = 0$, the asymptotic solution to (1.5) reads

\[ Q^{(0)}_0(u) = Q_+^{(0)}(u) Q_-^{(0)}\left(-\frac{i}{2}\right) + Q_-^{(0)}(u) Q_+^{(0)}\left(\frac{i}{2}\right). \]

(4.5)

in terms of the solution to the two-term recursion relations (4.4) written with the help of the roots $\delta_k$ of the transfer matrix $r^{(0)}(u) = 2(u - \delta_1)(u - \delta_2) \cdots (u - \delta_L)$ [34],

\[ Q_+^{(0)}(u) = 2^{L+2}u L \prod_{k=1}^L \Gamma(\pm iu + i\delta_k) / \Gamma(\pm iu + \frac{1}{2} \delta). \]

(4.6)

Now recall that we are interested only in the trajectory with the lowest energy only. The latter does not depend on the twist of the operator, i.e., it is $L$-independent. The reason for this being that for the corresponding state only the quadratic Casimir $q_2^{(0)}$ is large while all other integrals of motion become anomalously small. For the roots of the transfer matrix this is translated into the statement that just two roots $\delta_1 = \delta_L$ are much larger than the rest of $\delta$’s which are negligible [34], yielding the relation

\[ \delta_1^2 \simeq -q_2^{(0)}/2. \]

(4.7)

In this case the genus-$(L - 2)$ hyperelliptic Riemann surface parameterizing the magnet, with its moduli determined by the conserved charges $q_k^{(0)}$, degenerates into a sphere, i.e., the spectral curve of twist-two operators [34]. This implies that all zones but one of allowed classical motion in separated variables collapse into points. In this limit the transfer matrix reduces to

\[ r^{(0)}(u) \sim u^L t^{(0)}(u) = u^L(2 - N^2/u^2) \]

and the solutions to the recursion relations (4.4) becomes symmetric under the interchange $u \to -u$ and equal, $Q_+^{(0)}(u) = Q_-^{(0)}(u)$. In the infinite-spin limit, we then find that the leading behavior of the Baxter function is

\[ (i \ln Q_+^{(0)}(u))' = \psi(-iu + i\delta_1) + \psi(-iu - i\delta_1) + \cdots \]
\[ \simeq 2 \ln N + \cdots , \]

(4.8)

where in the last step we imposed the condition that the evaluation of the anomalous dimensions (1.4) requires $u \sim N^0$ and thus it can be neglected compared to $N$. This consideration immediately suggests that for the minimal-energy trajectory in the single-logarithmic asymptotics the dressing factors $u^L$ in the left-hand side of Eq. (4.4) are irrelevant. Thus they can be reduced to $u^L \to u^L$ and canceled with the factor extracted from the transfer matrix $r^{(0)}(u)$, making the equation $L$-independent, as expected. The latter is clearly seen in the quasiclassical approach when one assumes the spectral parameter to scale with $N$, i.e., $u = N\hat{u}$ and $\hat{u} \sim 1$. We will use the same argument below to write the all-loop Baxter equation for the lowest trajectory.

4.2. Beyond one loop

Let us find the equation for the minimal trajectory starting from the multi-loop Baxter equation (2.3). Again, we have
to separate only terms which generate leading behavior in
the large-spin limit. The transfer matrix degenerates on the
minimal trajectory to the one of twist-two operators, i.e.,
t(u) ∼ u^L−2(2u^2 + q_1u + q_2). Notice however that only \( O(g^0) \)
contributions to the charges \( q_{1,2}(g) \) can induce the leading effect
in the large-\( N \) limit since the quantum corrections grow at
most logarithmically with \( N \to \infty \). Therefore, we can replace
t(u) ∼ t^{(0)}(u) in the right-hand side of (2.3). Hence the reduced
Baxter equation admits the form
\[
e^{\sigma_{cusp}(x)}Q(u+i) + e^{\sigma_{cusp}(x)}Q(u-i) = \sigma^{(0)}(u)Q(u).
\]
Introducing again the ratio of the Baxter functions \( Q \) analogous
to Eq. (4.2), we can write again two asymptotic equations for
the two components of \( Q \). However, since we are interested
solely in the lowest trajectory, both equations generate the same
contributions to the anomalous dimension. Therefore, we may
consider only one of the resulting equations, e.g.,
\[
e^{\sigma_{cusp}(x)}Q(u+i) = \sigma^{(0)}(u)Q(u).
\]
Next, introducing the one- and all-loop Hamilton–Jacobi functions,
\[
S^{(0)}(u) = \ln Q^{(0)}(u), \quad S(u) = \ln Q(u),
\]
Eq. (4.10) can be rewritten by virtue of the one-loop degenerate
Baxter equation (4.4) for the lowest trajectory as follows
\[
S(u+i) - S^{(0)}(u+i) - S(u) + S^{(0)}(u) + \sigma_{cusp}(x) = 2\pi i m.
\]
(4.12)
Here \( m \) displays the ambiguity in choosing the branch of the
logarithm. Since the anomalous dimension (3.5) is expressed
in terms of the derivative of the Hamilton–Jacobi function, it
is instructive to differentiate both side of Eq. (4.12) with re-
spect to \( u \). Using the perturbative decomposition of the all-order
Hamilton–Jacobi function
\[
S(u) = S^{(0)}(u) + g^2 S_h(u), \quad S_h(u) = \sum_{\alpha=1}^{\infty} g^{2\alpha(n-1)} S^{(\alpha)}(u),
\]
and rescaling \( S_h \) by extracting its single logarithmic behavior
\[
i S_h(u) = \Sigma(u) \ln N,
\]
we finally arrive at the equation for the cusp anomaly
\[
\Sigma(u+i) - \Sigma(u) + \frac{1}{\sqrt{u^2 g^2}} \times \int_{-1}^{1} \frac{dt}{\pi} \sqrt{1-t^2} 2g^2 \Sigma (i/2 - gt) = 0.
\]
(4.15)
The cusp anomalous dimension is then found in terms of \( \Sigma \)
making use of Eq. (3.5) as
\[
\Gamma_{cusp}(g) = g^2 + \frac{g^4}{\pi} \int_{-1}^{1} \frac{dt}{\pi} \sqrt{1-t^2} \Sigma (i/2 - gt).
\]
As we will demonstrate below, there exists yet another ex-
pression for the cusp anomalous dimension in terms of the
rescaled Hamilton–Jacobi function \( \Sigma \) which leads to realiza-
tion of an iterative perturbative structure of \( \Gamma_{cusp} \) in gauge
theory. Eqs. (4.15) and (4.16) are the main results of this sec-
tion. If one shifts the spectral parameter as \( u \to u - \frac{1}{2} \), one
immediately realizes that the first two terms give the imagi-
ary part of \( \Sigma \) for real \( u \). Then the use of a dispersion relation
for the rescaled Hamilton–Jacobi function in the last term al-
low us to bring the equation into the form of a Fredholm
equation of the second kind. Then the large-\( x \) asymptotics of
the solution to this integral equation yields the cusp anomaly
\( 2\pi i m S_h(u + \frac{1}{2}) \to \infty = -[\Gamma_{cusp}(g)/g^2] \ln N \). However
below we choose a slightly different route to solve Eq. (4.15) at
weak coupling.

5. Weak-coupling expansion

We will seek the solution to the cusp equation (4.15) in the
form [17]
\[
\Sigma(u) = \int_{0}^{1} d\omega \omega^{iu-1} e^{-iu} \hat{\Sigma}(\ln \omega/\tilde{\omega}),
\]
(5.1)
with \( \tilde{\omega} = 1 - \omega \). This integral representation immediately diag-
onalizes the difference terms. The change of variables to \( p = \ln \omega/\tilde{\omega} \) brings Eq. (5.1) into the form of a Fourier transform.
However, before we proceed with the above transformation of
Eq. (4.15), we will manipulate it at first. We notice that the first
term in the infinite series expansion in Chebyshev polynomials
in Eq. (3.1) is determined by the all-order anomalous dimen-
sion. Therefore, we can separate it from the kernel and rewrite
the equation for the cusp anomaly \( \Gamma_{cusp} \) in a form which
immediately suggests yet another relation of the Hamilton–Jacobi
function to the cusp anomalous dimension. Performing these
steps, we find
\[
\sinh \left( \frac{p}{2} \right) \hat{\Sigma}(p) + \frac{\Gamma_{cusp}(g)}{g^3} J_1(gp)
\]
\[
+ \frac{gp}{2} \int_{0}^{\infty} dp' \frac{e^{-p'^2}}{p-p'} U(gp, gp') \hat{\Sigma}(p') = 0,
\]
(5.2)
where the kernel \( U \) is expressed in terms of the Bessel func-
tions,
\[
U(p, p') = J_1(p) \left[ J_0(p') - \frac{2}{p} J_1(p') \right]
- J_1(p') \left[ J_0(p) - \frac{2}{p} J_1(p) \right].
\]
(5.3)
An examination of Eq. (5.2) immediately suggests that the last
term dies out for \( p \to 0 \) much faster than the first two, which
scale linearly with \( p \). Therefore, we deduce yet another represen-
tation for \( \Gamma_{cusp} \) in terms of the solution \( \hat{\Sigma} \) to the cusp
equation (5.2), namely,
\[
\hat{\Sigma}(0) = -\frac{\Gamma_{cusp}(g)}{g^2}.
\]
(5.4)
At the same time, we can use Eq. (4.16) for the anomalous dimension in terms of the Hamilton–Jacobi function, such that we get

$$\hat{\Sigma}(0) = -1 - g \int_0^\infty \frac{dp}{p} e^{-p/2} J_1(gp) \hat{\Sigma}(p).$$  \hspace{1cm} (5.5)$$

This expression clearly displays the mixing of orders and thus exhibits an iterative structure of the perturbative series in coupling constant, i.e., the cusp anomaly at higher orders can be determined in terms of $\Sigma(p)$ at lower orders. Combining Eqs. (5.2)–(5.5) together we reproduce the cusp equation derived in Ref. [9].

Finally, let us solve the cusp equation perturbatively. Writing the expansion in coupling constant as

$$\hat{\Sigma}(p) = \frac{p/2}{\sinh p/2} \sum_{n=0}^\infty \beta_n \hat{\Sigma}_n(p),$$ \hspace{1cm} (5.6)

where the prefactor is extracted for the latter convenience, and substituting it into the cusp equation (5.2), we find for the few lowest order functions

$$\hat{\Sigma}_0(p) = -1,$$

$$\hat{\Sigma}_1(p) = \frac{\pi^2}{12} + \frac{1}{8} p^2,$$

$$\hat{\Sigma}_2(p) = -\frac{11}{720} \pi^4 + \frac{1}{8} \zeta(3) p - \frac{\pi^2}{96} p^2 - \frac{192}{7} p^4,$$

$$\hat{\Sigma}_3(p) = \frac{73\pi^6}{20160} - \zeta(3) \frac{p^3}{8} - \left( \frac{5}{16} \zeta(5) + \frac{\pi^4}{96} \right) \zeta(3) \frac{p^4}{96} + \frac{\pi^4}{480} p^2 - \frac{1}{96} \zeta(3) p^3 + \frac{\pi^2}{2304} p^4 + \frac{p^6}{9216},$$

The $p$-independent term in these expressions determines the cusp anomaly according to Eq. (5.4). The lowest six orders of $\Gamma_{\text{cusp}}$ read

$$\Gamma_{\text{cusp}}(g) = g^2 - \frac{\pi^2}{12} g^4 + \frac{11\pi^4}{720} g^6 - \left( \frac{73\pi^6}{20160} - \frac{\zeta(3)^2}{8} \right) g^8$$

$$+ \left( \frac{887\pi^8}{907200} - \frac{\pi^2}{48} \zeta(3)^2 - \frac{5}{8} \zeta(3) \zeta(5) \right) g^{10}$$

$$- \left( \frac{136883\pi^{10}}{479001600} - \frac{\pi^4}{240} \zeta(3)^2 - \frac{5\pi^2}{48} \zeta(3) \zeta(5) \right) g^{12}$$

$$- \left( \frac{51}{64} \zeta(5)^2 - \frac{105}{64} \zeta(3) \zeta(7) \right) g^{14}$$

$$+ \left( \frac{7680089\pi^{12}}{87178291200} - \frac{47\pi^6}{48384} \zeta(3)^2 + \frac{\zeta(3)^4}{64} \right) g^{16}$$

$$- \left( \frac{41\pi^4}{1920} \zeta(3) \zeta(5) - \frac{17\pi^2}{128} \zeta(5)^2 - \frac{35\pi^2}{128} \zeta(3) \zeta(7) \right) g^{18}$$

$$- \frac{273}{64} \zeta(5) \zeta(7) - \frac{147}{32} \zeta(3) \zeta(9) \right) g^{20} + \cdots.$$ \hspace{1cm} (5.7)

The two- and three-loop coefficients agree with Feynman diagram calculations of Refs. [19,20,15] and [22,23,12], respectively, and the rest with available predictions of Ref. [9]. The calculation can be extended to few dozens of terms in the series (5.6), but the results are too cumbersome to display here.

6. Outlook

In this note we proposed a multi-loop asymptotic Baxter equation for anomalous dimensions of arbitrary twist-$L$, spin-$N$ single-trace holomorphic Wilson operators in maximally supersymmetric Yang–Mills theory. We developed an approach for the asymptotic solution of the resulting equation for large values of spin $N$ and derived an all-order equation for the cusp anomaly which governs the Sudakov asymptotics of anomalous dimensions. The problem with the asymptotic nature of the equation was overcome by studying the lowest-energy trajectory which is insensitive to the twist of the operator in the single logarithmic regime $\text{Le}^L \ll N, L, N \to \infty$.

There are many questions which remain to be addressed. One has to constrain the amount of ambiguity left in restoration of higher loop effects from the lowest few terms of perturbative series for the dressing factors. The analysis of the strong-coupling expansion of $\Gamma_{\text{cusp}}$ is of special interest in light of available predictions for it from string theory [31]. A preliminary analysis reveals however that $g = \infty$ is an essential singularity of the cusp equation. Next, one has to understand how to incorporate wrapping effects to the Baxter equation (2.3) and to identify a putative microscopic spin chain standing behind it. An ultimate goal would be to generalize the all-order Baxter equation to all sectors of $N = 4$ super-Yang–Mills theory which is conceivably described by a long-range graded magnet.

Note added

Recently a new calculation was published of the four-loop cusp anomalous dimension using the unitarity technique [39]. Their numerical finding explicitly demonstrates that the prediction (5.7) based on the Baxter equation (2.3) with the dressing factor (3.1) is incorrect starting from four loops. In a companion paper [40], a modified form of the cusp equation was proposed which takes into account a nontrivial dressing factor in Bethe equations of Ref. [18].

Presently we use the result of Ref. [39] in order to fix the form of the four-loop correction to the Baxter equation (2.3) and find anomalous dimensions of local Wilson operators. It was suggested [39], that to reconcile within error bars the result of their numerical calculation with the one coming from the cusp equation, the sign of the $\zeta^2(3)$ in four-loop contribution of Eq. (5.7) has to be flipped. This requires the following additive modification of the four-loop cusp anomaly (5.7),

$$\Gamma_{\text{cusp}}(g) = -\frac{\zeta(3)^2}{4} g^8 + \mathcal{O}(g^{10}).$$ \hspace{1cm} (6.1)
In order to generate it from the cusp equation, one has to add the following term\(^4\) to the left-hand side of Eq. (5.2)
\[
\cdots + 2\alpha g^2 J_2(gp) \left( 1 + g \int_0^\infty dp' e^{-p'/2} \hat{\Sigma}(p') \frac{J_1(gp')}{p'} \right) + \mathcal{O}(g^3),
\]  
(6.2)
in agreement with Ref. [18]. Here the favored value of the constant is \(\alpha = \frac{1}{2}\zeta(3)\) [39,40]. This translates into a modification of the integrand in Eq. (4.15),
\[
\frac{1}{u_+ + gt} \to \frac{1}{u_+ + gt} + \frac{i\alpha g^4}{x_+^2} + \mathcal{O}(g^6).
\]  
(6.3)
A simple analysis allows to unambiguously restore the correction term to the dressing factors (3.1) of the Baxter equation (2.3). Namely, the former get shifted as
\[
\sigma_{\pm}(x) \to \sigma_{\pm}(x) + \Delta_{\pm}(x),
\]  
(6.4)
with
\[
\Delta_{\pm}(x) = \mp \frac{i\alpha g^6}{2x_+^2} \int_1^{\infty} \frac{dt}{\pi} \sqrt{1 - t^2} \ln \mathcal{Q} \left( \pm \frac{i}{2} - gt \right)^3 + \mathcal{O}(g^7).
\]  
(6.5)
Taking into account this extra term, the anomalous dimensions of Wilson operators acquire additional contributions. For instance, the four-loop term in Eq. (3.6) gets corrected by
\[
\gamma(g) = \cdots - \frac{5 \pm \sqrt{5}}{8} g^8 + \mathcal{O}(g^{10}).
\]  
(6.6)
This explicitly demonstrates that the attempt to rescue the principle of maximal transcendentality [23] in the cusp anomalous dimension with \(\alpha = \frac{1}{2}\zeta(3)\) results in breaking of the rational form of anomalous dimensions of local Wilson operators, i.e., they acquire transcendental addenda (6.6) in addition to rational terms (3.6). At the current state-of-the-art of higher loop calculations such terms are not ruled out yet. Within Mueller’s cut vertex technique [41], the main sources of transcendentals constants in local anomalous dimensions comes from virtual self-energy and vertex corrections with rational terms being generated by real cuts. The finiteness of maximally supersymmetric Yang–Mills theory, especially transparent in the light-cone gauge where Ward identities imply equality of the vanishing beta function with all field renormalization constants, seems to suggests the absence of transcendentals constants in local anomalous dimensions. This question deserves however a thorough study.

\footnote{In the unnumbered equations above (5.2), it yields corrections to the right-hand side the equations, i.e., \(\hat{\Sigma}_2(p) = \cdots + \frac{1}{2} \alpha p, \hat{\Sigma}_3(p) = \cdots + \frac{1}{2} \alpha \zeta(3) + \frac{1}{24} \alpha p(p^2 + p)\).}

**Acknowledgement**

This work was supported by the US National Science Foundation under grant No. PHY-0456520.

**References**

Renormalization group flows for the second $Z_5$ parafermionic field theory

Vladimir S. Dotsenko *, Benoit Estienne

Laboratoire de Physique Théorique et Hautes Energies, 1 Université Pierre et Marie Curie, Paris VI; CNRS, Université Denis Diderot, Paris VII, Boîte 126, Tour 25, 5ème étage, 4 place Jussieu, F-75252 Paris Cedex 05, France

Received 6 November 2006; accepted 11 November 2006

Editor: L. Alvarez-Gaumé

Abstract

Using the renormalization group approach, the Coulomb gas and the coset techniques, the effect of slightly relevant perturbations is studied for the second parafermionic field theory with the symmetry $Z_5$. New fixed points are found and classified.

© 2006 Elsevier B.V. All rights reserved.

While the first series of parafermionic conformal field theories [1] is well studied and applied in various domains [2–4], the second parafermionic series, with the symmetry $Z_5$, has been developed fairly recently [5–8] and it still awaits its applications.

In the case of the first series, to a given $N$ (of $Z_N$) is associated a single conformal theory. This is different for the second series: for a given $Z_N$, there exist an infinity of unitary conformal theories $Z_{N,p}$, with $p = N - 2 + k, k = 1, 2, 3, \ldots, \infty$. These theories correspond to degenerate representations of the corresponding parafermionic chiral algebra. They are much more rich in their content of physical fields, as compared to the theories of the first series. They are also much more complicated. But, on the other hand, the presence of the parameter $p$, for a given $Z_N$, opens a way to reliable perturbative studies. It allows in particular to study the renormalization group flows in the space of these conformal theory models, under various perturbations.

In this Letter we shall present results for a particular case of this problem: for the renormalization group flows of the $Z_{5,p}$ theories, being perturbed by two slightly relevant fields.

The details of the $Z_5$ parafermionic theory, the second one, could be found in [5]. The $q$ charge of $Z_5$ takes values $q = 0, \pm 1, \pm 2$, so that in the Kac table of this theory one finds the $Z_5$ neutral fields, of $q = 0$, the $q = \pm 1$ and the $q = \pm 2$ doublets, and the $Z_2$ disorder fields. The symmetry of the theory is actually $D_5$, which is made of $Z_5$ rotations and the $Z_2$ reflections in 5 different axes. These last symmetry elements amount to the charge conjugation symmetry: $q \rightarrow -q$.

We want to perturb by the $Z_5$ ($D_5$ in fact) neutral fields, in order to preserve this symmetry. Perturbatively well controlled domain of $Z_{5,p}$ theories is that of $p \gg 1$, giving a small parameter $\epsilon \sim 1/p$. This is similar to the original perturbative renormalization group treatment of minimal models for Virasoro algebra based conformal theory [9,10].

In this domain, i.e. for $p \gg 1$, one finds, in the lower part of the Kac table of $Z_5$ parafermionic theory, two $Z_5$ neutral fields which are slightly relevant and which close by the operator algebra. They are:

\begin{align*}
S &= \Phi_{(1,1)(3,1)}, \\
A &= A_{-2/5}\Phi_{(1,1)(1,3)}.
\end{align*}

The first one is a $Z_5$ singlet and the second is a parafermionic algebra descendant of a doublet $q = 1$ field. They both belong to the neutral $q = 0$ sector of $Z_5$ and they are both Virasoro algebra primaries.

Their labeling as $S$ and $A$ is just our shortened notations (in this Letter) for these fields.

In general, the parafermionic algebra primaries of the second $Z_5$ conformal theory are labeled by double indices of two
\( (\alpha_+ \text{ and } \alpha_-) \) lattices of the \( B_2 \) classical Lie algebra [5]:

\[
\Phi_{\{n_1,n_2|n'_1,n'_2\}}.
\]

The first and second couples of indices correspond respectively to the \( \alpha_+ \) and \( \alpha_- B_2 \) lattices. \( \alpha_+ \) and \( \alpha_- \) are the usual Coulomb gas type parameters. For the \( Z_{5,p} \) conformal theory they take the values:

\[
\alpha_+ = \sqrt{\frac{p+2}{p}}, \quad \alpha_- = -\sqrt{\frac{p}{p+2}}.
\]  

The formulas for the conformal dimensions of the fields (3) could be found in [5]. One could check that the dimensions of the fields \( S \) and \( A \) in (1) and (2) have the following values:

\[
\Delta_S = \frac{5}{2} \alpha^2 - \frac{3}{2} = 1 - 5\epsilon,
\]

\[
\Delta_A = \frac{3}{2} \alpha^2 - \frac{1}{2} = 1 - 3\epsilon.
\]

We have defined \( \epsilon \) as follows:

\[
\alpha^2 = \frac{p+2}{p} = 1 - 2 \epsilon, \quad \epsilon = \frac{1}{p+2} \simeq \frac{1}{p}.
\]  

Perturbing with the fields \( S \) and \( A \) corresponds to taking the action of the theory in the form:

\[
\Lambda = A_0 + \frac{2g}{\pi} \int d^2x \, S(x) + \frac{2h}{\pi} \int d^2x \, A(x),
\]

where \( g \) and \( h \) are the corresponding coupling constants; the additional factors \( \frac{2}{\pi} \) are added to simplify the coefficients of the renormalization group expansions which follow; \( A_0 \) is assumed to be the action of the unperturbed \( Z_{5,p} \) conformal theory.

It will be shown below that the operator algebra of the fields \( S \) and \( A \) is of the form:

\[
S(x') S(x) = \frac{D_1}{|x' - x|^4} A(x) + \cdots,
\]

\[
A(x') A(x) = \frac{D_2}{|x' - x|^3} A(x) + \cdots,
\]

\[
S(x') A(x) = \frac{D_1}{|x' - x|^2} S(x) + \cdots.
\]

Only the fields which are relevant for the renormalization group flows are shown explicitly in the r.h.s. of Eqs. (9)-(11). For instance, the identity operator is not shown in the r.h.s. of (9) and (10) while it is naturally present there. The operator algebra constants in (9) and (11) should obviously be equal, as the two equations could be related to a single correlation function \( \langle S(x_1) S(x_2) A(x_3) \rangle \).

Assuming the operator expansions in (9)-(11), one finds, in a standard way, the following renormalization group equations for the couplings \( g \) and \( h \):

\[
\frac{dg}{d\xi} = 2 \cdot 5\epsilon \cdot g - 4 D_1 gh,
\]

\[
\frac{dh}{d\xi} = 2 \cdot 3\epsilon \cdot h - 2 D_2 h^2 - 2 D_1 g^2.
\]

These are up to (including) the first non-trivial order of the perturbations in \( g \) and \( h \).

The problem now amounts to justifying the operator algebra expansions in (9)-(11) and to calculating the constants \( D_1 \) and \( D_2 \).

The efficient method for calculating the operator product expansions and defining the corresponding coefficients, is that of the Coulomb gas technique.

Calculating directly the expansions of the products of the operators (1), (2) encounters a problem: the explicit form of the Coulomb gas representation for the \( Z_5 \) theory is not known. We shall get around this problem by using the coset representation for the \( (\text{second} Z_5 \text{ theory and the related techniques. In particular, we shall generalize the method developed in papers [11,12] for the SU(2) cosets.})

The second \( Z_5 \) theory, which we shall denote as \( Z_{5,p}^{(2)} \), could be represented as the following coset construction [13]:

\[
Z_{5,p}^{(2)} = \frac{SO_k(5) \times SO_2(5)}{SO_{k+2}(5)}.
\]

Here \( SO_k(5) \) is the orthogonal affine algebra of level \( k \); \( p = 3 + k \). This coset could be rewritten as follows:

\[
Z_{5,p}^{(2)} \times \frac{SO_1(5) \times SO_1(5)}{SO_2(5)} = \frac{SO_k(5) \times SO_1(5)}{SO_k(5) \times SO_{k+1}(5) \times SO_1(5)}.
\]

The two coset factors in the r.h.s., as well as the additional coset factor in the l.h.s., correspond to the \( WB_2 \) theories [14]. For these theories the Coulomb gas representation is known. It is made of two bosonic fields, quantized with a background charge and the Ising model fields: \( \Psi \) (free fermion) and \( \sigma \) (spin operator) [14].

Eq. (15) could be rewritten as

\[
Z_{5,p}^{(2)} \times WB_{2,1} = WB_{2,k} \times WB_{2,k+1}.
\]

This equation relates the representations of the corresponding algebras. It could be reexpressed in terms of characters of representations, as is being usually done in the analyses of cosets. But this equation allows also to relate the conformal blocks of correlation functions. In doing so one relates the chiral (holomorphic) factors of physical operators. This later approach has been developed and analyzed in great detail in the papers [11,12], for the \( SU(2) \) coset theories.

As it was said above, the chiral factor operators are related to the conformal bloc functions, not to the actual physical correlators. On the other hand, the coefficients of the operator algebra expansions are defined by the three point functions. These latter are factorizable, into holomorphic–antiholomorphic functions. So that, when the relation is established on the level of chiral factor operators, for the holomorphic three point functions, this relation could then be easily lifted to the relation for the physical correlation functions. Saying it differently, with the relations for the chiral factor operators one should be able to define the square roots of the physical operator algebra constants.

By matching the conformal dimensions of operators on the two sides of the coset equation (16) one finds the following decompositions for the operators \( S \) and \( A \) in (1) and (2) (their
chiral factors in fact):
\[
\Phi_{(11|31)}(Z_{\psi}) \times \Phi_{(11|11)}(W_{B_2}) = \Phi_{(11|21)}(W_{B_2}) \times \Phi_{(21|31)}(W_{B_2}),
\]
(17)
\[
A_{-3/2} \Phi_{(11|13)}(W_{B_2}) \times \Phi_{(11|11)}(W_{B_2}) = a \Phi_{(11|11)}(W_{B_2}) \times \Phi_{(11|13)}(W_{B_2})
+ b \Phi_{(11|13)}(W_{B_2}) \times \Phi_{(13|33)}(W_{B_2}).
\]
(18)
The coefficients \(a\) and \(b\) in (18) are still to be determined. \(\Phi_{(11|11)}\) are the identity operators. They could actually be dropped. But we shall keep them sometimes, when this makes the decomposition more explicit. The operator \(\Phi_{(11|11)}\) in the l.h.s. of (17), (18) could be definitely suppressed.

By Eqs. (17), (18), one observes that to decompose the products \(SS\), \(AA\), \(SA\), as in Eqs. (9)–(11), one needs to know the decompositions of products of the operators of \(W_{B_2}k\) and \(W_{B_2,k+1}\) theories:
\[
\Phi_{(11|21)}(W_{B_2}) \times \Phi_{(11|21)}(W_{B_2}) \times \Phi_{(21|31)}(W_{B_2,k+1}) \times \Phi_{(21|31)}(W_{B_2,k+1}),
\]
eq \sum_{l_1,l_2} a(l_1,l_2) \Phi_{(n_1,n_2,l_1,l_2)} \times \Phi_{(n_1,n_2,l_1,l_2)}.
\]
(21)
The operators in this relation could be primaries or their descendants. Eqs. (17), (18) are two particular examples of Eq. (21). These decompositions will be discussed in more detail in [15]. They generalize the corresponding relations for the \(SU(2)\) cosets of [11,12].

The “diagonal” cross-products correspond to products of \(W_{B_2,k}\) and \(W_{B_2,k+1}\) operators of the type which appear in the r.h.s. of (21). The rest of possible cross-products have to be dropped when doing expansions.

The above features, square roots of constants and keeping the diagonal terms only, are due to the fact that we are dealing with the conformal bloc functions and not with the actual physical correlators. These features are discussed in much detail in the paper [12].

Equally, the overall factors of the resulting expressions (the expressions which should correspond to the decomposed \(Z_{\xi}^2\) operators) provide the square roots of the \(Z_{\xi}^2\) structure constants, and not the constants themselves. In this way one obtains the square roots of the values of \(D_1\), \(D_2\) in Eq. (20).

Substituting now the values of \(D_1\), \(D_2\) into the renormalization group equations (12), (13) and analysing them by the standard methods one obtains the following results.

The phase diagram of constants \(g\) and \(h\) contains:

1. the initial fixed point \(g_0^* = h_0^* = 0\);
2. the fixed point on the \(h\) axis: \(g_0^* = 0, h_0^* = \sqrt{10} \epsilon\);
3. two additional fixed points for non-vanishing values of the two couplings: \(g_2^* = \sqrt{\frac{3}{2}} \epsilon, h_2^* = \frac{\sqrt{10}}{\sqrt{2}} \epsilon\), and \(g_4^* = -\sqrt{\frac{3}{2}} \epsilon, h_4^* = \frac{\sqrt{10}}{\sqrt{2}} \epsilon\).

The renormalization group flows are shown in Fig. 1. They are symmetrical with respect to \(g \rightarrow -g\).
The value of the central charge at the point $g^*_1 = 0, h^*_1 = \sqrt{10} \epsilon$ agrees with that of the theory $Z^{(2)}_{5,p-2}$. This confirms the observation, made with the $SU(2)$ cosets [11, 12] and, more generally, with the cosets for the simply laced algebras [16], that the perturbation of a coset theory caused by an appropriate operator drives $p$ to $p - \Delta p$, $\Delta p$ being equal to the shift parameter of the coset. In our case the shift parameter of the coset is equal to 2, Eq. (14). Note that the algebra $B_2 \equiv SO(5)$ is not a simply laced one.

On the other hand, the appearance of two extra fixed points, $(g^*_2, h^*_2)$ and $(g^*_3, h^*_3)$, is somewhat surprising. By the value of the central charge, the two critical points correspond to the theory $Z^{(2)}_{5,p-1}$. This assignment has further been verified by calculating the critical dimension of the operator $\Phi(1, n | 1, n)$ at these points.

We observe that such additional fixed points do not appear in the parafermionic model $Z^{(2)}_3$: the second $Z_3$ parafermionic theory with $\Delta \Psi = 4/3$ [17]. This model could be realized by the $SU(2)$ cosets. Its perturbations, with two slightly relevant operators, have been analysed in [11, 12].

Further analysis and discussions will be left for the paper [15].

Acknowledgement

Useful discussion with V.A. Fateev is gratefully acknowledged.

References

A new method for calculating differential distributions directly in Mellin space

Alexander Mitov

Deutsches Elektronensynchrotron DESY, Platanenallee 6, D-15735 Zeuthen, Germany

Received 13 December 2005; accepted 30 October 2006

Available online 9 November 2006

Abstract

We present a new method for the calculation of differential distributions directly in Mellin space without recourse to the usual momentum-fraction (or z-) space. The method is completely general and can be applied to any process. It is based on solving the integration-by-parts identities when one of the powers of the propagators is an abstract number. The method retains the full dependence on the Mellin variable and can be implemented in any program for solving the IBP identities based on algebraic elimination, like Laporta. General features of the method are: (1) faster reduction, (2) smaller number of master integrals compared to the usual z-space approach and (3) the master integrals satisfy difference instead of differential equations. This approach generalizes previous results related to fully inclusive observables like the recently calculated three-loop space-like anomalous dimensions and coefficient functions in inclusive DIS to more general processes requiring separate treatment of the various physical cuts. Many possible applications of this method exist, the most notable being the direct evaluation of the three-loop time-like splitting functions in QCD.

© 2006 Elsevier B.V. All rights reserved.

1. Introduction

Achieving high precision in theoretical predictions is vital for the success of present and future collider experimental programs, as well as for the effective extraction of new physics from experimental data. A significant part of the theoretical work related to the experiment requires the evaluation of differential distributions, with most current research efforts focusing on the next-to-next-to-leading order (NNLO) or a higher level of precision. Examples of such distributions are the fully inclusive [1–7] and one-particle inclusive [8,9] DIS, the energy spectrum of hadrons in $e^+e^-$ collisions [10–12], the total partonic cross-section [13–16] and rapidity distribution [17] for Higgs and vector boson [18] production at hadron colliders, Drell–Yan [19–21], transverse distribution of hadrons at hadron colliders [22–26] or particle spectra in the decays of muon [27] or heavy flavors [28–33]. Another important class of distributions that are universal and thus underlay the description of many physical processes includes the space- and time-like splitting functions [34–38], heavy flavor matching conditions [39] and the heavy quark perturbative fragmentation function [40–42].

The various distributions can be classified according to the number of kinematical variables they involve. Clearly, the larger the number of variables, the more complicated the evaluation of a distribution becomes. In this Letter we will restrict our discussion to the case of distributions with a single kinematical variable. This class of distributions involves many important examples—some of them still significant open problems—like the three-loop time-like splitting functions in QCD. The extension of our discussion to cases with more than one variable will be rather transparent.

The choice of the most efficient approach to the evaluation of a particular single-scale distribution depends on its degree of “inclusiveness”. The fully inclusive observables, like the fully inclusive coefficient functions in DIS [6], allow a simplified treatment based on the optical theorem. This is however a rare
situation; most distributions of interest involve a specific final state, which requires that all contributing physical cuts of the relevant amplitudes be evaluated separately.

The purpose of this Letter is to present a conceptually new calculational method of general applicability. As will become clear from the subsequent discussion, this method builds a bridge between two very important and seemingly unrelated calculational approaches as it provides a new perspective on the calculation of single-scale distributions. Moreover, during all stages of calculation this method requires no custom work and utilizes tools, techniques and programs that are publicly available and easy to implement in practice. Our method relies heavily on the integration by parts (IBP) identities [43]. It has the important feature of being formulated in terms of variables that are the most natural ones for the effective solving of the IBP identities.

With the above-described applications in mind, let us properly introduce the type of distributions $\sigma(z)$ that we will be dealing with in this Letter. Such distributions depend on a single kinematical variable $z$. For example, $z$ can be the energy fraction of a parton produced in $e^+e^-$ annihilation. We will assume that this variable is conveniently normalized: $0 \leq z \leq 1$. The distribution $\sigma$ is a scalar that is typically of the following form:

$$\sigma(z) = \int dPS^{(m)} |M([\text{in}] \to [\text{out}])|^2 \delta(z - f).$$

(1)

The factor $dPS^{(m)}$ in Eq. (1) is related to the phase-space for the $m$-particle final state; it also contains the measure for the virtual integrations (if present). The precise form of this factor depends on the number of particles in the initial state. For a single-particle initial state processes with no virtual corrections it reads:

$$dPS^{(m)} = (2\pi)^d \delta(p_{\text{in}} - \sum p_{\text{out}}) \prod_{i=1}^{m}[dq_i].$$

$$[dq_i] = \frac{d^d q_i}{(2\pi)^d} \delta(q_i^2 - m_i^2).$$

(2)

In the case of processes with two particles in the initial state, $dPS^{(m)}$ has similar structure. It is detailed, for example, in [14].

Typically, expressions like Eq. (1) are UV and infrared divergent and in the following we assume that all divergences have been properly regulated by means of dimensional regularization. Besides $z$, the distribution $\sigma(z)$ can depend on other parameters. Since their presence is irrelevant to our discussion, we will assume in the following that these have some fixed values and we will suppress them in our notations. The function $f$ appearing in the argument of the $\delta$-function in Eq. (1) is a dimensionless scalar. Its form is specific for each particular observable.

As a typical example we will consider the evaluation of the single particle inclusive cross-section for massless quark production in the decay of a colorless particle $V \to q + X$ at $O(\alpha_S)$.

![Fig. 1. Real emission diagrams contributing to the decay of a colorless particle $V(p) \to q(p - q - k) + X$ at $O(\alpha_S)$.](image)

Figures and diagrams are not provided in the natural text format. Expressions and identities (i.e. no UV renormalization is performed) one has:

$$\frac{d\sigma}{dz} = \int [dp_q][dp_{\bar{q}}]M^{(0)}(V \to q + \bar{q})^2 \delta\left(z - \frac{2p_V \cdot p_q}{p_V^2}\right)$$

$$+ \int [dp_q][dp_{\bar{q}}] \frac{d^d p_{\bar{g}}}{(2\pi)^d} |M^{(1)}(V \to q + \bar{q})|^2$$

$$\times \delta\left(z - \frac{2p_V \cdot p_{\bar{g}}}{p_V^2}\right)$$

$$+ \int [dp_q][dp_{\bar{q}}][dp_{\bar{g}}] |M^{(1)}(V \to q + \bar{q} + g)|^2$$

$$\times \delta\left(z - \frac{2p_V \cdot p_{\bar{g}}}{p_V^2}\right).$$

(3)

In the example above, $|M^{(k)}|^2$ denotes the terms proportional to $\alpha_S^k$ in the squared matrix element for the process $V \to q + X$ (see Fig. 1). Clearly, the first line in Eq. (3) corresponds to the tree-level (Born) contribution while the second and the third lines respectively contain the contributions from the virtual and real-gluon emission corrections at order $\alpha_S$. On the above example, Eq. (1) stands for any one of the three lines in Eq. (3).

Perhaps the most elegant approach to date for the evaluation of distributions of the type Eq. (1) was proposed by Anastasiou and Melnikov [14] and further elaborated upon in [17,18]. Let us recall the salient features of this method. One uses the distributional identity:

$$2\pi i \delta(x) = \frac{1}{x + i\epsilon} - \frac{1}{x - i\epsilon},$$

to formally replace all $\delta$-functions appearing in Eq. (1) with propagators coinciding with the arguments of the $\delta$-functions, i.e. one introduces the invertible mapping $\hat{P}$ acting only on $\delta$-functions:

$$\hat{P}\left[c \prod_i \delta(x_i)\right] = c \prod_i \frac{1}{x_i},$$

(4)

with $c$ an arbitrary function of the propagators. The utility of the mapping (4) is that it allows one to treat the object on the right-hand side of Eq. (4) with the usual IBP identities [43]. By solving these identities one reduces the initial distribution $\sigma(z)$ to a combination of a small number of irreducible objects. Eventually, one performs the inverse mapping $\hat{P}^{-1}$, thus expressing $\sigma(z)$ as a linear combination (with simple known coefficients) of a small number of well defined master integrals. From the IBP identities it also follows that the master integrals satisfy a system of differential equations, which can be solved to obtain their $z$-dependence. To fully specify the solutions of
the differential equations, one has to prescribe the corresponding boundary conditions; these can be extracted from an explicit evaluation of the master integrals in a particular kinematical point like \( z = 1 \).

In the phenomenological applications one also needs the Mellin transform\(^1\) of the distribution in question:

\[
\sigma(n) = \int_{0}^{1} dz \, z^n \sigma(z). \quad (5)
\]

Performing the Mellin transform results in the evaluation of integrals over, typically, combinations of polylogarithms and rational functions of \( z \). At present, and certainly for the case of massless distributions, there exists a very good understanding of the mapping between the classes of basic functions in \( z \) and \( n \) spaces \([44–49]\). In the following discussion we will consider the knowledge of \( \sigma(z) \) as equivalent to that of \( \sigma(n) \) and vice versa, i.e. we will tacitly assume that one can always perform the needed Mellin or inverse Mellin transforms. That is definitely true for the massless case. In more complicated situations one may have to resort to numerical methods to perform the inverse Mellin transform \([50,51]\). In any case, we need not bother about that point here. We will consider our problem as solved, as long as we know either \( \sigma(z) \) or \( \sigma(n) \).

2. The method

In the present Letter we would like to advocate a new approach to the evaluation of the distribution in Eq. (1). It aims at the direct evaluation of \( \sigma(n) \) without calculating it first in \( z \)-space as is done at present. Our proposal is to explore the obvious possibility that one can integrate over \( z \) before performing the phase-space and/or virtual integrations:

\[
\sigma(n) = \int_{0}^{1} dz \, z^n \sigma(z)
= \int_{0}^{1} dz \, z^n \int dPS^{(m)} |M([\text{in}] \rightarrow [\text{out}])|^2 \delta(z - f)
= \int dPS^{(m)} |M([\text{in}] \rightarrow [\text{out}])|^2 \frac{1}{(f)^n}.
\quad (6)
\]

We see that as a result of the interchange of the order of integrations the invariant \( f \) enters the integrals as a propagator raised to power \( -n \). That power, however, should be treated as an abstract parameter that takes arbitrary and not fixed integer values.

By applying the mapping Eq. (4), and interchanging the order of integration as in Eq. (6), one can bring the original problem of calculating \( \sigma(z) \) to the following form:

\[
\hat{P}[\sigma(n)] = \sum \int d\mathcal{D} \frac{1}{P_1^{a_1}} \cdots \frac{1}{x_1^{a_1}} \cdots \frac{1}{f^{n+s}},
\quad (7)
\]

where \( \mathcal{D} \) represents the appropriate measure originating from the real and/or virtual integrations, \( P_i \) denote the propagators originating from the evaluation of the amplitude and \( x_i \) are the arguments of all phase-space \( \delta \)-functions (if present). The argument of the \( \delta \)-function that defines the observed fraction \( z \) is denoted by \( f \) and the powers \( a_i \) and \( s \) are some fixed integers.

This way, we have effectively reduced the problem of the calculation of the differential cross-section \( \sigma(z) \) to the problem of evaluation of functions of the following general form:

\[
S(a_1, \ldots, a_p) = \int \mathcal{D} \frac{1}{P_1^{a_1} \cdots P_n^{a_n}}.
\quad (8)
\]

The function \( S \) appearing in Eq. (8) can also depend on other fixed parameters.

In principle, scalar terms like the one in Eq. (8) can be simplified to a minimal set of terms by applying the integration by parts (IBP) identities \([43]\). Efficient, readily implementable methods for IBP’s solving presently exist only for the cases where all the powers \( a_1, \ldots, a_p \) are fixed integers. In the following, we present one very efficient approach for solving the IBP identities when one (or more) of the powers \( a_1, \ldots, a_p \) is an abstract parameter. The basic idea is to replace the problem with abstract power(s) with a problem having only fixed integer powers and then solve the latter with existing standard methods (like Laporta \([52]\) implemented in the program AIR \([53]\)). Previous works that have discussed the solving of the IBP’s in presence of abstract powers are \([54]\) and \([52]\).

3. Solving the IBP Identities in presence of an abstract power

It is very well known (see for example the book of Smirnov \([55]\) for detailed introduction) that when applied to the object \( S \) in Eq. (8), the IBP identities result in a system of linear homogeneous equations with rational coefficients that relate functions \( S \) with arguments shifted by \( \pm 1 \) relative to each other. If all \( a_i \)’s were fixed integers, then by successively relating terms that differ with \( \pm 1 \) one can eventually express the original function \( S(a_1, \ldots, a_p) \) through a linear combination of several, say \( m \), master integrals \( S_1(i_1) \), \ldots, \( S_m(i_m) \). The masters \( S_j \) are special cases of \( S(a_1, \ldots, a_p) \) with their arguments \( (i_j) \) taking special values. At present, the most popular method for solving the IBP identities is the one of Laporta \([52]\). It is based on solving the systems of linear homogeneous equations directly, through Gauss elimination.

Clearly, if one of the parameters \( a_i \) is not an integer this procedure cannot work, since: (1) with only integer steps one cannot relate the initial non-integer parameter to an element \( S(a_1, \ldots, a_p) \) with only integer \( a_i \)’s and (2) the number of steps in the Gauss elimination cannot even be specified when one of the parameters is an abstract number.

In the following, we detail a simple approach to solving this problem (at the end of this section we compare it to the one of Laporta \([52]\) for deriving difference equations for master integrals). To facilitate our discussion we shall assume that \( a_1, \ldots, a_p \) are integers having some specific values, while the last argument, \( a_p \), is an abstract parameter.

\(^1\) Note that usually the Mellin transform is defined through the variable \( N = n + 1, N \geq 1 \).
From Eq. (1) it is clear that $\sigma(n)$ is a sum of a number of terms of the type in Eq. (8) that have different values of their indexes $(a_1, \ldots, a_p)$. It is very important to observe, however, that the difference between any two values that the index $a_p$ can take is always an integer, i.e. $a_p - a'_p \in \mathbb{N}$.

This is a crucial observation, which one can use to modify the strategy for solving the IBP identities in the following way. First, one relaxes the requirement that the masters must have integer-valued indexes. Second, as we will explain in a moment, one can choose all masters in such a way that they all have the same value, say $a_p = r \notin \mathbb{N}$, of their last index i.e. the masters are all of the form:

$$ S(j_1, \ldots, j_{p-1}, r) = \int \frac{D}{p_1^a \ldots p_{p-1}^b} \frac{1}{f^r}, $$

with the same $r$, and the $j_i$’s being fixed integers specific to each master.

It is indeed possible to arrange that all masters have the same value of the non-integer-valued index $a_p$. That follows from the arbitrariness of the value of this parameter (we only assume that it is non-negative). Since there is no preferred value for that index, the IBP system has a sort of translational invariance along the index $a_p$. One can understand this by saying that $r$ and $r+k$, where $k$ is a fixed integer, are equally arbitrary. Therefore, we can take as a reference value for the index $a_p$ the number $r$ which we will consider abstract but having fixed value. Having done that, the “translational” invariance along the values of $a_p$ is now “broken”. Clearly, the value $r$ now plays the role of a zero reference point much like the value $a_p = 0$ in the usual case when all indexes take integer values. Therefore all one needs to do is to measure in integer units how much the value of the last index of an element $S$ is displaced from the reference point $r$.

Next, we give a practical recipe of how to implement the above idea. Let us work with the functions $2$ $B$:

$$ B(a_1, \ldots, a_{p-1}, k) = \int \frac{D}{p_1^{a_1} \ldots p_{p-1}^{a_{p-1}}} \frac{1}{f^{-a+k}}, \hspace{1cm} (9) $$

where as the “reference” point for the last index we take the Mellin variable $n$.

As follows from Eqs. (6) and (7) the distribution $\sigma(n)$ takes the following form:

$$ \sigma(n) = \sum_{a_1, \ldots, a_{p-1}} c_{a_1, \ldots, a_{p-1}} B(a_1, \ldots, a_{p-1}, k), \hspace{1cm} (10) $$

where $c_{a_1, \ldots, a_{p-1}}$ are some known coefficients. To construct the needed algebraic reductions, one first applies the IBP identities on a generic monomial of the form:

$$ \frac{1}{p_1^{v_1} \ldots p_{p-1}^{v_{p-1}}} f^{v_p}, \hspace{1cm} (11) $$

where all powers $v_i$ are treated as arbitrary parameters. Next, one identifies each term of the form (11) appearing in the IBP equations, with the function $B(v_1, \ldots, v_{p-1}, v_p + n)$, followed by the substitution $v_p \rightarrow v_p - n$. After this manipulation the Mellin variable $n$ is explicitly present as a parameter in the resulting equations. They can be solved in any approach available, including the one of Laporta.

A word of caution: one has to keep in mind that, as follows from Eq. (9), the functions $B$ implicitly depend on $n$. Therefore, one should not confuse the integer value $k$ in the last argument of the function $B$ with the absolute power of the corresponding propagator $1/f$, but should think of it as the “distance”—in integer units—from the reference power $n$.

Finally, one can map all integrals appearing in Eq. (10) to the masters obtained from the solving of the just-described reduction. This mapping is done in the standard way.

Next, we explain how one can extract the $n$ dependence of the master integrals. Assume that an element $B(b_1, \ldots, b_{p-1}, 0)$ is a master integral (with $b_1, \ldots, b_{p-1}$ some fixed integers). One can inspect the already solved IBP reduction and read off from there the result for the element $B(b_1, \ldots, b_{p-1}, -1)$. Note that this element differs from the master $B(b_1, \ldots, b_{p-1}, 0)$ only by the value of the last index. If the element $B(b_1, \ldots, b_{p-1}, -1)$ is not a masters itself, then it must be a linear combination of the master integrals:

$$ B(b_1, \ldots, b_{p-1}, -1) = c(n) B(b_1, \ldots, b_{p-1}, 0) + G(n). \hspace{1cm} (12) $$

Here $c(n)$ is a known, typically not very complicated function, and the term $G(n)$ is a homogeneous linear combination of all master integrals, except for the master $B(b_1, \ldots, b_{p-1}, 0)$. Eq. (12) is a first-order non-homogeneous difference equation of the type $F(n+1) = c(n) F(n) + G(n)$ (recall Eq. (9)) for the master $B(b_1, \ldots, b_{p-1}, 0)$. Clearly, repeating this procedure for each one of the master integrals found in the reduction run, one can derive a complete system of difference equations for all the masters. Typically, one observes certain hierarchy among the master integrals; the simplest ones satisfy homogeneous equations (i.e. $G(n) = 0$) that can be solved in terms of $\Gamma$-functions. These integrals then comprise the non-homogeneous terms for the equations of other masters, and so on.

In case the element $B(b_1, \ldots, b_{p-1}, -1)$ is also a master integral, one should read off from the reduction the result for the yet higher term $B(b_1, \ldots, b_{p-1}, -2)$. One should continue doing this until one reaches an element $B(b_1, \ldots, b_{p-1}, -k)$ which is not a master itself but all elements $B(b_1, \ldots, b_{p-1}, -s)$ with $0 \leq s < k$ are masters. The result from the reduction for the element $B(b_1, \ldots, b_{p-1}, -k)$ represents a $k$th order difference equation for the master integral $B(b_1, \ldots, b_{p-1}, 0)$.

To solve the resulting difference equations one can make use of existing techniques. Such equations were analyzed and successfully solved in the course of the evaluation of the three-loop anomalous dimensions in QCD [37,38] and of the two-[5] and three-loop [6] coefficient functions in DIS. In most cases of physical interest the resulting difference equations can be solved after expansion in $\epsilon$ in terms of harmonic sums or their generalizations. In simpler cases, one can even solve these equations in closed form in terms of hypergeometric and/or $\Gamma$-functions.
Difference equations were used by Laporta [52] for the high-precision numerical evaluation of master integrals. Applications of this idea include [56–58] and [59].

Upon solving the system of difference equations for the master integrals, one has achieved a complete extraction of the dependence of the masters on the Mellin variable \( n \). The only remaining thing to do is to specify the initial conditions for the solutions of the difference equations. Typically, that would be the value of the masters for \( n = 0 \). This is an important fact. It implies that to completely specify the master, one need to only evaluate integrals that are fully integrated over the available phase-space. These are pure numbers that do not depend on the kinematical variable \( n \). On the conceptual level, this is placing the evaluation of certain not-completely inclusive observables one step closer to the very familiar fully-inclusive case where, thanks to the optical theorem, one can significantly simplify the calculations by not considering separately all possible physical cuts.

Often, one can reduce the number of fixed-\( n \) integrals that have to be evaluated by hand. This follows from the property of the fixed-\( n \) reduction that not all integrals corresponding to initial conditions for the \( n \)-dependent masters are actually independent. To explore this fact one has to perform a separate fixed-\( n \) IBP reduction where \( n = 0 \) is taken from the very beginning. We have observed in simple one- and two-loop reductions as well as in rather complicated three-loop cases that this procedure indeed generates additional relations between the integrals corresponding to the initial conditions of the master integrals. One might wonder about the cost of such an additional run. That, however, should be of no concern since the fixed-\( n \) reduction is much simpler and faster than the general-\( n \) run one has to be able to perform anyway. The computer load pays off with the elimination of many of the integrals that otherwise have to be computed by hand. A fixed-\( n \) reduction can also be used as a cross-check of the general-\( n \) calculation. This is similar to the use of Mincer [60] in the three-loop DIS calculations in [36–38].

Before closing this section we would like to compare the approach for solving the IBP’s in the presence of an abstract power with the one used by Laporta in [52] to derive difference equations for master integrals. There are two basic differences. The first one is in the underlying algorithm used for solving the IBP’s. In [52], in particular step 4 in Algorithm 2, a modified algorithm for dealing with the case of an abstract power compared to the fixed-power case is proposed. That modification reflects the structuring of the software used in solving the recurrence relations. In this Letter, however, we use an approach which does not require such modification. Instead, our approach is based on the exact mapping of the problem with an abstract power to a problem with all powers being fixed integers. The second difference is in the way the master integrals in the problem are identified. In [52] one first solves a fixed power reduction and identifies the masters of that system. As a next step, one modifies that system to construct difference equations for the masters of the fixed-power reduction. As we noted previously, however, there can be a difference in the number of master integrals for, in our parlance, the fixed-\( n \) and general-\( n \) reductions. Thus, in the approach described in this Letter we propose a systematic way of first identifying the complete set of masters for the general-\( n \) reduction and then mapping the whole \( n \)-dependent problem to this set of masters.

4. Partial fractioning

Consider a case where during the evaluation of the amplitudes one gets a propagator that is not linearly independent from the constraint \( f \) defining \( z \) (see Eq. (7)). Clearly, that can happen in many ways and such linear dependence might even involve a group of several propagators. However, to simplify our point as much as possible, we will only consider a simple situation. Consider the function:

\[
F(n) = \int \mathcal{D} \times (\cdots) \times \frac{1}{(1-f)^n},
\]

(13)

where \( \cdots \) stay for powers of other possible propagators. If we were to evaluate this integral in \( z \)-space we would first replace \( f \) everywhere with \( z \), as is implied by the factor \( \delta(z-f) \). Then the factor \( 1/(1-f) \) becomes just the number \( 1/(1-z) \) and drops out of the integral. However, when we work in Mellin space the constraint \( f = z \) cannot be used anymore. If \( n \) were some fixed integer, we could have applied partial fractioning \( n \)-times and split the linearly-dependent propagators \( 1/f \) and \( 1/(1-f) \). For symbolic \( n \), however, that cannot be done and one should again resort to solving difference equations. This can be done in the following way. The identity:

\[
\frac{1}{(1-f)^{n+1}} = \frac{1}{(1-f)^n} \frac{1}{f^{n+1}},
\]

immediately translates into a difference equation for the function appearing in Eq. (13): \( F(n+1) = F(n) + G(n) \). This is a simple difference equation with non-homogeneous part given by:

\[
G(n) = \int \mathcal{D} \times (\cdots) \times \frac{1}{f^{n+1}}. \quad (14)
\]

The integral (14) does not contain linearly-dependent propagators and can be evaluated by using the procedures described previously.

Another way of eliminating the linear dependence among the propagators is to expand the propagator \( 1/(1-f) \) in geometric series. That would completely eliminate this propagator and one would end up with a standard problem where the index \( n \) is replaced by \( n + k, k \geq 0 \) (here one can apply the usual reduction since the index \( n + k \) is as arbitrary as the index \( n \) is). Finally, one would have to sum up the resulting expression over the index \( k \).

Our experience shows that the combination of the above methods is sufficient to eliminate the appearance of linearly-dependent propagators in any situation.

5. Concluding remarks

The method presented in this Letter represents a conceptually new approach for the evaluation of differential distrib-
NLO coefficient functions in $e^+e^-$ method opens up new venues for the application of the tech-
and are publicly available. They are capable of effectively deal-
ing of the IBP identities only on well established, multipurpose
general, applicability and it relies for the formulation and solv-
[5,6,36–38,61]. Still our method has wider, in fact completely
used in the inclusive DIS calculations at two- and three-loops.
the University of Hawaii. It was also supported in part by the
Award DE-FG03-94ER-40833 and by the start up funds of
DE-FG03-94ER-40833, the Outstanding Junior Investigator
I would like to thank Sven Moch for many useful con-
versations and the careful reading of the manuscript. I would
also like to thank Lance Dixon for related collaboration
and Kirill Melnikov for insightful discussions. The research
of A.M. is supported in part by the DOE under contract
DE-FG03-94ER-40833, the Outstanding Junior Investigator
Award DE-FG03-94ER-40833 and by the start up funds of
the University of Hawaii. It was also supported in part by the
Deutsche Forschungsgemeinschaft in Sonderforschungsbereich/Transregio 9.

Appendix A. Simple example

We present one example which is simple yet it demonstrates all non-trivial features of the method discussed above. We con-
sider the evaluation of the coefficient function in the decay of a
colorless object to a quark–antiquark pair $V \rightarrow q + X$ at order $a_S$. One constructs $|M|^2$ in the usual way; the relevant diagrams are shown on Fig. 1. At leading order, the differential observ-
able of interest is:

$$\sigma(z) = \frac{1}{\sigma(z)} \frac{d\sigma}{dz} = \delta(1 - z) + \mathcal{O}(a_S). \quad (A.1)$$

Note that we have chosen normalization where the coefficient of
the $\delta$-function in the leading term is exactly one to all orders
in $\epsilon, d = 4 - 2\epsilon$. In Mellin space, this corresponds to $\sigma(n) = 1$. The virtual corrections at order $a_S$ produce the same type of contributions.

Through order $a_S$, all non-trivial $n$-dependence of the distri-
bution Eq. (A.1) originates from the real gluon radiation dia-
grams. To be specific, we are interested in observing the final
state massless quark in the reaction $V(p) \rightarrow q(p - q - k) + X$, where
the unobserved massless antiquark and gluon carry momen-
ta $q$ and $k$ respectively. We take $p^2 = 1$. From the inde-
pendent momenta $p, q$ and $k$ one can construct five linearly
independent scalars that are needed to build the IBP reduc-
tion. As such we choose: $P_1 = (p - q)^2, P_2 = q^2, P_3 = k^2, P_4 = (p - q - k)^2, P_5 = 2 - 2p \cdot q - 2p \cdot k$.

After we perform the mapping (4) we have to deal with monomials of the type:

$$1 \frac{P_1^{\alpha_1}}{\prod_4^{\alpha_4}} \frac{P_2^{\alpha_2}}{\prod_6^{\alpha_6}} \frac{P_3^{\alpha_3}}{\prod_5^{\alpha_5}} \frac{P_4^{\alpha_4}}{\prod_5^{\alpha_5}} \frac{P_5^{\alpha_5}}{\prod_5^{\alpha_5}}. \quad (A.2)$$

Performing the IBP reductions in the Laporta’s method im-
plemented in the program AIR [53], we find that there is a single
master integral:

$$B(0, 1, 1, 1, 0) = \int d^d q_1 d^d q_2 \delta(P_1)\delta(P_2)\delta(P_3) \frac{1}{P_5^n}. \quad (A.3)$$

By inspecting the results for the element $B(0, 1, 1, 1, -1)$
from the solved IBP reduction, we derive the following differ-
ce equation for the only master integral:

$$B(0, 1, 1, 1, -1) = \frac{2 + n - 2\epsilon}{3 + n - 3\epsilon} B(0, 1, 1, 1, 0).$$

It is trivial to solve this recurrence relation (we modify the not-
tation in an obvious way to make completely transparent the
$n$-dependence):

$$B(0, 1, 1, 1, 0)(n) = \frac{\Gamma(2 + n - 2\epsilon)\Gamma(3 - 3\epsilon)}{\Gamma(3 - 3\epsilon + n)\Gamma(2 - 2\epsilon)} B(0, 1, 1, 1, 0)(n = 0).$$

The initial condition $B(0, 1, 1, 1, 0)(n = 0)$ is defined through
Eq. (A.3) after setting $n = 0$ there. It is a trivial to compute
number.
Our work is not quite done with the solving of the IBP reductions and the evaluation of the master integral since there also appears the propagator \( P_{add} = (q + k)^2 \) which does not belong to the set \( P_1, \ldots, P_5 \). This propagator results from the right diagram on Fig. 1; it is not linearly independent from the set \( P_1, \ldots, P_5 \) but it can appears downstairs together with these propagators. To resolve the situation one has to resort to the partial fractioning technique discussed in Section 4.

Exploiting the constraints implied by the three \( \delta \)-functions (with arguments \( P_{2,3,4} \)), one can easily establish that \( P_{add} = 1 - P_5 \). In this case one can apply the geometric series trick discussed in the previous section to all terms where \( P_{add} \) is present downstairs:

\[
\frac{1}{P_{add}} P_5^n = \frac{1}{1 - P_5} P_5^n = \sum_{i=0}^{\infty} P_5^i.
\]

Next one takes the summation outside the integrals; the resulting integrand is of the type in Eq. (A.2). With the help of the IBP reduction this integral can be reduced to the master integral discussed above. Inserting the explicit form of the master, one obtains very simple expression containing only \( \Gamma \)-functions. The summation over \( s \) of this product of \( \Gamma \)-functions can be easily performed and it results again in a product of the same type of functions. Thus, the result from the real emission radiation at order \( \alpha_s \) can be easily evaluated in closed form to all orders in \( \epsilon \). To use this expression in practical applications one has to decompose it in series in \( \epsilon \). This expansion can be easily automated with the help of the programs Summer [63] and XSummer [64].

References

[32] A. Czarnecki, M. Jezabek, J.H. Kuhn, Acta Phys. Pol. B 20 (1989) 961. To use this expression in practical applications one has to decompose it in series in \( \epsilon \). This expansion can be easily automated with the help of the programs Summer [63] and XSummer [64].
Three loop $\overline{\text{MS}}$ transversity operator anomalous dimensions for fixed moment $n \leq 8$

J.A. Gracey

Theoretical Physics Division, Department of Mathematical Sciences, University of Liverpool, P.O. Box 147, Liverpool, L69 3BX, United Kingdom

Received 2 November 2006; accepted 7 November 2006

Available online 15 November 2006

Editor: N. Glover

Abstract

We compute the anomalous dimensions of the transversity operator at three loops in the $\overline{\text{MS}}$ scheme for fixed moment $n$ where $n \leq 8$. The results for the RI' renormalization scheme are also provided for an arbitrary linear covariant gauge for $n \leq 7$.

© 2006 Elsevier B.V. All rights reserved.

Recently, the three loop anomalous dimension for the unpolarized twist-2 flavour non-singlet and singlet Wilson operators have been determined analytically for all values of the operator moment $n$, [1–4]. This was a formidable undertaking spanning ten years and relied on cutting edge computer algebra and symbolic manipulation techniques implemented on high performance computers. With the obvious necessity of such results to ensure the complete two loop evolution of the Wilson coefficients and obtain more precise estimates of quantities such as $\alpha_s(M_Z)$, there is also the need to extend such computations to problems involving spin. Indeed in this area, one quantity which will be of interest is that relating to transversity which was originally introduced in [5–7]. This corresponds to the probability of finding a quark in a transversely polarized nucleon polarized parallel to the nucleon versus that of the nucleon in the antiparallel polarization. Although experimentally it is harder to extract information on the transversity compared with usual deep inelastic scattering, one will still require the anomalous dimensions of the underlying operator to as high a loop order as is calculationally feasible for accurate renormalization group evolution. Currently the two loop anomalous dimensions are available for arbitrary moment, $n$, in the $\overline{\text{MS}}$ scheme, [8–12]. However, given the symbolic manipulation machinery now available [1–4,13], it is clearly only a matter of time before the full $n$-dependent three loop $\overline{\text{MS}}$ results are determined. Although the underlying transversity operator differs from that of the Wilson operators of [1–4], calculationally it is on a par with the non-singlet work of [14]. Prior to the full $n$-dependent results of [1–4] for the twist-2 Wilson operators, one approach was to carry out a fixed moment determination of the anomalous dimensions. Essentially the first even moments to $n = 16$ (apart from $n = 14$) were determined, [14–16]. Although eventually superseded by the analytic result, given the huge number of Feynman diagrams to evaluate by tedious recurrence relations, having information on the final results from an independent calculation provided an important crosscheck on the full $n$-dependent expressions. This is particularly the case when computations of gauge invariant quantities are simplified by choosing to work in the Feynman gauge. This substantially reduces the large number of integration by parts and hence the computation time is significantly smaller than, say, in an arbitrary linear covariant gauge. In other words working in a specific gauge means that the internal strong check of observing the gauge parameter cancellation for a gauge independent quantity is crucially absent.

Given the potential determination of analytic anomalous dimensions for the transversity operator in the foreseeable future, it is the purpose of this article to repeat the approach of [14] and provide the $\overline{\text{MS}}$ anomalous dimensions for fixed moments at three loops up to and including moment $n = 8$. This builds on the low moment results of [17,18] where the anomalous dimensions of the
tensor current and second moment were determined. More recently the results for moments \( n = 3 \) and \( 4 \) were provided in [19]. In [18,19] the primary aim was to provide the finite parts of a specific Green’s function in order to aid lattice measurements of the same quantity. In particular those measurements had to match onto the ultraviolet part of the Green’s function and the provision of the answer at three loops was necessary to help make the extraction of lattice results as precise as possible. However, the work also required performing the renormalization in the lattice renormalization scheme, known as the modified regularization invariant (RI’) scheme, [20,21]. It has a continuum definition which is discussed at length in [17,22]. As the anomalous dimension of a gauge invariant operator is gauge dependent in a mass dependent scheme, the computation had to be performed in an arbitrary linear covariant gauge. Though for practical reasons the lattice calculations were performed in the Landau gauge. Therefore, as a second thread to this article, we will also determine the transversity anomalous dimensions in the RI’ scheme in an arbitrary linear covariant gauge. Though for computational reasons this will be restricted to \( n \leq 7 \).

First, we discuss the basic properties of the transversity operator and outline our computational strategy. The operator is defined by, [8–10],

\[
\mathcal{O}^{\mu_{v_1} \cdots \mu_{v_n}} = S \bar{\psi} \sigma^{\mu_{v_1}} D^{\nu_2} \cdots D^{\nu_n} \psi,
\]

(1)

where \( \sigma^{\mu \nu} = \frac{1}{2} \{\gamma^\mu, \gamma^\nu\} \) and \( D_\mu \) is the covariant derivative involving the coupling constant \( g \). The operation \( S \) denotes symmetrization of the Lorentz indices \( \{v_1, \ldots, v_n\} \) as well as ensuring that the operator is traceless according to the rules

\[
\eta_{\mu v_i} \mathcal{O}^{\mu_{v_1} \cdots \nu_{v_i} \cdots \nu_{v_j} \cdots \nu_{v_n}} = 0 \quad (i \geq 2), \quad \eta_{v_i v_j} \mathcal{O}^{\mu_{v_1} \cdots \nu_{v_i} \cdots \nu_{v_j} \cdots \nu_{v_n}} = 0.
\]

(2)

To renormalize (1) we follow a procedure similar to [17,18] where the operator is inserted at zero momentum into a quark 2-point function, \( \langle \psi(p) \mathcal{O}(0) \bar{\psi}(-p) \rangle \), where \( p \) is the momentum. In order to apply the MINCER algorithm, [23], written in the symbolic manipulation language FORM, [24,25], one needs to saturate the Lorentz indices with the appropriate tensor. This is because the MINCER formalism can only be applied to massless three loop Lorentz scalar 2-point Feynman integrals, [23]. Here, since we are not interested in the finite part of this Green’s function, we merely multiply it by a Lorentz tensor which has the same symmetry and tracelessness properties as the original operator itself. This leads to an immediate algebraic simplification. When the fully symmetrized and traceless operator is inserted into the Green’s function, there is a part involving products of the tensor \( \eta^{\mu \nu} \). These derive from ensuring the tensor is overall traceless. However, when these terms multiply the projection tensor such terms will give zero. Therefore, in constructing the Feynman rules for the operator insertion for the current calculation, one needs only to consider the part of the operator which is independent of the \( \eta^{\mu \nu} \) tensors. In other words, the object \( \bar{\psi} \sigma^{\mu(\nu} D^{\nu_2} \cdots D^{\nu_n}) \psi \). For high moments, this represents a huge reduction in work such as the tedious but automatic derivation of the full Feynman rules, which can involve a significantly large number of terms. Instead the main work is in the construction of the projection tensor. However, this is achieved computer algebraically by writing down the complete set of independent objects built from one \( \sigma^{\mu \nu} \) tensor, together with the appropriate numbers of \( \eta^{\mu \nu} \) tensors and momenta \( \rho^\sigma \) such that the number of free Lorentz indices equates with that of the original operator. These independent tensors are then symmetrized automatically with respect to the indices \( \{v_1, \ldots, v_n\} \) and the arbitrary coefficients chosen so that the overall projection tensor is traceless according to (2). As was indicated in [18,19], there are three independent projections and therefore for the determination of the anomalous dimensions one needs only to select one of these for the projection procedure. It is preferable to choose the most algebraically compact one to minimize computation time. We note that this procedure is in contrast to the fixed moment strategy of [14] where the null vector \( \Delta_\mu \) was introduced. However, as was discussed in [14,26], one needs to handle the resulting integrals using a different projection technique. Finally, we note that our procedure automatically ensures that there is no operator mixing. [19]. The Feynman diagrams required for \( \langle \psi(p) \mathcal{O}(0) \bar{\psi}(-p) \rangle \) are generated automatically with the QGRAF package, [27], and converted to FORM input notation prior to the application of the three loop MINCER algorithm. There are 3 one loop, 37 two loop and 684 three loop Feynman diagrams and throughout we have used dimensional regularization in \( d = 4 - 2\epsilon \) dimensions.

We now record that the three loop \( \overline{\text{MS}} \) transversity anomalous dimensions, \( \gamma^{(n)}(a) \), for \( n = 5, 6 \) and 7 are

\[
\gamma^{(5)}(a) = \frac{92}{15} C_F a + \left[ 189515 C_A - 41674 C_F - 79810 T_F N_f \right] \frac{C_F a^2}{6750} \\
+ \left\{ (1192320000 \zeta(3) + 989903260) C_F T_F N_f - 83718800 T_F^2 N_f^2 \right\} \frac{C_F a^3}{12150000} + O(a^4),
\]

(3)
\[ \gamma^{(6)}(a) = \frac{34}{5} C_F a + [204770C_A - 42129C_F - 88810T_F N_f] \frac{C_F a^2}{6750} + \left[ \left( 707616000C_A - 312284800\zeta(3) + 1373507730 \right) C_A C_F \right. \\
\left. - (4626720000\zeta(3) + 1841332000) C_A T_F N_f + (1415232000\zeta(3) - 684744816) C_F^2 \right. \\
\left. + (4626720000\zeta(3) - 3910683210) C_F T_F N_f - 320975800T_F^2 N_f^2 \right] \frac{C_F a^3}{42525000} + O(a^4) \] (4)

and

\[ \gamma^{(7)}(a) = \frac{258}{35} C_F a + [75266555C_A - 15484767C_F - 33149830T_F N_f] \frac{C_F a^2}{2315250} + \left[ \left( 3517994592000\zeta(3) + 38365845513450 \right) C_A^2 \right. \\
\left. - (24084527040000\zeta(3) + 9039144860900) C_A T_F N_f + (7035989184000\zeta(3) - 4192441946262) C_F^2 \right. \\
\left. + (24084527040000\zeta(3) - 20698675427220) C_F T_F N_f - 1651311191600T_F^2 N_f^2 \right] \frac{C_F a^3}{204205050000} + O(a^4), \] (5)

where \( a = g^2/(16\pi^2) \), \( \zeta(n) \) is the Riemann zeta function, \( N_f \) is the number of quark flavours and the colour group Casimirs are given by

\[ \text{Tr}(T^a T^b) = T_F \delta^{ab}, \quad T^a T^a = C_F I, \quad f^{abc} f^{bcd} = C_A \delta^{ab} \] (6)

and \( T^a \) are the generators for the colour group with structure constants \( f^{abc} \). We have also calculated the \( n = 8 \) moment but modified the method used to find (3), (4) and (5). In order to reduce the computation time we performed the calculation in the Feynman gauge. Whilst this significantly reduces the number of integration by parts needed to be carried out by the MINCER algorithm, we have performed the two loop calculation in an arbitrary linear covariant gauge and reconstructed the gauge independent result of [8–12]. Thus, at three loops in the \( \overline{\text{MS}} \) scheme we have

\[ \gamma^{(8)}(a) = \frac{551}{70} C_F a + [1270588235C_A - 251839827C_F - 568470280T_F N_f] \frac{C_F a^2}{37044000} + \left[ \left( 57548150352000\zeta(3) + 651153115163775 \right) C_A^2 \right. \\
\left. - (41149067904000\zeta(3) + 148380288276500) C_A T_F N_f + (115096300704000\zeta(3) - 55777651074312) C_F^2 \right. \\
\left. + (41149067904000\zeta(3) - 356756758487220) C_F T_F N_f - 27902636165600T_F^2 N_f^2 \right] \frac{C_F a^3}{326728080000} + O(a^4), \] (7)

As further checks on our results, (3), (4), (5) and (7), we note that for \( \overline{\text{MS}} \) all the two loop expressions agree with those of [8–12], when they are evaluated for specific \( n \). Moreover, (3), (4) and (5) are clearly independent of the linear covariant gauge fixing parameter, \( \alpha \). Further, the coefficients of the leading order large \( N_f \) term at three loops, corresponding to the \( C_F T_F^2 N_f^2 \) term, agrees with the analytic evaluation at \( O(1/N_f) \) for arbitrary \( n \) given in [18]. A final check is provided by the fact that the triple and double poles in \( \epsilon \) of the resulting operator renormalization constant are reproduced precisely in agreement with the renormalization group equation prediction for all four cases. This is important since we followed the procedure of [28] for renormalizing operators in automatic symbolic manipulation programmes. For the specific case of the Lie group \( SU(3) \) we have

\[ \gamma^{(5)}(a) \bigg|_{SU(3)} = \frac{368}{45} a^2 - 2[119715N_f - 1538939] \frac{a^2}{30375} \right. \\
\left. - [188367300N_f^2 + 8942400000\zeta(3)N_f + 12843253410N_f^2 \right. \\
\left. - 954180000\zeta(3) - 144207743479] \frac{a^3}{82012500} + O(a^4), \] (8)

\[ \gamma^{(6)}(a) \bigg|_{SU(3)} = \frac{136}{15} a^2 - 2[44405N_f - 558138] \frac{a^2}{10125} \right. \\
\left. - [240731850N_f^2 + 11566800000\zeta(3)N_f + 16107360420N_f^2 \right. \\
\left. - 1179360000\zeta(3) - 183119500163] \frac{a^3}{95681250} + O(a^4), \] (9)
To summarize, this renormalization scheme is defined by ensuring that the finite part of the renormalized Green’s function \( \langle \psi(p)O(0)\bar{\psi}(-p) \rangle \), multiplied by the projector, is given purely by its tree value only, [17–21]. Further, it is important to note
that in (12), (13) and (14) the variables $a$ and $\alpha$ are to be regarded as RI\textsuperscript{prime} quantities. The relation to the $\overline{\text{MS}}$ variables are given to three loops in [17] for an arbitrary linear covariant gauge. Since our main motivation is to provide the $\overline{\text{MS}}$ anomalous dimensions, the $n = 8$ moment in the RI\textsuperscript{prime} scheme is clearly not available since the corresponding $\overline{\text{MS}}$ computation was restricted to the Feynman gauge.

We conclude by noting that we have provided higher moments of the transversity operator at three loops in both the $\overline{\text{MS}}$ and RI\textsuperscript{prime} schemes. Together with the earlier results of [18–21], there are now eight fixed moment $\overline{\text{MS}}$ anomalous dimensions available prior to an explicit $n$-dependent computation. This is similar to the situation with regard to the flavour non-singlet twist-2 operator where there were seven fixed moment anomalous dimensions available prior to the provision of the full $n$-dependent expression. Whilst it is still in principle possible to compute even higher moments using the method we have discussed here, we believe we have reached a computational limit beyond which it is not viable to proceed. For instance, for the three loop $n = 8$ Feynman gauge calculation, the necessary Feynman rules took around 36 hours to be generated electronically on a dual opteron 64 bit SMP machine (2 GHz), resulting in just under $5.5 \times 10^6$ terms. The former is an order of magnitude larger than the time required for the $n = 7$ case.

Acknowledgements

The author thanks Dr. P.E.L. Rakow and Dr. C. McNeile for valuable discussions.

References

Regulator dependence of the proposed UV completion of the ghost condensate

Donal O’Connell

Department of Physics, California Institute of Technology, Pasadena, CA 91125, USA

Received 20 October 2006; accepted 1 November 2006

Available online 10 November 2006

Editor: B. Grinstein

Abstract

Recently, it was shown that a renormalizable theory of heavy fermions coupled to a light complex boson could generate an effective action for the boson with the properties required to violate Lorentz invariance spontaneously through the mechanism of ghost condensation. However, there was some doubt about whether this result depended on the choice of regulator. In this work, we adopt a non-perturbative, unitary lattice regulator and show that with this regulator the theory does not have the properties necessary to form a ghost condensate. Consequently, the statement that the theory is a UV completion of the Higgs phase of gravity is regulator dependent.

© 2006 Elsevier B.V. All rights reserved.

1. Introduction

The ghost condensate proposal of [1] has received considerable attention recently [2–12]. The condensate is a mechanism for modifying gravity in the infrared. The starting point of the model is a scalar field, $\phi$, with a shift symmetry

$$\phi \rightarrow \phi + \alpha$$

such that the effective action for the scalar is of the form $\mathcal{L} = P(X)$, where $X = \partial_\mu \phi \partial^\mu \phi$. (We ignore terms such as $(\partial^2 \phi)^2$ as they will not be important in our discussion.) Moreover, we assume that $\phi$ is a ghost, so that $P(X)$ is of the form shown in Fig. 1. The origin, $\phi = 0$, is an unstable field configuration in this scenario. The ghost then condenses so that $(\partial \phi)^2$ has a value near the minimum of $P$. It is also possible that there is no ghost at the origin but a non-trivial minimum elsewhere, as shown in Fig. 2; in such a theory there would still be a ghost condensate near the minimum of $P$. This class of theories is of considerable phenomenological interest because a ghost condensate has equation of state $w = -1$ and could therefore be relevant for explaining the observed small but non-zero cosmological constant [1].

It is also of interest, however, to understand how the effective action $\mathcal{L} = P(X)$ could arise as a low energy effective theory of some more familiar UV quantum field theory [9]. Since the scalar field must have a shift symmetry, it is natural to seek a completion in which $\phi$ is the Goldstone boson of a spontaneously broken $U(1)$ symmetry. It was shown in [2] that it is impossible, classically, to generate a ghostly low energy effective action for such a Goldstone boson from a high energy theory with standard kinetic terms. However, the authors went on to find a theory in which a quantum correction could change the sign of the kinetic term of the Goldstone boson. In that proposal, all fields start out with standard kinetic terms. However, interactions between $\phi$ and certain heavy fermions correct the kinetic term of $\phi$. It was found that under certain assumptions, these corrections could produce an effective Lagrangian for $\phi$ of the form shown in Fig. 1 at scales much smaller than the fermion mass $m$. We do not expect to find an effective Lagrangian of the form shown in Fig. 2 because the higher order terms in the expansion of $P(X)$ are suppressed by powers of the cutoff.

The model described in [2] has some shortcomings. The high energy theory has a Landau pole. Moreover, in dimensional regularization it was found that to change the sign of the bosonic kinetic term, the mass of the fermions has to be close to the Landau pole. This circumstance may cause some concern that
the calculation could be regulator dependent. To alleviate these concerns, the authors demonstrated that their conclusion holds in a large class of momentum-dependent regulators, provided that the fermion masses were taken to be of order of the regulator. These regulators, however, violate unitarity, so again it is not clear to what extent the sign of the kinetic term is a well-defined quantity.

In this Letter, we re-examine the theory presented in [2] using a lattice regulator. This regulator is non-perturbatively valid and preserves unitarity. We will see that there is never a ghost when the theory is regulated in this way. As a consequence, it seems that the conclusions of [2] are regulator dependent.

2. Computation

We begin by describing the theory we will be working with in more detail. The candidate ghost field, \( \phi \), must have a shift symmetry so it is natural to suppose that it is a Goldstone boson associated with the breaking of some \( U(1) \) symmetry. Hence, following [2], we choose as the bosonic part of the Lagrangian the usual spontaneous symmetry breaking Lagrangian for a complex scalar field \( \Phi \),

\[
\mathcal{L}_b = \partial_\mu \Phi^* \partial^\mu \Phi - \frac{\lambda}{4} (|\Phi|^2 - v^2)^2. \tag{2}
\]

The Goldstone boson, \( \phi \), associated with the spontaneous symmetry breaking is the candidate ghost field. We couple \( \Phi \) to two families of fermions \( \psi_i \), \( i = 1, 2 \) of charges +1 and −1, respectively. We will assume that there are \( N \) identical fermions in each family, and that each fermion has the same mass \( m \). The fermions are coupled to \( \Phi \) by a Yukawa term with coupling \( g \).

Hence, the total Lagrangian density is

\[
\mathcal{L} = \mathcal{L}_b + \sum_{j=1}^N \sum_{i=1,2} \left[ (i \bar{\psi}_i^{(j)} \gamma^\mu \partial_\mu \psi_i^{(j)} - m \bar{\psi}_i^{(j)} \psi_i^{(j)}) 
- g \Phi \bar{\psi}_2^{(j)} \psi_1^{(j)} - g \Phi^* \bar{\psi}_1^{(j)} \psi_2^{(j)} \right]. \tag{3}
\]

The low energy effective action for \( \Phi \) is obtained by integrating the fermions out. The effective action can be written

\[
\mathcal{L}_{\text{eff}} = \Phi^* G(\bar{\Phi}^2) \Phi - V(|\Phi|), \tag{4}
\]

where

\[
G(p^2) = p^2 + g^2 N f(p^2). \tag{5}
\]

The function \( f(p^2) \) describes the effects of the quantum corrections to the bosonic kinetic term. If \( G(p^2) < 0 \) for some range of \( p^2 \), then the theory can have a ghost. This can only happen if \( g^2 N f(p^2) \) is negative and larger than the tree level term \( p^2 \).

Since this signals a breakdown in perturbation theory, we work in the large \( N \) limit with \( g^2 N \) fixed to maintain control over the calculation.

Let us now move on to compute \( f(p^2) \). To do so, we must evaluate the Feynman graph shown in Fig. 3. After Wick rotating both momenta into Euclidean space, we find

\[
f(p^2) = -4 \int \frac{d^4 k}{(2\pi)^4} \frac{m^2 - k \cdot (p + k)}{(k^2 + m^2)((p + k)^2 + m^2)}. \tag{6}
\]

This expression is divergent and requires regulation. We choose a lattice regulator with lattice spacing \( a \). Since we are working in the large \( N \) limit, the phenomenon of fermion doubling [13] will not pose a problem. Therefore, we will use naive lattice fermions. The (Euclidean) fermion propagator is given by [13]

\[
G(p) = a \sum_{\mu} \frac{-i \gamma^\mu (\rho_\mu a) + ma}{\sum_{\mu} \sin^2 p_\mu a + m^2 a^2}, \tag{7}
\]

where \( \gamma_\mu \) are Euclidean gamma matrices. On this lattice, momentum components lie in the first Brillouin zone, so \( -\pi < p_\mu a < \pi \). The regulated Feynman graph (Fig. 3) is

\[
f(p^2) = -4 \int_B \frac{d^4 k}{(2\pi)^4} a^2 \times \left[ m^2 a^2 - \sum_{\mu} \sin ak_\mu \cdot \sin a(p_\mu + k_\mu) \right]
\]

where

\[
|\Phi|^2 = \sum_{i=1}^N \sum_{j=1,2} \left| \bar{\psi}_i^{(j)} \psi_i^{(j)} \right|^2.
\]

The low energy effective action for \( \Phi \) is obtained by integrating the fermions out. The effective action can be written

\[
\mathcal{L}_{\text{eff}} = \Phi^* G(\bar{\Phi}^2) \Phi - V(|\Phi|), \tag{4}
\]

where

\[
G(p^2) = p^2 + g^2 N f(p^2). \tag{5}
\]

The function \( f(p^2) \) describes the effects of the quantum corrections to the bosonic kinetic term. If \( G(p^2) < 0 \) for some range of \( p^2 \), then the theory can have a ghost. This can only happen if \( g^2 N f(p^2) \) is negative and larger than the tree level term \( p^2 \).

Since this signals a breakdown in perturbation theory, we work in the large \( N \) limit with \( g^2 N \) fixed to maintain control over the calculation.

Let us now move on to compute \( f(p^2) \). To do so, we must evaluate the Feynman graph shown in Fig. 3. After Wick rotating both momenta into Euclidean space, we find

\[
f(p^2) = -4 \int \frac{d^4 k}{(2\pi)^4} \frac{m^2 - k \cdot (p + k)}{(k^2 + m^2)((p + k)^2 + m^2)}. \tag{6}
\]

This expression is divergent and requires regulation. We choose a lattice regulator with lattice spacing \( a \). Since we are working in the large \( N \) limit, the phenomenon of fermion doubling [13] will not pose a problem. Therefore, we will use naive lattice fermions. The (Euclidean) fermion propagator is given by [13]

\[
G(p) = a \sum_{\mu} \frac{-i \gamma^\mu (\rho_\mu a) + ma}{\sum_{\mu} \sin^2 p_\mu a + m^2 a^2}, \tag{7}
\]

where \( \gamma_\mu \) are Euclidean gamma matrices. On this lattice, momentum components lie in the first Brillouin zone, so \( -\pi < p_\mu a < \pi \). The regulated Feynman graph (Fig. 3) is

\[
f(p^2) = -4 \int_B \frac{d^4 k}{(2\pi)^4} a^2 \times \left[ m^2 a^2 - \sum_{\mu} \sin ak_\mu \cdot \sin a(p_\mu + k_\mu) \right]
\]

where

\[
|\Phi|^2 = \sum_{i=1}^N \sum_{j=1,2} \left| \bar{\psi}_i^{(j)} \psi_i^{(j)} \right|^2.
\]
\[
\times \left[ \sum_\nu \sin^2 \alpha k_\nu + m^2 a^2 \right]^{-1} \\
\times \left[ \sum_\rho \sin^2 \alpha (p_\rho + k_\rho) + m^2 a^2 \right]^{-1},
\]

where the integral is over the Brillouin zone \( B \). Note that as \( a \to 0 \), the regulated expression Eq. (8) reduces to the continuum expression Eq. (6).

In [2], there was a ghost at the origin. Since our goal is to check for potential regulator dependence of this statement, it suffices to extract the order \( p^2 \) part of \( f(p^2) \). Thus, we expand Eq. (8) in \( p_\mu \) and extract the second order term. We find

\[
f(p^2) \simeq -4a^2 \sum_\mu p_\mu p_\mu a^2 \int_B \frac{d^4k}{(2\pi)^4} \left( m^2 a^2 + \sum_\nu \sin^2 k_\nu a \right)^2 \\
\times \left[ \frac{m^2 a^2 - \sum_\nu \sin^2 k_\nu a}{m^2 a^2 + \sum_\nu \sin^2 k_\nu a} \cos 2k_\mu a \right] \\
+ \frac{1}{2} \sin^2 k_\mu a + \frac{1}{2} \frac{\sin^2 2k_\mu a}{m^2 a^2 + \sum_\nu \sin^2 k_\nu a} \\
+ \frac{m^2 a^2 - \sum_\nu \sin^2 k_\nu a}{(m^2 a^2 + \sum_\nu \sin^2 k_\nu a)^2} \sin^2 2k_\mu a.
\]

Now, in [2], the sign of the kinetic term was altered if the fermion masses large compared to the cutoff, analytic results were obtained demonstrating the presence of a ghost. In our case, we can obtain an analytic result when \( ma \gg 1 \). In this limit, the coefficient of \( p^2 \) induced by the quantum correction is given by

\[
f(p^2) = -4a^2 \left( \frac{1}{m^2 a^2} \right)^2 \int_B \frac{d^4k}{(2\pi)^4} \\
\times \sum_\mu \left[ \frac{1}{2} p_\mu p_\mu a^2 \sin^2 k_\mu a - p_\mu p_\mu a^2 \cos 2k_\mu a \right] \\
= -4a^2 \left( \frac{1}{m^2 a^2} \right)^2 \frac{p^2}{4a^2}.
\]

Rotating back into Euclidean space, we find

\[
G(p^2) \simeq p^2 + g^2 N \left( \frac{1}{ma} \right)^4 p^2.
\]

Clearly, this quantity never becomes negative, so the sign of the kinetic term does not change in this theory, at least when \( ma \gg 1 \).

To check for a sign change away from this limit, we have numerically integrated Eq. (9) to find the coefficient of \( p^2 \) induced by quantum corrections, as a function of \( x = 1/(ma) \). The result is shown in Fig. 4 for \( 0 \leq x \leq 2 \). Evidently, \( f(p^2)/p^2 \) is never negative, so there can be no change in the sign. For large \( x \), the fermion mass is much smaller than the cutoff so we need not worry about regulator dependence; therefore, we know from the results of [2] that there is no ghost in the region \( x > 2 \). This completes our demonstration that the sign of the kinetic term is always positive if the theory is regulated on a spacetime lattice.

3. Conclusions

We have examined the proposed high energy completion of the ghost condensate [2]. Using a lattice regulator, which is valid without invoking perturbation theory, and which is unitary, we have shown that this theory does not have a ghostly low energy effective action. The effect noted in [2], which involved changing the sign of the kinetic term for a scalar \( \phi \), appears to be a regulator dependent phenomenon. Thus, the search for a UV completion for the ghost condensate must continue.

Acknowledgements

We would like to thank Mark Wise for sharing his insights. We also thank Michael Graesser and Alejandro Jenkins for helpful conversations and for proof-reading the manuscript. Finally, we thank Árdís Elíasdóttir and André Walker-Loud for help with the manuscript.

References

Corrigendum

Corrigendum to: “Detection of SUSY in the stau-neutralino coannihilation region at the LHC”

Richard Arnowitt a, Bhaskar Dutta a,*, Teruki Kamon a, Nikolay Kolev b, David Toback a

a Department of Physics, Texas A&M University, College Station, TX 77843-4242, USA
b Department of Physics, University of Regina, Regina, SK S4S 0A2, Canada

Received 25 July 2006
Available online 2 August 2006

The correct affiliation of David Toback is Department of Physics, Texas A&M University, College Station, TX 77843-4242, USA.

DOI of original article: 10.1016/j.physletb.2006.05.090.
* Corresponding author.
E-mail address: dutta@physics.tamu.edu (B. Dutta).
Erratum

Erratum to: “Schwinger model in noncommutating space–time”

Anirban Saha\textsuperscript{a}, Anisur Rahaman\textsuperscript{b}, Pradip Mukherjee\textsuperscript{a,*}\textsuperscript{1}

\textsuperscript{a} Department of Physics, Presidency College, 86/1 College Street, Kolkata 700073, West Bengal, India
\textsuperscript{b} Department of Physics, Durgapur Government College, Durgapur 713214, Burdwan, West Bengal, India

Received 2 November 2006
Available online 13 November 2006

Abstract

A calculational error was crept in the expression of the Hamiltonian (Eq. (14)) in original article. We give the correct form of the Hamiltonian and also the correct form of the subsequent equations which depend on the Hamiltonian. The corrections are all of numerical nature and do not affect the physical conclusions in any fundamental way.

© 2006 Elsevier B.V. All rights reserved.

PACS: 11.15.-q; 11.10.Nx

The expression of the Hamiltonian (Eq. (14)) in [1] contains calculational error. The correct form will be

\[ H_{\text{CEV}} = \frac{\theta}{2} \left\{ \pi^2 \left( \phi'^2 - \phi''^2 \right) + e\phi \left( \phi'^2 - \phi''^2 \right) + e^2 \phi^3 - \left( \pi^1 \right)^3 + \frac{3}{2} e\phi \pi^1 \left( \pi^1 - e\phi \right) \right\} \]  \tag{1}

The error in the Hamiltonian percolates to the subsequent equations (18), (19), (20), (21) and (25) in the Letter the correct forms of which can be derived using the corrected Hamiltonian (1). Thus the corrected form of the reduced Hamiltonian is

\[ H_R = \left[ \frac{1}{2} \left( \pi^2 \phi + \phi'^2 + \phi''^2 \right) + \frac{e\theta}{2} \phi \left( \phi'^2 - \phi''^2 + e^2 \phi^2 \right) \right] \tag{2} \]

and the equations of motion for \( \phi \) and \( \pi_\phi \) are

\[ \dot{\phi} = (1 + e\theta \phi) \pi_\phi, \tag{3} \]

\[ \dot{\pi}_\phi = \phi'' - e^2 \phi - \frac{e\theta}{2} \left( \pi^2 \phi + \phi'^2 + 2\phi \phi'' + 3e^2 \phi^2 \right), \tag{4} \]

respectively. These two equations combine to give the second order equation that contains the physical contents of the theory

\[ (\Box + e^2)\phi = \frac{e\theta}{2} \left( \phi'^2 - \phi''^2 - 5e^2 \phi^2 \right). \tag{5} \]
The expression for the source of the interacting background of noncommutative origin is

\[
\mathcal{J}(x) = \frac{e\theta}{2} \int \frac{d\bar{p} \, d\bar{q}}{(8\pi^2)} e^{i(\bar{p} + \bar{q})\bar{x}} \left[ \left( -p_0 q_0 + \bar{p} \bar{q} - 5e^2 \right) \left\{ a(\bar{p}) a(\bar{q}) e^{-i(p^0 + q^0)x^0} + a^\dagger(-\bar{p}) a^\dagger(-\bar{q}) e^{i(p^0 + q^0)x^0} \right\} 
+ \left( p_0 q_0 - \bar{p} \bar{q} + 5e^2 \right) \left\{ a(\bar{p}) a^\dagger(-\bar{q}) e^{-i(p^0 - q^0)x^0} + a^\dagger(-\bar{p}) a(\bar{q}) e^{i(p^0 - q^0)x^0} \right\} \right].
\]

The equations (2), (3), (4), (5) and (6) are the corrected forms of the equations (18), (19), (20), (21) and (25) in [1], respectively. It is easy to appreciate that the corrections do not affect the conclusions of [1] in any essential way.

References