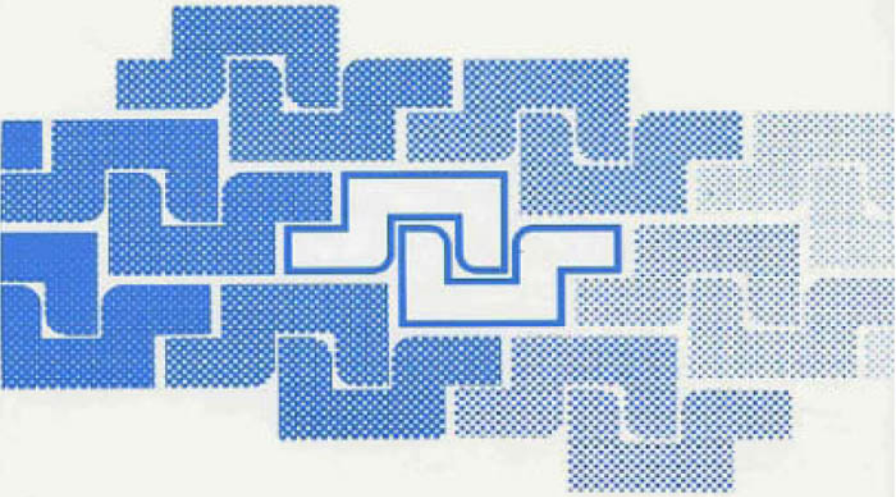


New Developments in Quantum Field Theory

Edited by
Poul Henrik Damgaard and
Jerzy Jurkiewicz

NATO ASI Series

Series B: Physics Vol. 366

New Developments in Quantum Field Theory

NATO ASI Series

Advanced Science Institutes Series

A series presenting the results of activities sponsored by the NATO Science Committee, which aims at the dissemination of advanced scientific and technological knowledge, with a view to strengthening links between scientific communities.

The series is published by an international board of publishers in conjunction with the NATO Scientific Affairs Division

A Life Sciences **B Physics**	Plenum Publishing Corporation New York and London
C Mathematical and Physical Sciences **D Behavioral and Social Sciences** **E Applied Sciences**	Kluwer Academic Publishers Dordrecht, Boston, and London
F Computer and Systems Sciences **G Ecological Sciences** **H Cell Biology** **I Global Environmental Change**	Springer-Verlag Berlin, Heidelberg, New York, London, Paris, Tokyo, Hong Kong, and Barcelona

PARTNERSHIP SUB-SERIES

1. Disarmament Technologies	Kluwer Academic Publishers
2. Environment	Springer-Verlag
3. High Technology	Kluwer Academic Publishers
4. Science and Technology Policy	Kluwer Academic Publishers
5. Computer Networking	Kluwer Academic Publishers

The Partnership Sub-Series incorporates activities undertaken in collaboration with NATO's Cooperation Partners, the countries of the CIS and Central and Eastern Europe, in Priority Areas of concern to those countries.

Recent Volumes in this Series:

Volume 364 — Quantum Fields and Quantum Space Time
edited by Gerard 't Hooft, Arthur Jaffe, Gerhard Mack, Pronob K. Mitter, and Raymond Stora

Volume 365 — Techniques and Concepts of High-Energy Physics IX
edited by Thomas Ferbel

Volume 366 — New Developments in Quantum Field Theory
edited by Poul Henrik Damgaard and Jerzy Jurkiewicz

Series B: Physics

New Developments in Quantum Field Theory

Edited by

Poul Henrik Damgaard

Niels Bohr Institute
Copenhagen, Denmark

and

Jerzy Jurkiewicz

Jagellonian University
Cracow, Poland

KLUWER ACADEMIC PUBLISHERS
NEW YORK, BOSTON, DORDRECHT, LONDON, MOSCOW

eBook ISBN 0-306-47075-6
Print ISBN 0-306-45816-0

Created in the United States of America

Visit Kluwer Online at: http://www.kluweronline.com
and Kluwer's eBookstore at: http://www.ebooks.kluweronline.com

PREFACE

Quantum field theory is one of most central constructions in 20th century theoretical physics, and it continues to develop rapidly in many different directions. The aim of the workshop "New Developments in Quantum Field Theory", which was held in Zakopane, Poland, June 14-20, 1997, was to capture a broad selection of the most recent advances in this field. The conference was sponsored by the Scientific and Environmental Affairs Division of NATO, as part of the Advanced Research Workshop series. This book contains the proceedings of that meeting.

Major topics covered at the workshop include quantized theories of gravity, string theory, conformal field theory, cosmology, field theory approaches to critical phenomena and the renormalization group, matrix models, and field theory techniques applied to the theory of turbulence.

One common theme at the conference was the use of large-N matrix models to obtain exact results in a variety of different disciplines. For example, it has been known for several years that by taking a suitable double-scaling limit, certain string theories (or two-dimensional quantum gravity coupled to matter) can be re-obtained from the large-N expansion of matrix models. There continues to be a large activity in this area of research, which was well reflected by talks given at our workshop. Remarkably, large-N matrix models have very recently – just a few months before our meeting – been shown to have yet another deep relation to string theory. This time the connection goes through the so-called M-theory, which can loosely be thought of as a unifying theory of strings. Also this very recent subject was covered at our workshop. At the very last moment Yuri Makeenko had to cancel his participation. He fortunately agreed to send his contribution to this volume.

The understanding of the rôle M-theory plays for the different string theories originates in some remarkable results concerning duality that have been uncovered within the last 2-3 years. While so-called T-duality of string theory has been known for years, it is now being seen in a new light, and also other kinds of dualities have been found. Simultaneously, exact or approximate dualities have been shown to be properties of certain highly non-trivial supersymmetric quantum field theories in four dimensions. Both these dualities, their origin in string theory, as well as direct analyses of T-duality in the σ-model language were discussed at the meeting.

Another recent application of large-N matrix model techniques has been in the description of certain exact features of field theories with spontaneous chiral symmetry breaking (such as Quantum Chromodynamics). A recent flurry of activity has revealed a number of surprising universal aspects of such quantum field theories, related to the spectrum of the Dirac operator. At the meeting new and impressive Monte Carlo results from lattice gauge theory simulations were presented. They appeared to be in complete

agreement with the theoretical predictions. Also other aspects of this computational framework of matrix models were discussed at the meeting, for example in connection with the behavior at finite temperature, or in the limiting case of no chiral symmetry breaking.

One final, and also surprising, application of large-N matrix models which was covered at the workshop concerns the derivation of exact results in the theory of turbulence. Enlightening lectures were also given on the use of quantum field theory techniques in general to solve problems related to turbulence, and on the application of magnetohydrodynamics on cosmological scales.

As testified by this volume, numerous other topics were discussed at our workshop. It left the participants with the distinct impression that despite the long history of the field, we are now witnessing an extremely fruitful period of developments in quantum field theory.

We take this opportunity to thank Yu. Makeenko, A. Polychronakos and J.F. Wheater for serving on the international advisory committee. Very special thanks go to M. Praszalowicz and B. Brzezicka for their tireless help both before and during the workshop, and to P. Bialas, Z. Burda, and P. Jochym for much assistance. We would in particular like to thank Z. Burda for his help in preparing this volume.

Copenhagen and Cracow

Poul H. Damgaard
Jerzy Jurkiewicz

CONTENTS

LECTURERS

The Structure of 2D Quantum Space-Time 1
J. Ambjørn

Scaling Laws in Turbulence 23
K. Gawędzki

Field Theory as Free Fall 33
A. Carlini and J. Greensite

Center Dominance, Center Vortices, and Confinement 47
L. Del Debbio, M. Faber, J. Greensite and Š. Olejník

Duality and the Renormalization Group 65
P.E. Haagensen

Unification of the General Non-Linear Sigma Model
and the Virasoro Master Equation 79
J. de Boer and M. Halpern

A Matrix Model Solution of the Hirota Equation 97
V.A. Kazakov

Lattice Approximation of Quantum Electrodynamics 113
J. Kijowski and Gerd Rudolph

Three Introductory Talks on Matrix Models of Superstrings 127
Y. Makeenko

New Developments in the Continuous Renormalization Group 147
T.R. Morris

Primordial Magnetic Fields and Their Development
(Applied Field Theory) 159
P. Olesen

Towards Matrix Models of IIB Superstrings 171
P. Olesen

Quantum Mechanics of the Electric Charge 179
A. Staruszkiewicz

Universal Fluctuations in Dirac Spectra 187
J. Verbaarschot

Determination of Critical Exponents and Equation
of State by Field Theory Methods 217
J. Zinn-Justin

SEMINAR SPEAKERS

Collective Dynamics of a Domain Wall 233
H. Arodź

Path Space Formulation of the BFV Theorem 243
K. Bering

Surplus Anomaly and Random Geometries 251
P. Bialas, Z. Burda and D. Johnston

Topological Contents of 3D Seiberg-Witten Theory 261
B. Broda

Free Strings in Non-Critical Dimensions 269
M. Daszkiewicz, Z. Hasiewicz and Z. Jaskólski

Seiberg-Witten Theory, Integrable Systems and D-Branes 279
A. Marshakov

Microscopic Universality in Random Matrix Models of QCD 287
S.M. Nishigaki

New Developments in Non-Hermitian Random Matrix Models 297
R.A. Janik, M.A. Nowak, G. Papp and I. Zahed

Potential Topography and Mass Generation 315
M. Kudinov, E. Moreno and P. Orland

Past the Highest-Weight, and What You Can Find There 329
A.M. Semikhatov

The Spectral Dimension on Branched Polymer Ensembles 341
T. Jonsson and J.F. Wheater

Solving the Baxter Equation in High Energy QCD 349
J. Wosiek

Participants 359

Index 363

This page intentionally left blank

THE STRUCTURE OF 2D QUANTUM SPACE-TIME

Jan Ambjørn

The Niels Bohr Institute
Blegdamsvej 17,
DK-2100, Copenhagen Ø, Denmark

INTRODUCTION

The free relativistic particle provides us with the simplest example of "quantum geometry". The action of a free relativistic particle is just the length of its world line* in R^d. The classical path between two space-time points x and y is just the straight line. The system is quantized by summing over all paths P_{xy} from x to y with the Boltzmann weight determined by the classical action, which is simply the length $L(P_{xy})$ of the path. We write for the relativistic two-point function:

$$G(x,y) = \int \mathcal{D}P_{xy}\, \mathrm{e}^{-mL(P_{xy})}, \tag{1}$$

where m is the mass of the particle. The measure on the set of *geometric paths* P_{xy} can be defined and are related in a simple way (see [1]) to the ordinary Wiener measure on the set of parameterized paths†. One of the main features of this measure is that a "typical" path has a length

$$L_{xy} \sim \frac{1}{\varepsilon}\,|x-y|^2, \tag{2}$$

where ε is some cut-off. We say that the fractal dimension of a typical *random path* is two.

The generalizations of (2) go in various directions: one can consider higher dimensional objects like strings. The action of a string will be the area A of the world sheet F swept out by the string moving in R^d. If we consider closed strings the quantum propagator between two boundary loops L_1 and L_2 will be

$$G(L_1,L_2) = \int \mathcal{D}F_{L_1L_2}\, \mathrm{e}^{-A(F_{L_1L_2})}, \tag{3}$$

where the integration is over all surfaces in R^d with boundaries L_1 and L_2. Alternatively, we can for manifolds of dimensions higher than one consider actions which depend only

*In the following we will always be working in Euclidean space-time.
†The geometric paths are just parameterized paths up to diffeomorphisms.

New Developments in Quantum Field Theory
Edited by Damgaard and Jurkiewicz, Plenum Press, New York, 1998

on the intrinsic geometry of the manifold. The simplest such action is the Einstein-Hilbert action, here written for a n-dimensional manifold M:

$$S(g) = \Lambda \int_{\mathcal{M}} \mathrm{d}^n\xi \sqrt{g(\xi)} - \frac{1}{16\pi G} \int_{\mathcal{M}} \mathrm{d}^n\xi \sqrt{g(\xi)}\, R(\xi), \tag{4}$$

where g is the metric on M and R the scalar curvature defined from g. Quantization of geometry means that we should sum over all geometries g with the weight $\mathrm{e}^{-S(g)}$. The partition function will be

$$Z(\Lambda, G) = \int \mathcal{D}[g]\, \mathrm{e}^{-S(g)}, \tag{5}$$

where the integration is over all equivalence classes of metrics, i.e. metrics defined up to diffeomorphisms. One can add matter coupled to gravity to the above formulation. Let $S_m(\phi, g)$ be the diffeomorphism invariant Lagrangian which describes the classical dynamics of the matter fields in a fixed background geometry defined by g and let λ denote the coupling constants of the scalar fields. The quantum theory will be defined by

$$Z(\Lambda, G, \lambda) = \int \mathcal{D}[g]\mathcal{D}\phi\, \mathrm{e}^{-S(g)-S_m(\phi,g)}. \tag{6}$$

Two-dimensional quantum gravity is particularly simple. As long as we do not address the question of topology changes of the underlying manifold M, the Einstein-Hilbert action (4) simplifies since the curvature term is just a topological constant, and we can write

$$S(g) = \Lambda \int_{\mathcal{M}} \mathrm{d}^2\xi \sqrt{g(\xi)} \quad \text{(two dimensions)}. \tag{7}$$

Classical string theory, as defined by the area action $A(F)$, has an equivalent formulation where an independent intrinsic metric $g(\xi)$ is introduced on the two-dimensional manifold corresponding to the world sheet and where the coordinates of the surface, $x(\xi) \in R^d$, are viewed as d scalar fields on the manifold with metric $g(\xi)$. The quantum string theory will then be a special case of two-dimensional quantum gravity coupled to matter, as defined by (6), with $S(g)$ given by (7). In the following we will study this theory, with special emphasis on pure two-dimensional quantum gravity, i.e. two-dimensional quantum gravity without any matter fields.

A TOY MODEL: THE FREE PARTICLE

It is instructive first to perform the same exercise for the free relativistic particle given by (1). In this case one can approximate the integration over random paths by the summation and integration over the class of piecewise linear paths where the length of each segment of the path is fixed to a, i.e. we make the replacement

$$\int \mathcal{D}P_{xy} \to \sum_{P_{xy}} \equiv \sum_n \int \prod \mathrm{d}\hat{e}_i\, \delta\Big(a \sum_i \hat{e}_i - (x - y)\Big), \tag{8}$$

where $\hat{e}_i$ denote unit vectors in R^d and $\sum_{P_{xy}}$ is a symbolic notation of the summation and integration over the chosen class of paths. The action is simply $m_0 \cdot na$ for a path with n "building blocks". A "discretized" two-point function is then defined by

$$G_a(x, y; m_0) = \sum_{P_{xy}} \mathrm{e}^{-m_0 L(P_{xy})} \equiv \sum_n \mathrm{e}^{-m_0 an} \int \prod \mathrm{d}\hat{e}_i\, \delta\Big(a \sum_i \hat{e}_i - (x - y)\Big). \tag{9}$$

The integration over the unit vectors is most easily performed by a Fourier transformation with removes the δ-function:

$$G_a(p;m_0) = \int dx\, e^{-ip\cdot(x-y)} G_a(x,y;m_0) = \sum_n e^{-m_0 an} \int \prod d\hat{e}_i\, e^{-iap\cdot\hat{e}_i}. \tag{10}$$

Since

$$\int d\hat{e}_i\, e^{-iap\cdot\hat{e}_i} = 2\pi^{d/2} \left[\frac{J_{(d-1)/2}(ap)}{(ap)^{(d-1)/2}} \right] \equiv f(ap), \tag{11}$$

the final expression for $G_a(p;m_0)$ becomes

$$G_a(p;m_0) = \sum_n \left(e^{m_0 a} f(ap) \right)^n = \frac{1}{1 - e^{-m_0 a} f(ap)}. \tag{12}$$

We only need the following properties of $f(ap)$:

$$f(ap) = f(0)(1 - c^2(ap)^2 + \cdots), \qquad f(0) > 0.$$

In order to obtain the continuum two-point function we have to take $a \to 0$ and this involves a *renormalization of the bare mass m_0 as well as a wave-function renormalization.* Let us define the *physical mass* m_{ph} by

$$e^{-m_0 a} f(ap) \to 1 - c^2 m_{ph}^2 a^2, \quad \text{i.e.} \quad m_0 = \frac{\log f(0)}{a} + c^2 m_{ph}^2 a. \tag{13}$$

With this fine tuning of the *bare mass* m_0 we obtain for $a \to 0$

$$G_a(p;m_0) \sim a^{-2}\, G_{cont}(p;m_{ph}), \tag{14}$$

where the continuum two-point function of the free relativistic particle is

$$G_{cont}(p;m_{ph}) \equiv \frac{1}{p^2 + m_{ph}^2}.$$

The prefactor $1/a^2$ in eq. (14) is a so-called wave-function renormalization. It is related to the short distance behavior of the propagator as will be discussed below.

Scaling Relations and Geometry

It is worth rephrasing the results obtained so far in terms of dimensionless quantities and in this way make the statistical mechanics aspects more visible. Introduce $\mu = m_0 a$ and $q = ap$ and view the coordinates in R^d as dimensionless. The steps in the discretized random walk will then be of length 1 and (12) reads

$$G_\mu(q) = \sum_n e^{-\mu n} f(q) = \frac{1}{1 - e^{-\mu} f(q)}. \tag{15}$$

It is seen that μ acts like a chemical potential for inserting additional sections in the piecewise linear random walk and that we have a *critical value* $\mu_c = \log f(0)$ such that the average number of steps of the random walk diverge for $\mu \to \mu_c$ from above. This is why we can take a continuum limit when $\mu \to \mu_c$. In fact, the relation (13) becomes

$$\mu - \mu_c = m_{ph}^2 a^2, \tag{16}$$

which defines a as a function of μ:

$$a(\mu) = m_{ph}^{-1} (\mu - \mu_c)^{1/2}. \tag{17}$$

Further, we see that the so-called *susceptibility* diverges as $\mu \to \mu_c$:

$$\chi(\mu) \equiv \int \mathrm{d}^d x \; G_\mu(x) = G_\mu(q=0) = \frac{1}{1 - \mathrm{e}^{-(\mu-\mu_c)}} \sim \frac{1}{\mu - \mu_c}. \tag{18}$$

These considerations can be understood in a more general framework. It is not difficult to show that $G_\mu(x)$ has to fall off exponentially for large x under very general assumptions concerning the probabilistic nature of the (discretized) random walk. It follows from standard sub-additivity arguments. In essence, they say that the random walks from x to y which pass through a given point z constitute a subset of the total number of random walks from x to y. This implies that

$$G_\mu(x,y) \sim \mathrm{e}^{-m(\mu)|x-y|} \quad \text{for} \quad |x-y| \gg \frac{1}{m(\mu)}. \tag{19}$$

Let us now *assume* that

$$m(\mu) \to (\mu - \mu_c)^\nu \quad \text{for} \quad \mu \to \mu_c. \tag{20}$$

In order that $G_\mu(x, y)$ has a non-trivial limit for $\mu \to \mu_c$ we have to introduce the following generalization of (16)

$$m(\mu) = m_{ph} a(\mu), \qquad x_{ph} = x\, a(\mu), \quad \text{i.e.} \quad a(\mu) \sim (\mu - \mu_c)^\nu. \tag{21}$$

It is clear that $m(\mu)$ has the interpretation as inverse correlation length (or a mass). If the mass $m(\mu)$ goes to zero as $\mu \to \mu_c$ the two-point function $G_\mu(x, y)$ will in general satisfy a power law for $|x - y|$ much less that the correlation length:

$$G_\mu(x,y) \sim \frac{1}{|x-y|^{d-2+\eta}} \quad \text{for} \quad |x-y| \ll \frac{1}{m(\mu)}. \tag{22}$$

Finally the susceptibility is defined as in (18):

$$\chi(\mu) = \int d^d x \; G_\mu(x,y) \sim \frac{1}{(\mu - \mu_c)^\gamma}, \tag{23}$$

where the *critical exponents* ν, η and γ (almost) by definition satisfy

$$\gamma = \nu(2 - \eta) \qquad \text{(Fisher's scaling relation).} \tag{24}$$

For the random walk representation of the free particle considered above we have:

$$\nu = \frac{1}{2}, \qquad \eta = 0, \qquad \gamma = 1. \tag{25}$$

Let us now show that $1/\nu$ is the extrinsic Hausdorff dimension of the random walk between x and y. The average length of a path between x and y is equal

$$\langle L_{xy} \rangle = \frac{\sum_{P_{xy}} L(P_{xy})\, \mathrm{e}^{-\mu L(P_{xy})}}{\sum_{P_{xy}} \mathrm{e}^{-\mu L(P_{xy})}} = -\frac{\mathrm{d} \log G_\mu(x,y)}{\mathrm{d}\mu}. \tag{26}$$

For $|x - y|$ sufficiently large, such that (19) can be used, we have

$$\langle L_{xy} \rangle \sim m'(\mu)|x-y|. \tag{27}$$

However, the continuum limit has to be taken in such a way that

$$m(\mu)|x-y| = m_{ph}|x_{ph} - y_{ph}|, \tag{28}$$

i.e. independent of μ for $\mu \to \mu_c$. From (20) and (28) we obtain

$$\langle L_{xy} \rangle \sim \frac{m'(\mu)}{m(\mu)} \sim \frac{1}{\mu - \mu_c} \sim |x - y|^{1/\nu}. \tag{29}$$

We define the extrinsic Hausdorff dimension by

$$\langle L_{xy} \rangle \sim |x - y|^{d_H^{(e)}}, \tag{30}$$

and we conclude that the critical exponent ν is related to the extrinsic Hausdorff dimension $d_H^{(e)}$ by

$$d_H^{(e)} = \frac{1}{\nu} \tag{31}$$

Summary

Above it has been shown how it is possible by a simple, appropriate choice of regularization of the set of geometric paths from x to y to define the measure DP_{xy}. One of the basic properties of this measure, namely that a generic path has $d_H^{(e)} = 2$ was easily understood. It is important that the regularization is performed directly in the set of geometric paths. In this way it becomes a reparameterization invariant regularization of DP_{xy}. The regularization can be viewed as a grid in the set of geometric paths, which becomes uniformly dense in the limit $\mu \to \mu_c$ or alternatively $a(\mu) \to 0$. The Wiener measure itself is defined on the set of *parameterized paths* and will not lead to the relativistic propagator.

THE FUNCTIONAL INTEGRAL OVER 2D GEOMETRIES

As described above the partition function for two-dimensional geometries is

$$Z(\Lambda) = \int \mathcal{D}[g]\, e^{-\Lambda V_g}, \qquad V_g \equiv \int_{\mathcal{M}} d^2\xi \sqrt{g(\xi)}. \tag{32}$$

It is sometimes convenient to consider the partition function where the volume V of space-time is kept fixed. We define it by

$$Z(V) = \int \mathcal{D}[g]\, \delta(V - V_g), \tag{33}$$

such that

$$Z(\Lambda) = \int_0^\infty dV\, e^{-V\Lambda} Z(V). \tag{34}$$

It is often said that two-dimensional quantum gravity has little to do with four-dimensional quantum gravity since there are no dynamical gravitons in the two-dimensional theory (the Lagrangian is trivial since it contains no derivatives of the metric). However, all the problems associated with the definition of reparameterization invariant observables are still present in the two-dimensional theory, and the theory is in a certain sense *maximal quantum*: from (33) it is seen that *each equivalence class of metrics is included in the path integral with equal weight,* i.e. we are as far from a classical limit as possible. Thus the problem of defining genuine reparameterization invariant observables in quantum gravity is present in two dimensional quantum gravity as well. Here we will discuss the so-called Hartle-Hawkings wave-functionals and the two-point functions. The Hartle-Hawking wave-functional is defined by

$$W(L; \Lambda)) = \int_L \mathcal{D}[g]\, e^{-S(g;\Lambda)} \tag{35}$$

where L symbolizes the *boundary* of the manifold M. In dimensions higher than two one should specify (the equivalence class of) the metric on the boundary and the functional integration is over all equivalence classes of metrics having this boundary metric. In two dimensions the equivalence class of the boundary metric is uniquely fixed by its length and we take L to be the length of the boundary. It is often convenient to consider boundaries with variable length L by introducing a *boundary cosmological term* in the action:

$$S(g;\Lambda,\Lambda_b) = \Lambda \int_{\mathcal{M}} \mathrm{d}^2\xi\sqrt{g(\xi)} + \Lambda_B \int_{\partial\mathcal{M}} \mathrm{d}s, \tag{36}$$

where ds is the invariant line element corresponding to the boundary metric induced by g and Λ_B is called the boundary cosmological constant. We can then define

$$W(\Lambda_B,\Lambda) = \int \mathcal{D}[g]\, \mathrm{e}^{-S(g;\Lambda,\Lambda_B)}. \tag{37}$$

The wave-functions $W(L;\Lambda)$ and $W(\Lambda_B,\Lambda)$ are related by a Laplace transformation in the boundary length:

$$W(\Lambda_B,\Lambda) = \int_0^\infty \mathrm{d}L\, \mathrm{e}^{-\Lambda_B L}\, W(L;\Lambda). \tag{38}$$

The two-point function is defined by

$$G(R;\Lambda) = \int \mathcal{D}[g]\, \mathrm{e}^{-S(g,\Lambda)} \iint \mathrm{d}^2\xi\sqrt{g(\xi)}\, \mathrm{d}^2\eta\sqrt{g(\eta)}\, \delta(D_g(\xi,\eta) - R), \tag{39}$$

where $D_g(\xi,\eta)$ denotes the geodesic distance between ξ and η in the given metric g. Again, it is sometimes convenient to consider a situation where the space-time volume V is fixed. This function, $G(R;V)$ will be related to (39) by a Laplace transformation, as above for the partition function Z:

$$G(R;\Lambda) = \int_0^\infty \mathrm{d}V\, \mathrm{e}^{-V\Lambda}\, G(R;V). \tag{40}$$

It is seen that $G(R;\Lambda)$ and $G(R;V)$ has the interpretation of partition functions for universes with two marked points separated a given geodesic distance R. If we denote the average volume of a spherical shell of geodesic radius R in the class of metrics with space-time volume V by $S_V(R)$, we have by definition

$$S_V(R) = \frac{G(R;V)}{VZ(V)}. \tag{41}$$

One can define an intrinsic fractal dimension, d_H, of the ensemble of metrics by

$$\lim_{R\to 0} S_V(R) \sim R^{d_H-1}(1 + O(R)). \tag{42}$$

Alternatively, one could take over the random walk definition of d_H. According to this definition

$$\langle V\rangle_R \sim -\frac{\partial \log G(R;\Lambda)}{\partial\Lambda} \sim R^{d_H} \tag{43}$$

for a suitable range of R related to the value of Λ. I will show that the two definitions agree in the case of pure gravity. Eq. (42) can be viewed as a "local" definition of d_H, while eq. (43) is "global" definition. Since the two definitions result in the same d_H two-dimensional gravity has a genuine fractal dimension over all scales.

Eq. (33) shows that the calculation of $Z(V)$ is basically a counting problem: each geometry, characterized by the equivalence class of metrics $[g]$, appears with the same weight. The same is true for the other observables defined above. One way of performing the summation is to introduce a suitable regularization of the set of geometries by means of a cut-off, to perform the summation with this cut-off and then remove the cut-off, like in the case of geometric paths considered above.

The Regularization

The integral over geometric paths were regularized by introducing a set of basic building blocks, "rods of length *a*", which were afterwards integrated over all allowed positions in R^d. Let us imitate the same construction for two-dimensional space-time [2, 3, 4]. The natural building blocks will be equilateral triangles with side lengths ε, but in this case there will be no integration over positions in some target space‡. We can glue the triangles together to form a triangulation of a two-dimensional manifold M with a given topology. If we view the triangles as flat in the interior, we have in addition a unique piecewise linear metric assigned to the manifold, such that the volume of each triangle is $dA_\varepsilon = \sqrt{3}\varepsilon^2/4$ and the total volume of a triangulation T consisting of N_T triangles will be $N_T dA_\varepsilon$, i.e. we can view the triangulation as associated with a Riemannian manifold (M,g). In the case of a one-dimensional manifold the total volume is the only reparameterization invariant quantity. For a two-dimensional manifold M the scalar curvature R is a *local* invariant. This local invariance in present in a natural way when we consider various triangulations. Each vertex v in a triangulation has a certain order n_v. In the context of two-dimensional piecewise linear geometry, curvature is located at the vertices and is characterized by a *deficit angle*

$$\Delta_v = \frac{\pi}{3}(6 - n_v), \tag{44}$$

such that the total curvature of the manifold is

$$\int \sqrt{g}R = \sum_v \Delta_v. \tag{45}$$

From this point of view a summation over triangulations of the kind mentioned above will form a grid in the class of Riemannian geometries associated with a given manifold M. The hope is that the grid is sufficient dense and uniform to be able the describe correctly the functional integral over all Riemannian geometries when $\varepsilon \to 0$.

We will show that it is the case by explicit calculations, where some of the results can be compared with the corresponding continuum expressions. They will agree. But the surprising situation in two-dimensional quantum gravity is that the analytical power of the regularized theory seems to exceed that of the formal continuum manipulations. Usually the situation is the opposite: regularized theories are either used in a perturbative context to remove infinities order by order, or introduced in a non-perturbative setting in order make possible numerical simulations. Here we will derive analytic (continuum) expressions with an ease which can presently not be matched by formal continuum manipulations.

The Hartle-Hawking Wave-Functional

Let us calculate the discretized version, $w(\lambda, \mu)$ of the Hartle-Hawking wave-functional $W(\Lambda_B,\Lambda)$, defined by (37). We assume the underlying manifold M has the topology of the disk. First note that the discretized action corresponding to (36) can be written as

$$S_T(\mu, \lambda) = \mu N_T + \lambda l_T, \tag{46}$$

where the given triangulation T also defines the metric, N_T and l_T denote the number of triangles and the number of links at the boundary of T, respectively, while μ and λ are

‡ We could introduce such embedding in R^d, but in that case we would not consider two-dimensional gravity but rather bosonic string theory, where the embedded surface was the world sheet of the string, as already mentioned above [3, 5].

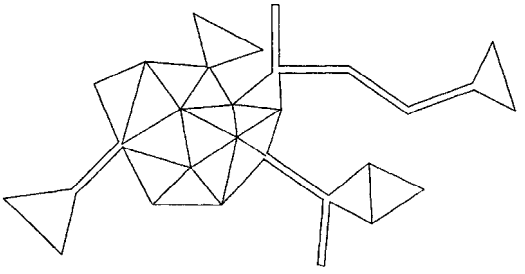


Figure 1. A typical unrestricted "triangulation".

the dimensionless "bare" cosmological and boundary cosmological coupling constants corresponding to Λ and Λ_B. We can now write

$$w(\lambda, \mu) = \sum_T e^{-S_T(\mu,\lambda)}, \tag{47}$$

where the summation is over all triangulations of the disk. Until now I have not specified the class of triangulations. The precise class should not be important, by universality, since any structure not allowed at the smallest scale by one class of triangulations can be imitated at a somewhat larger scale. Thus, it is convenient to choose a class of "triangulations" which results in the simplest equation. They are defined as the class of complexes homeomorphic to the disk that can be obtained by successive gluing together of triangles and a collection of double-links which we consider as (infinitesimally narrow) strips, where links, as well as triangles, can be glued onto the boundary of a complex both at vertices and along links. Gluing a double-link along a link makes no change in the complex. An example of such a complex is shown in fig. 1.

By introducing

$$g = e^{-\mu}, \qquad z = e^{\lambda}, \tag{48}$$

we can write (47) as

$$w(z, g) = \sum_{l,k} w_{l,k} g^k z^{-l-1} = \sum_l \frac{w_l(g)}{z^{l+1}}, \tag{49}$$

where $w_{k,l}$ is the number of triangulations of the disk with k triangles and a boundary of l links. We see that $w(z, g)$ is the *generating function* § for $\{w_{l,k}\}$. The generating function $w(z, g)$ satisfies the following equation, depicted graphically in fig. 2,

$$w(z, g) = zg\, w(z, g) + \frac{1}{z} w^2(z, g). \tag{50}$$

This equation is not correct from the smallest values of of the boundary-length l, as is clear from fig. (2), since all boundaries on the right-hand of the equation have a boundary length $l > 1$. Denote by $w_1(g)$ the generating function for triangulations of the disk with a boundary with only one link (see eq. (49)). The correct equation which replaces (50) is

$$w(z, g) = \frac{1}{z} + zg \left(w(z, g) - \frac{1}{z} - \frac{w_1(g)}{z^2} \right) + \frac{1}{z} w^2(z, g), \tag{51}$$

§ In (49) I have used $1/z$ rather than z as indeterminate for $\{w_{l,k}\}$ for later convenience, and for the same reason multiplied (49) by an additional factor $1/z$ relatively to (47).

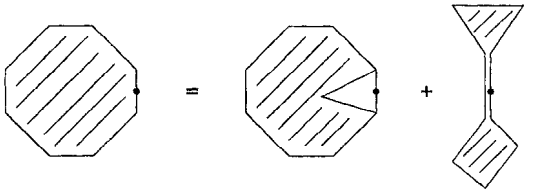


Figure 2. Graphical representation of eq. 51.

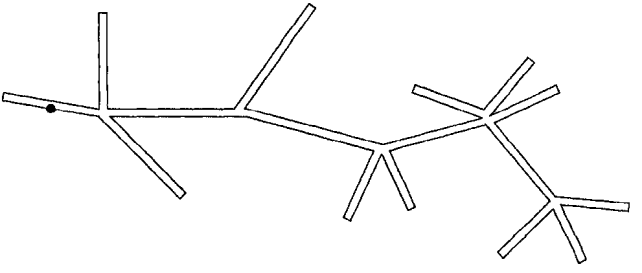


Figure 3. A boundary graph with no internal triangles.

if we use the normalization that a single vertex is represented by $1/z$. This equation is similar in spirit to the equation studied by Tutte in his seminal paper[6] from 1962, and it can by shown that it has a unique solution where all coefficients $w_{l,k}$ are positive. The solution is given by

$$w(z,g) = \frac{1}{2}\left(z - gz^2 + (gz - c_2(g))\sqrt{(z - c_+(g))(z - c_-(g))}\right), \tag{52}$$

where $c_-(g)$, $c_+(g)$ and $c_2(g)$ are analytic functions of g in a neighborhood of $g = 0$, with the initial conditions

$$c_2(0) = 1, \quad c_+(0) = 2, \quad , c_-(0) = -2. \tag{53}$$

Thus, for $g = 0$ we have

$$w(z) = \frac{1}{2}\left(z - \sqrt{z^2 - 4}\right) = \sum_{l=0}^{\infty} \frac{w_2 l}{z^{2l+1}}, \tag{54}$$

where the coefficients w_{2l} have the interpretation as the number of boundaries with no internal triangles, see fig. 3. We have

$$w_{2l} = \frac{2l!}{(l+1)!\,l!} = \frac{1}{\pi} l^{-3/2}\, 2^{2l}(1 + O(l)), \tag{55}$$

i.e. the number of such boundaries grows exponentially with the length l. We can view $1/z$ as the so-called fugacity[¶] for the number of boundary links, and the radius of convergence (here 1/2) can be viewed as the maximal allowed value of the fugacity.

[¶] The fugacity f is related to the chemical potential μ by $f = e^{-\mu}$.

When z approaches $z_c(0) = 2$ the average length of a typical boundary will diverge. In the same way g acts as the fugacity for triangles. As g increases the average number of triangles will increase, and at a certain critical value g_c some suitable defined average value of triangles will diverge. In terms of the coefficients $w_{l,k}$ in (49) it reflects an exponential growth of $w_{l,k}$ for $k \to \infty$, independent of l, i.e. the functions $w_l(g)$ all have the same radius of convergence g_c. For a given value $g < g_c$ we have a critical value $z_c(g)$ at which the average boundary length will diverge. As g increases towards g_c, $z_c(g)$ will increase towards $z_c \equiv z_c(g_c)$.

From the explicit solutions for $c_\pm(g)$ and $c_0(g)$ it is found that

$$c_+(g_c) = z_c = c_2(g_c)/g_c, \tag{56}$$

and near g_c we have, with $\Delta g \equiv g_c - g$:

$$c_+(g) = z_c\Big(1 + \frac{1}{2}\sqrt{\Delta g}\Big), \qquad c_2(g) = z_c g_c\Big(1 - \sqrt{\Delta g}\Big). \tag{57}$$

In particular, g_c is the radius of convergence for $c_+(g)$ and $c_2(g)$.

It is now possible to define a continuum limit of the above discretized theory by approaching the critical point in a suitable way:

$$g(\Lambda) = g_c(1 - \Lambda\varepsilon^2), \qquad z = z_c(1 + \Lambda_B\varepsilon). \tag{58}$$

If we return to the relations (48) between g and μ and z and λ, respectively, we can write (58) as follows:

$$-\mu_c = \Lambda\varepsilon^2, \qquad \lambda - \lambda_c = \Lambda_B\varepsilon, \tag{59}$$

where μ_c and λ_c correspond to g_c and z_c, respectively. We can now, as is standard procedure in quantum field theory, relate coupling constants μ and λ to Λ and Λ_B by an *additive renormalization*. The dimensionless coupling constants μ and λ are associated with so-called *bare coupling constants* Λ_0 and Λ_{B0} as follows:

$$\mu N_T + \lambda l_T = \frac{\mu}{\varepsilon^2}N_T\varepsilon^2 + \frac{\lambda}{\varepsilon}l\varepsilon \equiv \Lambda_0\int_{\mathcal{M}} \mathrm{d}^2\xi\sqrt{g} + \Lambda_{B0}\int_{\partial\mathcal{M}} ds. \tag{60}$$

We can now interpret (59) as an additive renormalization of the bare coupling constants:

$$\Lambda_0 = \frac{\mu_c}{\varepsilon^2} + \Lambda, \qquad \Lambda_{B0} = \frac{\lambda_c}{\varepsilon} + \Lambda_B. \tag{61}$$

This additive renormalization is to be expected from a quantum field theoretical point of view since both coupling constants have a mass-dimension.

Using the known behavior (57) of $c_\pm(g)$ and $c_2(g)$ in the neighborhood g_c, we get from (52) (except for the first two terms with are analytic in g and therefore "non-universal" terms[||] which can be shown to play no role for continuum physics):

$$w(z, g) \sim \varepsilon^{3/2}W(\Lambda_B, \Lambda) \tag{62}$$

where [7,8]

$$W(\Lambda_B, \Lambda) \sim (\Lambda_B - \frac{1}{2}\sqrt{\Lambda})\sqrt{\Lambda_B + \sqrt{\Lambda}}. \tag{63}$$

Again, the factor $\varepsilon^{3/2}$ has a standard interpretation in the context of quantum field theory: it is a wave-function renormalization.

By an inverse discrete Laplace transformation one obtains $w(l, g)$ from $w(z, g)$, and by an ordinary inverse Laplace transformation one obtains

$$W(L, \Lambda) = L^{-5/2}(1 + L\sqrt{\Lambda})\,\mathrm{e}^{-L\sqrt{\Lambda}}. \tag{64}$$

[||]Analytic terms are usually non-universal since trivail analytic redefinitions of the coupling constants can change these terms completely.

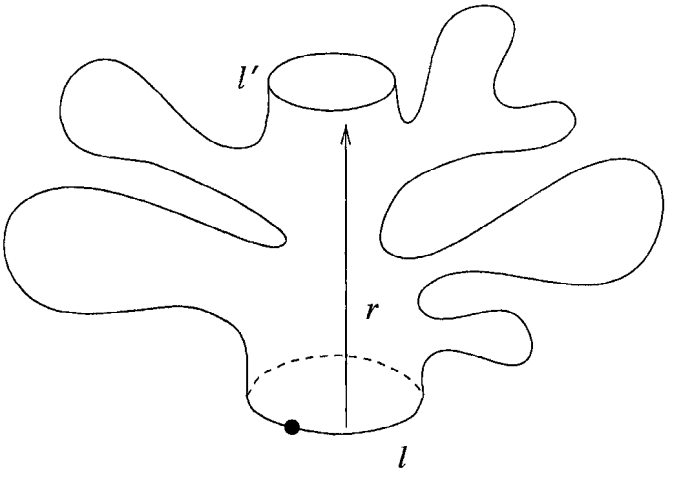


Figure 4. A typical surface contributing to $G_\mu(l, l' : r)$. The "dot" on the entrance loop signifies that the entrance loop has one marked link.

The Two-Point Function

Let us return to the calculation of $G(R; \Lambda)$. Using the regularization we define a *geodesic two-loop function* by

$$G_\mu(l_1, l_2; r) = \sum_T e^{-\mu N_T}, \tag{65}$$

and the class of triangulations which enters in the sum have the topology of a cylinder with an "entrance loop" of length l_1 and with one marked linked, and an "exit loop" of length l_2 and without a marked link, the loops separated by a geodesic distance r, see fig. 4. We say the geodesic distance between the exit loop and the entrance loop is r if each point on the exit loops has a minimal geodesic distance r to the set of points on the entrance loop. Note the asymmetry between exit and entrance loops in the definition. On the piecewise linear manifolds geodesic distances are uniquely defined. However, it is often convenient to use a graph-theoretical definition, since this makes combinatorial arguments easier. Here I define the geodesic distance between links (or vertices) as the shortest path along neighboring triangles.

$G_\mu(l_1, l_2 ; r)$ satisfies an equation [9], which is essentially equivalent to the equation satisfied by the Hartle-Hawking wave function $w(l, \mu)$ for a disk with boundary length l. It is obtained by a deformation of the entrance loop:

$$G_\mu(l_1, l_2; r) = g G_\mu(l + 1, l; r) + 2 \sum_{l=0}^{l_1-2} w(l, \mu)\, G_\mu(l_1 - l - 2, l_2; r). \tag{66}$$

In fig. 5 the possible elementary deformations of the entrance loops is shown. It is analogous to fig. 2. The second term in eq. (66) corresponds to the case where the surface splits in two after the deformation. We can view the process as a "peeling" of the surface, which occasionally chops off outgrows with disk topology as shown in fig. 6. The application of the one-step peeling l_1 times should on average correspond

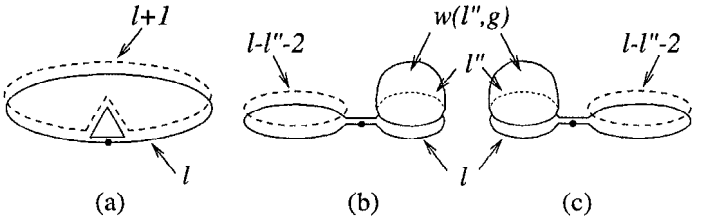


Figure 5. The "peeling" decomposition: a marked link on the entrance boundary can either belong to a triangle or to a "double" link. The dashed curved indicates the new entrance loop.

to cutting a slice (see fig. 6), of thickness one (or ε, which we have chosen equal 1 for convenience in the present considerations) from the surface. Thus we identify the change caused by one elementary deformation with

$$\left[\frac{\partial}{\partial r} G_\mu(l_1, l_2; r)\right] \frac{1}{l_1}, \tag{67}$$

forgetting for the moment that r is an integer. It follows that we can write

$$\begin{aligned}\frac{\partial}{\partial r} G_\mu(l_1, l_2; r) = & -l_1 G_\mu(l_1, l_2; r) + g l_1 G_\mu(l_1 + 1, l_2; r) \\ & + 2 \sum_l l w(l, \mu)\, G_\mu(l_1 - l - 2, l_2; r).\end{aligned} \tag{68}$$

To solve the combinatorial problem associated with (68) it is convenient (as for $w(l,\ \mu)$) to introduce the generating function $G_\mu(z_1, z_2; r)$ associated with (65):

$$G_\mu(z_1, z_2; r) = \sum_{l_1, l_2} \frac{G_\mu(l_1, l_2; r)}{z_1^{l_1+1} z_2^{l_2+1}}. \tag{69}$$

With this notation eq. (68) becomes

$$\frac{\partial}{\partial r} G_\mu(z_1, z_2; r) = \frac{\partial}{\partial z_1}\Big[\Big(z_1 - g z_1^2 - 2w(z, g)\Big)\, G_\mu(z_1, z_2; r)\Big]. \tag{70}$$

This differential equation can be solved since we know $w(z,\ g)$ (for details see [10, 9]). However, we are interested in the two-point function. It is obtained from the two-loop function be closing the exit loop with a "(cap" (i.e. the full disk amplitude $w(l,\ \mu)$) and shrinking the entrance loop to a point. The corresponding equation is

$$\begin{aligned}G_\mu(r) &= \sum_{l_2} G_\mu(l_1 = 1, l_2; r)\, l_2 w(l_2, g) \\ &= \oint \frac{\mathrm{d}z'}{2\pi i z'}\Big[z^2 G_\mu\Big(z, \frac{1}{z'}; r\Big)\Big]\ \Big[\frac{\partial}{\partial z'}[z' w(z', g)]\Big]\Big|_{z=\infty}.\end{aligned} \tag{71}$$

Since $w(z,\ g)$ and $G(z_1, z_2; r)$ are known we can find $G_\mu(r)$, see [11] for details. For $\mu \to \mu_c$, i.e. in the continuum limit, we obtain:

$$G_\mu(r) \sim (\mu - \mu_c)^{3/4} \frac{\cosh\Big[r\sqrt[4]{\mu - \mu_c}\Big]}{\sinh^3\Big[r\sqrt[4]{\mu - \mu_c}\Big]}. \tag{72}$$

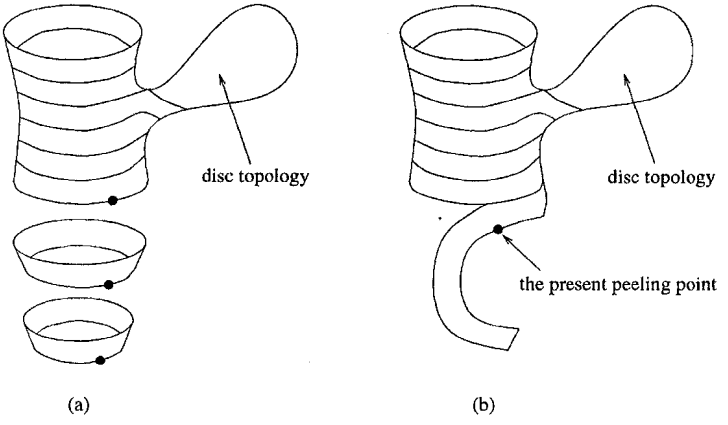


Figure 6. Decomposition of a surface by (a) slicing and (b) peeling.

If we introduce the following *continuum geodesic distance* $R = r\sqrt{\varepsilon}$, it follows that we can write:

$$G_\mu(r) \sim \varepsilon^{3/2} G(R;\Lambda), \qquad G(R;\Lambda) = \Lambda^{3/4} \frac{\cosh\left[R\sqrt[4]{\Lambda}\right]}{\sinh^3\left[R\sqrt[4]{\Lambda}\right]}. \tag{73}$$

The factor $\varepsilon^{3/2}$ is again a wave-function renormalization which connects the dimensionless, regularized $G_\mu(r)$ and the continuum two-point function $G(R;\ \Lambda)$.

We can compare the behavior of $G_\mu(r)$ (or $G(R;\ \Lambda)$) with that of the random walk two-point function. All conclusions and interpretations remain valid here, except that we only work with intrinsic geometric objects. First note that $G_\mu(r)$ falls off exponentially for large r (see (19) for the random walk). As for the random walk it follows from general sub-additive properties of $G_\mu(r)$. In addition the associated mass satisfies (20) since $m(\mu) \to 0$ for $\mu \to \mu_c$ as $(\mu - \mu_c)^\nu$ with $\nu = 1/4$. The behavior of $G_\mu(r)$ for $r << 1/m(\mu)$ is purely power-like corresponding to $\eta = 4$ in (22), and finally

$$\chi(\mu) = \int \mathrm{d}r\, G_\mu(r) \sim (\mu - \mu_c)^{1/2} + \text{less singular terms}, \tag{74}$$

i.e. $\gamma = -1/2$ according to definition (23). Needless to say, Fisher's scaling relation (24) is satisfied and the exponents for two-dimensional quantum gravity:

$$\nu = \frac{1}{4}, \qquad \eta = 4, \qquad \gamma = -\frac{1}{2}, \tag{75}$$

should be compared the the values for the random walk (see (25)). In particular it follows that *the intrinsic fractal dimension, d_H, of two-dimensional quantum space-time is*

$$d_H = \frac{1}{\nu} = 4. \tag{76}$$

This d_H is a "globally defined" Hausdorff dimension in the sense discussed below (43) as is clear from (72) or (73). We can determine the "local" d_H, defined by eq. (42), by performing the inverse Laplace transformation of $G(R;\ \Lambda)$ to obtain $G(R;\ V)$. The

average volume $S_V(R)$ of a spherical shell of geodesic radius R in the ensemble of universes with space-time volume V can then calculated from (41). One obtains

$$S_V(R) = R^3 F(R/V^{\frac{1}{4}}), \quad F(0) > 0, \tag{77}$$

where $F(x)$ can be expressed in terms of certain generalized hyper-geometric functions [1 2]. Eq. (77) shows that also the "local" $d_H = 4$.

Summary

It has been shown how it is possible to calculate the functional integral over two-dimensional geometries, in close analogy to the functional integral over random paths. One of the most fundamental results from the latter theory is that the generic random path between two points in R^d, separated a geodesic distance R, is *not* proportional to R but to R^2. This famous result has a direct translation to the theory of random two-dimensional geometries: the generic volume of a closed universe of radius R is *not* proportional to R^2 but to R^4.

2D GRAVITY COUPLED TO MATTER

While the fractal structure of pure two-dimensional quantum gravity can be analyzed in detail as described above, the change in the fractal structure when two-dimensional quantum gravity is coupled to matter is not fully understood. From an analytical point of view there have been two suggestions for the intrinsic Hausdorff dimension, as a function of the central charge c of a conformal matter theory coupled to gravity:

$$d_h(c) = 2 \times \frac{\sqrt{25-c}+\sqrt{49-c}}{\sqrt{25-c}+\sqrt{1-c}}, \tag{78}$$

and

$$\tilde{d}_h(c) = -\frac{2}{\gamma}, \tag{79}$$

where the string susceptibility γ is given by the famous KPZ formula:

$$\gamma(c) = \frac{1-c-\sqrt{(1-c)(25-c)}}{12}. \tag{80}$$

In the following I will review some of the arguments which lead to formula (78) and (79), respectively, and explain the present understanding of the formulas.

Liouville Diffusion

Consider the propagation of a massless scalar particle on a compact Riemannian manifold with metric g and total volume V. The scalar Laplacian is defined by

$$\Delta_g = \frac{1}{\sqrt{g}} \frac{\partial}{\partial \xi^\alpha} \sqrt{g} g^{\alpha\beta} \frac{\partial}{\partial \xi^\beta}, \tag{81}$$

and the diffusion process related to a scalar particle which is located at point ξ_0 at the diffusion time $\tau = 0$ can be expressed in terms of Δ_g by

$$K_g(\xi, \xi_0; \tau) = e^{\tau \Delta_g(\xi)} \frac{1}{\sqrt{g}} \delta(\xi - \xi_0). \tag{82}$$

The scalar propagator is related to the heat kernel K_g by

$$G_g(\xi,\xi_0) = \int_0^\infty d\tau\; K_g(\xi,\xi_0;\tau), \tag{83}$$

and the heat kernel has the following asymptotic τ expansion**

$$K_g(\xi,\xi_0,\tau) \sim \frac{e^{-d_g^2(\xi,\xi_0)/4\tau}}{\tau^{d/2}} \sum_r a_r(\xi,\xi_0)\,\tau^r, \tag{84}$$

where $d_g(\xi,\xi_0)$ denotes the geodesic distance between the points labeled ξ and ξ_0. As in flat space we have

$$\left\langle d_g^2(\tau)\right\rangle_\xi \equiv \frac{1}{V}\int d^2\xi\sqrt{g}\int d^2\xi'\sqrt{g}\; d_g^2(\xi,\xi')K_g(\xi,\xi';\tau) \sim \tau + 0(\tau^2). \tag{85}$$

$\left\langle d_g^2(\tau)\right\rangle_\xi$ as well as the the average of the return probability

$$rp(\tau) = \langle K_g(\tau)\rangle_\xi\,, \tag{86}$$

are clearly reparameterization invariant, and it makes sense to talk about the quantum average of such "observables". In the following we consider a fixed volume V and denote the functional average over geometries (i.e. over equivalence classes of metrics) by $\langle\,.\,\rangle_V$.

The calculation of d_h in Liouville theory is so far based on the assumption that for a given rescaling of the volume $V \to \lambda V$ of the universe there exists a rescaling $\tau \to \tau' = \tau'(\lambda)$ such that

$$\left\langle\, \langle K_g(\tau'(\lambda))\rangle_\xi \,\right\rangle_{\lambda V} = \left\langle\, \langle K_g(\tau)\rangle_\xi \,\right\rangle_V. \tag{87}$$

If one uses the representation (82) of K_g and performs a small τ expansion:

$$\left\langle \left\langle \left[(1+\tau\Delta_g + 0(\tau^2))\frac{1}{\sqrt{g}}\delta(\xi-\xi')\right]_{\xi'=\xi}\right\rangle_\xi \right\rangle_V, \tag{88}$$

it is seen that the first, trivial term satisfies (87). However, in order that the second term will satisfy (87) one is, due to the anomalous scaling of Δ_g in Liouville theory†† forced to the following λ-dependence of τ':

$$\tau' = \lambda^{-\alpha_{-1}/\alpha_1}\tau, \qquad \text{i.e.} \quad \dim[\tau] = \dim[A^{-\alpha_{-1}/\alpha_1}]. \tag{90}$$

Since we also expect that

$$\dim\left[\left\langle \left\langle d^2(\tau)\right\rangle_\xi \right\rangle_V\right] = \dim[A^{2/d_h}], \tag{91}$$

one finally obtains (78) from (85).

** We present here a slightly simplified version of the precise asymptotic expansion.

†† To be more specific the calculation proceed as follows: The Liouville field ø is introduced by the partial gauge fixing $g_{\alpha\beta} = e^{\varnothing}\hat{g}_{\alpha\beta}$, where $\hat{g}$ is a background metric. Field-operators with specific scaling properties when the volume $V \to \lambda V$ will pick up anomalous scaling in Liouville gravity. If the classical scaling is expected to be $\langle\Phi_n\rangle_{\lambda V} = \lambda^n\langle\Phi_n\rangle_V$, the quantum Liouville scaling will be

$$\langle\Phi_n\rangle_{\lambda V} = \lambda^{\alpha_n/\alpha_1}\langle\Phi_n\rangle_V\,, \qquad \alpha_n = \frac{2n\sqrt{25-c}}{\sqrt{25-c}+\sqrt{25-c-24n}}, \tag{89}$$

where c is the central charge of the conformal field theory coupled to two-dimensional quantum gravity.

Spin Boundaries

The derivation of the alternative formula (79) for the fractal dimension $\tilde{d}_h$ is best explained by considering the Ising model coupled to gravity in the context of dynamical triangulations. The model has a critical point as a function of the Ising coupling. Away from the critical point the geometric aspects of the combined theory coincide with those of pure gravity, but at the critical point, where the theory is believed to describe a conformal field theory of central charge $c = 1/2$ coupled to quantum gravity, the string susceptibility jumps from $-1/2$ (the value for pure gravity) to $-1/3$, the value for a $c = 1/2$ theory coupled to gravity according to the KPZ formula (80). At the same time the values of critical exponents related to magnetic properties of the spin system differ from the flat space values of the exponents, showing that gravity influences the critical properties of matter (the exact formulas being those of KPZ).

Ideally, one would like to follow the combinatorial approach for pure gravity in the determination of $\tilde{d}_h$. It has not yet been possible. The two-loop function defined above would involve an average over various spin configurations at the boundaries and it is not known how to perform such average analytically. However, it is possible to calculated the disk-amplitude (the Hartle-Hawking wave-functional) where all spins are aligned, as first noticed in [13]. In [14, 15] this was generalized to the calculation of a two-loop function somewhat similar to two-loop function $G_\mu(l_1, l_2; r)$ considered above for pure gravity. One difference is that the boundary loops l_1 and l_2 have aligned spins (which need not be the same for different boundaries). Another important difference is that the geodesic distance r is replaced by a "time" parameter τ defined by deformation along spin boundaries as will now be explained.

Let the spin be located at the vertices of the triangulation. A given spin configuration can be decomposed into spin clusters where spin either point up or down. The spin boundaries can be viewed as closed loops passing through the centers of the triangles and crossing the links (i.e. living on the lattice dual to the triangulation). The loop gas expansion of the Ising model (or more generally, the loop gas expansion of the $O(n)$ model on a lattice) is precisely an expansion in terms of such boundaries. For a given spin configuration the triangles either have no links or two links which belong to a spin boundary. We denote triangles with no links belonging to a spin boundary as type I and the other triangles as type II. Let us now define a modified geodesic distance in the following way. In the case of pure gravity the "peeling" decomposition along the boundary is defined in fig. 6. In the present case we proceed in the same way if we meet a triangle of type I. If we meet a triangle of type II it will be part of a closed loop of triangles, all of type II. We define all these triangles to have the *same* distance to the link from with we perform the deformation, and one step out will include all the triangles in the loop. One can view the step as if we "lasso" the loop and in this way create two new boundaries, each with a the same distance to the original link. The spins at the two boundaries will be opposite. The procedure is illustrated in fig. 7.

Following the above described procedure one can systematically proceed outwards from a given triangle where the three spins are aligned and define a "distance" or "time" τ to other triangles, or the distance from the initial boundary loop to the boundary created by the "peeling" (see fig. 5). It is of course not a genuine distance, since it does not satisfy the requirement that two triangles with zero distance are identical. However, it may serve as some approximate measure of distance which in principle could be proportional to the geodesic distance in the scaling limit after an average over spin configurations as well as over geometries. Below we will discuss the relation between geodesic distance r and the "time" τ.

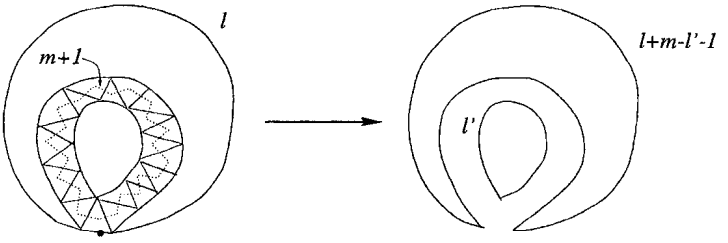


Figure 7. The deformation corresponding to fig. 5(a), but in the situation when the triangle is of type II.

Let us now by a heuristic argument show that one can expect

$$\dim[\tau] = \dim[V^{1/6}]. \tag{92}$$

The consequence of this dimension-assignment is that the fractal dimension of two-dimensional quantum gravity coupled to the critical Ising model is

$$\tilde{d}_h = 6, \tag{93}$$

provided τ is proportional to the geodesic distance.

In order to understand the relation (92) we first consider flat space, i.e. a triangular regular lattice (for instance with toroidal topology, such that each vertex can be of order 6). In this case the definitions given above for the "time" τ still apply. Recall the following facts from finite size scaling of the Ising model (or spin systems in general): if β_c denotes the critical point of the Ising model at infinite volume, there exists a "pseudo-critical point" $\beta_*(V) > \beta_c$ for a finite volume V, such that the system at β_c has an effective magnetization m per volume:

$$m \sim L^{-\tilde{\beta}/\nu} \sim V^{-\tilde{\beta}/\nu d}, \tag{94}$$

where we have used the notation $\tilde{\beta}$ for the critical exponent of the magnetization and L denotes the linear extension of the spin system. The total magnetization at β_c will be

$$M = mV \sim V^{1-\tilde{\beta}/\nu d}. \tag{95}$$

In 2d the magnetization at the critical point is determined by the largest cluster of spin. The clusters of spins are described by percolation theory and if p denotes probability that a site belongs to a largest cluster, it is known that

$$p \sim m, \qquad pV \sim M. \tag{96}$$

The largest spin clusters at the critical point will be fractal. In peculation theory one defines the fractal dimension D of a cluster by

$$pL^d = L^D, \qquad \text{i.e.} \quad pV = V^{D/d}. \tag{97}$$

We conclude that D is related to $\tilde{\beta}$ by

$$\frac{D}{d} = 1 - \frac{\tilde{\beta}}{\nu d} \quad \left(= \frac{15}{16} \quad \text{for the Ising model}\right). \tag{98}$$

Assume now that we use the definition of "time" or "geodesic distance" in terms of spin boundaries and step outwards from a single triangle boundary. Since it is known that the fractal dimension of the boundary of a large spin cluster is identical to the fractal dimension of the cluster itself we can write

$$V \sim V^{D/d} \times \tau \quad i.e. \quad \dim[\tau] = \frac{\tilde{\beta}}{\nu d} \quad (= \frac{1}{16} \text{ for the Ising model}). \tag{99}$$

This calculation can be taken over to the Ising model coupled to gravity since the change in $\tilde{\beta}$ and the change in νd after coupling to 2d quantum gravity can be calculated by KPZ:

$$\tilde{\beta} = \frac{1}{8} \to \frac{1}{2}, \qquad \nu d = 2 \to \nu d_h = 3. \tag{100}$$

Consequently we have for the Ising model coupled to 2d quantum gravity at the critical point:

$$\dim[\tau] = \frac{\tilde{\beta}}{\nu d} = \frac{1}{6}. \tag{101}$$

We conclude that the fractal dimension of 2d quantum space-time coupled to $c = 1/2$ conformal matter is 6, again *provided* τ can be identified with the geodesic distance.

The above argument can be generalized to any (p, q) conformal field theory coupled to gravity ([14, 16]) with the result that $\tilde{d}_h = -2/\gamma(c)$, where $\gamma(c)$ is the string susceptibility given by KPZ.

A Test for c=-2

As emphasized above the interpretation of $\tilde{d}_h(c) = -2/\gamma(c)$ relies on the identification of τ with the geodesic distance. The hypothesis can be tested numerically with high accuracy for $c = -2$. There are several reasons that $c = -2$ is an ideal test model. First it is possible to perform very good numerical simulations since one for $c = -2$ does not have to rely on Monte Carlo simulations, but can use a recursive algorithm for sampling configurations [17]. It implies that one can get much better statistics and that one can use much large triangulations. Further, the fact that $\tilde{d}_h(c) = 2$ is predicted for $c = -2$ means that there should be no finite size effects which invalidate the numerical results. If $\tilde{d}_h$ had been very large, as is the case for c close to 1 (for instance $\tilde{d}_h = 10$ for the three-states Potts model coupled to gravity), it would have been difficult to fulfill the criterion $1 << \tau << N^{1/\tilde{d}_h}$ for the number of triangles N available in computer simulations. Finally $c = -2$ belongs to the range of conformal field theories which are well described of $O(n)$ models coupled to gravity via the loop expansion. For these models the equations for two-point functions have been derived in detail starting from the discretized models ([18]). In particular, the two-point function for $c = -2$ can be found explicitly and one finds $\tilde{d}_h = 2$ for $c = -2$, again provided $\tau \sim r$ in average, where r denotes the geodesic distance [18].

The result of the numerical test leaves no doubt. Using the standard geodesic distances in such simulations (either triangle distance or link distance) one finds perfect agreement with formula (78), while the data are incompatible with (79). The prediction of (78) is

$$d_h(c = -2) = 3.56..., \tag{102}$$

and from the discussion of the two-point function for pure gravity one expects that a measurement of the average "volume" (in this case : average length) of spherical shells of geodesic radius r, measured on triangulations consisting of N triangles, will scale as

$$S_N(r) \sim N^{1-1/d_h} x^{d_h-1} F(x), \qquad x = \frac{r}{N^{1/d_h}}. \tag{103}$$

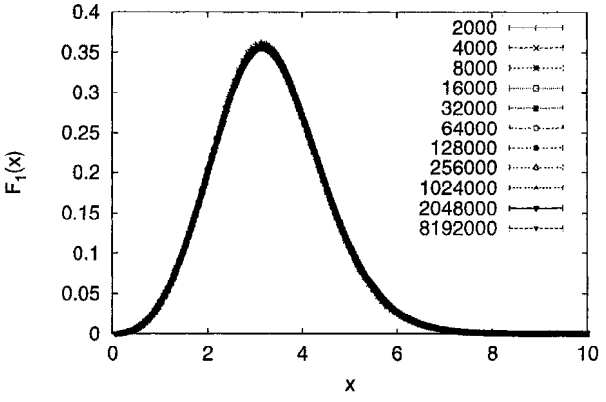


Figure 8. The scaling of the "volume" of spherical shells for a $c = -2$ theory coupled to gravity, according to formula (103), only with the parameter $x = (r + a)/N^{1/d_h}$, where a is a so-called shift parameter (see citemany for a detailed discussion.

I fig. 8 we have shown a verification of this scaling law for $c = -2$ and d_h given by (102) for N in the range from 2000 to 8,000.000 triangles. A detailed account of the best numerical determination of $d_h(c = -2)$ can be found in [19]. The result is $d_h = 3.58 \pm 0.03$, in perfect agreement with (102).

It is natural to conjecture from the outcome of the numerical simulations that

$$\dim[\tau] \neq \dim[r] \quad \text{for} \quad c \neq 0. \tag{104}$$

Unfortunately it is premature to conclude that (78) for all c. The numerical simulations for $c > 0$, i.e. for the Ising model ($c = 1/2$) and the three-states Potts model ($c = 4/5$), have not confirmed formula (78) in a convincing way. The results seem rather to indicate that $d_h = 4$ for $0 \leq c \leq 1$. First it should be mentioned that the quality of the numerical simulations do not match those of $c = -2$ since one has to use Monte Carlo simulations for $c \neq -2$. Next, although it seems strange that formula (78) should be valid for $c \leq 0$ and not valid for $c > 0$, one should be aware that precisely for $c < 0$ the cosmological term will not correspond to the most infrared dominant operator in the theory. This difference between $c < 0$ and $c > 0$ might be important for the validity of (78), but the details of such an mechanism are not yet understood.

Summary

Of two candidates for a fractal dimension of space-time, (78) and (79), it seems that (78) is correct at least for $c < 0$, while (79) never describes the correct fractal dimension of space-time when $c \neq 0$. It does not imply that the scaling implicit in (79) is not correct. Such a scaling indeed exists, as has independently verified recently in the context of the $O(n)$ model [18]. However, it has no relation the concept of geodesic distance and the fractal structure introduced via geometry. The analysis of the Ising model on a flat lattice highlighted such a scenario since the geodesic distance of course would result in $d_h = 2$ while the "distance" defined via spin boundaries resulted in $\tilde{d}_h = 16$! It is an unsolved and interesting question to what extend τ might anyway

serve as a kind of proper time in 2d quantum gravity. the question is also important since a string field theory of gravity coupled to matter can be formulated in relatively simple way using τ as evolution parameter.

DISCUSSION

I have presented our present understanding of some aspects of the fractal structure of two-dimensional quantum gravity coupled to matter. A few aspects of the fractal structure have not been touched upon in this review. In particular the so-called spectral dimension, which is still another, seeming independent, measure of fractal properties of space-time. The spectral dimension is closely related to the heat kernel for diffusion, which was discussed in connection with diffusion in Liouville theory. On a fixed manifold one has the asymptotic expansion

$$K_g(\xi, \xi_0, \tau) \sim \frac{e^{-d_g^2(\xi,\xi_0)/4\tau}}{\tau^{d/2}} \sum_r a_r(\xi, \xi_0)\, \tau^r, \tag{105}$$

of the heat kernel. The average return probability at "diffusion time τ" was defined as

$$rp(\tau) = \frac{1}{V} \int d^2\xi \sqrt{g}\; K_g(\xi, \xi; \tau) = \frac{1}{V} \mathrm{Tr}\, \hat{K}(\tau). \tag{106}$$

It follows that $rp(\tau)$ has the asymptotic expansion

$$rp(\tau) \sim \frac{1}{\tau^{d/2}} \sum_r A_r \tau^r, \tag{107}$$

where the coefficients A_r have a simple geometric representation, well-known from the theory of the heat kernel. One can define the diffusion on more general structures than manifolds. Under very general assumptions the associated heat kernel has an asymptotic expansion and the return probability can be written as in eq. (107), only with *the dimension d replaced by the so-called spectral dimension* d_s. As already discussed it makes perfect sense to take the functional average over geometries. After the functional average we *might* still have an asymptotic expansion like (107). Also in that case we denote the d appearing in the expansion the spectral dimension d_s, and it might be different from the dimension of the manifolds underlying the geometry.

Little is known about d_s from analytic calculation. It has been proven that $d_s = 2$ for $c = -2$ coupled to gravity [20], but even in the case of pure gravity there is presently no analytic calculation of d_s. However, if we turn to numerical "experiments" it seems that $d_s = 2$ for the central charge $c \in [-2, 1]$. From the point of view of diffusion, it seems that quantum gravity coupled to conformal matter with central charge c in this interval has the same spectral dimension as flat space-time. It would be very interesting to have an analytical proof of this fact. It is further interesting to note that when the central charge $c > 1$ it seems (again from computer simulations [12]) as if the individual triangulations have $d_s < 2$. One interpretation of this is that the matter interacts so strongly with the geometry that each individual manifold is teared apart and does no longer classify as a two-dimensional manifold.

Higher Dimensions

It is presently an open question how to generalize these results to higher dimensional geometries. In particular, our space-time world seems to be four-dimensional.

What is the genuine fractal dimension in the class of all four-dimensional geometries of fixed topology? Numerical simulations seem to indicate that *the typical four-dimensional spherical geometry has infinite intrinsic Hausdorff dimension.* This might well change when we take into account the Einstein-Hilbert action. Presently the numerical simulations indicate that we have a two-phase structure as a function of the bare gravitational coupling constant, the two phases being separated by a first order phase transition. For large bare gravitational coupling constant we are in the phase with infinite d_h (the limit of infinite bare gravitational coupling constant corresponds to not including the Einstein-Hilbert term in the action at all, i.e. the situation mentioned above), while a small gravitational coupling constant seemingly results in geometric structures which are more like branched polymers. It is presently not known how to extract interesting physics which resembles our four-dimensional wold from this scenario, in particular because the phase transition separating the two phases seems to be a first order transition. However, it might be possible to add new terms to the action and change this situation. This aspect is presently being investigated.

Another interesting question is whether the spectral dimension of such (discretized) manifolds is still four, or whether they are some kind of degenerate manifolds which cannot be viewed as representing genuine four-dimensional manifolds. Even if we have presently no well defined theory of four dimensional quantum gravity, questions like these can still be asked and they will clearly be important for the way we view functional integration over four-dimensional geometries.

REFERENCES

1. J. Ambjørn, B. Durhuus and T. Jonsson, *Quantum Geometry,* Cambridge Monographs on Mathematical Physics, Cambridge University Press, 1997.
2. F. David, Nucl.Phys. B257 (1985) 45; *Nucl. Phys.* B257 (1985) 543;
3. J. Ambjørn, B. Durhuus and J. Fröhlich, *Nucl. Phys.* B257 (1985) 433;
4. V.A. Kazakov, I.K. Kostov and A.A. Migdal, *Phys. Lett.* 157B (1985) 295.
5. J. Ambjørn, B. Durhuus J. Fröhlich and P. Orland, *Nucl. Phys.* B270 (1986) 457;
6. W.T. Tutte, *Can. J. Math.* 14 (1962) 21.
7. F. David, *Mod. Phys. Lett.* A5 (1990) 1019.
8. J. Ambjørn, J. Jurkiewicz and Y. M. Makeenko, *Phys. Lett.* B251 (1990) 517.
9. Y. Watabiki, *Nucl. Phys.* B441 (1995) 119
10. H. Kawai, N. Kawamoto, T. Mogami and Y. Watabiki, *Phys. Lett.* B306, (1993) 19.
11. J. Ambjørn and Y. Watabiki, *Nucl. Phys.* B445 (1995) 129.
12. J. Ambjørn, J. Jurkiewicz and Y. Watabiki, *Nucl. Phys.* B454 (1995) 313.
13. E. Gava and K.S. Narain (ICTP, Trieste), Phys.Lett. B263 (1991) 213.
14. N. Ishibashi and H. Kawai, Phys.Lett. B322 (1994) 67. hep-th/9312047.
15. R. Nakayama and T. Suzuki, Phys.Rev. D54 (1996) 3985. hep-th/9604073; Phys.Lett.B354 (1995) 69. hep-th/9503190.
16. J. Ambjørn and Y. Watabiki, *Noncritical string field theory for two-D quantum gravity coupled to (p, q) conformal fields.* hep-th/9604067.
17. N. Kawamoto, V. Kazakov, Y. Saeki and Y. Watabiki, Phys.Rev.Lett. 68 (1992) 2113.
18. J. ambjørn, C.F. Kristjansen and Y. Watabiki, *The Two point function of c = -2 matter coupled to 2-D quantum gravity.* hep-th/9705202.
19. J. Ambjørn, K. Anagnostopoulos, T. Ichihara, L. Jensen, N. Kawamoto, Y. Watabiki, K. Yotsuji., *The quantum space-time of c = –2 gravity.,* hep-lat/9706009; Phys.Lett. B397 (1997) 177. hep-lat/9611032.
20. D. Boulatov, Private communication.

This page intentionally left blank

SCALING LAWS IN TURBULENCE

Krzysztof Gawędzki

C.N.R.S., I.H.E.S.
91440 Bures-sur-Yvette, France

INTRODUCTION

We discuss the largely open problem of scaling laws in fully developed turbulence, stressing the similarities and the differences with scaling in field theory. A soluble model of the passive advection is examined in more detail in order to illustrate the principal ideas.

LANGEVIN VERSUS NAVIER-STOKES

Many dynamical problems in physics may be described by evolution equations of the type

$$\partial_t \Phi = -F(\Phi) + f \tag{1}$$

where $\Phi(t, \mathbf{x})$ represents local densities of physical quantities, $F(\Phi)$ is their nonlinear functional and f stands for an external source. We shall be interested in the situations where the source f is random. For concreteness, we shall assume it Gaussian with mean zero and covariance

$$\langle f(t,\mathbf{x}) f(s,\mathbf{y}) \rangle = \delta(t-s) C(\tfrac{\mathbf{x}-\mathbf{y}}{L}) \tag{2}$$

with L determining the scale on which $f(t, \mathbf{x})$ are correlated. An example of such an evolution law is provided by the Langevin equation describing the approach to equilibrium in systems of statistical mechanics or field theory[1]. In this case the nonlinearity is of the gradient type:

$$F(\Phi) = \frac{\delta S(\Phi)}{\delta \Phi} \tag{3}$$

with $S(\Phi)$ a local functional, e.g. $S(\Phi) = \frac{1}{2}\int(\nabla\Phi)^2 + \frac{1}{2}m^2\int\Phi^2 + \lambda\int\Phi^4$ in the Φ^4 field theory, and with L small so that $C(\mathbf{x}/L)$ is close to the delta-function $\delta(\mathbf{x})$ and regulates the theory on short distances $\lesssim L$.

New Developments in Quantum Field Theory
Edited by Damgaard and Jurkiewicz, Plenum Press, New York, 1998

Augmented by an initial condition $\Phi(t_0, \mathbf{x})$ (and, eventually, boundary conditions), the solution of Eq. (1) should define random field $\Phi(t, \mathbf{x})$. We are interested in the behavior of its correlation functions given by the mean values

$$\langle \prod_i \Phi(t_i, \mathbf{x}_i) \rangle . \tag{4}$$

Among the basic questions one may ask are the following ones:

1. Do the correlation functions become stationary (i.e. dependent only on time differences) when $t_0 \to -\infty$? If so, are the stationary correlators unique (independent of the initial condition)?

2. Do they obey scaling laws?

For the field theory case these questions are well studied with the use of powerful analytic tools as perturbative expansions and renormalization group and by numerical analysis (Monte Carlo simulations). The stationary correlators describe possibly different phases of the system. Universal (i.e. independent of the cutoff C) scaling laws of the type

$$\langle Q(\Phi(t, \mathbf{x})) Q(\Phi(t, \mathbf{y})) \rangle \sim |\mathbf{x} - \mathbf{y}|^{-\zeta_Q} \tag{5}$$

for $|\mathbf{x} - \mathbf{y}| \gg L$ with Q some local functions of Φ emerge at the points of the 2nd order phase transitions.

On the opposite pole of the field theoretic case are the hydrodynamical examples of the evolution equation (1). The best known of those is the Navier-Stokes equation

$$\partial_t \mathbf{v} = -P(\mathbf{v} \cdot \nabla)\mathbf{v} + \nu \Delta \mathbf{v} + \mathbf{f} \tag{6}$$

for the incompressible ($\nabla \cdot \mathbf{v} = 0$) velocity field $\mathbf{v}(t, \mathbf{x})$, with P standing for the orthogonal projection on such vector fields. ν denotes the viscosity and $\mathbf{f}$ is the external force which induces the fluid motion. In the fully developed turbulence one is interested in the regime where the stirring forces act on large distances (like the convective forces on scales of kilometers in the atmosphere) and we observe quite complicated (turbulent) motions on shorter distances down to scales on which the the dissipative term $\propto \nu$ becomes important ($\sim$ milimeters in the atmosphere). It is believed that the large scale details should not be essential for the statistics of the flow in this intermediate regime called the "inertial range". It is therefore common to model the stirring forces by a random Gaussian process with mean zero and covariance

$$\langle f^\alpha(t, \mathbf{x}) f^\beta(s, \mathbf{y}) \rangle = \delta(t - s) \mathcal{C}^{\alpha\beta}(\tfrac{\mathbf{x}-\mathbf{y}}{L}) \tag{7}$$

with $\partial_\alpha \mathcal{C}^{\alpha\beta} = 0$. L denotes now the large "integral scale" on which the random forces act. Note that, unlike for field theory, in this case the covariance $C(\mathbf{x}/L)$ is close to a constant, i.e. to a delta-function in the wavenumber space and not in the position space. Such regime in field theory would correspond to distances shorter than the ultraviolet cutoff with the behavior strongly dependent on the detailed form of the regularization. Another (related) difference is that in Eq. (6) the nonlinear term is not of a gradient type. Finally, the projection P renders it non-local which is another complication. All these differences make the Navier-Stokes problem (6) quite different from that posed by the Langevin equation and resistant to the methods employed successfully in the study of the latter.

KOLMOGOROV THEORY

The first major attempt to obtain universal scaling laws for the inertial range correlators is due to Kolmogorov[2]. Assuming the existence of homogeneous (i.e. translationally invariant) stationary correlators of velocities, one deduces the following relation at equal times

$$\tfrac{1}{4}\nabla_{\mathbf{x}}\cdot\langle(\mathbf{v}(\mathbf{x})-\mathbf{v}(\mathbf{y}))^2(\mathbf{v}(\mathbf{x})-\mathbf{v}(\mathbf{y}))\rangle+\tfrac{1}{2}\operatorname{tr}\mathcal{C}(\tfrac{\mathbf{x}-\mathbf{y}}{L})=\nu\langle\nabla\mathbf{v}(\mathbf{x})\cdot\nabla\mathbf{v}(\mathbf{y})\rangle. \tag{8}$$

Eq. (8) is obtained the following way. First, using Eq. (6), we write

$$\mathbf{v}(t+\delta t,\mathbf{x})=\mathbf{v}(t,\mathbf{x})+[-P(\mathbf{v}\cdot\nabla)\mathbf{v}(t,\mathbf{x})+\nu\Delta\mathbf{v}(t,\mathbf{x})]\delta t+\int_t^{t+\delta t}\mathbf{f}(s)\,ds+\mathcal{O}(\delta t^2) \tag{9}$$

and we instert this expansion into the equal-time 2-point function obtaining

$$\begin{aligned}&\langle\mathbf{v}(t+\delta t,\mathbf{x})\,\mathbf{v}(t+\delta t,\mathbf{y})\rangle=\langle\mathbf{v}(t,\mathbf{x})\,\mathbf{v}(t,\mathbf{y})\rangle\\&-\langle P(\mathbf{v}\cdot\nabla)\mathbf{v}(t,\mathbf{x})\cdot\mathbf{v}(t,\mathbf{y})\rangle\delta t-\langle\mathbf{v}(t,\mathbf{x})\cdot P(\mathbf{v}\cdot\nabla)\mathbf{v}(t,\mathbf{y}\rangle\delta t\\&+\nu\langle\Delta\mathbf{v}(t,\mathbf{x})\cdot\mathbf{v}(t,\mathbf{y})\rangle\delta t+\nu\langle\mathbf{v}(t,\mathbf{x})\cdot\Delta\mathbf{v}(t,\mathbf{y})\rangle\delta t+\operatorname{tr}\mathcal{C}(\tfrac{\mathbf{x}-\mathbf{y}}{L})\delta t+\mathcal{O}(\delta t^2).\end{aligned} \tag{10}$$

Vanishing of the $\mathcal{O}(\delta t)$ terms produces Eq. (8).

Under the limit $\mathbf{y}\to\mathbf{x}$ for positive ν, Eq. (8) becomes the identity

$$\tfrac{1}{2}\operatorname{tr}\mathcal{C}(0)\;=\;\langle\nu(\nabla\mathbf{v})^2\rangle\equiv\bar{\epsilon} \tag{11}$$

which expresses the energy balance: in the stationary state, the mean energy injection rate is equal to the mean rate of energy dissipation $\bar{\epsilon}$. On the other hand, performing the limit $\nu\to 0$ for $\mathbf{x}\neq\mathbf{y}$ one obtains

$$-\tfrac{1}{4}\nabla_{\mathbf{x}}\cdot\langle(\mathbf{v}(\mathbf{x})-\mathbf{v}(\mathbf{y}))^2(\mathbf{v}(\mathbf{x})-\mathbf{v}(\mathbf{y}))\rangle\;=\;\tfrac{1}{2}\operatorname{tr}\mathcal{C}(\tfrac{\mathbf{x}-\mathbf{y}}{L}). \tag{12}$$

For $|\mathbf{x}-\mathbf{y}|\ll L$ the right hand side is approximately constant and equal to $\frac{1}{2}\operatorname{tr}C(0)$, i.e. to $\bar{\epsilon}$. Assuming also isotropy (rotational invariance) of the stationary state one may then infer the form of the 3-point function in the inertial range:

$$\langle(v^\alpha(\mathbf{x})-v^\alpha(0))(v^\beta(\mathbf{x})-v^\beta(0))(v^\gamma(\mathbf{x})-v^\gamma(0))\rangle\;\cong\;-\frac{4\bar{\epsilon}}{d(d+2)}(\delta^{\alpha\beta}x^\gamma+\text{cycl.}). \tag{13}$$

The latter implies for the 3-point function of the component of $(\mathbf{v}(\mathbf{x})-\mathbf{v}(0))$ paralel to $\mathbf{x}$ the relation

$$\langle(\mathbf{v}(x)-\mathbf{v}(0))_{\|}^3\rangle\;\cong\;-\frac{12}{d(d+2)}\,\bar{\epsilon}\,|\mathbf{x}| \tag{14}$$

known as the Kolmogorov "$\frac{4}{5}$ law" ($\frac{4}{5}$ is the value of the coeffitient on the right hand side in 3 dimensions).

Under natural assumptions about the stationary state one may deduce a stronger version (sometimes called refined similarity) of the above relation which takes the form of the operator product expansion for the $\nu\to 0$ limit of the dissipation operator $\epsilon=\nu(\nabla\mathbf{v})^2$:

$$\epsilon(x)\;=\;-\tfrac{1}{4}\lim_{\mathbf{y}\to\mathbf{x}}\nabla_x\cdot[(\mathbf{v}(\mathbf{x})-\mathbf{v}(\mathbf{y}))^2(\mathbf{v}(\mathbf{x})-\mathbf{v}(\mathbf{y}))]\Big|_{\nu=0} \tag{15}$$

holding inside the expectations in the $\nu \to 0$ limit. Relation (15) expresses the dissipative anomaly: the dissipation ε whose definition involves a factor of ν does not vanish when $\nu \to 0$.

Kolmogorov postulated[2] that the scaling of general velocity correlators in the inertial range should be determined by universal relations involving only the distances and the mean dissipation rate $\bar{\epsilon}$. Such postulate leads to the scaling laws

$$\langle (\mathbf{v}(\mathbf{x}) - \mathbf{v}(0))_{||}^{n} \rangle \propto \bar{\epsilon}^{n/3} |\mathbf{x}|^{n/3} \tag{16}$$

for the n-point "structure functions" of velocity generalizing the (essentially rigorous) result (14) about the 3-point function. The right hand side of relation (16) is the only expression built from $\bar{\epsilon}$ and $|\mathbf{x}|$ with the right dimension.

The power law fits $\propto |\mathbf{x}|^{\zeta_n}$ for the structure functions measured in experiments and in numerical simulations lead to the values of the exponents slightly different from the Kolmogorov prediction $\zeta_n = n/3$ for $n \neq 3$. One obtains[3] $\zeta_2 \cong .70$, $\zeta_4 \cong 1.28$, $\zeta_6 \cong 1.77$, $\zeta_8 \cong 2.23$. The discrepencies indicate that the random variables $(\mathbf{v}(\mathbf{x}) - \mathbf{v}(0))_{||}$ are non-Gaussian for small $\mathbf{x}$ with the probability distribution functions decaying slower than in the normal distribution. Such a slow decay signals the phenomenon of frequent occurence of large deviations from the mean values called "intermittency". There exist many phenomenological models of intermittency of the inertial range velocity differences based on the idea that the turbulent activity is carried by a fraction of degrees of freedom with a self-similar ("multi-fractal") structure[4]. An explanation of the mechanism behind the observed intermittency starting from the first principles (i.e. from the Navier-Stokes equation) is, however, still missing and constitutes the main open fundamental problem of the fully developed turbulence.

KRAICHNAN MODEL OF PASSIVE ADVECTION

Recently some progress has been achieved[5,6,7] in understanding the origin of intermittency in a simple model[8,9] describing advection of a scalar quantity (temperature $T(t, \mathbf{x})$) by a random velocity field $\mathbf{v}(t, \mathbf{x})$. The evolution of the temperature is described by the equation

$$\partial_t T = -(\mathbf{v} \cdot \nabla) T + \kappa \Delta T + f \tag{17}$$

where κ denotes the molecular diffusivity and f is the external source which we shall take random Gaussian with mean zero and covariance (2). Following Kraichnan[8], we shall assume that $\mathbf{v}(t, \mathbf{x})$ is also a Gaussian process, independent of f, with mean zero and covariance

$$\langle v^{\alpha}(t, \mathbf{x}) v^{\beta}(s, \mathbf{y}) \rangle = \delta(t - s) \left(D_0 \delta^{\alpha\beta} - d^{\alpha\beta}(\mathbf{x} - \mathbf{y})\right) \tag{18}$$

with D_0 a constant, $d^{\alpha\beta}(\mathbf{x}) \propto |\mathbf{x}|^{\xi}$ for small $|\mathbf{x}|$ and with $\partial_\alpha d^{\alpha\beta} = 0$ in order to assure the incompressibility. Note the scaling of the 2n-point function of velocity differences with power $n\xi$ of the distance. The Komogorov scaling corresponds to $\xi = \frac{4}{3}$ (the temporal delta-function appears to have dimension $length^{\xi-1}$). The time decorrelation of the velocities is not, however, a very realistic assumption. In the Kraichnan model ξ is treated as a parameter running from 0 to 2.

Writing

$$T(t + \delta t, \mathbf{x}) = T(t, \mathbf{x}) - \int_{t}^{t+\delta t} [(\mathbf{v} \cdot \nabla) T(s, \mathbf{x}) - f(s, \mathbf{x})]\, ds + \kappa \Delta T(t, \mathbf{x}) \delta t + \mathcal{O}(\delta t^2), \tag{19}$$

we obtain the analogue of the relation (8) for the stationary state of the scalar (the latter may be shown to exist and to be independent of the initial condition decaying at spatial infinity):

$$-\tfrac{1}{2} d^{\alpha\beta}(\mathbf{x}-\mathbf{y})\, \partial_{x^\alpha}\partial_{y^\beta} \langle T(\mathbf{x})T(\mathbf{y})\rangle + \tfrac{1}{2}\mathcal{C}(\tfrac{\mathbf{x}-\mathbf{y}}{L}) = \kappa \langle \nabla T(\mathbf{x}) \cdot \nabla T(\mathbf{y})\rangle . \tag{20}$$

Letting in Eq. (20) $\mathbf{y} \to \mathbf{x}$ for $\kappa > 0$ produces, as before, the energy balance:

$$\tfrac{1}{2}\mathcal{C}(0) = \langle \kappa (\nabla T)^2 \rangle \equiv \bar{\epsilon}. \tag{21}$$

On the other hand, taking $\kappa \to 0$ for $\mathbf{x} \neq \mathbf{y}$ results in the equation analoguous to (12):

$$\tfrac{1}{2} d^{\alpha\beta}(\mathbf{x}-\mathbf{y})\, \partial_{x^\alpha}\partial_{y^\beta} \langle T(\mathbf{x})T(\mathbf{y})\rangle = \tfrac{1}{2}\mathcal{C}(\tfrac{\mathbf{x}-\mathbf{y}}{L}) \tag{22}$$

which may be easily solved exactly for the 2-point function of the scalar giving

$$\langle T(\mathbf{x})T(0)\rangle = A_{\mathcal{C}}\, L^{2-\xi} - \text{const.}\, \bar{\epsilon}\, |\mathbf{x}|^{2-\xi} + \mathcal{O}(L^{-2}) \tag{23}$$

or

$$\langle (T(\mathbf{x}) - T(0))^2 \rangle \cong \text{const.}\, \bar{\epsilon}\, |\mathbf{x}|^{2-\xi} \tag{24}$$

for $|\mathbf{x}| \ll L$. This is an analogue of the Kolmogorov $\frac{4}{5}$ law. It may be strengthen to the operator product expansion for the dissipation operator[10,15] $\epsilon = \kappa (\nabla T)^2$

$$\epsilon(x) = \tfrac{1}{2} \lim_{\mathbf{y}\to\mathbf{x}} d^{\alpha\beta}(\mathbf{x}-\mathbf{y})\, \partial_\alpha T(\mathbf{x})\, \partial_\beta T(\mathbf{y})\Big|_{\kappa=0} \tag{25}$$

valid inside the expectations in the limit $\kappa \to 0$. Eq. (25) expresses the dissipative anomaly in the Kraichnan model, analogous to the dissipative anomaly (15) for the Navier-Stokes case.

The natural question arises whether the higher structure functions of the scalar $\langle (T(\mathbf{x}) - T(0))^{2n} \rangle \equiv S_{2n}(\mathbf{x})$ scale with powers $n(2-\xi)$ as the dimensional analysis would suggest (Corrsin's analogue[11] of the Kolmogorov theory). The answer is no. Experiments show that the scalar differences display higher intermittency than that of the velocities[12]. Although, by assumption, in the Kraichnan model there is no intermittency in the distribution of the velocity differences, numerical studies[13, 14] indicate strong intermittency of the scalar differences signaled by anomalous (i.e. $\neq n(2-\xi)$) scaling exponents. Unlike in the Navier-Stokes case, we have now some analytic understanding of this phenomenon, although still incomplete and controvertial[9,14,16].

The simplifying feature of the Kraichnan model is that the insertion of expansion (19) into the higher point functions $\mathcal{F}_{2n}(\underline{\mathbf{x}}) \equiv \langle T(t,\mathbf{x}_1), \ldots, T(t,\mathbf{x}_{2n})\rangle$ leads to a system of (Hopf) equations which close:

$$\mathcal{M}_{2n}\, \mathcal{F}_{2n}(\underline{\mathbf{x}}) = \sum_{i<j} \mathcal{F}_{2n-2}(\mathbf{x}_1, \ldots, \hat{i}, \ldots, \hat{j}, \ldots, \mathbf{x}_{2n})\, \mathcal{C}(\tfrac{\mathbf{x}_i - \mathbf{x}_j}{L}) \tag{26}$$

where M_n are differential operators

$$\mathcal{M}_n = \sum_{1\le i<j\le n} d^{\alpha\beta}(\mathbf{x}_i - \mathbf{x}_j)\, \partial_{x_i^\alpha}\partial_{x_j^\beta} - \kappa \sum_{i=1}^{n} \Delta_{\mathbf{x}_i}. \tag{27}$$

In principle, the above equations permit to determine uniquely the stationary higher-point correlators of the scalar iteratively by inverting the positive elliptic operators

M_{2n} . By analyzing these operators whose symbols loose strict positivity when κ goes to zero, it was argued in[5,15] that at least for small ξ

1. $\lim_{\kappa\to 0} \mathcal{F}_{2n}(\underline{\mathbf{x}})$ exists and is finite,

2. $\mathcal{F}_{2n}(\underline{\mathbf{x}}) = A_{C,2n} L^{\rho_{2n}} \mathcal{F}^0_{2n}(\underline{\mathbf{x}}) + \mathcal{O}(L^{-2+\mathcal{O}(\xi)}) + \ldots$ at $\kappa = 0$

where $\rho_{2n} = \frac{2n(n-1)}{d+2}\xi + \mathcal{O}(\xi^2)$, $\mathcal{F}^0_{2n}$ is a scaling zero mode of the $\kappa = 0$ version $\mathcal{M}^0_{2n}$ of the operator $\mathcal{M}_{2n}$: $\mathcal{M}^0_{2n}\mathcal{F}^0_{2n} = 0$, $\mathcal{F}^0_{2n}(\lambda\underline{\mathbf{x}}) = \lambda^{n(2-\xi)-\rho_{2n}}\mathcal{F}^0_{2n}(\underline{\mathbf{x}})$. The . . .-terms do not depend on at least one of the vectors $\mathbf{x}_i$ and do not contribute to the correlators of scalar differences.

3. $S_{2n}(\mathbf{x}) \propto L^{\rho_{2n}} |\mathbf{x}|^{n(2-\xi)-\rho_{2n}}$ at $\kappa = 0$ and for $|\mathbf{x}| \ll L$.

The last relation, a simple consequence of the second one, shows appearence of intermittent exponents at least for small ξ (ρ_{2n} is their anomalous part and it is positive starting from the 4-point function). A similar analysis, consistent with the above one, has been done for large space dimensions[6, 17].

The above results about the "zero-mode dominance" of the correlators of the scalar differences show what degree of universality one may expect in the scaling laws of intermittent quantities: the amplitudes $A_{C,2n}$ in front of the dominant term depend on the shape of the covariance C i.e. on the details of the large scale stirring. But the zero modes of the dominant terms (and their scaling exponents) do not. In field theory, small-scale universality of the critical behavior finds its explication in the renormalization group analysis. Similarly, in the Kraichnan model there exists a renormalization group explanation of the observed long scale universality[18] . The renormalization group transformations eliminate subsequently the long scale degrees of freedom. In a sense, they consist of looking at the system by stronger and stronger magnifying glass so that the long distance details are lost from sight. The eliminated degrees of freedom induce an effective source for the remaining ones[18]. Whereas such an "inverse renormalization group" analysis may be implemented for more complicated turbulent systems rests an open problem.

DYNAMICS OF LAGRANGIAN TRAJECTORIES

What is the source of the zero mode dominance of the inertial range correlators of the scalar differences? In absence of the diffusion term in Eq. (17) the scalar density is given by the integral

$$T(t,\mathbf{x}) = \int_{-\infty}^{t} f(s;\mathbf{y}(s;t,\mathbf{x}))\,ds \tag{28}$$

where $\mathbf{y}(s; t, \mathbf{x})$ describes the Lagrangian trajectory, i.e. the solution of the equation

$$\frac{d\mathbf{y}}{ds} = \mathbf{v}(s,\mathbf{y}), \tag{29}$$

passing at time t through point $\mathbf{x}$ (for concretness, we have assumed the vanishing initial condition for T at $t = -\infty$). The Lagrangian trajectories describe the flow of the fluid elements. If the velocities are random, so are Lagrangian trajectories and we may ask the question about their joint probability distributions. Let $P_n(t,\underline{\mathbf{x}};s,\underline{\mathbf{y}})$ denote the probability that n Lagrangian trajectories starting at time s at points $(\mathbf{y}_1, \ldots, \mathbf{y}_n) \equiv \underline{\mathbf{y}}$ pass at time $t \geq s$ through points $(\mathbf{x}_1, \ldots, \mathbf{x}_n) \equiv \underline{\mathbf{x}}$. These probabilities may be

computed for the time-decorrelated Gaussian velocities. They appear to be given by the heat kernels of the singular elliptic operators M_n^0

$$P_n(t,\underline{\mathbf{x}};s,\underline{\mathbf{y}}) \;=\; \mathrm{e}^{-(t-s)\mathcal{M}_n^0}(\underline{\mathbf{x}};\underline{\mathbf{y}}) \tag{30}$$

(more exactly, this is true after averaging over the simultaneous translations of the initial points, i.e. for the probabilities of relative positions of the trajectories).

From the form of the operators M_n^0 we see that the (relative positions of) n Lagrangian trajectories undergo a diffusion process with distance-dependent diffusion coefficients. When two trajectories are close, the corresponding diffusion coefficient vanishes as the distance to power ξ slowing down the diffusive separation of the trajectories. When the trajectories eventually separate, the diffusion coefficient grows speeding up further separation. The result is a superdiffusive large time asymptotics:

$$\int f(\underline{\mathbf{x}})\, P_n(t,\underline{\mathbf{x}};0,\underline{\mathbf{y}})\, d\underline{\mathbf{x}} \;\propto\; t^{\frac{\sigma}{2-\xi}} \tag{31}$$

for a generic (translationally invariant) scaling function $f\,(\underline{\mathbf{x}})$ of scaling dimension $\sigma > 0$ (e.g. for $\sum_{i,j}|\mathbf{x}_i - \mathbf{x}_j|^2$ with the scaling dimension 2). Note the faster than diffusive growth in time for $\xi > 0$. There are, however, exceptions from this generic behavior. In particular, if f is a scaling zero mode of M_n^0 then

$$\int f(\underline{\mathbf{x}})\, P_n(t,\underline{\mathbf{x}};0,\underline{\mathbf{y}})\, d\underline{\mathbf{x}} \;=\; \int f(\underline{\mathbf{x}})\, \mathrm{e}^{-t\mathcal{M}_n^0}(\underline{\mathbf{x}};\underline{\mathbf{y}})\, d\underline{\mathbf{x}} \;=\text{const.}\;=\; f(\underline{\mathbf{y}}). \tag{32}$$

It can be shown[19] that each zero mode f_0 of M_n^0 of scaling dimension $\sigma_0 \geq 0$ generates descendent slow collective modes f_p, p=1,2, . . . , of scaling dimensions $\sigma_p = \sigma_0 + p(2-\xi)$ for which

$$\int f_p(\underline{\mathbf{x}})\, P_n(t,\underline{\mathbf{x}};0,\underline{\mathbf{y}})\, d\underline{\mathbf{x}} \;\propto\; t^p, \tag{33}$$

i.e. grows slower (if $\sigma_0 > 0$) than in (31). The descendants satisfy the chain of equations $M_n f_p = f_{p-1}$. The structure with towers of descendants over the primary zero modes resembles that in systems with infinite symmetries and may suggest presence of hidden symmetries in the Kraichnan model.

The slow modes f_p appear in the asymptotic expansion

$$P_n(t,\underline{\mathbf{x}};0,\underline{\mathbf{y}}/L) \;=\; \sum_{\substack{\text{zero modes}\\ f_0}} \sum_{p=0}^{\infty} L^{-\sigma_p}\, \overline{g_p(t,\underline{\mathbf{x}})}\, f_p(\underline{\mathbf{y}}) \tag{34}$$

valid for large L and describing the behavior of the trajectories starting close to each other. Since $P_n(t,\underline{\mathbf{x}};0,\underline{\mathbf{y}}/L) = P_n(t,\underline{\mathbf{y}}/L;0,\underline{\mathbf{x}})$, expansion (34) describes also probabilities that the trajectories approach each other after time t. By simple rescalings and the use of expansion (34), one obtains

$$\begin{aligned}\int f(\underline{\mathbf{x}})\, P_n(t,\underline{\mathbf{x}};0,\underline{\mathbf{y}})\, d\underline{\mathbf{x}} &= t^{\frac{\sigma}{2-\xi}} \int f(\mathbf{x})\, P_n(1,\underline{\mathbf{x}};0,\underline{\mathbf{y}}/t^{\frac{1}{2-\xi}})\, d\underline{\mathbf{x}} \\ &= \sum_{\substack{\text{zero modes}\\ f_0}} \sum_{p=0}^{\infty} t^{\frac{\sigma-\sigma_p}{2-\xi}}\, f_p(\underline{\mathbf{y}}) \int f(\underline{\mathbf{x}})\, \overline{g_p(t,\underline{\mathbf{x}})}\, d\underline{\mathbf{x}}.\end{aligned} \tag{35}$$

For generic f, the dominant term comes from the constant zero mode and is proportional to $t^{\frac{\sigma}{2-\xi}}$. However for f equal to one of the slow modes, the leading contribution is

given by the further terms in the expansion due to the orthogonality relations between f_p's and g_p's. The behaviors (31,32,33) result. The distributions $P_n(t,\underline{\mathbf{x}};\, 0,\underline{\mathbf{y}})$ enter the expressions for scalar correlators. For those of scalar differences, the contributions to expansion (34) of modes which do not depend on all $\mathbf{y}_i$ (like the constant mode) drop out leading to the dominance by non-trivial zero modes.

The asymptotic expansion (34), governed by the slow modes f_p, displays another important feature of the Lagrangian trajectories. Since $P_n(t,\underline{\mathbf{x}};\, 0,\underline{\mathbf{y}})$ is a joint probability distribution of the endpoints of n Lagrangian trajectories starting at points $\mathbf{y}_1, \ldots, \mathbf{y}_n$, we should expect that it becomes concentrated on the diagonal $\mathbf{x}_1 = \ldots = \mathbf{x}_n$ when the initial points $\mathbf{y}_1, \ldots, \mathbf{y}_n$ tend to each other. But this is not the case as $\lim_{L\to\infty} P_n(t,\underline{\mathbf{x}};\, 0,\underline{\mathbf{y}}/L) = P_n(t,\underline{\mathbf{x}};\, 0,0)$ is a finite function of $\underline{\mathbf{x}}$. It is the constant zero mode contribution to expansion (34). The whole expansion describes how exactly $P_n(t,\underline{\mathbf{x}};\, 0,\underline{\mathbf{y}}/L)$ fails to concentrate on the diagonal when $L \to \infty$. Similarly, $P_n(t,\, 0;\, 0,\underline{\mathbf{y}})$ is a non-singular function of $\underline{\mathbf{y}}$ showing that the probability that the trajectories collapse after time t to a single point is finite, contradicting the uniqueness of the Lagrangian trajectories passing through a given point. The solution of the paradox is as follows. The typical velocities in the ensemble that we consider have rough spatial behavior (and even rougher time behavior). As functions of $\mathbf{x}$ they are (essentially) Hölder continuous with exponent $\frac{\xi}{2}$. But for such velocities, the equation for the Lagrangian trajectories (29) does not have a unique solution, given the initial position. As a result, the Lagrangian trajectories loose deterministic character for a fixed velocity realization. Nevertheless, one may still talk about probability distribution $P_n(t,\, \underline{\mathbf{x}};\, 0,\underline{\mathbf{y}}\,|\,\mathbf{v})$ of their final points whose average over $\mathbf{v}$ gives $P_n(t,\, \underline{\mathbf{x}}\;\, 0,\underline{\mathbf{y}})$. In a more realistic description which takes into account a smoothing of the typical velocities at very short viscous scale, the same effect is due to the sensitive dependence of the now deterministic Lagrangian trajectories on the initial conditions, within the viscous scale, signaled by the positivity of the Lyapunov exponent[19].

Summarizing, intermittency in the Kraichnan model of the passive advection appears to be due to the slow collective modes in the otherwise superdiffusive stochastic Lagrangian flow. The presence of such modes is closely related to the breakdown of the deterministic character of Lagrangian trajectories for the fixed velocity configuration at high Reynolds numbers, due to the sensitive dependence of the trajectories on the initial conditions within the viscous scale. We expect both phenomena to be present also in more realistic turbulent velocity distributions and to be still responsible for the anomalous scaling and intermittency.

REFERENCES

1. B. I. Halperin and P. C. Hohenberg: *Theory of Dynamical Critical Phenomena.* Rev. Mod. Phys. **49** (1977), 435-479
2. A. N. Kolmogorov: *The Local Structure of Turbulence in Incompressible Viscous Fluid for Very Large Reynolds' Numbers.* C. R. Acad. Sci. URSS **30** (1941), 301-305
3. R. Benzi, S. Ciliberto, C Baudet and G. Ruiz Chavaria: *On the Scaling of Three Dimensional Homogeneous and Isotropic Turbulence.* Physica **D 80** (1995), 385-398
4. U. Frisch: *Turbulence: the Legacy of A. N. Kolmogorov.* Cambridge University Press, Cambridge 1995
5. K. Gawędzki and A. Kupiainen: *Anomalous scaling of the Passive Scalar.* Phys. Rev. Lett. **75** (1995), 3834-3837
6. M. Chertkov, G. Falkovich, I. Kolokolov and V. *Lebedev: Normal and Anomalous Scaling of the Fourth-Order Correlation Function of a Randomly Advected Scalar.* Phys. Rev. **E 52** (1995), 4924-4941

7. B. Shraiman and E. Siggia: *Anomalous Scaling of a Passive Scalar in Turbulent Flow*. C.R. Acad. Sci. **321** (1995), 279-284
8. R. H. Kraichnan: *Small-Scale Structure of a Scalar Field Convected by Turbulence*. Phys. Fluids **11** (1968), 945-963
9. R. H. Kraichnan: *Anomalous Scaling of a Randomly Advected Passive Scalar*. Phys. Rev. Lett. **72** (1994), 1016-1019
10. A. L. Fairhall, O. Gat, V. L'vov, I. Procaccia: *Anomalous Scaling in a Model of Passive Scalar Advection: Exact Results*. Phys. Rev. **E 53** (1996), 3518-3535
11. S. Corrsin: *On the Spectrum of Isotropic Temperature Fluctuations in an Isotropic Turbulence*. J. Appl. Phys. **22** (1951), 469-473
12. R. Antonia, E. Hopfinger, Y. Gagne and F. Anselmet: *Temperature Structure Functions in Turbulent Shear Flows*. Phys. Rev. **A 30** (1984), 2704-2707
13. R. H. Kraichnan, V. Yakhot and S. Chen: *Scaling relations for a Randomly Advected Passive Scalar Field*. Phys. Rev. Lett. **75** (1995), 240-243
14. A. L. Fairhall, B. Galanti, V. S. L'vov and I. Procaccia: *Direct Numerical Simulations of the Kraichnan Model: Scaling Exponents and Fusion Rules*. chao-dyn/9707003, submitted to Phys. Rev. Lett.
15. D. Bernard, K. Gawędzki and A. Kupiainen: *Anomalous Scaling in the N-Point Functions of a Passive Scalar*. Phys. Rev. **E 54** (1996), 2564-2572
16. R. H. Kraichnan: *Passive Scalar: Scaling Exponents and Realizability*. Phys. Rev. Lett. **78** (1997), 4922-4925
17. M. Chertkov and G. Falkovich: *Anomalous Scaling Exponents of a White-Advected Passive Scalar*. Phys. Rev. Lett **76** (1996), 2706-2709
18. K. Gawędzki: *Turbulence under a Magnifying Glass*. chao-dyn/9610003, to appear in the proceedings of the 1996 Cargèse Summer Institute
19. D. Bernard, K. Gawędzki and A. Kupiainen: *Slow Modes in Passive Advection*. cond-mat/9706035, submitted to J. Stat. Phys.

This page intentionally left blank

FIELD THEORY AS FREE FALL
A "Proper-Time" Approach to Classical and Quantum Gravity

A. Carlini[1] and J. Greensite[2*]

[1]Tokyo Institute of Technology
Oh-Okayama, Meguro-ku
Tokyo 152, Japan
[2]The Niels Bohr Institute
DK-2100 Copenhagen Ø, Denmark

INTRODUCTION

In this talk I would like to discuss the meaning, and possible utility, of the concept of "proper-time" when applied to the dynamics of the Universe as a whole. My talk is based on a series of papers by Alberto Carlini and myself.[1-3]

The motion of a point particle in free fall is described by the geodesic equation

$$g_{\mu\nu}\frac{d^2x^\nu}{dt^2} + \frac{1}{2}\left(\frac{\partial g_{\mu\alpha}}{\partial x^\beta} + \frac{\partial g_{\mu\beta}}{\partial x^\alpha} - \frac{\partial g_{\alpha\beta}}{\partial x^\mu}\right)\frac{dx^\alpha}{dt}\frac{dx^\beta}{dt} = 0 \tag{1}$$

where t is an affine parameter proportional to the proper-time of the particle trajectory. The classical field equations for gravity ($g_{\mu\nu}$) coupled to any number of other bosonic fields (ϕ^A) look, of course, quite different,

$$\begin{aligned} R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R &= -\kappa T_{\mu\nu} \\ \frac{\delta S}{\delta \phi^A} &= 0 \end{aligned} \tag{2}$$

where $S = S[g_{\mu\nu}\ \phi^A]$ is the action, and the time parameter corresponds to an arbitrary 3 + 1 decomposition of the Lorentzian 4-manifold. Nevertheless, I will show the following:

I. The Einstein + other bosonic field equations describe the free fall of a point particle in superspace, and can be put in the form of a geodesic equation.

II. Just as the mass m of a particle in free fall doesn't appear in the usual geodesic equation, there is a free parameter M in gravitational dynamics that does not appear in the classical field equations, but which *may* be important at the quantum level.

*Talk presented by J. Greensite.

New Developments in Quantum Field Theory
Edited by Damgaard and Jurkiewicz, Plenum Press, New York, 1998

III. Path integration of the "square-root" form of the gravitational action leads to quantum evolution of states in a proper-time parameter.

It is not yet clear whether this proper-time formulation will be useful. I will, however, mention some possible (albeit speculative) applications.

WORLDLINE ACTION FOR BOSONIC FIELDS

The classical motion of a freely falling, spinless particle of mass m (i.e. the geodesic equation) is derived from variation of the worldline action, which is proportional to the proper-time interval of the particle trajectory

$$\begin{aligned} S_p &= -m \int ds \\ &= -m \int d\tau \sqrt{-g_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau}} \end{aligned} \tag{3}$$

Removing the square-root by introduction of a Lagrange multiplier (lapse) N, we can write S_p in the form

$$S_p' = m \int d\tau \left[\frac{1}{2N} g_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} - \frac{1}{2} N \right] \tag{4}$$

and in 1st-order form

$$\begin{aligned} S_p'' &= \int d\tau \left[p_\mu \frac{dx^\mu}{d\tau} - NH \right] \\ H &= \frac{1}{2m} (g^{\mu\nu} p_\mu p_\nu + m^2) \end{aligned}$$

with $H = 0$ (which follows from variation w.r.t. N) the mass-shell condition. Quantum-mechanically, this condition is imposed as a constraint on physical states

$$H\psi = 0 \tag{5}$$

which is just the Klein-Gordon equation in curved spacetime. In an interacting theory the mass-shell condition is relaxed somewhat; it is required only of asymptotic states. The 4-momentum of a virtual particle is allowed to violate the mass-shell condition.

We would like to generalize the proper-time approach to the case of gravity in combination with any number of interacting bosonic fields; this calls for rewriting the gravitational action in the form

$$S_g = -\mathcal{M} \int ds \tag{6}$$

where s is an invariant length parameter in the space of all fields modulo spatial diffeomorphisms, i.e. superspace. The only reasonable candidate for s is the usual action of general relativity, so the problem is to reformulate that action as a proper time in superspace.

Let $\{q^a(x), p_a(x)\}$ represent a set of gravitational and other bosonic fields, and their conjugate momenta, with the fields scaled by an appropriate power of $\kappa^2 = 16\pi G_N$ so as to be dimensionless. The standard ADM action has the form

$$\begin{aligned} S_{ADM} &= \int d^4x \, [p_a \partial_0 q^a - N\mathcal{H}_x - N_i \mathcal{H}_x^i] \\ \mathcal{H}_x &= \kappa^2 G^{ab} p_a p_b + \sqrt{g} U(q) \\ \mathcal{H}_x^i &= O^{ia}[q, \partial_x] p_a \end{aligned} \tag{7}$$

where $\{q^a(x)\} = \{g_{ij}(x), \phi^A(x)\}$ is the set of 3-metric and other bosonic fields. In pure gravity, for example, in this condensed notation,

$$\begin{aligned} \{a = 1-6\} &\leftrightarrow \{(i,j),\ i \le j\} \\ q^a(x) &\leftrightarrow g_{ij}(x) \\ p_a(x) &\leftrightarrow \begin{cases} p^{ij}(x) & i = j \\ 2p^{ij}(x) & i < j \end{cases} \\ G_{ab}(x) &\leftrightarrow G^{ijnm}(x) \\ \sqrt{g}U &\leftrightarrow -\frac{1}{\kappa^2}\sqrt{g}R \end{aligned} \tag{8}$$

Now express momenta in terms of velocities, and solve the Hamiltonian constraint $H = 0$ for the lapse function, with a choice of shift functions $N_i = 0$. Insert the result into the ADM action to get the Baierlein-Sharp-Wheeler (BSW) form of the action: [5]

$$S_{BSW} = -\int d^4x \sqrt{-\frac{1}{\kappa^2}\sqrt{g}UG_{ab}\partial_0 q^a \partial_0 q^b} \tag{9}$$

This is not yet a worldline action. For that, we need some additional manipulations. First choose a time-parameter $t = x_0$ proportional to the BSW action, i.e.

$$S_{BSW} = -\int ds = -\sigma \int dt \tag{10}$$

where σ is an arbitrary parameter with dimensions of mass. Inspection of the BSW action gives

$$dt = \frac{1}{\sigma}\int d^3x \sqrt{-\frac{1}{\kappa^2}\sqrt{g}UG_{ab}dq^a dq^b} \tag{11}$$

or

$$1 = \frac{1}{\sigma}\int d^3x \sqrt{-\frac{1}{\kappa^2}\sqrt{g}UG_{ab}\partial_t q^a \partial_t q^b} \tag{12}$$

Let $\tilde{N}$ denote the lapse function (derived by solving the Hamiltonian constraint) associated with this time parameter t

$$\tilde{N} = \sqrt{-\frac{1}{4\kappa^2\sqrt{g}U}G_{ab}\partial_t q^a \partial_t q^b} \tag{13}$$

Then

$$1 = -\frac{1}{\sigma}\int d^3x\ \frac{1}{2\tilde{N}\kappa^2}G_{ab}\partial_t q^a \partial_t q^b \tag{14}$$

The condition $dt \propto ds$ imposes only one global restriction on the choice of $\tilde{N}$. Since

$$\int d^3x\ \tilde{N}\sqrt{g}U = \int d^3x \sqrt{-\frac{1}{4\kappa^2}\sqrt{g}UG_{ab}\partial_t q^a \partial_t q^b} \tag{15}$$

then from (12) we get the condition

$$\int d^3x\ \tilde{N}\sqrt{g}U = \frac{1}{2}\sigma \tag{16}$$

For any given $\tilde{N}$ satisfying this condition, there corresponds a time parameter t proportional to S_{BSW}. The condition is solved trivially by

$$\tilde{N} = \frac{\frac{1}{2}\sigma\mathcal{N}}{\int d^3x\ \mathcal{N}\sqrt{g}U} \tag{17}$$

where N is unrestricted. Then

$$1 = -\frac{1}{\sigma^2}\int d^3x\ \left[\int d^3x'\ \mathcal{N}\sqrt{g}U\right]\frac{1}{\mathcal{N}\kappa^2}G_{ab}\partial_t q^a \partial_t q^b \tag{18}$$

or

$$ds^2 = \int d^3x\ \left[\int d^3x'\ \mathcal{N}\sqrt{g}U\right]\frac{1}{\mathcal{N}\kappa^2}G_{ab}dq^a dq^b \tag{19}$$

Now we introduce a mixed discrete/continuous "coordinate index" (α, x) in superspace:

$$q^{(\alpha x)} \equiv q^\alpha(x) = \begin{cases} \mathcal{N}(x) & \alpha = 0 \\ q^a(x) & \alpha = a \neq 0 \end{cases} \tag{20}$$

We also define a degenerate supermetric

$$\begin{aligned} \mathcal{G}_{(ax)(by)} &= \left[\int d^3x'\ \mathcal{N}\sqrt{g}U\right]\frac{1}{\mathcal{N}(x)\kappa^2}G_{ab}(x)\delta^3(x-y) \\ \mathcal{G}_{(0x)(0y)} &= \mathcal{G}_{(ax)(0y)} = \mathcal{G}_{(0x)(by)} = 0 \end{aligned} \tag{21}$$

With these definitions,

$$ds^2 = -\mathcal{G}_{(\alpha x)(\beta y)}dq^{(\alpha x)}dq^{(\beta y)} \tag{22}$$

so finally

$$\begin{aligned} S_g &= -\mathcal{M}\int ds \\ &= -\mathcal{M}\int d\tau\ \sqrt{-\mathcal{G}_{(\alpha x)(\beta y)}\frac{dq^{(\alpha x)}}{d\tau}\frac{dq^{(\beta y)}}{d\tau}} \end{aligned} \tag{23}$$

This is the required world-line action for gravity coupled to any number of bosonic fields.

Variation of the action S_g w.r.t $q^{(\alpha x)}$ leads, in the usual way, to a geodesic equation

$$\mathcal{G}_{(\alpha x)(\beta y)}\frac{d^2q^{(\beta y)}}{ds^2} + \frac{1}{2}\left(\frac{\delta\mathcal{G}_{(\alpha x)(\beta y)}}{\delta q^{(\gamma z)}} + \frac{\delta\mathcal{G}_{(\alpha x)(\gamma z)}}{\delta q^{(\beta y)}} - \frac{\delta\mathcal{G}_{(\beta y)(\gamma z)}}{\delta q^{(\alpha x)}}\right)\frac{dq^{(\beta y)}}{ds}\frac{dq^{(\gamma z)}}{ds} = 0 \tag{24}$$

It is straightforward to verify that the $\alpha \neq 0$ components of the geodesic equation are the equations of motion

$$\underline{\frac{\partial}{\partial t}\left[\frac{1}{2\tilde{N}\kappa^2}G_{ab}\partial_t q^b\right] - \frac{1}{4\tilde{N}\kappa^2}\frac{\partial G_{cd}}{\partial q^a}\partial_t q^c \partial_t q^d + \int d^3x'\ \tilde{N}\frac{\delta}{\delta q^a(x)}(\sqrt{g}U) = 0} \tag{25}$$

while the $\alpha = 0$ component is the Hamiltonian constraint

$$\frac{1}{4\tilde{N}^2\kappa^2}G_{ab}\partial_t q^a \partial_t q^b + \sqrt{g}U = 0 \tag{26}$$

These equations are identical to those obtained from the ADM action, with the gauge choice $N_i = 0$ and $N = \tilde{N}$.† We have therefore interpreted the classical field equations of general relativity as describing the free fall of a point particle in superspace. We can

† The supermomentum constraints $H_i = 0$, may appear to be missing, but in fact those constraints can be derived, as a consistency condition, from the other equations of motion and the Hamiltonian constraint. [4]

now proceed, as in the relativistic particle case, to derive the proper-time Hamiltonian. Again introducing a Lagrange multiplier n to remove the square-root, we have

$$S'_g = \mathcal{M} \int d\tau \left[\frac{1}{2n} \mathcal{G}_{(ax)(by)} \frac{dq^{(ax)}}{d\tau} \frac{dq^{(by)}}{d\tau} - \frac{1}{2} n \right] \tag{27}$$

and the first-order form is

$$\begin{aligned} S''_g &= \int d\tau \left[p_{(ax)} \frac{dq^{(ax)}}{d\tau} - - - nH \right] \\ H &= \frac{1}{2\mathcal{M}} [\mathcal{G}^{(ax)(by)} p_{(ax)} p_{(by)} + \mathcal{M}^2] \\ &= \frac{1}{2\mathcal{M}} [Æ + \mathcal{M}^2] \end{aligned} \tag{28}$$

where

$$\begin{aligned} Æ &= \mathcal{G}^{(ax)(by)} p_{(ax)} p_{(by)} \\ &= \frac{\int d^3x \, \mathcal{N} \kappa^2 G^{ab} p_a p_b}{\int d^3x' \, \mathcal{N} \sqrt{g} U} \end{aligned} \tag{29}$$

Variation of the 1st order action with respect to $q^a(x,\tau)$, $p_a(x,\tau)$ and $N(x,\tau)$, $n(\tau)$ gives us, respectively, the set of Hamiltonian equations and constraints

$$\partial_\tau q^a(x) = \frac{n}{2\mathcal{M}} \frac{\delta Æ}{\delta p_a(x)} \quad , \quad \partial_\tau p_a(x) = \frac{n}{2\mathcal{M}} \frac{\delta Æ}{\delta q^a(x)} \quad , \quad \frac{\delta Æ}{\delta \mathcal{N}(x)} = 0 \quad , \quad Æ = -\mathcal{M}^2 \tag{30}$$

Setting $n = 1$, so that $\tau = s = \sigma t$; these equations are equivalent to the usual Hamiltonian equations of motion and Hamiltonian constraint

$$\begin{aligned} \partial_t q^a(x) &= \int d^3x' \, \tilde{N} \frac{\delta}{\delta p_a(x)} \mathcal{H} \\ \partial_t p_a(x) &= -\int d^3x' \tilde{N} \frac{\delta}{\delta q^a(x)} \mathcal{H} \\ \mathcal{H} &= \frac{\kappa^2}{\mathcal{M}} G^{ab} p_a p_b + \mathcal{M} \sqrt{g} U = 0 \end{aligned} \tag{31}$$

in the $N = \tilde{N}$, $N_i = 0$ gauge. We note that since N is arbitrary, this gauge choice does not imply any loss of freedom in the choice of constant-time hypersurfaces. The only restriction is on the labeling of those hypersurfaces, such that $\Delta t \propto \Delta S_{BSW}$.

The constant M in eq. (31) is implicitly set to $M = 1$, in the usual Hamiltonian formulation of general relativity, but there is no overwhelming reason to make this choice. We have seen that M appears as a constant multiplicative factor in the worldline action of the Universe, as does the mass m in the worldline action of a particle. Both of these constants drop out of the corresponding geodesic equations. In the context of the first-order formulation, the condition

$$Æ = -\mathcal{M}^2 \tag{32}$$

is in every sense analogous to the particle mass-shell condition $g^{\mu\nu} p_\mu p_\nu = -m^2$. We dignify this constraint with the title "Mass-Shell of the Universe."

Canonical quantization, in the "proper-time" gauge $n = 1$, leads to the proper-time dependent Schrödinger equation

$$i\hbar \frac{\partial \Psi}{\partial s} = \frac{1}{2\mathcal{M}} (Æ + \mathcal{M}^2) \Psi \tag{33}$$

which has the general s-dependent solution

$$\Psi[q,s] = \sum_{\mathcal{E}\beta} a_{\mathcal{E}\beta}\Phi_{\mathcal{E}\beta}[q]e^{i(\mathcal{E}-\mathcal{M}^2)s/(2\mathcal{M}\hbar)}$$
$$\text{Æ}\Phi_{\mathcal{E}\beta}[q] = -\mathcal{E}\Phi_{\mathcal{E}\beta}[q] \tag{34}$$

where the label β distinguishes among a linearly independent set of eigenstates of Æ with eigenvalue –ε. The classical constraint $\delta\text{Æ}/\delta N = 0$ becomes an operator constraint $\frac{\delta\text{Æ}}{\delta N}\Psi = 0$. Inserting the eigenstate expansion, we find that each eigenstate Φ_ε satisfies a Wheeler-DeWitt equation

$$\left[-\frac{\hbar^2}{\mathcal{E}}\kappa^2 G^{ab}\frac{\delta^2}{\delta q^a \delta q^b} + \sqrt{g}U\right]\Phi_{\mathcal{E}}[q] = 0 \tag{35}$$

associated with the parameter ε. Finally, if we also impose the mass-shell constraint

$$\text{Æ}\Psi = -\mathcal{M}^2\Psi \tag{36}$$

then the only physical states are those with $\varepsilon = M^2$, and the (classically indeterminate) constant M can be absorbed, via

$$\hbar_{eff} = \frac{\hbar}{\mathcal{M}} \tag{37}$$

into a rescaling of Planck's constant.

PATH-INTEGRAL FORMULATION

We will now derive the proper-time formulation in quite a different way, from direct path-integral quantization of the square-root form of the action. Although path-integral methods have been applied to, e.g., the square-root particle worldline action in Riemannian spacetime,[6] they have not, to our knowledge, been applied to square-root actions with Lorentzian metrics. The signature of the metric inside the square root introduces some new and quite non-trivial features in the quantization procedure, which will be discussed below.

In non-time-parametrized theories, the path-integral is built out of elementary integrals

$$\psi(q', t+\epsilon) = \int d^D q\ \mu_\epsilon \exp[iS[(q', t+\epsilon),(q,t)]/\hbar]\psi(q,t)$$
$$= U_\epsilon \psi(q', t) \tag{38}$$

where $S[(q', t+\epsilon),(q, t)]$ is the action of a classical trajectory between the points q at time t and q' at time $t + \varepsilon$. A unitary evolution operator

$$U_\epsilon = \exp[-iH\epsilon/\hbar] \quad + \quad O(\epsilon^2) \tag{39}$$

is obtained. Direct imitation of this construction doesn't work in time-parametrized theories, because

$$S[(q', t+\epsilon);(q,t)] = S[q', q] \tag{40}$$

which implies that U_ϵ is ϵ-independent (and in general non-unitary). Of course, there are well-known methods of evaluating the functional integral in certain cases, e.g. relativistic particles and strings propagating in a flat background, to obtain Green functions

$G[q', q]$. The easiest method is to introduce auxiliary fields, to remove the square root. But our aim here is rather different, and is related to the infamous "problem of time"[7] in time-diffeomorphism invariant theories: It is to describe the evolution of quantum states, in a time-diffeomorphism invariant theory, in terms of an appropriate proper-time parameter.

With this aim in mind, let us try to recover a unitary evolution operator of the form (39) by making a small change to the construction (38)

$$\begin{aligned}\psi(q', \tau + \epsilon) &= \int d^D q \; \mu_\epsilon \exp[c_\epsilon S(q', q)]\psi(q, \tau) \\ &= U_\epsilon \psi(q', \tau)\end{aligned} \tag{41}$$

where $c_\in$ is a complex constant to be chosen such that $U_\in$ is unitary. Begin with the relativistic particle action, where

$$\psi(x', \tau + \epsilon) = \int d^D x \; \mu_\epsilon \exp[-c_\epsilon m\sqrt{-\eta_{\mu\nu}\Delta x^\mu \Delta x^\nu}]\psi(x, \tau) \tag{42}$$

with measure

$$\mu_\epsilon^{-1} = \int d^D x \; \exp[-c_\epsilon m\sqrt{-\eta_{\mu\nu}\Delta x^\mu \Delta x^\nu}] \tag{43}$$

so that $U_\in \to 1$ as $\in \to 0$ Comparing to the corresponding expression for a free non-relativistic particle

$$\psi(x', \tau + \epsilon) = \int d^D x \; \mu_\epsilon \exp\left[m\frac{\delta_{ij}\Delta x^i \Delta x^j}{(-i\epsilon\hbar)}\right]\psi(x, t) \tag{44}$$

motivates us to try

(45)

The "time"-step $\in$ now has units of action.

To complete the definition of the functional integral, we must decide whether to integrate over *all* Δx, including both timelike and spacelike path segments, or only over timelike paths. The corresponding question, in quantum gravity, is whether to integrate over manifolds of Lorentzian *and* Euclidean signature, or just over Lorentzian manifolds. To address this question, let us calculate separately the contributions to $U_\in$ from timelike and spacelike paths. Following the usual steps leading from the path-integral to Schrodinger equation

$$\begin{aligned}\psi(x', \tau + \epsilon) &= \int d^D x \; \mu_\epsilon \exp[-m\sqrt{-\eta_{\mu\nu}\Delta x^\mu \Delta x^\nu}/\sqrt{-i\epsilon\hbar}] \\ &\quad \times \left[\psi(x', \tau) + \frac{\partial \psi}{\partial x'^\mu}\Delta x^\nu + \frac{1}{2}\frac{\partial^2 \psi}{\partial x'^\mu \partial x'^\nu}\Delta x^\mu \Delta x^\nu + \ldots\right] \\ &= U_\epsilon \psi(x', t)\end{aligned} \tag{46}$$

One finds

$$\begin{aligned}U_\epsilon &= 1 + \frac{1}{2(D-1)}\frac{I_B}{I_A}\partial^2 \quad + \quad \ldots \\ I_A &\equiv \int d^D x \; \exp[-m\sqrt{-\eta_{\mu\nu}x^\mu x^\nu}/\sqrt{-i\epsilon\hbar}] \\ I_B &\equiv \int d^D x \; r^2 \exp[-m\sqrt{-\eta_{\mu\nu}x^\mu x^\nu}/\sqrt{-i\epsilon\hbar}]\end{aligned} \tag{47}$$

where $x^\mu = \{t, \vec{x}\}$ and $r^2 = \vec{x} \cdot \vec{x}$. Divide I_A into timelike and spacelike contributions

$$I_A = \sigma \int_0^\infty dr\, r^{D-2} \left\{ \int_r^\infty dt \exp[-\frac{m\sqrt{t^2 - r^2}}{\sqrt{-i\epsilon\hbar}}] + \int_0^r dt \exp[-\frac{m\sqrt{r^2 - t^2}}{\sqrt{i\epsilon\hbar}}] \right\}$$
$$= \sigma \int_0^\infty dr\, r^{D-2} \{F_1(r) + F_2(r)\} \tag{48}$$

where

$$F_1(r) \equiv \int_0^\infty dy \frac{y}{\sqrt{y^2 + r^2}} e^{-my/\sqrt{-i\epsilon\hbar}} \sim O(\frac{1}{r}) \quad \text{as } r \to \infty$$
$$F_2(r) \equiv \int_0^r dy \frac{y}{\sqrt{r^2 - y^2}} e^{-my/\sqrt{i\epsilon\hbar}} \sim O(\frac{1}{r}) \quad \text{as } r \to \infty \tag{49}$$

therefore

$$\text{timelike paths contribution} = \sigma \int_0^\infty dr\, r^{D-2} F_1(r) \quad \text{is divergent}$$
$$\text{spacelike paths contribution} = \sigma \int_0^\infty dr\, r^{D-2} F_2(r) \quad \text{is divergent} \tag{50}$$

(and similarly for I_B). The interim conclusion is that an evolution operator defined over timelike paths alone, or spacelike paths alone, is meaningless. What about the sum? It turns out that

$$F_1 + F_2 = -\int_r^{i\infty} dy \frac{y}{\sqrt{r^2 - y^2}} e^{-my/\sqrt{i\epsilon\hbar}} = irK_1(\frac{mr}{\sqrt{i\epsilon\hbar}}) \tag{51}$$

which is exponentially damped at large r! We then find that

$$I_A = i\sigma 2^{D-2} (\frac{\sqrt{i\epsilon\hbar}}{m})^D \Gamma(\frac{D+1}{2}) \Gamma(\frac{D-1}{2})$$
$$I_B = i\sigma 2^D (\frac{\sqrt{i\epsilon\hbar}}{m})^{D+2} \Gamma(\frac{D+1}{2} + 1) \Gamma(\frac{D-1}{2} + 1) \tag{52}$$

and finally

$$U_\epsilon = 1 + i\epsilon\hbar \frac{(D+1)}{2m^2} \eta^{\mu\nu} \partial_\mu \partial_\nu + O(\epsilon^2)$$
$$= \exp[-i\text{Æ}\epsilon/\hbar] + O(\epsilon^2) \tag{53}$$

where

$$\text{Æ} = (D+1) \frac{\frac{1}{2m} \eta^{ab} p_a p_b}{m} \tag{54}$$

Note that $c_\in = 1/\sqrt{-i\epsilon\hbar}$ leads to the $i\in$ factor required for unitarity in $U_\in$. Our conclusion is that for the relativistic particle action, we *must* sum over timelike and spacelike contributions. A similar analysis for gravity implies that both Lorentzian and Euclidean 4-manifolds contribute to the evolution of the wavefunction.

The "unitary" path-integral formulation we have introduce here leads to the correct equation of motion (in this case the geodesic equation in flat space) in the classical limit. It does not impose a definite mass-shell condition, which is irrelevant to the geodesic equation. The mass-shell condition only pertains to stationary states of the quantum theory. These stationary states satisfy

$$\text{Æ}\psi = -\mathcal{E}\psi \tag{55}$$

which is simply the Klein-Gordon equation

$$[\eta^{ab}p_a p_b + m_{eff}^2]\psi = 0 \tag{56}$$

with a mass term

$$m_{eff}^2 = \frac{2\mathcal{E}}{D+1}m^2 \tag{57}$$

Non-stationary states are superpositions of Klein-Gordon states with different values of m_{eff}^2. It can be shown that as long as we only superimpose states with $\varepsilon > 0$, i.e. $m_{eff}^2 > 0$ non-tachyonic, the wavepacket follows a timelike trajectory.

Next, consider minisuperspace models of the form

$$\begin{aligned} S &= \int dt\ \{p_a \partial_t q^a - NH[p,q]\} \\ H &= \frac{1}{2m}G^{ab}(q)p_a p_b + mV(q) \end{aligned} \tag{58}$$

where the supermetric G^{ab} has Lorentzian signature (– + + + ... +). The "square-root" form of the action is obtained by solving for p_a in terms of the time-derivatives of the $\{q^a\}$

$$\begin{aligned} \partial_t q^a &= N\frac{\partial H}{\partial p_a} = \frac{N}{m}G^{ab}p_b \\ \Longrightarrow p_a &= \frac{m}{N}G_{ab}\partial_t q^b \end{aligned}$$

and then solving the Hamiltonian constraint for the lapse function

$$\begin{aligned} 0 &= \frac{1}{2m}G^{ab}(q)p_a p_b + mV(q) \\ &= \frac{m}{2N^2}G_{ab}\partial_t q^a \partial_t q^b + \frac{1}{2}mV(q) \\ \Longrightarrow N &= \sqrt{-\frac{1}{2V}G_{ab}\partial_t q^a \partial_t q^b} \end{aligned}$$

Substitute into the minisuperspace action, to find

$$S = -m\int dt\ \sqrt{-2VG_{ab}\partial_t q^a \partial_t q^b} \tag{59}$$

Define the modified supermetric

$$\mathcal{G}_{ab} \equiv 2VG_{ab} \tag{60}$$

so for Δq^a small

$$\begin{aligned} S[q',q] &= -m\int_0^{\Delta t} dt\ \sqrt{-\mathcal{G}_{ab}\partial_t q^a \partial_t q^b} \\ &= -m\sqrt{-\mathcal{G}_{ab}\Delta q^a \Delta q^b} \end{aligned}$$

The measure is

$$\mu_\epsilon^{-1}(q') = (\sqrt{\epsilon\hbar})^D \lim_{\epsilon\to 0}\int \frac{d^D q}{(\sqrt{\epsilon\hbar})^D}\ \exp[-m\sqrt{-\mathcal{G}_{ab}\Delta q^a \Delta q^b}/\sqrt{-i\epsilon\hbar}] \tag{61}$$

Introduce Riemann normal coordinates ξ^a around the point q'^a, which transforms the modified supermetric into the Minkowski metric $G_{ab} = \eta_{ab}$ at the point q' ($\xi = 0$). Then

$$\begin{aligned}\mu_\epsilon^{-1}(q') &= \int d^D\xi \left|\det[\frac{\partial \Delta q^a}{\partial \xi^b}]\right| \exp[-m\sqrt{-\eta_{ab}\xi^a\xi^b}/\sqrt{-i\epsilon\hbar}] \\ &= \frac{1}{\sqrt{|\mathcal{G}(q')|}} I_A \end{aligned} \tag{62}$$

so that

$$\begin{aligned}\psi(q', \tau+\epsilon) &= \frac{1}{I_A}\int d^D\Delta q \sqrt{|\mathcal{G}(q'+\Delta q)|}\exp[-m\sqrt{-\mathcal{G}_{ab}\Delta q^a \Delta q^b}/\sqrt{-i\epsilon\hbar}]\psi(q'+\Delta q, \tau) \\ &= \frac{1}{I_A}\int d^D\xi\ (1 - \frac{1}{6}\mathcal{R}_{ab}\xi^a\xi^b + ...)\exp[-m\sqrt{-\eta_{ab}\xi^a\xi^b}/\sqrt{-i\epsilon\hbar}] \\ &\quad \times \left\{\psi(q', \tau) + \frac{\partial\psi}{\partial\xi^c}\xi^c + \frac{1}{2}\frac{\partial^2\psi}{\partial\xi^c\partial\xi^d}\xi^c\xi^d + O(\xi^3)\right\} \\ &= \left[1 + \frac{1}{(D-1)}\frac{I_B}{I_A}\{\frac{1}{2}\partial^2 - \frac{1}{6}\mathcal{R} + O(\epsilon^2)\}\right]\psi(q'(\xi), \tau) \\ &= \left[1 + i\epsilon\hbar\frac{D+1}{2m^2}\eta^{ab}\frac{\partial^2}{\partial\xi^a\partial\xi^b} - i\epsilon\hbar\frac{D+1}{6m^2}\mathcal{R}\right]\psi(\xi, \tau)\end{aligned} \tag{63}$$

where R is the curvature scalar formed from the metric G_{ab}. Transforming back from Riemann normal coordinates, we have

$$\begin{aligned}\psi(q, \tau+\epsilon) &= \left[1 + i\epsilon\hbar\frac{D+1}{2m^2}\frac{1}{\sqrt{|\mathcal{G}|}}\frac{\partial}{\partial q^a}\sqrt{|\mathcal{G}|}\mathcal{G}^{ab}\frac{\partial}{\partial q^b} - i\epsilon\hbar\frac{D+1}{6m^2}\mathcal{R}\right]\psi(q, \tau) \\ &= U_\epsilon\psi(q, \tau)\end{aligned} \tag{64}$$

As in the relativistic particle case, $\exp[-i\text{Æ}\epsilon/\hbar]$ is a unitary operator, where

$$\text{Æ} = -\hbar^2\frac{D+1}{2m^2}\frac{1}{\sqrt{|\mathcal{G}|}}\frac{\partial}{\partial q^a}\sqrt{|\mathcal{G}|}\mathcal{G}^{ab}\frac{\partial}{\partial q^b} + \hbar^2\frac{D+1}{6m^2}\mathcal{R} \tag{65}$$

is obviously Hermitian in the measure μ_ϵ. Taking the $\epsilon \to 0$ limit, the wavefunction $\psi(q, \tau)$ satisfies a Schrodinger equation

$$i\hbar\partial_\tau\psi(x, \tau) = \text{Æ}\psi(q, \tau) \tag{66}$$

and the τ-independent Schrodinger equation $\text{Æ}\phi_\varepsilon = -\varepsilon\phi_\varepsilon$ is

$$\left[-\frac{\hbar^2}{\mathcal{E}}\frac{D+1}{4m^2}\frac{V}{\sqrt{|\mathcal{G}|}}\frac{\partial}{\partial q^a}\frac{\sqrt{|\mathcal{G}|}}{V}G^{ab}\frac{\partial}{\partial q^b} + \frac{\hbar^2}{\mathcal{E}}\frac{D+1}{6m^2}V\mathcal{R} + V\right]\phi_\mathcal{E}(q) = 0 \tag{67}$$

This is simply the Wheeler-DeWitt equation $H^\varepsilon\phi_\varepsilon = 0$ with a particular operator-ordering. In the classical limit

$$\begin{aligned}\text{Æ}_{cl}[q^a, p_a] &= \lim_{\hbar\to 0}\text{Æ}[q^a, -i\hbar\frac{\partial}{\partial q^a} \to p_a] \\ &= \frac{1}{2}(D+1)\frac{\frac{1}{2m}G^{ab}p_a p_b}{mV}\end{aligned} \tag{68}$$

which again has the "Kinetic/Potential" form of eq. (29). It can be readily verified that

$$\partial_\tau Q = \{Q, \text{Æ}_{cl}\} \tag{69}$$

gives the correct classical dynamics. The τ evolution parameter is a proper-time, for trajectories in minisuperspace with metric G_{ab}.

For quantum gravity, the starting point is the BSW action in eq. (9). Following steps analogous to the minisuperspace case leads to a proper-time Schrodinger equation

$$i\hbar\partial_\tau \Psi = \text{Æ}\Psi \tag{70}$$

where

$$\text{Æ} = \int d^3x \left[-\hbar^2 \tilde{N} \kappa^2 G^{ab} \frac{\delta^2}{\delta q^a \delta q^b} + \text{operator-ordering terms} \right] \tag{71}$$

and

$$\tilde{N}(x) \equiv \frac{\mathcal{N}(x)}{\int d^3x' \sqrt{g} \mathcal{N} U(q)} \tag{72}$$

Classically, this is the same Æ as in the "free fall" approach. For the details of the derivation, I must refer the reader to one of our articles.[3] The general solution of this evolution equation has again the form

$$\begin{aligned} \Psi[q^a(x), \tau] &= \sum_{\mathcal{E}} a_{\mathcal{E}\beta} \Phi_{\mathcal{E}\beta}[q^a(x)] e^{i\mathcal{E}\tau/\hbar} \\ \text{Æ}\Phi_{\mathcal{E}\beta} &= -\mathcal{E}\Phi_{\mathcal{E}\beta} \\ &\Longrightarrow \mathcal{H}^{\mathcal{E}}\Phi_{\mathcal{E}\beta} = 0 \end{aligned} \tag{73}$$

where H^ε is the Hamiltonian in eq. (31) with $\varepsilon = M^2$. As in the case of the relativistic particle, each stationary state $\Phi_{\varepsilon\beta}$ is a solution of a Hamiltonian constraint (in this case, the Wheeler-DeWitt equation), each with a separate value of the parameter ε.

Once again, unitarity of U_ε *requires* summing over all $q^a(x,\tau)$, including regions where the lapse function is imaginary. This means that we are integrating over manifolds of Euclidean, Lorentzian, and mixed Euclidean/Lorentzian signature. Why, then, does spacetime look Lorentzian? This has to do with Ehrenfest's principle: the WKB wavefunction will be concentrated on a manifold solving the Einstein equations. At both the classical and semi-classical level, metric signature depends on initial conditions. An initial 3-manifold can evolve via Einstein's equations into *either* a Lorentzian or Riemannian 4-manifold, depending on the initial choice of conjugate momenta. At the quantum level, the "average" metric signature depends on the initial state.

THE COSMOLOGICAL CONSTANT

In the spirit of third quantization,[8] we consider the effective super-gravitational action at one loop, where the "particle" going around the loop is, in this case, a virtual universe. This has the form

$$S_{eff}[\mathcal{G}_{ab}] = \frac{i\hbar}{2} \text{Trln}[\text{Æ} + \mathcal{M}^2] \tag{74}$$

where the trace runs over a basis of states satisfying the one-parameter family of Wheeler-DeWitt equations $H^\varepsilon \Phi_{\varepsilon\beta} = 0$. However, the supermetric G_{ab}, unlike the ordinary spacetime metric $g_{\mu\nu}$ is not arbitrary; it is constrained to be a particular functional of $U(q)$ and $g_{\mu\nu}$. But the form of $U(q)$ is *also* tightly constrained: it is the sum

of all possible potential terms that could appear in an ADM Hamiltonian. With this restriction, S_{eff} is just a function of the coupling constants of each possible interaction term, i.e.

$$S_{eff}[\mathcal{G}_{ab}] = S_{eff}[\lambda, e^2, g^2, ...] \tag{75}$$

and the couplings are now viewed as dynamical variables. Variation of S_{eff} with respect to the couplings could, in principle, determine their phenomenological values, very much in the spirit of Coleman's "Big Fix."[9]

As an illustration, consider the minisuperspace action representing a Friedman-Robertson-Walker (FRW) universe filled with a three-component, minimally coupled massless scalar field $\vec{\phi} \doteq (\phi_1, \phi_2, \phi_3)$, i.e.

$$S = \frac{1}{2}\int dt \left[-\frac{a\dot{a}^2}{N} + \frac{a^3 \dot{\vec{\phi}} \cdot \dot{\vec{\phi}}}{N} + N(a - \lambda a^3) \right] \tag{76}$$

With the choice of coordinates $q^0 = a$, $q^i = \phi^i$, the supermetric for the corresponding worldline action

$$S = -\mathcal{M}\int d\tau \sqrt{-\mathcal{G}_{ab}\dot{q}^a \dot{q}^b} \tag{77}$$

reads

$$\mathcal{G}_{\phi_i \phi_i} = -a^2 \mathcal{G}_{aa} = a^4(\lambda a^2 - 1) \quad ; \qquad i = 1,2,3 \tag{78}$$

As in ordinary gravity, the trace log in (74) has an adiabatic ("weak curvature") expansion

$$S_{eff}[\mathcal{G}_{ab}] = \int da \int d\vec{\phi}\, \sqrt{|\mathcal{G}|}\, [\Lambda_S + \kappa_S \mathcal{R} + ...] \tag{79}$$

Let us temporarily compactify the ranges of integration so that the scale factor runs from $a = 0$ to $a = \bar{a}$, and keep only the leading term in the adiabatic expansion

$$\begin{aligned} S_{eff} &\simeq \Lambda_S \int_0^{\bar{a}} da \int d^3\phi\; a^7(\lambda a^2 - 1)^2 \\ &= I_\phi \Lambda_S \bar{a}^8 \left(\frac{\lambda^2 \bar{a}^4}{12} - \frac{\lambda \bar{a}^2}{5} + \frac{1}{8} \right) \end{aligned} \tag{80}$$

where I_ϕ is the (compactified) ϕ-integral. Then it is easy to check that S_{eff} is stationary at

$$\frac{dS_{eff}}{d\lambda} = 0 \quad \Longrightarrow \quad \lambda = \frac{6}{5\bar{a}^2} \tag{81}$$

with the result that $\lambda \to 0^+$ as $\bar{a} \to \infty$.

Although the algebra is simplest in this case, the result seems to be fairly robust, and and survives inclusion of mass terms, supercurvature, and varying number of scalar fields.[1]

A NEW SOURCE OF DECOHERENCE?

As a second speculation, we may imagine the possibility that the Universe propagates slightly off-shell; the wavefunction of the Universe being a superposition of states with different values of $\varepsilon \neq \varepsilon_{av} = M^2$. The finite uncertainty $\Delta\varepsilon$ will be reflected in an apparent dispersion $\delta\hbar$ in the "effective" value of Planck's constant $\hbar_{eff}(\varepsilon) = \hbar/\sqrt{\mathcal{E}}$, i.e.

$$\delta\hbar = \left| \frac{d}{d\mathcal{E}} \hbar_{eff} \right|_{\mathcal{E}=\mathcal{M}^2} \Delta\mathcal{E} = \frac{1}{2}\hbar_{eff}\frac{\Delta\mathcal{E}}{\mathcal{E}} \tag{82}$$

If such a dispersion exists, it would manifest itself as a new source of decoherence between paths whose difference in action δS exceeds a certain bound. By arguments analogous to those in physical optics, the criterion for coherent interference between paths of action difference δS is[1]

$$\frac{\delta S}{\hbar_{eff}} < \frac{\hbar_{eff}}{\delta\hbar} \tag{83}$$

The signature of finite dispersion $\delta\hbar$ in the effective value of Planck's constant could be, e.g., an observed decoherence of particle beams in an ultra-sensitive particle interferometer, in a situation where standard time-energy considerations would imply that the beams should interfere coherently. Of course, we have no idea whether there really is a finite dispersion $\delta\hbar \neq 0$, nor, if there is such a dispersion, whether it is large enough to be detectable.

ACKNOWLEDGEMENTS

I would like to thank the organizers of the "New Developments" meeting for inviting me to Zakopane. I am also grateful for the hospitality of the high-energy theory group at Lawrence Berkeley National Laboratory, where some of this work was carried out. This research was supported in part by the U.S. Department of Energy, under Grant No. DE-FG03-92ER40711. Support has also been provided by Carlsbergfondet.

REFERENCES

1. A. Carlini and J. Greensite, Phys. Rev. D55 (1997) 3514.
2. J. Greensite, Class. Quantum Grav. 13 (1996) 1339.
3. A. Carlini and J. Greensite, Phys. Rev. D52 (1995) 6947.
4. V. Moncreif and C. Teitelboim, Phys. Rev. 6 (1972) 966.
5. R.F. Baierlein, D.H. Sharp, and J.A. Wheeler, Phys. Rev. 126 (1962) 1864;
   J. A. Wheeler, in *Relativity, Groups, and Topology*, ed. C. DeWitt and B. DeWitt (Gordon and Breach, New York, 1964).
6. J. Ambjørn, archive hep-th/9411179.
7. K. Kuchar, in *Proc. 4th Canadian Conference on General Relativity and Astrophysics*, ed. G. Kunstatter et al. (Singapore, World Scientific, 1992)
8. S. Giddings and A. Strominger, Nucl. Phys. B 321 (1989) 481;
   V. Rubakov, Phys. Lett. B214 (1988) 503;
   M. McGuigan, Phys. Rev. D38 (1988) 3031.
9. S. Coleman, Nucl. Phys. B310 (1988) 668.

This page intentionally left blank

CENTER DOMINANCE, CENTER VORTICES, AND CONFINEMENT

L. Del Debbio,[1] M. Faber,[2] J. Greensite,[3]* and Š Olejník[4]

[1]CPT, CNRS Luminy, Case 907, 13288 Marseille Cedex 9, France
[2]Inst. für Kernphysik, Tech. Univ. Wien, 1040 Vienna, Austria
[3]The Niels Bohr Institute, DK-2100 Copenhagen Ø, Denmark
[4]Inst. of Phys., Slovak Acad. of Sci., 842 28 Bratislava, Slovakia

INTRODUCTION

The results that I would like to discuss here are a collection of numerical data which strongly favor an old and, in recent years, somewhat neglected theory of quark confinement: the Z_N Vortex Condensation Theory. Some of this data (Figs. 1-8) was reported by our group late last year,[1] the rest is new.

The confinement region of an $SU(N)$ gauge theory really consists of at least two parts. The first is an intermediate distance region, extending from the onset of the linear potential up to some color-screening distance, which we call the **Casimir-Scaling regime.**[2,3] Many numerical experiments have shown that in this intermediate region flux tubes form, and a linear potential is established, between heavy quarks in any non-trivial representation of the gauge group. The string-tension is representation-dependent, and appears to be roughly proportional to the quadratic Casimir of the representation.[4] Thus, for an $SU(2)$ gauge theory,

$$\sigma_j \approx \frac{4}{3} j(j+1)\sigma_{1/2} \tag{1}$$

where σ_j is the string tension for a heavy quark-antiquark pair in representation j. Eventually, however, the color charge of higher-representation quarks must be screened by gluons, and the asymptotic string tension can then only depend on the transformation properties of the quarks under the center of the gauge group, i.e. on the "n-ality" of the representation. This **Asymptotic regime** extends from the color-screening length to infinity, and in the case of an $SU(2)$ gauge group the string tensions must satisfy

$$\sigma_j = \begin{cases} \sigma_{1/2} & j = \text{half-integer} \\ 0 & j = \text{integer} \end{cases} \tag{2}$$

In particular, the string between quarks in an adjoint representation must break, at some distance which presumably depends on the mass of "gluelumps" (i.e. the energy

*Talk presented by J. Greensite.

New Developments in Quantum Field Theory
Edited by Damgaard and Jurkiewicz, Plenum Press, New York, 1998

of a gluon bound to a massive adjoint quark). Also, since string-breaking is a $1/N^2$ suppressed process, the number of colors is relevant. The breaking of the adjoint string is difficult to observe in numerical experiments, although on general theoretical grounds one may be confident that the breaking *must* occur for sufficiently large adjoint quark separation.

The most popular theory of quark confinement is the abelian projection theory proposed by 't Hooft, which I will briefly describe in the next section. In past years our group has been highly critical of this theory (as well as the Z_N vortex theory), on the grounds that it fails to explain the existence of a linear potential between higher representation quarks in the Casimir scaling regime.[2, 3] This failure is very significant, because it is in the Casimir regime that the confining force replaces Coulombic behavior, and in fact it is only in this regime that the QCD string has been well studied numerically. If we don't understand Casimir scaling, then we don't really understand how flux tubes form.

A possible response to this criticism is simply to admit that the formation of flux tubes, at intermediate distances, remains to be understood, but that the abelian projection theory is nonetheless valid at very large distance scales, i.e. in the asymptotic regime. I will argue that there may be some truth to this response, but that the confining configurations relevant to the asymptotic regime seem to be Z_N vortices, rather than abelian monopoles.

ABELIAN DOMINANCE

One of the earliest ideas about confinement, known as "dual-superconductivity," was put forward independently by 't Hooft and Mandelstam in the mid-1970's. The idea is that the QCD vacuum resembles a superconductor with the roles of the $\vec{E}$ and $\vec{B}$ fields interchanged. Electric (rather than magnetic) fields are squeezed into vortices; electric (rather than magnetic) charges are confined. Magnetic monopoles are condensed; they play the role of the electrically charged Cooper pairs. The problem is to actually identify the magnetic monopoles of an unbroken non-abelian gauge theory, and to understand which non-abelian degrees of freedom play the role of electromagnetism.

A concrete suggestion along these lines was made by 't Hooft in 1981.[5] The proposal was to gauge fix part of the $SU(N)$ symmetry by diagonalizing some operator tranforming in the adjoint representation of the gauge group. This leaves a remnant $U(1)^{N-1}$ gauge symmetry, with gauge transformations g of the form

$$g = \mathrm{diag}[e^{i\alpha_1}, e^{i\alpha_2}, ..., e^{i\alpha_N}] \qquad \sum \alpha_n = 0 \tag{3}$$

The diagonal components of the vector potential, A_μ^{aa}, transform under the residual symmetry like abelian gauge fields, i.e.

$$\mathcal{A}_\mu'^{aa} = \mathcal{A}_\mu^{aa} + \partial_\mu \alpha^a \tag{4}$$

while the off-diagonal components transform like double (abelian) charged matter fields

$$\mathcal{A}_\mu'^{ab} = e^{i(\alpha^a - \alpha^b)} \mathcal{A}_\mu^{ab} \tag{5}$$

This gauge-fixed theory can therefore be regarded as an abelian gauge theory of "photons," charged matter fields, and magnetic monopoles. Monopole condensation confines abelian charged objects, and the abelian electric field forms a flux tube.

On the lattice, one can decompose the link variables U

$$U_\mu(x) = W_\mu(x) A_\mu(x) \tag{6}$$

into "abelian" link variables A, transforming under the residual symmetry as abelian gauge fields, and "matter" fields W

$$A'_\mu(x) = g(x)A_\mu(x)g^{-1}(x+\hat{\mu})$$
$$W'_\mu(x) = g(x)W_\mu(x)g^{-1}(x) \tag{7}$$

For $SU(2)$ lattice gauge theory, A is simply the diagonal part of U, rescaled to restore unitarity, i.e.

$$U = a_0 I + i\vec{a}\cdot\vec{\sigma} \quad , \qquad A = \frac{a_0 + ia_3\sigma^3}{\sqrt{a_0^2 + a_3^2}} \tag{8}$$

Monte Carlo studies of the abelian projection theory began with the work of Kronfeld et al., [6] who introduced a specific abelian projection gauge, the "maximal abelian gauge," [7] which has been used in most further studies. The maximal abelian gauge is defined as the gauge which maximizes the quantity

$$\sum_x \sum_\mu \mathrm{Tr}[U_\mu(x)\sigma^3 U_\mu^\dagger(x)\sigma^3] \tag{9}$$

This requires diagonalizing, at every site, the adjoint representation operator

$$X(x) = \sum_\mu [U_\mu(x)\sigma^3 U_\mu^\dagger(x) + U_\mu^\dagger(x-\hat{\mu})\sigma^3 U_\mu(x-\hat{\mu})] \tag{10}$$

This gauge choice makes the link variables as diagonal as possible, placing most of the quantum fluctuations in the abelian link variables. If the abelian projection idea is going to work at all, it ought to work best in this gauge. Other proposals (Polyakov-line gauge, Field-Strength gauge) have not, in fact, been very successful.

An important development was the finding, by Suzuki and collaborators, that if we fix to maximal abelian gauge and replace the full link variables U with the abelian link variables A (this is often termed "abelian projection"), and then calculate such quantities as Creutz ratios, Polyakov lines, etc., with the abelian links, the results very closely approximate those obtained with the full link variables.[8] The fact that the abelian link variables seem to carry most of the information about the infrared physics is known as "**abelian dominance,**" and it has stimulated a great deal of further work on the abelian projection theory.

CENTER DOMINANCE

Of course the abelian projection theory is not the *only* proposal for explaining the confining force; there have been many other suggestions over the years. One idea that was briefly popular in the late 1970's was the Vortex Condensation theory, put forward, in various forms, by 't Hooft,[9] Mack,[10] and by Nielsen and Olesen[11] (the "Copenhagen Vacuum"). The idea is that the QCD vacuum is filled with closed magnetic vortices, which have the topology of tubes (in 3 Euclidean dimensions) or surfaces (in 4 dimensions) of finite thickness, and which carry magnetic flux in the center of the gauge group (hence "center vortices"). The effect of creating a center vortex linked to a given Wilson loop, in an $SU(N)$ gauge theory, is to multiply the Wilson loop by an element of the gauge group center, i.e.

$$W(C) \to e^{i2\pi n/N} W(C) \qquad n = 1, \ldots, N-1 \tag{11}$$

Quantum fluctuations in the number of vortices linked to a Wilson loop can be shown to lead to an area law falloff, assuming that center vortex configurations are condensed in the vacuum.†

With one notable exception,[13] almost nothing has been done with this idea since the early 1980's, which was at the dawn of Monte Carlo lattice gauge simulations. It is therefore interesting to go back and study the vortex theory, using a numerical approach inspired by studies of the abelian projection theory.

In an $SU(2)$ lattice gauge theory, we begin by fixing to maximal abelian gauge and then go one step further, using the remnant $U(1)$ symmetry to bring the abelian link variables

$$A = \begin{bmatrix} e^{i\theta} & \\ & e^{-i\theta} \end{bmatrix} \tag{12}$$

as close as possible to the $SU(2)$ center elements $\pm I$ by maximizing $< \cos^2\theta >$, leaving a remnant Z_2 symmetry. This is the **(indirect) Maximal Center Gauge** (the center is maximized in A, rather than directly in U). We then define at each link

$$Z \equiv \text{sign}(\cos\theta) = \pm 1 \tag{13}$$

which is easily seen to transform like a Z_2 gauge field under the remnant Z_2 symmetry. "Center Projection" is the replacement $U \to Z$ of the full link variables by the center variables; we can then calculate Wilson loops, Creutz ratios, etc. with the center-projected Z-link variables.

Figure 1 is a plot of Creutz ratios vs. β, extracted from the center-projected configurations. The straight line is the asymptotic freedom prediction

$$\sigma a^2 = \frac{\sigma}{\Lambda^2}\left(\frac{6\pi^2}{11}\beta\right)^{102/121} \exp\left[-\frac{6\pi^2}{11}\beta\right] \tag{14}$$

with the value $\sqrt{\sigma}/\Lambda = 67$. What is remarkable about this plot, apart from the scaling, is that the Creutz ratios $\chi(R, R)$ at each β are almost independent of R. This means that the center projection sweeps away the Coulombic contribution, and the linear potential appears already at $R = 2$. This is seen quite clearly in Fig. 2, which compares the center-projected Creutz ratios (solid line), at $\beta = 2.4$, with the Creutz ratios of the full theory (dashed line). It is also interesting to compute the Creutz ratios derived from abelian link variables with the Z variable factored out, i.e. A/Z (dotted line). We note that, in this case, the string tension simply disappears.

It seems evident from this data that, just as the abelian A links are the crucial part of the full U link variables in maximal abelian gauge, so the Z center variables are the crucial part of the A links in maximal center gauge, carrying most of the information about the string tension. This is what we mean by "Center Dominance."

Should one then interpret center dominance to mean that the confining force is due to Z_2 center vortices, rather than $U(1)$ monopoles? That conclusion would be premature, in our view. In fact, our original interpretation of this data was that the success of center dominance suggests that *neither* abelian dominance *nor* center dominance has anything very convincing to say about quark confinement (and this fits very nicely with our further critique of abelian projection based on Casimir scaling).[2] Underlying that interpretation, however, was the belief that the "thin" Z_2 vortices of the center-projected configurations are probably irrelevant to the confining properties of the full, unprojected configurations. This belief is testable, however, and the result of the test is surprising.

†Some related ideas have also been put forward by Chernodub et al.[12]

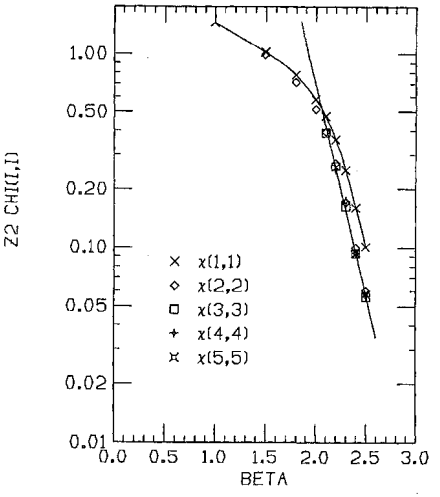


Figure 1. Creutz ratios from center-projected lattice configurations, in the (indirect) maximal center gauge.

VORTEX-LIMITED WILSON LOOPS

The only excitations of Z_2 lattice gauge theory with non-zero action are "thin" Z_2 vortices, which have the topology of a surface (one lattice spacing thick) in D=4 dimensions. We will call the Z_2 vortices, of the center projected Z-link configurations, "Projection-vortices" or just **P-vortices.** These are to be distinguished from the hypothetical "thick" center vortices, which might exist in the full, unprojected U configurations. A plaquette is pierced by a P-vortex if, upon going to maximal center gauge and center-projecting, the projected plaquette has the value –1. Likewise, a given lattice surface is pierced by n P-vortices if n plaquettes of the surface are pierced by P-vortices.

In a Monte Carlo simulation, the number of P-vortices piercing the minimal area of a given loop C will, of course, fluctuate. Let us define $W_n(C)$ to be the Wilson loop evaluated on a sub-ensemble of configurations, selected such that precisely n P-vortices, in the corresponding center-projected configurations, pierce the minimal area of the loop. It should be emphasized here that the center projection is used only to select the data set. The Wilson loops themselves are evaluated using the full, *unprojected* link variables. In practice, to compute $W_n(C)$, the procedure is to generate thermalized lattice configurations by the usual Monte Carlo algorithm, and fix to maximal center gauge by over-relaxation. For each independent configuration one then examines each rectangular loop on the lattice of a given size; those with n P-vortices piercing the loop are evaluated, the others are skipped. Creutz ratios $\chi_n(I, J)$ can then be extracted from the vortex-limited Wilson loops $W_n(C)$. In particular, if the presence or absence of P-vortices in the projected configuration is unrelated to the confining properties of the corresponding unprojected configuration, then we would expect

$$\chi_0(I, J) \approx \chi(I, J) \tag{15}$$

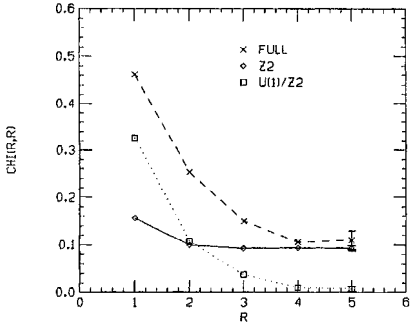


Figure 2. Creutz ratios $\chi(R, R)$ vs. R at $\beta = 2.4$ for full, center-projected, and $U(1)/Z_2$-projected lattice configurations.

at least for large loops.

The result of this test is shown in Fig. 3. Quite contrary to our expectations, the confining force vanishes if P-vortices are excluded. This does not necessarily mean that the confining configurations of $SU(2)$ lattice gauge theory are thick center vortices. It does imply, however, that the presence or absence of P-vortices in the projected gauge field is strongly correlated with the presence or absence of confining configurations (whatever they may be) in the unprojected gauge field.

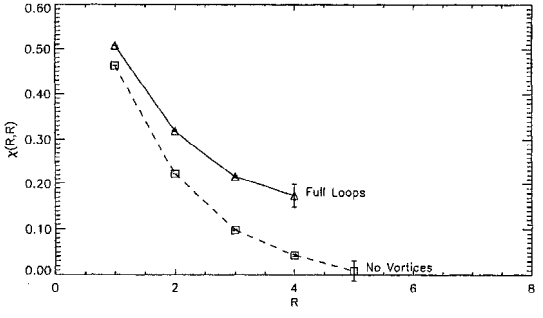


Figure 3. Creutz ratios $\chi_0(R, R)$ extracted from loops with no P-vortices, as compared to the usual Creutz ratios $\chi(R, R)$, at $\beta = 2.3$.

The next question is whether we can rule out the possibility that the confining configurations are, in fact, thick Z_2 center vortices. Suppose, for a moment, that to each P-vortex in the projected Z-link gauge field there corresponds a thick center vortex in the associated, unprojected, U-link gauge field. If that is the case, then in the limit of large loop area we expect

$$\frac{W_n(C)}{W_0(C)} \longrightarrow (-1)^n \tag{16}$$

The argument for this equation is as follows: Vortices are created by discontinuous

gauge transformations. Suppose loop C, parametrized by $x^\mu(\tau)$, $\tau \in [0, 1]$, encircles n vortices. At the point of discontinuity

$$g(x(0)) = (-1)^n g(x(1)) \tag{17}$$

The corresponding vector potential, in the neighborhood of loop C can be decomposed as

$$A_\mu^{(n)}(x) = g^{-1}\delta A_\mu^{(n)}(x)g + ig^{-1}\partial_\mu g \tag{18}$$

so that

$$\begin{aligned} W_n(C) &= < \mathrm{Tr}\exp[i\oint dx^\mu A_\mu^{(n)}] > \\ &= (-1)^n < \mathrm{Tr}\exp[i\oint dx^\mu \delta A_\mu^{(n)}] > \end{aligned} \tag{19}$$

In the region of the loop C, the vortex background looks locally like a gauge transformation. If all other fluctuations $\delta A_\mu^{(n)}$ are basically short-range, then they should be oblivious, in the neighborhood of the loop C, to the presence or absence of vortices in the middle of the loop. In that case, if we have correctly identified the vortex contribution, then

$$< \mathrm{Tr}\exp[i\oint dx^\mu \delta A_\mu^{(n)} > \approx < \mathrm{Tr}\exp[i\oint dx^\mu \delta A_\mu^{(0)}] > \tag{20}$$

for sufficiently large loops, and eq. (16) follows immediately. All we have to do is test this.

Figures 4 and 5 show our data for W_1/W_0 and W_2/W_0, respectively, at $\beta = 2.3$. Again, somewhat to our surprise, this data is entirely consistent with (16); it is consistent with the confining field configurations being center vortices, and in fact offers good evidence in favor of that possibility.

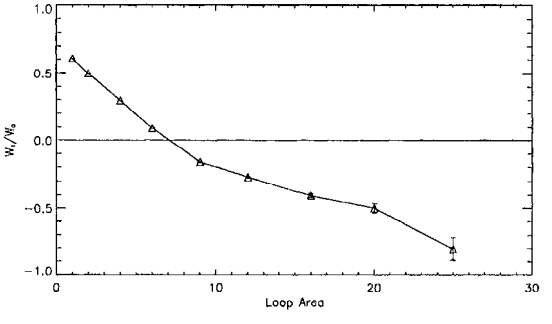


Figure 4. Ratio of the 1-Vortex to the 0-Vortex Wilson loops, $W_1(C)/W_0(C)$, vs. loop area at $\beta = 2.3$.

Of course, it could still be that we are looking at a rather small (and perhaps misleading) sample of the data, at least for the larger loops. Large loops will tend to be pierced by large numbers of P-vortices. As the area of a loop increases, the fraction of configurations in which no P-vortex (or exactly one, or exactly two P-vortices) pierces

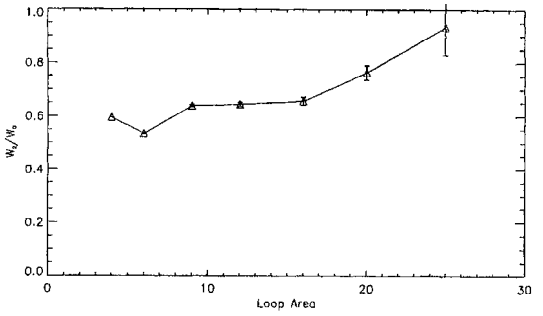


Figure 5. Ratio of the 2-Vortex to the 0-Vortex Wilson loops, $W_2(C)/W_0(C)$, vs. loop area at $\beta = 2.3$.

the loop will decrease, tending to zero in the limit. So let us instead consider $W_{evn}(C)$ and $W_{odd}(C)$, where $W_{evn}(C)$ denotes Wilson loops evaluated in configurations with an even (including zero) number of P-vortices piercing the loop, and $W_{odd}(C)$ denotes the corresponding quantity for odd numbers. Then

$$W(C) = P_{evn}(C)W_{evn}(C) + P_{odd}(C)W_{odd}(C) \tag{21}$$

where

$P_{evn}(C)$ = the fraction of configurations with an even (or zero) number of P-vortices piercing loop C

$P_{odd}(C)$ = the fraction of configurations with an odd number of P-vortices piercing loop C

One expects that for large loops, $P_{evn} \approx P_{odd} \approx 0.5$. According to the vortex condensation mechanism, neither W_{evn} nor W_{odd} falls with an area law; the area-law falloff is due to a delicate cancellation between the two terms in eq. (21). As loops become large, one should find $W_{odd} \to -W_{evn}$. The data, shown below in Figures 6-8, support these expectations. This time we are using essentially all of the data, since about half contributes to $W_{evn}(C)$, and the rest to $W_{odd}(C)$.

DIRECT MAXIMAL CENTER GAUGE

Along with the successes, there is one significant failure of center dominance in the data shown in Fig. 1. Despite the nice scaling of the data, the value of $\sqrt{\sigma}/\Lambda = 67$ is a little high, and in fact the center projected Creutz ratios are all significantly higher than the asymptotic string tension extracted from unprojected configurations, using "state-of-the-art" noise reduction techniques.

On the other hand, it is not so clear that the "indirect" maximal center gauge is the *true* maximal center gauge. What we have done up until now is to first fix to maximal abelian gauge, and then bring the abelian part A of link U as close as possible

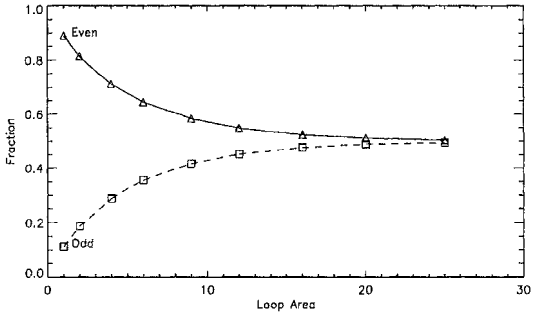


Figure 6. Fraction of link configurations containing even/odd numbers of P-vortices, at $\beta = 2.3$, piercing loops of various areas.

to $\pm I$. However, since we are emphasizing the role of the gauge group center, rather than the $U(1)$ subgroup, it really makes more sense to choose a gauge in which the entire link variable U is brought as close as possible to the center elements $\pm I$. With this motivation, let us define the **(direct) Maximal Center Gauge** of an $SU(N)$ gauge theory as the gauge which maximizes the quantity

$$Q = \sum_x \sum_\mu \mathrm{Tr}[U_\mu(x)]\mathrm{Tr}[U_\mu^\dagger(x)] \tag{22}$$

For the $SU(2)$ gauge group, we define

$$Z = \mathrm{sign}(\mathrm{Tr}[U]) \tag{23}$$

as the center-projected link variables; these again transform like Z_2 gauge fields under the remnant Z_2 gauge symmetry.

Using the direct maximal center gauge, we find the following results: Qualitatively, things look about the same, and plots of W_n / W_0, and W_{evn} vs. W_{odd}, look virtually identical to the previous data in the indirect maximal center gauge. Quantitatively, however, there is an improvement. We find that string tensions extracted from the center projection in the "direct" gauge are in much better agreement with the asymptotic string tension of the full theory, extracted by "state-of-the-art" methods. Figure 9 shows a plot of Creutz ratios vs. β The straight line is the usual scaling curve, but this time with a value $\sqrt{\sigma}/\Lambda = 58$. Figures 10-12 plot the center-projected Creutz ratios $\chi(R, R)$ at $\beta = 2.3,\ 2.4,\ 2.5$ respectively. The triangles are our data. The solid line is the asymptotic string tension of the unprojected configurations at these values of β, quoted by Bali et al.[14] The dashed lines are the error bars on the asymptotic string tension, which we have also taken from this reference.

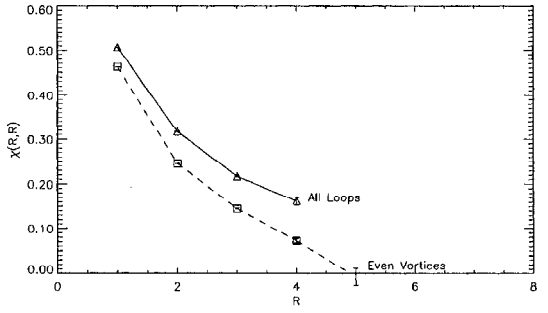


Figure 7. Creutz ratios $\chi_{evn}(R, R)$ extracted from Wilson loops $W_{evn}(C)$, taken from configurations with even numbers of P-vortices piercing the loop. The standard Creutz ratios $\chi(R, R)$ at this coupling (β = 2.3) are also shown.

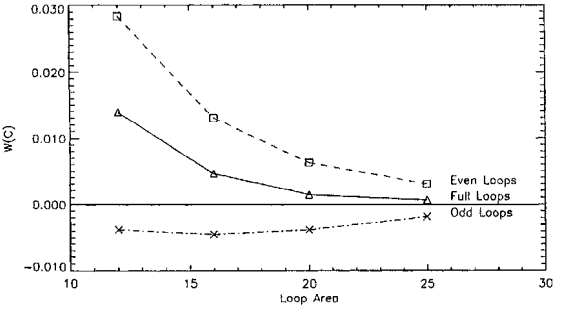


Figure 8. Wilson loops $W_{evn}(C)$, $W_{odd}(C)$ and $W(C)$ at larger loops areas, taken from configurations with even numbers of P-vortices, odd numbers of P-vortices, and any number of P-vortices, respectively, piercing the loop. Again β = 2.3.

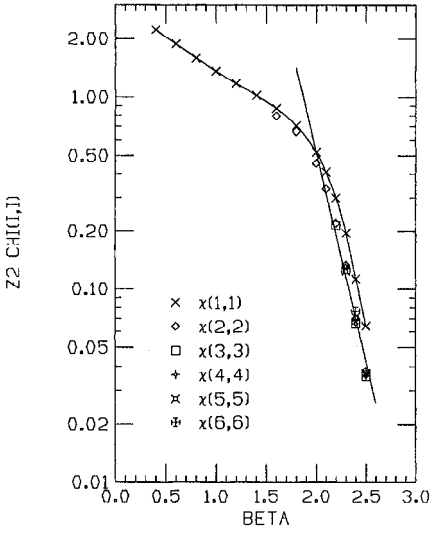


Figure 9. Creutz ratios from center-projected lattice configurations, in the direct maximal center gauge.

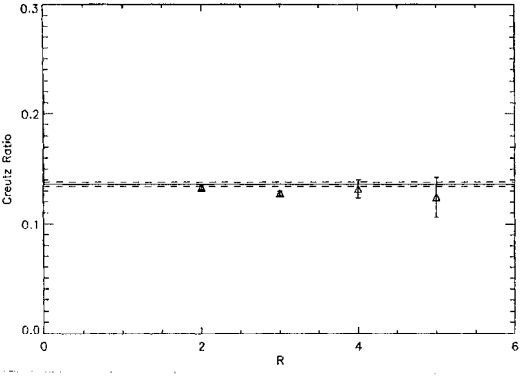


Figure 10. Center-projection Creutz ratios $\chi(R, R)$ vs. R at $\beta = 2.3$; direct maximal center gauge. Triangles are our data points. The solid line shows the value of the asympotic string tension of the unprojected configurations, and the dashed lines the associated error bars, quoted in Bali et al.[14]

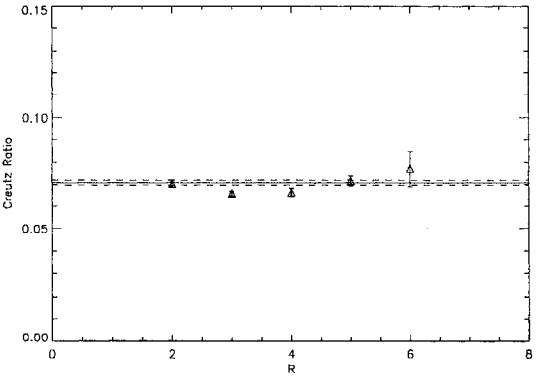


Figure 11. Same as Fig. 10, at $\beta = 2.4$.

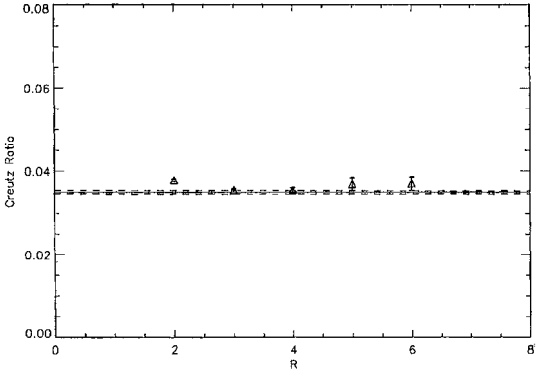
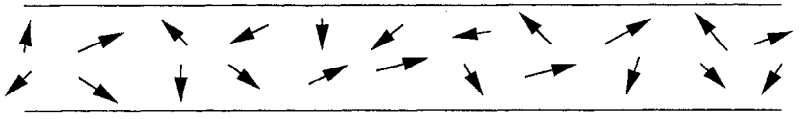


Figure 12. Same as Fig. 10, at $\beta = 2.5$.

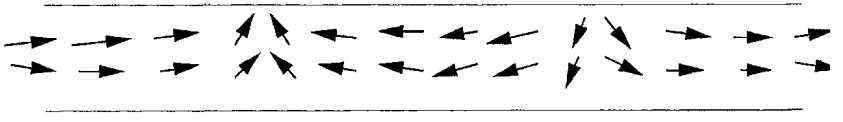
VORTICES VS. MONOPOLES

There is no denying that the data shown here, in support of the vortex condensation theory, is a little reminiscent of the data that has been put forward in support of the abelian projection theory. This raises a natural question: If the Yang-Mills vacuum is dominated, at long wavelengths, by Z_2 vortex configurations, then how do we explain the numerical successes of the abelian projection in maximal abelian gauge? In our opinion, the probable answer to this question is that a center vortex configuration, transformed to maximal abelian gauge and then abelian-projected, will appear as a chain of monopoles alternating with antimonopoles. These monopoles are essentially an artifact of the projection; they are condensed because the long vortices from which they emerge are condensed.

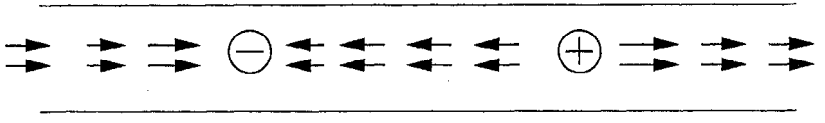
A little more graphically, the picture is as follows: Consider a center vortex at some constant time *t*. This time-slice of a thick vortex is then a tube of magnetic flux. Before gauge-fixing, the field-strength inside this tube points in arbitrary directions in color space



Fixing to maximal abelian gauge, the field strength tends to line up mainly (but not entirely) in the diagonal ($\pm\sigma^3$) color direction

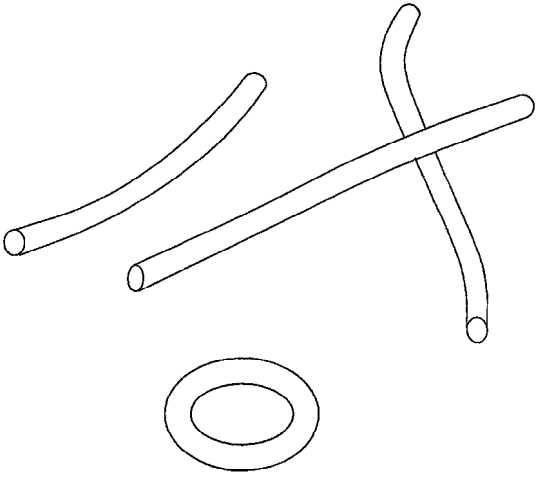


Upon abelian projection, the regions interpolating between $+\sigma^3$ and $-\sigma^3$ emerge as "monopoles." Their location is gauge (and Gribov copy) dependent.

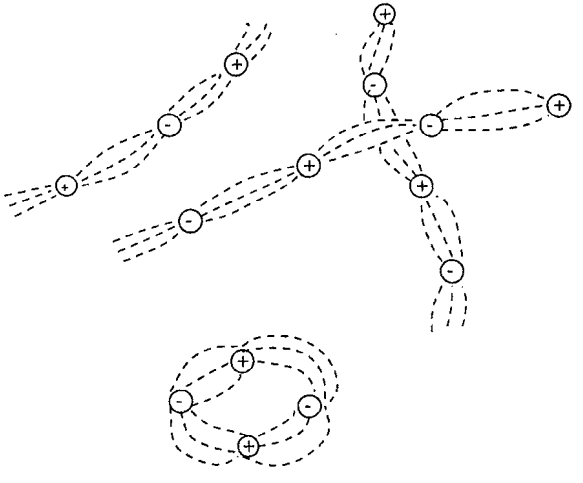


It is not difficult to construct examples of center vortices which behave in just this way, i.e. which are converted to monopole-antimonopole chains upon abelian projection in maximal abelian gauge.

If this picture is accurate, then the “spaghetti vacuum”

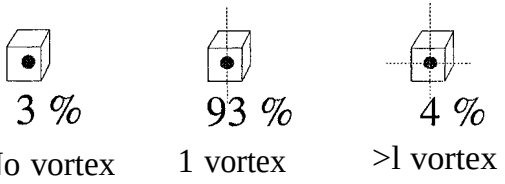


appears, under abelian projection, as a “monopole vacuum”



We have, in fact, obtained some preliminary evidence for this picture from Monte Carlo simulations. These simulations were carried out at $\beta = 2.4$, in the (indirect) maximal center gauge. We look at sites where the monopoles are "static," i.e. the monopole current is $j_0 = \pm 1$, $\vec{j} = 0$. The monopole charge is enclosed in a cube bounded by spacelike plaquettes. We find that:

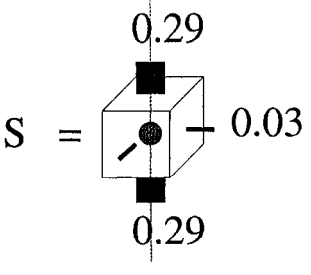
I: Almost all (93%) of monopole cubes are pierced by one, and only one, P-vortex.



II: The action of a monopole cube, pierced by a P-vortex, is highly asymmetric. Almost all the plaquette action

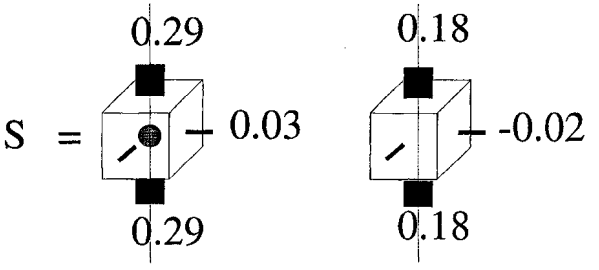
$$S = (1 - \frac{1}{2}\mathrm{Tr}[UUU^{\dagger}U^{\dagger}]) - S_0 \tag{24}$$

above the lattice average S_0, is oriented in the direction of the P-vortex. On each of the two plaquettes pierced by the P-vortex, at $\beta = 2.4$, the average action above S_0 is $S = 0.29$. On each of the four plaquettes which are not pierced by the vortex, $S = 0.03$ on average.‡



‡ Bakker et al. [15] have also studied the excess action of monopole cubes (but not the correlation with P-vortices) in maximal abelian gauge.

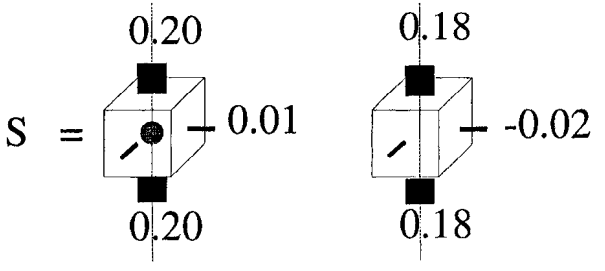
III: The (unprojected) action distribution of a monopole cube, pierced by a P-vortex, is similar to the action distribution of any other cube pierced by a P-vortex...



One Monopole
One Vortex

No Monopoles
One Vortex

...especially when we look at "isolated" monopoles (no neighboring monopole currents)



One Monopole
One Vortex

No Monopoles
One Vortex

In summary, abelian monopoles tend to lie along P-vortices. Isolated monopoles are hardly distinguished, in their (unprojected) field strength distribution, from other regions along the P-vortices.

This is in accordance with our intuitive picture.

CONCLUSIONS

We have developed a technique for locating center vortices in thermalized lattice gauge configurations, and have found evidence that center vortices account for the asymptotic string tension between static, fundamental representation, color charges. A "spaghetti vacuum" picture appears to be correct at sufficiently large scales.

On the other hand, string formation at intermediate distances, in the Casimir scaling regime, remains to be understood. This is a very important issue, especially since the Casimir scaling regime extends to infinity as $N_{colors} \to \infty$.[16] Casimir scaling suggests that center vortices, although they may be the crucial configurations asymptotically, are not the whole story. Since adjoint loops are oblivious to the gauge-group center, one may speculate that there are other types of configurations which contribute to the adjoint string tension. Or, possibly, the finite thickness and detailed inner structure of center vortices is a relevant issue, since adjoint loops which intersect the "core" of a center vortex *will* be affected by the vortex. Perhaps the gluon-chain model,[17] which I proposed some time ago, might be helpful in understanding the dynamics of the Casimir-scaling region.

We are currently in the process of repeating all our calculations for $SU(3)$ lattice gauge theory, and have already found evidence of center dominance on small lattices at strong couplings. If we also find that (i) center dominance persists on larger lattices at weaker couplings; (ii) the absence of P-vortices results in vanishing string tension, and (iii)

$$\frac{W_n(C)}{W_0(C)} \longrightarrow e^{2n\pi i/3} \tag{25}$$

then the combined evidence in favor of some version of the Z_N vortex condensation theory will be quite compelling.

One final note: Shortly after the Zakopane meeting, Tomboulis and Kovács reported on some new Monte Carlo. data they have obtained in support of the vortex condensation theory.[18] Their results are quite consistent with the work I have presented here.

ACKNOWLEDGEMENTS

I would like to thank the organizers of the "New Developments" meeting for inviting me to Zakopane. I am also grateful for the hospitality of the high-energy theory group at Lawrence Berkeley National Laboratory, where some of this work was carried out. This research was supported in part by the US. Department of Energy, under Grant No. DE-FG03-92ER40711. Support has also been provided by Carlsbergfondet.

REFERENCES

1. L. Del Debbio, M. Faber, J. Greensite, and Š. Olejník, Phys. Rev. D55 (1997) 2298, hep-lat/9610005.
2. L. Del Debbio, M. Faber, J. Greensite, and Š. Olejník, Nucl. Phys, Proc. Suppl. 53 (1997) 141, hep-lat/9607053.
3. L. Del Debbio, M. Faber, J. Greensite, and Š. Olejník, Phys. Rev. D53 (1996) 5897, hep-lat/9510028.
4. J. Ambjørn, P. Olesen, and C. Peterson, Nucl. Phys. B240 [FS12] (1984) 198; 533;
   C. Michael, Nucl. Phys. Proc. Suppl. 26 (1992) 417; Nucl. Phys. B259 (1985) 58;
   N. Cambell, L. Jorysz, and C. Michael, Phys. Lett. B167 (1986) 91;
   M. Faber and H. Markum, Nucl. Phys. Proc. Suppl. 4 (1988) 204;

M. Müller, W. Beirl, M. Faber, and H. Markum, Nucl. Phys. Proc. Suppl. 26 (1992) 423;
G. Poulis and H. Trottier, hep-lat/9504015.
5. G. 't Hooft, Nucl. Phys. B190 [FS3] (1981) 455.
6. A. Kronfeld, G. Schierholz, and U.-J. Wiese, Nucl. Phys. B293 (1987) 461.
7. A. Kronfeld, M. Laursen, G. Schierholz, and U.-J. Wiese, Phys. Lett. B198 (1987) 516.
8. T. Suzuki and Yotsuyanagi, Phys. Rev. D42 (1990) 4257;
S. Hioki et al., Phys. Lett. B272 (1991) 326.
9. G. 't Hooft, Nucl. Phys. B153 (1979) 141.
10. G. Mack, in *Recent Developments in Gauge Theories*, edited by G. 't Hooft et al. (Plenum, New York, 1980).
11. H. Nielsen and P. Olesen, Nucl. Phys. B160 (1979) 380.
12. M. Chernodub, M. Polikarpov, and M. Zubkov, Nucl. Phys. Proc. Suppl. 34 (1994) 256.
13. E. Tomboulis, Nucl. Phys. Proc. Suppl. 34 (1994) 192; Nucl. Phys. Proc. Suppl. 30 (1993) 549; Phys. Lett. B303 (1993) 103.
14. G. Bali, C. Schlichter, and K. Schilling, Phys. Rev. D51 (1995) 5165.
15. B. Bakker, M. Chernodub, and M. Polikarpov, hep-lat/9706007.
16. J. Greensite and M. Halpern, Phys. Rev. D27 (1983) 2545.
17. J. Greensite, Nucl. Phys. B315 (1989) 663.
18. T. Kovács and E. Tomboulis, proceedings of LATTICE 97, to appear in Nucl. Phys. Proc. Suppl.

DUALITY AND THE RENORMALIZATION GROUP

Peter E. Haagensen

Center for Theoretical Physics
Massachusetts Institute of Technology
77 Massachusetts Ave.
Cambridge, MA 02139
USA

ABSTRACT

The requirement that duality and renormalization group transformations commute as motions in the space of a theory has recently been explored to extract information about the renormalization flows in different statistical and field theoretical systems. After a review of what has been accomplished in the context of $2d$ sigma models, new results are presented which set up the stage for a fully generic calculation at two-loop order, with particular emphasis on the question of scheme dependence.

1. INTRODUCTION

Duality symmetries are typically transformations in the parameter space of a theory which leave the partition function and the correlators invariant (perhaps up to some known function of the parameters). Renormalization group (RG) transformations also act in this same space, with similar invariance properties. Given the generality of this observation, one expects that it may be possible to investigate the interplay between duality and the RG whenever a system presents a duality symmetry and a renormalization flow, regardless of whether it is a quantum field theory, a statistical system, a lattice theory, etc..

Indeed, such a nontrivial interplay has recently been verified in a number of different contexts,[1-5] and a requirement of consistency of duality symmetry and the RG has been used to obtain constraints on the RG flows. Spin systems, which generally enjoy a Kramers-Wannier duality symmetry, were considered in [1]. The Quantum Hall system, on the other hand, is strongly suspected to exhibit a much richer duality, under $SL(2,\mathbb{Z})$ or one of its (level 2) subgroups, and it was studied in [2]. For the spin systems considered in [1], the parameter space consisted of a single relevant coupling (the inverse temperature), while Kramers-Wannier duality forms the (colloquially speaking) simple group $\mathbb{Z}_2$, so that the constraints on the RG structure end up not being strong

New Developments in Quantum Field Theory
Edited by Damgaard and Jurkiewicz, Plenum Press, New York, 1998

Quantum Hall system, where the parameter space consists of the upper complex half-plane, and the requirements of holomorphy and modular symmetry naturally turn out to be considerably richer. Yet, even in that case, the existing results are not entirely conclusive: while on the one hand there is not enough experimental data confirming the precise symmetry group of the system, on the other hand by postulating a specific modular symmetry one still does not obtain unique RG flows.

Two-dimensional sigma models targeted on an arbitrary background of metric, antisymmetric tensor and dilaton fields also present a duality symmetry (when the background has an abelian isometry), and in that context, the situation is more favorable: while the symmetry group is, like for spin systems, also $\mathbb{Z}_2$, the parameter space is of course much larger, and the action of the group on it rather more involved, with geometry and torsion mixing in a nontrivial way. For the loop orders and backgrounds considered so far, this has in fact allowed for an essentially complete determination of the RG flows using only the requirement of consistency between duality and the RG (the qualification 'essentially' will be understood more clearly below). It is to these models and the relevant consistency requirements that the bulk of what follows will be dedicated.

To begin, we consider a system with a number of couplings, k^i, $i = 1, \ldots n$, and a duality symmetry, T, such that

$$Tk^i \equiv \tilde{k}^i = \tilde{k}^i(k) \tag{1}$$

represents a map between equivalent points in the parameter space (with equivalence taking the same meaning as, for instance, the order-disorder equivalence in the $2d$ Ising model). We will also assume the system has a renormalization group flow, R, encoded by a set of beta functions:

$$Rk^i \equiv \beta^i(k) = \mu \frac{d}{d\mu} k^i \ , \tag{2}$$

with μ some appropriate subtraction scale. On a generic function in the parameter space, $F(k)$, these operations act as follows:

$$\begin{aligned} TF(k) &= F(\tilde{k}(k)) \\ RF(k) &= \frac{\delta F(k)}{\delta k^i} \cdot \beta^i(k) \ . \end{aligned} \tag{3}$$

For a finite number of couplings the derivatives above should be understood as ordinary derivatives, whereas in the case of the sigma model these will be functional derivatives, and the dot will imply an integration over spacetime. The consistency requirement governing the interplay of duality and the RG can now be stated very simply:

$$[T, R] = 0 \tag{4}$$

or, in words, that duality transformations and RG flows commute as motions in the parameter space of the theory. This is the main concept to be explored, and from which most results will follow. It is easy to see that the above amounts to the following consistency conditions:

$$\beta^i(\tilde{k}) = \frac{\delta \tilde{k}^i}{\delta k^j} \cdot \beta^j(k) \ , \tag{5}$$

that is, under duality transformations the beta function must transform as a "form-invariant contravariant vector" (to avoid confusion: we are borrowing the language of General Relativity here, but of course duality transformations have nothing to do with diffeomorphisms!). It is this "form-invariance", *i.e.*, the fact that the functional form

of β^i on the l.h.s. above must be the same as the one on the r.h.s., that is mostly responsible for the severity of the constraints engendered.

For the 2*d* Ising model on a square lattice this yields a constraint which is nontrivially satisfied by the (known) beta function of the model, although it does not determine uniquely this beta function. In the Quantum Hall system, on the other hand, the resulting constraint is that the beta function transform as a weight –2 modular form (strictly speaking, negative weight modular forms do not exist, and this obstruction is then circumvented by slightly relaxing the condition of holomorphy, but these details will not concern us here).

In what follows, we will explore in detail the analogous constraints in the context of 2*d* bosonic sigma models, in order of increasing complexity: Sections 2 and 3 contain a review of previously published work[3,4,5] on, respectively, the fully generic one-loop case and the purely metric two-loop case, while Section 4 comprises results obtained in the course of more recent investigations,[6] and presents the setup for the fully generic two-loop case, in the presence of torsion. In this case, where all possible backgrounds are included, the issue of scheme dependence will also be discussed in some detail, as it arises for the first time to complicate matters in a considerable way.

2. SIGMA MODELS AT ONE-LOOP ORDER

Our starting point is the $d = 2$ bosonic sigma model on a generic D+1-dimensional background $\{g_{\mu\nu}(X),\ b_{\mu\nu}(X)\}$ of metric and antisymmetric tensor, respectively, where $\mu, \nu = 0, 1, \ldots, D = 0, i$, so that the $\mu = 0$ component is singled out. We shall assume this sigma model has an abelian isometry, which will enable duality transformations, and we shall consider the background above in the adapted coordinates, in which the abelian isometry is made manifest through independence of the background on the coordinate $\theta \equiv X^0$. The original sigma model action reads:

$$S = \frac{1}{4\pi\alpha'}\int d^2\sigma\ \Big[g_{00}(X)\partial_\alpha\theta\partial^\alpha\theta + 2g_{0i}(X)\partial_\alpha\theta\partial^\alpha X^i + g_{ij}(X)\partial_\alpha X^i\partial^\alpha X^j + i\varepsilon^{\alpha\beta}\left(2b_{0i}(X)\partial_\alpha\theta\partial_\beta X^i + b_{ij}(X)\partial_\alpha X^i\partial_\beta X^j\right)\Big]\ . \tag{6}$$

Throughout, all background tensors can depend only on target coordinates X^i, $i = 1, \ldots, D$, and not on θ.

The duality transformations in this model are also well-known:[7]

$$\tilde{g}_{00} = \frac{1}{g_{00}}\ ,\quad \tilde{g}_{0i} = \frac{b_{0i}}{g_{00}}\ ,\quad \tilde{b}_{0i} = \frac{g_{0i}}{g_{00}}\ ,$$
$$\tilde{g}_{ij} = g_{ij} - \frac{g_{0i}g_{0j} - b_{0i}b_{0j}}{g_{00}}\ ,\quad \tilde{b}_{ij} = b_{ij} - \frac{g_{0i}b_{0j} - b_{0i}g_{0j}}{g_{00}}\ . \tag{7}$$

The statement of classical duality is that the model defined on the dual background $\{\tilde{g}_{\mu\nu}, \tilde{b}_{\mu\nu}\}$ is simply a different parametrization of the same model, given that the manipulations used to derive the transformations essentially only involve performing trivial integrations in a different order starting from the path-integral in which the abelian isometry is gauged.

On a curved worldsheet, another background must be introduced, that of the dilaton $\phi(X)$, coupling to the worldsheet curvature scalar. The RG flow of background couplings is given by their respective beta functions:

$$\beta^g_{\mu\nu} \equiv \mu\frac{d}{d\mu}g_{\mu\nu}\ ,\ \beta^b_{\mu\nu} \equiv \mu\frac{d}{d\mu}b_{\mu\nu}\ ,\ \beta^\phi \equiv \mu\frac{d}{d\mu}\phi\ , \tag{8}$$

while the trace of the stress energy tensor is found from the Weyl anomaly coefficients[8]

$$\begin{aligned}\bar{\beta}^g_{\mu\nu} &= \beta^g_{\mu\nu} + 2\alpha'\nabla_\mu\partial_\nu\phi + \nabla_{(\mu}W_{\nu)}\ ,\\ \bar{\beta}^b_{\mu\nu} &= \beta^b_{\mu\nu} + \alpha' H_{\mu\nu}{}^\lambda\partial_\lambda\phi + H_{\mu\nu}{}^\lambda W_\lambda + \nabla_{[\mu}L_{\nu]}\ ,\\ \bar{\beta}^\phi &= \beta^\phi + \alpha'(\partial_\mu\phi)^2 + \nabla^\mu\phi W_\mu\ ,\end{aligned} \tag{9}$$

where W_μ and L_μ are some specific target vectors depending on $g_{\mu\nu}$ and $b_{\mu\nu}$, and $(\mu\nu) = \mu\nu + \nu\mu$, $[\mu\nu] = \mu\nu - \nu\mu$. For the loop orders and backgrounds considered in Sections 2 and 3, $W_\mu = L_\mu = 0$, and we will henceforth disregard them. Both the beta functions and the Weyl anomaly coefficients turn out to satisfy the consistency conditions, Eq. (5). However, while the latter satisfy them exactly, the former satisfy them up to a target reparametrization.[3,4] Since both encode essentially the same RG information, for simplicity we will consider RG motions as generated by the Weyl anomaly coefficients in what follows. Thus, in the present context, the couplings are denoted by $k^i = \{g_{\mu\nu}, \beta_{\mu\nu}, \phi\}$, with $i = g, b, \phi$ labeling metric, antisymmetric tensor and dilaton backgrounds, and our R operation will in this case be defined, on a generic functional $F[g, b, \phi]$ (and in principle at any loop order), to be

$$RF[g,b,\phi] = \frac{\delta F}{\delta g_{\mu\nu}}\cdot\bar{\beta}^g_{\mu\nu} + \frac{\delta F}{\delta b_{\mu\nu}}\cdot\bar{\beta}^b_{\mu\nu} + \frac{\delta F}{\delta\phi}\cdot\bar{\beta}^\phi\ , \tag{10}$$

where the dot also indicates a spacetime integration. Duality transformations are given by

$$TF[g,b,\phi] = F[\tilde{g},\tilde{b},\tilde{\phi}]\ , \tag{11}$$

where $\tilde{\phi}$ will be defined shortly.

The consistency conditions to be satisfied, Eq. (5), that obtain from Eq. (7) translate more explicitly into:

$$\begin{aligned}\bar{\beta}^{\tilde{g}}_{00} &= -\frac{1}{g_{00}^2}\bar{\beta}^g_{00}\ ,\\ \bar{\beta}^{\tilde{g}}_{0i} &= -\frac{1}{g_{00}^2}\left(b_{0i}\bar{\beta}^g_{00} - \bar{\beta}^b_{0i}g_{00}\right)\ ,\\ \bar{\beta}^{\tilde{b}}_{0i} &= -\frac{1}{g_{00}^2}\left(g_{0i}\bar{\beta}^g_{00} - \bar{\beta}^g_{0i}g_{00}\right)\ ,\\ \bar{\beta}^{\tilde{g}}_{ij} &= \bar{\beta}^g_{ij} - \frac{1}{g_{00}}\left(\bar{\beta}^g_{0i}g_{0j} + \bar{\beta}^g_{0j}g_{0i} - \bar{\beta}^b_{0i}b_{0j} - \bar{\beta}^b_{0j}b_{0i}\right)\\ &\quad + \frac{1}{g_{00}^2}\left(g_{0i}g_{0j} - b_{0i}b_{0j}\right)\bar{\beta}^g_{00}\ ,\\ \bar{\beta}^{\tilde{b}}_{ij} &= \bar{\beta}^b_{ij} - \frac{1}{g_{00}}\left(\bar{\beta}^g_{0i}b_{0j} + \bar{\beta}^b_{0j}g_{0i} - \bar{\beta}^g_{0j}b_{0i} - \bar{\beta}^b_{0i}g_{0j}\right)\\ &\quad + \frac{1}{g_{00}^2}\left(g_{0i}b_{0j} - b_{0i}g_{0j}\right)\bar{\beta}^g_{00}\ ,\end{aligned} \tag{12}$$

where, in a condensed notation, we take the quantities on the l.h.s. above to mean $\bar{\beta}^{\tilde{g}}_{\mu\nu} \equiv \bar{\beta}^g_{\mu\nu}[\tilde{g},\tilde{b},\tilde{\phi}]$, etc.. Both the dilaton duality transformation and its attendant consistency condition are still ostensibly missing, but will be determined shortly.

At loop order ℓ, the possible tensor structures $T_{\mu\nu}$ appearing in the beta function must scale as $T_{\mu\nu}(\Lambda g, \Lambda b) = \Lambda^{1-\ell}T_{\mu\nu}(g, b)$ under global scalings of the background fields. At $\mathcal{O}(\alpha')$ one may then have

$$\begin{aligned}\beta^g_{\mu\nu} &= \alpha'\left(A\,R_{\mu\nu} + B\,H_{\mu\lambda\rho}H_\nu{}^{\lambda\rho} + C\,g_{\mu\nu}R + D\,g_{\mu\nu}H_{\alpha\beta\gamma}H^{\alpha\beta\gamma}\right)\ ,\\ \beta^b_{\mu\nu} &= \alpha'\left(E\,\nabla^\lambda H_{\mu\nu\lambda}\right)\ ,\end{aligned} \tag{13}$$

where $H_{\mu\nu\lambda} = \partial_\mu b_{\nu\lambda}$ + cyclic permutations and with A, B, C, D, E being determined from one-loop Feynman diagrams. As found in [3], requiring Eq. (12) to be satisfied, and choosing $A = 1$ determines $B = -1/4$, $E = -1/2$, and $C = D = 0$, independently of any diagram calculations. The consistency conditions, Eq. (12), on $g_{\mu\nu}$ and $b_{\mu\nu}$ alone also allow for an independent determination of the dilaton transformation (or "shift") $\tilde{\phi} = \phi - \frac{1}{2}\ln g_{00}$. From this shift, and Eq. (5), one obtains yet another consistency condition,

$$\bar{\beta}^{\tilde{\phi}} = \bar{\beta}^{\phi} - \frac{1}{2g_{00}}\bar{\beta}^{g}_{00} \,, \tag{14}$$

from which one can finally find the dilaton beta function, thus completely determining all beta functions at this order:

$$\begin{aligned} \beta^{g}_{\mu\nu} &= \alpha' \left(R_{\mu\nu} - \frac{1}{4} H_{\mu\lambda\rho} H_{\nu}{}^{\lambda\rho} \right) \\ \beta^{b}_{\mu\nu} &= -\frac{\alpha'}{2} \nabla_\lambda H^{\lambda}{}_{\mu\nu} \\ \beta^{\phi} &= C - \frac{\alpha'}{2}\nabla^2\phi - \frac{\alpha'}{24} H_{\mu\nu\lambda} H^{\mu\nu\lambda} \,, \end{aligned} \tag{15}$$

up to the constant C.

3. TWO-LOOP ORDER WITH PURELY METRIC BACKGROUNDS

At the next order R is modified by the two-loop beta functions, and one must determine the appropriate modifications in T such that $[T, R] = 0$ continues to hold. We begin by working with restricted backgrounds of the form

$$g_{\mu\nu} = \begin{pmatrix} a & 0 \\ 0 & \bar{g}_{ij} \end{pmatrix} \,, \tag{16}$$

and $b_{\mu\nu} = 0$, so that no torsion appears in the dual background either. It is useful to define at this point the following two quantities: $a_i \equiv \partial_i \ln a$, and $q_{ij} \equiv \bar{\nabla}_i a_j + \frac{1}{2} a_i a_j$, where barred quantities here and below refer to the metric $\bar{g}_{ij}$ (also, indices $i, j, \ldots$, are contracted with the metric $\bar{g}_{ij}$). Within this class of backgrounds classical duality transformations reduce to the operation $a \to 1/a$, and it is simple to determine the possible corrections to T from a few basic requirements, spelled out in detail in [5]. For conciseness, we will directly present the final result for the corrected duality transformations: [9]

$$\begin{aligned} \ln \tilde{a} &= -\ln a + \lambda\alpha' a_i a^i \,, \\ \tilde{g}_{ij} &= g_{ij} = \bar{g}_{ij} \,, \\ \tilde{\phi} &= \phi - \frac{1}{2}\ln a + \frac{\lambda}{4}\alpha' a_i a^i \,, \end{aligned} \tag{17}$$

where λ is a constant that cannot be determined from the basic requirements. The consistency conditions that follow from the above are:

$$\begin{aligned} \frac{1}{\tilde{a}}\,\bar{\beta}^{\tilde{g}}_{00} &= -\frac{1}{a}\bar{\beta}^{g}_{00} + 2\lambda\alpha' \left[a^i \partial_i \left(\frac{1}{a}\bar{\beta}^{g}_{00} \right) - \frac{1}{2} a^i a^j \bar{\beta}^{g}_{ij} \right] \,, \\ \bar{\beta}^{\tilde{g}}_{ij} &= \bar{\beta}^{g}_{ij} \,, \\ \bar{\beta}^{\tilde{\phi}} &= \bar{\beta}^{\phi} - \frac{1}{2a}\bar{\beta}^{g}_{00} + \frac{\lambda}{2}\alpha' \left[a^i \partial_i \left(\frac{1}{a}\bar{\beta}^{g}_{00} \right) - \frac{1}{2} a^i a^j \bar{\beta}^{g}_{ij} \right] \,. \end{aligned} \tag{18}$$

The terms scaling correctly under $g \to \Lambda g$ at this order, and thus possibly present in the beta function, are

$$\beta^{g(2)}_{\mu\nu} = A_1 \nabla_\mu \nabla_\nu R + A_2 \nabla^2 R_{\mu\nu} + A_3 R_{\mu\alpha\nu\beta} R^{\alpha\beta} + A_4 R_{\mu\alpha\beta\gamma} R_\nu{}^{\alpha\beta\gamma} + A_5 R_{\mu\alpha} R_\nu{}^\alpha + A_6 R_{\mu\nu} R + A_7 g_{\mu\nu} \nabla^2 R + A_8 g_{\mu\nu} R^2 + A_9 g_{\mu\nu} R_{\alpha\beta} R^{\alpha\beta} + A_{10} g_{\mu\nu} R_{\alpha\beta\gamma\delta} R^{\alpha\beta\gamma\delta} \quad (19)$$

(we have used Bianchi identities to reduce from a larger set of tensor structures).

It will suffice in fact to study the consistency conditions for the (ij) components, $\bar{\beta}^{\bar{g}}_{ij} = \bar{\beta}^{g}_{ij}$, in order to determine the only structure satisfying all the consistency conditions.

We write

$$\bar{\beta}^{g}_{ij} = \alpha' \left(\beta^{g(1)}_{ij} + 2 \bar{\nabla}_i \partial_j \phi \right) + \alpha'^2 \beta^{g(2)}_{ij} \ , \quad (20)$$

where $\beta^{g(1)}_{ij} = R_{ij} = \bar{R}_{ij} - \frac{1}{2} q_{ij}$ is the one-loop beta function, and perform the duality transformation (17), keeping terms to $\mathcal{O}(\alpha'^2)$. Using the fact that the one-loop Weyl anomaly coefficient satisfies the one-loop consistency conditions (12), we arrive at

$$\beta^{\bar{g}(2)}_{ij} = \beta^{g(2)}_{ij} - \frac{1}{4} \lambda a_{(i} \partial_{j)} (a^k a_k) \ , \quad (21)$$

where the duality transformation of $\beta^{g(2)}_{ij}$ is given simply by $a \to 1/a$ without α' corrections, since this is already $\mathcal{O}(\alpha'^2)$. Separating the possible tensor structures into even and odd tensors under $a \to 1/a$,

$$\beta^{g(2)}_{ij} = E_{ij} + O_{ij} \ , \quad \tilde{E}_{ij} = E_{ij} \ , \quad \tilde{O}_{ij} = -O_{ij} \ , \quad (22)$$

the even structures drop out of Eq. (21) and we are left with

$$O_{ij} = \frac{1}{8} \lambda a_{(i} \partial_{j)} (a^k a_k) \ . \quad (23)$$

We now perform a standard Kaluza-Klein reduction on the ten terms in (19) to identify which if any satisfy this condition. The results can be obtained using the formulas in the Appendix of [4], and are as follows:

$$\begin{aligned}
&(1) \ : \ \nabla_i \nabla_j R = \bar{\nabla}_i \bar{\nabla}_j (\bar{R} - q_n{}^n) \ , \\
&(2) \ : \ \nabla^2 R_{ij} = (\bar{\nabla}^2 + \frac{1}{2} a_k \bar{\nabla}^k)(\bar{R}_{ij} - \frac{1}{2} q_{ij}) - \frac{1}{4} a_i a_j q_n{}^n \\
&\qquad\qquad - \frac{1}{4} a^k a_{(i} \left(\bar{R}_{j)k} - \frac{1}{2} q_{j)k} \right) \ , \\
&(3) \ : \ R_{i\alpha j\beta} R^{\alpha\beta} = \frac{1}{4} q_{ij} q_n{}^n + \bar{R}_{injm} (\bar{R}^{nm} - \frac{1}{2} q^{nm}) \ , \\
&(4) \ : \ R_{i\alpha\beta\gamma} R_j{}^{\alpha\beta\gamma} = \frac{1}{2} q_{ik} q_j{}^k + \bar{R}_{iknm} \bar{R}_j{}^{knm} \ , \\
&(5) \ : \ R_{i\alpha} R_j{}^\alpha = \bar{R}_{ik} \bar{R}_j{}^k - \frac{1}{2} \bar{R}_{k(i} q_{j)}{}^k + \frac{1}{4} q_{ik} q_j{}^k \ , \\
&(6) \ : \ R_{ij} R = (\bar{R}_{ij} - \frac{1}{2} q_{ij})(\bar{R} - q_n{}^n) \ , \\
&(7) \ : \ g_{ij} \nabla^2 R = \bar{g}_{ij} \left[\frac{1}{2} a^k \partial_k (\bar{R} - q_m{}^m) \right. \\
&\qquad\qquad \left. + \bar{\nabla}^k \partial_k (\bar{R} - q_m{}^m) \right] \ , \\
&(8) \ : \ g_{ij} R^2 = \bar{g}_{ij} \left(\bar{R} - q_m{}^m \right)^2 \ , \\
&(9) \ : \ g_{ij} R_{\alpha\beta} R^{\alpha\beta} = \bar{g}_{ij} \left[\frac{1}{4} (q_m{}^m)^2 + (\bar{R}_{km} - \frac{1}{2} q_{km})^2 \right] \ , \\
&(10) \ : \ g_{ij} R_{\alpha\beta\gamma\delta} R^{\alpha\beta\gamma\delta} = \bar{g}_{ij} \left[q_{km} q^{km} + \bar{R}_{k\ell mn} \bar{R}^{k\ell mn} \right] \ .
\end{aligned} \quad (24)$$

The respective odd parts are

$$
\begin{aligned}
O^{(1)}_{ij} &= -\bar{\nabla}_i\bar{\nabla}_j\bar{\nabla}_n a^n \ , \\
O^{(2)}_{ij} &= \frac{1}{2}a_k\bar{\nabla}^k\bar{R}_{ij} - \frac{1}{2}\bar{\nabla}^2\bar{\nabla}_i a_j - \frac{1}{4}a_i a_j\bar{\nabla}_k a^k \ , \\
O^{(3)}_{ij} &= -\frac{1}{2}\bar{R}_{injm}\bar{\nabla}^n a^m + \frac{1}{8}a_n a^n\bar{\nabla}_i a_j + \frac{1}{8}a_i a_j\bar{\nabla}_n a^n \ , \\
O^{(4)}_{ij} &= \frac{1}{4}a_k a_{(i}\bar{\nabla}_{j)}a^k \ , \\
O^{(5)}_{ij} &= -\frac{1}{2}\bar{R}_{k(i}\bar{\nabla}_{j)}a^k + \frac{1}{8}a_k a_{(i}\bar{\nabla}_{j)}a^k \ , \\
O^{(6)}_{ij} &= -\frac{1}{2}\bar{R}\bar{\nabla}_i a_j - \bar{R}_{ij}\bar{\nabla}_n a^n + \frac{1}{4}a_i a_j\bar{\nabla}_n a^n \\
&\quad + \frac{1}{4}a_n a^n\bar{\nabla}_i a_j \ , \\
O^{(7)}_{ij} &= \bar{g}_{ij}\left[\frac{1}{2}a^k\partial_k(\bar{R} - \frac{1}{2}a_m a^m) - \bar{\nabla}^k\partial_k(\bar{\nabla}_m a^m)\right] \ , \\
O^{(8)}_{ij} &= \bar{g}_{ij}\left[-2(\bar{\nabla}^k a_k)\bar{R} + (\bar{\nabla}^k a_k)a^m a_m\right] \ , \\
O^{(9)}_{ij} &= \bar{g}_{ij}\left[\frac{1}{4}(\bar{\nabla}^k a_k)a^m a_m - (\bar{\nabla}_k a_m)\bar{R}^{km}\right. \\
&\quad \left. + \frac{1}{4}(\bar{\nabla}_k a_m)a^k a^m\right] \ , \\
O^{(10)}_{ij} &= \bar{g}_{ij}(\bar{\nabla}_k a_m)a^k a^m \ .
\end{aligned}
\tag{25}
$$

It is fortunate that none of these tensors contain purely even structures, since such structures are left unconstrained (and thus undetermined) by duality. The only odd term of the form (23) comes from $A_4 R_{\mu\alpha\beta\gamma}R_\nu{}^{\alpha\beta\gamma}$, and a detailed inspection shows that no linear combination of the other terms gives rise to odd tensors generically of the form (23). This determines that, with the requirement of covariance of duality under the RG, the $O(\alpha'^2)$ term in the beta function is

$$\beta^{g(2)}_{\mu\nu} = \lambda R_{\mu\alpha\beta\gamma}R_\nu{}^{\alpha\beta\gamma} \ . \tag{26}$$

One should now check that the corresponding (00) component also satisfies its consistency condition. A straightforward computation shows that it does, and the determination of the two-loop beta function is thus complete.

Although we treated a restricted class of metric backgrounds, the final result is valid for a generic metric, since none of the possible tensor structures are built out of the off-block-diagonal g_{0i} elements alone (in which case our consistency conditions would be blind to them, just as they are to the even terms E_{ij}).

Simply using the requirements that duality and the RG commute as motions in the parameter space of the sigma model, we have thus been able to determine the two-loop beta function to be

$$\beta_{\mu\nu} = \alpha' R_{\mu\nu} + \alpha'^2\lambda R_{\mu\alpha\beta\gamma}R_\nu{}^{\alpha\beta\gamma} \ , \tag{27}$$

for an entirely generic metric background, again without any Feynman diagram calculations. Because we used an extremely restrictive class of backgrounds, it was not possible to determine the value of λ (the correct value is $\lambda = \frac{1}{2}$). However, we expect that, similarly to what happens at $O(\alpha')$, once a more generic background is used in the consistency conditions, even this constant should be determined. We now examine how to go about calculating in such a generic background in an efficient way.

4. SETUP FOR THE FULLY GENERIC TWO-LOOP CASE

The inclusion of torsion at two-loop order brings with it a number of complications which one should try to minimize as much as possible. First, the number of new terms appearing in the beta functions is greatly increased as compared to the purely metric case. Moreover, there is now one more beta function to worry about, for the antisymmetric tensor. Also, it is clear that there will be several new terms in the corrections to duality transformations, and the general arguments used in [5] will not be sufficient to determine them. To finally complicate the situation further, the scheme dependence present leaves a lot more latitude to what the correct expressions for these beta functions are, as well as which duality transformations should make them transform covariantly.

It thus seems that a direct guessing of the corrected duality transformations, attempting to keep the two-loop beta functions covariant, would be an extremely arduous task, and we will try rather to first streamline the calculations involved by going through what may seem at first a longer path.

We start with the observation that there is a connection between the Weyl anomaly coefficients and the string background effective action ("EA" in what follows), whereby one establishes a direct relation between the former and the equations of motion of the latter. In principle, there is thus the possibility that the duality transformation properties of one imply the transformation properties of the other. Should this be the case, one might save considerable effort by studying the effective action alone, since this is a scalar function on the parameter space, invariant under duality, whereas the beta function represents a vector field in that space, with nontrivial transformation properties.

Unfortunately, we will see that the transformation properties of the EA under duality will *not* allow us to deduce the transformation properties of the Weyl anomaly coefficients. However, the detailed consideration of this relationship will still be useful, on the one hand to limit the possible transformations under which the Weyl anomaly coefficients behave covariantly, and on the other, to clarify the messy issue of scheme dependence.

We begin by examining the situation at one-loop order. The EA is given by

$$S \equiv \alpha' S_0 = \alpha' \int d\theta \; d^D x \; \sqrt{g} e^{-2\phi} \left(R + 4(\nabla\phi)^2 - \frac{1}{12} H_{\mu\nu\lambda} H^{\mu\nu\lambda} \right) \; . \tag{28}$$

Given the one-loop expressions for the beta functions, Eq. (15), it is simple to see that this EA is actually equivalent to (with $C = 0$)

$$S = R \; V = V_{,i} \cdot \vec{\beta}^i \; , \tag{29}$$

where $V = \sqrt{g} \exp -2\phi$, and $V_{,i} \equiv \delta V / \delta k^i$, in the notation of the introduction. Because we know the one-loop Weyl anomaly coefficients transform contravariantly under duality (cf. Eq. (5)), the gradient $\delta/\delta k^i$ transforms covariantly, and V is invariant, it immediately follows that S as defined above is invariant under duality transformations. Similarly, at higher loop orders, if we are able to find the corrected duality transformations under which the higher-loop Weyl anomaly coefficients transform contravariantly as in Eq. (5), and if we are able to find a scheme in which the EA continues to be given by Eq. (29), then we are guaranteed duality invariance of the EA. But that is actually opposite to the direction we are seeking. Can we attempt to argue also conversely? At first sight, Eq. (29) does seem to give this converse result, that once a duality transformation can be found at some loop order that keeps S invariant, that will imply the

sought for contravariance of the Weyl anomaly coefficients, and thus the statement that $[T, R] = 0$.

That is not correct, however, for $V_{,i}$, which is given more explicitly by

$$V_{,i} = \begin{pmatrix} V_{,g} \\ V_{,b} \\ V_{,\phi} \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} V g^{\mu\nu} \\ 0 \\ -2V \end{pmatrix} , \tag{30}$$

(we are omitting a delta function coming from the functional derivative) has an enormous amount of "zero modes", given by

$$z^i = \begin{pmatrix} z^g \\ z^b \\ z^\phi \end{pmatrix} = \begin{pmatrix} z^g_{\mu\nu} \\ z^b_{\mu\nu} \\ \frac{1}{4} g^{\mu\nu} z^g_{\mu\nu} \end{pmatrix} , \tag{31}$$

with $z^g_{\mu\nu}, z^b_{\mu\nu}$ arbitrary, so that $V_{,i}\, z^i = 0$. This implies that, at some loop order, if there is a set of duality transformations that keeps S as defined in Eq. (29) invariant, then the Weyl anomaly coefficients are seen to transform not as in Eq. (5), but as

$$\beta^i(\tilde{k}) + \tilde{z}^i(\tilde{k}) = \frac{\delta \tilde{k}^i}{\delta k^j} \cdot \left(\beta^j(k) + z^j(k) \right) , \tag{32}$$

with $z^i(k)$ and $\tilde{z}^i(\tilde{k})$ some specific vectors of the form (31), not necessarily zero. Naturally, this does not represent any covariance property at all.

Could some other reasoning rescue the possibility that the invariance of the EA might imply the contravariance of the Weyl anomaly coefficients? For instance, one immediate criticism that may be applied to the argument above is that one is not sure *a priori* that the EA should really be given by Eq. (29). This brings us to the issue of scheme ambiguities.

An independent definition of the EA corresponds to the field theory action that generates the massless sector of the (string) tree level string S-matrix. That EA contains a large number of terms that are ambiguous in that they can be modified with field redefinitions of the EA, and it contains a smaller number of terms that are invariant under field redefinitions, and thus unambiguous. Field redefinitions also affect the beta functions, and stemming from their definition, Eq. (8), it is simple to see that they must transform under field redefinitions as contravariant vectors (now we *are* talking about diffeomorphisms). Parenthetically, we note that a subset of these field redefinitions correspond to what is typically referred to as a change of subtraction scheme in the renormalization of the sigma model: if, say, minimal subtraction corresponds to the subtraction of a divergent term $1/\varepsilon\ T_{\mu\nu}$, then a different, nonminimal scheme corresponds to the subtraction of a term (const. + $1/\varepsilon$) $T_{\mu\nu}$, which in turn is equivalent to a field redefinition by the term (const.) $T_{\mu\nu}$. With such a notion of scheme ambiguity, it can be seen that the two-loop beta function in the purely metric case is actually scheme independent, a property which is lost when torsion is included. More generally, however, because the sigma model is not renormalizable in the usual sense, one is also allowed finite subtractions of terms not originally present in the action, and these correspond to arbitrary field redefinitions. In order not to propagate semantic confusion, we will refrain from using the standard (and more restrictive) notion of scheme ambiguity, and will always consider instead the full generality of arbitrary field redefinitions, referring to different redefinitions as different choices of scheme.

At any rate, we realize from the discussion above that there is an unambiguous and independent notion to the EA, and that to each different "realization" of it, or

choice of scheme, there corresponds also a choice of scheme for the beta functions. It is expected that in any scheme there should be a relation between the equations of motion of the EA and the Weyl anomaly coefficients of the form

$$\frac{\delta S}{\delta k^i} = G_{ij} \cdot \bar{\beta}^j \ , \tag{33}$$

with G_{ij} invertible, in the sense that the equations of motion imply the vanishing of the Weyl anomaly coefficients, and vice-versa (an even stronger requirement would be the positivity of G_{ij}, in order to connect the EA to a c-function, but that will not concern us here). Could this relation between the EA and $\bar{\beta}^i$'s allow us to deduce the contravariance of the latter from invariance of the former?

The notation certainly is very suggestive, with G_{ij} naturally appearing to change an object transforming contravariantly into an object transforming covariantly. However, insofar as G_{ij} itself has no independent meaning,* but is *devised* to have the above equation satisfied (in the sense that it just represents the particular linear combinations of $\bar{\beta}^i$'s that yield the equations of motion of the EA), the answer is again unfortunately negative, and it is well exemplified by the situation at one loop order already. With the beta functions given by Eq. (15), and the EA by Eq. (28), it is simple to find that G_{ij} will be

$$\begin{aligned} G_{ij} &= \begin{pmatrix} G^{\mu\nu\lambda\rho}_{gg} & G^{\mu\nu\lambda\rho}_{gb} & G^{\mu\nu}_{g\phi} \\ G^{\mu\nu\lambda\rho}_{bg} & G^{\mu\nu\lambda\rho}_{bb} & G^{\mu\nu}_{b\phi} \\ G^{\lambda\rho}_{\phi g} & G^{\lambda\rho}_{\phi b} & G_{\phi\phi} \end{pmatrix} \\ &= -\sqrt{g}e^{-2\phi} \begin{pmatrix} \frac{1}{2}g^{\mu(\lambda}g^{\rho)\nu} - \frac{1}{2}g^{\mu\nu}g^{\lambda\rho} & 0 & 2g^{\mu\nu} \\ 0 & \frac{1}{2}g^{\mu[\lambda}g^{\rho]\nu} & 0 \\ 2g^{\lambda\rho} & 0 & -8 \end{pmatrix} . \end{aligned} \tag{34}$$

We already know that both the one-loop Weyl anomaly coefficients and EA transform in the proper way, and thus one *must* find, if one were to check explicitly the transformation properties of the particular G_{ij} given above, that it transforms like a rank 2 form-invariant covariant tensor under duality. If we did not know how the $\bar{\beta}^i$'s transformed, however, all we could tell from the invariance of S is that G_{ij} has to cancel whatever (possibly completely wrong) transformation property of $\bar{\beta}^j$, and yield the transformation rule for a covariant vector. Thus if, say, the antisymmetric tensor Weyl anomaly coefficient were twice its correct value, Eq. (33) would still hold if we multiplied G_{bb} by 1/2, and yet the "new" $\bar{\beta}^i$ (with the wrong coefficient of $\bar{\beta}^b$) would certainly *not* satisfy Eq. (5), and consequently $[T, R] = 0$ would also not be satisfied. Accordingly, the "new" G_{ij} would also not transform like a rank 2 covariant tensor. Another clear, and even more pertinent, example of this can be seen with G_{ij} at two-loop order: if Eq. (33) is expanded to $O(\alpha'^2)$, the r.h.s. will contain, at $O(\alpha'^2)$, a term given by the contraction of G_{ij} at two-loop order with $\bar{\beta}^j$ at one-loop order. Such an expansion is considered in [10], and the authors note there that because G_{ij} at two-loop order has the same tensor structures as $\bar{\beta}^j$ at one-loop order, whenever a term appears containing (roughly speaking) the square of a one-loop $\bar{\beta}^j$, it becomes impossible to determine which piece belongs to G_{ij}, and which to $\bar{\beta}^j$. Of course, any choice other than the correct one will lead to a violation of $[T, R] = 0$, even though Eq. (33) is perfectly

*Again, in the context of a c-theorem it would, but scheme dependence in the present context complicates matters too much to allow one at this point to seriously conjecture G_{ij} to be the Zamolodchikov metric. This may well turn out to be true eventually, in some scheme, but we shall simply not assume it here.

satisfied whichever way these terms are split up. Incidentally, the authors of [10] suggest the only way to resolve this indefinition in G_{ij} is by going one order higher; we would suggest instead that the present considerations involving duality will eventually resolve this problem without the need to go to three-loop order.

So far, it has seemed that the invariance of the EA cannot really be of any help in determining the covariance properties of the Weyl anomaly coefficients. But, in fact, the above has not been entirely in vain: we know that, in the scheme in which the EA is given by Eq. (29) at higher loop order, if there exists at all any duality transformations that respect $[T, R] = 0$, *i.e.*, such that $\bar{\beta}^i$ transforms contravariantly, then these transformations must keep the EA invariant; thus, if there is only one set of transformations that keep the EA invariant, these are the only transformations that have a chance of satisfying $[T, R] = 0$. So, we are not guaranteed that the transformations that keep S invariant satisfy $[T, R] = 0$, but if we know the only transformations that keep S invariant, we are at least not groping in the dark trying to guess which transformations we should be testing on the Weyl anomaly coefficients.

Recently, a set of corrected duality transformations has been found[11] that keep invariant the two-loop EA in a particular scheme. In that scheme, it is claimed that the matrix G_{ij} connecting the equations of motion and the Weyl anomaly coefficients is purely numerical, that is, it contains no spacetime derivative operators acting on $\bar{\beta}^i$.[10,12] That is certainly a crucial advantage if one is interested in studying a c-theorem for generalized sigma models. For our purposes, however, the disadvantage of that scheme is the fact that the expression for the EA is very complicated, containing a large number of scheme dependent terms as compared to the "minimal" EA that reproduces string scattering amplitudes. Furthermore, that EA does not satisfy Eq. (29), a property we would like to maintain; instead, the so-called "minimal" EA, S_{min}, does.[13] We would therefore like to obtain all our results in that scheme if possible.

In order to do this, we will study the general problem of scheme dependence, to determine whether we can find a set of duality transformations that keeps an EA invariant in one scheme if another set of transformations is given that keeps the EA invariant in another scheme.

In the generic notation of the introduction, we assume we are given an EA in one particular scheme to two-loop order,

$$S(k) = \alpha' S_0(k) + \alpha'^2 S_1(k) , \tag{35}$$

and a set of (two-loop corrected) duality transformations

$$\tilde{k}^i(k) = \tilde{k}_0^i(k) + \alpha' \tilde{k}_1^i(k) , \tag{36}$$

such that $S(\tilde{k}) = S(k)$ to $\mathcal{O}(\alpha'^2)$. Thus, $S_0(k)$ is given by Eq. (28), $S_1(k)$ may be for instance the two-loop EA in the scheme considered in [10,11]; $\tilde{k}_0(k)$ is given by Eq. (7), and $\tilde{k}_1(k)$ would then be the corrections to duality found in [11]. We now make a field redefinition

$$\hat{k}^i(k) = k^i + \alpha' f^i(k) , \tag{37}$$

with $f^i(k)$ some functional of the couplings with the appropriate dimensions. The field redefined EA, to $O(\alpha'^2)$, will be

$$\hat{S}(k) \equiv S(\hat{k}) = S(k) + \alpha' f^i(k) \cdot \frac{\delta S(k)}{\delta k^i} . \tag{38}$$

To the order considered, the invariance $S(\tilde{k}) = S(k)$ assumed implies

$$S_0(k) = S_0(\tilde{k}_0)$$

$$S_1(k) = S_1(\tilde{k}_0) + \tilde{k}_1^i(k) \cdot \frac{\delta S_0(\tilde{k}_0)}{\delta \tilde{k}_0^i} \ . \tag{39}$$

We now write

$$\tilde{\kappa}^i(k) = \tilde{k}_0^i(k) + \alpha' \ \tilde{\kappa}_1^i(k) \tag{40}$$

for the duality transformations that will keep the field redefined EA, $\hat{S}(k)$, invariant:

$$\hat{S}(\tilde{\kappa}(k)) = \hat{S}(k) \ . \tag{41}$$

To determine these field redefined duality transformations, one must now substitute Eq. (40) into Eq. (41), using Eqs. (37,38,39), and keeping terms to $O(\alpha'^2)$. This is done in a straightforward way, and we simply state the final result:

$$\tilde{\kappa}_1^i(k) = \tilde{k}_1^i(k) + \left(\frac{\delta \tilde{k}_0^i}{\delta k^j} \cdot f^j(k) - f^i(\tilde{k}_0) \right) \ . \tag{42}$$

This is the result we sought: given a set of transformations keeping the EA invariant in some scheme, we can explicitly construct the set of transformations keeping the EA invariant in any other scheme. It is interesting to note that the term in parenthesis on the r.h.s. above represents precisely the one-loop consistency conditions $[T, R] = 0$, but acting on the field redefinitions rather than on the Weyl anomaly coefficients. In other words, in changing from one scheme to another through a field redefinition, the duality transformations that keep the redefined EA invariant differ from the original transformations by a term which "corrects" for how much off the field redefinitions themselves are from satisfying the one-loop consistency conditions.

The minimal EA, $S_{min} = \alpha' S_0 + \alpha'^2 S_{1\min}$,

$$S_{1min} = \frac{1}{4} \int d\theta \ d^D x \ \sqrt{g} e^{-2\phi} \left(R_{\mu\nu\lambda\rho} R^{\mu\nu\lambda\rho} - \frac{1}{2} R^{\mu\nu\lambda\rho} H_{\mu\nu\sigma} H_{\lambda\rho}{}^{\sigma} \right.$$
$$\left. + \frac{1}{24} H_{\mu\nu\lambda} H^{\nu}{}_{\rho\alpha} H^{\rho\sigma\lambda} H_{\sigma}{}^{\mu\alpha} - \frac{1}{8} H_{\mu\alpha\beta} H_{\nu}{}^{\alpha\beta} H^{\mu\rho\sigma} H^{\nu}{}_{\rho\sigma} \right) \ , \tag{43}$$

the field redefinition taking the nonminimal action of [10,11] into it, and the beta functions in several different subtraction schemes, can all be gleaned from the literature.[10,13,14] The task at hand is now to find the duality transformations that keep Eq. (43) invariant and, using those as the operation T, and the beta functions in the appropriate scheme to define R, verify whether $[T, R] = 0$ holds at two-loop order. It should be noted that what we have done above *guarantees* that there exists a set of duality transformations that keeps S_{min} invariant; however, the constructive procedure, in Eq. (42), of obtaining these transformations starting from the transformations found in [11], is very likely not the most efficient way to proceed, and we have opted instead for direct guessing and verification on S_{min}. We expect to report on this in the near future.[6]

5. CONCLUSIONS

The requirement that duality and the RG commute as motions in the parameter space of a model is a very basic one, and it has been shown not only to be verified in the instances it has been tested, but also to yield important constraints on the RG flows in the context of 2*d* sigma models.

While at one-loop order this interplay between duality and the RG in the sigma model has been thoroughly investigated, at two-loop order the analysis has not been

exhaustive so far. To help us in achieving this complete analysis, we have available first of all a set of duality transformations keeping a string background effective action invariant. [11] We have shown that there is no guarantee that the set of transformations that keeps this effective action invariant will also turn out to satisfy the duality consistency conditions $[T, R] = 0$. However, we have also seen that if any transformations at all exist that do satisfy the consistency conditions, they must also keep the effective action invariant (at least in the "minimal" scheme), so that by finding the transformations that keep the effective action invariant one is selecting the one set of transformations that has a chance of satisfying $[T, R] = 0$.

We believe this basic statement, $[T, R] = 0$, to be a more fundamental feature of the models in question than the invariance of the string background effective action, which follows from it (and which only is defined for sigma models). This has represented sufficient motivation for us to delve into the question of its validity in full generality at two-loop order, with the encouragement that the existence of a duality invariance of the string background effective action has already been shown in the same context.

Field redefinition ambiguities enter at this loop order as an added complication. We have taken the first step in comprehensively accounting for them by establishing that duality symmetry is a well-defined notion over and above the presence of such ambiguities, in the sense that if it is present in one choice of scheme, it may be modified but will nonetheless also be present in any other scheme.

REFERENCES

1. P.H. Damgaard and P.E. Haagensen, cond-mat/9609242, *J. Phys.* **A30** (1997) 4681.
2. C.P. Burgess and C.A. Lütken, cond-mat/9611118, to be published in *Nucl. Phys.* **B.** See also C.A. Lütken, *Nucl. Phys.* **B396** (1993) 670 for earlier considerations of how modular symmetry may constrain RG flows.
3. P.E. Haagensen, *Phys. Lett.* **382B** (1996) 356.
4. P.E. Haagensen and K. Olsen, hep-th/9704157, to be published in *Nucl. Phys.* **B.**
5. P.E. Haagensen, K. Olsen and R. Schiappa, hep-th/9705105.
6. P.E. Haagensen, S. Penati, A. Refolli and D. Zanon, work in progress.
7. T.H. Buscher, *Phys. Lett.* **194B** (1987) 59; *Phys. Lett.* **201B** (1988) 466.
8. A.A. Tseytlin, *Phys. Lett.* **178B** (1986) 34; *Nucl. Phys.* **B294** (1987) 383.
9. These were first written down explicitly in A.A. Tseytlin, *Mod. Phys. Lett.* **A6** (1991) 1721. Corrections to duality transformations were also considered in J. Panvel, *Phys. Lett.* **284B** (1992) 50, and in J. Balog, P. Forgács, Z. Horváth and L. Palla, *Nucl. Phys.* (Proc. Suppl.) **B49** (1996) 16; *Phys. Lett.* **388B** (1996) 121. In the case of D target isometries, duality at two loop order was also considered in K. Meissner, *Phys. Lett.* **392B** (1997) 298.
10. I. Jack and D.R.T. Jones, *Nucl. Phys.* **B303** (1988) 260.
11. N. Kaloper and K. Meissner, hep-th/9705193.
12. N.E. Mavromatos and J.L. Miramontes, *Phys. Lett.* **201B** (1988) 473.
13. R.R. Metsaev and A.A. Tseytlin, *Nucl. Phys.* **B293** (1987) 385.
14. D. Zanon, *Phys. Lett.* **191B** (1987) 363.

This page intentionally left blank

UNIFICATION OF THE GENERAL NON-LINEAR SIGMA MODEL AND THE VIRASORO MASTER EQUATION

J. de Boer* and M.B. Halpern†

Department of Physics, University of California at Berkeley
366 Le Conte Hall, Berkeley, CA 94720-7300, U.S.A.
and
Theoretical Physics Group, Mail Stop 50A–5101
Ernest Orlando Lawrence Berkeley National Laboratory
Berkeley, CA 94720, U.S.A.

1. INTRODUCTION

There have been two broadly successful approaches to the construction of conforma1 field theories,

- The general affine-Virasoro construction[1–7]
- The general non-linear sigma model[8–13] (1)

but, although both approaches have been formulated as Einstein-like systems[12, 2], the relation between the two has remained unclear.

This talk summarizes recent work[14] which unifies these two approaches, following the organization of Fig. 1. The figure shows the two developments (1) with the left column (the general affine-Virasoro construction) as a special case of the right column (the general non-linear sigma model). Our goal here is to explain the unification shown in the lower right of the figure.

In the general affine-Virasoro construction, a large class of exact Virasoro operators[1, 3]

$$T = L^{ab} {}^{*}_{*} J_a J_b {}^{*}_{*} + iD^a \partial J_a, \qquad \text{a, b} = 1 \ldots \dim(g) \tag{2}$$

are constructed as quadratic forms in the currents J of the general affine Lie algebra[15,16]. The coefficients $L^{ab} = L^{ba}$ and D^a are called the inverse inertia tensor and the improvement vector respectively. The general construction is summarized[1, 3] by the (improved) Virasoro master equation (VME) for L and D, and this approach is the basis of irrational conformal field theory[7] which includes the affine-Sugawara[16–19] and coset

*e-mail address: deboer@theorm.lbl.gov
†e-mail address: halpern@theor3.lbl.gov

New Developments in Quantum Field Theory
Edited by Damgaard and Jurkiewicz, Plenum Press, New York, 1998

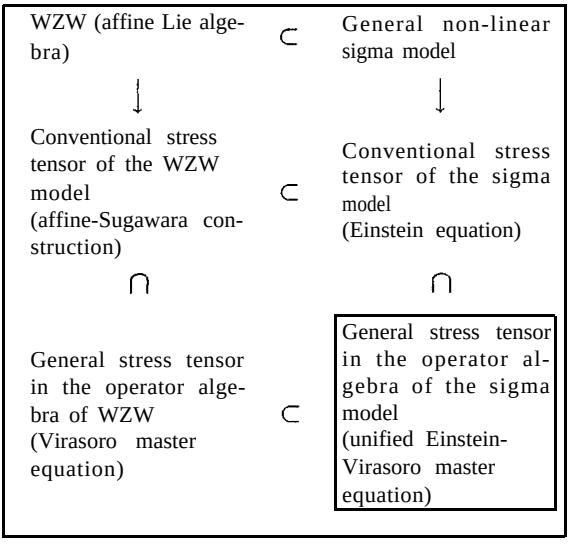


Figure 1. Conformal Field Theory

constructions[16, 17, 20] as a small subspace. The construction (2) can also be considered as the general Virasoro construction in the operator algebra of the WZW model[21, 22], which is the field-theoretic realization of the affine algebras. See Ref. 7 for a more detailed history of affine Lie algebra and the affine-Virasoro constructions.

For each non-linear sigma model, a Virasoro operator[23]

$$T = -\frac{1}{2\alpha'} G_{ij}\partial x^i \partial x^j + \mathcal{O}(\alpha'^0) = -\frac{1}{2\alpha'} G^{ab}\Pi_a \Pi_b + \mathcal{O}(\alpha'^0) \tag{3a}$$

$$G^{ab} = e_i{}^a G^{ij} e_j{}^b, \qquad \Pi_a = G_{ab} e_i{}^b \partial x^i, \qquad i,j,a,b = 1,\ldots,\dim(M) \tag{3b}$$

is constructed in a semiclassical expansion on an arbitrary manifold M, where G_{ij} is the metric on M and G^{ab} is the inverse of the tangent space metric. This is the canonical or conventional stress tensor of the sigma model and this construction is summarized[12, 23] by the Einstein equations of the sigma model, which couple the metric G, the antisymmetric tensor field B and the dilaton Φ. In what follows we refer to these equations as the conventional Einstein equations of the sigma model, to distinguish them from the generalized Einstein equations obtained below.

In this paper, we unify these two approaches, using the fact that the WZW action is a special case of the general sigma model. More precisely, we study the general Virasoro construction

$$T = -\frac{1}{\alpha'} L_{ij}\partial x^i \partial x^j + \mathcal{O}(\alpha'^0) = -\frac{1}{\alpha'} L^{ab}\Pi_a \Pi_b + \mathcal{O}(\alpha'^0) \tag{4a}$$

$$i,j,a,b = 1,\ldots,\dim(M) \tag{4b}$$

at one loop in the operator algebra of the general sigma model, where L is a symmetric second-rank spacetime tensor field, the inverse inertia tensor, which is to be determined. The unified construction is described by a system of equations which we call

- the Einstein-Virasoro master equation

of the general sigma model. This geometric system, which resides schematically in the lower right of Fig. 1, describes the covariant coupling of the spacetime fields L, G, B and Φ_a, where the vector field Φ_a generalizes the derivative $\nabla_a \Phi$ of the dilaton Φ.

The unified system contains as special cases the two constructions in (1): For the particular solution

$$L^{ab} = L^{ab}_G = \frac{G^{ab}}{2} + \mathcal{O}(\alpha'), \qquad \Phi_a = \Phi^G_a = \nabla_a \Phi \tag{5}$$

the general stress tensors (4) reduce to the conventional stress tensors (3) and the Einstein-Virasoro master equation reduces to the conventional Einstein equations of the sigma model. Moreover, the unified system reduces to the general affine-Virasoro construction and the VME when the sigma model is taken to be the WZW action. In this case we find that the contribution of Φ_a to the unified system is precisely the known improvement term of the VME.

More generally, the unified system describes a space of conformal field theories which is presumably much larger than the sum of the general affine-Virasoro construction and the sigma model with its canonical stress tensors.

2. BACKGROUND

To settle notation and fix concepts which will be important below, we begin with a brief review of the two known constructions in (1), which are the two columns of Fig. 1.

2.1. The General Affine-Virasoro Construction

The improved VME

The general affine-Virasoro construction, which is the left column of Figure 1, begins with the currents of a general affine Lie algebra [15, 16]

$$J_a(z)J_b(w) = \frac{G_{ab}}{(z-w)^2} + \frac{if_{ab}{}^c J_c(w)}{z-w} + \text{reg.} \tag{6}$$

where $a, b = 1 \ldots \dim g$ and $f_{ab}{}^c$ are the structure constants of g. For simple g, the central term in (6) has the form $G_{ab} = k\eta_{ab}$ where η_{ab} is the Killing metric of g and k is the level of the affine algebra. Then the general affine-Virasoro construction is [1]

$$T = L^{ab} {}^{*}_{*} J_a J_b {}^{*}_{*} + iD^a \partial J_a \tag{7}$$

where the coefficients $L^{ab} = L^{ba}$ and D^a are the inverse inertia tensor and the improvement vector respectively. The stress tensor T is a Virasoro operator

$$T(z)T(w) = \frac{c/2}{(z-w)^4} + \frac{2T(w)}{(z-w)^2} + \frac{\partial_w T(w)}{z-w} + \text{reg.} \tag{8}$$

iff the improved Virasoro master equation[1]

$$L^{ab} = 2L^{ac}G_{cd}L^{db} - L^{cd}L^{ef}f_{ce}{}^a f_{df}{}^b - L^{cd}f_{ce}{}^f f_{df}{}^{(a}L^{b)e} - f_{cd}{}^{(a}L^{b)c}D^d \tag{9a}$$

$$D^a(2G_{ab}L^{be} + f_{ab}{}^d L^{bc} f_{cd}{}^e) = D^e \tag{9b}$$

$$c = 2G_{ab}(L^{ab} + 6D^a D^b) \tag{9c}$$

is satisfied‡ by L and D, and the central charge of the construction is given in (9c). The unimproved VME[1,3] is obtained by setting the improvement vector D to zero.

K-conjugation covariance

A central property of the VME at zero improvement is K-conjugation covariance[16, 17, 20, 1] which says that all solutions come in K-conjugate pairs L and $\tilde{L}$,

$$L^{ab} + \tilde{L}^{ab} = L_g^{ab}, \qquad T + \tilde{T} = T_g, \qquad c + \tilde{c} = c_g \tag{10a}$$

$$T(z)\tilde{T}(w) = \text{reg.} \tag{10b}$$

whose K-conjugate stress tensors T, $\tilde{T}$ commute and add to the affine-Sugawara construction [15–18] on g

$$T_g = L_g^{ab} \,{}_*^*\, J_a J_b \,{}_*^*. \tag{11}$$

For simple g, the inverse inertia tensor of the affine-Sugawara construction is

$$L_g^{ab} = \frac{\eta^{ab}}{2k + Q_g} = \frac{\eta^{ab}}{2k} + \mathcal{O}(k^{-2}) = \frac{G^{ab}}{2} + \mathcal{O}(k^{-2}) \tag{12}$$

where η^{ab} is the inverse Killing metric of g and Q_g is the quadratic Casimir of the adjoint. K-conjugation covariance can be used to generate new solutions $\tilde{L} = L_g - L$ from old solutions L and the simplest application of the covariance generates the coset constructions [16, 17, 20] as $\tilde{L} = L_g - L_h = L_{g/h}$.

Semiclassical expansion

At zero improvement, the high-level or semiclassical expansion[24, 7] of the VME has been studied in some detail. On simple g, the leading term in the expansion has the form

$$L^{ab} = \frac{P^{ab}}{2k} + \mathcal{O}(k^{-2}), \qquad c = \text{rank}(P) + \mathcal{O}(k^{-1}) \tag{13a}$$

$$P^{ac}\eta_{cd}P^{db} = P^{ab} \tag{13b}$$

where P is the high-level projector of the L theory. These are the solutions of the classical limit of the VME,

$$L^{ab} = 2L^{ac}G_{cd}L^{db} + \mathcal{O}(k^{-2}) \tag{14}$$

but a semiclassical quantization condition[24] provides a restriction on the allowed projectors. In the partial classification of the space of solutions by graph theory [5, 25, 7], the projectors P are closely related to the adjacency matrices of the graphs.

Irrational conformal field theory

Given also a set of antiholomorphic currents $\bar{J}_a$, $a = 1 \ldots \dim(g)$, there is a corresponding antiholomorphic Virasoro construction

$$\bar{T} = L^{ab} \,{}_*^*\, \bar{J}_a \bar{J}_b \,{}_*^*\, + iD^a \partial \bar{J}_a \tag{15}$$

‡ Our convention is $A^{(a}B^{b)} = A^aB^b + A^bB^a$, $A^{[a}B^{b]} = A^aB^b - A^bB^a$.

with $\bar{c} = c$. Each pair of stress tensors T and $\bar{T}$ then defines a conformal field theory (CFT) labelled by L and D. Starting from the modules of affine $g \times g$, the Hilbert space of a particular CFT is obtained [26, 27, 7] by modding out by the local symmetry of the Hamiltonian.

It is known that the CFTs of the master equation have generically irrational central charge, even when attention is restricted to the space of unitary theories, and the study of all the CFTs of the master equation is called irrational conformal field theory (ICFT), which contains the affine-Sugawara and coset constructions as a small subspace.

In ICFT at zero improvement, world-sheet actions are known for the following cases: the affine-Sugawara constructions (WZW models [21, 22]), the coset constructions (spin-one gauged WZW models [28]) and the generic ICFT (spin-two gauged WZW models [26, 29, 30]). The spin-two gauge symmetry of the generic ICFT is a consequence of K-conjugation covariance.

See Ref. 7 for a comprehensive review of ICFT, and Ref. 31 for a recent construction of a set of semiclassical blocks and correlators in ICFT.

In this talk, we restrict ourselves to holomorphic stress tensors, and the reader is referred to Ref. 14 for the antiholomorphic version.

WZW model

The left column of Fig. 1 can be considered as the set of constructions in the operator algebra of the WZW model, which is affine Lie algebra.

The WZW action is a special case of the general nonlinear sigma model, where the target space is a group manifold G and g is the algebra of G.

2.2. The General Non-Linear Sigma Model

The general non-linear sigma model (the right column of Fig. 1) has been extensively studied [32, 33, 8, 34, 9, 10, 35, 36, 11, 12, 37, 38, 23, 13]

The Euclidean action of the general non-linear sigma model is

$$S = \frac{1}{2\alpha'} \int d^2z (G_{ij} + B_{ij}) \partial x^i \bar{\partial} x^j \tag{16a}$$

$$d^2z = \frac{dx dy}{\pi}, \qquad z = x + iy, \qquad H_{ijk} = \partial_i B_{jk} + \partial_j B_{ki} + \partial_k B_{ij}. \tag{16b}$$

Here x^i, $i = 1 \ldots \dim(M)$ are coordinates with the dimension of length on a general manifold M and α', with dimension length squared, is the string tension or Regge slope. The fields G_{ij} and B_{ij} are the (covariantly constant) metric and antisymmetric tensor field on M.

We also introduce a covariantly constant vielbein $e_i{}^a$, $a = 1 \ldots \dim(M)$ on M and use it to translate between Einstein and tangent-space indices, e.g. $G_{ij} = e_i{}^a G_{ab} e_j{}^b$, where G_{ab} is the covariantly constant metric on tangent space. Covariant derivatives are defined as usual in terms of the spin connection, $R_{ija}{}^b$ is the Riemann tensor and $R_{ab} = R_{acb}{}^c$ is the Ricci tensor. It will also be convenient to define the generalized connections and covariant derivatives with torsion,

$$\hat{\nabla}_i^{\pm} v_a = \partial_i v_a - \hat{\omega}_{ia}^{\pm b} v_b \tag{17a}$$

$$\hat{\omega}_{ia}^{\pm b} = \omega_{ia}{}^b \pm \frac{1}{2} H_{ia}{}^b \tag{17b}$$

$$\hat{R}_{ija}^{\pm}{}^b = (\partial_i \hat{\omega}_j^{\pm} - \partial_j \hat{\omega}_i^{\pm} - [\hat{\omega}_i^{\pm}, \hat{\omega}_j^{\pm}])_a{}^b \tag{17c}$$

where $\hat{\omega}^{\pm}_{iab}$ is antisymmetric under (a, b) interchange and $\hat{R}^{\pm}_{ijab}$ is pairwise antisymmetric in (i, j) and (a, b).

Following Banks, Nemeschansky and Sen[23], the canonical or conventional stress tensors of the general sigma model have the form

$$T_G = -\frac{G_{ij}}{2\alpha'}\partial x^i \partial x^j + \partial^2\Phi + T_1 + \mathcal{O}(\alpha') \tag{18a}$$

$$= -\frac{G^{ab}}{2\alpha'}\Pi_a\Pi_b + \partial^2\Phi + T_1 + \mathcal{O}(\alpha') \tag{18b}$$

$$\Pi_a = G_{ab}e_i{}^b\partial x^i, \qquad \bar{\Pi}_a = G_{ab}e_i{}^b\bar{\partial}x^i \tag{18c}$$

where Φ is the dilaton and T_1 is a finite one-loop counterterm which depends on the renormalization scheme. The condition that T_G is one-loop conformal reads [12]

$$R_{ij} + \frac{1}{4}(H^2)_{ij} - 2\nabla_i\nabla_j\Phi = \mathcal{O}(\alpha') \tag{19a}$$

$$\nabla^k H_{kij} - 2\nabla^k\Phi H_{kij} = \mathcal{O}(\alpha') \tag{19b}$$

$$c_G = \bar{c}_G = \dim(M) + 3\alpha'(4|\nabla\Phi|^2 - 4\nabla^2\Phi + R + \frac{1}{12}H^2) + \mathcal{O}(\alpha'^2) \tag{19c}$$

where (19a) and (19b) are the conventional Einstein equations of the sigma model and (19c) is the central charge of the construction. The result for the central charge includes two-loop information, but covariant constancy of the field-dependent part of the central charge follows by Bianchi identities from the Einstein equations, so all three relations in (19) can be obtained with a little thought from the one-loop calculation. It will also be useful to note that the conventional Einstein equations (19a), (19b) can be written in either of two equivalent forms

$$\hat{R}^{\pm}_{ij} - 2\hat{\nabla}^{\pm}_i\hat{\nabla}^{\pm}_j\Phi = \mathcal{O}(\alpha') \tag{20}$$

by using the generalized quantities (17) with torsion.

WZW data

The WZW action is a special case of the general sigma model (16a) on a group manifold G. Identifying the vielbein e on M with the left-invariant vielbein e on G, we find that $J_a = \frac{i}{\sqrt{\alpha'}}\Pi_a$ are the classical currents of WZW and

$$G_{ab} = k\eta_{ab}, \qquad H_{ab}{}^c = \frac{1}{\sqrt{\alpha'}}f_{ab}{}^c. \tag{21}$$

Here $f_{ab}{}^c$ and η_{ab} are the structure constants and the Killing metric of g and k is the level of the affine algebra. From this data, one also computes

$$\omega_{ab}{}^c = -\frac{1}{2\sqrt{\alpha'}}f_{ab}{}^c \tag{22a}$$

$$\hat{\omega}^{+}_{ab}{}^c = 0, \qquad \hat{\omega}^{-}_{ab}{}^c = -\frac{1}{\sqrt{\alpha'}}f_{ab}{}^c \tag{22b}$$

$$\hat{R}^{\pm}_{ija}{}^b = 0. \tag{22c}$$

Manifolds with vanishing generalized Riemann tensors are called parallelizable[35, 37].

2.3. Strategy

As seen in Fig. 1, our strategy here is a straightforward generalization of the VME to the sigma model, following the relation of the general affine-Virasoro construction to the WZW model. In the operator algebra of the general sigma model, we use the technique of Banks et al.[23] to study the general Virasoro construction

$$T = -\frac{L_{ij}}{\alpha'}\partial x^i \partial x^j + \mathcal{O}(\alpha'^0) = -\frac{L^{ab}}{\alpha'}\Pi_a \Pi_b + \mathcal{O}(\alpha'^0) \tag{23a}$$

$$\bar{\partial} T = 0 \tag{23b}$$

$$< T(z)T(w) >= \frac{c/2}{(z-w)^4} + \frac{2 < T(w) >}{(z-w)^2} + \frac{< \partial T(w) >}{(z-w)} + \text{reg.} \tag{23c}$$

where the dilatonic contribution is included at $\mathcal{O}(\alpha'^0)$ and L is a symmetric second-rank spacetime tensor field (the inverse inertia tensor) to be determined.

It is clear that this one-loop construction includes the conventional stress tensor T_G of the general sigma model, as well as the general affine-Virasoro construction when the sigma model is chosen to be WZW.

3. CLASSICAL PREVIEW OF THE CONSTRUCTION

The classical limit of the general construction (23a) can be studied with the classical equations of motion of the general sigma model

$$\bar{\partial}\Pi_a + \bar{\Pi}_b \Pi_c \hat{\omega}^{+bc}{}_a = 0 \tag{24}$$

where $\Pi, \bar{\Pi}$ are defined in (18c) and $\hat{\omega}^{\pm}$ are the generalized connections (17b) with torsion.

One then finds that the classical stress tensor is holomorphic

$$T = -\frac{L^{ab}}{\alpha'}\Pi_a \Pi_b, \qquad \bar{\partial} T = 0 \tag{25}$$

iff the inverse inertia tensor is covariantly constant

$$\hat{\nabla}^+_i L^{ab} = 0 \tag{26}$$

where $\hat{\nabla}^{\pm}$ are the generalized covariant derivatives (17a) with torsion. Further discussion of this covariant-constancy condition is found in Sections 5.2 and especially 5.5, which places the relation in a more geometric context.

To study the classical Virasoro conditions, we introduce Poisson brackets in Minkowski space, and study the classical chiral stress tensor

$$T_{++} = \frac{1}{8\pi\alpha'} L^{ab} J^+_a J^+_b \tag{27}$$

where J^+_a is the Minkowski-space version of Π_a. This stress tensor satisfies the equal-time Virasoro algebra iff

$$L^{ab} = 2L^{ac} G_{cd} L^{db}, \tag{28}$$

which is the analogue on general manifolds of the high-level or classical limit (14) of the VME on group manifolds.

4. THE UNIFIED EINSTEIN-VIRASORO MASTER EQUATION

We summarize here the results obtained by enforcing the Virasoro condition (23c) at one loop. Details of the relevant background field expansions, Feynman diagrams and dimensional regularization can be found in Ref. 14.

Including the one-loop dilatonic and counterterm contributions, the holomorphic stress tensor T is

$$T = -L^{ab}\left(\frac{\Pi_a\Pi_b}{\alpha'} + \frac{1}{2}\Pi_c\Pi_d H_{ae}{}^{c}H_b{}^{ed}\right) + \partial(\Pi_a\Phi^a) + \mathcal{O}(\alpha') \tag{29a}$$

$$a, b = 1, \ldots, \dim(M) \tag{29b}$$

where $L^{ab} = L^{ba}$ is the inverse inertia tensor and Π_a is defined in (18c). The second term in T is a finite one-loop counterterm which characterizes our renormalization scheme. The quantity Φ^a in (29a) is called the dilaton vector, and we will see below that the dilaton vector includes the conventional dilaton as a special case.

The necessary and sufficient condition that T satisfies the Virasoro algebra is the unified Einstein-Virasoro master equation

$$L^{cd}\hat{R}^{+}_{acdb} + \hat{\nabla}^{+}_{a}\Phi_b = \mathcal{O}(\alpha') \tag{30a}$$

$$\Phi_a = 2L_a{}^{b}\Phi_b + \mathcal{O}(\alpha') \tag{30b}$$

$$\hat{\nabla}^{+}_{i}L^{ab} = \mathcal{O}(\alpha') \tag{30c}$$

$$\begin{aligned} L^{ab} =& 2L^{ac}G_{cd}L^{db} \\ &- \alpha'(L^{cd}L^{ef}H_{ce}{}^{a}H_{df}{}^{b} + L^{cd}H_{ce}{}^{f}H^{(a}_{df}L^{b)e}) \\ &- \alpha'(L^{c(a}G^{b)d}\nabla_{[c}\Phi_{d]}) + \mathcal{O}(\alpha'^2) \end{aligned} \tag{30d}$$

$$c = 2G_{ab}L^{ab} + 6\alpha'(2\Phi_a\Phi^a - \nabla_a\Phi^a) + \mathcal{O}(\alpha'^2) \tag{30e}$$

where the first line of (30d) is the classical master equation in (28).

In what follows, we refer to (30a) as the generalized Einstein equation of the sigma model, and equation (30b) is called the eigenvalue relation of the dilaton vector. Equation (30d) is called the generalized Virasoro master equation (VME) of the sigma model. The central charge (30e) is consistent[14] by Bianchi identities with the rest of the unified system. The $\mathcal{O}(\alpha')$ corrections to the covariant-constancy condition (30c) can be computed in principle from the solutions of the generalized VME.

Some simple observations

1. Algebraic form of the generalizedVME. In parallel with the VME, the generalized VME (30d) is an algebraic equation for L. This follows because any derivative of L can be removed by using the covariant-constancy condition (30c).

2. Semiclassical solutions of the generalized VME. The solutions of (30c) and (30d) have the form

$$L_a{}^{b} = \frac{1}{2}P_a{}^{b} + \mathcal{O}(\alpha') \tag{31a}$$

$$\hat{\nabla}^{+}_{i}P_a{}^{b} = 0 \tag{31b}$$

where P is a covariantly constant projector, in parallel with the form (13) of the high-level solutions of the VME. The solutions of (31b) are further discussed in Section 5.5.

3. Correspondence with the VME. The non-dilatonic terms of the generalized VME (30d) have exactly the form of the unimproved VME (see eq. (9a)), after the covariant substitution

$$f_{ab}{}^{c} \to \sqrt{\alpha'} H_{ab}{}^{c} \tag{32}$$

for the general sigma model. This correspondence is the inverse of the WZW datum in (21),

$$H_{ab}{}^{c} = \frac{1}{\sqrt{\alpha'}} f_{ab}{}^{c} \tag{33}$$

which means that, for the special case of WZW, the non-dilatonic terms of the generalized VME will reduce correctly to those of the unimproved VME. We return to complete the WZW reduction in Section 5.2.

4. Dilaton solution for the dilaton vector. According to the classical limit (28) of the generalized VME, one solution of the eigenvalue relation (30b) for the dilaton vectors is

$$\Phi^{a}(\Phi) \equiv 2L^{ab}\nabla_{b}\Phi \tag{34}$$

In what follows, this solution is called the dilaton solution, and we shall see in the following section that the scalar field Φ is in fact the conventional dilaton of the sigma model.

5. PROPERTIES OF THE UNIFIED SYSTEM

5.1. The Conventional Stress Tensors of the Sigma Model

In this section, we check that the conventional stress tensors of the sigma model are correctly included in the unified system.

In the full system, the conventional stress tensor T_G of the sigma model corresponds to the particular solution of the generalized VME whose classical limit is

$$L^{ab} = L_G^{ab} = \frac{G^{ab}}{2} + \mathcal{O}(\alpha') \tag{35}$$

where G^{ab} is the inverse of the metric in the sigma model action. The covariant-constancy condition (30c) is trivially solved to this order because $\hat{\nabla}_i^{\pm} G_{ab} = 0$.

To obtain the form of T_G through one loop, we must also take the dilaton solution (34) for the dilaton vector, so that the dilaton contributes to the system as

$$\Phi_a = \Phi_a^G = \nabla_a \Phi + \mathcal{O}(\alpha'), \qquad \nabla_{[a}\Phi_{b]}^G = \mathcal{O}(\alpha'). \tag{36}$$

The relations (35) and (36) then tell us that the generalized Einstein equation (30a) simplifies to the conventional Einstein equation

$$\hat{R}_{ab}^{\pm} - 2\hat{\nabla}_a^{\pm}\hat{\nabla}_b^{\pm}\Phi = \mathcal{O}(\alpha'). \tag{37}$$

Moreover, eq. (36) tells us that the dilaton terms do not contribute to the generalized VME in this case, and we may easily obtain

$$L^{ab} = L_G^{ab} = \frac{G^{ab}}{2} - \frac{\alpha'}{4}(H^2)^{ab} + \mathcal{O}(\alpha'^2) \tag{38a}$$

$$\hat{\nabla}_i^{\pm} L_G^{ab} = -\frac{\alpha'}{4}\hat{\nabla}_i^{\pm}(H^2)^{ab} + \mathcal{O}(\alpha'^2) \tag{38b}$$

$$T_G(\Phi) = -\frac{G^{ab}}{2\alpha'}\Pi_a\Pi_b + \partial^2\Phi + \mathcal{O}(\alpha') \tag{38c}$$

by solving the generalized VME through the indicated order. In this case, the stress tensor counterterm in (29a) cancels against the $\mathcal{O}(\alpha)$ correction to L_G, and (38c) are consistent with (18). In what follows, the stress tensor $T_G(\Phi)$ is called the conventional stress tensor of the sigma model.

To complete the check, we evaluate the central charge $c = c_G(\Phi)$ in this case,

$$c_G(\Phi) = 2G_{ab}\left(\frac{G^{ab}}{2} - \frac{\alpha'}{4}(H^2)^{ab}\right) + 6\alpha'(2|\nabla\Phi|^2 - \nabla^2\Phi) + \mathcal{O}(\alpha'^2) \tag{39a}$$

$$= \dim(M) + 3\alpha'(4|\nabla\Phi|^2 - 2\nabla^2\Phi - \frac{1}{6}H^2) + \mathcal{O}(\alpha'^2) \tag{39b}$$

$$= \dim(M) + 3\alpha'(4|\nabla\Phi|^2 - 4\nabla^2\Phi + R + \frac{1}{12}H^2) + \mathcal{O}(\alpha'^2) \tag{39c}$$

which agrees with the conventional central charge in (19c). To obtain the usual form in (39c), we used the conventional Einstein equations (19a) in the form $R = 2\nabla^2\Phi - \frac{1}{4}H^2$.

We also note the form of the system for $L = L_G$ with general dilaton vector Φ_a^G,

$$T_G(\Phi_a) = -\frac{G^{ab}}{2\alpha'}\Pi_a\Pi_b + \partial(\Pi^a\Phi_a^G) + \mathcal{O}(\alpha') \tag{40a}$$

$$c_G(\Phi_a) = \dim(M) + 3\alpha'(4\Phi_a^G\Phi_G^a - 4\nabla_a\Phi_G^a + R + \frac{1}{12}H^2) + \mathcal{O}(\alpha'^2) \tag{40b}$$

$$\hat{R}^+_{ab} - 2\hat{\nabla}^+_a\Phi_b^G = \mathcal{O}(\alpha') \tag{40c}$$

where Φ_a^G is unrestricted because its eigenvalue equation is trivial.

5.2. WZW and the Improved VME

In this section we check that, for the special case of WZW, the unified system reduces to the improved VME (9a), where the improvement vector D is constructed from the general dilaton vector.

Using the WZW datum above we find that the generalized VME (30d) has the form

$$L^{ab} = (\text{usual } L^2 \text{ and } L^2f^2 \text{ terms}) + \sqrt{\alpha'}f_{cd}{}^{(a}L^{b)c}\Phi^d + \mathcal{O}(\alpha'^2) \tag{41}$$

when the sigma model is taken as WZW. The terms in parentheses are the usual terms (see eq. (9a)) of the unimproved VME. Next, we solve the generalized Einstein equation (30a) to find (using $\hat{R}^{\pm} = 0$) that the dilaton vector is a constant

$$\partial_i\Phi^a = 0. \tag{42}$$

It follows that the dilaton vector can be identified with the improvement vector of the VME in (9a)

$$D^a \equiv -\sqrt{\alpha'}\Phi^a = \text{ constant.} \tag{43}$$

Moreover, the solution of the covariant-constancy condition (30c) is

$$\partial_i L^{ab} = 0, \qquad L^{ab} = \text{ constant} \tag{44}$$

because $\hat{\omega}^+ = 0$. This completes the recovery of the improved VME in (9a).

The central charge reduces in this case to

$$c = 2G_{ab}(L^{ab} + 6D^aD^b) + \mathcal{O}(k^{-2}) \tag{45}$$

in agreement with the central charge (9c) of the improved VME. We finally note that the eigenvalue relation (30b) of the dilaton vector can be rewritten with (43) as

$$2L^{ab}G_{bc}D^c = D^a + \mathcal{O}(k^{-2}) \tag{46}$$

which is recognized as the leading term of the exact eigenvalue relation (9b) of the improved VME. This completes the one-loop check of the unified Einstein-Virasoro master equation against the improved VME.

5.3. Alternate Forms of the Central Charge

Using the generalized Einstein equation and the generalized VME, the central charge (30e) can be written in a variety of forms,

$$c = 2L_a{}^a + 6\alpha'(2\Phi_a\Phi^a - \hat{\nabla}_a^+\Phi^a) + \mathcal{O}(\alpha'^2) \tag{47a}$$

$$= 4L_a{}^bL_b{}^a + 2\alpha'\left[L_b{}^eL_d{}^fH^{bda}H_{efa} + 3(2\Phi_a\Phi^a - \hat{\nabla}_a^+\Phi^a)\right] + \mathcal{O}(\alpha'^2) \tag{47b}$$

$$= \mathrm{rank}(P) + 2\alpha'\left[L_a{}^bL_c{}^e(4L_d{}^f - 3\delta_d{}^f)H^{acd}H_{bef} + 3(2\Phi_a\Phi^a - \hat{\nabla}_a^+\Phi^a)\right] + \mathcal{O}(\alpha'^2) \tag{47c}$$

$$= \mathrm{rank}(P) + 2\alpha'\left[3L^{ab}\hat{R}_{ab}^+ + L_a{}^bL_c{}^e(4L_d{}^f - 3\delta_d{}^f)H^{acd}H_{bef} + 6(\Phi_a\Phi^a - \hat{\nabla}_a^+\Phi^a)\right] + \mathcal{O}(\alpha'^2) \tag{47d}$$

$$= 4L_a{}^bL_b{}^a + 2\alpha'\left[3L^{ab}\hat{R}_{ab}^+ + L_a{}^bL_c{}^eH^{acd}H_{bed} + 6(\Phi_a\Phi^a - \hat{\nabla}_a^+\Phi^a)\right] + \mathcal{O}(\alpha'^2). \tag{47e}$$

The first form in (47a) is the 'affine-Virasoro form' of the central charge. The form in (47d), with the first occurence of the generalized Ricci tensor, is called the 'conventional form' of the central charge because it reduces easily to the central charge of the conventional stress tensor

$$c_G(\Phi) = \dim(M) + 3\alpha'(4|\nabla\Phi|^2 - 4\nabla^2\Phi + R + \frac{1}{12}H^2) + \mathcal{O}(\alpha'^2) \tag{48}$$

when $P = G$ and $\Phi_a^G = \nabla_a\Phi$. The conventional form is also the form in which we found [14] it most convenient to prove the constancy of c

$$\partial_i c = \hat{\nabla}_i^+ c = \mathcal{O}(\alpha'^2) \tag{49}$$

using the Bianchi identities and the rest of the unified system. The final form of c in (47e) is the form which we believe comes out directly from the two-loop computation.

5.4. Solution Classes and a Simplification

Class I and Class II solutions

The solutions of the unified system (30) can be divided into two classes:

Class I. T conformal but $T_G(\Phi_a)$ not conformal

Class II. T and $T_G(\Phi_a)$ both conformal.

The distinction here is based on whether or not (in addition to the generalized Einstein equation) the dilaton-vector Einstein equation in (40a) is *also* satisfied. In the case when the dilaton solution $\Phi_a(\Phi)$ in (34) is taken for the dilaton vector, the question is whether or not the background sigma model is itself conformal in the conventional sense.

In Class I, we are constructing a conformal stress tensor T in the operator algebra of a sigma model whose conventional stress tensor $T_G(\Phi_a)$ with general dilaton vector is not conformal. This is a situation not encountered in the general affine-Virasoro construction because the conventional stress-tensor T_g of the WZW model is the affine-Sugawara construction, which is conformal. It is expected that Class I solutions are generic in the unified system, since there are so many non-conformal sigma models, but there are so far no non-trivial[§] examples (see however Ref. 40, which proposes a large set of candidates).

In Class II, we are constructing a conformal stress tensor T in the operator algebra of a sigma model whose conventional stress tensor $T_G(\Phi_a)$ with general dilaton vector is conformal. This class includes the case where the conventional stress tensors $T_G(\Phi)$ are conformal so that the sigma model is conformal in the conventional sense. The general affine-Virasoro construction provides a large set of non-trivial examples in Class II when the sigma model is the WZW action. Other examples are known from the general affine-Virasoro construction which are based on coset constructions, instead of WZW. In particular, Halpern et al. [41] construct exact Virasoro operators in the Hilbert space of a certain class of *g/h* coset constructions, and we are presently studying these conformal field theories as Class II solutions in the sigma model description of the coset constructions (see also the Conclusions).

It is also useful to subdivide Class II solutions into Class IIa and IIb. In Class IIb, we require the natural identification

$$\Phi_a = 2L_a{}^b\Phi_b^G + \mathcal{O}(\alpha') \tag{50}$$

which solves (30b), and Class IIa is the set of solutions in Class II without this identification. Note in particular that Class IIb contains all solutions in Class II with the dilaton solution $\Phi_a(\Phi)$ in (34).

Simplification for Class IIb with the dilaton solution

A simplification in Class IIb follows for the dilaton solution $\Phi_a(\Phi)$. In this case the unified system reads

$$R_{ij} + \frac{1}{4}(H^2)_{ij} - 2\nabla_i\nabla_j\Phi = \mathcal{O}(\alpha') \tag{51a}$$

$$\nabla^k H_{kij} - 2\nabla^k\Phi H_{kij} = \mathcal{O}(\alpha') \tag{51b}$$

$$\hat{\nabla}_i^+ L^{ab} = \mathcal{O}(\alpha'), \qquad \hat{\nabla}_i^- \bar{L}^{ab} = \mathcal{O}(\alpha') \tag{51c}$$

$$\begin{aligned} L^{ab} =& 2L^{ac}G_{cd}L^{db} \\ &- \alpha'(L^{cd}L^{ef}H_{ce}{}^aH_{df}{}^b + L^{cd}H_{ce}{}^fH_{df}^{(a}L^{b)e}) \\ &- \alpha'(L^{c(a}G^{b)d}\nabla_{[c}\Phi_{d]}) + \mathcal{O}(\alpha'^2) \end{aligned} \tag{51d}$$

$$c = 2G_{ab}L^{ab} + 6\alpha'(2\Phi_a\Phi^a - \nabla_a\Phi^a) + \mathcal{O}(\alpha'^2) \tag{51e}$$

$$\Phi_a = \Phi_a(\Phi) = 2L_a{}^b\nabla_b\Phi. \tag{51f}$$

This simplified system is close in spirit to the VME of the general affine-Virasoro construction: The solution of the conventional Einstein equation in (5la), (51b) provides

[§] Trivial examples in Class I are easily constructed as tensor products of conformal and non-conformal theories.

a conformal background, in which we need only look for solutions of the generalized VME in the form

$$L_a{}^b = \frac{P_a{}^b}{2} + \mathcal{O}(\alpha') \tag{52}$$

where P is a covariantly constant projector. Moreover, as in the VME, it has been shown[14] that all solutions of the simplified system (51) exhibit K-conjugation covariance, so that

$$\tilde{T} \equiv T_G - T, \qquad \tilde{c} = c_G - c \tag{53}$$

is also a conformal stress tensor when T is conformal.

5.5. Integrability Conditions

The inverse inertia tensor L^{ab} is a second-rank symmetric spacetime tensor, and we know that its associated projector P is covariantly constant

$$\hat{\nabla}_i^+ P_a{}^b = 0 \tag{54}$$

Operating with a second covariant derivative, we find that the integrability conditions

$$\hat{R}_{cd}^{+\,ae} P_e{}^b + \hat{R}_{cd}^{+\,be} P_e{}^a = 0 \tag{55}$$

follow as necessary conditions for the existence of solutions to (54).

On any manifold, there is always at least one solution to the covariant-constancy condition (54) and its integrability conditions (55), namely

$$P^{ab} = G^{ab} \tag{56a}$$

$$\hat{R}_{cd}^{\pm ab} + \hat{R}_{cd}^{\pm ba} = 0 \tag{56b}$$

$$L^{ab} = L_G^{ab} = \frac{G^{ab}}{2} + \mathcal{O}(\alpha') \tag{56c}$$

where G_{ab} is the metric of the sigma model action. This solution corresponds to the classical limit of the conventional sigma model stress tensor, as discussed in Section 5.1. For WZW, the integrability conditions (55) are also trivially satisfied (because $\hat{R}^{\pm}_{abcd} = 0$) and the general solutions of the covariant-constancy conditions were obtained for this case in Section 5.2.

In general we are interested in the classification of manifolds with at least one more solution P^{ab}, beyond G^{ab}. In what follows, we outline the sufficient and necessary condition for this phenomenon.

In a suitable basis, any projector P can be written as

$$P = \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & 0 \end{array}\right). \tag{57}$$

Inserting this form in (54) and (55) shows then that $\hat{R}^+$ and $\hat{\omega}^+$ must be 'block diagonal' in the same basis, i.e. they can be written as

$$(\hat{R}_{cd}^+)_a^b = \left(\begin{array}{c|c} A_{cd} & 0 \\ \hline 0 & D_{cd} \end{array}\right), \qquad (\hat{\omega}_i^+)_a^b = \left(\begin{array}{c|c} D_i & 0 \\ \hline 0 & E_i \end{array}\right) \tag{58}$$

for some matrices A_{cd}, B_{cd}, D_i, E_i. Thus, a necessary condition for new solutions to the covariant-constancy condition to exist is that $\hat{R}$ and $\hat{\omega}$ should be block diagonal.

Conversely, given a block diagonal $\hat{\omega}^+$, we can construct a new solution to the covariant-constancy condition with P given in (57). In fact, with $\hat{\omega}^+$ written in terms of the smallest possible blocks we can classify all possible solutions to the covariant constancy condition. If we denote the smallest diagonal blocks of $\hat{\omega}^+$ by $D_1, \ldots, D_k$, then the most general covariantly constant projector is

$$P = p_1 \mathbf{1}_1 + \ldots + p_k \mathbf{1}_k \tag{59}$$

where $p_i \in \{0, 1\}$ and $\mathbf{1}_j$ is the matrix which consists of the identity matrix in the j^{th} block and zeroes everywhere else. In the case when one of the blocks in $\hat{\omega}^+$ is zero, say D_j, then $p_j \mathbf{1}_j$ can be replaced by an arbitrary projector P_j in the j^{th} subspace.

New solutions obtained following this procedure are discussed in the Conclusions.

Mathematically, the problem of finding block-diagonal curvatures is the problem of finding manifolds with reducible holonomy. In the absense of torsion, it is known that block-diagonal curvatures exist only on product manifolds, but in the presence of torsion the question of manifolds with a block-diagonal curvature is an unsolved problem, except for the group manifolds discussed above (where $\hat{R}^+ = 0$), and the new examples given in the Conclusions.

6 . CONCLUSIONS

We have studied the general Virasoro construction

$$T = -\frac{L_{ij}}{\alpha'} \partial x^i \partial x^j + \mathcal{O}(\alpha'^0) \tag{60}$$

at one loop in the operator algebra of the general non-linear sigma model, where L is a spin-two spacetime tensor field called the inverse inertia tensor. The construction is summarized by a unified Einstein-Virasoro master equation which describes the covariant coupling of L to the spacetime fields G, B and Φ_a, where G and B are the metric and antisymmetric tensor of the sigma model and Φ_a is the dilaton vector, which generalizes the derivative $\nabla_a \Phi$ of the dilaton Φ. As special cases, the unified system contains the Virasoro master equation of the general affine-Virasoro construction and the conventional Einstein equations of the canonical sigma model stress tensors. More generally, the unified system describes a space of conformal field theories which is presumably much larger than the sum of these two special cases.

In addition to questions posed in the text, we list here a number of other important directions.

1. New solutions. It is important to find new solutions of the unified system, beyond the canonical stress tensors of the sigma model and the general affine-Virasoro construction.

Although it is not in the original paper[14], we have recently discovered a large class of new solutions of the covariant-constancy condition: It is not hard to see that the spin connection in the sigma model description of the g/h coset constructions has the form

$$(\hat{\omega}_i^+)_a{}^b = N_i{}^A f_{Aa}{}^b \tag{61}$$

where A is an h-index and a, b are g/h-indices, and $f_{Aa}{}^b$ are the structure constants of g. The structure constants and hence the spin connection can be taken block diagonal, where the blocks correspond to irreducible representations of h. As discussed in Section 5.5, this allows us to classify all possible covariantly-constant projectors on these manifolds. More work remains to be done in this case, including the solution of the generalized VME, but there are indications that the resulting conformal field theories

may be identified as the set of local Lie h-invariant conformal field theories[41] on g, which have in fact been studied in the Virasoro master equation itself.

2. Duality. The unified system contains the coset constructions in two distinct ways, that is, both as $G_{ab} = k\eta_{ab}$, $L^{ab} = L^{ab}_{g/h}$ in the general affine-Virasoro construction and among the canonical stress tensors of the sigma model with the sigma model metric that corresponds to the coset construction. This is an indicator of new duality transformations in the system, possibly exchanging L and G, which may go beyond the coset constructions. Indeed, if the conjecture of the previous paragraph holds, this duality would extend over all local Lie h-invariant conformal field theory, and perhaps beyond.

In this connection, we remind the reader that the VME has been identified[2] as an Einstein-Maxwell system with torsion on the group manifold, where the inverse inertia tensor is the inverse metric on tangent space. Following this hint, it may be possible to cast the unified system on group manifolds as two coupled Einstein systems, with exact covariant constancy of both G and L.

3. Non-renormalization theorems. The unified Einstein-Virasoro master equation is at present a one-loop result, while the Virasoro master equation is exact to all orders. This suggests a number of possibly exact relations[14] to all orders in the WZW model and in the general non-linear sigma model.

4. Spacetime action and/or C-function. These have not yet been found for the unified system, but we remark that they are known for the special cases unified here: The spacetime action[12, 42] is known for the conventional Einstein equations of the sigma model, and, for this case, the C-function is known[13] for constant dilaton. Moreover, an exact C-function is known[43] for the special case of the unimproved VME.

5. World-sheet actions. We have studied here only the Virasoro operators constructible in the operator algebra of the general sigma model, but we have not yet worked out the world-sheet actions of the corresponding new conformal field theories, whose beta functions should be the unified Einstein-Virasoro master equation. This is a familiar situation in the general affine-Virasoro construction, whose Virasoro operators are constructed in the operator algebra of the WZW model, while the world-sheet actions of the corresponding new conformal field theories include spin-one[28] gauged WZW models for the coset constructions and spin-two[26, 29, 30] gauged WZW models for the generic construction.

As a consequence of this development in the general affine-Virasoro construction, more or less standard Hamiltonian methods now exist for the systematic construction of the new world-sheet actions from the new stress tensors, and we know for example that K-conjugation covariance is the source of the spin-two gauge invariance in the generic case. At least at one loop, a large subset of Class IIb solutions of the unified system exhibit K-conjugation covariance, so we may reasonably expect that the world-sheet actions for generic constructions in this subset are spin-two gauged sigma models. For solutions with no K-conjugation covariance, the possibility remains open that these constructions are dual descriptions of other conformal sigma models.

6. Superconformal extensions. The method of Ref. 23 has been extended[44–46] to the canonical stress tensors of the supersymmetric sigma model. The path is therefore open to study general superconformal constructions in the operator algebra of the general sigma model with fermions. Such superconformal extensions should then include and generalize the known $N = 1$ and $N = 2$ superconformal master equations[47] of the general affine-Virasoro construction.

In this connection, we should mention that that the Virasoro master equation is the true master equation, because it includes as a small subspace all the solutions of

the superconformal master equations. It is reasonable to expect therefore that, in the same way, the unified system of this paper will include the superconformal extensions.

Acknowledgements

For many helpful discussions, we would like to thank our colleagues: L. Alvarez-Gaumé, I. Bakas, K. Clubok, S. Deser, A. Giveon, R. Khuri, E. Kiritsis, P. van Nieuwenhuizen, N. Obers, A. Sen, K. Sfetsos, K. Skenderis, A. Tseytlin and B. de Wit. We also thank the organizers of the Zakopane Conference for the opportunity to summarize this work.

This research was supported in part by the Director, Office of Energy Research, Office of High Energy and Nuclear Physics, Division of High Energy Physics of the U.S. Department of Energy under Contract DOE-AC03-76SF00098 and in part by the National Science Foundation under grant PHY-951497. JdB is a fellow of the Miller Institute for Basic Research in Science.

REFERENCES

1. M.B. Halpern and E. Kiritsis, *Mod, Phys. Lett.* A4:1373 (1989); Erratum *ibid.* A4:1797 (1989) .
2. M.B. Halpern and J.P. Yamron, *Nucl. Phys.* B332:411 (1990).
3. A.Yu Morozov, A.M. Perelomov, A.A. Rosly, M.A. Shifman and A.V. Turbiner, *Int. J. Mod. Phys.* A5:803 (1990).
4. M.B. Halpern, E. Kiritsis, N.A. Obers, M. Porrati and J.P. Yamron, *Int. J. Mod. Phys.* A5:2275 (1990).
5. M.B. Halpern and N.A. Obers, *Comm. Math. Phys.* 138:63 (1991).
6. M.B. Halpern and N.A. Obers, *J. Math. Phys.* 36:1080 (1995).
7. M.B. Halpern, E. Kiritsis, N.A. Obers and K. Clubok, *"Irrational Conformal Field Theory", Physics Reports* 265:1 (1996).
8. D. Friedan, "Nonlinear Models in 2 $+\epsilon$ Dimensions", PhD thesis, LBL-11517 (1980); *Phys. Rev. Lett.* 45:1057 (1980).
9. L. Alvarez-Gaumé, D.Z. Freedman and S. Mukhi, *Ann. of Phys.* 134:85 (1981).
10. C. Lovelace, *Phys. Lett.* 135B:75 (1984).
11. E.S. Fradkin and A.A. Tseytlin, *Nucl. Phys.* B261:1 (1985).
12. C.G. Callan, D. Friedan, E.J. Martinec and M.J. Perry, *Nucl. Phys.* B262:593 (1985).
13. A.A. Tseytlin, *Int. J. Mod. Phys.* A4:1257 (1989).
14. J. de Boer and M.B. Halpern, *Int. J. Mod. Phys.* A12:1551 (1997).
15. V.G. Kac, *Funct, Anal. App.* 1:328 (1967); R.V. Moody, *Bull. Am. Math. Soc.* 73:217 (1967).
16. K. Bardakçi and M.B. Halpern, *Phys. Rev.* D3:2493 (1971).
17. M.B. Halpern, *Phys. Rev.* D4:2398 (1971).
18. R. Dashen and Y. Frishman, *Phys. Rev.* D11:2781 (1975).
19. V.G. Knizhnik and A.B. Zamolodchikov, *Nucl. Phys.* B247:83 (1984).
20. P. Goddard, A. Kent and D. Olive, *Phys. Lett.* B152:88 (1985).
21. S.P. Novikov, Usp. Math. Nauk. 37:3 (1982).
22. E. Witten, *Comm. Math. Phys.* 92:455 (1984).
23. T. Banks, D. Nemeschansky and A. Sen, *Nucl. Phys.* B277:67 (1986).
24. M.B. Halpern and N.A, Obers, *Nucl. Phys.* B345:607 (1990).
25. M.B. Halpern and N.A. Obers, *J. Math. Phys.* 33:3274 (1992).
26. M.B. Halpern and J.P. Yamron, *Nucl. Phys.* B351:333 (1991).
27. M.B. Halpern, Recent progress in irrational conformal field theory, *in*: "Strings 1993", M.B. Halpern et al., eds., World Scientific, Singapore (1995).
28. K. Bardakçi, E. Rabinovici and B. Säring, *Nucl. Phys.* B299:151 (1988); D. Altschuler, K. Bardakçi and E, Rabinovici, *Comm. Math. Phys.* 118:241 (1988).
29. A.A. Tseytlin, *Nucl. Phys.* B418:173 (1994).
30. J. de Boer, K. Clubok and M.B. Halpern, *Int, J. Mod. Phys.* A9:2451 (1994).
31. M.B. Halpern and N.A. Obers, *Int. J. Mod. Phys.* A11:4837 (1996).

32. G. Ecker and J. Honercamp, *Nucl. Phys*, B35:481 (1971).
33. J. Honerkamp, *Nucl. Phys.* B36:130 (1972).
34. E.S. Fradkin and A.A. Tseytlin, *Phys. Lett.* 106B:63 (1981).
35. T.L. Curtright and C.K. Zachos, *Phys. Rev. Lett.* 53:1799 (1984).
36. S. Gates, C. Hull and M. Roček, *Nucl. Phys.* B248:157 (1984).
37. E. Braaten, T.L. Curtright and C.K. Zachos, *Nucl. Phys.* B260:630 (1985).
38. S. Mukhi, *Nucl. Phys.* B264:640 (1986).
39. M.B. Green, J.H. Schwarz and E. Witten, "Superstring Theory", Cambridge University Press, Cambridge (1987).
40. K. Clubok and M.B. Halpern, The generic world-sheet action of irrational conformal field theory, *in*: "Strings'95", I. Bars et al., eds., World Scientific, Singapore (1996).
41. M.B. Halpern, E. Kiritsis and N.A. Obers, *in*: "Infinite Analysis"; *Int. J. Mod. Phys.* A7, [Suppl. 1A]:339 (1992).
42. G. Horowitz, The dark side of string theory, *in*: "String Theory and Quantum Gravity 1992", hep-th/9210119.
43. A. Giveon, M.B. Halpern, E. Kiritsis and N.A. Obers, *Nucl. Phys.* B357:655 (1991).
44. G. Aldazabal, F. Hussain and R. Zhang, *Phys. Lett.* 185B:89 (1987).
45. T. Itoh and M. Takao, *Int. J. Mod. Phys.* A5:2265 (1990).
46. G. Aldazabal and J.M. Maldacena, *Int. J. Mod. Phys.* A8:3359 (1993).
47. A. Giveon, M.B. Halpern, E. Kiritsis and N.A. Obers, *Int. J. Mod. Phys.* A7:947 (1992).

This page intentionally left blank

A MATRIX MODEL SOLUTION OF THE HIROTA EQUATION

Vladimir A. Kazakov

[1] Laboratoire de Physique Théorique de l'École Normale Supérieure *
24 rue Lhomond, F-75231 Paris Cedex 05, France
and
Yukawa Institute for Theoretical Physics
Kyoto University, Sakyo-ku, Kyoto 606-01, Japan

ABSTRACT

We present a hermitian matrix chain representation of the general solution of the Hirota bilinear difference equation of three variables. In the large N limit this matrix model provides some explicit particular solutions of continuous differential Hirota equation of three variables. A relation of this representation to the eigenvalues of transfer matrices of 2D quantum integrable models is discussed.

INTRODUCTION

The Hirota bilinear equations (HE) [3] provide, may be, the most general view on the world of exactly solvable models, from integrable hierarchies of differential and difference equations equations, like KdV equations or Toda chains, to, as it was recently shown, the transfer matrices of 2D models of statistical mechanics and quantum field theory [4, 5, 1] integrable by the Bethe ansatz (BA) techniques. Indeed, in the last case, it was shown that the eigenvalues of transfer matrices must obey the general HE with 3 discrete variables corresponding to the rapidity, rank and level of a representation with rectangular Young tableaux, used in the fusion procedure. The Thermodynamical Bethe Ansatz (TBA) equations discovered by Lee and Yang and widely used in the last years for the description of thermal properties and finite size effects of 2D integrable models follow almost directly from this HE [6].

HE has its own history in connection with the matrix models (MM). The general solitonic solutions of HE [7] bear a striking resemblance with the matrix models with logarithmic potentials. The discovery of double scaling limit in the matrix models [8, 9, 10] (corresponding to the big size of matrices and a special tuning of potentials)

* Unité Propre du Centre National de la Recherche Scientifique, associée à l'École Normale Supérieure et à l'Université de Paris-Sud

New Developments in Quantum Field Theory
Edited by Damgaard and Jurkiewicz, Plenum Press, New York, 1998

showed that the matrix models are closely related to KdV and KP hierarchies of integrable differential equations [11]. The study following it provided even more general examples of this correspondence: the multi matrix models before taking any large N limit appeared to be related the classical Toda chains [12]. Another manifestation of these connections are the Schwinger-Dyson equation for MM which can be written in terms of Virasoro constraints [13, 14] (see [15] for a modern account of this approach).

We would also recall one rather mysterious coincidence: in [16] an open string amplitude for the (1+1)D string with mixed (Dirichlet-Neumann) boundary conditions appeared to have the same form as the Sine-Gordon S-matrix.

All this suggests that matrix models could have something to do with the quantum 2D integrable models and the way from one to another might go through HE.

In this paper we will propose a matrix chain representation of the general solution of Hirota difference equation. We will use for that the analogy between the so called Bazhanov-Reshetikhin determinant representation [17] of the transfer-matrix eigenvalues of integrable models (which obeys the HE) and the determinant representation of matrix chain, in terms of eigenvalues of the matrices. The potentials acting at every site of the chain will depend on the eigenvalues of the matrices and on the coordinate of the site. Hence the arbitrary potential is a function of 2 variables which is in general enough to parameterize any solution of a difference equation of 3 variables.

The solutions of HE relevant to TBA must obey very special boundary conditions. It can include, for example, the condition on the maximal possible size of the fusion representation (reflecting the invariance of the initial model with respect to some continuous symmetry), and Lorentz invariance of the spectrum of physical particles emerging on the top of physical (dressed) vacuum.

It is not easy to extract the physical information from HE or the corresponding TBA equations as it is not easy to solve the nonlinear difference or integral equations with specific boundary conditions. The results obtained on this way are quite limited: they mostly concern the calculations of central charges and dimensions of operators in the conformal (ultraviolet) limits and various asymptotic expansions corresponding to high energies (see for example [17, 18, 19]). It even appears to be difficult to reproduce the first loop calculation for the asymptotically free models in finite temperature, apart from some simplified models [20]), although numerically the TBA equations work quite well. Another challenge is to find the planar (large N) limit of such an interesting 2D quantum field theory as the principal chiral field. The model was formally integrated by the BA approach in [21, 22]. It has been solved rather explicitly in the large N limit in case of zero temperature and arbitrarily big external field [23, 24], but the attempts to generalize it to finite temperatures were not successful.

On the other hand, the general solution of HE follows from its integrability and can be represented in terms of a τ-function. In case of a general difference HE it coincides with the determinant representation of Bazhanov-Reshetikhin. It might be a good idea to use this representation, and hence our matrix chain representation, to get some hand on HE and TBA. Of course, for a finite rank N of matrices (and hence finite dimensional groups of symmetry of corresponding models) our representation hardly could offer some breakthrough. But for big N we can try to use the machinery of the matrix models (like orthogonal polynomials, character expansions and various saddle point techniques) to calculate the corresponding infinite determinant.

We were not able to find any solutions of HE satisfying correct TBA-like boundary conditions. A natural way to impose these boundary conditions is the most important drawback of our representation. Leaving it to future studies we propose here some particular solutions of continuous (differential) Hirota equations which describe our

matrix chain in the large N limit. They correspond to a particular choice of parameters (potentials) of the chain.

In the next section we will briefly review how the Hirota equation is connected with the integrable models of the 2D quantum field theory. The continuous differential version of them will be presented.

In section 3 we will propose the matrix chain representation of the difference Hirota equation based on the Bazhanov-Reshetikhin determinant representation of the fusion rules for transfer-matrices.

In section 4 we will present some particular examples of solution of the continuous differential HE, given by the one matrix model and the matrix oscillator with the specific boundary conditions.

In section 5 we will sketch out a general solution of the differential HE for an arbitrary time dependent matrix chain potential and consider a more explicit solution for the particular case of time independent potential.

The last section will be devoted to conclusions and prospects.

TBA, FUSION RULES AND HIROTA DIFFERENCE EQUATION

To set a more physical background for our construction let us briefly review how the HE appears from the Bethe ansatz. We will mostly follow in this section the framework and the notations of [1,2].

The transfer-matrix of an integrable 2D model with periodic boundary conditions depends (apart from a number of fixed parameters, like volume, temperature or the anisotropy q) on three variables: rapidity $u = i\theta$, rank ("color") a and level ("string" length) s. The variables a and s have the meaning of a representation of elementary spins filling the bare vacuum of the model, given by the rectangular Young tableau of the size $a \times s$. The corresponding transfer matrix is called $\hat{T}^a_s(u)$.

The integrability imposes the commutativity of transfer-matrices for different values of all three variables playing thus the role of spectral parameters:

$$[\hat{T}^a_s(u), \hat{T}^{a'}_{s'}(u')] = 0 \tag{1}$$

It follows from (1) that we can always work with the eigenvalues $T^a_s(u)$ instead of the transfer-matrix itself and view them as usual functions.

The transfer-matrices, as well as their eigenvalues, obey a set of relations known as fusion rules, originally found as the relations between S-matrices for particles with different spins in integrable QFT. They can be summarized in the so called Bazhanov-Reshetikhin formula [17] (BR) presenting the function of three variables $T^a_s(u)$ in terms of the function of only two variables $T^1_s(u)$:

$$T^a_s(u) = det_{1\le i,j\le a} T^1_{s+i-j}(u + i + j + a) \tag{2}$$

Actually, there exists a more general BR formula, expressing the transfer-matrix eigenvalue of any skew representation h/h' through $T^1_s(u)$:

$$T_{h/h'}(u) = det_{1\le i,j\le a} T^1_{h_i-h'_j}(u + h_i + h'_j) \tag{3}$$

where $h_i = m_i + a - i$ and $h'_i = m'_i + a - i$ are the so called shifted highest weight components of two representations R and R' of GL(N) characterized by the usual highest weight components $R = (m_1,...m_a)$ and $R' = (m'_1,...m'_a)$, so that they obey the inequalities $h_i < h_{i-1}$, $h'_i < h'_i$ and also $h'_i < h_i$. But the transfer-matrices with

rectangular Young tableaux play an exceptional role since they obey a closed set of fusion rules given by the difference Hirota equation:

$$T_s^a(u+1)T_s^a(u-1) - T_{s+1}^a(u)T_{s-1}^a(u) = T_s^{a+1}(u)T_s^{a-1}(u) \tag{4}$$

It follows directly from (2) in virtue of the Jacobi identity for determinants. It contains little information since it is true for any function of two variables $T_s^1(u)$ in (2). To specify it further to some particular integrable model we have to impose some boundary conditions on solutions of the eq. (4).

One of these conditions specifies the group of symmetry of the model. To make it, say, SU(N) (or A_{N-1} in terms of underlying algebra) we put:

$$T_s^a(u) = 0, \qquad for \quad a < 0 \; and \quad a > N \tag{5}$$

It is not enough since it leaves us with an infinite discrete set of possible solutions (like in quantum mechanics, fixing the boundary conditions on a wave function we are still left with infinitely many wave functions corresponding to different energy levels). We have to specify some analytical properties of solutions.

There are two ways to do it in the case of BA.

One is related to the so called bare BA where one specifies $T_s^0(u)$ and $T_s^N(u)$ to be some given polynomials in the variable u whose zeroes specify completely a model, where as the functions $T_s^a(u)$, for $1 \le a \le N-1$ are polynomials whose zeros we have to find. The details of the analyticity conditions for the bare BA can be found for example in (1).

Another way to fix analytical properties corresponds to the dressed BA where the elementary excitations are already the real physical particles. To precise them let us derive from (4) the TBA equations. For that we introduce the function:

$$Y_s^a(u) = \frac{T_{s+1}^a(u)T_{s-1}^a(u)}{T_s^{a+1}(u)T_s^{a-1}(u)} \tag{6}$$

which, in virtue of (4), satisfies the equation sometimes called Y-system [6]:

$$\frac{Y_s^a(u+1)Y_s^a(u-1)}{Y_s^{a+1}(u)Y_s^{a-1}(u)} = \frac{[1+Y_{s+1}^a(u)][1+Y_{s-1}^a(u)]}{[1+Y_s^{a+1}(u)][1+Y_s^{a-1}(u)]} \tag{7}$$

Note that this system is symmetric under the change: $Y \to Y^{-1}$, $a \to s$, $s \to a$ (rank-level duality).

To make it a little bit more symmetric let us introduce the functions:

$$U_s^a(u) = 1 + Y_s^a(u), \qquad \tilde{U}_s^a(u) = 1 + \frac{1}{Y_s^a(u)} \tag{8}$$

Then the eq. (7) can be rewritten as

$$\frac{U_s^a(u+1)U_s^a(u-1)}{U_{s+1}^a(u)U_{s+1}^a(u)} = \frac{\tilde{U}_s^a(u+1)\tilde{U}_s^a(u-1)}{\tilde{U}_s^{a+1}(u)\tilde{U}_s^{a-1}(u)} \tag{9}$$

Taking logarithm of both sides of this equation and applying the operator

$$\int_{-\infty}^{\infty} d\theta' \frac{1}{\cosh\frac{\pi}{2}(\theta-\theta')} \cdots \tag{10}$$

where $\theta = iu$ we obtain:

$$C_{ss'}(\theta) * \ln(1+Y_{s'}^a(\theta)) = C^{aa'}(\theta) * \ln(1+\frac{1}{Y_{s'}^a(\theta)}) \tag{11}$$

where we introduced the so called "baxterized" Cartan matrices:

$$C^{aa'}(\theta) = \delta_{a,a'} - \frac{1}{2\cosh\frac{\pi}{2}(\theta)}(\delta_{a,a'+1} + \delta_{a+1,a'}) \tag{12}$$

and similarly for $C_{ss'}$. The $*$ sign defines the usual convolution operation: $f(\theta) * g(\theta) = \int_{-\infty}^{\infty} d\theta' f(\theta - \theta')g(\theta')$.

Let us now act on both sides of eq. (11) by the operator inverse to (12):

$$A_N^{aa'} = C_{aa'}^{-1}(\theta) = \frac{2}{\pi}\int_0^{\infty} dp \cos(p\theta)\coth(p)\frac{\sinh[(N - max(aa'))p]\sinh[min(aa')p]}{\sinh(pN)} \tag{13}$$

Note that this inverse is respecting the boundary conditions restricting the values of a, a' to $1 \le a, a' \le N - 1$.

Note also that the operator $C^{aa'}$ has zero modes:

$$C^{aa'}(\theta) * \sigma(s) m_0 \sin(a'\pi k/N)\cosh(\theta\pi k/N) = 0 \tag{14}$$

for any integer k and any function $\sigma(s)$. So in acting by (13) on both sides of (11) we might be obliged to add one of zero modes. The choice of zero mode and the function $\sigma(s)$ defines completely the boundary conditions and hence the model. If we want to respect the 1+1 dimensional relativistic invariance we can add the zero mode with $k = 1$, since only it will lead to the relativistic spectrum of energies elementary excitations (which are described by this zero mode) of a type:

$$\sigma(s) m_0 \sin(a\pi/N)\cosh(\theta\pi/N) \tag{15}$$

It gives a typical mass and energy spectrum of physical particles for integrable relativistic models of 2D QFT. The choice of $\sigma(s)$ and the range of s define a particular relativistic model. For example, for $\sigma(s) = \delta_{s,1}$, $s \ge 1$ corresponds to the chiral Gross-Neveu model, whereas $\sigma(s) = \delta_{s,0}$, $-\infty \le s \le \infty$ corresponds to the principal chiral field (PCF) with the SU(N) symmetry.

With all these settings the final TBA equation (or similar equations for the ground state of the finite length system with the periodic boundary conditions) takes a familiar form (say, for the PCF):

$$A^{aa'}(\theta) * C_{ss'}(\theta) * \ln(1 + Y_{s'}^a(u)) - \ln\left(1 + \frac{1}{Y_{s'}^a(u)}\right) = \delta_{s,0} m_0 \sin(a\pi/N)\cosh(\theta\pi/N) \tag{16}$$

where $\epsilon(x, y, \tau) = \log Y_s^a(u)$ plays the role of the energy density of the excitations characterized by the rank a and level s.

At the end of this section let us comment on the large N limit of the TBA equations. It is not obvious to us how to simplify the eq. (16) in this limit, but the difference equation equation (7) after introducing the rescaled continuous variables

$$\tau = u/N, \quad \eta = s/N, \quad \nu = a/N \tag{17}$$

becomes a second order differential equation for the quantity $\epsilon(\eta, \nu, \tau) = \log Y_s^a(u)$ †:

$$(\partial_\nu^2 - \partial_\tau^2)\epsilon = (\partial_\nu^2 - \partial_\eta^2)\ln(1 + \exp\epsilon) \tag{18}$$

This is an integrable classical equation, as it is a consequence of the general HE and the determinant representation (2) for its solution, although the determinant becomes functional in the large N limit.

† I thank P. Zinn-Justin for this comment

Let us give also another form of this equation in terms of new variables l, m and a (and the corresponding rescaled variables) λ , μ and ν defined through the original variables u, s and a as follows:

$$\lambda = \frac{l}{N} = \frac{u+s+a}{2N}, \quad \mu = \frac{m}{N} = \frac{u-s+a}{2N}, \quad \nu = \frac{a}{N} \tag{19}$$

Note that λ and μ play the role of "light-cone" variables.

In their terms we can represent the function $Y = \ln \epsilon$ in the large N limit as

$$\epsilon = \ln \frac{1}{\exp F''_{\lambda\mu} - 1} \tag{20}$$

where

$$F = lim_{N\to\infty} \frac{1}{N^2} \log T. \tag{21}$$

The continuous HE (18) for this function looks in the new variables as

$$\partial_\lambda \partial_\mu \ln(\exp[-F''_{\lambda\mu}] - 1) = \partial_\nu(\partial_\nu - \partial_\lambda - \partial_\mu)F''_{\lambda\mu}] \tag{22}$$

This form of HE will be useful for our matrix model representation of its solution.

MATRIX MODEL CHAIN AS SOLUTION OF THE BILINEAR DIFFERENCE HIROTA EQUATION

In this section we propose to parametrize the general solution of HE (4) by means of so called matrix chain integral - a matrix model widely used and investigated in the literature (see [25] for the details).

Let us define the following Green's function:

$$K_a(T, \phi(0), \phi(T)) = \int \prod_{k=1}^{T-1} d^{a^2}\phi(k) \exp tr[\sum_{k=0}^{T-1} \phi(k)\phi(k+1) - \sum_{k=0}^{T} V(k, \phi_k)] \tag{23}$$

where $\phi(k)$ - are $a \times a$ hermitian matrices, each corresponding to its site k of the chain, with matrix elements $\phi_{ij}(k)$, $i, j = 1, 2, \cdots, N$, $k = 0, 1, 2, \cdots, T$. Two matrices at the ends of the chain are fixed.

The continuous analogue of the chain would be just the matrix quantum mechanical Green's function on the interval of time $(0, T)$:

$$K_a(T, \phi(0), \Phi(T)) = \int D^{a^2}\phi(t) \exp -tr \int_0^T dt[\frac{1}{2}\dot{\phi}^2 + V(t, \phi)] \tag{24}$$

with the end point values of the matrix fields also fixed. The matrix quantum mechanics was introduced and solved in (36).

Let us now define a new quantity:

$$Z_a(l, m) = \int d^{a^2}\phi(0) \int d^{a^2}\phi(T)(\det \phi(0))^l K_a(T, \phi(0), \phi(T))(\det \phi(T))^m \tag{25}$$

where K could be any of both Green's functions (23) or (24). We have added two logarithmic potentials at the ends of the chain to describe the dependence on l and m introduced by (19).

We claim that the function of three discrete variables

$$T^a_s(u) = Z_a(\frac{u+s+a}{2}, \frac{u-s+a}{2}) \tag{26}$$

obeys the HE (4). More than that, it gives the most general solution of HE parameterized by the function of 2 variables - the potential V(t,x).

The proof goes as follows. If we start, say, from the discrete version we can diagonalize each of the matrices in the chain by the unitary rotation:

$$(\phi_a)_{ij} = \sum_k (\Omega_a^+)_{ik}(z_a)_k(\Omega_a)_{kj} \tag{27}$$

and integrate over the relative "angles" between two consecutive matrices in the chain by means of the Itzykson-Zuber-Harish-Chandra formula. This concerns only the first term in the exponent in the r.h.s of (23). The potentials, including two determinants at the ends, depend only on the eigenvalues. The overall result after the angular integration will be (see (25) for the details of this calculation):

$$Z_a(l,m) = \det_{1\le i,j\le a} \int dp \int dq K(T,p,q) p^{l+i-1} q^{m+j-1} \tag{28}$$

where

$$K(T,p,q) = \int \prod_{k=1}^{T-1} dz(k) \exp\{\sum_{k=0}^{T-1} z(k)z(k+1) - \sum_{k=0}^{T} V(k,z(k))\} \tag{29}$$

for the discrete chain, or

$$K(T,p,q) = \int Dz(t) \exp -\{\int_0^T dt[\frac{\dot{z}^2}{2} + V(t,z)]\} \tag{30}$$

for the continuous quantum mechanics, where $z(0) = p$, $z(T) = q$.

Now we see that due to the determinant representation (28) the function $T_s^a(u)$ defined by the eq. (26) has the same determinant form as the BR formula (2) and hence it obeys the HE (4), if we identify

$$T_s^1(u) = \int dp \int dq K(T,p,q) p^{\frac{u+s}{2}} q^{\frac{u-s}{2}} \tag{31}$$

The formula (28) generally defines an arbitrary function of two variables *s* and *u*. It is clear from the fact that it is just the Mellin transform of an arbitrary function (30) (or (29)) in two variables, which is in our case the Greens function of a quantum mechanical particle in an arbitrary time and space dependent potential. This potential obviously gives enough of freedom to define $K(T, p, q)$ as an arbitrary function of two variables *p* and *q*. So we proved (or at least made rather obvious) our statement about the generality of this representation of solution of the HE.

We can provide a more general matrix integral giving the parameterization of the most general BR formula (3):

$$T_{h/h'} = \int d^{a^2}\phi_0 \int d^{a^2}\phi_T \chi_{[h]}(\phi_T) K_a(T,\phi_0,\phi_T) \chi_{[h']}(\phi_T) \tag{32}$$

where $\chi_{[h]}(\phi)$ is the GL(a) character of the representation characterized by the highest weight $[h]$. It can be easily proved when written in terms of eigenvalues with the use of the Weyl formula for characters.

For any finite N all this seems to be on the edge of triviality: we just defined in a sophisticated way an arbitrary function of two variables and built from it the necessary determinant. Naturally, we don't expect this representation to be of big use for a finite N. The boundary conditions of TBA will be as difficult to satisfy as before. What is our major hope is the large N limit of this matrix model which should correspond to the

large N limit of the integrable models of the type of principal chiral field. In this case the determinant in the BR formula is essentially functional, and the matrix models give a rich variety of methods for the calculation of such determinants. That why our strategy will be the following: we investigate the matrix integrals of the type (25) in the large N limit for various potentials and look for physically interesting regimes. Things might become much more universal in the large N limit, and it could exist a classification of interesting regimes, like it was done for the multi-critical points in the matrix models. This paper represents of course only a few modest steps in this direction.

Let us make an important remark concerning the large N limit of the representation (28): the Y variable introduced in the previous section and presented by the formula (20) in the large N limit (with the matrix chain partition function Z instead of the transfer-matrix T) obeys the differential equation (18) or, in new variables (19), (22). It is almost clear from our definitions and it will be demonstrated in the following sections,

Another more formal but interesting application of this method could be the search for new solutions in the integrable equations of the type (18). In the next section we will show some particular examples of it.

EXAMPLES OF SOLUTIONS OF THE CONTINUOUS HIROTA EQUATION

We will demonstrate here on some examples limited to particular choices of the matrix chain potentials how this relation between the HE and MM works.

One matrix model and GL(n) character

Let us start from the one matrix model partition function with an extra logarithmic potential

$$Z_a(l) = \int d^{a^2}\phi (\det \phi)^l \exp -NtrV(\phi) \tag{33}$$

which, after going to the eigenvalue representation, becomes

$$Z_a(l) = \det_{1\le i,j\le a} \int dp\, p^{l+i+j-2} \exp -NV(x) \tag{34}$$

It is well know that if one chooses:

$$\exp -V(x) = \prod_{k=1}^{a} (b_k - x)^{-1} \tag{35}$$

and performs the integral in (34) along the contour encircling all these poles one will identify this partition function with the GL(a) character $\chi_l^a(b)$ of the $a \times l$ rectangular Young tableau given by the Weyl determinant formula. So, much of our next formulas is valid for the characters as well.

It is easy to see from the Jacobi identity for determinants that the function

$$t^a(l) = Z_a(a - l) \tag{36}$$

satisfies a simplified version of the general HE (4):

$$t^a(l+1)t^a(l-1) - t^{a+1}(l)t^{a-1}(l) = [t^a(l)]^2 \tag{37}$$

Introducing the variable

$$Y^a(l) = \frac{t^{a+1}(l)t^{a-1}(l)}{[t^a(l)]^2} \tag{38}$$

we obtain from (37) a simplified version of the general Y-system (7):

$$\frac{Y^a(l+1)Y^a(l-1)}{[Y_a(l)]^2} = \frac{[1+Y^{a+1}(l)][1+Y^{a-1}(l)]}{[1+Y^a(l)]^2} \tag{39}$$

If we go to the large N limit and introduce the continuous variables (19) we obtain from the previous equation a simplified version of the differential HE (18):

$$\partial_\lambda^2 \log Y = \partial_\nu^2 \log(1+Y) \tag{40}$$

which is again integrable, as it is obvious from the above determinant formulas. In the last equations we changed a bit the definition of the function Y rescaling the variables by $1/N$.

Take first a simple example of the potential $V(x) = |x|$. The direct calculation of (34) gives:

$$t_a(l) = \prod_{k=1}^{a} (l-a+k)!(k-1)! \tag{41}$$

which yields $Y_a(l)$ as

$$Y_a(l) = \frac{a}{l-a+1} \tag{42}$$

or, in the large N limit:

$$Y_\nu(\lambda) = \frac{\nu}{\lambda-\nu} \tag{43}$$

which perfectly satisfies the eq. (40).

To find a general (up to some comments which will follow) solution of the eq. (40) we just have to apply the well known formulas for the saddle point approximation in the one matrix model with the potential which is now $V(x) + (\lambda - \nu)\log x$. Omitting the standard calculations (see for example [9]) we give the result: the function $Y_\nu(\lambda)$ obeys the following system of equations on Y and an intermediate variable S:

$$\int_{-1}^{1} \frac{du}{\pi} \frac{V'(\sqrt{Y}u+S)}{\sqrt{1-u^2}} = \frac{\lambda-\nu}{\sqrt{S^2-4Y}} \tag{44}$$

$$\int_{-1}^{1} \frac{du}{\pi} \frac{(\sqrt{Y}u+S)V'(\sqrt{Y}u+S)}{\sqrt{1-u^2}} = \lambda+\nu \tag{45}$$

So we have obtained the solution of the eq. (40) in terms of a system of ordinary equations. For example, for a polynomial potential the equations will become algebraic. In general they are functional.

Not every potential is compatible with this solution. We restricted ourselves to the so called one cut solution implying the existence of one classically stable well in the potential. This restricts our solution to some parametrically general but still limited class of solutions of (40). The generalization to the multi-cut solution which is straightforward should in principle yield the most general solution of (40).

The equations (44-45) look like the characteristics method of solution of the eq. (40). In the next sections we shall see to what extent we can generalize it to the full differential HE (4).

Gaussian Chain Solution of HE

Now we shell consider another particular example of solution of the HE (4) by restricting all the potentials in the matrix chain to be Gaussian. Then all the integrals over $\phi(1), ..., \phi(T)$ in (23) can be easily performed and we are left in (25) with the following two matrix integral over the endpoint variables:

$$Z_a(l,m) = \int d^{a^2}\phi \int d^{a^2}\phi' \, (\det \phi)^l \exp -tr(\frac{1}{2}\phi^2 + \frac{1}{2}\phi'^2 - c\phi\phi')(\det \phi')^m \qquad (46)$$

If we do the same with the continuous quantum mechanical integral (24) we arrive (up to a trivial coefficient typical for the Green's function of the harmonic oscillator) at the same two matrix integral, with $c = 1/\cosh(\omega T)$, where ω is the frequency of the corresponding oscillator.

An important comment is in order. Although this partition function satisfies the HE for three variables it should a little bit modified for finite N: note that $Z_a(l,m)$ in (46) after passing to eigenvalues splits into the product of two determinants corresponding to i, j both even or both odd (the matrix elements with different parities of i, j are zero, see the formula (28) with the gaussian kernel). Each of these determinants satisfies the same difference HE with the shifts of discrete variables by ± 2 and not by ± 1. The continuous (large N) version of HE will be the same as before.

This two-matrix model has the only complication with respect to the ordinary one, containing usually only polynomial potentials: its potentials contain logarithmic parts, like in the well know one matrix Penner model. To solve it the ordinary method of orthogonal polynomials does not look convenient. We propose here another, rather powerful method worked out in a series of papers [26, 27, 28, 29, 30, 31] and capable to solve some even more sophisticated models than the present one.

First we perform the integral over the relative "angles" of two matrices in (46) by means of the character expansion [32]:

$$\int (d\Omega)_{U(N)} \exp[c \; tr(\phi\Omega^+\phi'\Omega)] = \sum_{0\le h_a<\cdots<h_1<\infty} \frac{c^{\sum_k h_k - a(a-1)/2}}{\prod_k h_k!}\chi_h(\phi)\chi_h(\phi') \qquad (47)$$

We dropped here some unessential overall coefficient.

Plugging this formula into the eq. (46) we encounter two identical independent Gaussian integrals over ϕ and ϕ' with the characters as pre-exponentials. These integrals can be calculated (they slightly generalize the similar integrals appearing in [33, 27, 28, 29] to the case of $l, m \neq 0$). The result (again up to some unessential factor) is:

$$\int d^{a^2}\phi e^{-\frac{1}{2}Tr\phi^2}\chi_h(\phi)(\det\phi)^l = \frac{\prod_i (h_i^e + l - 1)!! \prod_j (h_j^o + l)!!}{h_j^o!}\Delta(h^e)\Delta(h^o) \qquad (48)$$

where we denote by $h^e (h^o)$ the even(odd) highest weights whose numbers should be equal. $\Delta(h)$ is the Van-der-Monde determinant of h's. We chose here l to be even; for l odd one only has to exchange h^e and h^o in (48).

Putting all this together we obtain for (46) a representation in terms of the multiple sum over h's (we dropped a h-independent coefficient):

$$Z_a(l,m) = \sum_{\{h^e,h^o\}} \Delta^2(h^e)\Delta^2(h^o) c^{\sum_k (h_k^e + h_k^o)} \times \qquad (49)$$
$$\times \frac{\prod_i (h_i^e + l - 1)!! \prod_j (h_j^o + l)!! \prod_i (h_i^e + m - 1)!! \prod_j (h_j^o + m)!!}{\prod_n h_n!}$$

We chose here l, m to be both even. Different parity for them is forbidden by the symmetry $\phi \to -\phi$ of (46). This will not be important in the large N limit. As we see, the sums over h^e and h^o are decoupled and can be calculated independently.

The method of calculation of these multiple sums in the large N limit was proposed in [26] and further elaborated in [43, 27, 28, 29] and is based on the saddle point approximation of this sum. One introduces the resolvent function of shifted highest weights:

$$H(h) = \sum_{k=1}^{a} \frac{1}{h - h_k} \tag{50}$$

In what follows we change h by h/N (since the highest weights are supposed to be of the order N in the large N limit). So, in the large N limit:

$$H(h) = \int_0^d dh' \frac{\tilde{\rho}(h')}{h - h'} = H_+(h) \pm i\pi\tilde{\rho}(h) \tag{51}$$

where $H_+(h)$ is the symmetric part of the function $H(h)$ on the cut defined by the distribution of h's and $\tilde{\rho}(h)$ is the density of h's along this cut. In the large N limit we can calculate the multiple sum by the saddle point method. The saddle point condition defines the most probable Young tableau shaped by the density $\tilde{\rho}(h)$:

$$\fint_0^d dh' \frac{\tilde{\rho}(h')}{h - h'} = -\frac{1}{2} \ln \left(\frac{c^2(h+\lambda)(h+\mu)}{h^2} \right) \tag{52}$$

One has to remember that a part of the most probable Young tableau is in general empty (some of the highest weight components m_k are equal to zero, see [26, 27] for the details). So, the function $\tilde{\rho}(h)$ is equal to one on the interval $(0,b)$ and to some nontrivial function $\rho(h)$ on the interval (b, d). This yields, instead of (52), the equation:

$$\fint_b^d dh' \frac{\rho(h')}{h - h'} = -\frac{1}{2} \ln \left(\frac{c^2(h+\lambda)(h+\mu)}{(h-b)^2} \right) \tag{53}$$

This linear integral equation has a one-cut solution:

$$\begin{aligned} H(h) = \ln \left[c^{-1}(d+b)^2 h \sqrt{(h+\lambda)(h+\mu)} \right] \\ - \ln \left[(h+\lambda)^2 - (\sqrt{(d+\lambda)(b+\lambda)} - \sqrt{(h-d)(h-b)})^2 \right] \\ - \ln \left[(h+\mu)^2 - (\sqrt{(d+\mu)(b+\mu)} - \sqrt{(h-d)(h-b)})^2 \right] \end{aligned} \tag{54}$$

To fix d and b we should recall the asymptotic of $H(h)$ with respect to large h:

$$H(h) = \nu/h + < h > /h^2 + 0(1/h^3) \tag{55}$$

following from (50). Here $< h >$ is the average shifted highest weight in the most probable Young tableau. Expanding $H(h)$ in $1/h$ up to the terms $O(1/h^2)$ we obtain a system of equations defining d and b:

$$\begin{aligned} \left[2\lambda + d + b + 2\sqrt{(d+\lambda)(b+\lambda)} \right] \left[2\mu + d + b + 2\sqrt{(d+\mu)(b+\mu)} \right] \\ = c^{-1}(d+b)^2 \end{aligned} \tag{56}$$

$$\frac{(d+b)\sqrt{(d+\lambda)(b+\lambda)}+(d+b)\lambda+2db}{2\sqrt{(d+\lambda)(b+\lambda)}+d+b+2\lambda} + \frac{(d+b)\sqrt{(d+\mu)(b+\mu)}+(d+b)\mu+2db}{2\sqrt{(d+\mu)(b+\mu)}+d+b+2\mu} + \frac{\lambda+\mu}{2} = \nu \tag{57}$$

¿From (50) we deduce the following formula for the solution of the continuous HE (22):

$$F''_{\lambda\mu} = \frac{1}{4} << \ln(h+\lambda), \ln(h+\mu) >> \tag{58}$$

where by $<< A, B >>$ we denoted the connected average of any two h-dependent functions A and B. Note that this average has a finite large N limit, as it should be.

This solution can be brought into a more explicit form: the explicite formula for such correlators in the one matrix models was given in [34, 35] for $W(z,z') = << \frac{1}{z-h}, \frac{1}{(z'-h)} >>$:

$$W(z,z') = \frac{1}{2}\left[\sqrt{\frac{(z-d)(z'-b)}{(z'-d)(z-b)}} + \sqrt{\frac{(z'-d)(z-b)}{(z-d)(z'-b)}} - \frac{1}{(z-z')^2}\right] \tag{59}$$

Integrating it in z and z' and putting $z = -\lambda$, $z' = -\mu$ we obtain a rather explicit solution of differential HE (22)

$$F''_{\lambda\mu} = \frac{1}{8}[g(\lambda,b,d)g(\mu,d,b) + g(\lambda,d,b)g(\mu,b,d) + \ln(\lambda+\mu)] \tag{60}$$

where

$$g(z,b,d) = \sqrt{(z+d)(z+b)} + \frac{d-b}{2}\sqrt{\frac{z+d}{z+b}}\ln[2\sqrt{(z+d)(z+b)} - 2z - d - b] \tag{61}$$

and d,b are defined by the eqs. (58).

In the next section we will give the solution of (18) in the case of a general time-independent potential $V(x)$. We will also reduce the search for the most general solution of continuous Hirota equation (18) defined by the time dependent potential $V(x,t)$ to a simpler problem.

COLLECTIVE FIELD METHOD FOR THE DIFFERENTIAL HE

Now let us briefly describe how our method works in the case of a the general potential in the matrix quantum mechanics defined by (24). In this case we can apply the collective coordinate method of Jevicki and Sakita [44] which is valid in the large N limit and can be applied to the non-stationary saddle point solutions which are needed in our case. The details of this approach can be found in [44, 42, 45, 37]. We will use the results of it and apply them to our case.

In terms of this method the effective action for (24) can be written for the density $\rho(x,t)$ of eigenvalues x_k and its conjugate momentum $P(x,t)$ developing in time as:

$$S_{eff}[\rho,P] = \int dx \int_0^T dt[\dot{\rho}P - \frac{1}{2}\rho P'^2 + \frac{\pi^2}{6}\rho^3 + \rho V(x,t)] \tag{62}$$

reflecting our specific boundary conditions at the ends of the interval $(0, T)$ following from (25).

The equations of motion corresponding to the action (62) are:

$$\dot{\rho} + \partial_x(P'\rho) = 0 \tag{63}$$

$$\dot{P} + \frac{1}{2}P'^2 = \frac{\pi^2}{2}\rho^2 + V(x,t) \tag{64}$$

Differentiating the second one in x and defining the function $f(x, t) = P' + i\pi\rho$ we rewrite this system as only one forced Hopf equation on this complex function:

$$\partial_t f + f\partial_x f = -V'(x,t) \tag{65}$$

We have to impose at any moment, say, at $t = 0$ the normalization condition

$$\int dx\rho(x,0) = \nu \tag{66}$$

Then it will be true at any t due to the condition (63).

We have excluded the end points of the interval $(0, T)$ in the last equation. The logarithmic potentials and the two Van-der-Monde determinants left at the ends of the interval can be taken into account as the boundary conditions:

$$Ref(x,0) = \frac{\lambda}{x} + \fint_{b(0)}^{d(0)} dx' \frac{\rho(x',0)}{x-x'}, \quad b(0) \le x \le d(0) \tag{67}$$

$$Ref(x,T) = -\frac{\mu}{x} - \fint_{b(T)}^{d(T)} dx' \frac{\rho(x',T)}{x-x'}, \quad b(T) \le x \le d(T) \tag{68}$$

or, introducing the resolvent: $R_\pm(x,t) = \int_{b(t)}^{d(t)} dx' \frac{\rho(x',t)}{x-x'} \pm i\pi\rho(x,t)$,

$$f(x,0) = \frac{\lambda}{x} + R_+(x,0) \tag{69}$$

$$f(x,T) = -\frac{\mu}{x} - R_-(x,T) \tag{70}$$

Hence we reduced the problem of the virtually general (since it is parameterized by a general potential depending on 2 variables) solution of the continuous HE (4) on the function of 3 variables v, λ et μ to the solution of the differential equation of the first order (65) (assuming that the complex function f is analytical in their variables) supplemented by the boundary conditions (67-68). The variables v, λ et μ appear here only as fixed parameters.

It is still quite complicated, although simpler than the original problem and, may be, physically more transparent. We don't know how we could simplify it further in general case. So let us consider an interesting particular example of the time independent potential $V(x)$. In that case the forced Hopf equation

$$\partial_t f + f\partial_x f = -V'(x) \tag{71}$$

becomes completely integrable by the characteristics method [‡].

We will choose $\lambda = \bar{\mu}$, which does not look as restriction if we assume the analyticity in λ and μ Then due to the time reversal symmetry we have $f(x, T) = -\bar{f}(x, 0)$. So

[‡] I am grateful to A. Matytsin for the explanation of this method and its application to the forced Hopf equation

we can say that also $\rho(x, T) = \rho(x, 0)$. Hence only one of these two boundary conditions is independent.

The first part of the problem is to find the solution of eq. (71) with fixed endpoint density $\rho(x, T) = \rho(x, 0)$. The result can be formulated as the following equation on the functions already at the endpoints

$$T = \int_x^{G(x)} \frac{dy}{\sqrt{2(g_0(x) - V(y))}} \tag{72}$$

where $G(x) = x_T(g_T(x))$ and $x_T(g)$ is the function to be found by solution of

$$g(x_T) = \frac{1}{2} f^2(x_T, T) + V(x_T) \tag{73}$$

Note that the function $\bar{G}(x) = x_0(g_0(x))$ defined by the solution of

$$g = \frac{1}{2} f^2(x_0, 0) + V(x_0) \tag{74}$$

is the functional inverse of the function $G(x)$ itself which gives the equation of A. Matytsin [37]:

$$G(\bar{G}(x)) = x \tag{75}$$

Although the dynamics of the forced Hopf equation is summarized by the relation (72) the last equation leads to a strong constraint on the analytical structure of the function $G(x)$.

Once we found the functional $f_{\rho(x, 0)}(x, 0)$ as the solution of eqs. (72-75) we have to much it with our boundary conditions:

$$Re f_{\rho(x,0)}(x, 0) = \frac{\lambda}{x} + \int_{b(0)}^{d(0)} dx' \frac{\rho(x', 0)}{x - x'}, \quad b(0) \le x \le d(0) \tag{76}$$

With a given $V(x)$ this defines the end-point density $\rho(x, 0)$ which is the only non-trivial information we need to find the quantity (25). Probably it is convenient to represent this boundary condition as a condition on the large x asymptotic of $f(x, 0)$:

$$f(x, 0) \to_{x \to \infty} \frac{\nu + \lambda}{x} \tag{77}$$

We thus reduced the solution of the continuous HE (22) to some simpler functional problem in a particular but rather representative case of the time independent potential. The solution with the harmonic oscillator potential obtained in the previous section should be also reproducible by this method.

We can also use these equation to produce more explicit solutions of the HE (22) by the method proposed in [37]: one chooses two conjugated roots $G(x)$ and $\bar{G}(x)$ of an algebraic equation $x(G) = 0$, where $x(G)$ is some polynomial. These two roots satisfy by construction the eq. (75). Then one has to plug them into (72) and solve it as an integral equation for $V(x)$. Of course the choice should be limited by the boundary conditions (76) or (77).

It would be interesting to analyze the case of the inverted oscillator potential corresponding to the 1+1 dimensional non-critical string theory. But this question lies beyond the scope of this paper.

CONCLUSIONS AND PROSPECTS

In this paper we proposed a matrix model representation for the solution of the general difference Hirota equation. For its continuous analog of the differential HE of second order on three variables the solution is represented by the large N limit of the corresponding matrix chain. It gives an effective framework for solving the continuous HE in a rather explicit way, at least for some particular cases. In particular, the problem can be reduced to the forced Hopf equation with specific boundary conditions.

Many things remain to be understood. First of all it is not clear how to find the solutions satisfying the boundary conditions of various 1+1 dimensional quantum field theories solvable by Bethe ansatz. Especially how to choose the matrix potentials to get the relativistic spectrum for the physical particles and to fix the symmetries of original models. Another question: is there some physical interpretation in terms of these integrable theories of the time variable t and of the eigenvalue variable x in our matrix representation similar to the non-critical (1+1) dimensional string theory (where these variables describe the target space of the string [42, 41])? It might be for example that the time can be considered as the physical space dimension of the corresponding integrable theory. But it remains to be proved. We hope that the formalism proposed here at least sets a convenient framework for the attempts to find the physically interesting solutions.

A more formal use of our method might be the search for new solitonic solutions for the well known integrable equations. For example, the Toda equation $\partial_\nu^2 F = \partial_\lambda \partial_\mu \exp F$ is just a particular limit of the continuous HE (18). To our knowledge, the solitonic solutions to this equation are not yet found.

Another interesting question is how the double scaling limit in matrix models is related to HE? For example, how to find the corresponding solution for the inverted harmonic oscillator giving the description of the (1+1) dimensional string field theory. To answer this question as well as many others we have to learn how to deal with the non-stationary forced Hopf equation with our specific boundary conditions. The methods worked out in the papers [37, 38, 39, 40] could be useful for that.

Acknowledgments

I would like to thank P. Wiegmann, I. Kostov, A. Matytsin and especially P. Zinn-Justin for many valuable discussions. I am also grateful to the Theory Division of CERN where a part of this work was done for the kind hospitality.

REFERENCES

1. I. Krichever, O. Lipan, P. Wiegmann, A. Zabrodin, hep-th/9604080
2. A. Zabrodin, hep-th9610039.
3. R. Hirota, *J. Phys. Soc. Japan 43* (1977) 1424.
4. A. Klumper and P. Pierce, *Physica A183* (1992) 304.
5. A. Kuniba and T. Nakanishi, preprint HUTP-92/A046, hep-th/9210025.
6. Al. Zamolodchikov, *Phys. Lett. B253* (1991) 391.
7. T. Miwa, *Proc. Japan Acad. 58.A* (1982) 9.
8. E. Brézin and V. Kazakov, *Phys.Lett. 236B* (1990) 144.
9. D. Gross and A. Migdal, *Phys.Rev.Lett. 64* (1990) 127.
10. M. Douglas and S. Shenker, *Nucl.Phys. B335* (1990) 635.
11. M. Douglas, *Phys.Lett. B348* (1991) 507.
12. S.Kharchev, A. Marshakov, A. Mironov, A. Orlov and A. Zabrodin, *Nucl. Phys. B366* (1991) 569.
13. H. Itoyama and Y. Matsuo, *Phys.Lett.B255* (1991) 202.

14. S. Kharchev, A. Marshakov, A. Mironov, A. Morozov and S. Pakuliak, *Nucl. Phys. B404* (1993) 717.
15. I. Kostov, Saclay preprint, hep-th/9604011.
16. V. Kazakov and I. Kostov, *Nucl. Phys. B386* (1992) 520.
17. V. Bazhanov and N. Reshetikhin, *J. Phys. A23* (1990) 1477.
18. Y. Zhou, P. Pearce, *Nucl. Phys. B446* (1995) 485.
19. V. Bazhanov, S. Lukyanov and A. Zamolodchikov, *Nucl. Phys. B489* (1977) 487.
20. Al. Zamolodchikov, *Nucl. Phys. B358* (1991) 497
21. A. Polyakov and P. Wiegmann, *Phys. Lett. 131B* (1984) 121.
22. P. Wiegmann, *Phys. Lett.141B* (1984) 217.
23. V.A. Fateev, V.A. Kazakov and P.B. Wiegman, *Phys.Rev. Lett 73* (1994) 1750.
24. V.A. Fateev, V.A. Kazakov and P.B. Wiegman, *Nucl. Phys. B424* (1994) 505.
25. V. Kazakov, Lecture given at Cargese Workshop, Cargese, France, May 27 - Jun 1, 1990. In *Cargese 1990, Proceedings, Random surfaces and quantum gravity* 269-306.
26. M.R. Douglas and V.A. Kazakov, *Phys. Lett. 319B* (1993) 219.
27. V. Kazakov, M. Staudacher and T. Wynter, *Comm. Math. Phys. 177* (1996) 451.
28. V. Kazakov, M. Staudacher and T. Wynter, *Comm, Math. Phys. 179* (1996) 235.
29. V. Kazakov, M. Staudacher and T. Wynter, *Nucl.Phys.B471* (1996) 309.
30. I. Kostov, M. Staudacher and T. Wynter, hep-th/9703189.
31. I. Kostov, M. Staudacher, *Phys.Lett. B394* (1997) 75.
32. C. Itzykson and J.-B. Zuber, *J.Math.Phys 21* (1980) 411.
33. P. Di Francesco and C. Itzykson, *Ann. Inst. Henri Poincaré Vol. 59, no. 2* (1993) 117.
34. J. Ambjorn, J. Jurkevich and Yu. Makeenko, *Phys. Lett. B251* (1990) 517.
35. I. Kostov, (2-loop propagator in the asymmetric case, private communication).
36. E. Brézin, C. Itzykson, G. Parisi and J.-B. Zuber, *Math. Phys 59* (1978) 35.
37. A. Matytsin *Nucl. Phys. B411* (1994) 805.
38. A. Matytsin and P. Zaugg, *Nucl. Phys. B497* (1997) 658.
39. A. Matytsin and P. Zaugg, *Nucl. Phys. B497* (1997) 699.
40. A. Dhar, G. Mandal and S. Wadia, *Int.J.Mod.Phys.A8*(1993) 325.
41. J. Polchinski, *Nucl. Phys. B346* (1990) 253.
42. S. Das and A. Jevicki, *Mod. Phys. Lett. A5* (1990) 1639.
43. V. Kazakov and T. Wynter, *Nucl. Phys. B440* (1995) 407.
44. A. Jevicki and B, Sakita, *Nucl. Phys. Bl65* (1980) 511.
45. J. Avan and A. Jevicki, *Mod.Phys.Lett.A7* (1992) 357.

LATTICE APPROXIMATION OF QUANTUM ELECTRODYNAMICS

Jerzy Kijowski,[1] Gerd Rudolph,[2]

[1]Center for Theoretical Physics, Polish Academy of Sciences
al. Lotników 32/46, 02-668 Warsaw, Poland
[2] Institut für Theoretische Physik, Universität Leipzig
Augustusplatz 10/11, 04109 Leipzig, Germany

INTRODUCTION

There is a growing interest in discrete approximations of quantum gauge field theories, using real (and not *Euclidean*) time. This is partially due to the successes of Quantum Gravity in the Ashtekar – Lewandowski – Rovelli – Smolin formulation [1, 2]. In this approach we organize the field Cauchy data on a given 3-D space-like hypersurface into the phase space of a physical system which is going to be quantized. Since we are not able to quantize directly a system with infinitely many degrees of freedom (unless it is linear!), we use finite, 3-dimensional lattices (or *networks* in Ashtekar – Lewandowski terminology) to select a finite number of degrees of freedom and then we are able to provide a quantum description of such a truncated subsystem of our original system. The way we select such subsystems from the original physical system is fully *gauge invariant.*

The fundamental idea in the above approach is to reconstruct both the quantum kinematics and dynamics of the complete system (i.e. to construct the continuum quantum gauge field theory) *via* a suitable limiting procedure from the above quantum theories of subsystems.

Such an approach should be first tested on Quantum Electrodynamics, where we are still far from understanding the fundamental structures of the theory beyond the perturbative picture. In particular, the gauge structure is rather poorly understood.

In this talk we present an attempt to construct Quantum Electrodynamics in a fully gauge invariant, non-perturbative way. It is based on the ideas published earlier [3,4], where we have shown that the functional integrals of QED and QCD can be reformulated in terms of local gauge invariant quantities. Recently[5,6], we were able to implement similar ideas on the level of the 3-D lattice approximation of QED in the Hamiltonian approach, where the real tirne variable (and not the Euclidean time) remains a continuous parameter. The main result of these papers is a complete description of the observable algebra and an explicit construction of the physical Hilbert space as a direct sum of charge superselection sectors. A similar programme for the continuum theory has been formulated by many authors a long time ago. Since the problems arising for

New Developments in Quantum Field Theory
Edited by Damgaard and Jurkiewicz, Plenum Press, New York, 1998

the continuum case are extremely difficult, only partial results have been obtained until now [7 – 11].

The algebra of observables (gauge invariants) in QED is generated by the electromagnetic field operators $\hat{E}$ and $\hat{B}$ together with the invariants which are bilinear in fermionic fields:

$$\hat{\mathcal{W}}_{\gamma}^{ba} := \hat{\psi}^{b*}(x)\, \exp(-ig \int_{\gamma} \hat{A})\, \hat{\psi}^{a}(y)\ , \tag{1}$$

where γ is an oriented path starting at x and ending at y. Dividing the bi-spinor field $\hat{\psi}^{a}$ into the positron degrees of freedom $\hat{\phi}^{K}$ and the electron degrees of freedom $\hat{\varphi}_{L}^{*}$, $K, L = 1, 2$, we may consider, in particular, the "pair annihilation operators":

$$\hat{\mathcal{M}}_{\gamma;L}{}^{K} = \hat{\varphi}_{L}(y) \exp\left(- ig \int_{\gamma} \hat{A}\right) \hat{\phi}^{K}(x)\ , \tag{2}$$

(they annihilate a positron with helicity K at x and an electron with helicity L at y). It turns out[5] that these operators – together with their conjugate "pair creation operators" – are sufficient to reconstruct *all* the operators $\hat{W}$ defined above.

The purpose of this talk is to present the structure of the observable algebra generated by these operators in a lattice version of the theory. It turns out, that the algebra splits in a natural way into a tensor product of the algebra of electromagnetic observables and the algebra generated by the above operators $\hat{\mathcal{M}}$. The first component carries the structure of a finitely generated Heisenberg algebra. The latter is a finite-dimensional C^{*}–algebra isomorphic to the enveloping algebra of $SL(2N, C)$ factorized by a certain ideal. We are able to find all irreducible representations of this algebra and, thus, the decomposition of the physical Hilbert space into charge superselection sectors. In the last part of the talk we discuss a possible strategy how to reconstruct the continuum theory. Our strategy differs slightly from the one proposed by Ashtekar and Lewandowski. The two constructions may lead to physically non-equivalent sectors of the theory.

SECOND QUANTIZATION ON THE LATTICE

A classical field configuration of the continuum Maxwell-Dirac theory consists of a U(1)-gauge potential (A_{μ}) and a four-component spinor field (ψ^{a}), where $a, b, \ldots = 1,2,3,4$ denote bispinor indices and $\mu, \nu, \ldots = 0, 1, 2, 3$ spacetime indices. The classical Lagrangian of the theory is given by

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - m\psi^{a*}\, \beta_{ab}\, \psi^{b} - \hbar\ \mathrm{Im}\ \left\{ \psi^{a*}\, \beta_{ab}\, (\gamma^{\mu})^{b}{}_{c}\, D_{\mu} \psi^{c} \right\}, \tag{3}$$

where $F_{\mu\nu} = \partial_{\mu} A_{\nu} - \partial_{\nu} A_{\mu}$ and $D_{\mu}\psi^{a} = \partial_{\mu}\psi^{a} + igA_{\mu}\psi^{a}$. The star denotes complex conjugation, β_{ab} denotes the canonical Hermitean structure in bispinor space and (γ^{μ}) are the Dirac matrices. For a given Cauchy hyperplane $\Sigma = \{t = \text{const}\}$ in Minkowski space, the above Lagrangian gives rise to an infinite-dimensional Hamiltonian system in variables $(A_{k}, E^{k}, \psi^{a}, \psi^{a*})$ with the Hamiltonian given by

$$\mathcal{H} = \frac{1}{2} \left(E_{k}\, E^{k} + B_{k}\, B^{k} \right) + m\psi^{a*}\, \beta_{ab}\, \psi^{b} + \hbar \mathrm{Im}\ \left\{ \psi^{a*}\, \beta_{ab}\, \left(\gamma^{k}\right)^{b}{}_{c}\, D_{k} \psi^{c} \right\}, \tag{4}$$

where $B = \mathrm{curl}\ A$.

Let us take a finite regular cubic lattice Λ contained in Σ, with lattice spacing a, and let us denote the set of n-dimensional lattice elements by Λ^{n}, $n = 0, 1, 2, 3$. Such

elements are (in increasing order of n) called sites, links, plaquettes and cubes. We approximate every continuous configuration $(A_k, E^k, \psi^a, \psi^{a*})$ in the following way:

$$\Lambda^0 \ni x \longrightarrow \psi^a_x := a^{\frac{3}{2}}\ \psi^a(x) \in \mathbb{C}\,, \tag{5}$$

$$\Lambda^1 \ni (x, x+\hat{k}) \longrightarrow A_{x,x+\hat{k}} := \int_{(x,x+\hat{k})} A_k\, dl \in \mathbb{R}\,, \tag{6}$$

$$\Lambda^1 \ni (x, x+\hat{k}) \longrightarrow E_{x,x+\hat{k}} := \int_{\sigma(x,x+\hat{k})} E^k\, d\sigma_k \in \mathbb{R}\,. \tag{7}$$

Here $\sigma(x,\ x+\hat{k})$ denotes a plaquette of the dual lattice, dual to the link $(x,\ x+\hat{k}) \in \Lambda^1$. Note that we have chosen the non-compact lattice approximation, where the potential and the field strength remain Lie-algebra-valued on the lattice level.

We define the second quantization for the lattice theory by postulating the following canonical (anti-) commutation relations for the lattice quantum field operators:

$$\left[\hat{\psi}^a_x, \hat{\psi}^{b*}_y\right]_+ = \delta^{ab}\delta_{xy}\ , \tag{8}$$

$$\left[\hat{A}_{x,x+\hat{k}}, \hat{E}_{y,y+\hat{l}}\right] = i\hbar\delta_{xy}\delta_{\hat{k}\hat{l}}\ . \tag{9}$$

The remaining (anti-) commutators have to vanish.

All irreducible representations in the strong (Weyl) sense of the above algebra are equivalent [12, 13]. In particular, the bosonic quantities (A, E) may be described by the Schrödinger representation, in the Hilbert space of wave functions Ψ depending on parameters A. Operators $\hat{A}$ are thus multiplication operators and canonically conjugate momenta are represented by derivatives

$$\hat{E}_{x,x+\hat{k}} := \frac{\hbar}{i}\frac{\partial}{\partial A_{x,x+\hat{k}}}\ . \tag{10}$$

For the fermion fields we use the following decomposition into Weyl spinors:

$$\psi^a = \begin{pmatrix} \phi^K \\ \varphi^*_L \end{pmatrix}\ ,\quad K, L = 1, 2\,. \tag{11}$$

We take the anti-holomorphic representation for the upper part and the holomorphic representation for the lower part of ψ. Thus, we represent the "classical" Grassmann algebra valued quantities (ϕ^{K*}, φ^*_L) as multiplication operators $(\hat{\phi}^{K*}, \hat{\varphi}^*_L)$ in the space of all functions (polynomials) of these variables. It follows that the adjoint operators $(\hat{\phi}^K, \hat{\varphi}_L)$ satisfy relations (8). They may be represented as derivatives [14]:

$$\hat{\phi}^K := \frac{\partial}{\partial \phi^{K*}}\,,\quad \hat{\varphi}_L := \frac{\partial}{\partial \varphi^*_L}\,. \tag{12}$$

The tensor product of all these representations is, therefore, defined in the space of wave functions

$$\Psi = \Psi(\{A_{x,x+\hat{k}}\}, \{\phi^{K*}_x\}, \{\varphi^*_{x;L}\})\ , \tag{13}$$

which are polynomials in the anticommuting variables (ϕ^*, φ^*) with coefficients being functions of variables A. The Hilbert space structure is defined by the L^2-norm. Integration over the Grassmann variables is understood in the sense of Berezin, which means that the set of all different monomials in these variables forms an orthonormal basis.

Obviously, the algebra generated by (8) and (9) contains a lot of unphysical (gauge-dependent) elements. Moreover, the above electric field operators do not satisfy the Gauss law. In what follows we will present an explicit construction of the algebra of observables (gauge invariant operators satisfying the Gauss law), together with a complete classification of its irreducible representations.

GAUGE INVARIANCE, CONSTRAINTS AND BOUNDARY DATA

A local gauge transformation of a lattice configuration is given by:

$$\tilde{\psi}_x = \exp(-ig\lambda_x)\,\psi_x\,, \tag{14}$$

$$\tilde{A}_{x,x+\hat{k}} = A_{x,x+\hat{k}} + \lambda_{x+\hat{k}} - \lambda_x\,, \tag{15}$$

where $\Lambda^0 \ni x \longrightarrow \lambda_x \in \mathbb{R}$.

At this point we could have also used the more familiar compact description, in which the lattice gauge potentials are group-valued quantities $\exp(igA_{x,x+\hat{k}})$ (parallel transporters). On the level of physical observables, this would lead to replacing the magnetic field *B*, see forthcoming formula (24), by $\exp(igB)$ (the *Wilson loop*). The *compact description* of gauge fields is also used by Ashtekar and Lewandowski in Quantum Gravity. For the purposes of QED we prefer, however, to keep the above non-compact description which leads to the topology of the phase space which better resembles the corresponding structure of the continuum theory. Also the classification of irreducible representations of the observable algebra which we obtain in our approach is much simpler: we avoid the so called θ-representations for the electromagnetic field operators $\exp(igB)$ and *E*, which occur in the compact representation (they have no physical meaning and arise merely as artifacts of the compact formalism used).

Local gauge transformations act on wave functions in the following way:

$$(U(\{\lambda_x\})\Psi)(\{A_{x,x+\hat{k}}\},\{\phi_x^{K*}\},\{\varphi^*_{x;L}\}) = \Psi(\{\tilde{A}_{x,x+\hat{k}}\},\{\tilde{\phi}_x^{K*}\},\{\tilde{\varphi}^*_{x;L}\})\,. \tag{16}$$

This induces the transformation law for the field operators $\hat{A}$ and $\hat{\psi}$, formally identical with (14) and (15). To calculate the generator $\hat{\mathcal{G}}_x$ of infinitesimal local gauge transformations at *x* we take the derivative of $\Psi(\{\tilde{A}_{x,x+\hat{k}}\},\{\tilde{\phi}_x^{K*}\},\{\tilde{\varphi}^*_{x;L}\})$ with respect to λ_x, at $\lambda_x = 0$:

$$\begin{aligned}\hat{\mathcal{G}}_x\Psi &:= -\sum_{\hat{k}} \frac{\partial\Psi}{\partial A_{x,x+\hat{k}}} + ig\sum_K \phi_x^{K*}\frac{\partial\Psi}{\partial\phi_x^{K*}} - ig\sum_L \varphi^*_{x;L}\frac{\partial\Psi}{\partial\varphi^*_{x;L}} = \\ &= \frac{i}{\hbar}\left\{-\sum_{\hat{k}}\hat{E}_{x,x+\hat{k}} + g\hbar\sum_K\left(\hat{\phi}_x^{K*}\hat{\phi}_x^K - \hat{\varphi}^*_{x;K}\hat{\varphi}_{x;K}\right)\right\}\Psi\,.\end{aligned} \tag{17}$$

We conclude that the generator $\hat{\mathcal{G}}_x$ of local gauge transformations is given by

$$\hat{\mathcal{G}}_x := -\frac{i}{\hbar}\left\{\sum_{\hat{k}}\hat{E}_{x,x+\hat{k}} - \hat{j}^0_x\right\}\,, \tag{18}$$

where, the operator $\hat{j}^0_x$ of electric charge at *x* is automatically obtained in the "normally ordered" form:

$$\hat{j}^0_x =: \; e\sum_a \hat{\psi}_x^{a*}\hat{\psi}_x^a \; := e\sum_K\left(\hat{\phi}_x^{K*}\hat{\phi}_x^K - \hat{\varphi}^*_{x;K}\hat{\varphi}_{x;K}\right)\,, \tag{19}$$

and $e := g\hbar$ is the elementary charge.

The necessary and sufficient condition for gauge invariance of the wave function is provided by the following "Gauss law constraint":

$$\hat{\mathcal{G}}_x\,\Psi = 0\,. \tag{20}$$

Unfortunately, wave functions fulfilling (20) *are not* square integrable with respect to the standard measure on the configuration space, because they are constant on non-compact gauge orbits. A possible strategy to circumvent this difficulty consists in looking for an appropriate Hilbert space structure in the space of gauge invariant wave functions. As far as we know, there is no unique construction of such a structure. In our approach, we construct explicitly the algebra of observables and find all its irreducible representations. This way, *no ambiguity* in the definition of the scalar product in the space of physical, gauge-invariant functions arises. It is *uniquely* implied by the structure of the algebra of observables.

If we sum up equations (20) over all $x \in \Lambda^0$ we see that, heuristically, the total charge $\hat{Q}$ should vanish, when acting on gauge invariant wave functions Ψ:

$$\hat{Q}\Psi = \sum_{x \in \Lambda^0} \hat{j}^0_x \, \Psi = 0 \, . \tag{21}$$

Thus, nontrivial values of the total charge $\hat{Q}$ can only arise from nontrivial boundary data, which we are now going to introduce. For this purpose we consider also external links of our finite lattice Λ, connecting lattice sites belonging to the boundary $\partial\Lambda$ with "the rest of the world". This way we can treat Λ as part of a bigger (maybe infinite) lattice. We denote these external links by (x,∞) and allow the wave functions Ψ from the beginning to depend on the corresponding potentials $A_{x,\infty}$. Moreover, we put $\partial\Lambda^n := \partial\Lambda \cap \Lambda^n$. Now gauge invariance does no longer imply vanishing of the total charge, because the electric fields on external links remain when we sum up equations (20) over all sites of Λ:

$$\hat{Q}\Psi = \left(\sum_{x \in \Lambda^0} \hat{j}^0_x \right) \Psi = \left(\sum_{x \in \partial\Lambda^0} \hat{E}_{x,\infty} \right) \Psi \, . \tag{22}$$

We stress that the external fluxes $\hat{E}_{x,\infty}$ *are not* dynamical quantities in this approach, they play the role of *prescribed* boundary conditions. (On the other hand, one could possibly treat them dynamically *à la* Staruszkiewicz – see his contribution in the present volume.) It will be shown in the sequel that the charge operator $\hat{Q}$ defines a superselection rule. Thus, we have $\hat{Q} = Q\mathbb{1}$ on every superselection sector. Consequently, the only consistent choice for the external fluxes is $\hat{E}_{x,\infty} = E_{x,\infty}\,\mathbb{1}$ on every superselection sector, where $E_{x,\infty}$ are c-numbers fulfilling

$$Q = \sum_{x \in \partial\Lambda^0} E_{x,\infty} \, . \tag{23}$$

In principle, we could distinguish between representations characterized by the same value Q, but corresponding to different external flux distributions fulfilling (23). This would lead to additional superselection rules. Here, we have chosen another option, which is motivated by the fact that different boundary conditions corresponding to the same value Q give equivalent representations. Thus, for any value of Q we have an equivalence class of boundary data and we choose a representative, e.g. the "most symmetric" distribution of the values $\hat{E}_{x,\infty}$ on $\partial\Lambda$.

ALGEBRA OF GAUGE INVARIANT OPERATORS

Consider the algebra $\hat{\mathcal{O}}$ of gauge invariant operators acting in the auxiliary Hilbert space H^0 of square integrable wave functions (13), admitting (possibly) nontrivial

boundary data $\hat{E}_{x,\infty}$. This algebra is defined as the commutant of $\{U(\{\lambda_x\})\}$ in $B(H^0)$ We are going to list below the set of relations between its generators. This set, supplemented by the Gauss law, will be taken as a set of axioms for the generators of the algebra of observables, which we are going to define in the next Section.

We see that arbitrary bounded functions of (self-adjoint) electric flux operators $\hat{E}$ are gauge invariant. To each oriented plaquette $(x; \hat{k}, \hat{l})$ we may also assign the (self-adjoint) magnetic flux operator:

$$\hat{B}_{x;\hat{k},\hat{l}} := \hat{A}_{x,x+\hat{k}} + \hat{A}_{x+\hat{k},x+\hat{k}+\hat{l}} + \hat{A}_{x+\hat{k}+\hat{l},x+\hat{l}} + \hat{A}_{x+\hat{l},x} \ . \tag{24}$$

Again, arbitrary bounded operator functions of $\hat{B}$ are gauge invariant – due to (15). Observe that magnetic flux operators are subject to the following constraint: the total magnetic flux through the boundary of a lattice cube vanishes as a consequence of (24),

$$\mathrm{div}_{x;\hat{k},\hat{l},\hat{n}}\,\hat{B} := \hat{B}_{x;\hat{k},\hat{l}} + \hat{B}_{x;\hat{n},\hat{k}} + \hat{B}_{x;\hat{l},\hat{n}} + \hat{B}_{x+\hat{n};\hat{l},\hat{k}} + \hat{B}_{x+\hat{l};\hat{k},\hat{n}} + \hat{B}_{x+\hat{k};\hat{n},\hat{l}} = 0 \ . \tag{25}$$

Moreover, we consider bilinear invariants (1). Obviously, any such $\hat{\mathcal{W}}$ is bounded in H^0 and hence it belongs to $\tilde{\mathcal{O}}$.

Proposition 1 *If γ is a path connecting x with y, β is a path connecting y with z and if the degrees of freedom (x, a), (y, b) and (z, c) differ from each other, then we have the following identity:*

$$\left[\hat{\mathcal{W}}^{cb}_{\beta}\,,\,\hat{\mathcal{W}}^{ba}_{\gamma}\right] = \hat{\mathcal{W}}^{ca}_{\beta\gamma} \ , \tag{26}$$

where $\beta\gamma$ is the composition of γ and β, a path connecting x with z.

Proof:

$$\begin{aligned}\left[\hat{\mathcal{W}}^{cb}_{\beta}\,,\,\hat{\mathcal{W}}^{ba}_{\gamma}\right] &= \left(\hat{\psi}^{c*}_{z}\hat{\psi}^{b}_{y}\hat{\psi}^{b*}_{y}\hat{\psi}^{a}_{x} - \hat{\psi}^{b*}_{y}\hat{\psi}^{a}_{x}\hat{\psi}^{c*}_{z}\hat{\psi}^{b}_{y}\right)\,\exp(-ig\int_{\beta\gamma}\hat{A}) \\ &= \left(\hat{\psi}^{b}_{y}\hat{\psi}^{b*}_{y} + \hat{\psi}^{b*}_{y}\hat{\psi}^{b}_{y}\right)\hat{\mathcal{W}}^{ca}_{\beta\gamma} = \hat{\mathcal{W}}^{ca}_{\beta\gamma} \ .\end{aligned}$$

There are three types of operators $\hat{\mathcal{W}}$:

$$\hat{\mathcal{L}}_{\gamma}{}^{LK} := \hat{\phi}^{L*}_{y}\,\exp(-ig\int_{\gamma}\hat{A})\,\hat{\phi}^{K}_{x} \ , \tag{27}$$

$$\hat{\mathcal{M}}_{\gamma;L}{}^{K} := \hat{\varphi}_{y;L}\,\exp(-ig\int_{\gamma}\hat{A})\,\hat{\phi}^{K}_{x} \ , \tag{28}$$

$$\hat{\mathcal{R}}_{\gamma;LK} := \hat{\varphi}_{y;L}\,\exp(-ig\int_{\gamma}\hat{A})\,\hat{\varphi}^{*}_{x;K} \ , \tag{29}$$

(the fourth type coincides with the adjoint operators $\hat{\mathcal{M}}_{\gamma;L}{}^{K*}$). Observe that $\hat{\mathcal{M}}_{\gamma;L}{}^{K}$ annihilates a positron at x and an electron at y, whereas $\hat{\mathcal{M}}_{\gamma;L}{}^{K*}$ creates such a pair. We call them "pair annihilation operators" (respectively, "pair-creation operators"). A non-diagonal operator $\hat{\mathcal{L}}_{\gamma}{}^{LK}$ (i. e. such that $(x, K) \neq (y, L)$, where x is the beginning and y is the end of γ) annihilates a positron at x and creates another one at y, whereas $\hat{\mathcal{R}}_{\gamma;LK}$ does the same with electrons. Finally, we have diagonal operators corresponding to trivial paths. We denote them by $\hat{L}_{x}{}^{KK} = \hat{\phi}^{K*}_{x}\hat{\phi}^{K}_{x}$ (the projector representing the number of positrons at x with helicity K) and by $\hat{R}_{x;LL}$ (the projector $\mathbb{1} - \hat{R}_{x;LL} = \hat{\varphi}^{*}_{y;L}\hat{\varphi}_{y;L}$ describes the number of electrons at x with helicity L).

The following statement was proved[5]:

Proposition 2 *The operators* $\hat{\mathcal{L}}$ *and* $\hat{\mathcal{R}}$ *can be expressed in terms of the operators* $\hat{\mathcal{M}}$ *and* $\hat{\mathcal{M}}^*$.

It follows from this Proposition that also every local charge operator (19) can be expressed in terms of $\hat{\mathcal{M}}$ and $\hat{\mathcal{M}}^*$. We see, therefore, that the algebra $\tilde{\mathcal{O}}$ is generated by the family $\{\hat{E}, \hat{B}, \hat{\mathcal{M}}, \hat{\mathcal{M}}^*\}$. These generators are, however, not independent. Below, we list a number of operator identities between them, which may be easily verified by inspection.

Proposition 3 *We have*

1.
$$\mathrm{div}_{x;\hat{k},\hat{l},\hat{n}}\,\hat{B} = 0\ . \tag{30}$$

2. *Any pair of operators* $\hat{\mathcal{M}}$ *commutes:*
$$\left[\hat{\mathcal{M}}_{\alpha;L}{}^{K}, \hat{\mathcal{M}}_{\beta;N}{}^{M}\right] = 0\ . \tag{31}$$

3. *Pair-annihilation operators along two different paths* γ *and* β, *having common ends x and y, are related by:*
$$\hat{\mathcal{M}}_{\gamma;L}{}^{K} = \exp(-ig\hat{B}_{(\gamma\beta^{-1})})\hat{\mathcal{M}}_{\beta;L}{}^{K}\ , \tag{32}$$
where $\hat{B}_{(\gamma\beta^{-1})}$ *denotes the magnetic flux through the closed path* $\gamma\beta^{-1}$.

4.
$$\left(\hat{\mathcal{M}}_{\alpha;L}{}^{K}\right)^2 = 0\,. \tag{33}$$

5. *Let* α *connect x with y,* β *connect z with u and* γ *connect t with w. Denote by* $\delta_{(x,K)(z,M)}$ *the Kronecker symbol, which vanishes if* $(x, K) \neq (z, M)$ *and takes the value equal to one if both points and indices coincide. Then:*
$$\begin{aligned}\left[\hat{\mathcal{M}}_{\alpha;L}{}^{K}, \hat{\mathcal{M}}_{\beta;N}{}^{M*}, \hat{\mathcal{M}}_{\gamma;R}{}^{P}\right] &= \delta_{(x,K)(z,M)}\delta_{(u,N)(w,R)}\hat{\mathcal{M}}_{\alpha\beta^{-1}\gamma;L}{}^{P} \\ &+ \delta_{(y,L)(u,N)}\delta_{(t,P)(z,M)}\hat{\mathcal{M}}_{\gamma\beta^{-1}\alpha;R}{}^{K}\ ,\end{aligned} \tag{34}$$
where we have denoted $[\cdot,\cdot,\cdot] := [[\cdot,\cdot],\cdot] \equiv [\cdot,[\cdot,\cdot]]$, *(both definitions are equivalent, because the first and the last variable commute).*

6.
$$\left[\hat{\mathcal{M}}_{\alpha;L}{}^{K}, \hat{B}_{x;\hat{k},\hat{l}}\right] = 0\ , \tag{35}$$
$$\left[\hat{\mathcal{M}}_{\alpha;L}{}^{K}, \hat{E}_{x,x+\hat{k}}\right] = e\hbar\delta_{\alpha,(x,x+\hat{k})}\hat{\mathcal{M}}_{\alpha;L}{}^{K}\ , \tag{36}$$
$$\left[\hat{B}_{x;\hat{k},\hat{l}}, \hat{E}_{y,y+\hat{n}}\right] = i\hbar\delta_{\partial(x;\hat{k},\hat{l}),(y,y+\hat{n})}\ . \tag{37}$$
Here $\delta_{\alpha,(x,x+\hat{k})} = 0$ *if* $(x, x+\hat{k}) \notin \alpha$, $\delta_{\alpha,(x,x+\hat{k})} = 1$ *if* $(x, x+\hat{k}) \in \alpha$ *and has the same orientation as* α *and* $\delta_{\alpha,(x,x+\hat{k})} = -1$ *otherwise. In the last formula we have denoted by* $\partial(x;\hat{k},\hat{l})$ *a closed path – the boundary of the oriented plaquette* $(x;\hat{k},\hat{l})$.

ALGEBRA OF OBSERVABLES AND TREE DECOMPOSITION

The algebra $\tilde{\mathcal{O}}$ is, of course, unphysical, because it does not respect the Gauss law. Its representation in H^0 is highly reducible. To construct the physical observable algebra O we additionally impose the Gauss law

$$\mathrm{div}_x \hat{E} = \hat{j}^0_x \,, \tag{38}$$

where $\hat{j}^0_x$ is expressed in terms of $\hat{\mathcal{M}}$ according to Proposition 2.

Thus, let us start with the $*$-algebra generated by the family of abstract elements $\{\hat{E}, \hat{B}, \hat{\mathcal{M}}, \hat{\mathcal{M}}^*\}$ (with $\hat{E}$ and $\hat{B}$ being self-adjoint), satisfying axioms (30) – (37) together with (38). We will show in this section that these axioms (with canonical commutation relations understood in the sense of Weyl) define, indeed, uniquely a von Neumann algebra, which we shall call *Algebra of Observables* of our model and denote by O.

Our main tool will be the notion of a *tree*. By a tree we mean a pair $(x_0, \mathcal{T})$, where $x_0 \in \Lambda^0$ is a lattice site, which we call the *root* of the tree and $\mathcal{T} \subset \Lambda^1$ is a subset of links, having the following property: for every site $x \in \Lambda^0$ there is one and only one path composed of links belonging to $\mathcal{T}$, which connects the root x_0 with x. We denote this unique path by $(x_0, x)_\mathcal{T}$.

The simplest example of a tree is obtained as follows. Choose any root and take as $\mathcal{T}$ the following collection of links

1. all the links $(x, x + \hat{3})$ belonging to the x^3-axis passing through the root,

2. all the links $(x, x + \hat{2})$, belonging to the two-dimensional plane (x^2, x^3) passing through the root,

3. all the links $(x, x + \hat{1})$.

With every *off-tree* link, $(x, x + \hat{k}) \notin \mathcal{T}$, we associate the unique closed path composed of the tree-path $(x_0, x)_\mathcal{T}$, the link $(x, x + \hat{k})$ itself and the inverse tree-path $(x + \hat{k}, x_0)_\mathcal{T}$. For any surface (finite number of plaquettes), such that the above closed path is its boundary, we denote the operator of total magnetic flux through it, i.e. the sum of all operators $\hat{B}_{x,\hat{k},\hat{l}}$ corresponding to this surface, by $\hat{\mathcal{B}}_{x,x+\hat{k}}(\mathcal{T})$. Of course, this quantity does not depend upon the choice of the surface, because the divergence of $\hat{B}$ vanishes. We call these quantities "along-tree magnetic fluxes. From (37) we have

$$\left[\hat{\mathcal{B}}_{x,x+\hat{k}}(\mathcal{T}), \hat{E}_{y,y+\hat{l}}\right] = i\hbar\delta_{xy}\delta_{\hat{k}\hat{l}} \,, \tag{39}$$

which means that the operators canonically conjugate to the along-tree magnetic fluxes are equal to the off-tree electric fluxes. From now on we denote them by $\hat{\mathcal{E}}_{y,y+\hat{l}}(\mathcal{T})$.

Finally, let us describe all electron and positron degrees of freedom as follows: For operators $\hat{\phi}^K_x$ and $\hat{\varphi}_{y;L}$ we write $\hat{\phi}_i$ and $\hat{\varphi}_j$, where $i = (x, K)$, $j = (y, L)$ label all possible values of indices K and L, and all lattice sites x and y. We denote by $\hat{m}_{ij}$ those generators $\hat{\mathcal{M}}$, which correspond to the "on-tree paths":

$$\hat{m}_{ij} := \hat{\mathcal{M}}_{\gamma;L}{}^K \,, \tag{40}$$

where $i = (x, K)$, $j = (y, L)$ and γ denotes the unique "on-tree" path connecting the lattice site x with the lattice site y.

Definition 1 *For a given tree* $\mathcal{T}$ *and a given family* $\{\hat{E}, \hat{B}, \hat{\mathcal{M}}, \hat{\mathcal{M}}^*\}$ *of generators fulfilling our axioms, we call the family* $\{\hat{\mathcal{E}}_{y,y+\hat{l}}(\mathcal{T}), \hat{\mathcal{B}}_{x,x+\hat{k}}(\mathcal{T}), \hat{m}_{ij}, \hat{m}^*_{ij}\}$ *the tree data of* $\{\hat{E}, \hat{B}, \hat{\mathcal{M}}, \hat{\mathcal{M}}^*\}$.

Observe that the tree data inherit the following properties from axioms (31) – (37):

Proposition 4

$$[\hat{m}_{ij}, \hat{m}_{kl}] = 0\ , \tag{41}$$

$$(\hat{m}_{ij})^2 = 0\ , \tag{42}$$

$$[\hat{m}_{ij}, \hat{m}^*_{kl}, \hat{m}_{rs}] = \delta_{ki}\,\delta_{ls}\,\hat{m}_{rj} + \delta_{kr}\,\delta_{lj}\,\hat{m}_{is}\ . \tag{43}$$

$$\left[\hat{m}_{ij}, \hat{\mathcal{B}}_{x;x+\hat{k}}(\mathcal{T})\right] = 0\,, \tag{44}$$

$$\left[\hat{m}_{ij}, \hat{\mathcal{E}}_{x,x+\hat{k}}(\mathcal{T})\right] = 0\,, \tag{45}$$

$$\left[\hat{\mathcal{B}}_{x;x+\hat{k}}(\mathcal{T}), \hat{\mathcal{E}}_{y,y+\hat{l}}(\mathcal{T})\right] = i\hbar\delta_{xy}\delta_{\hat{k}\hat{l}}\,. \tag{46}$$

It turns out[6] that the $*$-algebra O_m generated by operators $\hat{m}$ and $\hat{m}^*$, fulfilling axioms (41) – (43), is finite dimensional. Moreover, the following is true:

Theorem 1 *Let there be given a tree* $\mathcal{T}$. *Every family* $\{\hat{E}, \hat{B}, \hat{\mathcal{M}}, \hat{\mathcal{M}}^*\}$ *of generators fulfilling axioms (30) – (37) together with (38), is in one-to-one correspondence with its tree data, fulfilling identities (41) – (46).*

Due to this Theorem, the $*$-algebra generated by $\{\hat{E}, \hat{B}, \hat{\mathcal{M}}, \hat{\mathcal{M}}^*\}$ fulfilling axioms (30) – (37), is isomorphic to the $*$-algebra generated by the tree data, fulfilling identities (41) – (46). Due to (39), the first two components, $\hat{\mathcal{E}}(\mathcal{T})$ and $\hat{\mathcal{B}}(\mathcal{T})$, fulfil the commutation relations of a finitely generated Heisenberg algebra. We take the corresponding Weyl-algebra generated by them and denote its strong closure by $O_{\mathrm{e-m}}(\mathcal{T})$. This algebra may be represented as $B(L^2(\mathcal{B}(\mathcal{T})))$, the algebra of bounded operators acting on the Hilbert space of L^2-integrable functions depending on classical variables $\mathcal{B}(\mathcal{T})$, with $\hat{\mathcal{B}}_{x,x+\hat{k}}(\mathcal{T})$ defined as multiplication and $\hat{\mathcal{E}}_{x,x+\hat{k}}(\mathcal{T})$ as differentiation operators (Schrödinger representation). It follows from (44) and (45) that O_{m} and $O_{\mathrm{e-m}}(\mathcal{T})$ commute. Thus, the C^*-algebra generated by $(\hat{\mathcal{E}}(\mathcal{T}), \hat{\mathcal{B}}(\mathcal{T}), \hat{m}, \hat{m}^*)$ is the tensor product of $O_{\mathrm{e-m}}(\mathcal{T})$ and O_{m}. Consequently, we have the following

Definition 2 *The observable algebra O is defined as*

$$\mathcal{O} := \mathcal{O}_{\mathrm{e-m}}(\mathcal{T}) \otimes \mathcal{O}_{\mathrm{m}}\,. \tag{47}$$

Due to Theorem 1 we may identify elements of O reconstructed from tree data corresponding to different trees. It is easy to check that this identification is an isomorphism of algebras. Hence, the algebras O generated from data corresponding to different trees coincide. Consequently, the definition of O does not depend upon the tree.

The following result easily follows from the above algebraic structure of the theory.

Theorem 2 *The total charge operator defines a superselection rule in* O . *We have*

$$\mathcal{O} = \bigoplus_{Q} \{\mathcal{O}_{\mathrm{e-m}}(\mathcal{T}) \otimes \mathcal{O}_{\mathrm{m}}(Q)\}\ , \tag{48}$$

where $O_{\mathrm{m}}(Q)$ *are the central decomposition components of* O_{m}.

UNIQUENESS OF IRREDUCIBLE REPRESENTATIONS AND CHARGE SUPERSELECTION SECTORS

We restrict ourselves to strongly continuous representations of the Weyl relations. They are unitarily equivalent to at most a countable sum of copies of the Schrödinger representation [12]. Thus, all representations of the observable algebra O may be constructed from the Schrödinger representation and from representations of O_{m}. In this Section we will describe all irreducible representations of the latter algebra.

In a first step, we are going to define a family of *canonical* representations of O_{m}. It turns out that every irreducible representation is isomorphic to one of them. For this purpose we denote $N = \{1,2, \ldots, N\}$, where N is – as in Section – the number of all positron degrees of freedom $\hat{\phi}_i$ (and also the number of all electron degrees of freedom $\hat{\varphi}_j$). For any integer $Z \in [-N + 1, N - 1]$ we define a finite-dimensional representation of O_{m} in the following way.

We denote by S_N the set of all subsets of N and take the free vector space

$$\mathcal{H}_{\mathrm{m}}(\mathcal{T}) := \bigoplus_{(I,J)\in S_N\times S_N} \mathcal{H}_{(I,J)}\,, \qquad \mathcal{H}_{(I,J)} \cong \mathbb{C}\,,$$

over $S_N \times S_N$. Next we restrict ourselves to the subspace H_z of such vectors that the number $\#I$ of elements of I differs from the number $\#J$ of elements of J exactly by Z:

$$\mathcal{H}_Z := \bigoplus_{\#I-\#J=Z} \mathcal{H}_{(I,J)}\,. \tag{49}$$

We endow H_Z with a Hilbert space structure by choosing in each subspace $H_{(I,J)} \cong \mathbb{C}$ the unit number $\Omega_{(I,J)} := 1_{(I,J)} \in H_{(I,J)}$ and treating $\{\Omega_{(I,J)}\}$ as an orthonormal basis. We have, of course,

$$\mathcal{H}_{\mathrm{m}}(\mathcal{T}) = \bigoplus_Z \mathcal{H}_Z\,. \tag{50}$$

We define the irreducible representation of the operators $\hat{m}$ and $\hat{m}^*$ on H_Z as follows:

$$\hat{m}_{ij}\Omega_{(I,J)} = \begin{cases} \mathrm{sgn}_{((i,j),(I\setminus\{i\},J\setminus\{j\}))}\ \Omega_{(I\setminus\{i\},J\setminus\{j\})} & \text{if } i \in I \text{ and } j \in J \\ 0 & \text{otherwise}\,, \end{cases} \tag{51}$$

$$\hat{m}^*_{kl}\Omega_{(I,J)} = \begin{cases} \mathrm{sgn}_{((k,l),(I,J))}\ \Omega_{(I\cup\{k\},J\cup\{l\})} & \text{if } k \notin I \text{ and } l \notin J \\ 0 & \text{otherwise}\,. \end{cases} \tag{52}$$

By $\mathrm{sgn}_{((k,l),(I,J))}$ ($k \notin I$ and $l \notin J$) we denote the parity (± 1) of the permutation, which is necessary to reestablish the canonical order of the sequence $(k, l, i_1 \ldots i_k, j_1 \ldots j_1)$. It is easy to check that the operators $\hat{m}$ represented this way fulfil the defining relations of the algebra O_{m}.

The above constructed representations will be called Z-representations.

The Hilbert space H_Z constructed above may also be treated as a subspace of the fermionic Fock space defined by N (abstract) positron degrees of freedom ϕ_i and N electron degrees of freedom φ_j:

$$\mathcal{H}_Z \subset \mathcal{F}(\mathbb{C}^N) \bigotimes \mathcal{F}(\mathbb{C}^N)\,, \tag{53}$$

where by $\mathcal{F}(\mathbb{C}^N)$ we denote the fermionic Fock space with N generators, satisfying the canonical anticommutation relations. In this space, vectors $\Omega_{(I,J)}$ may be represented as canonically ordered monomials $\phi^*_{i_1} \cdots \phi^*_{i_k} \varphi^*_{j_1} \cdots \varphi^*_{j_l}$, (where $i_1 < \cdots < i_k$ and $j_1 <$

… < *ji* are all elements of *I* and *J* respectively) of Grassmannian (anticommuting) variables ϕ_* and φ_*. They are obtained from the Fock vacuum by the action of creation operators:

$$\phi^*_{i_1}\cdots\phi^*_{i_k}\varphi^*_{j_1}\cdots\varphi^*_{j_l} := \hat{\phi}^*_{i_1}\cdots\hat{\phi}^*_{i_k}|\omega_p> \otimes \hat{\varphi}^*_{j_1}\cdots\hat{\varphi}^*_{j_l}|\omega_e> , \tag{54}$$

and $|\omega_p>$ (respectively $|\omega_e>$) denotes the (fermionic) Fock vacuum for positrons (resp. electrons). The subspace H_Z corresponds to the value $Q = eZ$ of the total charge. In the polynomial representation of H_Z, operators $\hat{m}^*$ may be identified with multiplication operators by the Grassmannian 2-nd order quantity $m^*_{ij} := \phi^*_i\varphi^*_j$. The coefficient arising in (52) is chosen in such a way that it reestablishes the canonical order in the product $\phi^*_k\varphi^*_l\phi^*_{i_1}\cdots\phi^*_{i_k}\varphi^*_{j_1}\cdots\varphi^*_{j_l}$.

The following, fundamental result was proved[5,6]:

Theorem 3 *Every irreducible representation of* O_{m} *is unitarily equivalent to one of the Z-representations defined by (51) and (52).*

LOCAL OBSERVABLES AND LATTICE QUANTUM HAMILTONIAN

The operators $\hat{m}$ are non-local. But due to (43) they can be expressed in terms of local quantities, namely annihilation operators of pairs located at the same lattice site or pairs separated by at most one lattice link:

$$\hat{u}_{x;L}{}^K := \hat{\varphi}_{x;L}\hat{\phi}^K_x \tag{55}$$

$$\hat{w}_{x,x+\hat{k};L}{}^K := \hat{\varphi}_{x+\hat{k};L}\exp(-igA_{x,x+\hat{k}})\hat{\phi}^K_x . \tag{56}$$

Using (34) we can express all pair creation and annihilation operators along any *long* path γ as a multiple commutator of the above *short* pair creation operators assigned to links and points belonging to γ. This way, the observable algebra *O* may be viewed as the algebra generated by the following set of local generators: $\{\hat{E}, \hat{B}, \hat{u}, \hat{w}\}$.

The quantum evolution of the field is governed by the second quantized Hamiltonian (4). Using standard lattice approximation recipes one gets its lattice approximation:

$$\hat{H} = \hat{H}_{\text{e-m}} + \hat{H}_{\mathrm{m}} + \hat{H}_{\mathrm{kin}}, \tag{57}$$

$$\hat{H}_{\text{e-m}} = \frac{1}{2}\sum_{(x,x+\hat{k})}\left(\hat{E}_{x,x+\hat{k}}\right)^2 + \frac{1}{2}\sum_{x;\hat{k},\hat{l}}\left(\hat{B}_{x;\hat{k},\hat{l}}\right)^2 , \tag{58}$$

$$\hat{H}_{\mathrm{m}} = m\sum_x [(\mathrm{tr}\,\hat{u}_x)^*, \mathrm{tr}\,\hat{u}_x] , \tag{59}$$

$$\hat{H}_{\mathrm{kin}} = -\frac{\hbar}{a}\sum_{x,\hat{k}}\mathrm{sgn}\,\hat{k}\cdot\mathrm{Im}\left\{\sigma^{kL}{}_K\cdot\hat{w}_{x,x+\hat{k};L}{}^K\right\} . \tag{60}$$

The above Hamiltonian is bounded from below, because $\hat{H}_{\text{e-m}}$ is positive definite and the remaining terms are bounded operators. Hence, we may define the *dynamical vacuum* as the minimal energy state in the vacuum sector $Q = 0$ and the notion of a *dressed particle* as the minimal energy state in the $Q = e$ sector, with appropriately chosen boundary data $E_{x,\infty}$. We stress that these states have nothing to do with the *perturbative vacuum* and the notion of a *bare particle* in the perturbative approach.

There is, probably, no way to obtain an exact, analytic expression for these states, even on the lattice level, and a numerical analysis will remain as the only tool to

investigate the spectrum of the operator $\hat{H}$. There is, however, an interesting idea[15] to consider $\hat{H}_{\mathrm{kin}}$ as the perturbation to the operator $\hat{H}_0 := \hat{H}_{\mathrm{e\text{-}m} } + \hat{H}_{\mathrm{m}}$. We stress that this type of perturbative approach has nothing to do with "switching the interaction off", because quantum states corresponding to $g = 0$ and to $g \neq 0$ belong to completely different Hilbert spaces.

TOWARDS CONTINUUM THEORY

Here, we present some ideas concerning the construction of the full continuum quantum theory. Heuristically, the algebra of observables of the continuum theory should be constructed as an inductive limit[1,16] of our algebras O_Λ, each of them describing a finite number of degrees of freedom, related to a finite lattice Λ. For this purpose an order relation "$\prec$" in the set of finite lattices has to be chosen. We say that the lattice Λ_2 is "later" than Λ_1 (or $\Lambda_1 \prec \Lambda_2$) if it describes *more* field degrees of freedom than Λ_1 does. Thus, being "later" means being "bigger" respectively "finer", or both.

Given a pair $\Lambda_1 \prec \Lambda_2$, there is, obviously, a natural embedding

$$P_{\Lambda_2,\Lambda_1} : \mathcal{O}_{\Lambda_1} \to \mathcal{O}_{\Lambda_2} \ , \tag{61}$$

which preserves the properties (30) – (37). The Gauss law (38), however, is no longer true in its original version but in a new version, with the charge $\hat{j}^0_x$ replaced by the sum of all charges $\hat{j}^0_{x_i}$ corresponding to the sites x_i of Λ_2 which are contained in the same cell of the dual lattice as x. There is a natural compatibility relation, $P_{\Lambda_3,\Lambda_2} P_{\Lambda_2,\Lambda_1} = P_{\Lambda_3,\Lambda_1}$, if $\Lambda_1 \prec \Lambda_2 \prec \Lambda_3$

The inductive limit of our observable algebras describes, in principle, degrees of freedom of the continuum theory. To avoid singular objects, we may smear the fields $\hat{E}$, $\hat{B}$ and $\hat{u}$ with sufficiently regular test functions and obtain this way "observable-valued-distributions". The field $\hat{w}$ cannot be smeared *directly*, because of its non-additive character. A natural way to encode the information about the field $\hat{w}$ in an "observable-valued-distribution" consists in replacing it by the field

$$\hat{v}_{k;L}{}^K(x) := \lim_{a\to 0} \frac{1}{a} \left(\hat{w}_{x,x+\hat{k};L}{}^K - \hat{u}_{x;L}{}^K \right) \simeq \left(D_k \hat{\varphi}_{x;L}(x)\right) \hat{\phi}^K_x(x) \ . \tag{62}$$

Our previous work[3] suggest that $\hat{v}$, together with ultra-local invariants $\hat{u}_L{}^K(x) := \hat{\varphi}_{x;L} \hat{\phi}^K_x$, might be the fundamental fields of the continuum theory.

Once the observable algebra of the continuum theory is given, its Hilbert space representations may be constructed *via* the GNS construction – provided a vacuum state is given. This idea is also followed by Ashtekar and Lewandowski, but they propose to use a *perturbative* vacuum, constructed in a fully kinematic way (in the case of bosonic degrees of freedom their construction of the Hilbert space[1] consists, in fact, in choosing the Gaussian wave function as a vacuum state or the constant function in case of the compact representation of gauge fields). In our opinion such a choice might possibly lead to an unphysical sector of the theory, because it is not plausible that the perturbative vacuum belongs to the physical sector. A possible way to avoid this difficulty consists in approximating the vacuum state of the continuum theory by the *true* vacuum states of its lattice approximations[16]. For this purpose observe that the space of states S_Λ (not necessarily *pure states*, but all the *mixed states*) may be treated as being dual to the observable algebra O_Λ. This implies that we have a family of dual mappings

$$P^*_{\Lambda_2,\Lambda_1} : \mathcal{S}_{\Lambda_1} \leftarrow \mathcal{S}_{\Lambda_2} \ , \tag{63}$$

defined for $\Lambda_1 \prec \Lambda_2$.

Physically, O_{Λ_1} may be thought of as a *subsystem* of a bigger physical system O_{Λ_2}, containing more degrees of freedom. The above mapping assigns, to every state of a bigger system, a mixed state of the subsystem. This state is obtained by "forgetting" about those degrees of freedom which are not contained in the subsystem.

Let $\omega_\Lambda \in S_\Lambda$ denote the vacuum state, corresponding to the minimum of the Hamiltonian (57). Let us project the vacuum from "finer lattices" backwards to "coarser lattices" and define

$$\omega_{\Lambda_1,\Lambda_2} := P^*_{\Lambda_2,\Lambda_1}\omega_{\Lambda_2} \in \mathcal{S}_{\Lambda_1} \ . \tag{64}$$

Suppose that the limit $\Omega_{\Lambda_1} = \lim_{\Lambda_2} \omega_{\Lambda_1,\Lambda_2}$ exists. If this is true for every Λ_1, then the compatibility condition $\Omega_{\Lambda_1} = P^*_{\Lambda_2,\Lambda_1}\Omega_{\Lambda_2}$ is automatically fulfilled and the state $\Omega := \{\Omega_\Lambda\}$ belongs to the projective limit of spaces S_Λ. Therefore, it defines a state on the inductive limit of the algebras O_Λ. As a limit of approximate vacuum states, it is a natural candidate for the non-perturbative vacuum of the continuum theory and the starting point for the GNS construction of the Hilbert space of its quantum states.

The existence of the above limit of vacuum states may be extremely difficult to prove. A realistic attitude consists, therefore, in a detailed analysis of lattice approximations of the theory presented in this paper: if the continuum limit of these approximations does exist, the numerical results obtained on the level of a sufficiently "late" lattice Λ should approximate the true physical quantities.

REFERENCES

1. A. Ashtekar, J. Lewandowski, "Differential geometry on the space of connections via graphs and projective limits", Journ. Geom. Phys. **17** (1995), p. 191–230
2. R. Matthews, "Beyond space and time", New Scientist **154** No 2082 (1997) p. 38-42
3. J. Kijowski, G. Rudolph, M. Rudolph, "Functional integral of QED in terms of gauge invariant quantities", Lett. Math. Phys **33** (1995) p. 139–146
4. J. Kijowski, G. Rudolph, M. Rudolph, "Effective bosonic degrees of freedom for one-flavour chromodynamics", Ann. Inst. H. Poincaré (in print)
5. J. Kijowski, G. Rudolph and A. Thielman, "The algebra of observables and charge superselection sectors for QED on the lattice", Commun. Math. Phys. (in print), prepr. 40/1996 Leipzig University
6. J. Kijowski, G. Rudolph, C. Śliwa, "On the structure of the observable algebra for QED on the lattice", submitted to Lett. Math. Phys.
7. F. Strocchi and A. Wightman, "Proof of the charge superselection rule in local relativistic quantum field theory", J. Math. Phys. 15 (1974) 2198
8. F. Strocchi, "Spontaneous symmetry breaking in local gauge quantum field theory; the Higgs mechanism", Commun. Math. Phys. 56 (1977) 57
9. J. Fröhlich, "The charged sectors of quantum electrodynamics in a framework of local observables", Commun. Math. Phys. 66 (1979), 223
10. D. Buchholz, "The physical state space of quantum electrodynamics", Commun. Math. Phys. 85 (1982) 49
    D. Buchholz, "Gauss' law and the infraparticle problem", Phys. Lett. B174 (1986) 331
11. K. Fredenhagen and M. Marcu, "Charged states in Z_2 gauge theories", Commun. Math. Phys. 92 (1983) 81
12. J. v. Neumann, "Die Eindeutlichkeit der Schrödingerschen Operatoren", Math. Ann., bd. 104 (1931) 570
13. P. Jordan and E. P. Wigner, "Über das Paulische Äquivalenzverbot" Zeitschr. f. Phys. 47 (1928) 631
14. F. A. Berezin, "The Method of Second Quantization", Academic Press, New York and London 1966
15. I. Bialynicki-Birula, "The hamiltonian of quantum dynamics", in "QED and Quantum Optics", Ed. A. O. Barut, Plenum Press, N. Y. 1984

16. J. Kijowski, “Symplectic geometry and second quantization”, Rep. Math. Phys. 11 (1977) 97
S. Zakrzewski, “On the lattice approximation and Feynman path integrals for gauge fields”, Thesis, Warsaw 1980, unpublished
S. Zakrzewski, “Unitary Relations”, Publicações de Fisica Matemática 4 (1985), Coimbra

THREE INTRODUCTORY TALKS ON MATRIX MODELS OF SUPERSTRINGS

Yuri Makeenko

Institute of Theoretical and Experimental Physics
B. Cheremushkinskaya 25, 117259 Moscow, Russia
and
The Niels Bohr Institute
Blegdamsvej 17, 2100 Copenhagen Ø, Denmark

PREFACE

These introductory talks on the matrix models of Superstrings and M theory were not given at the Nato Advanced Research Workshop on Theoretical Physics in Zakopane for the reason which pretty much reminds the one widely known from the times of the Former Soviet Union. I am grateful to the Organizers who nevertheless invited me to make a contribution to the Proceedings. I used for this purpose the notes of three lectures at 5th Nordic Meeting on Supersymmetric Field and String Theories in Helsinki (March 10–12, 1997) which exist in the e-Print Archive but has never been published.

Each of the three talks is mostly concentrated around one of the three selected papers[1,2,3]. The references in the text are only to the results quoted. More complete list of references can be found clicking a mouse on the number of citations to the pioneering paper of Banks, Fischler, Shenker and Susskind[1] in HEP database at SLAC. An up-to-date survey of the subject can be found in the talk by Poul Olesen at this Workshop.

Contents:

1. M(atrix) theory of BFSS,
2. From IIA to IIB with IKKT,
3. The NBI matrix model.

INTRODUCTION

The standard non-perturbative approach to bosonic (Polyakov) string, which is based on discretized random surfaces and matrix models, exists since the middle of the eighties[4,5,6]. The main result of the investigations (both analytical and numerical) within this approach is that the bosonic string is not in the stringy phase but rather in a *branched polymer* phase when the dimension D of the embedding space is larger

New Developments in Quantum Field Theory
Edited by Damgaard and Jurkiewicz, Plenum Press, New York, 1998

than one*. This is the way how the tachyonic problem is resolved for $D > 1$. In other words the perturbative vacuum with the tachyon is unstable and the system chooses a stable vacuum which is *not* associated with strings.

A question immediately arises what about superstrings where the GSO-projection kills tachyons (at least perturbatively). This is a strong argument supporting the expectation for superstrings to live in a stringy phase, which agrees with the common belief that fermions smooth out the dynamics.

The attempts (not quite successful until very recent time) of discretizing superstrings are performed starting from[8]. The problem resides, roughly speaking, in the fact that is not easy to discretize the target-space supersymmetry (SUSY). A progress had been achieved only for the simplest case of pure two dimensional supergravity which can be associated with a supereigenvalue model[9]. For a more detail review, see[7].

The dramatic recent progress in a non-perturbative formulation of superstrings by supersymmetric matrix models, which has occurred during last few months, is the subject of these notes. I shall mostly concentrate on ten dimensional superstrings practically leaving outside presumably most interesting question of constructing the fundamental Lagrangian of eleven dimensional M theory in the language of matrix models.

M(ATRIX) THEORY OF BFSS[1]

Eleven dimensional M theory combines different ten dimensional superstring theories (IIA, IIB, . . .), which are in fact related by duality transformations, into a single fundamental theory. BFSS proposed[1] to describe it by a supersymmetric matrix quantum mechanics in the limit of infinite matrices. This construction is called M(atrix) theory.

The Set Up

The point of interest of [1] is $D = 10 + 1$ dimensional M theory (characterized by its Planck's length, l_p). The eleven coordinates

$$x^\mu = \left(t, x^i, x^{11}\right) \qquad (i = 1, \ldots, 9)$$

are split into time, t, the nine *transverse* ones, x^i or $x^\perp$, and the *longitudinal* one, called x^{11}, which is compactified:

$$x^{11} = x^{11} + 2\pi R .$$

The radius of compactification R plays the role of an infrared cutoff in the theory.

The system is considered in the Infinite Momentum Frame (IMF), which is more or less the same as the light cone frame, boosting along the longitudinal axis. The role of the light cone variables is played by $t \pm x^{11}$. The advantage of using IMF is that only positive momenta p_{11} are essential while systems with zero or negative p_{11} do not appear as independent dynamical degrees of freedom. A price for this is the absence of manifest Lorentz invariance.

Due to compactness all systems have (positive) longitudinal momentum

$$p_{11} = \frac{N}{R} \qquad (N > 0), \tag{1}$$

*For a review see [7].

where $N > 0$ is integer. At the end of calculations R should tend to infinity,

$$R \to \infty, \tag{2}$$

to get *uncompactified* infinite momentum limit of 11D theory. N will be identified in what follows with the Ramond-Ramond (RR) charge of the system.

10D Versus 11D Language

M theory with compactified x^{11} is by construction type *IIA superstring* in $D = 9+1$ dimensions. The parameters R and l_p of eleven dimensional M theory and those g_s and l_s of the ten dimensional superstring are related by

$$R = g_s^{2/3} l_p\,, \qquad l_s = g_s^{-1/3} l_p\,, \tag{3}$$

where g_s is the *string coupling constant* and $l_s \equiv \sqrt{\alpha'}$ is the string length scale related to the *string tension* T by

$$T = \frac{1}{2\pi\alpha'}\,. \tag{4}$$

The 11D M theory is in turn a strong coupling limit of 10D IIA superstring, since (2) is guaranteed as $g_s \to \infty$.

No perturbative string states carry RR charge $\Rightarrow$ they are associated with vanishing momentum p_{11}. 1 unit of RR charge is carried by D0-brane of Polchinski[10] for which

$$p_{11} = \frac{1}{R} \tag{5}$$

in accord with Eq. (1) at $N = 1$.

The *low-energy* limit of M theory is 11D supergravity having 256 massless states: 44 gravitons, 84 three-forms and 128 gravitinos. These 256 states are referred to as *supergravitons* which are *massless* as 11D objects $\Rightarrow$ they are Bogomolny–Prasad–Sommerfield (BPS) saturated states in 10D theory. Their 10D mass $\sim 1/R$.

States with $N \neq 1$ are not associated with elementary D0-branes. The states with $N > 1$ are bound composites of N D0-branes as is discussed in the next subsection.

The Appearance of Matrices

The world-volume of a p-brane is parametrized by $p + 1$ coordinates $\xi_0, \ldots, \xi_p$. The p-branes emerge as classical solutions in 10D supergravities† which describe low-energy limits of 10D superstrings. They possess an intrinsic abelian gauge field $A_\alpha(\xi)$ ($\alpha = 0, \ldots, p$) which can be viewed as tangent (to p-brane) components of 10D abelian gauge field reduced to p-brane. Otherwise, the remaining $9 - p$ components of the 10D abelian gauge field, which are orthogonal to the p-brane, are associated with its coordinates[10]

$$X_i(\xi) = 2\pi\alpha' A_i(\xi) \qquad (i = p+1, \ldots, 9)\,. \tag{6}$$

A D(irichlet) p-brane can emit a fundamental open string which has the Dirichlet boundary condition on a $p + 1$ dimensional hyperplane and the Neumann boundary condition in the $9 - p$ dimensional bulk of space. This string can end either on the same Dp-brane or on another one as is illustrated by Fig. 1.

If one has N parallel Dp-branes separated by some distances in the $9 - p$ dimensional space, then massless vector states emerge only when the string begins and ends

† For a review, see [11].

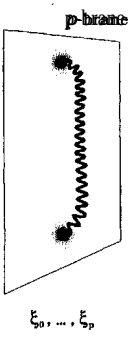


Figure 1. Dp-brane (depicted by a hyperplane parametrized by the coordinates $\xi_0, \ldots, \xi_p$) and the fundamental string.

at the same brane, so the gauge group $U(1)^N$ appears in a natural way. Since the energy of strings stretched between different D-branes is

$$M \sim T\,|X^i - X^j|\,, \tag{7}$$

more massless vector states appear when the branes are practically on the top of each other. Since the string is oriented, all possible massless states when the string begins and ends either on same or different Dp-branes form a U(N) multiplet when strings are very short. The example of $N = 2$ is illustrated by Fig. 2. This is how hermitian $N \times N$ matrices appear in the description of bound composites of N Dp branes according to Witten[12].

For our case of N D0-branes, their coordinates $X_i(t)$ become 9 Hermitean $N \times N$ matrices $X_i^{ab}(t)$ (accompanied by the fermionic superpartners $\theta_\alpha^{ab}(t)$ which are 16 component nine dimensional spinors). They can be thought as spatial components of the vector field in ten dimensional super Yang–Mills theory after reduction to zero space dimension (same for the superpartners). N is associated with the value of the RR charge of these states.

The Fundamental Lagrangian

The possibility of formulating the fundamental Lagrangian of M theory as a matrix model is stated in [1] as the

Conjecture: M theory in IMF is a theory with the only dynamical degrees of freedom of D0-branes.

In other words *all* systems are composed of D0-branes. Therefore, the fundamental Lagrangian of M theory is completely expressed via the hermitian $N \times N$ matrices $X_i^{ab}(t)$ describing coordinates of D0-branes (and their fermionic superpartners $\theta_\alpha^{ab}(t)$), so that

$$\boxed{\text{M theory} = \text{M(atrix) theory}}.$$

M(atrix) theory is described (in units of $l_s = 1$) by the Lagrangian

$$L = \frac{1}{2g_s}\,\mathrm{tr}\,\left(\dot{X}^i\dot{X}^i + 2\theta^T\dot{\theta} - \frac{1}{2}[X^i, X^j]^2 - 2\theta^T\gamma_i[\theta, X^i]\right). \tag{8}$$

Here $N \to \infty$ in order to satisfy (2).

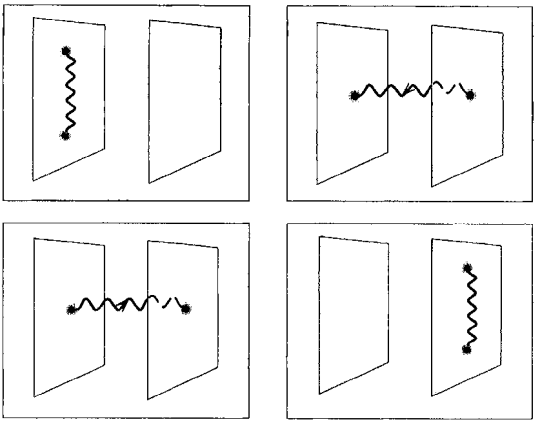


Figure 2. Appearance of matrices in the example of bound states of two parallel D-branes (N = 2). The fundamental string can begin and end either at the same or different D-branes. Since the string is oriented, there are four massless vector states when the branes are practically on the top of each other. They form a representation of U(2).

Changing the units to those where eleven dimensional l_p = 1 and introducing

$$Y = \frac{X}{g_s^{1/3}},$$

Eq. (8) can be rewritten as

$$L = \mathrm{tr}\left(\frac{1}{2R} D_t Y^i D_t Y^i - \frac{1}{4} R\,[Y^i, Y^j]^2 - \theta^T D_t \theta - R\,\theta^T \gamma_i [\theta, Y^i]\right), \tag{9}$$

where

$$D_t = \partial_t + iA_0 \tag{10}$$

is the covariant derivative with respect to the A_0 field. Equation (8) is written in the A_0 = 0 gauge.

The Lagrangian (9) is invariant under two *SUSY transformations*

$$\delta_{\mathrm{SUSY}} X^i = -2\epsilon^T \gamma^i \theta, \tag{11}$$

$$\delta_{\mathrm{SUSY}} \theta = \frac{1}{2}\left(D_t X^i \gamma_i + \gamma_- + \frac{1}{2}[X^i, X^j]\,\gamma_{ij}\right)\epsilon + \epsilon', \tag{12}$$

$$\delta_{\mathrm{SUSY}} A_0 = -2\epsilon^T \theta, \tag{13}$$

where ϵ and ϵ' are two independent 16 component (t-independent) parameters. It is seen from this formula that A_0 is needed to close the SUSY algebra.

Matrix Quantum Mechanics

The *Hamiltonian* which is associated with the Lagrangian (9) reads

$$H = R\,\mathrm{tr}\left\{\frac{\Pi_i \Pi_i}{2} + \frac{1}{4}\,[Y^i, Y^j]^2 + \theta^T \gamma_i [\theta, Y^i]\right\}, \tag{14}$$

where Π_i is the canonical conjugate to Y^i. As is usual for fermions, a half of θ_α^{ab} plays the role of coordinates and the other half plays the role of canonical conjugate momenta in the language of 1st quantization.

All finite energy states of the 10D Hamiltonian (14) acquire infinite energy as $R \to \infty$, *i. e.* in the uncompactified 11D limit. Only the states whose energy $\sim 1/N$ as $N \to \infty$ yield

$$H \sim \frac{R}{N} = \frac{1}{p_{11}} \tag{15}$$

as is expected since $p^2 = 2Ep_{11} - p_\perp^2$ in 11D IMF, so that

$$p^2 = 0 \ \text{(in 11D)} \implies E = \frac{p_\perp^2}{2p_{11}} \ \text{(in 10D)} \tag{16}$$

in 10D.

The simplest states of the Hamiltonian (14) is when the matrices Y^i are diagonal with only one nonvanishing diagonal component and all θ's equal zero. For nonvanishing $p_\perp$ — the eigenvalue of $\Pi_\perp$ — Eq. (14) yields

$$E = \frac{R}{2} p_\perp^2 = \frac{p_\perp^2}{2p_{11}} \tag{17}$$

since the commutators vanish. Thus we get Eq. (5) with $N = 1$ and this state corresponds to a single D0-brane in 10D language.

Each of these states is accompanied by the fermionic superpartners and they form a representation of the algebra of 16 θ's with

$$2^{16/2} = 2^8 = 256$$

components. They are exactly 256 states of supergraviton in 11D. In the 10D language these are BPS states of the mass $\sim 1/R$ which become massless in the uncompactified limit $R \to \infty$.

A more general eigenstate of the Hamiltonian (14) has a form of the *diagonal* $N \times N$ matrix

$$Y_i = \begin{pmatrix} Y_i^{(1)} & & \\ & \ddots & \\ & & Y_i^{(N)} \end{pmatrix}. \tag{18}$$

The commutator obviously vanishes in this case.

It is convenient to split the U(N) group as U(1) $\otimes$ SU(N) and to associate the U(1) part with the center mass coordinate

$$Y_i(cm) = \frac{1}{N} \operatorname{tr} Y_i . \tag{19}$$

Then

$$p_i(cm) = \operatorname{tr} \Pi_i = \frac{N}{R} \dot{Y}_i(cm) , \tag{20}$$

and using $p_{11} = N/R$ we get the usual relation

$$\frac{1}{p_{11}} p_i(cm) = \dot{Y}_i(cm) \tag{21}$$

between transverse velocity and momentum.

Interaction states are described in this construction by *non-diagonal* matrices. They correspond to *scattering states* of supergravitons in 11D. The interaction of supergravitons at the tree level is correctly reproduced within M(atrix) theory.

a

The Hamiltonian (14) of the $N \to \infty$ supersymmetric quantum mechanics looks pretty much like the one[13] for a 11D supermembrane in IMF. While there are no truly stable finite energy membranes in the decompactified limit, there exist very long lived classical membranes.

The membrane action can be derived in the Weyl basis on gl(N), which is given by two unitary $N \times N$ matrices g and h (clock and shift operators) obeying

$$hg = \omega gh\,, \quad \omega = \mathrm{e}^{2\pi i/N}\,, \tag{22}$$

$$h^N = 1 = g^N\,. \tag{23}$$

Any hermitian $N \times N$ matrix Z can be expanded in this basis as

$$Z = \sum_{n,m=1}^{N} Z_{n,m} g^m h^n\,. \tag{24}$$

As $N \to \infty$, we can introduce a pair of canonical variables q and p, so that

$$g = \mathrm{e}^{ip}\,, \qquad h = \mathrm{e}^{iq}\,, \tag{25}$$

$$[q\,,\, p] = \frac{2\pi i}{N}\,. \tag{26}$$

As usual in quantum mechanics, the last equality is possible only as $N \to \infty$. Then, we have

$$\mathrm{tr}\, Z \Rightarrow N \int dpdq Z(p,q)\,, \tag{27}$$

$$[X\,,\, Y] \Rightarrow \frac{i}{N}\{\partial_q X \partial_p Y - \partial_p X \partial_q Y\} \tag{28}$$

for the trace and the commutator, and finally[1]

$$\boxed{\text{M(atrix) action} \Rightarrow \text{Supermembrane action}}$$

as $N \to \infty$.

A special comment is needed concerning the continuum spectrum of the supermembrane[14]. From the point of view of the M(atrix) theory, it is as a doctor ordered for describing the supergraviton scattering states. The conjecture of M(atrix) theory is that there exists a normalizable bound state at the beginning of the continuum spectrum at $p^2 = 0$.

The emergence of membrane states in M(atrix) theory can be seen from the classical equations of motion

$$\left[Y^i, \left[Y^j\,,\, Y^i\right]\right] = 0\,, \qquad \left[Y^i\,,\, (\gamma_i\theta)_\alpha\right] = 0 \tag{29}$$

which are satisfied by static configurations.

An infinite membrane stretched out in the 8,9 plane is given by[1]

$$Y^8 = R_8\sqrt{N}p\,, \quad Y^9 = R_9\sqrt{N}q\,,$$
$$\text{all other Y's and } \theta\text{'s} = 0, \tag{30}$$

where p and q are $N \to \infty$ matrices (operators), and R_8 and R_9 are (large enough) compactification radii. Equations (29) are satisfied by (30) because

$$[\mathrm{Y}^8\,,\, \mathrm{Y}^9\,] = \text{c-number}. \tag{31}$$

The membrane in this picture is built out of infinitely many D0-branes.

The interaction between these membrane configurations has been studied [15, 16, 17] and compared with the superstring results.

FROM IIA TO IIB WITH IKKT[2]

M(atrix) theory naturally describes ten dimensional IIA superstring. IKKT proposed[2] another matrix model associated with IIB superstring, which is in spirit of the Eguchi–Kawai large-N reduced ten dimensional super Yang–Mills theory. This non-perturbative formulation of IIB superstring is called the IKKT matrix model.

Preliminaries

IIB superstring differs from IIA superstring by chiralities of the fermionic superpartners. They are opposite for IIA superstring and same for IIB superstring.

As a consequence of this, Dp-branes of even p (p = 0, 2, 4, . . .) are consistently incorporated by type IIA superstring theory while type IIB superstring is associated[10] with Dp-branes of *odd* p (p = –1, 1, 3, 5, . . .). This is due to the rank of the antisymmetric field which is odd for IIA superstring and even for IIB superstring. Correspondingly, the analog of D0-brane (associated with p = 0 in the IIA case) is D-instanton (associated with p = –1 in the IIB case) and the analog of D-membrane (associated with p = 2 in the IIA case) is D-string (associated with p = 1 in the IIB case).

In analogy with [1] where the fundamental Lagrangian is expressed in terms of D0-branes, one might expect that IIB superstring is described in terms of D-instanton variables, *i.e.* by the ten dimensional super Yang–Mills dimensionally reduced to a point[12].

Schild Formulation of IIB Superstring

The starting point in the IKKT approach is the Green–Schwartz action of type IIB superstring theory with fixed κ-symmetry:

$$S_{\rm GS} = -T \int d^2\sigma \left\{ \sqrt{-\boldsymbol{\sigma}^2} + 2i\varepsilon^{ab}\partial_a X^\mu \bar{\Psi}\gamma_\mu \partial_b \Psi \right\}, \tag{32}$$

where

$$\boldsymbol{\sigma}^{\mu\nu} = \varepsilon^{ab}\partial_a X^\mu \partial_b X^\nu , \tag{33}$$

the vector index μ of $X^\mu\ (\sigma_1, \sigma_2)$ runs from 0 to 9 and the spinor index α of $\Psi_\alpha(\sigma_1, \sigma_2)$ runs from 1 to 32. The fermion Ψ is a Majorana–Weyl spinor in 10D which satisfies the condition $\gamma_{11}\Psi = \Psi$, so that only 16 components effectively remain.

The action (32) is invariant under the N = 2 *SUSY transformation*

$$\begin{aligned} \delta_{\rm SUSY}\Psi_\alpha &= \frac{1}{2\sqrt{-\frac{1}{2}\boldsymbol{\sigma}^2}} \boldsymbol{\sigma}^{\mu\nu} (\gamma_{\mu\nu}\epsilon)_\alpha + \xi_\alpha , \\ \delta_{\rm SUSY} X^\mu &= 4i\bar{\epsilon}\gamma^\mu \Psi \end{aligned} \tag{34}$$

whose parameters ϵ and ξ do not depend on σ_1 and σ_2.

The action (32) can be rewritten in the Schild form

$$S_{\rm Schild} = \int d^2\sigma \left\{ \sqrt{g}\,\alpha \left(\frac{1}{4}\{X^\mu, X^\nu\}^2 - \frac{i}{2}\bar{\Psi}\gamma^\mu \{X_\mu, \Psi\} \right) + \beta\sqrt{g} \right\}, \tag{35}$$

where $\sqrt{g(\sigma_1, \sigma_2)}$ is positive definite scalar density (which is considered as an independent dynamical variable) and the Poisson bracket is defined by

$$\{X, Y\} \equiv \frac{1}{\sqrt{g}} \varepsilon^{ab}\partial_a X \partial_b Y . \tag{36}$$

Note that $\sqrt{g}$ cancels in the fermionic term in the action.

The equivalence of (32) and (35) at the classical level can be proven by using the classical equation of motion for $\sqrt{g}$. Varying the Schild action (35) with respect to $\sqrt{g}$, we get

$$-\frac{1}{4}\alpha\frac{1}{\left(\sqrt{g}\right)^2}\left(\varepsilon^{ab}\partial_a X^\mu \partial_b X^\nu\right)^2 + \beta = 0\,. \tag{37}$$

Substitution of the solution

$$\sqrt{g} = \frac{1}{2}\sqrt{\frac{\alpha}{\beta}}\sqrt{\left(\varepsilon^{ab}\partial_a X^\mu \partial_b X^\nu\right)^2} \tag{38}$$

into (35) restores the Nambu–Goto form (32) of the Green–Schwartz action:

$$S_{\mathrm{NG}} = T\int d^2\sigma\left\{\sqrt{\alpha\beta}\sqrt{\left(\varepsilon^{ab}\partial_a X^\mu \partial_b X^\nu\right)^2} - \frac{i}{2}\,\alpha\,\varepsilon^{ab}\partial_a X^\mu \bar{\Psi}\gamma_\mu \partial_b \Psi\right\}. \tag{39}$$

The action (39) is invariant under the $N = 2$ SUSY transformation

$$\begin{aligned}\delta_{\mathrm{SUSY}}\Psi_\alpha &= -\frac{1}{2}\sqrt{g}\{X_\mu, X_\nu\}\left(\gamma^{\mu\nu}\epsilon\right)_\alpha + \xi_\alpha,\\ \delta_{\mathrm{SUSY}}X^\mu &= i\bar{\epsilon}\gamma^\mu\Psi\,,\end{aligned} \tag{40}$$

where the parameters ϵ and ξ do not depend again on σ_1 and σ_2.

Finally the *partition function* in the Schild formulation of IIB superstring is defined by the path integral over the positive definite function $\sqrt{g}$, and over X^μ and Ψ_α:

$$Z_{\mathrm{Schild}} = \int D\sqrt{g}\,DX^\mu\,D\Psi_\alpha\,\mathrm{e}^{-S_{\mathrm{Schild}}}\,. \tag{41}$$

It is invariant under the SUSY transformation (40) since both the action (35) and the measure $DX^\mu D\Psi_\alpha$ are invariant.

Equations (35) and (41) represent[2] IIB superstring in the Schild formalism with fixed κ-symmetry.

In addition to the $N = 2$ SUSY transformation (40), the partition function is invariant at fixed $\sqrt{g}$ under area-preserving or *symplectic diffeomorphisms*

$$\delta_{\mathrm{sdiff}}X^\mu = \{X^\mu, \Omega\}\,, \qquad \delta_{\mathrm{sdiff}}\Psi_\alpha = \{\Psi_\alpha, \Omega\} \tag{42}$$

which is only a part of the whole reparametrization (or diffeomorphism) transformations. The invariance of the string theory under the whole group of reparametrizations is restored when $\sqrt{g}$ is transformed. The symmetry (42) reminds the non-abelian gauge symmetry in Yang–Mills theory and is to be fixed for doing perturbative calculations.

The IKKT Matrix Model

The IKKT matrix model can be obtained from the representation (41) of IIB superstring in the Schild formalism by replacing

$$X_\mu(\sigma_1, \sigma_2) \Longrightarrow A_\mu^{ab}\,, \tag{43}$$

$$\Psi_\alpha(\sigma_1, \sigma_2) \Longrightarrow \psi_\alpha^{ab}\,, \tag{44}$$

where A_μ^{ab} and ψ_α^{ab} are hermitian $n \times n$ bosonic and fermionic matrices, respectively.

The IKKT matrix model is defined by the partition function

$$Z = \sum_{n=1}^{\infty} \int dA_\mu \, d\psi_\alpha \, \mathrm{e}^{-S} , \tag{45}$$

which is of the type of 2nd quantized (euclidean) field theory, with the action

$$S = \alpha \left(-\frac{1}{4} \operatorname{tr} [A_\mu, A_\nu]^2 - \frac{1}{2} \operatorname{tr} (\bar{\psi} \gamma^\mu [A_\mu, \psi]) \right) + \beta n . \tag{46}$$

The summation over the matrix size n in Eq. (45) implies that n is a dynamical variable (an analog of $\sqrt{g}$ in Eq. (41)).

The action (46) and the measure $dA_\mu \, d\psi_\alpha$ in (45) are invariant under the $N = 2$ SUSY transformation

$$\begin{aligned} \delta_{\mathrm{SUSY}} \psi_\alpha^{ab} &= \frac{i}{2} [A_\mu, A_\nu]^{ab} (\gamma^{\mu\nu} \epsilon)_\alpha + \xi_\alpha \delta^{ab} , \\ \delta_{\mathrm{SUSY}} A_\mu^{ab} &= i \bar{\epsilon} \gamma_\mu \psi^{ab} , \end{aligned} \tag{47}$$

where the parameters ϵ and ξ are numbers rather than matrices, as well as under the SU(n) gauge transformation

$$\begin{aligned} \delta_{\mathrm{gauge}} A_\mu &= i [A_\mu, \omega] , \\ \delta_{\mathrm{gauge}} \psi_\alpha &= i [\psi_\alpha, \omega] . \end{aligned} \tag{48}$$

The formulas (47) and (48) look like as if ten dimensional super Yang–Mills theory is reduced to a point. For instance only the commutator is left in the non-abelian field strength

$$f_{\mu\nu} = i [A_\mu , A_\nu] \tag{49}$$

and there are no space-time derivatives. However, the action (46) coincides with the one of 10D super Yang–Mills dimensionally reduced to zero dimensions only if $\beta = 0$ and n is fixed. This differs the IKKT matrix model from a pure D-instanton matrix model.

As was argued in [2], if large values of n and *smooth* matrices A_μ^{ab} and ψ_α^{ab} dominate in (45), one substitutes

$$[\cdot , \cdot] \Longrightarrow i \{ \cdot , \cdot \} \tag{50}$$

$$\operatorname{tr} \ldots \Longrightarrow \int d^2\sigma \sqrt{g} \ldots \tag{51}$$

similarly to what is discussed above for M(atrix) theory. Then the formulas (45) to (48) for the IKKT matrix models reproduce the ones (35) to (42) for the Schild formulation of IIB superstring.

This passage from the IKKT matrix model to the Schild formulation of IIB superstring can be formalized introducing the matrix function

$$L(\sigma_1, \sigma_2)^{ab} = \sum_{m_1, m_2} j_{m_1, m_2} (\sigma_1, \sigma_2) \, J_{m_1, m_2}^{ab} , \tag{52}$$

where J_{m_1, m_2}^{ab} form a basis for gl(∞) and $j_{m_1, m_2} (\sigma_1, \sigma_2)$ form a basis in the space of functions of σ_1 and σ_2. An explicit form of j's depends on the topology of the σ-space. Explicit formulas are available for a sphere and a torus.

With the aid of (52) we can relate matrices with functions of σ_1 and σ_2 by

$$A_\mu = \int d^2\sigma \sqrt{g}\, X_\mu\, L\,, \tag{53}$$

$$X_\mu = \mathrm{tr}\, A_\mu\, L\,. \tag{54}$$

These formulas result for smooth configurations in Eqs. (50) and (51). The word "smooth" means that configurations can be reduced by a gauge transformation to the form where high modes are not essential in the expansions (53) or (54).

The commutators of J's coincide with the Poisson brackets of j's as $n \to \infty$. This demonstrates the equivalence between the group of symplectic diffeomorphisms and the gauge group SU(∞) for smooth configurations.

D-Strings as Classical Solutions

The classical equations of motion for the Schild action (35) read

$$\{X^\mu, \{X_\mu, X_\nu\}\} = 0\,, \qquad \left\{X^\mu, (\gamma_\mu \Psi)_\alpha\right\} = 0\,. \tag{55}$$

Their matrix model counterparts are

$$[A^\mu, [A_\mu, A_\nu]] = 0\,, \qquad [A^\mu, (\gamma_\mu \psi)_\alpha] = 0\,, \tag{56}$$

which are to be solved for $n \times n$ matrices A_μ at infinite n.

Since Eqs. (56) look like Eq. (29) for M(atrix) theory, they possess *operator-like* solutions of the form (30), which are now associated with D-strings[2]. The solution associated with static D-string along 1st axis reads

$$A_\mu^{\mathrm{cl}} = \left(\frac{T}{2\pi} q, \frac{L}{2\pi} p, 0, \ldots, 0\right), \qquad \psi_\alpha^{\mathrm{cl}} = 0\,, \tag{57}$$

where the (infinite) $n \times n$ matrices p and q obey the canonical commutation relation (26), while $T/2\pi$ and $L/2\pi$ are (large enough) compactification radii.

The arguments in favor of identification of the classical solution (57) with static D-string are

- It is one dimension less than D-membrane of [1];
- Interaction between the two D-strings is reproduced at large distances[2];
- It is a BPS state (a proper central charge of SUSY algebra exists [18,19]);
- It can be extended [19, 20] to $p = 3, 5$.

Zoo of Dp-Branes

A solution associated with two D-strings has a block-diagonal form and is built out of the ones given by Eq. (57) for single D-strings.

The solution for two *parallel* static D-strings separated by the distance b along 2nd axis reads[2]

$$A_0^{\mathrm{cl}} = \begin{pmatrix} Q & 0 \\ 0 & Q \end{pmatrix}, \quad A_1^{\mathrm{cl}} = \begin{pmatrix} P & 0 \\ 0 & P \end{pmatrix}, \quad A_2^{\mathrm{cl}} = \begin{pmatrix} b/2 & 0 \\ 0 & -b/2 \end{pmatrix},$$
$$A_3^{\mathrm{cl}} = \ldots = A_9^{\mathrm{cl}} = 0\,, \tag{58}$$

where we have denoted

$$Q \equiv \frac{T}{2\pi} q\,, \qquad P \equiv \frac{L}{2\pi} p\,. \tag{59}$$

The solution associated with two *anti-parallel* static D-strings separated by the distance b along 2nd axis is

$$A_0^{\rm cl} = \begin{pmatrix} Q & 0 \\ 0 & Q \end{pmatrix}, \quad A_1^{\rm cl} = \begin{pmatrix} P & 0 \\ 0 & -P \end{pmatrix}, \quad A_2^{\rm cl} = \begin{pmatrix} b/2 & 0 \\ 0 & -b/2 \end{pmatrix},$$
$$A_3^{\rm cl} = \ldots = A_9^{\rm cl} = 0\,. \tag{60}$$

The solution associated with two static D-strings *rotated* through the angle θ in the 1,2 plane and separated by the distance b along 3rd axis is

$$A_0^{\rm cl} = \begin{pmatrix} Q & 0 \\ 0 & Q \end{pmatrix}, \quad A_1^{\rm cl} = \begin{pmatrix} P & 0 \\ 0 & P\cos\theta \end{pmatrix}, \quad A_2^{\rm cl} = \begin{pmatrix} P & 0 \\ 0 & P\sin\theta \end{pmatrix},$$
$$A_3^{\rm cl} = \begin{pmatrix} b/2 & 0 \\ 0 & -b/2 \end{pmatrix}, \quad A_4^{\rm cl} = \ldots = A_9^{\rm cl} = 0\,. \tag{61}$$

The solution associated with one Dp-brane, which extends (57) to $p > 1$, is given by

$$A_\mu^{\rm cl} = \left(P_1, Q_1, \ldots, P_{\frac{p+1}{2}}, Q_{\frac{p+1}{2}}, 0, \ldots, 0\right), \tag{62}$$

where P's and Q's form $(p+1)/2$ pairs of operators (infinite matrices) as in Eq. (59) obeying canonical commutation relation on a torus associated with compactification (of large enough radii $L_a/2\pi$) along the axes $0, \ldots, p$, so that

$$\omega_k = \frac{L_{2k-2} L_{2k-1}}{2\pi n^{\frac{2}{p+1}}} \qquad \left(k = 1, \ldots, \frac{p+1}{2}\right) \tag{63}$$

is fixed as $n \to \infty$. This is because of the fact that the full Hilbert space of the dimension n is represented [18] as the tensor product of $(p+1)/2$ Hilbert spaces of the dimension $n^{\frac{2}{p+1}}$ each. The value of n is related to the $p+1$ dimensional volume

$$V_{p+1} \equiv L_0 L_1 \cdots L_p \tag{64}$$

of the p-brane by

$$n = V_{p+1} \prod_{i=1}^{\frac{p+1}{2}} (2\pi\omega_i)^{-1}\,. \tag{65}$$

These formulas allows one to extract world-volume characteristics of Dp-branes from the matrix model.

A general multi-brane solution has a block-diagonal form and is built out of single p-brane solutions (62) quite similar to (58)–(61).

One-Loop Effective Action

The calculation of the one-loop effective action in the IKKT matrix model at fixed n can be performed for an arbitrary background, $A_\mu^{\rm cl}$ and $\psi_\alpha^{\rm cl} = 0$, obeying the classical equations of motion (56). The calculation is quite similar to the one in the Eguchi–Kawai reduced model.

Expanding around the classical solution

$$A_\mu = A_\mu^{\rm cl} + a_\mu \tag{66}$$

and adding the gauge fixing and ghost terms to the action (46):

$$S_{\text{g.f.}} = -\operatorname{tr}\left(\frac{1}{2}\left[A_\mu^{\text{cl}}, a_\mu\right]^2 + \left[A_\mu^{\text{cl}}, b\right]\left[A_\mu^{\text{cl}}, c\right]\right), \tag{67}$$

where the matrices b and c represent ghosts, we get[2]

$$W = \frac{1}{2}\operatorname{Tr}\ln(P^2\delta_{\mu\nu} - 2iF_{\mu\nu}) - \frac{1}{4}\operatorname{Tr}\ln\left(\left(P^2 + \frac{i}{2}F_{\mu\nu}\gamma^{\mu\nu}\right)\left(\frac{1+\gamma_{11}}{2}\right)\right) - \operatorname{Tr}\ln P^2 . \tag{68}$$

Here the *adjoint* operators P_μ and $F_{\mu\nu}$ are defined on the space of matrices by

$$P_\mu = \left[A_\mu^{\text{cl}}, \cdot\right], \qquad F_{\mu\nu} = \left[f_{\mu\nu}^{\text{cl}}, \cdot\right] = i\left[\left[A_\mu^{\text{cl}}, A_\nu^{\text{cl}}\right], \cdot\right]. \tag{69}$$

For the solution (62), Im W vanishes for $p = 1, 3, 5, 7$ since $P_\mu = 0$ at least in one direction.

The first term on the right hand side of Eq. (68) comes from the quantum fluctuations of A_μ, the second and third terms which come from fermions and ghosts have the minus sign for this reason. The extra factor 1/2 in the first and second terms is because the matrices A and ψ are hermitian.

If A_μ^{cl} is diagonal

$$A_\mu^{\text{cl}} = \operatorname{diag}\left(p_\mu^{(1)}, \ldots, p_\mu^{(n)}\right), \qquad \psi_\alpha^{\text{cl}} = 0, \tag{70}$$

which is a solution of Eq. (56) associated with the flat space-time, then $F_{\mu\nu} = 0$ and

$$W = \left(\frac{1}{2}\cdot 10 - \frac{1}{4}\cdot 16 - 1\right)\operatorname{Tr}\ln P^2 = 0. \tag{71}$$

The plane vacuum is a BPS state.

The same is true (to all loops) for any A_μ^{cl} whose commutator is diagonal:

$$\left[A_\mu^{\text{cl}}, A_\nu^{\text{cl}}\right] = c_{\mu\nu}\mathbf{1}_n, \tag{72}$$

where $c_{\mu\nu}$ are c-numbers rather than matrices. Such solutions preserve[2, 18] a half of SUSY and are BPS states. The solution (58) associated with parallel D-strings is an example of such a BPS state.

For a general background A_μ^{cl}, the matrix $F_{\mu\nu}$ can always be represented in the canonical (Jordan) form

$$F_{\mu\nu} = \begin{pmatrix} 0 & -\omega_1 & & & & \\ \omega_1 & 0 & & & & \\ & & \ddots & & & \\ & & & 0 & -\omega_5 \\ & & & \omega_5 & 0 \end{pmatrix}, \tag{73}$$

so that

$$\operatorname{Tr}\ln(P^2\delta_{\mu\nu} - 2iF_{\mu\nu}) = \sum_{i=1}^{5}\operatorname{Tr}\ln((P^2)^2 - 4\omega_i^2) \tag{74}$$

and

$$\operatorname{Tr}\ln\left(\left(P^2 + \frac{i}{2}F_{\mu\nu}\gamma^{\mu\nu}\right)\left(\frac{1+\gamma_{11}}{2}\right)\right) = \sum_{\substack{s_1,\ldots,s_5=\pm 1 \\ s_1\ldots s_5=1}} \operatorname{Tr}\ln\left(P^2 - \sum_i \omega_i s_i\right) \tag{75}$$

There are 16 terms on the right hand side of Eq. (75) representing the trace over γ-matrices. Equations (74) and (75) are most useful in practical calculations for the background of Dp-brane given by (62), when only $(p+1)/2$ of 5 omegas are nonvanishing.

Brane-Brane Interaction

The interaction between two Dp-branes is calculated by substituting the proper classical solutions into (68) and using Eqs. (74), (75).

For parallel Dp-branes, $W = 0$ is accordance with the general arguments of the previous subsection.

For anti-parallel Dp-branes, we get ([2] for $p = 1$, [19] for $p \geq 3$)

$$W = -2n \int_0^\infty \frac{ds}{s} e^{-b^2 s} \left[\sum_{i=1}^{\frac{p+1}{2}} \left(\cosh 4\omega_i s - 1 \right) - 4 \left(\prod_{i=1}^{\frac{p+1}{2}} \cosh 2\omega_i s - 1 \right) \right] \prod_{i=1}^{\frac{p+1}{2}} \frac{1}{2 \sinh 2\omega_i s}. \tag{76}$$

The asymptotics of this formula at large b:

$$W = -\frac{1}{16} n \Gamma\left(\frac{7-p}{2}\right) \left[2 \sum_{i=1}^{\frac{p+1}{2}} \omega_i^4 - \left(\sum_{i=1}^{\frac{p+1}{2}} \omega_i^2 \right)^2 \right] \prod_{i=1}^{\frac{p+1}{2}} \omega_i^{-1} \left(\frac{2}{b}\right)^{7-p} + O\left(\frac{1}{b^{9-p}}\right), \tag{77}$$

agrees with the superstring calculation at large distances.

However, the superstring result [10, 21, 22] for the interaction between two anti-parallel Dp-branes at arbitrary distances b, which is given by the annulus diagram in the open-string language or by the cylinder diagram in the closed-string language,

$$W = -V_{p+1} \int_0^\infty \frac{dt}{t} \frac{1}{(8\pi^2 \alpha' t)^{\frac{p+1}{2}}} e^{-b^2 t / 2\pi\alpha'} q^{-1} \frac{\prod_{n=1}^\infty (1 - q^{2n-1})^8}{\prod_{n=1}^\infty (1 - q^{2n})^8} \tag{78}$$

with $q = e^{-\pi t}$, does *not* coincide with the matrix-model result (76). There is no agreement even if one truncates to the lightest open string modes.

A way out could be to interpret [23] the classical solutions in the IKKT matrix model as D-branes with magnetic field, in analogy with previous work [16] on M(atrix) theory[1]. An alternative possibility is to modify the IKKT matrix model to better reproduce the superstring calculation.

THE NBI MATRIX MODEL[3]

Calculations of the brane-brane interaction in the matrix model can be extended to the case of moving and rotated static Dp-branes. The results agree with the superstring calculations for empty branes only at large distances between them. This was one of the motivations of [3] to modify the IKKT matrix model introducing (instead of n) an additional dynamical variable — a positive definite hermitian matrix Y^{ab} — which is the direct analog of $\sqrt{g}$ in the Schild formulation of IIB superstring. Integration over Y^{ab} results in the Non-abelian Born–Infeld (NBI) action which reproduces the Nambu–Goto version of the Green–Schwarz action of IIB superstring.

Parallel Moving Branes

The operator-like solution to Eqs. (56), which is associated with two parallel branes separated by the distance b along the $(p+2)$-th axis and *moving* with velocities v and $-v$ along the $(p+1)$-th axis, can be obtained by boosting the one for parallel branes

(see (58)) along the (p+1)-th axis:

$$
\begin{aligned}
A_0^{\rm cl} &= \begin{pmatrix} B_0\cosh\epsilon & 0 \\ 0 & B_0\cosh\epsilon \end{pmatrix}, \\
A_a^{\rm cl} &= \begin{pmatrix} B_a & 0 \\ 0 & B_a \end{pmatrix}, \quad a = 1,\dots p, \\
A_{p+1}^{\rm cl} &= \begin{pmatrix} B_0\sinh\epsilon & 0 \\ 0 & -B_0\sinh\epsilon \end{pmatrix}, \\
A_{p+2}^{\rm cl} &= \begin{pmatrix} \frac{b}{2} & 0 \\ 0 & -\frac{b}{2} \end{pmatrix}, \\
A_i^{\rm cl} &= 0, \quad i = p+3,\dots 9 .
\end{aligned}
\tag{79}
$$

Here

$$
v = \tanh\epsilon , \tag{80}
$$

and we have denoted

$$
B_0 \equiv Q_1 , \quad B_1 \equiv P_1 , \quad \dots \quad B_{p-1} \equiv Q_{\frac{p+1}{2}} , \quad B_p \equiv P_{\frac{p+1}{2}} . \tag{81}
$$

The substitution of (79) into the one-loop (euclidean) effective action (68) yields[3]

$$
W = -n^{\frac{2p}{p+1}} \prod_{a\neq 1} L_a^{-1} \int_0^\infty \frac{ds}{s} \left(\frac{\pi}{s}\right)^{\frac{p}{2}} {\rm e}^{-b^2 s} \frac{(\cosh(4\omega_1 s\sinh\epsilon) - 4\cosh(2\omega_1 s\sinh\epsilon) + 3)}{\cosh\epsilon\ \sinh(2\omega_1 s\sinh\epsilon)} . \tag{82}
$$

Using Eqs. (63), (64) and Wick rotating back to Minkowski space-time, we get for the phase shift

$$
\delta = -\frac{V_p}{(2\pi)^p}\omega_1 \prod_{i=1}^{l} \frac{1}{\omega_i^2} \int_0^\infty \frac{ds}{s} \left(\frac{\pi}{s}\right)^{\frac{p}{2}} {\rm e}^{-b^2 s} \frac{(\cos(4\omega_1 s\sinh\epsilon) - 4\cos(2\omega_1 s\sinh\epsilon) + 3)}{\cosh\epsilon\ \sin(2\omega_1 s\sinh\epsilon)} \tag{83}
$$

where

$$
V_p = \prod_{a=1}^{p} L_a . \tag{84}
$$

This result was shown[3] to agree with the superstring calculation of Bachas[24] at large b for the real part of δ. Analogously, the imaginary part of (83) which comes from the poles at zeros of the denominator agrees at small v providing $\omega_i = 2\pi\alpha'$.

Rotated Branes

Taking the configuration of two parallel Dp-branes separated by the distance b along the (p+2)-th axis and rotating them in the opposite directions in the (p, p+1) plane through the angle $\theta/2$, one obtains the following solution to Eq. (56)

$$
\begin{aligned}
A_a^{\rm cl} &= \begin{pmatrix} B_a & 0 \\ 0 & B_a \end{pmatrix}, \quad a = 0,\dots,p-1, \\
A_p^{\rm cl} &= \begin{pmatrix} B_p\cos\frac{\theta}{2} & 0 \\ 0 & B_p\cos\frac{\theta}{2} \end{pmatrix}, \\
A_{p+1}^{\rm cl} &= \begin{pmatrix} B_p\sin\frac{\theta}{2} & 0 \\ 0 & -B_p\sin\frac{\theta}{2} \end{pmatrix}, \\
A_{p+2}^{\rm cl} &= \begin{pmatrix} \frac{b}{2} & 0 \\ 0 & -\frac{b}{2} \end{pmatrix}, \\
A_i^{\rm cl} &= 0, \quad i = p+3,\dots,9 ,
\end{aligned}
\tag{85}
$$

which extends (61) to $p > 1$. This looks pretty much like an analytic continuation of Eq. (79) ($\epsilon \to i\theta/2$).

The interaction between two rotated Dp-branes is given by[3]

$$W = -4n^{\frac{2p}{p+1}} \frac{1}{\cos\frac{\theta}{2}} \prod_{a\neq p-1} L_a^{-1} \times \int_0^\infty \frac{ds}{s} \left(\frac{\pi}{s}\right)^{\frac{p}{2}} e^{-b^2 s} \tanh\left(\omega_{\frac{p+1}{2}} s \sin\frac{\theta}{2}\right) \sinh^2\left(\omega_{\frac{p+1}{2}} s \sin\frac{\theta}{2}\right). \quad (86)$$

It can be obtained from (82) substituting $\epsilon = i\theta/2$.

Expanding in $1/b^2$ and using Eq. (63), one gets

$$W = -\Gamma\left(\frac{6-p}{2}\right) \frac{4V_p}{(4\pi)^{\frac{p}{2}}} \omega_{\frac{p+1}{2}}^4 \prod_i \frac{1}{\omega_i^2} \frac{\sin^3\frac{\theta}{2}}{\cos\frac{\theta}{2}} \frac{1}{b^{6-p}} + O\left(\frac{1}{b^{8-p}}\right) \quad (87)$$

for large distances, which agrees with the supergravity result. For $p = 1$ this is first shown in [2].

The NBI Action

In the IKKT model the matrix size n is considered as a dynamical variable, so the partition function (45) includes the summation over n. This sum is expected to recover the integration over $\sqrt{g}$ in (41) while the proof is missing. Even at the classical level, the minimization of Eq. (46) with respect to n does not result in a nice matrix-model action which could be associated with the Nambu–Goto action (39).

These problems can be easily resolved by a slight modification of the IKKT matrix model. Let us introduce a positive definite $N \times N$ hermitian matrix Y^{ab} which would play the role of a dynamical variable instead of n. In other words, the matrix size N is fixed (to be distinguished from fluctuating n) while the elements of Y *fluctuate.*

The classical action has the form

$$S^{\rm cl} = -\alpha\left(\frac{1}{4}\,{\rm tr}\, Y^{-1}[A_\mu, A_\nu]^2 + \frac{1}{2}\,{\rm tr}\,(\bar\psi\gamma^\mu[A_\mu, \psi])\right) + \beta\,{\rm tr}\, Y\,, \quad (88)$$

which yields the following classical equation of motion for the Y-field:

$$\frac{\alpha}{4}\left(Y^{-1}[A_\mu, A_\nu]^2 Y^{-1}\right)_{ij} + \beta\delta_{ij} = 0\,. \quad (89)$$

The solution to Eq. (89) reads

$$Y = \frac{1}{2}\sqrt{\frac{\alpha}{\beta}}\sqrt{-[A_\mu, A_\nu]^2}\,. \quad (90)$$

Here $-[A_\mu\,, A_\nu\,]^2$ is positive definite, since the commutator is anti-hermitian (cf. Eq. (49)). The square root in (90) is unique, provided Y is positive definite which is the case. After the substitution of (90), the classical action (88) reduces to

$$S^{\rm cl}_{\rm NBI} = \sqrt{\alpha\beta}\,{\rm tr}\,\sqrt{-[A_\mu, A_\nu]^2} - \frac{\alpha}{2}\,{\rm tr}\,(\bar\psi\gamma^\mu[A_\mu, \psi])\,. \quad (91)$$

The bosonic part of (91) coincides with the strong field limit of the Non-abelian Born–Infeld (NBI) action. The action (91) is called for this reason the NBI action. Notice that it is *field-theoretic* rather than widely discussed stringy NBI action which has a different structure[23].

The formulas above in this subsection are very similar to the ones above for the Schild formulation. Thus the hermitian matrix Y^{ab} with positive definite eigenvalues is the direct analog of $\sqrt{g(\sigma_1, \sigma_2)}$ so that

$$Y^{ab} \Longrightarrow \sqrt{g(\sigma_1, \sigma_2)} \tag{92}$$

in the same sense as in (43), (44). In the next subsection we discuss that it is possible to choose such a measure of integration over Y which reproduces the Nambu–Goto version of the Green–Schwarz action even at the quantum level.

The NBI Model of IIB Superstring

The NBI matrix model is defined by the action

$$S_{\mathrm{NBI}} = -\alpha \left(\frac{1}{4} \operatorname{tr} Y^{-1} [A_\mu, A_\nu]^2 + \frac{1}{2} \operatorname{tr} \left(\bar{\psi} \gamma^\mu [A_\mu, \psi] \right) \right) + V(Y) , \tag{93}$$

where Y is a hermitian $N \times N$ matrix with *positive* eigenvalues. The potential is

$$V(Y) = \beta \operatorname{Tr} Y = \gamma \operatorname{Tr} \ln Y, \tag{94}$$

where

$$\gamma = N - \frac{1}{2} . \tag{95}$$

The partition function is then given by the matrix integral[3]

$$Z_{\mathrm{NBI}} = \int dA_\mu \, d\psi_\alpha \, dY \, \mathrm{e}^{-S_{\mathrm{NBI}}} . \tag{96}$$

The action (93) is invariant under the SUSY transformation

$$\begin{aligned} \delta_{\mathrm{SUSY}} \psi &= \frac{i}{4} \left[Y^{-1}, [A_\mu, A_\nu] \right]_+ \gamma^{\mu\nu} \epsilon + \xi \\ \delta_{\mathrm{SUSY}} A_\mu &= i \bar{\epsilon} \gamma_\mu \psi, \end{aligned} \tag{97}$$

in the limit $N \to \infty$, where $[\cdot, \cdot]_+$ stands for the anticommutator. Y is not changed under this transformation.

The action (93) differs from its classical counterpart (88) by the second term on the right-hand side of Eq. (94). It is associated with the measure of integration over Y rather than with the classical action. The classical action (88) can be obtained from (93) in the limit $\alpha \sim \beta \to \infty$, $\alpha / \beta \sim 1 \sim \gamma / N$. This corresponds to the usual classical limit in string theory since[2] $\alpha \sim \beta \sim g_s^{-1}$.

The matrix Y can be always brought to the diagonal form

$$Y = \Omega^\dagger \operatorname{diag} (y_1, \ldots, y_N) \, \Omega \qquad (y_1, \ldots, y_N \geq 0) , \tag{98}$$

where Ω is unitary. The measure for integration over Y reads explicitly

$$\int dY \ \ldots = \int_0^\infty \prod_{i=1}^N dy_i \, \Delta^2[y] \cdot d\Omega \ \ldots \tag{99}$$

with

$$\Delta[y] = \prod_{i>j} (y_i - y_j) \tag{100}$$

being the Vandermonde determinant.

The integral over Y in (96) can be done. Let us mention that the fermionic term in (93) is Y-independent and denote

$$\mathcal{F}(z) = \int dY \, \mathrm{e}^{-\alpha \, \mathrm{tr}\, Y^{-1} z^2/4 - \beta \, \mathrm{tr}\, Y - \gamma \, \mathrm{tr} \ln Y}, \tag{101}$$

where $z^2 = -[A_\mu \, , \, A_\nu]^2$. This matrix integral looks like an external field problem for the Penner matrix model.

Doing the Itzykson–Zuber integral over the "angular" variable Ω, (101) takes the form

$$\mathcal{F}(z) \propto \int_0^\infty \prod_{i=1}^N dy_i \, \frac{\Delta^2[y]}{\Delta[1/y]\Delta[z^2]} \mathrm{e}^{-\alpha \sum_i y_i^{-1} z_i^2/4 - \beta \sum_i y_i - \gamma \sum_i \ln y_i} \propto \frac{\Delta[z]}{\Delta[z^2]} \mathrm{e}^{-\sqrt{\alpha\beta} \sum_i z_i}, \tag{102}$$

where z_i^2 stands for the eigenvalues of z^2.

Hence, it is shown that

$$\int dA_\mu \, d\psi_\alpha \, dY \, \mathrm{e}^{-S_{\mathrm{NBI}}} = \int \frac{dA_\mu \, d\psi_\alpha}{\prod_{i>j}(z_i + z_j)} \mathrm{e}^{-S^{\mathrm{cl}}_{\mathrm{NBI}}}. \tag{103}$$

Thus the NBI action $S^{\mathrm{cl}}_{\mathrm{NBI}}$ defined by Eq. (91) is reproduced modulo the change of the measure for integration over A_μ.

The significance of this result is that it can be explicitly shown that

$$S^{\mathrm{cl}}_{\mathrm{NBI}} \Longrightarrow S_{\mathrm{NG}}$$

given by Eq. (39), where the arrow is in the same sense as in (43), (44) and (92). Analogously, the Schild action (35) can be reproduced from the model (96) with the additional integration over Y (without explicitly doing it).

A proposal of [3] is to modify the measure for the integration over A_μ from the outset to get

$$\int dA_\mu \, d\psi_\alpha \, dY \prod_{i>j}(z_i + z_j) \, \mathrm{e}^{-S_{\mathrm{NBI}}} = \int dA_\mu \, d\psi_\alpha \mathrm{e}^{-S^{\mathrm{cl}}_{\mathrm{NBI}}}$$
$$\overset{N=\infty}{=} \int DX^\mu \, D\Psi_\alpha \mathrm{e}^{-S_{\mathrm{NG}}}. \tag{104}$$

Then the Nambu–Goto version of the Green–Schwartz action of IIB superstring is exactly reproduced by the NBI matrix model.

Remark on D-Brane Solutions in the NBI Model

The classical solutions (62) associated with D-brane configurations are also classical solutions to the NBI matrix model whose classical equations of motion, which result from the variation of the action (91) with respect to A_μ and ψ_α, read

$$\left[A^\mu, \left[Y^{-1}, [A_\mu, A_\nu]\right]_+\right] = 0 \, , \qquad [A_\mu \, , (\gamma^\mu \psi)_\alpha] = 0 \, . \tag{105}$$

The reason is that these classical solutions are BPS states and the commutator $[A_\mu, A_\nu]$ is proportional to the unit matrix (see Eq. (72)).

A more general property holds in the large–N limit when any classical solution of the IKKT matrix model is simultaneously a solution of the classical equations of motion of the NBI model. However, the structure of the classical equations (89) and (105) in the NBI matrix model is, generally speaking, richer that Eq. (56) in the IKKT

model, since Y^{cl} may have some nontrivial distribution of eigenvalues (typical for the large–N saddle points).

One of most urgent checks of the NBI model would be to perform the calculation of the brane-brane interaction to compare with the superstring result. This calculation will take into account the fact that Y is a dynamical field while the ones described above for the IKKT matrix model are done at fixed n, *i.e.* without considering n as a dynamical variable.

Conclusion

It is now too early to make any definite conclusions since it is not yet clear whether or not this formulation of superstrings, which is based on the supersymmetric matrix models, would survive. Nevertheless, such an approach to M theory looks most promising among those proposed so far.

This situation reminds me somewhat of the one with QCD in the very beginning of the seventies about the times when the QCD Lagrangian was introduced. Before that there existed the approach to the theory of strong interaction based on strings and dual resonance models, while the new theory looked quite different and was most convenient to study strong interaction at small distances. Once again, it is now too early to predict whether the same could happen with superstrings in the nearest future, but this option should not be immediately excluded.

One of the simplest checks of the matrix models of superstrings is the study of the interaction between D-branes. It should answer, in particular, the question whether the classical operator-like solutions of the matrix models are associated with empty D-branes or D-branes carrying magnetic field.

A more serious problem is to show how string perturbation theory emerges from the matrix models. The NBI matrix model is very promising from this point of view since it reproduces the Nambu–Goto version of the Green–Schwartz action.

While the proposed matrix models of IIB superstring are of the type of reduced ten-dimensional super Yang–Mills, they have additional degrees of freedom which are essential to have strings. This differs the situation from the one in large–N QCD where the fundamental Lagrangian is fixed, and the problem to obtain strings in the Eguchi–Kawai reduced model is almost as difficult as in whole QCD. Now, for the matrix models of superstrings, the true model is not know from the outset. The reader is still free to introduce his/her own model to describe superstrings in the best way

Acknowledgments

I thank I. Chepelev, A. Fayyazuddin, P. Olesen, D. Smith and K. Zarembo for the pleasant collaboration and illuminating discussions. This work was supported in part by the grants INTAS 94–0840, CRDF 96–RP1–253 and RFFI 97–02–17927.

REFERENCES

1. T. Banks, W. Fischler, S.H. Shenker and L. Susskind, M theory as a matrix model: a conjecture, *Phys. Rev.* **D 5 5**:5112 (1997), hep-th/9610043.
2. N. Ishibashi, H. Kawai, Y. Kitazawa and A. Tsuchiya, A large–N reduced model as superstring, *Nucl. Phys.* **B498**:467 (1997), hep-th/9612115.
3. A. Fayyazuddin, Y. Makeenko, P. Olesen, D.J. Smith and K. Zarembo, Towards a non-perturbative formulation of IIB superstring by matrix models, *Nucl. Phys.* **B499**:159 (1997), hep-th/9703038.

4. V.A. Kazakov, Bilocal regularization of models of random surfaces, *Phys. Lett.* **150B:** 282 (1985).
5. F. David, Planar diagrams, two-dimensional lattice gravity and surface models, *Nucl. Phys.* **B257**[FS14]:45 (1985).
6. J. Ambjørn, B. Durhuus and J. Frölich, Diseases of triangulated random surfaces, and possible cures, *Nucl. Phys.* **B257** [FS14]:433 (1985).
7. Y.M. Makeenko, Discretizing Superstring, Talk at QUARKS'96, Yaroslavl May 5–11, 1996, ITEP–TH–42/96 (September, 1996).
8. A. Miković and W. Siegel, Random superstrings, *Phys. Lett.* **240B:** 363 (1990).
9. L. Alvarez-Gaume *et al.*, Superloop equations and two dimensional supergravity, *Int. J. Mod. Phys.* **A7:** 5337 (1992), hep-th/9112018.
10. J. Polchinski, Dirichlet branes and Ramond–Ramond charges, *Phys. Rev. Lett.* **75:** 4724 (1995), hep-th/9510017.
11. K.S. Stelle, Lectures on supergravity p-branes, hep-th/9701088.
12. E. Witten, Bound states of strings and p-branes, *Nucl. Phys.* **B460:** 335 (1995), hep-th/9510135.
13. B. de Wit, J. Hoppe and H. Nicolai, On the quantum mechanics of supermembranes, *Nucl. Phys.* **B305** [FS23]:545 (1988).
14. B. de Wit, M. Luscher and H. Nicolai, The supermembrane is unstable, *Nucl. Phys.* **B320:** 135 (1989).
15. O. Aharony and M. Berkooz, Membrane dynamics in M(atrix) theory, *Nucl. Phys.* **B491:** 184 (1997), hep-th/9611215.
16. G. Lifschytz and S.D. Mathur, Supersymmetry and membrane interactions in M(atrix) theory, hep-th/9612087.
17. G. Lifschytz, Four-brane and six-brane interactions in M(atrix) theory, hep-th/9612223.
18. T. Banks, N. Seiberg and S. Shenker, Branes from matrices, *Nucl. Phys.* **B490:** 91 (1997), hep-th/9612157.
19. I. Chepelev, Y. Makeenko and K. Zarembo, Properties of D-branes in matrix model of IIB superstring, *Phys. Lett.* **B400:** 43 (1997), hep-th/9701151.
20. A. Fayyazuddin and D.J. Smith, P-Brane solutions in IKKT IIB matrix theory, hep-th/9701168.
21. J. Polchinski, TASI lectures on D-branes, hep-th/9611050.
22. M.B. Green and M. Gutperle, Light-cone supersymmetry and D-branes, *Nucl. Phys.* **B476:** 484 (1996), hep-th/9604091.
23. A.A. Tseytlin, On non-Abelian generalization of Born–Infeld action in string theory, hep-th/9701125.
24. C. Bachas, D-brane dynamics, *Phys. Lett.* **B374:** 37 (1996), hep-th/9511043.

NEW DEVELOPMENTS IN THE CONTINUOUS RENORMALIZATION GROUP

Tim R. Morris

Department of Physics
University of Southampton
Highfield
Southampton, SO17 1BJ
U.K.

INTRODUCTION

Over the last several years, there has been a resurgence of interest in using non-perturbative approximation methods based on Wilson's continuous Renormalization Group (RG), in quantum field theory.[1] The reason is simple: on the one hand there is a desperate need for better analytic approximation methods to understand truly non-perturbative situations in quantum field theory (i.e. where there are no small parameters in which one can fruitfully expand). On the other hand, Wilson's framework offers many possibilities for systematic approximations which preserve a crucial defining property of a quantum field theory – namely the existence of a continuum limit.[2,3,4]

In this lecture, I review progress in the use and understanding of Wilson's continuous renormalization group[1] particularly in the past year. I try not to overlap too much with reviews given in RG96[4,5]. I concentrate on progress in the understanding of the structure of the continuous RG since this is of fundamental importance to all research in this area, and is an aspect that I have been particularly involved with, but I hope that the references collected at the end are a representative list of just last years research in this area.[6–26] These papers deal with – amongst, other issues – chiral symmetry breaking,[6,8] chiral anomalies,[9] finite temperature,[8,10,11,12] reparametrization invariance,[13,14] gauge invariance,[6,15] perturbation theory,[9,16,17] gravity and supergravity,[15] phase transitions,[8,11,13,14,18,20] novel continuum limits,[19] massive continuum limits,[20] local potential approximation,[21,22,24] large N limits,[21,22,23,24] C functions,[24,25] and with applications from condensed matter[18] to cosmology[26].

The basic idea behind the (continuous) Wilsonian RG, is to construct the partition function in two steps. Rather than integrate over all momentum modes q in one go, one first integrates out modes between a cutoff scale Λ_0 and a very much lower energy scale Λ. Both of these scales are introduced by hand. The remaining integral from Λ to zero may again be expressed as a partition function, but the bare action S_{Λ_0} (which is typically chosen to be as simple a functional as possible) is replaced by a very

New Developments in Quantum Field Theory
Edited by Damgaard and Jurkiewicz, Plenum Press, New York, 1998

complicated effective action S_Λ^{tot}, and the overall cutoff Λ_0 by the effective cutoff Λ Differential RG flow equations of the generic form

$$\Lambda\frac{\partial}{\partial\Lambda}S_\Lambda^{tot}[\varphi] = \mathcal{F}[S_\Lambda^{tot}] \quad , \tag{1}$$

can be written down that determine S_Λ^{tot} in such a way that the resulting effective partition function gives precisely the same results for all correlators as the original partition function.

It may seem at first sight that such a partial integration step merely complicates the issue. For example, we have had to replace the (generally) simple S_{Λ_0} by a complicated S_Λ^{tot}. However, for the most part the complicated nature of S^{tot} merely expresses the fact that quantum field theory itself is complicated: on setting $\Lambda = 0$, S_Λ^{tot} becomes equivalent to the generating function for all connected Green functions. To see this, note that the effective cutoff Λ can be regarded, either as an effective ultraviolet cutoff for the remaining modes q – as just described, or, from the point of view of the modes q that have already been integrated out, Λ behaves as an infrared cutoff. These intuitive statements can be formalised and proved[2].* We introduce Λ by modifying propagators $\sim 1/q^2$ to $\Delta_{UV} = C_{UV}(q,\Lambda)/q^2$, where C_{UV} is a profile that acts as an ultra-violet cutoff [2 9]; i.e. $C_{UV}(0,\Lambda) = 1$ and $C_{UV} \to 0$ (sufficiently fast) as $q \to \infty$. If we introduce the interaction part of the effective action as

$$S_\Lambda^{tot}[\varphi] = \frac{1}{2}\varphi\cdot\Delta_{UV}^{-1}\cdot\varphi + S_\Lambda[\varphi] \quad , \tag{2}$$

then Polchinski's form[29] of Wilson's RG[1] is,

$$\frac{\partial S_\Lambda}{\partial\Lambda} = \frac{1}{2}\frac{\delta S_\Lambda}{\delta\varphi}\cdot\frac{\partial\Delta_{UV}}{\partial\Lambda}\cdot\frac{\delta S_\Lambda}{\delta\varphi} - \frac{1}{2}\mathrm{tr}\frac{\partial\Delta_{UV}}{\partial\Lambda}\cdot\frac{\delta^2 S_\Lambda}{\delta\varphi\delta\varphi} \quad . \tag{3}$$

On the other hand, $\exp -S_\Lambda^{tot}$ is itself given by a partition function Z_Λ, in which the effective cutoff appears as an infrared cutoff as I have already mentioned.[2] Propagators $\sim 1/q^2$ are in this partition function replaced by $\Delta_{IR} = C_{IR}(q,\Lambda)/q^2$, where it can be shown that $C_{IR}(q,\Lambda) = 1 - C_{UV}(q,\Lambda)$. It is easy to see from the above properties of C_{UV}, that C_{IR} indeed behaves as an infrared cutoff.

¿From this partition function, Z_Λ, one can construct the Legendre effective action $\Gamma_\Lambda^{tot}[\varphi^c]$, and it is immediately clear that there must be a close relation between the two effective actions Γ^{tot} and S^{tot}. Indeed, if we write the interaction parts of Γ_Λ^{tot} as

$$\Gamma_\Lambda^{tot}[\varphi^c] = \frac{1}{2}\varphi^c\cdot\Delta_{IR}^{-1}\cdot\varphi^c + \Gamma_\Lambda[\varphi^c] \quad , \tag{4}$$

(where φ^c is the so-called 'classical' field) then it can be shown that the following Legendre transform type relation exists between the two effective actions[2]

$$S_\Lambda[\varphi] = \Gamma_\Lambda[\varphi^c] + \frac{1}{2}(\varphi^c - \varphi)\cdot\Delta_{IR}^{-1}\cdot(\varphi^c - \varphi) \quad . \tag{5}$$

This relationship is important, for it means that the $\Lambda \to 0$ limit of the Wilsonian effective action can be related to the standard Legendre effective action $\Gamma[\varphi^c] = \lim_{\Lambda\to 0}\Gamma_\Lambda[\varphi^c]$ and hence to Green functions, S matrices, classical effective potentials, and so forth. On the other hand, by relating the infrared cutoff Legendre effective action to the Wilsonian effective action, it provides physical justification for the existence of

*Somewhat similar statements have appeared elsewhere [27, 28]

fixed points and self-similar behaviour in the solutions for Γ_Λ. By differentiating Z_Λ, one readily obtains a flow equation for the infrared cutoff Legendre effective action: [2,30,28,34]

$$\frac{\partial\Gamma_\Lambda}{\partial\Lambda} = -\frac{1}{2}\mathrm{tr}\left[\frac{1}{\Delta_{IR}}\frac{\partial\Delta_{IR}}{\partial\Lambda}\cdot\left(1+\Delta_{IR}\cdot\frac{\delta^2\Gamma_\Lambda}{\delta\varphi^c\delta\varphi^c}\right)^{-1}\right] \quad . \tag{6}$$

It is straightforward to show that substitution of the Legendre transform relationship converts one flow equation into the other.

One should note also that Wilson's original form of the continuous RG[1] is related to Polchinski's by some simple changes of variables[30,4], and that the Wegner-Houghton sharp cutoff RG[31] is nothing but the sharp cutoff limit of Polchinski's equation[2,32]. Since these cover all present forms of the continuous RG, this might give the impression that no other forms are possible. This is not the case. But they are the simplest forms of continuous RG (related to the fact that the cutoff Λ may be placed entirely within the inverse propagator). Other more complex versions may eventually prove more useful for certain applications, e.g. gauge theory.

STRUCTURE OF THE CONTINUUM LIMIT – I

This RG method of describing quantum field theory becomes advantageous when we consider the continuum limit. I will indicate how one can solve the flow equations in this case, directly in the continuum, dispensing with the standard, but for quantum field theory, actually artificial and extraneous, scaffolding of imposing an overall cutoff Λ_0, finding a sufficiently general bare action $S_{\Lambda 0}$, and then tuning to a continuum limit as $\Lambda_0 \to \infty$. The solutions for the effective actions, being sensitive only to momenta of magnitude $\sim \Lambda$, may be expressed *directly* in terms of renormalized quantities. In this case one finds that the effective action may be expressed as a self-similar flow of the relevant and marginally-relevant couplings, say g^1 to g^n, about some fixed point:

$$S_\Lambda[\varphi] = S[\varphi]\left(g^1(\Lambda),\cdots,g^n(\varphi)\right) \quad . \tag{7}$$

Actually, we also require to change to renormalised fields of course,[20] but, for clarities sake we leave this implicit. Also, to see this self-similar behaviour, it is necessary to add the other essential ingredient of an RG step: scaling back the cutoff to the original size. Simpler and equivalent, is to ensure that all variables are 'measured' in units of Λ, i.e. we change variables to ones that are dimensionless, by dividing by Λ raised to the power of their scaling dimensions.[4] ¿From now on, we will assume that this has been done.

A fixed point $S_\Lambda[\varphi] = S_*[\varphi]$, i.e. such that $\partial S_*/\partial\Lambda = 0$, being thus completely scale free, corresponds to a massless continuum limit. We can arrange that the coupling constants by definition vanish at the corresponding fixed point

$$S[\varphi](0,\cdots,0) = S_*[\varphi] \quad , \tag{8}$$

and are conjugate to the eigen-perturbations (i.e. integrated operators of definite scaling dimension) at the fixed point:

$$\frac{\partial S}{\partial g^i}\Big|_{g=0} = \mathcal{O}_i[\varphi] \quad . \tag{9}$$

We can easily isolate a set of non-perturbative beta-functions[20] $\beta^i(g) \equiv \Lambda\partial g^i/\partial\Lambda$, whose perturbative expansion begins as $\beta^i(g) = \lambda_i g^i + O(g^2)$, once a definition (a.k.a. renormalization) condition consistent with (9) is chosen for the couplings. Here, the $\lambda_i \geq 0$

are the corresponding RG eigenvalues. This follows simply by comparing coefficients of the left and right hand sides of (1). Suppose for example, that $\mathcal{O}_4 = \varphi^4$, and a coupling g^4 is defined to be $c - c_*$ where $c(\Lambda)$ is the coefficient of the four-point vertex at zero external momentum, and c_* its fixed point value; then the zero-momentum part of the four-point vertex on the left hand side of (1) is nothing but β^4, and the corresponding term on the right hand side yields an expression for it in terms of S_Λ itself. In this way one readily converts the original Wilsonian RG into a self-similar flow for the underlying relevant couplings, in close analogy to the usual field theory perturbative RG,† although here the β functions are defined non-perturbatively. Finite massive continuum limits then follow providing that the couplings $g^i(\Lambda)$ are themselves finite at some finite physical scale $\Lambda \sim \mu$.[20]

Of course this translation to 'manifest' self-similar flow is just a rewriting, if the solution S_Λ is already known. On the other hand the solution S_Λ is determined completely (e.g. numerically) from (1), once the fixed point solution S_* and relevant (and marginally-relevant) eigenperturbations are determined, since these provide the complete boundary conditions via

$$S_\Lambda = S_* + \sum_i \alpha^i (\mu/\Lambda)^{\lambda_i} \mathcal{O}_i \qquad \text{as} \quad \Lambda \to \infty \quad , \tag{10}$$

where the α^i are the finite integration constants. (10) follows from (8), (9) and the β^i to first order in g. Note that this more subtle boundary condition is required because $S_\Lambda = S_*$ is a singular point for the continuous RG, regarded as a first-order-in-Λ differential equation. In Wilsonian terms this establishes the initial position and direction of the Renormalised Trajectory, which thus is sufficient to determine the entire trajectory. In this form, or even better in the self similar form of (8), (9) and the $\beta^i(g)$, we have dispensed entirely with the usual tuning procedure required to reach a continuum limit. (We still need to compute the infrared limit $\Lambda \to 0$, however this involves no tuning, and the asymptotic behaviour in this limit is straightforward to derive analytically since it corresponds to 'freeze-out' of finite dimensionful quantities on the scale of $\Lambda \ll \mu$.[20]) It is worth remarking also, that for a Gaussian fixed point these equations (including the β^i) are soluble analytically by iteration, directly in terms of renormalised perturbation theory.[2,24,16]

The remaining basic structural question then, is to understand how the (generically) finite number of fixed point solutions and relevant perturbations arise from equations such as (3) and (6), when these quantities are determined by functional differential equations (the right hand sides of (3) or (6), and their perturbations) with apparently a continuous infinity of solutions. Let us return to this after first considering some possible approximations. Then the so-called Local Potential Approximation will be used to illustrate the general solution to these questions.[20]

APPROXIMATIONS

Clearly, except in very special simple cases, these flow equations are not exactly soluble. However, virtually any approximation of the flow equations that preserves the fact that they are non-linear (in S_Λ) will continue to have fixed points and self-similar asymptotic solutions about the fixed point of the form (7), i.e. preserve renormalisability. This is in direct contrast to other frameworks for approximations such as the use of

† with the important difference however that physical Green functions are here obtained only when the scale used to define the beta function, tends to zero

Dyson-Schwinger equations.[2] The difficulty then, is 'only' one of finding approximations that are sufficiently reliable and accurate for the purpose at hand.

The simplest form of approximation is to truncate the effective action S_Λ so that it contains just a few operators. The Λ dependent coefficients of these operators then have flow equations determined by equating coefficients on the left and right hand side of (3), after rejecting from the right hand side of (3) all terms that do not 'fit' into this set (this being the approximation). The difficulty with this approximation is that it inevitably results‡ in a truncated expansion in powers of the field φ (about some point), which can only be sensible if the field φ does not fluctuate very much, i.e. is close to mean field.[4] This is precisely the opposite regime from the truly non-perturbative one that concerns us here. In most situations, it is difficult in practice to be systematic about the choice of which operators to include, while even if this is possible, in this regime one finds generically that higher orders cease to converge and thus yield limited accuracy, while there is also no reliability – even qualitatively – since many spurious fixed points are generated.[35]

One exception to this rule deserves comment. As is well known, fluctuations of $\rho = \varphi^a\varphi_a$ (when appropriately scaled) disappear in the limit $N \to \infty$. (Here φ^a is an N component scalar field in an $O(N)$ invariant theory for example. More general large N soluble systems also have this property.) In this limit, it can be shown that the flow equations for the (infinitely many) operators in Γ_Λ which can be written in such a way that they only involve ρ, form a closed set, so that truncation to all such operators involves no approximation.[23] Furthermore specialization to just the non-derivative operators (thus forming the effective potential) also involves no approximation.[31,21,22,23] In fact, the further truncation of these flow equations to a *finite* power expansion in ρ about this minimum, is *still* exact in this limit.[22] This gives some insight into why these latter truncations give good results also at finite N,[22,36,37,23] improving at high orders of truncation to as much as 8 digits accuracy, before succumbing to the generic pattern for finite truncations as outlined above.[7,35]

A less severe, and more natural, expansion, closely allied to the succesful truncations in real space RG of spin systems,[4] is rather to perform a 'short distance expansion'[4,32] of the effective action Γ_Λ, which for smooth cutoff profiles corresponds to a derivative expansion.[30] As well as evidently always allowing a systematic expansion (each level of approximation merely corresponding to discarding all terms with more than a specified number of derivatives), and robust and reasonably accurate approximations,[4] it also preserves enough of the continuous RG to address the structural question posed above§ – namely how (generically) discrete sets of fixed point solutions and eigenperturbations arise. This is because the second order functional differential equations that determine these, reduce under the derivative expansion to second order ordinary differential equations[30,38,4] which thus retain the property that they have a continuum of solutions.

The simplest such approximation is the so-called Local Potential Approximation (LPA), introduced by Nicoll, Chang and Stanley:[39]

$$S_\Lambda \sim \int d^D x \left\{ \frac{1}{2}(\partial_\mu \varphi)^2 + V(\varphi, \Lambda) \right\} \tag{11}$$

It has since been rediscovered by many authors,[32] notably Hasenfratz and Hasenfratz.[40] As a concrete example, consider the case of sharp cutoff. The flow equations may be

‡ unlike the case with truncations in the real space renormalization group of simple spin systems, where such a procedure has proved very succesful[4]

§ in contrast to the case of truncations to a finite set of operators

shown to reduce to [39,40, 2, 35, 32]

$$\frac{\partial}{\partial t}V(\varphi,t)+\frac{1}{2}(D-2)\,\varphi V'-DV=\ln(1+V'') \quad , \tag{12}$$

where $' \equiv \frac{\partial}{\partial\varphi}$ and $t = \ln(\mu/\Lambda)$. Actually, the analogous $N = \infty$ case of (12) was already derived by Wegner and Houghton in their paper introducing the sharp-cutoff flow equation.[31] As already pointed out above, in this limit the LPA is effectively exact.[23] It can be shown[23,20,2] that V coincides with the Legendre effective potential in the limit $\Lambda \to 0$, i.e. $t \to \infty$.

STRUCTURE OF THE CONTINUUM LIMIT – II

Thus in this example, the fixed point potential satisfies

$$\frac{1}{2}(D-2)\,\varphi V_*'(\varphi)-DV_*(\varphi)=\ln(1+V_*'') \quad . \tag{13}$$

This equation has indeed a continuum of solutions, in fact a continuous two-parameter set. However generically, all but a countable number of these solutions are singular![35,41] ($D = 2$ dimensions is an exception.[38]) It is the discrete set of non-singular solutions that are approximations to the exact fixed points. A review was given in RG96.[4,41] Since this scenario generalises to higher orders of the derivative expansion (with an increasingly larger dimension parameter space of solutions),[30,4] it seems reasonable to suppose that it applies also to the exact RG equations (e.g. 3,6) in the sense that, although the second order non-linear functional differential equations governing the fixed points have a full functional space worth of solutions, only typically a discrete number are well defined for all $\varphi(\mathbf{x})$.[20]

As reviewed in RG96,[4] this structure of the derivative expansion equations, and (13), allows straightforwardly for systematic searches of all possible continuum limits within this approximation. In the traditional approaches, this would require systematic searches of the infinite dimensional space of all possible bare actions!

For large field φ the only consistent behaviour (with $D > 2$) for the fixed point potential in (13) is

$$V_*(\varphi) \sim A\varphi^{D/d} \quad , \tag{14}$$

where $d = \frac{1}{2}(D–2)$, and A is a constant determined by the equations. This simply solves the left hand side of (13), these terms arising from purely dimensional considerations, and neglects the right hand side of the flow equation – which encodes the quantum corrections. Or in other words, (14) is precisely what would be expected by dimensions (since V's mass-dimension is D and φ's is d) providing only that any dependence on Λ, and thus the remaining quantum corrections, can be neglected. Requiring the form (14) to hold for both $\varphi \to \infty$ and $\varphi \to -\infty$, provides the necessary two boundary conditions for the second order ordinary differential equation (13), so we should indeed generally expect at most a discrete set of globally non-singular solutions. These considerations generalise to any order of the derivative expansion, and indeed we expect them to hold also for the exact RG, with one modification: beyond LPA, $d = \frac{1}{2}(D - 2 + \eta)$, where η is the anomalous dimension at the fixed point.[20]

(η is set artificially to zero by the LPA because in the LPA the momentum dependent terms in (11) remain uncorrected. There are subtleties to do with reparametrization invariance, if higher orders in the derivative expansion are to determine a discrete set of solutions for η, as properly to be expected. We will not review them again

here. The preservation of such an invariance under the derivative expansion, requires a particular power-law form of cutoff profile.[4,3,30] There has been recent progress in understanding the ramifications of reparametrization invariance for derivative expansions of the Polchinski equation.[14])

Now consider the determination of the eigenoperators. For this we set (by separation of variables),

$$V(\varphi,t) = V_*(\varphi) + \alpha e^{\lambda t} v(\varphi) \quad , \tag{15}$$

and expand to first order in α (c.f. (10)), obtaining

$$\lambda v + d\,\varphi v' - Dv = \frac{v''}{1+V_*''} \quad . \tag{16}$$

In this case we again have a one parameter continuum of solutions (after some choice of normalisation), which are guaranteed globally well defined since (16) is linear, and this for each value of λ. How can this be squared with the fact that experiment, simulation etc., typically only uncover a discrete spectrum of such operators? The answer is that only the discrete set of solutions for $v(\varphi)$ that behave as a power of φ for large field, can be associated with a corresponding renormalised coupling $g(t)$ and thus the universal self-similar flow (7) which is characteristic of the continuum limit.[41,20]

Indeed we see from (16) and (14), that those solutions that behave as a power for large φ must do so as

$$v(\varphi) \sim \varphi^{(D-\lambda)/d} \quad , \tag{17}$$

this being again the required power to balance scaling dimensions (with $[g(t)] = \lambda$) if the remaining quantum corrections may be neglected in this regime. Once again for $\varphi \to \pm\infty$, this supplies two boundary conditions, but this time, since (16) is linear, this overdetermines the equations, and generically allows only certain quantized values of λ.

On the other hand if v does not behave as a power of φ for large φ, then from (16) and (14), we obtain that instead

$$v(\varphi) \sim \exp\left\{A(D-d)\varphi^{D/d}\right\} \quad . \tag{18}$$

Actually, in the large φ regime, we may solve (12) without linearising in α. Indeed we may solve it non-perturbatively. This is because, just as before, we may neglect the quantum corrections in this regime. These are given by the right hand side of (12), and thus $V(\varphi, t)$ follows mean-field-like evolution:

$$V(\varphi,t) \sim e^{Dt} V(\varphi e^{-dt}, 0) \quad . \tag{19}$$

If we take $V(\varphi, 0)$ to be given by (15) at $t = 0$, and use (19), (14) and (17), then we see that for large φ, we recover the t dependence of (15) even without linearising in α Thus for power law v (17) in the large φ regime, we may absorb the t dependence into the self-similar flow of a corresponding coupling $g(t) = \alpha e^{\lambda t}$, even for finite α. But using (19) and the non-power-law behaviour (18), results in the t dependence being 'stuck' in the exponential in (18). In this case then, the scale dependence cannot be combined with α into a corresponding coupling and the RG flow is not self-similar. In fact one can further show that perturbations of the form (18) collapse on t-evolution into an infinite sum of the quantized power-law perturbations, and thus the non-power law eigenperturbations are entirely irrelevant for continuum physics.[20]

Again, these considerations generalise to higher orders of the derivative expansion and thus presumably to the exact RG, the non-quantized perturbations growing

faster than a power for large φ and consequently not associated with renormalised couplings. (The precise form of the large φ dependence of the non-power-law perturbations however depends on non-universal details including the level of derivative expansion approximation used, if any.)[20]

I finish the review with a couple of applications of these ideas.

A NUMERICAL EXAMPLE

It is worth stating again that the derivative expansion gives fair numerical approximations in practice.[4] For example, we have previously reviewed[4] the numerical results for *N* component scalar field theory at the non-perturbative Wilson-Fisher fixed point in three dimensions,[13] and the impressive numerical results for the sequence of multicritical scalar field theories in two dimensions (where all standard methods fail).[38] As an example of the application of all the above concepts to a calculation, we here review the Ising model scaling equation of state, in the symmetric phase, in three dimensions.[20] This is of current interest since it allows direct comparison with the recent progress in accurate calculations from resummed perturbation theory.[43,44,45] The Ising model equation of state follows from the Legendre effective potential of the massive non-perturbative field theory of a single scalar field built around the Wilson-Fisher fixed point. Such a theory has only one relevant eigenperturbation and thus only one coupling – which simply sets the scale. If all quantities are measured in terms of this mass-scale, the results are thus pure numbers and universal. (This is of course the basis for the universality of critical phenomena in the Ising model class.)

A second order derivative expansion, 'O(∂^2)',

$$\Gamma_t = \int d^D x \left\{ \frac{1}{2}(\partial_\mu \varphi)^2 K(\varphi, t) + V(\varphi, t) \right\} \tag{20}$$

with a certain careful choice of smooth cutoff (~ power-law as alluded to earlier) results in the flow equations,[30]

$$\frac{\partial V}{\partial t} + \frac{1}{2}(1+\eta)\varphi V' - 3V = -\frac{1-\eta/4}{\sqrt{K}\sqrt{V''+2\sqrt{K}}} \tag{21}$$

and

$$\frac{\partial K}{\partial t} + \frac{1}{2}(1+\eta)\varphi K' + \eta K = \left(1-\frac{\eta}{4}\right)\left\{ \frac{1}{48}\frac{24KK''-19(K')^2}{K^{3/2}(V''+2\sqrt{K})^{3/2}} \right. \tag{22}$$

$$\left. -\frac{1}{48}\frac{58V'''K'\sqrt{K}+57(K')^2+(V''')^2K}{K(V''+2\sqrt{K})^{5/2}} + \frac{5}{12}\frac{(V''')^2K+2V'''K'\sqrt{K}+(K')^2}{\sqrt{K}(V''+2\sqrt{K})^{7/2}} \right\} .$$

These equations enjoy the following reparametrization symmetry, $\varphi \mapsto \Omega^{5/2}\varphi$, $V \mapsto \Omega^3 V$, $K \mapsto \Omega^{-4}K$, which turns the fixed point equations into non-linear eigenvalue equations for the anomalous dimension η.[30,4] On the other hand, by discarding (22), and setting $K = 1$ and $\eta = 0$ in (21), one obtains the flow equation associated with the lowest order in the derivative expansion (with this cutoff), 'O(∂^0)'.

It should be pointed out that one method of extracting the universal characteristics of the fixed point behaviour in the above equations is simply to solve them in a traditional way, by choosing an appropriate bare potential and bare K (typically $K = 1$), at some initial point t_0 (effectively the cutoff Λ_0) and tuning the potential as the cutoff is removed, so as to recover a continuum limit. Nevertheless, the numerical methods based on the insight of the previous sections are certainly faster, more elegant, and more accurate.

Table 1. Universal coupling constant ratios for the three dimensional Ising model universality class.

Approximation	α_4^p	F_5	F_7	F_9
Sharp	1.514(2)	.0155(3)	$3.6(5) \times 10^{-4}$	$-1.7(5) \times 10^{-5}$
$O(\partial^0)$	1.3012(2)	.01638(1)	$4.68(3) \times 10^{-4}$	$-2.4(1) \times 10^{-5}$
$O(\partial^2)$	.8635(5)	.01719(4)	$4.9(1) \times 10^{-4}$	$-5.2(3) \times 10^{-5}$
∂ exp^n	.86(15)	.0172(3)	$4.9(1) \times 10^{-4}$	$-5(1) \times 10^{-5}$
D=3 [43]	.988(2)	.01712(6)	$4.96(49) \times 10^{-4}$	$-6(4) \times 10^{-5}$
D=3 [44]		.0168 - .0173	$4.1 - 16.2 \times 10^{-4}$	
ϵ−exp. [43]	1.2	.0176(4)	$4.5(3) \times 10^{-4}$	$-3.2(2) \times 10^{-5}$
ϵ−exp. [44]		.0176		
ERG [36,46]	1.20	.016	4.3×10^{-4}	
HT [47]	.99(6)	.0205(52)		
HT [48]	1.019(6)	.01780(15)		
HT [49]	.987(4)	.017(1)	$5.4(6) \times 10^{-4}$	$-2(1) \times 10^{-5}$
MC [50]	.97(2)	.0227(26)		
MC [51]	1.020(8)	.027(2)	.00236(40)	

We will not detail the numerical methods used to implement the above ideas,[13,4,38,30] but simply quote the results. We found just one non-trivial non-singular fixed point solution with $\eta = .05393208\cdots$, which is thus identified with the Wilson-Fisher fixed point. [30,20] Our resulting $O(\partial^2)$ value for η should be compared with the (combined) worlds best estimates [42,30] $\eta = .035(3)$. Extracting from the quantized spectrum of power-law eigenperturbations, the relevant operator and the least irrelevant operator, one obtains the critical exponents ν and ω (via the formulae[1] $\nu = 1/\lambda$ and $\omega = -\lambda$). We found at $O(\partial^0)$, $\nu = .6604$, $\omega = .6285$, and at $O(\partial^2)$, $\nu = .6181$ and $\omega = .8972$. These should be compared to combined worlds best estimates of $\nu = .631(2)$ and $\omega = .80(4)$. [42,30]

Having transformed to renormalised variables and integrating out along the renormalised trajectory we have that the Legendre effective potential $V(\varphi)$ is given by the $t \to \infty$ limit of $V(\varphi, t)$. Writing in terms of the physical dimensionful variables

$$V(\varphi) = \frac{1}{2}m^2\varphi^2 + u_4\varphi^4 + u_6\varphi^6 + \cdots, \tag{23}$$

the ratios $\alpha_{2k}^p = u_{2k}m^{k-3}$ are dimensionless and universal. By Taylor expansion of the equations (21,22), careful choice of closure ansatze, and reworking the equations to expose the self-similar flow along the renormalised trajectory (in the way indicated earlier), we obtained ordinary differential equations which could be straightforwardly integrated to obtain the α_{2k}^p. [20] (We performed all the calculations within the Maple package.) The results are displayed in table 1. The six-point coupling and higher are written in terms of $F_{2k-1} = 2k\alpha_{2k}^p/(24\alpha_4^p)^{k-1}$, which allows a direct comparison with the most accurate recent perturbative results. [43] In row "Sharp" of this table, we show also the results obtained from eqn.(12). In the row "∂ exp^n", we use both orders of the derivative expansion, and a comparison of our results for the critical exponents ω and ν with the best determinations, to estimate an error. [20] It should not be taken as seriously as the very careful error analyses possible, and performed, [43] in the large order perturbation theory calculations. Those rows labelled D=3, ϵ –exp., ERG, HT, and MC give results respectively from resummed perturbation theory, ϵ expansion, another exact RG approximation, high temperature series and Monte-Carlo estimates.

It can be seen from the table, that while perturbative methods are more powerful than the derivative expansion for low order couplings, the derivative expansion even-

tually wins out. The reason for this is that the derivative expansion at these lowest orders, is crude in comparison to the perturbation theory methods, however the perturbative methods suffer from being asymptotic – which in particular results in rapidly worse determinations for higher order couplings. The derivative expansion does not suffer from this, since it is not related at all to an expansion in powers of the field. Indeed, it may be shown that[2] even at the level of the LPA, Feynman diagrams of all topologies are included.

Let us mention that we also obtained 'for free' some estimates for F_{11} and F_{13}, and universal coefficient ratios in the $O(\partial^2)$ function $K(\varphi)$.[20] These latter correspond to universal information about the $O(p^2)$ terms of the one particle irreducible Green functions. At present, we know of no other estimates with which these can be compared.

THEORETICAL EXAMPLE – A C FUNCTION

Following Zamolodchikov's celebrated *c*-theorem[52] for two dimensional quantum field theory, a number of groups have sought to generalise these ideas to higher dimensions.[25,53] The motivation behind this, is not only to demonstrate irreversibility of renormalization group flows, and thus prove that exotic flows such as limit cycles, chaos, etc, are missing in these cases, but perhaps more importantly to provide an explicit, and useful, geometric framework for the space of quantum field theories,. Zamolodchikov established three important properties for his *c* function:

(i) There exists a function $c(g) \geq 0$ of such a nature that $\frac{d}{dt}c \equiv \beta^i(g)\frac{\partial}{\partial g^i}c(g) \leq 0$, the equality being reached only at fixed points $g(t) = g_*$.
(ii) $c(g)$ is stationary at fixed points, i.e. $\beta^i(g) = 0$ for all i, implies $\partial c/\partial g^i = 0$.
(iii) The value of $c(g)$ at the fixed point g_* is the same as the corresponding Virasoro algebra central charge.[54] (This property thus only makes sense in two dimensions.)

Within the LPA to Wilson's, or Polchinski's exact RG,¶

$$\frac{\partial}{\partial t}V(\varphi,t) + \frac{1}{2}(D-2)\varphi_a\frac{\partial V}{\partial \varphi_a} - DV = \frac{\partial^2 V}{\partial\varphi_a\partial\varphi_a} - \left(\frac{\partial V}{\partial\varphi_a}\right)^2 \tag{24}$$

we have discovered[24] a *c*-function which has the first two properties in any dimension *D*. Our *c*-function has a counting property which generalises property (iii): it is extensive at fixed points, i.e. additive in mutually non-interacting degrees of freedom, and counts one for each Gaussian scalar and zero for each infinitely massive scalar (corresponding to a High Temperature fixed point). These properties are shared by the two dimensional Virasoro central charge. It is probably not possible within the LPA, to establish a more concrete link to Zamolodchikov's *c*.

The idea is very simple and builds on the observation of Zumbach,[55] that (24) may be written as a gradient flow

$$a^N G\frac{\partial\rho}{\partial t} = -\frac{\delta\mathcal{F}}{\delta\rho} \quad , \tag{25}$$

where we have introduced $G = \exp -\frac{1}{4}(D-2)\varphi_a^2$, $\rho = \mathrm{e}^{-V}$, and a measure normalization factor a. The functional F is given by

$$\mathcal{F}[\rho] = a^N\int d^N\varphi\; G\left\{\frac{1}{2}\left(\frac{\partial\rho}{\partial\varphi_a}\right)^2 + \frac{D}{4}\rho^2(1-2\ln\rho)\right\} \quad . \tag{26}$$

¶for *N* scalar fields, with no symmetry implied, and (almost) any smooth cutoff.[24,56]

Fixed points $\rho = \rho_*$ are then given by $\delta F/\delta\rho = 0$. If one substitutes this equation back into F one obtains[24]

$$\mathcal{F}[\rho_*] = \frac{D}{4} a^N \int d^N\varphi \, G\rho_*^2 \quad . \tag{27}$$

Now, if the field content splits into two mutually non-interacting sets, the potential splits in two, and thus ρ_* factorizes. We see then that the integral in (27) itself factorizes. Thus if we define a c function through the logarithm of F, $c = \ln(4F/D)/A$ (where A is another normalization factor), it will be additive in mutually non-interacting degrees of freedom. Furthermore, we may use the normalization constants a and A, to normalise the c function so that it counts one for Gaussian scalars ($V_* = 0$) and zero for the high temperature fixed point ($V_* = \frac{1}{2}\varphi^2 - 1/D$ per scalar[21,24]).

The other properties follow from the geometrization of (25), by rewriting the flow in terms of a complete set of coupling constants $g^i(t)$, $V \equiv V(\varphi, g)$. Close to a fixed point, these need only be the relevant and marginal couplings, as we discussed earlier. Elsewhere, they have to be infinite in number to span the space of all potentials. Following Zamolodchikov[52] we generalise (9) to the whole space by writing $O_i(g) = \partial_i V(\varphi, g)$, where $\partial_i \equiv \partial/\partial g^i$. Then, after some straightforward manipulation, we obtain

$$\partial_i c(g) = -\mathcal{G}_{ij}\beta^j(g) \quad , \tag{28}$$

where G_{ij} is our analogue of the so-called Zamolodchikov metric on coupling constant space:[52]

$$\mathcal{G}_{ij}(g) = (\mathcal{F}\ln A) a^N \int d^N\varphi \, G\rho^2 \mathcal{O}_i\mathcal{O}_j \quad . \tag{29}$$

Property (ii) follows immediately from (28), while, because G_{ij} is positive definite, multiplication of (28) by β^i establishes property (i).

I close with two remarks. Firstly, it would be very interesting of course, to understand if this LPA c function generalizes to higher orders in the derivative expansion[38] (which likely would allow a direct comparison with Zamolodchikov's c), or indeed generalizes to an exact expression along the present lines. Secondly, the geometric structure immediately suggests a variational approximation method for (24): Restricting the flows to some finite dimensional submanifold parametrized (in some way) say by $g^1, \cdots, g^M$, corresponds to ansatzing a subspace of potentials $V \equiv V(\varphi, g)$, where g stands now only for the M parameters. Approximations to the fixed points then follow from the variational equations $\partial_i c(g) = 0$. Our investigations indicate that this approximation method is rather powerful.[24,57] This illustrates once again that purely theoretical insights into the structure of the continuous RG and its approximations, tend in turn to suggest yet more elegant and more powerful methods of approximation.

REFERENCES

1. K. Wilson and J. Kogut, Phys. Rep. 12C (1974) 75.
2. T.R. Morris, Int. J. Mod. Phys. A9 (1994) 2411.
3. T.R. Morris, in *Lattice '94*, Nucl. Phys. B(Proc. Suppl.)42 (1995) 811.
4. T.R. Morris, in *RG96*, hep-th/9610012.
5. Yu. Kubyshin, in *RG96*, hep-th/9702082.
6. K.-I. Aoki et al, Prog.Theor.Phys. 97 (1997) 479.
7. K.-I. Aoki, in *SCGT 96*, hep-ph/9706204.
8. J. Berges et al, hep-ph/9705474.
9. M. Bonini and F. Vian, hep-th/9707094.
10. M. D'Attanasio and M. Pietroni, hep-ph/9612283, hep-th/9611038.
11. N. Tetradis, Nucl.Phys.B488 (1997) 92.
12. J.D. Shafer and J.R. Shepard, Phys.Rev.D55 (1997) 4990.

13. T.R. Morris and M. Turner, hep-th/9704202, to be published in Nucl. Phys. B.
14. J. Comellas, hep-th/9705129.
15. S. Falkenberg and S.D. Odintsov, hep-th/9612019;
L.N. Granda and S.D. Odintsov, hep-th/9706062.
16. M. Bonini and M. Simionato, hep-th/9705146.
17. U. Ellwanger, hep-ph/9702309; C. Wierczerkowski, hep-th/9705096, 9612226, 9612214.
18. C. Bagnuls and C. Bervillier, J. Phys. Stud. 1 (1997) 366; A.P.C. Malbouisson et al, cond-mat/9609035, cond-mat/9705228; J. Berges et al, Phys. Lett. B393 (1997) 387.
19. A. Bonanno and D. Zappala, Catania preprint 97-196; K. Langfeld and H. Reinhardt, hep-ph/9702271; K. Sailer and W. Greiner, hep-th/9610143.
20. T.R. Morris, Nucl.Phys. B495 (1997) 477.
21. J. Comellas and A. Travesset, Nucl. Phys. B498 (1997) 539.
22. K.-I. Aoki et al, Prog. Theor. Phys. 95 (1996) 409.
23. M. D'Attanasio and T.R. Morris, hep-th/9704094, to be published in Phys. Lett. B.
24. J. Generowicz, C. Harvey-Fros and T.R. Morris, hep-th/9705088, to appear in Phys. Lett. B.
25. J. Gaite and D. O'Connor, Phys, Rev. D54 (1996) 5163; J. Gaite, C96-08-26.1, hep-th/9610040; R.C. Myers and V. Periwal, hep-th/9611132.
26. H.J. de Vega et al, astro-ph/9609005, Phys.Rev.D54 (1996) 6008.
27. M. Salmhofer, Nucl. Phys. B (Proc. Suppl.) 30 (1993) 81;
M. Bonini et al, Nucl. Phys. B418 (1994) 81.
28. C. Wetterich, Phys. Lett. B301 (1993) 90.
29. J. Polchinski, Nucl. Phys. B231 (1984) 269.
30. T.R. Morris, Phys. Lett. B329 (1994) 241.
31. F.J. Wegner and A. Houghton, Phys. Rev. A8 (1973) 401.
32. T.R. Morris, Nucl. Phys. B458[FS] (1996) 477.
33. T.R. Morris, Phys. Lett. B357 (1995) 225, Phys. Rev. D53 (1996) 7250;
M. D'Attanasio and T.R. Morris, Phys. Lett. B378 (1996) 213.
34. J.F. Nicoll and T.S. Chang, Phys. Lett. 62A (1977) 287;
M. Bonini et al, Nucl. Phys. B409 (1993) 441.
35. T.R. Morris, Phys. Lett. B334 (1994) 355.
36. N. Tetradis and C. Wetterich, Nucl. Phys. B422 (1994) 541.
37. M. Alford, Phys. Lett. B336 (1994) 237.
38. T.R. Morris, Phys. Lett. B345 (1995) 139.
39. J.F. Nicoll et al, Phys. Rev. Lett. 33 (1974) 540.
40. A. Hasenfratz and P. Hasenfratz, Nucl. Phys. B270 (1986) 687.
41. T.R. Morris, Phys. Rev. Lett. 77 (1996) 1658.
42. E.g. J. Zinn-Justin, "Quantum Field Theory and Critical Phenomena" (1993) C.P., Oxford.
43. R. Guida and J. Zinn-Justin, Nucl. Phys. B489 (1997) 626.
44. A.I. Sokolov, Phys. Solid State 38 (1996) 354; A. I. Sokolov et al, J. Phys. Stud. 3 (1997) 1, Phys. Lett. A227 (1997) 255.
45. See the first lecture by J. Zinn-Justin, this conference.
46. J. Berges, N. Tetradis and C. Wetterich, Phys. Rev. Lett. 77 (1996) 873.
47. T. Reisz, Phys. Lett. B360 (1995) 77.
48. S. Zinn, S.-N. Lai, and M.E. Fisher, Phys. Rev. E 54 (1996) 1176.
49. P. Butera and M. Comi, preprint IFUM 545/TH (1996).
50. M.M. Tsypin, Phys. Rev, Lett. 73 (1994) 2015.
51. J-K Kim and D.P. Landau, hep-lat/9608072.
52. A.B. Zamolodchikov, JETP Lett. 43 (1986) 730.
53. J.L. Cardy, Phys. Lett. B215 (1988) 749; H. Osborn, Phys. Lett. B222 (1989) 97; I. Jack and H. Osborn, Nucl. Phys. B343 (1990) 647; A. Cappelli et al, Nucl. Phys. B352 (1991) 616, B376 (1992) 510; D. Freedman et al, Mod. Phys. Lett. A6 (1991) 531; G.M. Shore, Phys. Lett. B253 (1991) 380, B256 (1991) 407; A.H. Castro Neto and E. Fradkin, Nucl. Phys. B400 (1993) 525; P. Haagensen et al, Phys. Lett. B323 (1994) 330; X. Vilasís-Cardona, Nucl. Phys. B435 (1995) 635; A.C. Petkou, Phys. Lett. B359 (1995) 101; V. Periwal, Mod. Phys. Lett. A10 (1995) 1543; F. Bastianelli, Phys. Lett. B369 (1996) 249; J. Comellas and J.I. Latorre, UB-ECM-PF 96/1, hep-th/9602123; J.I. Latorre and H. Osborn, DAMTP-97-1, hep-th/9703196.
54. A.A. Belavin, A.M. Polyakov and A.B. Zamolodchikov, Nucl. Phys. B241 (1984) 333.
55. G. Zumbach, Nucl. Phys. B413 (1994) 754, Phys. Lett. A190 (1994) 225.
56. R.D. Ball et al, Phys. Lett. B347 (1995) 80.
57. work in progress.

PRIMORDIAL MAGNETIC FIELDS AND THEIR DEVELOPMENT (APPLIED FIELD THEORY)

P. Olesen

The Niels Bohr Institute
Blegdamsvej 17
DK-2100 Copenhagen Ø
Denmark

MOTIVATION

In this talk I discuss the non-linear development of magnetic fields in the early universe. Since this is based on a classical field theory, which turns out to have a rather complex structure, I thought it could be of interest for this meeting, under the heading of "applied field theory". I very briefly mention a number of particle physics mechanisms for generating magnetic fields in the early universe, but the emphasis is on the field theoretic aspects of the developments of these fields, mainly on the occurence of inverse cascades, i.e. generation of order from disorder. In this connection it is also discussed how the Silk effect (photon diffusion) is counteracted by the inverse cascade, which moves energy from smaller to larger scales.

Many galaxies (including our own) are observed to have magnetic fields. One way to observe such fields is to study the polarization of light passing the galaxy. Due to the interaction with the field and the plasma there is a *Faraday rotation* of the polarization vector, proportional to the field and to the square of the wave length of the light. In this way fields are found to have the order of magnitude $10^{-6} - 10^{-8}$ Gauss on a scale of 100 kpc* If you have forgotten what a G(auss) is: the mean field on the sun is approximately one G.

Usually the galactic magnetic field is explained by the *dynamo effect*: turbulence (e.g. differential rotation) in the galaxy enhances the magnetic field exponentially up to some saturation value, corresponding to equipartition between kinetic and magnetic energy. The dynamics which governs these phenomena is called *magnetohydrodynamics*, abbreviated as MHD, which is essentially the Maxwell plus Navier-Stokes equations. The dynamo can produce an enhancement factor of several orders of magnitude. An important feature is that the dynamo needs a seed field. It appears reasonable to assume that this field is of *primordial origin*, i.e. it has existed already in the early universe. Astrophysicists often say as a joke that a primordial magnetic field is a field

*A p(arse)c is an astronomical unit, which has the physical value 1 pc ≈ 3.26 light year.

New Developments in Quantum Field Theory
Edited by Damgaard and Jurkiewicz, Plenum Press, New York, 1998

which has existed for so long that everybody has forgotten how it was created. However, in particle physics we must be more serious since we have knowledge of the early universe, and hence we should explain the origin of these fields.

PRIMORDIAL MAGNETIC FIELDS IN THE EARLY UNIVERSE

In natural units magnetic fields have dimension (mass)^2. At the electroweak scale, assuming the Higgs mass to be of order m_W, there is essentially only one mass, m_W, and we may therefore expect something like

$$B_{\rm EW} \lesssim m_W^2 \approx 10^{24}\ {\rm G} \tag{1}$$

on a scale $\sim 1/m_W$. This is a huge field, far larger than anything one has ever seen or produced on this earth. How does this compare with the rather weak fields found in galaxies?

In the standard cosmological model all distances are blown up by the scale factor $R(t)$. It is useful for estimates that the scale factor is proportional to the inverse temperature. Thus, $R_{\rm now}/R_{\rm EW} = T_{\rm EW}/T_{\rm now} \approx 10^{15}$. Hence, an initial correlation length of order $\sim 1/m_W$ is of order 1 cm today, which has no astrophysical interest. We need fields on a scale of order 100 kpc $\approx 3\times10^{23}$cm.

If we assume that B is essentially random, we can estimate the field at any distance from a simple random walk. We have the field at the initial correlation length, but we want it at $\approx 10^{23}$ times this length. Thus, in d dimensions we have

$$< B_{\rm EW} >_{\rm scale\ 10^{23}/m_W} \approx 10^{24}\ {\rm G}/(10^{23})^{d/2}. \tag{2}$$

So for $d = 3$ we get $< B_{\rm EW} > \approx 10^{-10}$G, whereas for $d = 2$ and $d = 1$ we have $< B_{\rm EW} >$ approximately equal 10 G and 10^{12}G, respectively, on the scale of $10^{23}/m_W$.

In order to see if these fields are reasonable, we need to know the cosmological developments of $< B >$. From MHD (with viscosity ignored) one has the result that the flux through a surface bounded by a curve following the fluid of charged particles is conserved. Since such a surface increases like R^2, it follows that $< B >$ decreases like $1/R^2$ when the universe expands with the scale factor R[†]. It therefore follows that if today we need e.g. a primordial field of order 10^{-15}G on a scale of 100 kpc, then on the corresponding scale $10^{23}/m_W$ at the electroweak phase transition, we need $< B > \approx 10^{15}$ G. Thus, from the random walk estimates above we see that only the case $d = 1$ comes near this value, although a factor 10^3 is missing. Actually one could argue that the case $d = 1$ is relevant, because in observing the magnetic field by Faraday rotation, a one dimensional average is made along the line of sight. However, this argument is not really convincing, since the dynamo effect is three dimensional, and hence the field relevant for this effect is the very small $3d$ average.

The conclusion is thus that if fields of the order of m_W^2 can be generated at the EW-scale, then there could still be missing a factor of order 10^x, where x is of order 3. However, it should be emphasized that a field of order m_W^2 is very large, and is not obtained in most mechanisms for creation of primordial fields. Hence, in most cases x is larger than 3.

[†]The metric is

$$d\tau^2 = dt^2 - R(t)^2\left[\frac{dr^2}{1-kr^2} + r^2 d\Omega^2\right], \tag{3}$$

where $k = +1, 0, -1$ for a closed, flat or open universe, respectively.

MAGNETIC FIELDS FROM PARTICLE PHYSICS MECHANISMS

In this section we very briefly discuss a number of proposed mechanisms for the generation of primordial fields. The list is by no means exhaustive.

Fields From Inflation

An inflationary creation of primordial magnetic fields has the advantage that the coherence scale is larger than in other mechanisms. As an example, we mention the work by Gasperini, Giovannini and Veneziano [1], which is based on a pre-big-bang cosmology inspired by superstrings, which is an alternative to the usual slow roll inflation. The dilaton field ϕ in the Lagrangian

$$\mathcal{L} = -\sqrt{g}\, e^{-\phi}(R + \frac{1}{2}\,\partial_\mu\phi\,\partial^\mu\phi + \frac{1}{4}F_{\mu\nu}F^{\mu\nu}) \tag{4}$$

amplifies the quantum fluctuations of $F_{\mu\nu}$. The magnetic energy spectrum behaves like $\sim k^{0.8}$. The resulting magnetic fields are of the right order of magnitude on a 100 kpc scale. Recently, however, Turner and Weinberg [2] have argued that this scenario requires fine tuning of the initial conditions in order to get enough inflation to solve the flatness and horizon problems.

Bubble Formation at the EW Phase Transition and Magnetic Fields

In a first order EW phase transition bubbles of new vacuum are formed. This was used by Baym, Bödecker and McLerran [3] to obtain the generation of a magnetic field. The main point is that the bubbles, although overall neutral, have a dipole charge layer on the surface, so rotating bubbles generate a field. Although the field from each bubble is very small, there is a large number of bubbles, so depending on the subsequent development of $< B >$, in the end a reasonable magnitude can be produced. A different mechanism was considered by Kibble and Vilenkin [4]: when the bubbles collide, the phase of the Higgs field varies, giving rise to currents and a magnetic field. Again, in this case one can get a reasonable magnitude provided the subsequent development of $< B >$ is favourable.

Superconducting Cosmic Strings and Other Mechanisms

There exists a number of other mechanisms for the generation of magnetic fields. Vachaspati pointed out that if the gradients of Higgs fields fluctuate, they can induce a magnetic field[5] at the electroweak scale. The statistical averaging involved in this scenario was discussed in details by Enqvist and me[6]. As mentioned already in the first section, the conclusion is that if line averaging is relevant, one obtains nearly the right order of magnitude, since by this mechanism the field at genesis is of order m_W^2 on a scale of $1/m_W$. However, this scenario operates with physically motivated fluctuations in gauge dependent quantities like gradients of Higgs fields. Such a procedure is not very clear to me.

Recently there has been discussions of generation of primordial magnetic fields from a network of Witten's superconducting cosmic strings[7,8]. These strings are current carrying, and hence produce magnetic fields. It turns out that if the strings are created at the GUT phase transition, where the current is very large, they can produce a field which is large enough over a sufficient scale, assuming that MHD does not give rise to any trouble. On the other hand, superconducting strings created at the EW phase transition cannot generate sufficient fields.

It has also been proposed that a Savvidy-type vacuum, where the energy is lowered relative to the trivial vacuum by having a magnetic field, can generate enough field[9]. For SU(N) the field produced at a temperature T is of order

$$B \sim T^2 \exp\left(-\frac{48\pi^2}{11Ng^2}\right). \tag{5}$$

At the EW transition (N = 3), this field is far too small. At the GUT transition, however, it produces a large enough field for N = 5, due to the strong sensitivity of the exponent with respect to N. Whether this field is acceptable depends on the subsequent development according to MHD.

It was proposed long ago by Harrison that magnetic fields could be generated from vorticity present in eddies of plasma in the early universe[10]. This idea was criticized, and the eddies were replaced by irrotational density fluctuations by Rees[10]. A more modern version of this scenario is due to Vachaspati and Vilenkin[10], where the magnetic field is generated by vorticity arizing in the wakes of ordinary (i.e. not superconducting) cosmic strings.

Finally we mention that recently Joyce and Shaposhnikov[11] have presented a scenario which has the potential of leading to quite large fields. The standard model has charges with abelian anomaly only (e.g. right-handed electron number) which are essentially conserved in the very early universe, until a short time before the EW transistion. A state with finite chemical potential of such a charge is unstable to the generation of hypercharge U(1) fields. Such fields can turn into large magnetic fields, depending on their subsequent development.

It is clear that the physical validity of most, if not all, of these scenarios, depends on the subsequent non-linear development of the primordial field, due to MHD. This will be discussed in the next section. In the end of this talk we shall also discuss Silk diffusion, which is a mechanism for destroying magnetic fields by turning it into heat. We shall show that this linear diffusion is in fact counteracted by the non-linear terms of MHD.

INVERSE CASCADE FROM MAGNETOHYDRODYNAMICS

We shall now investigate what happens subsequently to a primordial magnetic field generated in the early universe. To simplify things, here we give the details only for the non-relativistic case, and mention without giving the arguments, what happens in the general relativistic case.

The Non-Relativistic MHD Equations

In the rest frame of a plasma consisting of charged particles with current **j**, we have Ohm's law,

$$\mathbf{j}_{\text{rest}} = \sigma\, \mathbf{E}_{\text{rest}}, \tag{6}$$

where σ is the conductivity. The universe is a good conductor, so σ is very large[‡]. Thus it follows that

$$\mathbf{E}_{\text{rest}} \approx 0. \tag{7}$$

In a frame moving with bulk velocity $\mathbf{v}$ one therefore has

$$\mathbf{E} \approx -\mathbf{v} \times \mathbf{B}. \tag{8}$$

The induction equation $\partial \mathbf{B}/\partial t = -\nabla \times \mathbf{E}$ therefore gives

$$\partial \mathbf{B}/\partial t \approx \nabla \times (\mathbf{v} \times \mathbf{B}). \quad \text{(MHD I)} \tag{9}$$

This is one (out of two) of the fundamental MHD equations. It tells us that the magnetic field is influenced by the velocity, and also, if you start from $\mathbf{B} = \mathbf{0}$ no magnetic field can be generated. Therefore a seed field is needed in the dynamo mechanism.

The second fundamental MHD equation is the Navier-Stokes equation with the Lorentz force $\mathbf{j} \times \mathbf{B}$ on the right hand side. Here $\mathbf{j}$ can be estimated from the Maxwell equation $\mathbf{j} + \partial \mathbf{E}/\partial t = \nabla \times \mathbf{B}$. The time derivative can be estimated to be small in the non-relativistic cases[§], so $\mathbf{j} \approx \nabla \times \mathbf{B}$, so the Navier-Stokes equation with the Lorentz force is given by

$$\partial \rho \mathbf{v}/\partial t + (\mathbf{v}\nabla)\rho\mathbf{v} \approx -\nabla(p + \frac{1}{2}B^2) + (\mathbf{B}\nabla)\mathbf{B}. \quad \text{(MHD II)} \tag{10}$$

Here for simplicity we have ignored the viscosity. For σ large, this can be generalized to the relativistic case at the expense of a considerable increase in the complexity of the equations.

Why and When does MHD have an Inverse Cascade: A Simple Scaling Argument

Now let us suppose that by some particle physics mechanism a primordial magnetic field is generated. At the genesis the field has some correlation length, and the crucial question is then what happens as time passes. If the correlation length grows smaller, corresponding to a *cascade*, the situation is quite bad, even for the inflationary scenario. Such a cascade would appear if the system develops into a more chaotic direction. If, on the other hand, we have an *inverse cascade*, the correlation length grows and the system develops towards more order. In an inverse cascade, energy is thus transferred from smaller to larger scales.

It turns out that[13] the situation depends on the initial spectrum. Roughly speaking, if the spectrum is concentrated at short (large) distances, it will develop into large (short) distances. To see this, one can make use of the fact that the MHD equations are invariant under the "self-similarity" equations,

$$\mathbf{x} \to l\mathbf{x}, \; t \to l^{1-h}t, \; \mathbf{v} \to l^h\mathbf{v}, \; \mathbf{B} \to l^h\mathbf{B}, \tag{11}$$

[‡] In the relativistic era this can be seen from the following estimate: The current is defined by $\mathbf{j} = ne\mathbf{v}$. The velocity is given essentially by the Newtonian expression $\mathbf{v}/\tau \approx e\mathbf{E}/E$, where E is the relativistic energy, and τ is the average time between collisions. Thus, $\tau \approx 1/n\sigma_\times$, where $\sigma_\times$ is a typical relativistic cross section. Thus, $\mathbf{j} \approx ne(\tau e\mathbf{E}/E)$, so $\sigma \approx ne^2\tau/E \approx e^2/E\sigma_\times$. A relativistic cross section goes like e^4/T^2, since the temperature T is a typical momentum transfer. Also, $E \sim T$. Thus $\sigma \approx T/e^2$, which is very large in the early universe, because the temperature is very high. At later stages the universe is still a good conductor, for different reasons[12].

[§] We have $E \sim vB$, so $\partial E/\partial t \sim (v/l)vB \sim (B/l)v^2$. But $|\nabla \times \mathbf{B}| \sim B/l$, where l is some typical length, so the time derivative of the electric field can be ignored relative to the curl of B.

and if the viscosity ν and Ohmic resistance are included, we further have

$$\nu \to l^{1+h}\nu, \ \sigma \to l^{-1-h}\sigma. \tag{12}$$

¿From this it is very easy to show that the magnetic and kinetic energy densities (ε), given in 3+1 dimensions by expressions like

$$\mathcal{E}(k,t)\ V = \frac{2\pi k^2}{(2\pi)^3}\int d^3x\ d^3y\ e^{i\mathbf{k}(\mathbf{x}-\mathbf{y})} < \mathbf{B}(\mathbf{x},t)\mathbf{B}(\mathbf{y},t) >, \text{ with } V = \int d^3x, \tag{13}$$

where

$$V\int dk\ \mathcal{E}(k,t) = \frac{1}{2}\int d^3x < \mathbf{B}^2 >= \text{ total magnetic energy}, \tag{14}$$

must satisfy the scaling relation[13]

$$\mathcal{E}(k/l, l^{1-h}t) = l^{1+2h}\ \mathcal{E}(k,t). \tag{15}$$

This is valid in the *inertial range*, where viscosity and Ohmic resistance can be ignored. The general solution of this equation is

$$\mathcal{E}(k,t) = k^p\ \psi(k^{(3+p)/2}t), \tag{16}$$

with $p = -1 - 2h$ and ψ some function of the single argument $k^{(3+p)/2}t$. The interpretation of this equation is that, if at the initial time $t = 0$ the spectrum is k^p (from some particle physics mechanism), then at later times it will be governed by the function ψ. Hence the wave vector scales like

$$k \sim t^{-2/(3+p)} \tag{17}$$

Thus, if $p > -3$ there is an *inverse cascade*, because k moves towards smaller values, whereas for $p < -3$, there is a cascade. Thus, if initially we have a random system corresponding to $p \geq 0$, then later the system becomes more ordered, as already announced. For $p = -3$ it follows from eq. (15) that the k and t dependence of the energy density become uncorrelated.

In the case when general relativity is included one obtains for a flat, expanding universe[13]

$$R(t)^4\ \mathcal{E}(k,t) = k^p\ \psi(k^{(3+p)/2}\ \tilde{t}), \tag{18}$$

where t is the Hubble time, $\tilde{t} = \int dt/R(t)$ is the conformal time, $\tilde{t} \propto \sqrt{t}$, and where k is the *comoving* wave vector, so that the physical wave vector is $k_{\text{phys}} = k/R(t)$. Therefore the physical wave vector scales like

$$k_{\text{phys}} \sim \tilde{t}^{-2/(3+p)}/R(t) \propto t^{-(5+p)/2(3+p)}. \tag{19}$$

Thus, if the spectrum starts out with $p = 2$, corresponding to a Gaussian random initial field¶, we have a scaling of the physical wave vector by $t^{-0.7}$, instead of $t^{-0.5}$

¶ If the initial field is given by

$$< B_i(\mathbf{x},0)B_k(\mathbf{y},0) >= \lambda\left(\delta_{ik} - \frac{\partial_i\partial_k}{\partial^2}\right)\delta^3(\mathbf{x}-\mathbf{y}), \tag{20}$$

then the initial energy is given by

$$\mathcal{E}(k,0) = \lambda(\delta_{ii} - k_ik_i/k^2)k^2 = 2\lambda k^2. \tag{21}$$

Thus the general relativistic scaling goes as $k_{\text{phys}} \sim t^{-3/5}$.

from pure expansion. If p decreases, the effective expansion increases. Thus, if the initial spectrum is characterized by $p = 1$, we get a scaling $k_{\text{phys}} \sim t^{-0.75}$, and for $p = 0$ we have $k_{\text{phys}} \sim t^{-0.8}$. These examples imply physically that the typical size of an eddy increases like $t^{0.2}$, $t^{0.25}$ and $t^{0.3}$, respectively, on top of the expansion factor. Finally we mention that the "scale invariant" initial spectrum with $p = -1$ (dk/k =scale inv.) has an increase of the typical eddy size by t, which means that the eddies follow the horizon.

Numerical Simulations in 2+1 Dimensions

From the general scaling arguments presented above one cannot deduce the value of the scaling function ψ. Here numerical investigations are needed, since realistic analytic solutions of MHD are not known. However, a problem arises, since the Reynold number‖ is very large in the early universe. For example, in the paper by Brandenburg, Enqvist and me[14] the magnetic Reynold number was estimated to be of order 10^{17}. In numerical simulations one cannot reach this value, no matter how much computertime is used. We therefore did numerical simulations in 2+1 dimensions with an unrealistically low Reynold number[14], taken to be 10. So the non-linear terms are approximately ten times as important as the diffusion terms. These terms are, however, needed to achieve numerical stability (they act as a short distance cutoff).

We used the general relativistic MHD equations, which are considerably more complicated than the MHD equations discussed in a previous section. We took the initial conditions that the velocity vanishes and **B** is Gaussian random, so the magnetic energy spectrum goes like k. Also, we took the energy density to be $\rho = \text{const}/R^4$, and the pressure $p = \rho/3$.

In fig. 1 the numerical results[14] are displayed. At the initial time there is a rather chaotic state, where the magnetic flux lines are either long random walk curves, or small closed loops. We used periodic boundary conditions, and satisfied div**B** = 0. We see that in a short time the typical lenght scale increases considerably. In the end of the simulation there are quite large eddies. Therefore we clearly see an inverse cascade, where order is produced from chaos, in contrast to the usual paradigm.

Also, the initial velocity **v** = 0 acquires a spectrum which shows an inverse cascade. The velocity is initially induced by the Navier-Stokes equation through the Lorentz force. The velocity generated this way then influences the magnetic field through the induction equation, etc. etc.

Numerical Simulations in 3+1 Dimensions: The Shell Model

As already mentioned, simulations of MHD with large Reynold numbers is not possible with present day computers**. The situation gets worse when we go from 2+1 to 3+1 dimensions. Therefore one needs to make a model which has as many features of the real Navier-Stokes as is compatible with practical tractability. In recent years the so-called GOY (Gledzer, Ohkitani and Yamada) model has become increasingly popular. Another name for this model is the "shell model". It gives results in good agreement with experiments, especially as far as the subtle intermittency effects are concerned. The model captures a basic feature of turbulence, namely the coupling of many different length scales. It is not known whether the model has relation to the real

‖ This number can be understood as the ratio between "typical" non-linear terms and the linear viscosity term. Thus, if *Re* is large, turbulence is important.

** This also applies to hydrodynamics, and is perhaps the reason why weather forecasts are pretty bad, at least in Denmark.

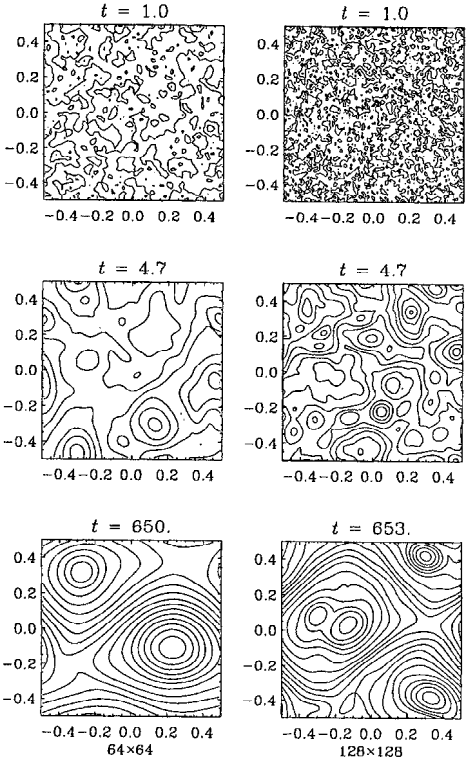


Figure 1. Left column: magnetic field lines at different times at low resolution (64 × 64 meshpoints). Right column: magnetic field lines at different times at higher resolution (128 × 128 meshpoints). This figure is taken from ref. 14.

Navier-Stokes and MHD. But it nicely illustrates the behaviour of a system in which numerical simulations are made difficult by the effect of a huge number of couplings between the different length scales. Also, real world conservation laws (energy, helicity) are buildt into the model.

To motivate the model, let us mention that in the Navier-Stokes equations and MHD one has terms like

$$(\mathbf{v}\nabla)\mathbf{v}, \quad \nabla \times (\mathbf{v} \times \mathbf{B}), \quad (\mathbf{B}\nabla)\mathbf{B}, \tag{22}$$

etc. In Fourier space they e.g. have the form

$$(\mathbf{v}\nabla)\mathbf{v} \to \int d^3p \; v_i(\mathbf{p})(p_i - k_i)v_j(\mathbf{p} - \mathbf{k}). \tag{23}$$

Experience with numerical simulations show that the largest contributions come from triangles in **k**–space with similar side lengths. This is taken as a "phenomenological" input in the shell model.

At this stage, for numerical purposes, one would discretize **k**–space. In the shell model one of the basic ingredients is a hierachical structure, where $|\mathbf{k}|$–space is divided into shells

$$k_n = \lambda^n k_0, \quad n = 1, 2, ..., N. \tag{24}$$

Here λ is often taken to be 2. There furthermore exists a complex "velocity mode" $v_n = v(k_n)$, which can be considered as the Fourier transform of the velocity difference $|\mathbf{v}(|\mathbf{x}| + 2\pi/\lambda) - \mathbf{v}(|\mathbf{x}|)|$. Since k_n increases exponentially, it covers a wide range of corresponding length scales. The model then assumes couplings between neighbours and next nearest neighbours,

$$(\mathbf{v}\nabla)\mathbf{v} \to \sum_{i,j=-2}^{i,j=2} c_{ij}\ v_{n+i} k_n v_{n+j}, \tag{25}$$

where the sum is over neighbours and/or next nearest neighbours to n. The couplings C_{ij} in this sum should be made such that energy is conserved in the absence of diffusion. Thus, energy conservation

$$\int (v^2 + B^2) d^3x = \text{const} \tag{26}$$

now corresponds to

$$\sum_{n=1}^{n=N} (|v_n|^2 + |B_n|^2) = \text{const.} \tag{27}$$

Thus, we need to satisfy

$$\sum_{n=1}^{n=N} \left(v_n \frac{dv_n^*}{d\tilde{t}} + B_n \frac{dB_n^*}{d\tilde{t}} + \text{complex conj.} \right) = 0, \tag{28}$$

where, as before, $\tilde{t}$ is the conformal time, $\tilde{t} = \int dt/R(t)$. In this approach the vectorial character is thus lost, but the conservation of energy is kept as an essential feature.

We should now find equations for the time derivatives respecting the conservation of energy. Taking into account some factors from general relativity in an expanding universe (the expansion factor as well as the energy density and pressure) we get[14]

$$\begin{aligned} \tfrac{8}{3} dv_n/d\tilde{t} \ &= ik_n(A + C)(v_{n+1}^* v_{n+2}^* - B_{n+1}^* B_{n+2}^*) \\ &+ ik_n(B - \tfrac{1}{2}C)(v_{n-1}^* v_{n+1}^* - B_{n-1}^* B_{n+1}^*) \\ &- ik_n(\tfrac{1}{2}B + \tfrac{1}{4}A)(v_{n-2}^* v_{n-1}^* - B_{n-2}^* B_{n-1}^*), \end{aligned} \tag{29}$$

$$\begin{aligned} dB_n/d\tilde{t} \ &= ik_n(A - C)(v_{n+1}^* B_{n+2}^* - B_{n+1}^* v_{n+2}^*) \\ &+ ik_n(B + \tfrac{1}{2}C)(v_{n-1}^* B_{n+1}^* - B_{n-1}^* v_{n+1}^*) \\ &- ik_n(\tfrac{1}{2}B - \tfrac{1}{4}A)(v_{n-2}^* B_{n-1}^* - B_{n-2}^* v_{n-1}^*), \end{aligned} \tag{30}$$

where with A, B, C arbitrary constants energy is conserved. In 3+1 dimensions, magnetic helicity is also conserved. In the continuum helicity is given by

$$H = \int d^3x\ \mathbf{A}\mathbf{B}, \tag{31}$$

where **A** is the vector potential. This conservation is trivial in 2+1 dimensions, since there $H = 0$. To mimic conservation of H in the shell model we require that the quantity

$$H_{\text{shell}} = \sum_{n=1}^{n=N} (-1)^n k_n^{-1} B_n^* B_n \tag{32}$$

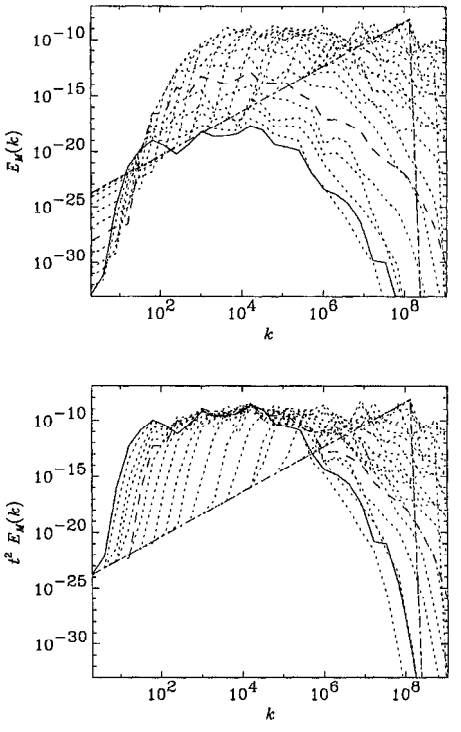


Figure 2. Spectra of the magnetic energy at different times. The straight dotted-dashed line gives the initial condition ($t_0 = 1$), the solid line gives the final time ($t = 3 \times 10^4$), and the dotted curves are for intermediate times (in uniform intervals of $\Delta \log(t - t_0) = 0.6$). $A = 1$, $B = -1/2$, and $C = 0$. This figure is from ref. 14.

is conserved. The reason is that $k_n^{-1} B_n$ is like the vector potential. The factor $(-1)^n$ is a more "phenomenological" factor. The corresponding conservation in hydrodynamics ($\int \mathbf{v}(\nabla \times \mathbf{v}) d^3x$ =const) has been studied, and it was found that the integrand oscillates in sign. This is then taken into account in the shell model by the oscillating factor.

The requirement that helicity is conserved thus corresponds to taking into account 3+1 dimensions, and it leads to the following values for the otherwise arbitrary constants A, B, C,

$$A = 1, \quad B = -1/4, \quad C = 0. \tag{33}$$

Using these values, we have $2N$ coupled set of equations. In our calculations we took N =30, corresponding to solving 60 coupled equations. The resulting spectra at different times are shown in fig. 2. Again we see a nice inverse cascade, because as functions of the comoving wave vector k the spectra clearly move towards $k = 0$.

To give a more precise picture of the change of the spectrum towards large distances, we also computed a correlation length defined by averaging over the magnetic

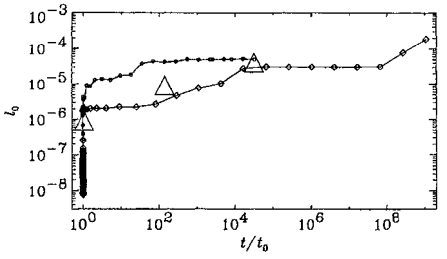


Figure 3. The correlation length (the "integral scale") as a function of time. The two curves are for slightly different models. This figure is from ref. 14.

energy density,

$$l_0 \equiv \int dk \frac{2\pi}{k} \mathcal{E}(k,t) \left[\int dk\ \mathcal{E}(k,t) \right]^{-1} . \tag{34}$$

In turbulence theory this quantity is called the "integral scale". It is a measure of the characteristic size of the largest eddies of turbulence.

The result is shown in fig. 3. We see that initially the system moves extremely rapidly towards larger scales. Clearly MHD (in the shell version) does not like the initial Gaussian random state for the magnetic field! The scaling arguments in eq. (18) predicts an increase in the eddy size like $t^{0.2}$. This cannot directly be compared to the integral scale l_0, since the integrations in eq. (33) are limited by an ultraviolet cutoff, which also becomes scaled. However, a fit in ref. 14 gives $l_0 \sim t^{0.25}$, if the steep initial increase in fig. 3 is ignored. Taking into account some uncertainty in the fitting, this is in good agreement with the scaling in eq. (18).

EFFECTS OF DIFFUSION: SILK DAMPING

The effect of diffusion has been ignored in the above discussion, except as a short distance cutoff in the numerical calculations. However, this is not realistic, as was pointed out by Siegl, Olinto, and Jedamzik[15]. This is connected to Silk damping, which occurs in the charged plasma because radiation can penetrate the plasma and carry away momentum by scattering off the charged particles. Around the time of recombination photon diffusion became very important and corresponded to a very large photon mean free path. The diffusion coefficient is proportional to the photon mean free path, and hence photon diffusion at that time cannot be ignored[15]. In a linear approximation of MHD it was clearly demonstrated that the magnetic field must be destroyed, the magnetic energy beeing turned into heat[15]. Silk diffusion would therefore remove the hope of understanding primordial magnetic fields from most points of view!

All hope is not lost, however, since the *non-linear* inverse cascade, discussed in the previous section, *counteracts Silk diffusion.* While the latter is buzy removing magnetic energy at shorter scales, the former is active in removing the energy frorn these scales to large scales. As we have seen in fig. 3, this happens very quickly. Therefore, without doing any calculations it is clear that these two mechanisms compete against one another.

To be more precise, numerical simulations are needed. This was done by Brandenburg, Enqvist, and me[16], and the result is that even if Silk diffusion is included, the

inverse cascade is strong enough to make the magnetic field survive, at least until close to recombination. This should be enough for the dynamo effect to start to operate. We refer to the original paper[16] for a full discussion of this.

CONCLUSIONS

In conclusion we mention that there are several particle physics models which can produce primordial magnetic fields. Of course, they are based on assumptions which may not turn out to be ultimately true. For example, there may not be a first order EW phase transition, superconducting or ordinary cosmic strings may not exist, etc. etc. So when the dust settles, there may not be so many mechanisms which survive. Also, it should be remembered that without the inverse cascade, there is no hope to produce large enough background fields (this does, of course, not apply to the inflationary mechanism), and it may be that for some or all of these models, the inverse cascade is not large enough.

Thus, in estimating the effect of various models one should take into account the combined effect of the inverse cascade and Silk diffusion. This will perhaps require rather complicated numerical calculations, although some results might conceivably be obtained or guessed from simple scaling arguments, as discussed in ref. 13.

Finally, we mention that there is a very interesting proposal for direct observation of a primordial background field[17]. The idea is that gamma rays arising from strong sources can scatter in a background field, making pair production and delayed photons. The spectrum of these photons could then be observed, provided the field is of order 10^{-24} G or larger[17]. If this is technically feasible, important information on the spectrum would be obtained, which could then be compared with different models.

REFERENCES

1. M. Gasperini, M. Giovannini, and G. Veneziano, *Phys. Rev. Lett.* **75:**3796 (1995); M. Giovannini, hep-th/9706201.
2. M. Turner and E. J. Weinberg, hep-th/970535
3. G. Baym, D. Bödeker, and L. McLerran, *Phys. Rev.* D **53:**662 (1996).
4. T. Kibble and A. Vilenkin, hep-ph/9704334; J. T. Ahonen and K. Enqvist, hep-ph/9704334; D. Grasso and A. Riotto, hep-ph/9707265.
5. T. Vachaspati, *Phys. Lett.* **B265:**258 (1991).
6. K. Enqvist and P. Olesen, *Phys. Lett.* **B319:**178 (1993).
7. K. Dimopoulos, hep-ph/9706513.
8. C. J. A. P. Martins and E. P. S. Shellard, astro-ph/9706287.
9. K. Enqvist and P. Olesen, *Phys. Lett.* **B329:**195 (1994).
10. E. R. Harrison, *Phys. Rev. Lett.* **30:**188 (1973); M. J. Rees, *J. R. Astr. Soc.* **28:**197 (1987); T. Vachaspati and A. Vilenkin, *Phys. Rev. Lett.* **67:** 1057 (1991).
11. M Joyce and M. Shaposhnikov, astro-ph/9703005.
12. M. S. Turner and L. M. Widrow, *Phys. Rev.* D **37:**2743 (1988).
13. P. Olesen, *Phys. Lett.* **B398:**321 (1997).
14. A. Brandenburg, K. Enqvist, and P. Olesen, *Phys. Rev.* D **54:**1291 (1996).
15. G. Siegl, A. V. Olinto, and K. Jedamzik, *Phys. Rev.* D **55:**4582 (1996)
16. A. Brandenburg, K. Enqvist, and P. Olesen, *Phys. Lett.* **B391:**395 (1997)
17. R. Plaga, *Nature* **374:**430 (1996).

TOWARDS MATRIX MODELS OF IIB SUPERSTRINGS

P. Olesen

The Niels Bohr Institute
Blegdamsvej 17
DK-2100 Copenhagen Ø
Denmark

INTRODUCTION

Recently there has been a lot of papers on matrix models and superstrings, induced by the work of Banks, Fischler, Shenker, and Susskind (hep-th/ 9706168). I refer to Makeenko's talk at this meeting for a general review of this subject.

Most of the work* reported in this talk has been done together with Fayyazuddin, Makeenko, Smith, and Zarembo[1]. As explained in Makeenko's (virtual) talk at this meeting, we started from the work by Ishibashi, Kawai, Kitazawa, and Tsuchiya[2], who proposed that type IIB superstrings in 10 dimensions are described by the reduced action,

$$S_{\rm IKKT} = \alpha \left(-\frac{1}{4}{\rm Tr}\ [A_\mu, A_\nu]^2 - \frac{1}{2}{\rm Tr}\ (\bar{\psi}\Gamma^\mu [A_\mu, \psi]) \right), \tag{1}$$

where A_μ and ψ_α are $n \times n$ matrices. A sum over n is implied, with weight $\exp(-\beta n)$. Later the sum over n has been replaced by a double scaling limit[3].

In our paper[1] we discussed various problems associated with eq. (1), and we proposed a different model with action

$$S_{\rm NBI} = -\frac{\alpha}{4}{\rm Tr}\left(Y^{-1}[A_\mu, A_\nu]^2\right) + V(Y) - \frac{\alpha}{2}{\rm Tr}\ (\bar{\psi}\Gamma^\mu [A_\mu, \psi]), \tag{2}$$

where the potential is given by

$$V(Y) = \beta\ {\rm Tr} Y + \gamma\ {\rm Tr} \ln Y. \tag{3}$$

The partition function is thus given by

$$Z = \int dA_\mu d\psi dY\ e^{-S_{\rm NBI}}. \tag{4}$$

* The paper in reference 1 has been published in Nuclear Physics B. Unfortunately, the editors of that journal used an early draft of the manuscript, which contained several typhos. For this reason I cannot recommend the published version, but refer to the version in the Archives.

New Developments in Quantum Field Theory
Edited by Damgaard and Jurkiewicz, Plenum Press, New York, 1998

We select the constant γ in such a way that the result of the Y-integration is as close to the superstring as is possible. This turns out to mean

$$\gamma = n - \frac{1}{2}, \tag{5}$$

as we shall see later.

Physically the model S_{NBI} is motivated by a GUT scenario: Suppose one has a field theory valid down to the GUT scale. Then, in our model the group is SU(n), with n large. As we shall see, this type of GUT model then leads to superstrings if $n \to \infty$. For n finite, supersymmetry is broken, as is expected for energies below the GUT energy. Thus, superstrings can emerge from a GUT type of model. Of course, the model with action S_{NBI} is not a realistic GUT model.

Under the SUSY transformations

$$\delta\psi = \frac{i}{4}\{Y^{-1}, [A_\mu, A_\nu]\}\Gamma^{\mu\nu}\epsilon, \quad \delta A_\mu = i\bar{\epsilon}\Gamma_\mu\psi, \tag{6}$$

the action transforms like

$$\delta S_{\mathrm{NBI}} \propto \epsilon^{\mu\alpha\beta\lambda_1\ldots\lambda_7}\ \mathrm{Tr}\left(\psi_m(\Gamma^0\Gamma^{11}\Gamma^{\lambda_1}\ldots\Gamma^{\lambda_7})_{mp}\epsilon_p\{[A_\alpha, A_\beta], [A_\mu, Y^{-1}]\}\right). \tag{7}$$

It can be shown that

$$\delta S_{\mathrm{NBI}} \to 0 \quad \text{for} \quad n \to \infty, \tag{8}$$

so the action is supersymmetric in the limit $n \to \infty$, but for finite n the symmetry is broken.

THE Y-INTEGRATION

The integration over Y can be done exactly. Consider

$$\mathcal{F}(z) = \int dY\ \exp\left(-\frac{\alpha}{4}\mathrm{Tr}(Y^{-1}z^2) - \beta\ \mathrm{Tr}Y - \gamma\ \mathrm{Tr}\ln Y\right), \quad z^2 \equiv -[A_\mu, A_\nu]^2. \tag{9}$$

The "angular" integration is of the Itzykson-Zuber type, so we get

$$\mathcal{F}(z) = \text{const.}\prod_{i=1}^{i=n}\int dy_i \frac{\Delta^2(y)}{\Delta(1/y)\Delta(z^2)}\ e^{-\alpha\sum_i z_i^2/4y_i - \beta\sum_i y_i - \gamma\sum_i \ln y_i}. \tag{10}$$

Here the z_i's and y_i's are the eigenvalues, and

$$\Delta(x) = \prod_{i>j}(x_i - x_j) = \det_{ki} x_i^{k-1} \tag{11}$$

is the Vandermonde determinant. We only integrate over the positive eigenvalues of Y. Thus we get

$$\Delta(z^2)\mathcal{F}(z) = \text{const.}\int_0^\infty \prod_i dy_i\ y_i^{n-1}\prod_{i>j}(y_i - y_j)\ e^{-\alpha\sum_i z_i^2/4y_i - \beta\sum_i y_i - \gamma\sum_i \ln y_i}. \tag{12}$$

This can be rewritten as a determinant

$$\begin{aligned}\Delta(z^2)\mathcal{F}(z) &= \text{const.}\ \det_{ki}\int \frac{dy}{\sqrt{y}} y^{k-1}e^{-\alpha z_i^2/4y - \beta y} \\ &= \text{const.}\ \det_{ki}\left[(-1)^{k-1}\frac{\partial^{k-1}}{\partial\beta^{k-1}}\left(\sqrt{\frac{\pi}{\beta}}\ e^{-\sqrt{\alpha\beta}z_i}\right)\right].\end{aligned} \tag{13}$$

Here z_i is by definition the positive square root of z_i^2. This determinant can be evaluated using basic properties of determinants, and the result is[1]

$$\Delta(z^2)\mathcal{F}(z) = \text{const. } \Delta(z)\ e^{-\sqrt{\alpha\beta}\ \Sigma_i z_i}. \tag{14}$$

This result is exact, and hence it is valid for any n.

The sum over eigenvalues in the exponent has the following interpretation,

$$\sum_i z_i = \sum_i \sqrt{z_i^2} = \frac{1}{4i\sqrt{\pi}} \int_{-\infty}^{(0+)} \frac{dt}{t^{3/2}} \sum_i e^{tz_i^2}, \tag{15}$$

where we used an integral representation of the square root. Thus,

$$\sum_i z_i = \frac{1}{4i\sqrt{\pi}} \int_{-\infty}^{(0+)} \frac{dt}{t^{3/2}}\ \mathrm{Tr}\ \exp(-t[A_\mu, A_\nu]^2) = \mathrm{Tr}\ \sqrt{-[A_\mu, A_\nu]^2}, \tag{16}$$

valid in Euclidean space.

The partition function therefore becomes

$$\begin{aligned} Z &= \int dA_\mu d\psi dY\ \exp\left(\frac{\alpha}{4}\mathrm{Tr}(Y^{-1}[A_\mu, A_\nu]^2) - \beta\ \mathrm{Tr}Y - \gamma\ \mathrm{Tr}\ln Y\right) \\ &= \text{const.} \int \frac{dA_\mu d\psi}{\prod_{i>j}(z_i + z_j)} \exp\left(-\sqrt{\alpha\beta}\ \mathrm{Tr}\sqrt{-[A_\mu, A_\nu]^2} - \frac{\alpha}{2}\mathrm{Tr}\ (\bar{\psi}\Gamma^\mu[A_\mu, \psi])\right). \end{aligned} \tag{17}$$

This is the exact result of the Y-integration. In order to get the square root it is important to use the value of γ given in eq. (5).

This result can be expressed in an alternative form, at the cost of introducing an auxillary Hermitean field M. We use the identity

$$\frac{1}{\prod_{i<j}(z_i + z_j)} = \text{const.}\sqrt{\det z} \int dM e^{-\mathrm{Tr}\ zM^2}, \tag{18}$$

to obtain

$$\begin{aligned} Z = \text{const.} &\int dA_\mu d\psi dM \left(\det\sqrt{-[A_\mu, A_\nu]^2}\right)^{1/2} \\ &\times \exp\left(-\mathrm{Tr}\left((\sqrt{\alpha\beta} + M^2)\sqrt{-[A_\mu, A_\nu]^2}\right) - \frac{\alpha}{2}\mathrm{Tr}\ (\bar{\psi}\Gamma^\mu[A_\mu, \psi])\right). \end{aligned} \tag{19}$$

The field M is essentially trivial, with a "classical equation of motion" $M = 0$.

ON THE WEYL REPRESENTATION AND THE APPROACH OF THE COMMUTATOR TO THE POISSON BRACKET

The square root occuring in the result above is somewhat reminisent of the Nambu-Goto square root. If we could replace the commutator in the square root in eq. (17) by the corresponding Poisson bracket, we would have a partition function which is very similar to the one for the superstring.

This problem has been discussed by Hoppe[4], and in different settings by a number of other authors [5,6,7]. It turns out that making some assumptions, one has the limit

$$\text{“ } [A, B] \to i\{A, B\}_{\mathrm{PB}} \text{ “, \quad for } \ n \to \infty. \tag{20}$$

Here $\{,\}_{PB}$ denotes the usual Poisson bracket.

We refer to the literature for a detailed discussion. Here we shall follow Bars[7], and consider a torus (although this restriction is probably not important[8]). The case of a sphere was discussed in ref. 5. A Hermitean matrix can be expanded in a Weyl basis,

$$(A_\mu)^i_j = C_1 \sum_{\mathbf{k}} a^{\mathbf{k}}_\mu (l_{\mathbf{k}})^i_j, \quad \text{with} \quad \mathbf{k} = (k_1, k_2). \tag{21}$$

The matrix $l_{\mathbf{k}}$ can be expressed in terms of the $n \times n$ (n =odd) Weyl matrices h and g, which satisfy

$$h^n = g^n = 1 \quad \text{and} \quad gh = \omega hg, \quad \text{with} \quad \omega = e^{4\pi i/n}. \tag{22}$$

The explicit form of these matrices are

$$\begin{aligned} h &= \text{diag}(1, \omega, \omega^2, \ldots, \omega^{n-1}) \\ g^i_j &= \delta^i_{j+1}, \; i = 1, 2, \ldots, n-1, \; g^n_j = 0 \text{ except for } g^n_1 = 1. \end{aligned} \tag{23}$$

The SU(n)-generators $l_{\mathbf{k}}$ are then constructed as

$$l_{\mathbf{k}} = \frac{n}{4\pi} \omega^{k_1 k_2/2} h^{k_1} g^{k_2}, \tag{24}$$

since the powers of h and g are linearly independent for $k_1, k_2 = 1, 2, \ldots, n-1$, are unitary, close under multiplication, and are traceless. Using that

$$\text{Tr } h^{k_1} g^{k_2} = n \delta_{k_1,0} \delta_{k_2,0}, \tag{25}$$

we easily see that

$$\text{Tr } l_{\mathbf{k}} l_{\mathbf{p}} = \frac{n^3}{(4\pi)^2} \delta_{\mathbf{k}+\mathbf{p},0}. \tag{26}$$

Thus the expansion coefficients in eq. (21) are given by

$$a^{-\mathbf{p}}_\mu = (a^{\mathbf{p}}_\mu)^* = \frac{(4\pi)^2}{n^3 C_1} \text{Tr } (l_{\mathbf{p}} A_\mu). \tag{27}$$

Also, from the relation

$$g^{k_1} h^{k_2} = \omega^{k_1 k_2} h^{k_2} g^{k_1} \tag{28}$$

we get

$$[l_{\mathbf{p}}, l_{\mathbf{k}}] = i \frac{n}{2\pi} \sin\left(\frac{2\pi}{n} \mathbf{p} \times \mathbf{k}\right) l_{\mathbf{p}+\mathbf{k}}, \tag{29}$$

where

$$\mathbf{p} \times \mathbf{k} = p_1 k_2 - k_1 p_2. \tag{30}$$

Using the expansion (21) we get

$$[A_\mu, A_\nu]^j_i = C_1^2 \sum_{\mathbf{p},\mathbf{q}} a^{\mathbf{p}}_\mu a^{\mathbf{q}}_\nu \frac{n}{2\pi} \sin\left(\frac{2\pi}{n} \mathbf{p} \times \mathbf{q}\right) (l_{\mathbf{p}+\mathbf{q}})^j_i. \tag{31}$$

This can be compared with the similar expression for the string variables $X_\mu(\sigma, \tau)$, where we have the expansion

$$X_\mu(\sigma, \tau) = C_2 \sum_{\mathbf{m}} a^{\mathbf{m}}_\mu \exp(i\sigma m_1 + i\tau m_2), \tag{32}$$

leading to

$$\{X_\mu, X_\nu\}_{\mathrm{PB}} = C_2^2 \sum_{\mathbf{p},\mathbf{q}} a_\mu^{\mathbf{p}} a_\nu^{\mathbf{q}} \, (\mathbf{p} \times \mathbf{q}) \exp(i(\mathbf{p}+\mathbf{q})\sigma). \tag{33}$$

Now it is clear that *if*

$$\lim_{n\to\infty} \sum_{\text{modes}} \ldots = \sum_{\text{modes}} \lim_{n\to\infty} \ldots, \tag{34}$$

then we have by use of (26)

$$\mathrm{Tr}\,[A_\mu, A_\nu]^2 \to \text{const.} \int d\sigma d\tau \{X_\mu, X_\nu\}^2 \quad \text{for} \quad n \to \infty. \tag{35}$$

It is obvious that the commutativity (34) is only valid if the infinite modes are unimportant. This is, however, not true e.g. for the bosonic string. If we fix the end points of this string at some distance, then there is a critical distance (essentially the inverse tachyon mass) at which the string oscillates so wildly that this behaviour can only be reproduced with an infinite number of modes. Below this distance the "string" becomes a branched polymer, and hence is no longer a string.

For superstrings this problem does not arize, and hence there is at least no obvious reason why the limits cannot be intechanged as in eq. (35). In the following we assume that eq. (35) is correct for type IIB superstrings.

Since we are interested in the square root of the squared commutator, the result (35) is not enough. Using eq. (26) and repeated applications of the formula

$$l_{\mathbf{m}} l_{\mathbf{r}} = (n/4\pi) \exp(2\pi i(\mathbf{m} \times \mathbf{r})/n) l_{\mathbf{m}+\mathbf{r}} \tag{36}$$

one can easily derive

$$\mathrm{Tr}\; l_{\mathbf{m}_1} l_{\mathbf{m}_2} \ldots l_{\mathbf{m}_s} \to n(n/4\pi)^s \, \delta_{\mathbf{m}_1+\mathbf{m}_2+\ldots+\mathbf{m}_s,0}, \quad \text{for } n \to \infty \tag{37}$$

to leading order in *n*. Using this in eq. (16) we obtain

$$\begin{aligned} \mathrm{Tr}\,\sqrt{-[A_\mu, A_\nu]^2} &= \frac{1}{4i\sqrt{\pi}} \int_{-\infty}^{(0+)} \frac{dt}{t^{3/2}} \mathrm{Tr}\, \exp(-t[A_\mu, A_\nu]^2) \\ &\to \frac{1}{4i\sqrt{\pi}} \int_{-\infty}^{(0+)} \frac{dt}{t^{3/2}} \int d\sigma d\tau \exp(t \cdot \text{const.} \{X_\mu, X_\nu\}_{\mathrm{PB}}^2) \\ &= \text{const.} \int d\sigma d\tau \sqrt{\{X_\mu, X_\nu\}_{\mathrm{PB}}^2} \quad \text{for } n \to \infty. \end{aligned} \tag{38}$$

Thus we see that in the leading order the Nambu-Goto square root arises as the limit of the square root of the corresponding commutator. However, it should be remembered what was said before about strings with tachyons. They do not allow the interchange of limits as in (34), and hence the result (38) is not valid in that case †.

† For such strings where the end points are actually separated by a *large* distance, the interchange of limits in (34) is probably allowed. Thus at large distances the string picture is most likely right for the NBI matrix model even without supersymmetry. At shorter distances near the critical one, this picture breaks down, and the sine function in (31) cannot be approximated by its first term in a power series expansion.

We can now summarize our results in the following rather long formula,

$$\begin{aligned} Z &= \int dA_\mu d\psi dY \ \exp\left(\frac{\alpha}{4}\mathrm{Tr}(Y^{-1}[A_\mu, A_\nu]^2) - \beta\ \mathrm{Tr}Y - (n-\frac{1}{2})\ \mathrm{Tr}\ln Y\right) \\ &= \text{const.} \int dA_\mu d\psi dM \left(\det\sqrt{-[A_\mu, A_\nu]^2}\right)^{1/2} \\ &\quad \times \exp\left(-\mathrm{Tr}\left((\sqrt{\alpha\beta}+M^2)\sqrt{-[A_\mu, A_\nu]^2}\right) - \frac{\alpha}{2}\mathrm{Tr}\ (\bar{\psi}\Gamma^\mu[A_\mu, \psi])\right) \\ &\to \text{const.} \int dX_\mu d\psi dM \left(\det\sqrt{\{X_\mu, X_\nu\}^2}\right)^{1/2} \\ &\quad \times \exp\left(-\int d\sigma d\tau \left((\sqrt{\alpha\beta}+M^2)\sqrt{\{X_\mu, X_\nu\}^2} - \frac{i\alpha}{2}(\bar{\psi}\Gamma^\mu\{X_\mu, \psi\})\right)\right), \end{aligned} \tag{39}$$

where the last expression is valid for $n \to \infty$. In this expression the fields ψ and M have expansions similar to eqs. (21) and (32), and some normalization constants have been absorbed in α and β in the last formula above.

The functional integration over M in eq. (39) is just Gaussian and can of course easily be performed,

$$\begin{aligned} z &\to \text{const.} \int dX_\mu d\psi \left(\frac{\det\sqrt{\{X_\mu, X_\nu\}^2}}{\mathcal{D}et\sqrt{\{X_\mu, X_\nu\}^2}}\right)^{1/2} \\ &\quad \times \exp\left(-\int d\sigma d\tau \left(\sqrt{\alpha\beta}\sqrt{\{X_\mu, X_\nu\}^2} - \frac{i\alpha}{2}(\bar{\psi}\Gamma^\mu\{X_\mu, \psi\})\right)\right). \end{aligned} \tag{40}$$

The two determinants in this expression arises from different types of Gaussian integrations, the "det" beeing defined through eq. (18) and the subsequent limit $n \to \infty$, whereas the *"Det"* determinant comes from the continuum integral over *M*. Naively one would tend to identify these two determinants, so that the fraction containing them is just one,

$$\frac{\det\sqrt{\{X_\mu, X_\nu\}^2}}{\mathcal{D}et\sqrt{\{X_\mu, X_\nu\}^2}} \to 1. \tag{41}$$

If so, the NBI action gives exactly the Nambu-Goto version of the Green-Schwarz type IIB superstring.

However, Zarembo [9] has pointed out to me that the situation can be more complicated. For example, from eq. (18) one sees that the "det" determinant (= det z for $n \to \infty$) is subdominant relative to the factor $\prod(z_i + z_k)$ occurring in eq. (18), and hence can be ignored in the limit $n \to \infty$. In this case, one has instead of (41)

$$\frac{\det\sqrt{\{X_\mu, X_\nu\}^2}}{\mathcal{D}et\sqrt{\{X_\mu, X_\nu\}^2}} \to \frac{1}{\mathcal{D}et\sqrt{\{X_\mu, X_\nu\}^2}}. \tag{42}$$

If so, the *Det* determinant survives in the measure. However, even if this is so, this factor is rather harmless: correlation functions are invariant, since the measure is multiplied by a constant factor under reparametrizations. This factor does not depend on the fields and cancels in the correlation functions [9].

Perhaps the right answer depends on how exactly the continuum limit is constructed, because in order to interpret *Det* a regulator is needed.

It should also be mentioned that Chekhov and Zarembo [10] have discussed models somewhat different from the NBI model, and have also discussed the measure in more details.

A SADDLE POINT AND THE VIRTUAL EULER NUMBER

We shall now study the saddle point of the NBI action. By variation of the A_μ –fields we obtain the classical equation of motion

$$[A_\mu, \{Y^{-1}, [A_\mu, A_\nu]\}] = 0. \tag{43}$$

This equation was studied by Kristjansen and me[11]. The solution is

$$[A_\mu, A_\nu]^i_j = i m_{\mu\nu}\ Y^i_j, \tag{44}$$

where $m_{\mu\nu}$ is a matrix with repect to the space indices. In the saddle point the action has the value

$$S^{\text{saddle}}_{\text{NBI}} = (\beta + m^2_{\mu\nu}\alpha/4)\ \text{Tr}Y + (n - 1/2)\ \text{Tr}\ln Y. \tag{45}$$

In order to have a non-trivial $n \to \infty$ limit, it is necessary that α and β are of order n. It should be stressed that this does *not* imply the usual classical limit in string theory, as explained in details in ref. 11.

In addition to the terms exhibited above, there are of course subdominant terms arising from the expansion of A_μ around the classical solution. These terms are ignored in the following. Therefore, at the A_μ -saddle point we have the integral[11] (α/n and β/n are both of order one)

$$Z^{\text{saddle}} = \int dY \exp[-n\{(\beta + m^2_{\mu\nu}\alpha/4)/n\ \text{Tr}Y + t\ \text{Tr}\ln Y\}],\ \text{with}\ \ t = 1 - 1/2n. \tag{46}$$

This functional integral is of the Penner type[12]. For the value of the parameter t needed in the saddle point, the Y-integral actually diverges. However, by analytic continuation á la the gamma function for negative argument one can start by defining the integral for negative t, and then ultimately continue back to positive t. It the turns out that $t = 1$ is a critical point, and in the vicinity of this point one can define a double scaling limit with the "cosmological constant"

$$\mu = (1 - t)n = \text{ fixed}. \tag{47}$$

We see that with the value of t given by eq. (46), μ =1/2=fixed quite automatically! Thus, we do not need to make any special assumptions in order to have this double scaling limit in the NBI model.

What is the meaning of the Penner model in the double scaling limit? An asymptotic expansion in μ can be made. Consider the "free energy" F, $Z^{\text{saddle}} \equiv e^F$, then

$$F(\mu) = F_0(\mu) + F_1(\mu) + \sum_{g=2}^{\infty} \chi_g \mu^{2-2g}, \quad \mu = 1/2. \tag{48}$$

Here χ_g is the "virtual Euler number" for moduli space of Riemann surfaces with genus g, which is well known to be relevant for strings. One has[12]

$$\chi_g = \frac{B_{2g}}{2g(2g-2)}, \tag{49}$$

where B_{2g} are the Bernoulli numbers. These have positive sign, and blow up factorially, so the sum defining F is not Borel summable. This is also well known to be the case for genus expansion of string theories.

The physical interpretation of this result is that the field Y captures the Euler characteristic of moduli space of Riemann surfaces. Therefore it is quite likely that the NBI model encodes non-perturbative information on Riemann surfaces generated by moduli space. It should be remembered that in string theories one usually sum the functional integral over g, however here this seems to be already included. It must be admitted that the virtual Euler number represents very global properties of moduli space, and certainly more details are needed before one can claim a good understanding of the non-perturbative nature of this model.

Recently Soloviev[13] has commented on "a curious relation" between Siegel's model[14] of random lattice strings and the above saddle point approximation to the NBI model. This comes about if one starts from Siegel's T-self-dual matrix model

$$S = \mathrm{Tr}\left(\frac{1}{2}\Phi^2 + n\ \ln(1 - g\Phi)\right), \tag{50}$$

where Φ is a Hermitean $n \times n$ matrix and g is a constant. It was then pointed out by Soloviev[13] that if one makes the substitution

$$gY = 1 - g\Phi, \tag{51}$$

and perform the limit $n \to \infty$, $g \to 0$, gn =fixed, then one obtains

$$S \to n\ \mathrm{Tr}\ (\mathrm{const.}Y + n \ln Y) + \text{irrelevant const.}. \tag{52}$$

This is, however, precisely the saddle point expression (46) for the NBI model. This sadle point is therefore a weak string coupling limit of the Siegel matrix model[13]. For arbitrary coupling there is, however, an additional Y^2-term in the Siegel action, and hence it was suggested that perhaps the potential (3) should have an additional $\mathrm{Tr}\,Y^2$ term[13]. Of course, a similar statement can be made about the NBI model, where there are various corrections to the saddle point expansion.

REFERENCES

1. A. Fayyazuddin, Y. Makeenko, P. Olesen, D. J. Smith, and K. Zarembo, hep-th/9703038.
2. N. Ishibashi, H. Kawai, Y. Kitazawa, and A. Tsuchiya, hep-th/9612115.
3. M. Fukuma, H. Kawai, Y. Kitazawa, and A. Tsuchiya, hep-th/9705128.
4. J. Hoppe, Ph. D. Thesis, MIT (1982);J. Hoppe, *Int. J. Mod. Phys.* **A 4**:5235.
5. E. G. Floratos, j. Iliopoulos, and G. Tiktopoulos, *Phys. Lett.* **B217**:285(1989).
6. D. Fairlie, P. Fletcher, and G. Zachos, *Phys. Lett.* **B218**:203 (1989);D. Fairlie and G. Zachos, *Phys. Lett.* **B224**: 101 (1989);D. Fairlie, P. Fletcher, and G. Zachos, *J. Math. Phys.* **31**:1088 (1990).
7. I. Bars, *Phys. Lett.* **B245**:35 (1990).
8. I. Bars, hep-th/9706177.
9. K. Zarembo, private communication (1997).
10. L. Chekhov and K. Zarembo, hep-th/9705014.
11. C. F. Kristjansen and P. Olesen, hep-th/9704017.
12. R. C. Penner, *J. Diff. Geom.* **27**:35 (1988).
13. O. A. Soloviev, hep-th/9707043.
14. W. Siegel, *Phys. Lett.* **B252**:558 (1990);W. Siegel, *Phys. Rev.* D **54**:2797 (1996); W. Siegel, hep-th/9603030.

QUANTUM MECHANICS OF THE ELECTRIC CHARGE

Andrzej Staruszkiewicz

Institute of Physics, Jagellonian University
Reymonta 4, 30-059 Kraków, Poland

INTRODUCTION

P.A.M. Dirac described in his last published work[1] a certain representation of the proper, ortochronous Lorentz grow up and added the following comment: "*This is the simplest example of a pathological representation of the Lorentz group. It may very well be that this pathological representation is essential for the physics of the future. Then one will be unable to make any important advance without it.*" My aim in this paper is to elucidate these prophetic words, namely to place Dirac's "pathological" representation within the established body of knowledge and to reveal its physical relevance.

GEOMETRY OF THE LIGHT CONE

To this end I have to describe the geometry of the light cone. This is a textbook material [2] but unfortunately not known among the physicists.

The light cone is a figure in space-time given by the equation

$$kk \equiv (k^0)^2 - (k^1)^2 - (k^2)^2 - (k^3)^2 = 0\,, \quad k^0 \geq 0\,.$$

Its internal geometry consists of the following three elements:

the (degenerate) metric $g_{\mu\nu}dk^\mu dk^\nu\big|_{kk=0}$;

the projective distance along the rays

$$\left.\frac{dk^0}{k^0}\right|_{k^1:k^2:k^3=\text{const}}\,;$$

the volume of the set of rays

$$d^2k = \frac{k^1[dk^2dk^3] + k^2[dk^3dk^1] + k^3[dk^1dk^2]}{k^0}\,.$$

New Developments in Quantum Field Theory
Edited by Damgaard and Jurkiewicz, Plenum Press, New York, 1998

The invariant volume

$$dk = \frac{[dk^1 dk^2 dk^3]}{k^0}$$

can be written as the outer product of the projective distance along the rays and the volume of the set of rays:

$$dk \equiv \frac{[dk^1 dk^2 dk^3]}{k^0} = \left[\frac{dk^0}{k^0} d^2k\right].$$

UNITARY IRREDUCIBLE REPRESENTATIONS OF THE PROPER, ORTOCHRONOUS LORENTZ GROUP

These representations have been discovered by Gelfand and Neumark [3] and by Harish-Chandra[4] who was a student of Dirac[5]. They consist of the *main series* and the *supplementary series* [2]. Representations from the main series are spanned by positive frequency solutions of the wave equation $\Box f = 0$, which are homogeneous of degree $-1 - iv$, where v is a real number:

$$f(\lambda x) = \lambda^{-1-iv} f(x) \quad \text{for all} \quad \lambda > 0.$$

They can be represented as

$$f(x) = \int dk e^{-ikx} f(k),$$

where the function $f(k)$ is homogeneous of degree $-1 + iv$. Hence the scalar product

$$\langle f|f\rangle = \int d^2k \overline{f(k)} f(k) = \int d^2k |f(k)|^2$$

is manifestly Lorentz invariant and positive definite. The supplementary series is obtained by putting $v = +i\sigma$, $0 < \sigma < 1$. We choose $v = +i\sigma$ to increase the real part of the degree of homogeneity *i.e.* to make the process more diffuse or less well localized in space-time. The scalar product for the supplementary series [2],

$$\langle f|f\rangle = \int \frac{d^2k d^2l}{(kl)^{1-\sigma}} \overline{f(k)} f(l),$$

is manifestly Lorentz invariant. Its positivity, however, is not obvious at all. It can be proven as follows.

Introduce the spherical coordinates

$$k^0 = \omega, \quad k^i = \omega n^i, \quad i = 1, 2, 3,$$

where n is the unit Euclidean vector orthogonal to the sphere $(k^1/k^0)^2 + (k^2/k^0)^2 + (k^3/k^0)^2 = 1$. Then

$$\begin{aligned} f(k) &= \omega^{-1-\sigma} f(n) \\ &= \omega^{-1-\sigma} \sum_{lm} f_{lm} Y_{lm}(n). \end{aligned}$$

The invariant kernel $(kl)^{\sigma-1}$ reproduces the spherical functions $Y_{lm}(n)$. Hence

$$\langle f|f\rangle = \frac{8\pi}{2^\sigma} \sum_{lm} \frac{\Gamma(l+1-\sigma)\Gamma(\sigma)}{\Gamma(l+1+\sigma)\Gamma(1-\sigma)} |f_{lm}|^2,$$

which is obviously positive definite for $0 < \sigma < 1$.

THE LIMIT $\sigma = 1$

The scalar product for the supplemantary series remains well defined for $\sigma = 1$ but ceases to be positive definite:

$$\lim_{\sigma\to 1}\langle f|f\rangle = \int d^2k d^2l \overline{f(k)} f(l) = \left|\int d^2k f(k)\right|^2 .$$

Only the spherically symmetric part has a nonvanishing norm! Assume, however, that $\int d^2k f(k) = 0$. This is a Lorentz invariant condition because for $\sigma = 1$ $f(k)$ is homogeneous of degree –2. Divide the norm $\langle f|f\rangle$ $1 - \sigma$ and work out the limit $\sigma \to 1$ from de l'Hôpital's rule; the result is

$$\begin{aligned}\lim_{\sigma\to 1}\frac{\langle f|f\rangle}{1-\sigma} &= -\int d^2k d^2l \ln(kl)\overline{f(k)} f(l) \\ &= 4\pi \sum_{\substack{lm \\ l\geq 1}} \frac{1}{l(l+1)}\left|f_{lm}\right|^2 .\end{aligned}$$

It is thus seen that for $\sigma = 1$ we have *two* scalar products: the degenerate scalar product

$$\langle f|f\rangle_{(1)} = \left|\int d^2k f(k)\right|^2$$

applicable for all states $f(k)$ and the positive definite scalar product

$$\langle f|f\rangle_{(2)} = -\int d^2k d^2l \ln(kl)\overline{f(k)} f(l)$$

applicable if and only if $\int d^2k f(k) = 0$. If $\int d^2k f(k) \neq 0$, the symbol $\langle f|f\rangle_{(2)}$ is obviously meaningless.

The representation with $\sigma = 1$ is exactly Dirac's "pathological" representation. There is, however, nothing pathological about it. It is an infinite-dimensional analogue of the Galilean structure of space-time well known from the classical Newtonian mechanics. In the Newtonian mechanics there are also *two* measures of distance: the lapse of the Newtonian time, which is applicable without any restrictions and the Euclidean distance along the hyperplanes of simultaneity; for events which are not simultaneous the Euclidean distance cannot be defined.

THE PROBLEM OF CHARGE UNIVERSALITY

The electric charges of all elementary particles are exactly the same. In the case of the electron and the proton their charges are of equal magnitude with experimental accuracy like $1 : 10^{-21}$. It is completely obvious that coincidence of two independent quantities, which holds always and with such fantastic accuracy, is a manifestation of some deep law of Nature. Recall, for example, that inertial and gravitational masses of macroscopic bodies are equal with accuracy like $1 : 10^{-12}$ and this gave rise to the General Theory of Relativity.

There are many statements in the literature on the problem of charge universality but they are not really helpful. Heisenberg said in the early thirties that to understand the charge universality it is necessary to have the theory of elementary particles, something we do not have even now [6]. Weinberg said roughly the same 50 years later during the 2nd Shelter Island conference [7].

Heisenberg might after all be right, but the argument implicit in his statement is plainly wrong. To see the fallacy in Heisenberg's argument consider another physical

quantity which is exactly the same for the electron and the proton: the spin. It is certainly curious that two completely different objects rotate always so that they produce exactly the same amount of angular momentum. Someone ignorant of the group theory of angular momentum could repeat *verbatim* Heisenberg's argument, replacing only the word "charge" by the word "spin". In reality, however, to understand the universality of spin it is enough to know elementary quantum mechanics of angular momentum, as described *e.g.* in the third chapter of the Condon and Shortley's book. The fallacy in Heisenberg's argument is thus seen to consist in this: Heisenberg assumes implicitly that the electric charge is a *dynamical* quantity, ignoring the possibility that it is, like spin, a purely *kinematical* property of space-time.

In the next section I will try to justify the following statement: *the "pathological" representation of P.A.M. Dirac is the mathematical structure underlying the electric charge.* This statement has exactly the same meaning as the statement that the SU(2) group is the mathematical structure underlying the spin.

The relevance of the "pathological" representation can be seen as follows: for $\sigma = 1$ the degree of homogeneity is exactly zero. Thus the representation describes a phenomenon completely insensitive to increase of distance from some fixed origin, which is the basic property of the electric charge as determined from the Gauss law.

THE STRUCTURE OF THE ELECTROMAGNETIC FIELD AT THE SPATIAL INFINITY

The structure of the electromagnetic field at the spatial infinity has been described by Alexander and Bergmann [8]. It can be summarized as follows. At the spatial infinity the field $F_{\mu\nu}(x)$ is homogeneous of degree –2:

$$F_{\mu\nu}(\lambda x) = \lambda^{-2} F_{\mu\nu}(x), \quad A_\mu(\lambda x) = \lambda^{-1} A_\mu(x) \text{ for all } \lambda > 0.$$

$x^\nu F_{\mu\nu}(x)$ and $(1/2)\varepsilon^{\mu\nu\rho\sigma} x_\nu F_{\rho\sigma}(x)$ are gradients of homogeneous of degree zero solutions of the wave equation:

$$x^\nu F_{\mu\nu}(x) = \partial_\mu e(x), \ (1/2)\varepsilon^{\mu\nu\rho\sigma} x_\nu F_{\rho\sigma}(x) = \partial^\mu m(x),$$
$$\Box e(x) = 0, \ e(x) = x^\mu A_\mu(x), \ \Box m(x) = 0,$$
$$e(\lambda x) = e(x), \ m(\lambda x) = m(x) \text{ for all } \lambda > 0.$$

A simple argument given in [9] shows that it is prudent to put $m(x) = 0$. In this way the following statement is seen to be true: *the electromagnetic field at the spatial infinity is completely determined by a single, homogeneous of degree zero solution of the wave equation.* This is exactly Dirac's "pathological" case.

QUANTUM MECHANICS OF THE ELECTRIC CHARGE

Basing upon the above observations I proposed some time ago [10] the following Quantum Mechanics of the Electric Charge:

$$\boxed{\begin{array}{l}[Q, S(x)] = ie, \\ S(x) = -ex^\mu A_\mu(x).\end{array}} \qquad (\hbar = c = 1)$$

Here Q is the operator of the total electric charge, e is a constant and $A_\mu(x)$ is the homogeneous of degree –1 part of Maxwell's field. This is a *closed* kinematical scheme

akin to the theory of angular momentum. It looks very simple and certainly falls short of anything, which either Heisenberg or Weinberg could have in mind. For this reason I would like to elucidate the epistemological status of both equations.

The equation $[Q, S(x)] = ie$, where $S(x)$ is a *phase i.e.* a field such that the linear combination $eA_\mu(x) + \partial_\mu S(x)$ is gauge invariant, is a theorem in QED [10]. In my present context it is simply an implicit definition of the constant e. Hence it is completely unambiguous.

The equation $S(x) = -ex^\mu A_\mu(x)$ is a hypothesis, which identifies the *phase* at the spatial infinity. It cannot be derived from anything more fundamental. Let me elucidate the nature of this hypothesis. The potential $A_\mu(x)$ is measured in the Gaussian units while the phase $S(x)$ is measured in radians. Hence there must be a constant which converts the Gaussian units into radians, just like the constant c converts seconds into centimeters. This follows from the dimensional analysis and *per se* contains no assumption at all. The hypothesis consists in the assumption that the constant e in the equation $[Q, S(x)] = ie$ is identical with the constant e in the definition of the phase $S(x) = -e x^\mu A_\mu(x)$.

Let me give some circumstantial evidence that this hypothesis is indeed true. Interaction of the classical charge Q with the potential $A_\mu(x)$ is described by the action

$$-Q \int A_\mu(x)\dot{x}^\mu ds\,.$$

Here Q is a constant *i.e.* a parameter in the Lagrangian. Assuming, in the spirit of the present theory, that Q is a dynamical variable I put it behind the sign of integral:

$$-\int QA_\mu(x)\dot{x}^\mu ds + \ldots\,.$$

Dots denote additional terms which must be there to make the action gauge invariant. I omit them since they are purely hypothetical and not relevant for my present purpose. Integrating by parts and dropping the total derivative I obtain the action

$$\int (\dot{Q}A_\mu x^\mu + Q\dot{A}_\mu x^\mu)ds\,.$$

Hence the momentum canonically conjugated with the total charge Q is

$$p_Q = -\frac{\partial L}{\partial \dot{Q}} = -x^\mu A_\mu(x)\,.$$

The minus sign must be there because I use the space-time metric with the signature $(+ - - -)$; hence $p_\mu = -\partial L/\partial \dot{x}^\mu$. Imposing the canonical commutation relation

$$[Q, p_Q] = i = [Q, -x^\mu A_\mu(x)]$$

I see that it will be consistent with the theorem $[Q, S(x)] = ie$ if $S(x) = -ex^\mu A_\mu(x)$.

The phase $S(x) = -ex^\mu A_\mu(x)$, where $A_\mu(x)$ is the homogeneous of degree -1 part of Maxwell's field, contains only infinitely slowly oscillating part of the field. Is it justified to treat such an infinitely soft object as. a quantum object satisfying the commutation relation $[Q, S(x)] = ie$? The answer is emphatically yes! Berestetsky, Lifshitz, and Pitaevsky [11] say that the field $F_{\mu\nu}(x)$ is approximately classical if ($\hbar = 1 = c$)

$$\left(\Delta x^0\right)^2 \sqrt{F_{01}^2 + F_{02}^2 + F_{03}^2} \gg 1\,.$$

Here Δx^0 is the observation time over which the field can be averaged without being significantly changed. Fields $F_{\mu\nu}(x)$ homogeneous of degree -2, which are emitted when a

charge Q changes its four-velocity, are localized entirely outside the light cone. This may be seen in each textbook of QED dealing with the so called infrared catastrophe. Hence Δx^0 is limited by the opening of the light cone: $|\Delta x^0| < 2r$, $r = \sqrt{(x^1)^2 + (x^2)^2 + (x^3)^2}$. The field is simply the Coulomb field Q/r^2 multiplied by a kinematical, velocity dependent factor, which is clearly irrelevant for the present analysis, and which can be made of order 1 by a suitable choice of the velocity change. Hence the Berestetsky, Lifshitz, and Pitaevsky inequality takes on the form

$$(2r)^2 \frac{|Q|}{r^2} \gg 1$$

i.e.

$$|Q| \gg \tfrac{1}{4},$$

which certainly does not hold for a single electron, for which

$$|Q| = \frac{1}{\sqrt{137}} < \frac{1}{4}.$$

The experimentally observed value of the fine structure constant shows beyond any doubt that the phase at the spatial infinity, $S(x) = -ex^\mu A_\mu(x)$, is not a classical object.

THE SPECTRAL CONTENT OF THE COULOMB FIELD

There are big books on the quantum mechanics of angular momentum. I could write an even bigger book on the Quantum Mechanics of the Electric Charge, because it contains the theory of angular momentum as a part. I imagine, however, that this would not be very interesting. The really interesting question is whether this theory says something about the magnitude of the constant e. It turns out that it does, although the results I have obtained up to now are of a very peculiar nature.

It is possible to associate with each four-velocity u a state $|u\rangle$, which is an eigenstate of the total charge Q, $Q|u\rangle = e\,|u\rangle$, is spherically symmetric in the rest frame of u and contains no transversal excitations. These three properties determine uniquely the state $|u\rangle$, which is the quantum counterpart of the classical Coulomb field moving with the four-velocity u. In [12] I proved the following theorem: the state $|u\rangle$, when decomposed into irreducible unitary representations of the proper, ortochronous Lorentz group, contains

— only the main series if $e^2/\pi > 1$;

— the main series and a single representation from the supplementary series with

$$C_1 = -\frac{1}{2} M_{\mu\nu} M^{\mu\nu} = \frac{e^2}{\pi}\left(2 - \frac{e^2}{\pi}\right)$$

if $0 < e^2/\pi < 1$. The scalar product, which corresponds to this particular value of the Casimir operator C_1, is

$$\langle f|f\rangle = \int \frac{d^2k d^2l}{(kl)^{e^2/\pi}} \overline{f(k)} f(l).$$

Thus e^2/π is the $(1 - \sigma)$ of mathematicians. The proof of this theorem is based on analytical properties of a new class of special functions; this might indicate that I am covering here a completely new ground.

REFERENCES

1. P. A. M. Dirac. *Int. J. Theor. Phys.* 23:677 (1984).
2. I. M. Gelfand, M. I. Graev, and N. Ya. Vilenkin, "Generalized Functions", Vol. 5, Moscow (1962).
3. I. M. Gelfand and M. A. Neumark, *Izv. Akad. Nauk SSSR* 11:411 (1947).
4. Harish-Chandra, *Proc. Roy. Soc. London*, Ser. A 189:372 (1947).
5. H. Kragh, "Dirac: a Scientific Biography", Cambridge University Press (1992). (This book should be consulted with some care as it contains an unbelievable number of misprints.)
6. I quote Heisenberg from the Russian collection of papers on Sommerfeld edited by Ya. A. Smorodinskij, Nauka, Moscow (1973), page 299. I have not seen Heisenberg's original text but I think I have rendered correctly the meaning of his words.
7. Proceedings of the 1983 Shelter Island conference on Quantum Field Theory and the Fundamental Problems of Physics, R. Jackiw, N. N. Khuri, S. Weinberg, and E. Witten, ed., The MIT Press, Cambridge (Mass.) (1985), page 24.
8. M. Alexander and P. G. Bergmann, *Found. Phys.* 14:925 (1984).
9. A. Staruszkiewicz, an essay in honour of Yakir Aharonov in "Quantum Coherence and Reality", J. S. Anandan and J. L. Safko, ed., World Scientific, Singapore (1994), page 90. See also Banach Center Publications Vol. 41, Part II, Warsaw (1997), page 257.
10. A. Staruszkiewicz, *Ann. Phys*, N.Y. 190:354 (1989).
11. W. B. Berestetsky, E. M. Lifshitz, and L. P. Pitaevsky, "Relativistic Quantum Theory", Moscow (1968), page 30.
12. A. Staruszkiewicz, *Acta Phys. Pol.* B23:591 (1992). This paper is marred by several misprints, which are there because the journal did not take into account my proofs. For corrected version see ERRATUM, *Acta Phys. Pol.* B23:959 (1992).

This page intentionally left blank

UNIVERSAL FLUCTUATIONS IN DIRAC SPECTRA

Jacobus Verbaarschot

Department of Physics and Astronomy
University at Stony Brook, SUNY
Stony Brook, NY 11794, USA

INTRODUCTION

It is generally accepted that QCD with two massless quarks undergoes a chiral phase transition (see reviews by DeTar[1], Ukawa[2] and Smilga[3] for recent results on this topic). This leads to important observable signatures in the real world with two light quarks. The order parameter of the chiral phase transition is the chiral condensate $\langle\bar{\psi}\psi\rangle$. Below the critical temperature chiral symmetry is spontaneously broken with $\langle\bar{\psi}\psi\rangle \neq 0$. One consequence is that parity doublets are absent in the hadronic spectrum. For example, the pion mass and the δ mass are very different. To a good approximation, as dictated by the Goldstone theorem, the pion is massless. According to lattice QCD simulations, chiral symmetry is restored for $T > T_c$ where $\langle\bar{\psi}\psi\rangle = 0$. For two light flavors the critical temperature is expected to be $T_c \approx 140$ *MeV*. In the restored phase, parity doublets are present, and a massive pion is degenerate with the σ-meson. It is well known that the QCD Lagrangian has two chiral symmetries: the $U_A(1)$ symmetry and the $SU(N_f) \times \mathrm{SU}(N_f)$ symmetry. As was in particular pointed out by Shuryak[4] not necessarily both symmetries are restored at the same temperature[5]. This may lead to interesting physical consequences.

According to the Banks-Casher formula[6], the chiral condensate is directly related to the average spectral density of the Dirac operator near zero virtuality. However, the eigenvalues of the Dirac operator fluctuate about their average position over the ensemble of gauge field configurations. The main question we wish to address in these lectures is to what extent such fluctuations are universal. If that is the case, they do not depend on the full QCD dynamics and can be obtained from a much simpler chiral Random Matrix Theory (chRMT) with the global symmetries of the QCD partition function.

This conjecture has its origin in the study of spectra of complex systems[7]. According to the Bohigas conjecture, spectral correlations of classically chaotic quantum systems are given by RMT. A first argument in favor of universality in Dirac spectra came from the analysis of the finite volume QCD partition function[8]. As has been shown by Gasser and Leutwyler[9], for box size L in the range

$$1/\Lambda \ll L \ll 1/m_\pi, \tag{1}$$

New Developments in Quantum Field Theory
Edited by Damgaard and Jurkiewicz, Plenum Press, New York, 1998

(Λ is a typical hadronic scale and m_π is the pion mass) the mass dependence of the QCD partition function is completely determined by its global symmetries. As a consequence, fluctuations of Dirac eigenvalues near zero virtuality are constrained by, but not determined by, an infinite family of sum rules[8] (also called Leutwyler-Smilga sum rules). For example, the simplest Leutwyler-Smilga sum rule can be obtained from the microscopic spectral density[10] (the spectral density near zero virtuality on a scale of a typical eigenvalue spacing). On the other hand, the infinite family of Leutwyler-Smilga sum rules is not sufficient to determine the microscopic spectral density. The additional ingredient is universality. A priori there is no reason that fluctuations of Dirac eigenvalues are in the same universality class as chRMT. Whether or not QCD is inside this class is a dynamical question that can only be answered by full scale lattice QCD simulations. However, the confidence in an affirmative answer to this question is greatly enhanced by universality studies within chiral Random Matrix Theory. The aim of such studies is to show that spectral fluctuations do not depend of the details of the probability distribution. Recently, it has been shown that the microscopic spectral density is universal for a wide class of probability distributions[11, 12, 13, 14, 15, 16]. We will give an extensive review of these important new results.

The fluctuations of Dirac eigenvalues near zero virtuality are directly related to the approach to the thermodynamic limit of the chiral condensate. In particular, knowledge of the microscopic spectral density provides us with a quantitative explanation[17] of finite size corrections to the valence quark mass of dependence of the chiral condensate[18].

Because of the $U_A(1)$ symmetry of the Dirac operator two types of spectral fluctuations can be distinguished. Spectral fluctuations near zero virtuality and spectral fluctuations in the bulk of the spectrum (Actually, there is a third type: spectral fluctuations near the end-points of lattice QCD Dirac spectra. However, this region of the spectrum is completely unphysical, and it will not be considered in these lectures.)

Recently, it has become possible to obtain *all* eigenvalues of the lattice QCD Dirac operator on reasonably large lattices[19, 20], making a direct verification of the above conjecture possible. This is one of the main objectives of these lectures. This is easiest for correlations in the bulk of the spectrum. Under the assumption of spectral ergodicity[21], eigenvalue correlations can be studied by spectral averaging instead of ensemble averaging[22, 23]. On the other hand, in order to study the microscopic spectral density, a very large number of independent gauge field configurations is required. First lattice results confirming the universality of the microscopic spectral density have been obtained recently[20].

At this point I wish to stress that there are two different types of applications of Random Matrix Theory. In the first type, fluctuations of an observable are related to its average. Because of universality it is possible to obtain exact results. In general, the average of an observable is not given by Random Matrix Theory. There are many examples of this type of universal fluctuations ranging from atomic physics to quantum field theory (a recent comprehensive review was written by Guhr, Müller-Groeling and Weidenmüller[24]). Most of the examples are related to fluctuations of eigenvalues. Typical examples are nuclear spectra[25], acoustic spectra[26], resonances in resonance cavities[27], S-matrix fluctuations[28, 29] and universal conductance fluctuations[30]. In these lectures we will discuss correlations in the bulk of Dirac spectra and the microscopic spectral density. The second type of application of Random Matrix Theory is as a schematic model of disorder. In this way one obtains *qualitative* results which may be helpful in understanding some physical phenomena. There are numerous examples in this category. We only mention the Anderson model of localization[31], neural networks[32], the Gross-Witten model of QCD[33] and quantum gravity[34]. In these lectures we will

discuss chiral random matrix models at nonzero temperature and chemical potential. In particular, we will review recent work by Stephanov[35] on the quenched approximation at nonzero chemical potential.

At nonzero chemical potential the QCD Dirac operator is nonhermitean with eigenvalues scattered in the complex plane. As was first pointed out by Fyodorov *et al.*[36], this leads to the possibility of a new type of universal behavior. Characteristic features of Dirac spectra will be discussed at the end of this lecture. For a review of nonhermitean random matrices, we refer to the talk by Nowak[37] in these proceedings.

In the first lecture we will review some general properties of Dirac spectra including the Banks-Casher formula. From the zeros of the partition function we will show that there is an intimate relation between chiral symmetry breaking and correlations of Dirac eigenvalues. Starting from Leutwyler-Smilga sum-rules the microscopic spectral density will be introduced. We will discuss the statistical analysis of quantum spectra. It will be argued that spectral correlations of 'complex' systems are given by Random Matrix Theory. We will end the first lecture with the introduction of chiral Random Matrix Theory.

In the second lecture we will compare the chiral random matrix model with QCD and discuss some of its properties. We will review recent results that show that the microscopic spectral density and eigenvalue correlations near zero virtuality are strongly universal. Lattice QCD results for the microscopic spectral density and correlations in the bulk of the spectrum will be discussed in detail. We will end the second lecture with a review of chiral Random Matrix Theory at nonzero chemical potential. New features of spectral universality in nonhermitean matrices will be discussed.

THE DIRAC SPECTRUM

Introduction

The Euclidean QCD partition function is given by

$$Z_{\rm QCD}(m,\theta) = \int dA \det(\gamma D + m) e^{-S_{\rm YM}/\hbar + i\theta\nu}, \tag{2}$$

where $\gamma D = \gamma_\mu (\partial_\mu + iA_\mu)$ is the anti-Hermitean Dirac operator and $S_{\rm YM}$ is the Yang-Mills action. The integral over field configurations includes a sum over all topological sectors with topological charge ν. Each sector is weighted by $\exp(i\theta\nu)$. Phenomenologically the value of the vacuum θ-angle is consistent with zero. We use the convention that the Euclidean gamma matrices are Hermitean with $\{\gamma_\mu, \gamma_\nu\} = 2\delta_{\mu\nu}$. The integral is over all gauge field configurations, and for definiteness, we assume a lattice regularization of the partition function.

Our main object of interest is the spectrum of the Dirac operator. The eigenvalues λ_k are defined by

$$\gamma D \phi_k = i\lambda_k \phi_k. \tag{3}$$

The spectral density is given by

$$\rho(\lambda) = \sum_k \delta(\lambda - \lambda_k). \tag{4}$$

Correlations of the eigenvalues can be expressed in terms of the two-point correlation function

$$\rho_2(\lambda,\lambda') = \langle \rho(\lambda)\rho(\lambda') \rangle, \tag{5}$$

where $\langle \ldots \rangle$ denotes averaging with respect to the QCD partition function (2). The connected two-point correlation function is obtained by subtraction of the product of the average spectral densities

$$\rho_c(\lambda, \lambda') = \rho_2(\lambda, \lambda') - \langle \rho(\lambda) \rangle \langle \rho(\lambda') \rangle. \tag{6}$$

Because of the $U_A(1)$ symmetry

$$\{\gamma_5, \gamma D\} = 0, \tag{7}$$

the eigenvalues occur in pairs $\pm\lambda$ or are zero. The eigenfunctions are given by ϕ_k and $\gamma_5\phi_k$, respectively. If $\gamma_5\phi_k = \pm\phi_k$, then necessarily $\lambda_k = 0$. This happens for a solution of the Dirac operator in the field of an instanton. In a sector with topological charge ν the Dirac operator has ν exact zero modes with the same chirality. In order to represent the low energy sector of the Dirac operator for field configurations with topological charge ν, it is natural to choose a chiral basis with n right-handed states and $m \equiv n + \nu$ left-handed states. Then the Dirac matrix has the block structure

$$\begin{pmatrix} 0 & T \\ T^\dagger & 0 \end{pmatrix}, \tag{8}$$

where T is an $n \times m$ matrix. For $m = 2$ and $n = 1$, one can easily convince oneself that the Dirac matrix has exactly one zero eigenvalue. We leave it as an exercise to the reader to show that in general the Dirac matrix has $|m - n|$ zero eigenvalues.

In terms of the eigenvalues of the Dirac operator, the QCD partition function can be rewritten as

$$Z_{\rm QCD}(m, \theta) = \sum_\nu e^{i\nu\theta} \prod_f m_f^{|\nu|} \int_\nu dA \prod_f \prod_k (\lambda_k^2 + m_f^2) e^{-S_{YM}/\hbar} \tag{9}$$

where $\int_\nu dA$ denotes the integral over field configurations with topological charge ν, and $\prod_f$ is the product over N_f flavors with mass m_f. The partition function in the sector of topological charge ν is obtained by Fourier inversion

$$Z_\nu(m) = \frac{1}{2\pi} \int_0^{2\pi} d\theta e^{-i\nu\theta} Z_{\rm QCD}(m, \theta). \tag{10}$$

The fluctuations of the eigenvalues of the QCD Dirac operator are induced by the fluctuations of the gauge fields. Formally, one can think of integrating out all gauge field configurations for fixed values of the Dirac eigenvalues. The transformation of integration variables from the fields, A, to the eigenvalues, λ_k, leads to a nontrivial "Jacobian". Universality in Dirac spectra has its origin in this "Jacobian".

The free Dirac spectrum can be obtained immediately from the square of the Dirac operator. For a box of volume $L_1 \times L_2 \times L_3 \times L_4$ one finds

$$\lambda_{\vec{n}} = 2\pi \left((\frac{n_1}{L_1})^2 + (\frac{n_2}{L_2})^2 + (\frac{n_3}{L_3})^2 + (\frac{n_4 + 0.5}{L_4})^2 \right)^{1/2} \tag{11}$$

where we have used periodic boundary conditions in the spatial directions and anti-periodic boundary conditions in the time direction. The spectral density is obtained by counting the total number of eigenvalues in a sphere of radius $\lambda L/2\pi$. The result is

$$\rho^{\rm free}(\lambda) \sim V\lambda^3. \tag{12}$$

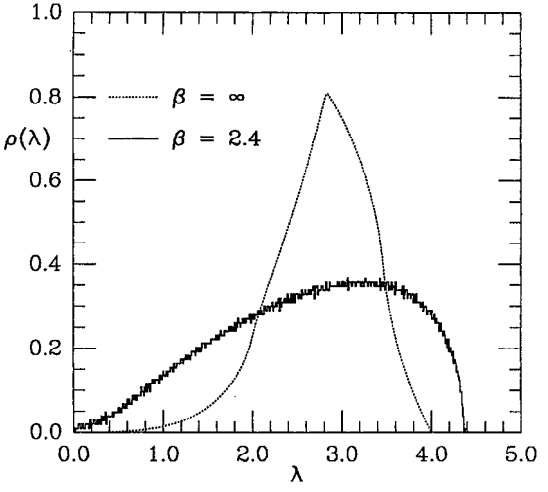


Figure 1. The free Dirac spectral density (dotted curve) and the spectral density of the Kogut-Susskind Dirac operator for a gauge field configuration generated with $\beta = 2.4$ (histogram). Both spectral densities are on a 12^4 lattice and are normalized to unit area.

For future reference, we note that in the generic case, when the sides of the hypercube are related by an irrational number, the eigenvalues are uncorrelated, i.e.

$$\rho_2(\lambda, \lambda') = \langle \rho(\lambda) \rangle \langle \rho(\lambda') \rangle. \tag{13}$$

Two examples of Dirac spectra are shown in Fig. 1. The dotted curve represents the free Kogut-Susskind Dirac spectrum on a 12^4 lattice with periodic boundary conditions in the spatial directions anti-periodic boundary conditions in the time direction. For an $N_1 \times N_2 \times N_3 \times N_4$ lattice this spectrum is given by

$$\lambda_{\vec{n}} = 2\left(\sin^2\left(\frac{\pi n_1}{N_1}\right) + \sin^2\left(\frac{\pi n_2}{N_2}\right) + \sin^2\left(\frac{\pi n_3}{N_3}\right) + \sin^2\left(\frac{\pi(n_4 + 0.5)}{N_4}\right)\right)^{1/2}. \tag{14}$$

Here, $n_i = 0,1, \ldots, [N_i/2]$ ($i = 1, 2, 3$) and $n_4 = 0, 1, \ldots, [(N_4 + 1)/2]$. The Kogut-Susskind Dirac spectrum for an $SU(2)$ gauge field configuration with $\beta = 2.4$ on the same size lattice is shown by the histogram in the same figure (full curve). We clearly observe an accumulation of small eigenvalues.

The Banks-Casher Relation

The order parameter of the chiral phase transition, $\langle \bar{\psi}\psi \rangle$, is nonzero only below the critical temperature. As was shown by Banks and Casher[6], $\langle \bar{\psi}\psi \rangle$ is directly related to the eigenvalue density of the QCD Dirac operator per unit four-volume

$$\Sigma \equiv |\langle \bar{\psi}\psi \rangle| = \lim \frac{\pi \langle \rho(0) \rangle}{V}. \tag{15}$$

It is elementary to derive this relation. The chiral condensate follows from the partition function (9) (all quark mass are chosen equal),

$$\langle \bar{\psi}\psi \rangle = -\lim \frac{1}{V N_f} \partial_m \log Z(m) = -\lim \frac{1}{V} \langle \sum_k \frac{2m}{\lambda_k^2 + m^2} \rangle. \tag{16}$$

If we express the sum as an integral over the average spectral density, and take the thermodynamic limit before the chiral limit, so that many eigenvalues are less than m, we recover (15). The order of the limits in (15) is important. First we take the thermodynamic limit, next the chiral limit and, finally, the field theory limit. As can be observed from (16) the sign of $\langle\bar{\psi}\psi\rangle$ changes if m crosses the imaginary axis.

An important consequence of the Bank-Casher formula (15) is that the eigenvalues near zero virtuality are spaced as

$$\Delta\lambda = 1/\rho(0) = \pi/\Sigma V. \tag{17}$$

This should be contrasted with the eigenvalue spectrum of the non-interacting Dirac operator. Then one obtains from (12) an eigenvalue spacing equal to $\Delta\lambda \sim 1/V^{1/4}$. Clearly, the presence of gauge fields leads to a strong modification of the spectrum near zero virtuality. Strong interactions result in the coupling of many degrees of freedom leading to extended states and correlated eigenvalues. On the other hand, for uncorrelated eigenvalues, the eigenvalue distribution factorizes, and for $\lambda \neq 0$, we have $\rho(\lambda) \sim \lambda^{2N_f+1}$ in the chiral limit, i.e. no breaking of chiral symmetry. One consequence of the interactions is level repulsion of neighboring eigenvalues. Therefore, the two smallest eigenvalues of the Dirac operator, $\pm\lambda_{\min}$ repel each other, and the Dirac spectrum will have a gap at $\lambda = 0$ with a width of the order of $1/\Sigma V$.

Spectral Correlations and Zeros of the Partition Function

The study of zeros of the partition function has been a fruitful tool in statistical mechanics [38, 39]. In QCD, both zeros in the in the complex fugacity plane and the complex mass plane have been studied[40, 41]. Since the QCD partition function is a polynomial in m it can be factorized as (all quark masses are taken to be equal to m)

$$Z_{\rm QCD}(m,\theta) = \prod_k (m - m_k). \tag{18}$$

Because configurations of opposite topological charge occur with the same probability, the coefficients of this polynomial are real, and the zeros appear in complex conjugate pairs. For an even number of flavors the zeros occur in pairs $\pm m_k$. In a sector with topological charge V, this is also the case for even $N_f \times v$. The chiral condensate is given by

$$\Sigma(m) = -\lim \frac{1}{VN_f}\partial_m \log Z_{\rm QCD}(m,\theta) = -\lim \frac{1}{VN_f}\sum_k \frac{1}{m - m_k}. \tag{19}$$

For an even number of flavors, $\Sigma(m)$ is an odd function of m. In order to have a discontinuity at $m = 0$, the zeros in this region have to coalesce into a cut along the imaginary axis in the thermodynamic limit.

In the hypothetical case that the eigenvalues of the Dirac operator do not fluctuate the zeros are located at $m_k = \pm i\lambda_k$. In the opposite case, of uncorrelated eigenvalues, the eigenvalue distribution factorizes and all zeros are located at $\pm i\sigma$, where $\sigma^2 = \langle\lambda_k^2\rangle$. As a result, the chiral condensate does not show a discontinuity across the imaginary axis and is equal to zero.

We hope to convince the reader that the presence of a discontinuity is intimately related to correlations of eigenvalues[42]. Let us study the effect of pair correlations for one flavor in the sector of zero topological charge. The fermion determinant can be

written as

$$\langle \prod_k (m^2 + \lambda_k^2) \rangle = \sum_k \begin{pmatrix} N \\ k \end{pmatrix} m^{2(N-k)} \langle \lambda_1^2 \cdots \lambda_k^2 \rangle. \tag{20}$$

There are

$$\begin{pmatrix} k \\ 2l \end{pmatrix} (2l-1)!!$$

ways of selecting l pairs from $\lambda_1^2 \ldots \lambda_k^2$. The average of each pair of different eigenvalues is given by

$$\langle \lambda_m^2 \lambda_n^2 \rangle = \sigma^4 + C_2, \tag{21}$$

where σ^2 is the expectation value of λ_k^2 and C_2 is the connected correlator

$$C_2 = \langle \lambda_m^2 \lambda_n^2 \rangle - \langle \lambda_m^2 \rangle \langle \lambda_n^2 \rangle, \qquad m \neq n. \tag{22}$$

This result in the partition function

$$Z(m) = \sum_{k=0}^{N} \sum_{l=0}^{[\frac{k}{2}]} m^{2(N-k)} \begin{pmatrix} N \\ k \end{pmatrix} \begin{pmatrix} k \\ 2l \end{pmatrix} (2l-1)!! C_2^l \sigma^{2(k-2l)}. \tag{23}$$

After interchanging the two sums, one can easily show that Z(m) can be expressed as a multiple of a Hermite polynomial

$$Z(m) = (-\frac{C_2}{2})^{N/2} \mathrm{H}_N((\sigma^2 + m^2)/\sqrt{-2C_2}). \tag{24}$$

In terms of the zeros of the Hermite polynomials, z_k, the zeros of the partition function are located at

$$m_k^2 = z_k \sqrt{-2C_2} - \sigma^2. \tag{25}$$

Asymptotically, the zeros of the Hermite polynomials are given by $z_k \approx \pi k / 2\sqrt{N}$. In order for the zeros to join into a cut in the thermodynamic limit, they have to be spaced as $\sim 1/N$. This requires that

$$C_2 \sim -\frac{1}{N}. \tag{26}$$

The density of zeros is then given by

$$\frac{dk}{dm} \sim Nm. \tag{27}$$

We conclude that pair correlations are sufficient to generate a cut of $Z(m)$ in the complex m-plane, but the chiral symmetry remains unbroken. Pair correlations alone cannot suppress the effect of the fermion determinant.

Leutwyler-Smilga Sum Rules

We have shown that pair-correlations are not sufficient to generate a discontinuity in the chiral condensate. In this subsection we start from the assumption that chiral symmetry is broken spontaneously, and look for consistency conditions this imposes on the Dirac spectrum. As has been argued by Gasser and Leutwyler[9] and Leutwyler

and Smilga[8], in the mesoscopic range (1), the mass dependence of the QCD partition function is given by (for simplicity, all quark masses have been taken equal)

$$Z^{\text{eff}}(m,\theta) \sim \int_{U\in G/H} dU e^{mV\Sigma \text{Re}\,\text{Tr}\, U e^{i\theta/N_f}}. \tag{28}$$

The integral is over the Goldstone manifold associated with chiral symmetry breaking from G to H. For three or more colors with fundamental fermions $G/H = SU(N_f) \times SU(N_f)/SU(N_f)$. The finite volume partition function in the sector of topological charge ν follows by Fourier inversion according to (10). The partition function for $\nu = 0$ is thus given by (28) with the integration over $SU(N_f)$ replaced by an integral over $U(N_f)$. The case of $N_f = 1$ is particularly simple. Then only a $U(1)$ integration remains, and the partition function is given by[8] $Z^{\text{eff}}_{\nu=0}(m) = I_0(mV\Sigma)$. Its zeros are regularly spaced along the imaginary axis in the complex m-plane, and, in the thermodynamic limit, they coalesce into a cut.

The Leutwyler-Smilga sum-rules are obtained by expanding the partition function $Z_\nu(m)$ in powers of m before and after averaging over the gauge field configurations and equating the coefficients. This corresponds to an expansion in powers of m of both the QCD partition function (2) and the finite volume partition function (28) in the sector of topological charge ν. As an example, we consider the coefficients of m^2 in the sector with $\nu = 0$. This results in the sum-rule

$$\langle {\sum_k}' \frac{1}{\lambda_k^2} \rangle = \frac{\Sigma^2 V^2}{4N_f}, \tag{29}$$

where the prime indicates that the sum is restricted to nonzero positive eigenvalues.

The next order sum rules are obtained by equating the coefficients of order m^4. They can be combined into

$$\langle {\sum_{k,l}}' \frac{1}{\lambda_k^2\lambda_l^2} \rangle - \langle {\sum_k}' \frac{1}{\lambda_k^2} \rangle \langle {\sum_l}' \frac{1}{\lambda_l^2} \rangle = \frac{\Sigma^4 V^4}{16N_f^2(N_f^2-1)}. \tag{30}$$

We conclude that chiral symmetry breaking leads to correlations of the inverse eigenvalues. However, if one performs an analysis similar to the one in previous section, it can be shown easily that pair correlations given by (30) do not result in a cut in the complex m-plane. Apparently, chiral symmetry breaking requires a subtle interplay of all types of correlations.

For two colors with fundamental fermions or for adjoint fermions the pattern of chiral symmetry breaking is different. Sum rules for the inverse eigenvalues can be derived along the same lines. The general expression for the simplest sum-rule can be summarized as[43, 44]

$$\langle {\sum_k}' \frac{1}{\lambda_k^2} \rangle = \frac{\Sigma^2 V^2}{4(|\nu| + (\dim(G/H)+1)/N_f)}. \tag{31}$$

The Leutwyler-Smilga sum-rules can be expressed as an integral over the average spectral density and spectral correlation functions. For the sum rule (29) this results in

$$\frac{1}{V^2\Sigma^2} \int \frac{\langle \rho(\lambda) \rangle d\lambda}{\lambda^2} = \frac{1}{4N_f}. \tag{32}$$

If we introduce the microscopic variable

$$u = \lambda V \Sigma, \tag{33}$$

this integral can be rewritten as

$$\int \frac{1}{V\Sigma}\langle\rho(\frac{u}{V\Sigma})\rangle\frac{du}{u^2} = \frac{1}{4N_f}. \tag{34}$$

The thermodynamic limit of the combination that enters in the integrand,

$$\rho_S(u) = \lim_{V\to\infty}\frac{1}{V\Sigma}\langle\rho(\frac{u}{V\Sigma})\rangle, \tag{35}$$

will be called the microscopic spectral density[10]. This limit exists if chiral symmetry is broken. Our conjecture is that $\rho s(u)$ is a universal function that only depends on the global symmetries of the QCD partition function. Because of universality it can be derived from the simplest theory with the global symmetries of the QCD partition function. Such theory is a chiral Random Matrix Theory which will be introduced later in these lectures.

We emphasize again that the U_A (1) symmetry of the QCD Dirac spectrum leads to two different types of eigenvalue correlations: spectral correlations in the bulk of the spectrum and spectral correlations near zero virtuality. The simplest example of correlations of the latter type is the microscopic spectral density defined in (35).

We close this subsection with two unrelated side remarks. First, the QCD Dirac operator is only determined up to a constant matrix. We can exploit this freedom to obtain a Dirac operator that is maximally symmetric. For example, the Wilson lattice QCD Dirac operator, D^W, is neither Hermitean nor anti-Hermitean, but $\gamma_5 D^W$ is Hermitean.

Second, the QCD partition function can be expanded in powers of m^2 before or after averaging over the gauge field configurations. In the latter case one obtains sum rules for the inverse zeros of the partition function. As an example we quote,

$$\sum \frac{1}{m_k^2}\Big|_{\nu=0} = \frac{\Sigma^2 V^2}{4}, \tag{36}$$

where we have averaged over field configurations with zero topological charge.

SPECTRAL CORRELATIONS IN COMPLEX SYSTEMS

Statistical Analysis of Spectra

Spectra for a wide range of complex quantum systems have been studied both experimentally and numerically (a excellent recent review has been given by Guhr, Müller-Groeling and Weidenmüller[24]). One basic observation is that the scale of variations of the average spectral density and the scale of the spectral fluctuations separate. This allows us to unfold the spectrum, i.e. we rescale the spectrum in units of the local average level spacing. Specifically, the unfolded spectrum is given by

$$\lambda_k^{\rm unf} = \int_\infty^{\lambda_k}\langle\rho(\lambda')\rangle d\lambda', \tag{37}$$

with unfolded spectral density

$$\rho_{\rm unf}(\lambda) = \sum_k \delta(\lambda - \lambda_k^{\rm unf}). \tag{38}$$

The fluctuations of the unfolded spectrum can be measured by suitable statistics. We will consider the nearest neighbor spacing distribution, P(S), and moments of the

number of levels in an interval containing n levels on average. In particular, we will consider the number variance, $\Sigma_2(n)$, and the first two cumulants, $\gamma_1(n)$ and $\gamma_2(n)$. Another useful statistic is the $\Delta_3(n)$-statistic introduced by Dyson and Mehta[45]. It is related to $\Sigma_2(n)$ via a smoothening kernel. The advantage of this statistic is that its fluctuations as a function of n are greatly reduced. Both $\Sigma_2(n)$ and $\Delta_3(n)$ can be obtained from the pair correlation function defined as

$$Y_2(\lambda, \lambda') = -\langle \rho_{\text{unf}}(\lambda)\rho_{\text{unf}}(\lambda')\rangle + \langle \rho_{\text{unf}}(\lambda)\rangle\langle \rho_{\text{unf}}(\lambda')\rangle \tag{39}$$

Analytical expressions for the above statistics can be obtained for the eigenvalues of the invariant random matrix ensembles. They are defined as ensembles of Hermitean matrices with Gaussian independently distributed matrix elements, i.e. with probability distribution given by

$$P(H) \sim e^{-\frac{N\beta}{2}\mathrm{Tr}H^\dagger H}. \tag{40}$$

Depending on the anti-unitary symmetry, the matrix elements are real, complex or quaternion real. They are called the Gaussian Orthogonal Ensemble (GOE), the Gaussian Unitary Ensemble (GUE) and the Gaussian Symplectic Ensemble (GSE), respectively. Each ensemble is characterized by its Dyson index β which is defined as the number of independent variables per matrix element. For the GOE, GUE and the GSE we thus have $\beta = 1$, 2 and 4, respectively.

Independent of the value of β, the average spectral density is the semicircle,

$$\langle \rho(\lambda)\rangle = \frac{N}{\pi}\sqrt{2-\lambda^2}. \tag{41}$$

Analytical results for all spectral correlation functions have been derived for each of the three ensembles[46] via the orthogonal polynomial method. We only quote the most important results. The nearest neighbor spacing distribution, which is known exactly in terms of a power series, is well approximated by

$$P(S) \sim S^\beta \exp(-a_\beta S^2), \tag{42}$$

where a_β is a constant of order one. The asymptotic behaviour of the pair correlation functions is given by[46]

$$Y_2(\lambda, \lambda') \sim \frac{1}{\pi^2(\lambda-\lambda')^2} \quad \text{for} \quad \beta = 1, \tag{43}$$

$$Y_2(\lambda, \lambda') \sim \frac{\sin^2 \pi(\lambda-\lambda')}{\pi^2(\lambda-\lambda')^2} \quad \text{for} \quad \beta = 2, \tag{44}$$

$$Y_2(\lambda, \lambda') \sim -\frac{\cos 2\pi(\lambda-\lambda')}{4(\lambda-\lambda')} + \frac{1+(\pi/2)\sin 2\pi(\lambda-\lambda')}{4\pi^2(\lambda-\lambda')^2} \quad \text{for} \quad \beta = 4. \tag{45}$$

The $1/(\lambda-\lambda')^2$ tail of the pair correlation function results in a logarithmic dependence of the asymptotic behavior of $\Sigma_2(n)$ and $\Delta_3(n)$,

$$\Sigma_2(n) \sim (2/\pi^2\beta)\log n \quad \text{and} \quad \Delta_3(n) \sim \beta\Sigma_2(n)/2. \tag{46}$$

Characteristic features of random matrix correlations are level repulsion at short distances and a strong suppression of fluctuations at large distances.

For uncorrelated eigenvalues the level repulsion is absent and one finds

$$P(S) = \exp(-S), \tag{47}$$

and

$$\Sigma_2(n) = n \quad \text{and} \quad \Delta_3(n) = n/15. \tag{48}$$

Spectral Universality

The main conclusion of numerous studies of eigenvalue spectra of complex systems is that spectral correlations of classically chaotic systems are given by RMT[24]. As illustration we mention three examples from completely different fields. The first example is the nuclear data ensemble in which the above statistics are evaluated by superimposing level spectra of many different nuclei[25]. The second example concerns correlations of acoustic resonances in irregular quartz blocks[26]. In both cases the statistics that were considered are, within experimental accuracy, in complete agreement with the GOE statistics. The third example pertains to the zeros of Riemann's zeta function, Extensive numerical calculations[47] have shown that asymptotically, for large imaginary part, the correlations of the zeros are given by the GUE.

The Gaussian random matrix ensembles introduced above can be obtained[46] from two assumptions: i) The probability distribution is invariant under unitarity transformations; ii) The matrix elements are statistically independent. If the invariance assumption is dropped it can be shown with the theory of free random variables[48] that the average spectral density is still given by a semicircle if the variance of the probability distribution is finite. For example, if the matrix elements are distributed according to a rectangular distribution, the average spectral density is a semicircle. On the other hand, if the independence assumption is released the average spectral density is typically not a semicircle. For example, this is the case if the quadratic potential in the probability distribution is replaced by a more complicated polynomial potential $V(H)$. Using the supersymmetric method for Random Matrix Theory, it was shown by Hackenbroich and Weidenmüller[49] that the same supersymmetric nonlinear σ-model is obtained for a wide range of potentials $V(H)$. This implies that spectral correlations of the unfolded eigenvalues are independent of the potential. Remarkably, this result could be proved for all three values of the Dyson index.

An explicit construction of the correlation functions using orthogonal polynomials could only be performed[50, 51, 52, 53] for $\beta = 2$. Several examples have been considered where both the invariance assumption and the independence assumption are relaxed. We mention $H \to H + A$, where A is an arbitrary fixed matrix, and the probability distribution of H is given by a polynomial $V(H)$. It was shown by P. Zinn-Justin[54] that also in this case the spectral correlations are given by the invariant random matrix ensembles. For a Gaussian probability distribution the proof was given by Brézin and Hikami[55].

The domain of universality has been extended in the direction of real physical systems by means of the Gaussian embedded ensembles[56, 57]. The simplest example is the ensemble of matrix elements of n-particle Slater determinants of a two-body operator with random two-particle matrix elements. It can be shown analytically that the average spectral density is a Gaussian. However, according to substantial numerical evidence, the spectral correlations are in complete agreement with the invariant random matrix ensembles[56].

Universal spectral correlations are obtained in the thermodynamic (or semi-classical) limit, $N \to \infty$, with $(E - E')N$ fixed. Alternatively, one can take the thermodynamic limit with $E - E'$ fixed. This leads to the Ambjorn-Jurkiewicz-Makeenko (AJM-) universality[58] for smoothened correlation functions[52]. They are obtained from the exact correlation functions by replacing the oscillating terms by their average over a scale much larger than $1/N$, but small compared to the secular variation of the average spectral density. The result for the two-point correlation function is given by a smoothening of the leading order asymptotic result[52]. However, correlations at macroscopic distance

in energy should include another ingredient. Namely, correlations resulting from the compactness of the support of the spectrum. Indeed, Ambjorn, Jurkiewicz and Makeenko showed[58] that the smoothened correlation functions for an arbitrary potential are determined completely by the endpoints of the spectrum. This theorem was proved for all three ensembles by Beenakker[60] and it was generalized to multi-cut potentials by Akemann and Ambjorn[59]. A general expression for the smoothened correlation function in terms of one-point Greens functions has been derived for different deformations of the invariant random matrix ensembles[56, 61, 62]. An interesting question is whether correlations given by this general expression are also found in real physical systems. There are several indications that the answer to this question is negative. First of all, AJM universality is closely related to the compactness of the support of the spectrum. In real physical systems, the spectral density usually increases with energy so that the average resolvent does not even exist (only the difference of two resolvents enters in the general expression). Second, in fully chaotic systems, with microscopic correlations given by Random Matrix Theory, it was shown by Berry[63] that the asymptotics of the two-point correlation function, as measured by the Δ_3-statistic, is determined by the shortest periodic trajectory.

Smoothened correlators are obtained via a perturbative expansion of the correlation functions. In Random Matrix Theory this is equivalent to a loop expansion in $1/N$. The full non-perturbative result cannot be obtained this way. That requires the use of orthogonal polynomials or the super-symmetric method of Random Matrix Theory.

Another type of universal behavior is given by correlations in the neighborhood of the largest eigenvalue. It was shown by Kanzieper and Freilikher[64] that for an arbitrary potential the spectral correlations at the soft edge of the spectrum are given by the Airy kernel. However, we are not aware of any physical applications of such correlations. Certainly, this type of universality is restricted to systems with a resolvent that is defined by a finite cuts on the real axis. For example, for the embedded ensembles with a Gaussian spectral density, we do not expect such type of universality.

We have seen a large number of examples that fall into one of universality classes of the invariant random matrix ensembles. This calls out for a more general approach. Naturally, one thinks in terms of the renormalization group. This approach was pioneered by Brézin and Zinn-Justin[65]. The idea is to integrate out rows and columns of a random matrix and to show that the Gaussian ensembles are a stable fixed point. This was made more explicit in a paper by Higuchi *et al.*[66]. However, much more work is required to arrive at a natural proof of spectral universality.

Although the above mentioned universality studies provide support for the validity of the Bohigas conjecture, the ultimate goal is to derive it directly from the underlying classical dynamics. An important first step in this direction was made by Berry[63]. He showed that the asymptotics of the two-point correlation function is related to sum-rules for isolated classical trajectories. Another interesting approach was introduced by Andreev *et al.*[67] who where able to obtain a supersymmetric nonlinear sigma model from spectral averaging. In this context we also mention the work of Altland and Zirnbauer[68] who showed that the kicked rotor can be mapped onto a supersymmetric sigma model.

CHIRAL RANDOM MATRIX THEORY

Introduction of the Model

In this section we will introduce an instanton liquid[69, 70] inspired chiral RMT for the QCD partition function. In the spirit of the invariant random matrix ensembles we construct a model for the Dirac operator with the global symmetries of the QCD partition function as input, but otherwise Gaussian random matrix elements. The chRMT that obeys these conditions is defined by[10, 71, 72, 73]

$$Z^{\beta}_{N_f,\nu}(m_1,\cdots,m_{N_f}) = \int DW \prod_{f=1}^{N_f} \det(\mathcal{D}+m_f) e^{-\frac{N\Sigma^2\beta}{4}\mathrm{Tr}W^{\dagger}W}, \tag{49}$$

where

$$\mathcal{D} = \begin{pmatrix} 0 & iW \\ iW^{\dagger} & 0 \end{pmatrix}, \tag{50}$$

and W is a $n \times m$ matrix with $\nu = |n-m|$ and $N = n+m$. The matrix elements of W are either real ($\beta = 1$, chiral Gaussian Orthogonal Ensemble (chGOE)) , complex ($\beta = 2$, chiral Gaussian Unitary Ensemble (chGUE)), or quaternion real ($\beta = 4$, chiral Gaussian Symplectic Ensemble (chGSE)). As is the case in QCD, we assume that ν does not exceed $\sqrt{N}$, so that, to a good approximation, $n = N/2$.

This model reproduces the following symmetries of the QCD partition function:

- The U_A (l) symmetry. All nonzero eigenvalues of the random matrix Dirac operator occur in pairs $\pm\lambda$ or are zero.
- The topological structure of the QCD partition function. The Dirac matrix has exactly $|\nu| \equiv |n-m|$ zero eigenvalues. This identifies ν as the topological sector of the model.
- The flavor symmetry is the same as in QCD. For $\beta = 2$ it is $SU(N_f) \times SU(N_f)$, for $\beta = 1$ it is $SU(2N_f)$ and for $\beta = 4$ it is $SU(N_f)$.
- The chiral symmetry is broken spontaneously with chiral condensate given by

$$\Sigma = \lim_{N\to\infty} \pi\rho(0)/N. \tag{51}$$

  (N is interpreted as the (dimensionless) volume of space time.) The symmetry breaking pattern is[43] $SU(N_f) \times SU(N_f)/SU(N_f)$, $SU(2N_f)/Sp(N_f)$ and $SU(N_f)/O(N_f)$ for $\beta = 2$, 1 and 4, respectively, the same as in QCD[74].

- The anti-unitary symmetries. For three or more colors with fundamental fermions the Dirac operator has no anti-unitary symmetries, and generically, the matrix elements of the Dirac operator are complex. The matrix elements W_{kl} of the corresponding random matrix ensemble are chosen arbitrary complex as well ($\beta = 2$). For $N_c = 2$, the Dirac operator in the fundamental representation satisfies

$$[C\tau_2 K, i\gamma D] = 0, \tag{52}$$

  where C is the charge conjugation matrix and K is the complex conjugation operator. Because, $(C\tau_2 K)^2 = 1$, the matrix elements of the Dirac operator can always be chosen real, and the corresponding random matrix ensemble is defined

with real matrix elements ($\beta = 1$). For two or more colors with gauge fields in the adjoint representation the anti-unitary symmetry of the Dirac operator is given by

$$[CK, i\gamma D] = 0 \ . \tag{53}$$

Because $(CK)^2 = -1$, it is possible to rearrange the matrix elements of the Dirac operator into real quaternions. The matrix elements W_{kl} of the corresponding random matrix ensemble are chosen quaternion real ($\beta = 4$).

Together with the invariant random matrix ensembles, the chiral ensembles are part of a larger classification scheme. Apart from the random matrix ensembles discussed in this review, this classification also includes random matrix models for disordered super-conductors[75]. As pointed out by Zirnbauer[76], all known universality classes of Hermitean random matrices are tangent to the large classes of symmetric spaces in the classification given by Cartan. There is a one-to-one correspondence between this classification and the classification of the large families of Riemannian symmetric superspaces[76].

Selected Results for the Chiral Random Matrix Ensembles

The joint eigenvalue distribution follows from (49) by choosing the eigenvalues and eigenvectors as new integration variables. For N_f flavors and topological charge ν it is given by[71]

$$\rho_\beta(\lambda_1, \cdots, \lambda_n) = C_{\beta,n} \prod_{k<l} |\lambda_k^2 - \lambda_l^2|^\beta \prod_k \lambda_k^{2N_f+\beta\nu+\beta-1} \exp\left(-\frac{n\beta\Sigma^2}{2} \sum_k \lambda_k^2\right), \tag{54}$$

where $C_{\beta,n}$ is a normalization constant and $\nu \geq 0$. For $\beta = 2$ the average spectral density and the spectral correlation functions can be derived from (54) with the help of the orthogonal polynomial method[46]. The associated polynomials are the generalized Laguerre polynomials. That is why this ensemble is also known as the Laguerre ensemble[77, 78]. The spectral density and the two-point correlation function were also derived within the framework of the supersymmetric method of Random Matrix Theory[79]. The calculation of the average spectral density and the spectral correlations functions for $\beta = 1$ and $\beta = 4$ is much more complicated. However, with the help of skew-orthogonal polynomials[80, 81, 82] exact analytical results for finite N can be obtained as well.

¿From the properties of the Laguerre polynomials it can be shown that, independent of the value of β, the average spectral density is a semi-circle

$$\rho(\lambda) = (n\Sigma^2/\pi)\sqrt{4/\Sigma^2 - \lambda^2}. \tag{55}$$

The microscopic spectral density can be derived from the limit (35) of the exact spectral density for finite N. For $N_c = 3$, N_f flavors and topological charge ν it is given by[71]

$$\rho_S(z) = \frac{z}{2}\left(J_a^2(z) - J_{a+1}(z)J_{a-1}(z)\right), \tag{56}$$

where $a = N_f + |\nu|$. The expressions for $SU(2)$ with fundamental fermions ($\beta = 1$) are much more complicated. In this case we find the microscopic spectral density[83]

$$\rho_S(z) = \frac{\Sigma}{4} J_{2a+1}(z\Sigma) + \frac{\Sigma}{2}\int_0^\infty dw (zw)^{2a+1} \epsilon(z-w) \left(\frac{1}{w}\frac{d}{dw} - \frac{1}{z}\frac{d}{dz}\right) \times \frac{wJ_{2a}(z\Sigma)J_{2a-1}(w\Sigma) - zJ_{2a-1}(z\Sigma)J_{2a}(w\Sigma)}{(zw)^{2a}(z^2-w^2)}, \tag{57}$$

where a is the combination

$$a = N_f - \frac{1}{2} + \frac{\nu}{2}. \tag{58}$$

The microscopic spectral density in the symmetry class with $\beta = 4$ was first calculated by Nagao and Forrester [84]. It is given by

$$\rho_S(z) = 2z^2 \int_0^1 du u^2 \int_0^1 dv \ [J_{4a-1}(2uvz)J_{4a}(2uz) - vJ_{4a-1}(2uz)J_{4a}(2uvz)] \tag{59}$$

with $4a = N_f + 2|\nu| + 1$.

The spectral correlations in the bulk of the spectrum are given by the invariant random matrix ensemble with the same value of β. For $\beta = 2$ this was already shown three decades ago by Fox and Kahn[77]. For $\beta = 1$ and $\beta = 4$ this was only proved recently [82] .

Duality between Flavor and Topology

As one can observe from the joint eigenvalue distribution, for $\beta = 2$ the dependence on N_f and ν enters only through the combination $N_f + \nu$. This allows for the possibility of trading topology for flavors. In this section we will work out this duality for the finite volume partition function. This relation completes the proof of the conjectured expression [85] for the finite volume partition function for different quark masses and topological charge ν.

In terms of the eigenvalues the partition function (49) is given by

$$Z^{\beta=2}_{N_f,\nu}(m_1, \cdots, m_{N_f}) = \int d\lambda \Delta^2(\lambda_k^2) \prod_k \lambda_k^{2(N_f+\nu)+1} \prod_f m_f^\nu \prod_{f,k} (\lambda_k^2 + m_f^2) e^{-\frac{\Sigma^2 N}{2} \sum_k \lambda_k^2}, \tag{60}$$

where the Vandermonde determinant is defined by

$$\Delta(\lambda_k^2) = \prod_{k,l} (\lambda_k^2 - \lambda_l^2). \tag{61}$$

By inspection we have

$$\frac{Z_{N_f,\nu}(m_1, \cdots, m_{N_f})}{\prod_f m_f^\nu} \sim Z_{N_f+\nu,0}(m_1, \cdots, m_{N_f}, 0, \cdots, 0), \tag{62}$$

where the argument of the last factor has ν zeros.

As an example, the simplest nontrivial identity of this type is given by

$$Z_{1,1}(m) \sim m Z_{2,0}(m, 0). \tag{63}$$

Let us prove this identity without relying on Random Matrix Theory. According to the definition (28) we have

$$Z_{1,1}(m) \sim \int d\theta e^{i\theta} e^{mV\Sigma \cos\theta}, \tag{64}$$

and

$$Z_{2,0}(m, 0) \sim \int_{U \in U(2)} dU e^{V\Sigma \mathrm{Re}\, \mathrm{Tr}(MU)}, \tag{65}$$

where M is a diagonal matrix with diagonal elements m and 0. The integral over U can be performed by diagonalizing U according to $U = U_1 e^{i\phi_k} U_1^{-1}$, and choosing U_1 and ϕ_k as new integration variables. The Jacobian of this transformation is

$$J \sim \Delta^2(e^{i\phi_k}). \tag{66}$$

The integral over U_1 can be performed using the Itzykson-Zuber formula. This results in

$$Z_{2,0} \sim \int d\phi_1 d\phi_2 \frac{|e^{-i\phi_1} - e^{-i\phi_2}|^2}{m(\cos\phi_1 - \cos\phi_2)} (e^{mV\Sigma\cos\phi_1} - e^{mV\Sigma\cos\phi_2}) \tag{67}$$

Both terms in the last factor result in the same contribution to the integral. Let us consider only the first term $\sim \exp(mV\Sigma\cos\phi_1)$. Then the integral over ϕ_2 has to be defined as a principal value integral. If we use the identity

$$\frac{|e^{-i\phi_1} - e^{-i\phi_2}|^2}{(\cos\phi_1 - \cos\phi_2)} = 2\left(\cos\phi_1 - \frac{\sin\phi_1}{\tan((\phi_1+\phi_2)/2)}\right) \tag{68}$$

the ϕ_2-integral of the term proportional to $\sin\phi_1$ gives zero because of the principal value prescription. The term proportional to $\cos\phi_1$ trivially results in $Z_{1,1}$. We leave it as an exercise to the reader to generalize this proof to arbitrary N_f and ν.

The group integrals in finite volume partition function (28) were evaluated by Leutwyler and Smilga[8] for *equal* quark masses. An expression for *different* quark masses was obtained by Sener *et al.*[85]. The expression could only be proved for $\nu = 0$. The above duality can be used to relate a partition function at arbitrary ν to a partition function at $\nu = 0$. This completes the proof of the conjectured expression for arbitrary topological charge.

UNIVERSALITY IN CHIRAL RANDOM MATRIX THEORY

In the chiral ensembles, two types of universality studies can be performed. First, the universality of correlations in the bulk of the spectrum. As discussed above, they are given by the invariant random matrix ensembles. Second, the universality of the microscopic spectral density and the eigenvalue correlations near zero virtuality. The aim of such studies is to show that microscopic correlations are stable against deformations of the chiral ensemble away from the Gaussian probability distribution. Recently, a number of universality studies on microscopic correlations have appeared. They will be reviewed in this section.

In a first class of universality studies one considers probability distributions that maintain unitary invariance, i.e.

$$P(W) \sim \exp(-N\beta\frac{\Sigma^2}{4}\sum_{k=1}^{\infty} a_k \mathrm{Tr}(W^\dagger W)^k). \tag{69}$$

The first study of this kind was performed by Brézin, Hikami and Zee[13]. They considered a potential with only a_1 and a_2 different from zero and showed that the microscopic spectral density is independent of a_2. A general proof valid for arbitrary potential was given by Nishigaki[11, 86]. The essence of the proof is a remarkable generalization of the identity for the Laguerre polynomials,

$$\lim_{n\to\infty} L_n(\frac{x}{n}) = J_0(2\sqrt{x}) \ , \tag{70}$$

to orthogonal polynomials determined by an arbitrary potential *V*. This relation was proved by deriving a differential equation from the continuum limit of the recursion relation for orthogonal polynomials. In a consecutive work, Akemann, Damgaard, Magnea and Nishigaki[12] extended this proof to all microscopic correlation functions.

In a second class of universality studies one considers deformations of the Gaussian random matrix ensemble that violate unitary invariance. In particular, one has considered the case where the matrix *W* in (50) is replaced by

$$W \to W + A \tag{71}$$

where, because of the unitary invariance, the matrix *A* can always be chosen diagonal. The simplest case with $A = \pi T$ times the identity was considered by Sener *et al.*[15]. This model provides a schematic model of the chiral phase transition. For large matrices, the average resolvent defined by

$$G(z) = \mathrm{Tr}\langle \frac{1}{z + i\epsilon - \mathcal{D}} \rangle \tag{72}$$

obeys the cubic equation[87, 88, 15] (the parameter $\Sigma = 1$ in (49))

$$G^3 - 2zG^2 + G(z^2 - \pi^2 T^2 + 1) - z = 0. \tag{73}$$

This equation was first obtained using a diagrammatic method [87,15]. Later, this derivation was rewritten[89] in terms of the so called blues function [90]. The average spectral density given by

$$\rho(\lambda) = -\frac{1}{\pi}\mathrm{Im}G(z = \lambda) \tag{74}$$

is a semicircle at $\pi T = 0$ and splits into two arcs at $\pi T = 1$. For the spectral density at zero one obtains $\rho(0) = \sqrt{1 - \pi^2 T^2}$, and therefore chiral symmetry is broken for $\pi T < 1$, and is restored above this temperature. In spite of this drastic change in average spectral density, it could be shown[15] with the help of a supersymmetric formulation of Random Matrix Theory that the microscopic spectral density does not depend on *T*.

The super-symmetric method in the first paper by Sener *et al.*[15] is not easily generalizable to higher order correlation functions. A natural way to proceed is to employ the super-symmetric method introduced by Guhr[91]. In the case of $\beta = 2$ this method results in an expression for the kernel determining all correlation functions. This approach was followed in two papers, one by Guhr and Wettig[14] and one by Sener *et al.*[16]. The latter authors studied microscopic correlation functions for *A* in (71) proportional to the identity, whereas Guhr and Wettig considered an arbitrary diagonal matrix *A*. It was shown that independent of the matrix *A*, the correlations are given by the Bessel kernel[92]. Of course, a necessary condition on the matrix *A* is that chiral symmetry is broken. Guhr and Wettig also showed that correlations in the bulk of the spectrum are insensitive to *A*.

The main ingredient of the proof was a supersymmetric generalization [93] of the Itzykson-Zuber-Harish-Chandra type integral[93, 85, 94]

$$\int d\mu(U,V) \exp(\mathrm{Re}\,\mathrm{Tr} U^\dagger S V R) = C \frac{\det_{k,l}(I_\nu(r_k s_l))}{\prod_{k=1}^{N_2}(r_k s_k)^\nu \Delta(S^2)\Delta(R^2)}. \tag{75}$$

Here, $U \in U(N_1)$, $V \in U(N_2)$, $\nu = N_1 - N_2 \geq 0$. The positive square roots of the eigenvalues of $S^\dagger S$ and $R R^\dagger$ are denoted by, s_k and r_k, respectively. The integral is

over the invariant measure. The constant C in the r.h.s. can be evaluated, and the I_ν are modified Bessel functions.

The deformation $W \to W + A$ with the probability distribution for W given by an arbitrary invariant potential has not yet been considered. We have no doubt that universality proofs along the lines of methods developed by P. Zinn-Justin[54] can be given.

We wish to emphasize that all universality studies for the chiral ensembles have been performed for $\beta = 2$. The reason is that $\beta = 1$ and $\beta = 4$ are mathematically much more involved. It certainly would be of interest to extend the above results to these cases as well.

In addition to the above analytical studies the universality of the microscopic spectral density also follows from numerical studies of models with the symmetries of the QCD partition function. In particular, we mention strong support in favor universality from a different branch of physics, namely from the theory of universal conductance fluctuations. In that context, the microscopic spectral density of the eigenvalues of the transmission matrix was calculated for the Hofstadter[95] model, and, to a high degree of accuracy, it agrees with the random matrix prediction[96]. Other studies deal with a class random matrix models with matrix elements with a diverging variance. Also in this case the microscopic spectral density is given by the universal expressions[97] (57) and (56).

The conclusion that emerges from all numerical and analytical work on modified chiral random matrix models is that the microscopic spectral density and the correlations near zero virtuality exhibit a strong universality that is comparable to the stability of microscopic correlations in the bulk of the spectrum.

Of course, QCD is much richer than chiral Random Matrix Theory. One question that should be asked is at what scale (in virtuality) QCD spectral correlations deviate from RMT. This question has been studied by means of instanton liquid simulations. Indeed, at macroscopic scales, it was found that the number variance shows a linear dependence instead of the logarithmic dependence observed at microscopic scales[98]. More work is needed to determine the point where the crossover between these two regimes takes place.

LATTICE QCD RESULTS

Recently, the Dirac spectrum in lattice QCD received a good deal of attention. In particular, the connection between the topology of field configurations and the spectrum of the Wilson Dirac operator has been studied in detail[99, 100, 101]. Other studies are related to the connection between the Wilson Dirac spectrum and the localization properties of the eigenfunctions[102].

In this section we will focus ourselves on the spectral correlations of the lattice QCD Dirac operator. Both correlations in the bulk of the spectrum and the microscopic spectral density will be studied. Consistent with universality arguments presented above, we find that spectral correlations are in complete agreement with chiral Random Matrix Theory.

Correlations in the Bulk of the Spectrum

Recently, Kalkreuter[19] calculated *all* eigenvalues of the $N_c = 2$ lattice Dirac operator both for Kogut-Susskind (KS) fermions and Wilson fermions for lattices as large as 12^4. For the Kogut-Susskind Dirac operator, D^{KS}, we use the convention that it is

anti-Hermitean. Because of the Wilson-term, the Wilson Dirac operator, D^W, is neither Hermitean nor anti-Hermitean. Its Hermiticity relation is given by $D^{W\dagger} = \gamma_5 D^W \gamma_5$. Therefore, the operator $\gamma_5 D^W$ is Hermitean. However, it does not anti-commute with γ_5, and its eigenvalues do not occur in pairs $\pm\lambda_k$.

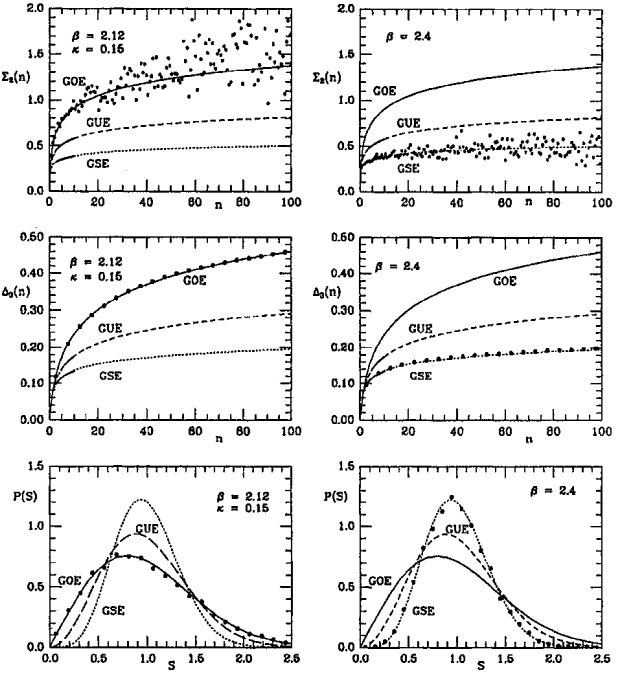


Figure 2. Spectral correlations of Dirac eigenvalues for Wilson fermions (left) and KS-fermions (right). Results are shown for the number variance, $\Sigma_2(n)$, the Δ_3 –statistic and the nearest neighbor spacing distribution, $P(S)$. The full, dashed and dotted curves represent the analytical result for the GOE, GUE and GSE, respectively.

In the the case of $SU(2)$, the anti-unitary symmetry of the Kogut-Susskind and the Wilson Dirac operator is given by[103,22],

$$[D^{KS}, \tau_2 K] = 0, \quad \text{and} \quad [\gamma_5 D^W, \gamma_5 C K \tau_2] = 0. \tag{76}$$

Because

$$(\tau_2 K)^2 = -1, \quad \text{and} \quad (\gamma_5 C K \tau_2)^2 = 1, \tag{77}$$

the matrix elements of the KS Dirac operator can be arranged into real quaternions, whereas the Wilson Dirac operator can be expressed into real matrix elements. Therefore, we expect that eigenvalue correlations in the bulk of the spectrum are described by the GSE and the GOE, respectively[22]. The microscopic correlations for KS fermions are described by the chGSE. However, the microscopic correlations for Wilson fermions are

not described by the chGOE but rather by the GOE. Because of the anti-unitary symmetry, the eigenvalues of the KS Dirac operator are subject to the Kramers degeneracy, i.e. they are double degenerate.

In both cases, the Dirac matrix is tri-diagonalized by Cullum's and Willoughby's Lanczos procedure[104] and diagonalized with a standard QL algorithm. This improved algorithm makes it possible to obtain *all* eigenvalues. This allows us to test the accuracy of the eigenvalues by means of sum-rules for the sum of the squares of the eigenvalues of the lattice Dirac operator. Typically, the numerical error in the sum rule is of order 10^{-8}.

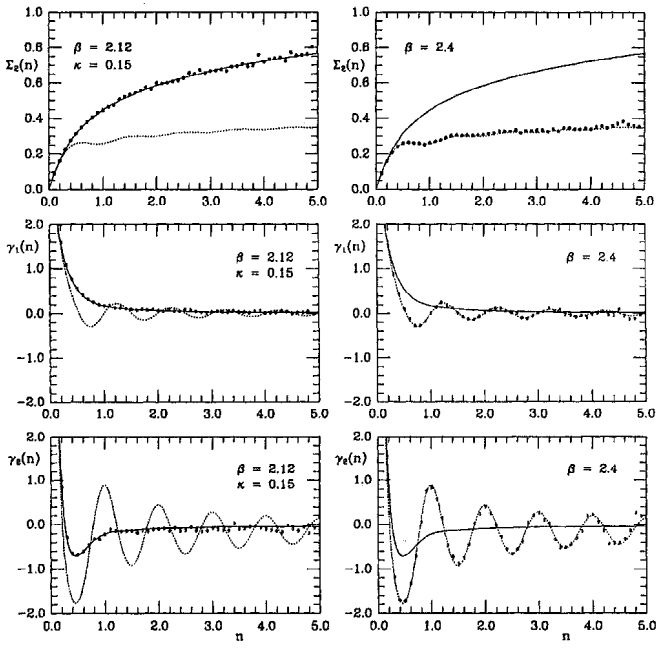


Figure 3. The number variance, $\Sigma_2(n)$ and the first two cumulants, $\gamma_1(n)$ and $\gamma_2(n)$ as a function of n for eigenvalues of the Wilson Dirac operator (left) and the Kogut-Susskind Dirac operator (right). The full and dotted curves represent the analytical result for the GOE and the GSE, respectively.

As an example, in Fig. 1 we show a histogram of the overall Dirac spectrum for KS fermions at $\beta = 2.4$. Results for the spectral correlations are shown in Figs. 2 and 3. The results for KS fermions are for 4 dynamical flavors with $ma = 0.05$ on a 12^4 lattice. The results for Wilson fermions were obtained for two dynamical flavors on a $8^3 \times 12$ lattice. For the values of β and κ we refer to the labels of the figure. For $\beta > 2.4$, with our lattice parameters for KS fermions, the Dirac spectrum near zero virtuality develops a gap. Of course, this is an expected feature of the weak coupling domain. For small enough value of κ the Wilson Dirac spectrum shows a gap at $\lambda = 0$ as well. In the scaling domain the value of κ is just above the critical value of κ. A qualitative description of the Wilson Dirac spectrum can be obtained with a random matrix model with the structure of the Wilson Dirac operator[105].

The eigenvalue spectrum is unfolded by fitting a second order polynomial to the integrated spectral density of a stretch of 500-1000 eigenvalues. The results for $\Sigma_2(n)$, $\Delta_3(n)$ and $P(S)$ in Fig. 2 show an impressive agreement with the RMT predictions. The fluctuations in $\Sigma_2(n)$ are as expected from RMT. The advantage of Δ_3-statistic is well illustrated by this figure. We also investigated[23] the n dependence of the first two cumulants of the number of levels in a stretch of length n. Results presented in Fig. 4 show a perfect agreement with RMT.

Spectra for different values of β have been analyzed as well. It is probably no surprise that random matrix correlations are found at stronger couplings. What is surprising, however, is that even in the weak-coupling domain ($\beta = 2.8$) the eigenvalue correlations are in complete agreement with Random Matrix Theory. Finally, we have studied the stationarity of the ensemble by analyzing level sequences of about 200 eigenvalues (with relatively low statistics). No deviations from random matrix correlations were observed all over the spectrum, including the region near $\lambda = 0$. This justifies the spectral averaging which results in the good statistics in Figs. 2 and 3.

In the case of three or more colors with fundamental fermions, both the Wilson and Kogut-Susskind Dirac operator do not possess any anti-unitary symmetries. Therefore, our conjecture is that in this case the spectral correlations in the bulk of the spectrum of both types of fermions can be described by the GUE. In the case of two fundamental colors the continuum theory and Wilson fermions are in the same universality class. It is an interesting question of how spectral correlations of KS fermions evolve in the approach to the continuum limit. Certainly, the Kramers degeneracy of the eigenvalues remains. However, since Kogut-Susskind fermions represent 4 degenerate flavors in the continuum limit, the Dirac eigenvalues should obtain an additional two-fold degeneracy. We are looking forward to more work in this direction.

The Microscopic Spectral Density

The advantage of studying spectral correlations in the bulk of the spectrum is that one can perform spectral averages instead of ensemble averages requiring only a relatively small number of equilibrated configurations. This so called spectral ergodicity cannot be exploited in the study of the microscopic spectral density. In order to gather sufficient statistics for the microscopic spectral density of the lattice Dirac operator a large number of independent configurations is needed. One way to proceed is to generate instanton-liquid configurations which can be obtained much more cheaply than lattice QCD configurations. Results of such analysis[106] show that for $N_c = 2$ with fundamental fermions the microscopic spectral density is given by the chGOE. For $N_c = 3$ it is given by the chGUE. One could argue that instanton-liquid configurations can be viewed as smoothened lattice QCD configurations. Roughening such configurations will only improve the agreement with Random Matrix Theory.

Of course, the ultimate goal is to test the conjecture of microscopic universality for realistic lattice QCD configurations. In order to obtain a very large number of independent gauge field configurations one is necessarily restricted to relatively small lattices. The first study in this direction was reported recently [20,107]. In this work, the quenched $SU(2)$ Kogut-Susskind Dirac operator is diagonalized for lattices with linear dimension of 4, 6, 8 and 10, and a total number of configurations of 9978, 9953, 3896 and 1416, respectively. The results were compared with predictions from the chGSE.

We only show results for the largest lattice. For more detailed results, including results for the two-point correlation function, we refer to the original work. In Fig. 4 we show

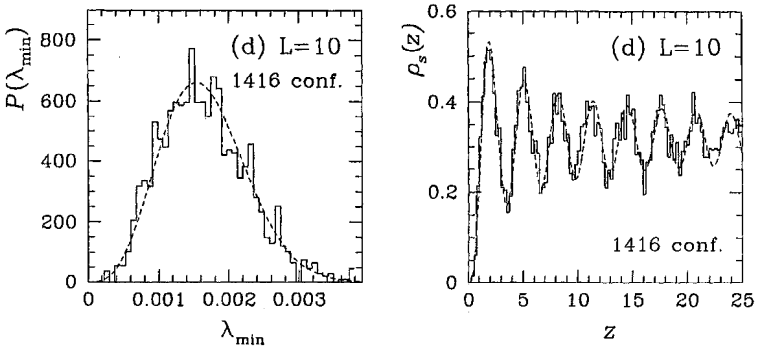


Figure 4. The distribution of the smallest eigenvalue (left) and the microscopic spectral density (right) of the Kogut-Susskind Dirac operator for two colors and $\beta = 2.0$. Lattice results are represented by the histogram, and the analytical results for the chGSE are given by the dashed curves.

the distribution of the smallest eigenvalue (left) and the microscopic spectral density (right). The lattice results are given by the full line. The dashed curve represents the random matrix results. The distribution of the smallest eigenvalue was derived by Forrester [108] and is given by

$$P(\lambda_{\min}) = \alpha\sqrt{\frac{\pi}{2}}(\alpha\lambda_{\min})^{3/2} I_{3/2}(\alpha\lambda_{\min}) e^{-\frac{1}{2}(\alpha\lambda_{\min})^2}, \tag{78}$$

where $\alpha = V\Sigma$. The random matrix result for the microscopic spectral density is given in eq. (59). We emphasize that the theoretical curves have been obtained without any fitting of parameters. The input parameter, the chiral condensate, is derived from the same lattice calculations. The above simulations were performed at a relatively strong coupling of $\beta = 2$. Recently, the same analysis[109] was performed for $\beta = 2.2$ and for $\beta = 2.5$ on a 16^4 lattice. In both cases agreement with the random matrix predictions was found[109].

An alternative way to probe the Dirac spectrum is via the valence quark mass dependence of the chiral condensate[18] defined as

$$\Sigma(m) = \frac{1}{N}\int d\lambda \langle\rho(\lambda)\rangle \frac{2m}{\lambda^2 + m^2}. \tag{79}$$

The average spectral density is obtained for a fixed sea quark mass. For masses well beyond the smallest eigenvalue, $\Sigma(m)$ shows a plateau approaching the value of the chiral condensate Σ. In the mesoscopic range (1), we can introduce $u = \lambda mN$ and $x = mN\Sigma$ as new variables. Then the microscopic spectral density enters in $\Sigma(m)$. For three fundamental colors the microscopic spectral density for $\beta = 2$ (eq. (56)) applies and the integral over λ in (79) can be performed analytically. The result is given by[17],

$$\frac{\Sigma(x)}{\Sigma} = x(I_a(x)K_a(x) + I_{a+1}(x)K_{a-1}(x)), \tag{80}$$

where $a = N_f + |\nu|$. In Fig. 2 we plot this ratio as a function of x (the 'volume' V is equal to the total number of Dirac eigenvalues) for lattice data of two dynamical flavors

with mass $ma = 0.01$ and $N_c = 3$ on a $16^3 \times 4$ lattice. We observe that the lattice data for different values of β fall on a single curve. Moreover, in the mesoscopic range this curve coincides with the random matrix prediction for $N_f = \nu = 0$.

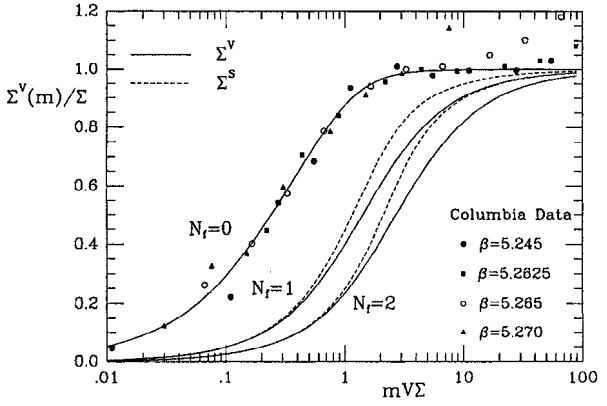


Figure 5. The valence quark mass dependence of the chiral condensate $\Sigma^V(m)$ plotted as $\Sigma^V(m)/\Sigma$ versus $mV\Sigma$. The dots and squares represent lattice results by the Columbia group[18] for values of β as indicated in the label of the figure.

Apparently, the zero modes are completely mixed with the much larger number of nonzero modes. For eigenvalues much smaller than the sea quark mass, one expects quenched ($N_f = 0$) eigenvalue correlations. In the same figure the dashed curves represent results for the quark mass dependence of the chiral condensate (i.e. the mass dependence for equal valence and sea quark masses). In the sector of zero topological charge one finds[110, 9, 8]

$$\frac{\Sigma^S(u)}{\Sigma} = \frac{I_1(u)}{I_0(u)} \quad \text{for} \quad N_f = 1, \tag{81}$$

and

$$\frac{\Sigma^S(u)}{\Sigma} = \frac{I_1^2(u)}{u(I_0^2(u) - I_1^2(u))} \quad \text{for} \quad N_f = 2. \tag{82}$$

We observe that both expressions do not fit the data. Also notice that, according to Göckeler *et al.* [111], eq. (81) describes the valence mass dependence of the chiral condensate for non-compact QED with quenched Kogut-Susskind fermions. However, we were not able to derive their result (no derivation is given in the paper).

CHIRAL RANDOM MATRIX THEORY AT $\mu \neq 0$

At nonzero temperature T and chemical potential μ a *schematic* random matrix matrix model of the QCD partition function is obtained by replacing the Dirac operator in (49) by[87, 112, 88, 35]

$$\mathcal{D} = \begin{pmatrix} 0 & iW + i\Omega_T + \mu \\ iW^\dagger + i\Omega_T + \mu & 0 \end{pmatrix}. \tag{83}$$

Here, $\Omega_T = T \otimes_n (2n+1)\pi\mathbf{1} + \mu$ are the matrix elements of $\gamma_0\partial_0 + \mu\gamma_0$ in a plane wave basis with anti-periodic boundary conditions in the time direction. Below, we will discuss a model with Ω_T absorbed in the random matrix and $\mu \neq 0$. The aim of this model is to explore the effects of the non-Hermiticity of the Dirac operator. For example, the random matrix partition function (49) with the Dirac matrix (83) is well suited for the study of zeros of this partition function in the complex mass plane and in the complex chemical potential plane. For a complete analytical understanding of the location of such zeros we refer to the work by Halasz *et al.*[113].

The term $\mu\gamma_0$ does not affect the anti-unitary symmetries of the Dirac operator. This is also the case in lattice QCD where the color matrices in the forward time direction are replaced by $U \to e^{\mu} U$ and in the backward time direction by $U^{\dagger} \to e^{-\mu}U^{\dagger}$. For this reason the universality classes are the same as at zero chemical potential.

The Dirac operator that will be discussed in this section is thus given by

$$\mathcal{D}(\mu) = \begin{pmatrix} 0 & iW + \mu \\ iW^{\dagger} + \mu & 0 \end{pmatrix}, \tag{84}$$

where the matrix elements W are either real ($\beta = 1$), complex ($\beta = 2$) or quaternion real ($\beta = 4$). For all three values of β the eigenvalues of $D(\mu)$ are scattered in the complex plane.

Since many standard random matrix methods rely on convergence properties based on the Hermiticity of the random matrix, direct application of most methods is not possible. The simplest way out is the Hermitization[114] of the problem, i.e we consider the Hermitean operator

$$D^H(z, z^*) = \begin{pmatrix} \kappa & z - \mathcal{D}(\mu) \\ z^* - \mathcal{D}^{\dagger}(\mu) & \kappa \end{pmatrix}. \tag{85}$$

For example, the generating function in the supersymmetric method of Random Matrix Theory[115, 116] is then given by [117, 118, 119, 120]

$$Z(J, J^*) = \langle \frac{\det(D^H(z + J, z^* + J^*)}{\det(D^H(z, z^*))} \rangle. \tag{86}$$

The determinants can be rewritten as fermionic and bosonic integrals. Convergence is assured by the Hermiticity and the infinitesimal increment κ. The resolvent follows from the generating function by differentiation with respect to the source terms

$$G(z, z^*) = \mathrm{Tr}\langle \frac{1}{z - \mathcal{D}(\mu)} \rangle = \frac{\partial}{\partial J} Z(J, J^*)\Big|_{J=J^*=0}. \tag{87}$$

Notice that, after averaging over the random matrix, the partition function depends in a non-trivial way on both z and z^*. The spectral density is then given by

$$\rho(z) = \frac{1}{\pi}\frac{\partial}{\partial z^*} G(z). \tag{88}$$

Alternatively, one can use the replica trick[121, 35] with generating function given by

$$\langle \det^{N_f} D_H(z, z^*) \rangle. \tag{89}$$

The idea is to perform the calculation for integer values of N_f and perform the limit $N_f \to 0$ at the end of the calculation. Although the replica limit, $N_f \to 0$, fails in general[122], it is expected to work for $\det D_H(z, z^*)$ because it is positive definite (or

zero). Then the partition function is a smooth function of N_f. For other techniques addressing nonhermitean matrices we refer to the recent papers by Feinberg and Zee[114] and Nowak and co-workers[123]. One recent method that does not rely on the Hermiticity of the random matrices is the method of complex orthogonal polynomials[36]. This method was used by Fyodorov *et al.*[36] to calculate the number variance and the nearest neighbor spacing distribution in the regime of weakly nonhermitean matrices. As surprising new result, they found an $S^{5/2}$ repulsion law.

In the physically relevant case of QCD with three colors, the fermion determinant is complex for nonzero chemical potential. Its phase prevents the convergence of fully unquenched Monte-Carlo simulations (see Kogut *et al.*[124] for the latest progress in this direction). However, it is possible to perform quenched simulations. In such calculations it was found that the critical chemical potential $\mu_c \sim \sqrt{m}$, instead of a third of the nucleon mass[125]. This phenomenon was explained analytically by Stephanov[35] with the help of the above random matrix model. He could show that for small μ the eigenvalues are distributed along the imaginary axis in a band of width $\sim \mu^2$ leading to a critical chemical potential of $\mu_c \sim m^2$. As has been argued above, the quenched limit is necessarily obtained from a partition function in which the fermion determinant appears as

$$\lim_{N_f \to 0} |\det \mathcal{D}(\mu)|^{N_f}, \tag{90}$$

instead of the same expression without the absolute value signs. The partition function with the absolute value of the determinant can be interpreted as a partition function of an equal number of fermions and conjugate fermions. The critical value of the chemical potential, equal to half the pion mass, is due to Goldstone bosons with net a baryon number consisting of a quark and conjugate quark. The reason that the quenched limit does not correspond to the standard QCD partition function is closely related to the failure of the replica trick in the case of a determinant with a nontrivial phase.

Both for $\beta = 1$ and $\beta = 4$ the fermion determinant, $\det(D(\mu) + m)$, is real. This is obvious for $\beta = 1$. For $\beta = 4$ the reality follows from the identity $q^* = \sigma_2 q \sigma_2$ for a quaternion real element q, and the invariance of a determinant under transposition. We thus conclude that quenching works for an even number of flavors. Consequently, chiral symmetry will be restored for arbitrarily small nonzero μ, whereas a condensate of a quark and a conjugate quark develops. Indeed, this phenomenon has been observed in the strong coupling limit of lattice QCD with two colors[127].

In the quenched approximation, the spectral properties of the random matrix ensemble (84) can be easily studied numerically by simply diagonalizing a set of matrices with probability distribution (49). In Fig. 6 we show numerical results[128] for the eigenvalues of a few 100×100 matrices for $\mu = 0.15$ and $\mu = 0.5$. The dots represent the eigenvalues in the complex plane. The full line is the analytical result[35] for the boundary of the eigenvalues which is given by the algebraic curve

$$(y^2 + 4\mu^2)(\mu^2 - x^2)^2 + x^2 = 4\mu^2(\mu^2 - x^2). \tag{91}$$

This result was obtained by Stephanov for $\beta = 2$ with the help of the fermionic replica trick. This amounts to rewriting the determinants in (89) as integrals over Grassmann variables. Since Grassmann integrals are always convergent the infinitesimal increment κ can be put equal to zero. This method can be extended[128] to $\beta = 1$ and $\beta = 4$. Although the effective partition function is much more complicated, it can be shown without too much effort that the solutions of the saddle point equations are the same if the variance of the probability distribution is scaled as $1/\beta$. In particular, the boundary

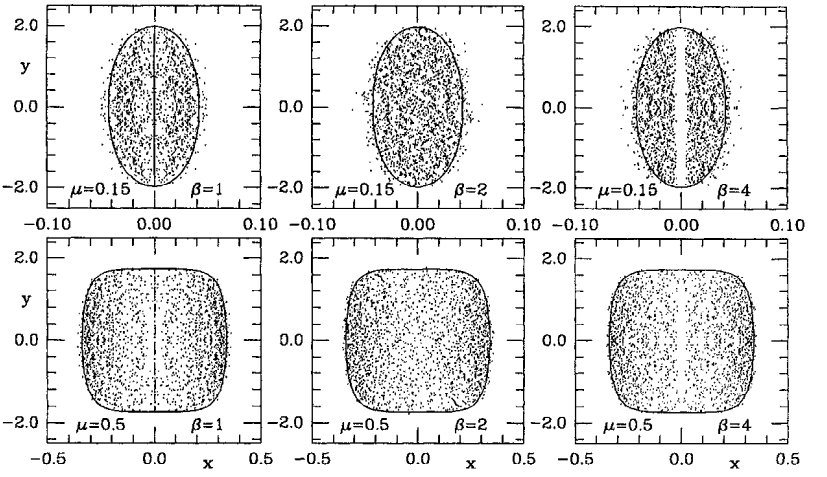


Figure 6. Scatter plot of the real (x), and the imaginary parts (y) of the eigenvalues of the random matrix Dirac operator. The values of β and μ are given in the labels of the figure. The full curve shows the analytical result for the boundary.

of the domain of eigenvalues is the same in each of the three cases. However, as one observes from Fig. 6, for $\beta = 1$ and $\beta = 4$ the spectral density deviates significantly from the saddle-point result. For $\beta = 1$ we find an accumulation of eigenvalues on the imaginary axis, whereas for $\beta = 4$ we find a depletion of eigenvalues in this domain. This depletion can be understood as follows. For $\mu = 0$ all eigenvalues are doubly degenerate. This degeneracy is broken at $\mu \neq 0$ which produces the observed repulsion of the eigenvalues.

The number of purely imaginary eigenvalues for $\beta = 1$ appears to scale as $\sqrt{N}$. This explains that this effect is not visible in a leading order saddle point analysis. From a perturbative analysis of (89) one obtains a power series in $1/N$. Clearly, the $\sqrt{N}$ dependence requires a truly nonperturbative analysis of the partition function (49) with the Dirac operator (84). Such a $\sqrt{N}$ scaling behavior is typical for the regime of weak non-hermiticity first identified by Fyodorov *et al.*[119]. Using the supersymmetric method for the generating function (86) the $\sqrt{N}$ dependence was obtained analytically by Efetov[120].

A similar cut below a cloud of eigenvalues was found in instanton liquid simulations[129] for $N_c = 2$ at $\mu \neq 0$ and in a random matrix model of arbitrary real matrices[118]. The depletion of the eigenvalues along the imaginary axis was observed earlier in lattice QCD simulations with staggered fermions[130]. Obviously, more work has to be done in order to arrive at a complete characterization of the universal features[36] in the spectrum of nonhermitean matrices.

CONCLUSIONS

We have argued that there is an intimate relation between correlations of Dirac eigenvalues and the breaking of chiral symmetry. In the chiral limit, the fermion determinant suppresses gauge field configurations with small Dirac eigenvalues. Correlations counteract this suppression, and are a necessary ingredient of chiral symmetry break-

ing. From the study of eigenvalue correlations in strongly interacting systems, we have concluded that they are described naturally with by Random Matrix Theory with the global symmetries of the physical system. In QCD, this led to the introduction of chiral Random Matrix Theories. They provided us with an analytical understanding of the statistical properties of the eigenvalues on the scale of a typical level spacing. In particular, impressive agreement between lattice QCD and chiral Random Matrix Theory was found for the microscopic spectral density and for spectral correlations in the bulk of the spectrum. An extension of this model to nonzero chemical potential explains some intriguing properties of previously obtained lattice QCD Dirac spectra and instanton liquid Dirac spectra.

Acknowledgements

This work was partially supported by the US DOE grant DE-FG-88ER40388. Poul Damgaard and Jerzy Jurkiewicz are thanked for organizing this workshop. NATO is acknowledged for financial support. We benefitted from discussions with T. Wettig, and M. Halasz is thanked for a critical reading of the manuscript. Finally, I thank my collaborators on whose work this review is based.

REFERENCES

1. C. DeTar, *Quark-gluon plasma in numerical simulations of QCD,* in *Quark gluon plasma 2,* R. Hwa ed., World Scientific 1995.
2. A. Ukawa, Nucl. Phys. Proc. Suppl. **53** (1997) 106.
3. A. Smilga, *Physics of thermal QCD,* hep-ph/9612347.
4. E. Shuryak, Comments Nucl. Part. Phys. **21** (1994) 235.
5. C. Bernard, T. Blum, C. DeTar, S. Gottlieb, U. Heller, J. Hetrick, K. Rummukainen, R. Sugar, D. Toussaint and M. Wingate, Phys. Rev. Lett. **78** (1997) 598.
6. T. Banks and A. Casher, Nucl. Phys. **B169** (1980) 103.
7. O. Bohigas, M. Giannoni, Lecture notes in Physics **209** (1984) 1; O. Bohigas, M. Giannoni and C. Schmit, Phys. Rev. Lett. **52** (1984) 1.
8. H. Leutwyler and A. Smilga, Phys. Rev. **D46** (1992) 5607.
9. J. Gasser and H. Leutwyler, Phys. Lett. **l88B** (1987) 477.
10. E.V. Shuryak and J.J.M. Verbaarschot, Nucl. Phys. **A560** (1993) 306.
11. S. Nishigaki, Phys. Lett. **387 B** (1996) 707.
12. G. Akemann, P. Damgaard, U. Magnea and S. Nishigaki, Nucl. Phys. **B 487**[**FS**] (1997) 721.
13. E. Brézin, S. Hikami and A. Zee, Nucl. Phys. **B464** (1996) 411.
14. T. Guhr and T. Wettig, *Universal spectral correlations of the Dirac operator at finite temperature,* hep-th/9704055, Nucl. Phys. **B** (in press).
15. A.D. Jackson, M.K. Sener and J.J.M. Verbaarschot, Nucl. Phys. **B479** (1996) 707.
16. A.D. Jackson, M.K. Sener and J.J.M. Verbaarschot, *Universality of correlation functions in random matrix models of QCD,* hep-th/9704056, Nucl. Phys. **B** (in press).
17. J.J.M. Verbaarschot, Phys. Lett. **B368** (1996) 137.
18. S. Chandrasekharan, Nucl. Phys. Proc. Suppl. **42** (1995) 475; S. Chandrasekharan and N. Christ, Nucl. Phys. Proc. Suppl. **42** (1996) 527; N. Christ, Lattice 1996.
19. T. Kalkreuter, Phys. Lett. **B276** (1992) 485; Phys. Rev. **D48** (1993) 1; Comp. Phys. Comm. **95** (1996) 1.
20. M.E. Berbenni-Bitsch, S. Meyer, A. Schäfer, J.J.M. Verbaarschot, and T. Wettig, *Microscopic Universality in the spectrum of the lattice Dirac operator,* hep-lat/9704018.
21. A. Pandey, Ann. Phys. **134** (1981) 119.
22. M.A. Halasz and J.J.M. Verbaarschot, Phys. Rev. Lett. **74** (1995) 3920.
23. M.A. Halasz, T. Kalkreuter and J.J.M. Verbaarschot, Nucl. Phys. Proc. Suppl. **53** (1997) 266.
24. T. Guhr, A. Müller-Groeling and H.A. Weidenmüller, *Random Matrix Theories in quantum physics: Common concepts,* cond-mat/9707301, Phys. Rep. (in press).
25. R. Haq, A. Pandey and O. Bohigas, Phys. Rev. Lett. **48** (1982) 1086.

26. C. Ellegaard, T. Guhr, K. Lindemann, H.Q. Lorensen, J. Nygard and M. Oxborrow, Phys. Rev. Lett. **75** (1995) 1546.
27. S. Deus, P. Koch and L. Sirko, Phys. Rev. **E 52** (1995) 1146; H. Gräf, H. Harney, H. Lengeler, C. Lewenkopf, C. Rangacharyulu, A. Richter, P. Schardt and H.A. Weidenmüller, Phys. Rev. Lett. **69** (1992) 1296,
28. T. Ericson, Phys. Rev. Lett. **5** (1960) 430.
29. H.A. Weidenmüller, Ann. Phys. (N.Y.) **158** (1984) 78; in *Proceedings of T. Ericson's 60th birthday*.
30. Y. Imry, Europhysics Lett. **1** (1986) 249; B.L. Altshuler, P.A. Lee and R.A. Webb (eds.), *Mesoscopic Phenomena in Solids*, North-Holland, New York, 1991; S. Iida, H.A. Weidenmüller and J. Zuk, Phys. Rev. Lett. **64** (1990) 583; Ann. Phys. (N.Y.) **200** (1990) 219; C.W.J. Beenakker, Rev. Mod. Phys. **69** (1997) 731.
31. P.W. Anderson, Phys. Rev. **109** (1958) 1492.
32. H. Sommers, A. Crisanti, H. Sompolinsky and Y. Stein, Phys. Rev. Lett. **60** (1988) 1895.
33. D. Gross and E. Witten, Phys. Rev. **D21** (1980) 446; S. Chandrasekharan, Phys. Lett. **B395** (1997) 83.
34. P. Di Francesco, P. Ginsparg, and J. Zinn-Justin, Phys, Rep. **254** (1995) 1.
35. M. Stephanov, Phys. Rev. Lett. **76** (1996) 4472.
36. Y. Fyodorov, B. Khoruzhenko and H. Sommers, *Almost-Hermitian Random Matrices: Crossover from Wigner-Dyson to Ginibre eigenvalue statistics*, cond-mat/9703152.
37. M.A. Nowak, this proceedings.
38. C.N. Yang and T.D. Lee, Phys. Rev. **87** (1952) 104, 410.
39. V. Matteev and R. Shrock, J. Phys. A: Math. Gen. **28** (1995) 5235.
40. J. Vink, Nucl. Phys, **B323** (1989) 399.
41. I. Barbour, A. Bell, M. Bernaschi, G. Salina and A. Vladikas, Nucl. Phys. **B386** (1992) 683.
42. J. Osborn and J.J.M. Verbaarschot, in preparation.
43. A. Smilga and J.J.M. Verbaarschot, Phys. Rev. **D51** (1995) 829.
44. M.A. Halasz and J.J.M. Verbaarschot, Phys. Rev. **D52** (1995) 2563.
45. F. Dyson and M. Mehta, J. Math. Phys. **4** (1963) 701.
46. M. Mehta, *Random Matrices*, Academic Press, San Diego, 1991.
47. A.M. Odlyzko, Math. Comput. **48** (1987) 273.
48. D. Voiculescu, K. Dykema and A. Nica, *Free Random Variables*, Am. Math. Soc., Providence RI, 1992.
49. G. Hackenbroich and H.A. Weidenmüller, Phys. Rev. Lett. **74** (1995) 4118.
50. V. Freilikher, E. Kanzieper and I. Yurkevich, Phys. Rev. **E53** (1996) 2200.
51. B. Eynard, *Eigenvalue distribution of large random matrices, from one matrix to several coupled matrices*, cond-mat/9707005.
52. E. Brézin and A. Zee, Nucl. Phys. **B402** (1993) 613.
53. N. Deo, *Orthogonal polynomials and exact correlation functions for two cut random matrix models*, cond-mat/9703136.
54. P. Zinn-Justin, *Universality of correlation functions of hermitean random matrices in an external field*, cond-mat/9705044; Nucl. Phys. **B497** (1997) 725.
55. E. Brézin and S. Hikami, *An extension of level spacing universality*, cond-mat/9702213.
56. T.A. Brody, J. Flores, J.B. French, P.A. Mello, A. Pandey and S.S.M. Wong, Rev. Mod. Phys. **53** (1981) 385.
57. J.J.M. Verbaarschot and M.R. Zirnbauer, Ann. Phys. (N.Y.) **158** (1984) 78.
58. J. Ambjorn, J. Jurkiewicz and Y. Makeenko, Phys. Lett. B251 (1990) 517.
59. J. Ambjorn and G. Akemann, J. Phys. A29 (1996) L555; Nucl. Phys. **B482** (1996) 403.
60. C.W.J. Beenakker, Nucl.Phys. **B422** (1994) 515.
61. J.J.M. Verbaarschot, H.A. Weidenmüller and M.R. Zirnbauer, Ann. Phys. (N.Y.) **153** (1984) 367.
62. E. Brézin and A. Zee, Nucl. Phys. **B453** (1995) 531.
63. M. Berry, Proc. Roy. Soc. London **A 400** (1985) 229.
64. E. Kanzieper and V. Freilikher, Phys. Rev. Lett. **78** (1997) 3806.
65. E. Brézin and J. Zinn-Justin, Phys. lett. **B288** (1992) 54.
66. S. Higuchi, C.Itoi, S.M. Nishigaki and N. Sakai, *Renormalization group approach to multiple arc random matrix models*, hep-th/9612237.
67. A.V. Andreev, O. Agam, B.D. Simons and B.L. Altshuler, Nucl. Phys. **B482** (1996) 536.
68. A. Altland and M. Zirnbauer, Phys. Rev. Lett. **77** (1996) 4536.
69. T. Schäfer and E. Shuryak, *Instantons in QCD*, hep-ph/9610451, Rev. Mod. Phys. (1997).

70. D.I. Diakonov and V.Yu. Petrov, Nucl. Phys. **B272** (1986) 457.
71. J.J.M. Verbaarschot, Phys. Rev. Lett. **72** (1994) 2531; Phys. Lett. **B329** (1994) 351.
72. J.J.M. Verbaarschot and I. Zahed, Phys. Rev. Lett. **70** (1993) 3852.
73. J.J.M. Verbaarschot, Nucl. Phys. Proc. Suppl. **53** (1997) 88.
74. M. Peskin, Nucl. Phys. **B175** (1980) 197; S. Dimopoulos, Nucl. Phys. **B168** (1980) 69; M. Vysotskii, Y. Kogan and M. Shifman, Sov. J. Nucl. Phys. **42** (1985) 318; D.I. Diakonov and V.Yu. Petrov, Lecture notes in physics, **417,** Springer 1993.
75. A. Altland, M.R. Zirnbauer, Phys. Rev. Lett. **76** (1996) 3420; *Novel Symmetry Classes in Mesoscopic Normal-Superconducting Hybrid Structures,* cond-mat/9602137.
76. M.R. Zirnbauer, J. Math. Phys. **37** (1996) 4986; F.J. Dyson, Comm. Math. Phys. 19 (1970) 235.
77. D. Fox and P.Kahn, Phys. Rev. **134** (1964) B1152; (1965) 228.
78. B. Bronk, J. Math. Phys. **6** (1965) 228.
79. A.V. Andreev, B.D. Simons, and N. Taniguchi, Nucl. Phys **B432** [FS] (1994) 487.
80. F. Dyson, J. Math. Phys. **13** (1972) 90.
81. G. Mahoux and M. Mehta, J. Phys. I France **I** (1991) 1093.
82. T. Nagao and M. Wadati, J. Phys. Soc. Japan **60** (1991) 2998; J. Phys. Soc. Jap. **60** (1991) 3298; J. Phys. Soc. Jap. bf 61 (1992) 78; J. Phys. Soc. Jap. bf 61 (1992) 1910. **61** (1992) 78, 1910.
83. J.J.M. Verbaarschot, Nucl. Phys. B426 (1994) 559.
84. T. Nagao and P.J. Forrester, Nucl. Phys. **B435** (1995) 401.
85. A.D. Jackson, M.K. Sener and J.J.M. Verbaarschot, Phys. Lett. B **387** (1996) 355.
86. S. Nishigaki, this proceedings.
87. A.D. Jackson and J.J.M. Verbaarschot, Phys. Rev. **D53** (1996) 7223.
88. M. Stephanov, Phys. Lett. **B275** (1996) 249; Nucl. Phys. Proc. Suppl. **53** (1997) 469.
89. M.A. Nowak, G. Papp and I. Zahed, Phys. Lett. B389 (1996) 137.
90. A. Zee, Nucl. Phys. **B474** (1996) 726.
91. T. Guhr, J. Math. Phys. **32** (1991) 336.
92. C. Tracy and H. Widom, Comm. Math. Phys. **161** (1994) 289.
93. T. Guhr and T. Wettig, J. Math. Phys. **37** (1996) 6395.
94. F.A. Berezin and F.I. Karpelevich, Doklady Akad. Nauk SSSR **118** (1958) 9.
95. D. Hofstadter, Phys. Rev. **B14** (1976) 2239.
96. K. Slevin and T. Nagao, Phys. Rev. Lett. **70** (1993) 635.
97. J. Kelner and J.J.M. Verbaarschot, in preparation.
98. J. Osborn and J.J.M. Verbaarschot, in progress.
99. T. Ivanenko, *Study of Instanton Physics in lattice QCD,* Thesis, Massachusetts Institute of Technology, 1997.
100. W. Bardeen, A. Duncan, E. Eichten, G. Hockney and H. Thacker, *Light quarks, zero modes, and exceptional configurations,* hep-lat/9705008; W. Bardeen, A. Duncan, E. Eichten and H. Thacker, *Quenched approximation artifacts: a detailed study in two-dimensional QED,* hep-lat/9705002.
101. C.R. Gattringer, I. Hip and C.B. Lang, *Topological charge and the spectrum of the fermion matrix in lattice QED in two-dimensions,* hep-lat/9707011.
102. K. Jansen, C. Liu, H. Simma and D. Smith, Nucl. Phys. Proc. Supp. **53,** 262 (1997).
103. S. Hands and M. Teper, Nucl. Phys. **B347** (1990) 819.
104. J. Cullum and R.A. Willoughby, J. Comp. Phys. **44** (1981) 329.
105. J. Jurkiewicz, M.A. Nowak and I. Zahed, Nucl. Phys. **B478** (1996) 605.
106. J.J.M. Verbaarschot, Nucl. Phys. **B427** (1994) 534.
107. T. Wettig, T. Guhr, A. Schäfer and H. Weidenmüller, *The chiral phase transition, random matrix models, and lattice data,* hep-ph/9701387.
108. P. Forrester, Nucl. Phys. **B**[**FS**]**402** (1993) 709.
109. T. Wettig, private communication 1997.
110. T. Jolicoeur and A. Morel, Nucl. Phys. **B262** (1985) 627.
111. M. Göckeler, R. Horsley, E. Laermann, P. Rakow, G. Schierholtz, R. Sommer and U.-J. Wiese, Nucl. Phys. **B334** (1990) 527.
112. T. Wettig, A. Schäfer and H. Weidenmüller, Phys. Lett. **B367** (1996) 28.
113. M.A. Halasz, A.D. Jackson and J.J.M. Verbaarschot, Phys. Lett. **B395** (1997) 293; *Fermion determinants in matrix models of QCD at nonzero chemical potential,* hep-lat/9703006, *Phys.* Rev. **D** (in press).
114. J. Feinberg and A. Zee, *Non-Hermitean Random Matrix Theory: method of hermitization,* cond-mat/9703118; *Nongaussian nonhermitean random matrix theory: phase transition and addition*

formalism, cond-mat/9704191; *Nonhermitean random matrix theory: method of hermitean reduction,* cond-mat/9703087.
115. K. Efetov, Adv. Phys. **32,** (1983) 53.
116. J.J.M. Verbaarschot, H.A. Weidenmüller, and M.R. Zirnbauer, Phys. Rep. **129,** (1985) 367.
117. Y. Fyodorov and H. Sommers, JETP Lett. **63** (1996) 1026.
118. B. Khoruzhenko, J. Phys. A: Math. Gen. **29,** (1996) L165.
119. Y. Fyodorov, B. Khoruzhenko and H. Sommers, Phys. Lett. **A 226,** (1997) 46.
120. K. Efetov, *Directed quantum chaos,* cond-mat/9702091; *Quantum disordered systems with a direction,* cond-mat/9706055.
121. V.L. Girko, *Theory of random determinants,* Kluwer Academic Publishers, Dordrecht, 1990.
122. J.J.M. Verbaarschot and M.R. Zirnbauer, J. Phys. **A17** (1985) 1093.
123. R. Janik, M.A. Nowak, G. Papp, J. Wambach, and I. Zahed, Phys. Rev. **E55** (1997) 4100; R. Janik, M.A. Nowak, G. Papp and I. Zahed, *Nonhermitean random matrix models. 1,* cond-mat/9612240.
124. I. M. Barbour, S. E. Morrison, E. G. Klepfish, J. B. Kogut and M.-P. Lombardo, *Results on Finite Density QCD,* hep-lat/9705042.
125. I. Barbour, N. Behihil, E. Dagotto, F. Karsch, A. Moreo, M. Stone and H. Wyld, Nucl. Phys. **B275** (1986) 296; M.-P. Lombardo, J.B. Kogut and D.K. Sinclair, Phys. Rev. **D54** (1996) 2303.
126. I.M. Barbour, S. Morrison and J. Kogut, *Lattice gauge theory simulation at nonzero chemical potential in the chiral limit,* hep-lat/9612012.
127. E. Dagotto, F. Karsch and A. Moreo, Phys. Lett. **169 B,** (1986) 421.
128. M.A. Halasz, J. Osborn and J.J.M. Verbaarschot, *Random matrix triality at nonzero chemical potential,* hep-lat/9704007, Phys. Rev. **D** (in press).
129. Th. Schäfer, *Instantons and the Chiral Phase Transition at non-zero Baryon Density,* hep-ph/9708256.
130. C. Baillie, K.C. Bowler, P.E. Gibbs, I.M. Berbour and M. Rafique, Phys. Lett. **197B,** (1987) 195.

DETERMINATION OF CRITICAL EXPONENTS AND EQUATION OF STATE BY FIELD THEORY METHODS

J. Zinn-Justin*

CEA-Saclay, Service de Physique Théorique†,
F-91191 Gif-sur-Yvette
Cedex, FRANCE

ABSTRACT

We review here the methods, based on renormalized ϕ_3^4 quantum field theory and renormalization group, which have led to an accurate calculation of critical exponents of the N-vector model, and more recently of the equation of state of the 3D Ising model. The starting point is the perturbative expansion for RG functions or the effective potential to the order presently available. Perturbation theory is known to be divergent and its divergence has been related to instanton contributions. This has allowed to characterize the large order behaviour of perturbation series, an information that can be used to efficiently "sum" them. Practical summation methods based on Borel transformation and conformal mapping has been developed, leading to the most accurate results available probing field theory in a non-perturbative regime. We illustrate the methods with a detailed discussion of the scaling equation of state of the 3D Ising model [1]. Compared to exponents its determination involves a few additional (non-trivial) technical steps.

A general reference on the topic is

J. Zinn-Justin, 1989, *Quantum Field Theory and Critical Phenomena*, in particular chap. 28 of third ed., Clarendon Press (Oxford 1989, third ed. 1996).

Many relevant articles are reprinted in

Large Order Behaviour of Perturbation Theory, Current Physics vol. 7, J.C. Le Guillou and J. Zinn-Justin eds., (North-Holland, Amsterdam 1990).

1. INTRODUCTION

Let us first briefly review one line of arguments which, starting from some statistical model (like the Ising model), leads, in the critical domain, to quantum field theory.

*email: zinn@spht.saclay.cea.fr

†Laboratoire de la Direction des Sciences de la Matière du Commissariat à l'Energie Atomique

New Developments in Quantum Field Theory
Edited by Damgaard and Jurkiewicz, Plenum Press, New York, 1998

The starting point is mean field theory which allows to describe phase transitions and explore the neighbourhood of the critical temperature. In the case of second order phase transitions, continuous phase transitions where the correlation length diverges, this leads to the concept of super-universality. The latter is summarized in Landau's theory of critical phenomena. A number of quantities, like the exponents which characterize the singular behaviour of physical observables near the critical temperature, are universal, i.e. independent of the system (provided it has only short range interactions), and even the dimension of space. However empirical evidence, exact solutions of 2D models, and finally an analysis of corrections to mean field theory, had shown that a universality of such general nature could not be true. Empirical evidence also suggested that universality could still hold, but in a more limited sense. Since the failure of mean field theory originated from the impossibility of describing critical systems with a purely macroscopic theory, it could have been feared that critical properties would be short distance sensitive. Therefore the existence of even a more restricted universality was puzzling. It took many years to develop the renormalization group (RG) ideas [2] which explain the origin of universality: it relies on the existence of IR fixed points of RG transformations.

From the technical point of view, RG arguments, as well as an analysis, near dimension four, of the most divergent terms appearing in the expansion around mean field theory [3], then indicated that all universal quantities quantities could be calculated from renormalizable or super-renormalizable quantum field theories. In particular for an important class of physical systems and models (with short range interactions) one is led to a ϕ^4-like euclidean field theory with $O(N)$ symmetry. Among those let us mention statistical properties of polymers, liquid–vapour and binary mixtures transitions, superfluid Helium, ferromagnets... For these systems we shall explain how critical exponents and other universal quantities have been calculated with field theory techniques. To simplify notation we shall concentrate on the universality class of the Ising model (models with Z_2 symmetry), but the arguments are more generally applicable.

1.1 The effective quantum field theory

As indicated above, after some analysis, one is led to a ϕ^4 field theory with an action of the form

$$\mathcal{H}(\phi) = \int \mathrm{d}^d x \left[\frac{1}{2}c(\nabla\phi(x))^2 + \frac{1}{2}a\phi^2(x) + b\frac{1}{4!}\phi^4(x)\right], \tag{1}$$

with a, b and c being *regular* functions of the temperature for T close to T_c. Actually in perturbation theory the coefficient of ϕ^2 is the most sensitive to a variation of the temperature. Calling a_c its value at T_c, we can take $a - a_c \propto T - T_c$ as a linear measure of the deviation from the critical temperature.

The field ϕ can be considered as some local average of the initial spin variable.

The hamiltonian 1 generates a perturbative expansion of field theory type which can be described in terms of Feynman diagrams. These have to be calculated with a momentum cut-off of order 1, reflection of the microscopic scale of the initial physical model (like the lattice spacing in lattice models).

The critical domain, where universality can be expected, is characterized by

$$\text{physical massm} = \xi^{-1} \ll 1 \Rightarrow \ |a - a_c \propto T - T_c| \ll 1$$

$$\text{distances} >> 1 \text{ or momenta} << 1, \tag{2}$$

$$\text{magnetization } M \equiv \langle \phi(x) \rangle << 1.$$

Taking the microscopic scale as the length unit is not specially convenient if one wants to describe only large distance properties. It is more convenient, as is usually done in particle physics, to fix the large distance scale one wants to describe and thus to study the small microscopic scale limit. One therefore rescales the cut-off in momentum space to a large value Λ and all distances correspondingly

$$x \mapsto \Lambda x. \tag{3}$$

For similar reasons one also rescales the field $\phi(x)$ in such a way that the coefficient of $[\nabla\phi(x)]^2$ becomes the standard 1/2

$$\phi(x) \mapsto \zeta\phi(x). \tag{4}$$

The choice of normalization for the gradient term implies:

$$\zeta = c^{-1/2}\Lambda^{1-d/2}, \tag{5}$$

which shows that ϕ now has in terms of Λ its canonical dimension $d/2 - 1$. In this way the critical domain corresponds to $\langle \phi(x) \rangle = O(1)$.

After this resealing all quantities have a dimension in units of Λ. The action $H(\phi)$ becomes:

$$\mathcal{H}(\phi) = \int \mathrm{d}^d x \left\{ \frac{1}{2} [\nabla\phi(x)]^2 + \frac{1}{2} r\phi^2(x) + \frac{1}{4!} g_0 \Lambda^{4-d} \phi^4(x) \right\}, \tag{6}$$

with:

$$r = a\Lambda^2/c, \qquad g_0 = b/c^2. \tag{7}$$

The study of the critical domain thus reduces to the study of the large cut-off behaviour, i.e. to renormalization theory and the corresponding renormalization group. However one extremely important feature of this action distinguishes it from field theory, at least in the form it is usually presented in particle physics: the coefficient of ϕ^4 has a dependence in the cut-off Λ which is given *a priori.* In particular in the dimensions of interest, $d < 4$, the "bare" coupling constant diverges, though the field theory, being super-renormalizable, requires only a mass renormalization.

1.2 The divergence of the bare coupling constant

To circumvent the problem of the large coupling constant two strategies have been devised. The first is the famous Wilson–Fisher ε-expansion [4]. One considers the dimension d as a continuous variable. Setting $d = 4 - \varepsilon$ one expands both in g_0 and ε. Divergences then behave like in four dimensions, they are only logarithmic and can be dealt with. Moreover it is possible to work directly in the massless (or critical) theory.

The second, initiated by Parisi [5], relies in working at fixed dimension $d < 4$, in the massive theory (the massless theory is IR divergent). The reason one has to face this problem is of practical nature: it is easier to calculate Feynman diagrams in dimension three than in generic dimensions, and thus more perturbative orders are available.

The strategy then is to take first the large cut-off limit at $u_0 = g_0 \Lambda^{4-d}$ fixed, and then the infinite u_0 limit. This implies that one first tunes the initial parameters of the model to remain artificially close to the unstable $u_0 = 0$ gaussian fixed point, When m the physical mass or inverse correlation length goes to zero near T_c one varies the dimensionless bare quantity g_0 as

$$g_0^{1/(4-d)} \propto m/\Lambda,$$

This approach has sometimes be questioned, but the final results and their comparison with other data have shown the consistency of the method.

Still one is confronted with a serious problem from the technical point of view: perturbation theory is finite but one is interested in the infinite coupling limit, which is obviously outside the perturbative domain. Parisi's idea was to introduce coupling constant and field renormalizations, $\phi = Z^{1/2}\phi_R$, as in four dimensions. They are implicitly defined by the renormalization conditions:

$$\begin{aligned} \Gamma_R^{(2)}(p; m, g) &= m^2 + p^2 + O(p^4), \\ (p_i = 0; m, g) &= m^{4-d} g. \end{aligned} \tag{8}$$

The role of these renormalizations, however, is here different. When the initial coupling constant u_0 becomes large the new dimensionless coupling g has a finite limit provided the corresponding Callan–Symanzik β-function has an IR stable zero g^*:

$$u_0 \to \infty \Rightarrow \beta(g^*) = 0 \quad \text{and} \, \omega \equiv \beta'(g^*) > 0 .$$

Indeed we recall the relations

$$\begin{aligned} u_0 = g_0 \Lambda^{(4-d)} &= m^{(4-d)} G(g), \\ \beta(g) &= (d-4) G(g)/G'(g). \end{aligned}$$

It follows that at g_0 fixed, $\Lambda/m \to \infty$,

$$g - g^* \propto (m/\Lambda)^{\omega}. \tag{9}$$

Therefore the renormalized coupling g is a more suitable expansion parameter then u_0.

The form of the β-function in d dimension is

$$\beta(g) = -(4-d)g + a_2(d) g^2 + \cdots . \tag{10}$$

For $\varepsilon = 4 - d$ small one finds a zero of order ε and therefore perturbation is applicable. One recovers the principle of the ε-expansion.

By contrast at fixed dimension three or two, there is no small parameter. Therefore an accurate determination of g^* and all quantities depends on the analytic properties of the series, in addition to the number of terms available. A semi-classical analysis, based on instanton calculus, unfortunately indicates that perturbation in quantum field theory is always divergent. Therefore to extract any information from perturbation theory a summation method is required. Note then that at any finite order the results for universal quantities become renormalization scheme dependent in contrast with the results of the ε-expansion.

2. CRITICAL EXPONENTS

The most studied quantities are critical exponents, because they are easier to calculate, and because they have been used to test RG predictions by comparing them with other results (experiments, high or low temperature series expansion, Monte-Carlo simulations) [6]. The first accurate determination of exponents of the $O(N)$ symmetric N-vector model has been presented in [7] using six-loop series for RG functions reported in [8]. The basic ideas in [7] to sum the series, are Borel transformation and conformal mapping. The same ideas have been later applied to the ε-expansion where five loop calculations have been performed, and more recently to the equation of state. With time the method has been refined and the efficiency improved by various tricks but the basic principles have not changed.

2.1 Borel transformation and conformal mapping

Let $R(g)$ be any quantity given by a perturbation series

$$R(g) = \sum_{k=0} R_k g^k. \tag{11}$$

One knows from large order behaviour analysis (instantons) in the ϕ^4 theory that R_k at large order k behaves like

$$R_k \underset{k\to\infty}{\propto} k^s(-a)^k k!\,.$$

The value of $a > 0$ has been numerically determined (while s is known analytically). One introduces $B(b, g)$, the Borel (rather Borel–Leroy) transform of $R(g)$, which is defined by

$$B(b,g) = \sum_{k=0} \frac{R_k}{\Gamma(b+k+1)} g^k, \tag{12}$$

where b is a free parameter. Formally, i.e. in the sense of series expansion, $R(g)$ can be recovered from

$$R(g) = \int_0^\infty t^b \mathrm{e}^{-t} B(b, gt)\mathrm{d}t\,, \tag{13}$$

Using the large order estimate in (12) one verifies that $B(z)$ is analytic at least in a circle of radius $1/a$ and its singularity closest to the origin is located at $z = -1/a$. Therefore $B(z)$, in contrast with $R(z)$, is determined by its series expansion. However, for relation (13) to make sense as a relation between functions, and not simply between formal series, one must know $B(z)$ on the whole real positive axis. This implies that $B(z)$ must be analytic near the axis, a result which has been proven in constructive field theory (as well as the property that the function $R(g)$ is indeed by given by (13)) [9]. Moreover it is necessary to continue analytically the function from the circle to the real positive axis. Consideration of more general instanton contributions strongly suggest that the Borel transform actually is analytic in a cut-plane, the cut being on the real negative axis, at the left of $-1/a$. Therefore a method to perform the analytic continuation is to conformally map the cut-plane onto a circle:

$$z \mapsto u(z) = \frac{\sqrt{1+az}-1}{\sqrt{1+za}+1}\,. \tag{14}$$

Table 1
Estimates of critical exponents in the $O(N)$ symmetric $(\phi^2)_3^2$ field theory.

N	0	1	2	3
g^*_{Ni}	1.413 ± 0.006	1.411 ± 0.004	1.403 ± 0.0030	1.390 ± 0.003
g^*	26.63 ± 0.11	23.64 ± 0.07	21.16 ± 0.05	19.06 ± 0.04
γ	1.1597 ± 0.0020	1.2398 ± 0.0012	1.3170 ± 0.0015	1.3893 ± 0.0029
ν	0.5882 ± 0.0009	0.6303 ± 0.0012	0.6701 ± 0.0012	0.7071 ± 0.0016
η	0.0285 ± 0.002	0.033 ± 0.002	0.035 ± 0.002	0.035 ± 0.002
β	0.3025 ± 0.0007	0.3257 ± 0.0008	0.3469 ± 0.0008	0.3660 ± 0.0013
α	0.235 ± 0.003	0.109 ± 0.004	–0.010 ± 0.004	–0.121 ± 0.005
ω	0.812 ± 0.016	0.800 ± 0.009	0.790 ± 0.007	0.785 ± 0.007
$\theta = \omega\nu$	0.48 ± 0.01	0.50 ± 0.01	0.53 ± 0.01	0.56 ± 0.01

The function $R(g)$ is then given by the new, hopefully convergent, expansion

$$R(g) = \sum_{k=0}^{\infty} B_k(b) \int_0^{\infty} t^b \mathrm{e}^{-t} [u(gt)]^k \, \mathrm{d}t \,. \tag{15}$$

The parameter b, as well as a few other parameters introduced in variants, are used to improve the apparent convergence and test the sensitivity of results to their variations. Moreover the value of b has to stay within a reasonable range around the value s predicted by the large order behaviour. Finally it is to be expected that the summation method will be efficient if the coefficients R_k are already approaching the asymptotic large order regime.

2.2 Exponents

The values of critical exponents obtained from field theory have remained after about twenty years among the most accurate determinations. Only recently have consistent, but significantly more accurate experimental results been reported (in low gravity superfluid experiments [10]). Also various numerical simulations [11] and high temperature expansions on the lattice have claimed similar accuracies.

The RG values of critical exponents are now in the process of being updated [12] because one additional term has been added to two of the three RG functions. Preliminary results are displayed in table 1. Compared with the previously published values the main improvement concerns the exponent η which was poorly determined, and the new lower value of γ for $N = 0$ (polymers).

3. ISING MODEL: THE SCALING EQUATION OF STATE

We now consider the scaling equation of state for the $N = 1$, $d = 3$ case (Ising-like systems). Before discussing the actual calculation at $d = 3$ fixed, let us however recall a few properties of the equation of the state in the critical domain, as well as the results which have been previously obtained within the framework of the ε-expansion.

The equation of state is the relation between magnetic field H, magnetization $M = \langle\phi\rangle$ (the "bare" field expectation value) and the temperature which is represented by the parameter $t \propto a - a_c$ (see (2)). It is related to the free energy per unit volume F, in field theory language the generating functional $\Gamma(\phi)$ of 1PI correlation functions restricted to constant fields, i.e the effective potential,

$$\mathcal{F}(M) = \Gamma(M)/\text{vol.}\,,$$

by

$$H = \frac{\partial \mathcal{F}}{\partial M}\,. \tag{16}$$

In the critical domain the equation of state has Widom's scaling form

$$H(M,t) = M^{\delta} f(t/M^{1/\beta}), \tag{17}$$

a form initially conjectured and which renormalization group has justified.

One property of the function $H(M, t)$ which plays an essential role in the analysis is *Griffith's analyticity*: it is regular at $t = 0$ for $M > 0$ fixed, and simultaneously it is regular at $M = 0$ for $t > 0$ fixed. *Amplitude ratios.* Universal amplitude ratios are numbers characterizing the behaviour of thermodynamical quantities near T_c. Several of them commonly considered in the literature can be directly derived from the scaling equation of state. Let us just give two examples.

The singular part of the specific heat, i.e. the ϕ^2 2-point correlation function at zero momentum, behaves like

$$C_H = A^{\pm}\,|t|^{-\alpha}\,, \qquad t \propto T - T_c \to \pm 0\,. \tag{18}$$

The ratio A^+/A^- then is universal.

The magnetic susceptibility χ in zero field, i.e. the ϕ 2-point function at zero momentum, diverges like

$$\chi = C^{\pm}\,|t|^{-\gamma}\,, \qquad t \to \pm 0\,. \tag{19}$$

The ratio C^+/C^- then is also universal.

3.1 The ε-expansion

Let us recall the results concerning the equation of state which have been obtained within the framework of the $\varepsilon = 4 - d$ expansion.

The ε-expansion of the scaling equation of state has been determined up to order ε^2 for the general $O(N)$ model, [13] and order ε^3 for $N = 1$ [14]. We give here the functions for $N = 1$ only up to order ε^2 to display the structure. The function $f(x = t/M^{1/\beta})$ of eq.(17) can be written:

$$f(x) = 1 + x + \frac{1}{6}\varepsilon(x+3)L + \varepsilon^2\left[\frac{1}{72}(x+9)L^2 + \frac{25}{324}(x+3)L\right] + O\left(\varepsilon^3\right) \tag{20}$$

with $L = \log(x + 3)$.

The expression (20) is not valid for x large, i.e. for small magnetization M. In this regime the magnetic field H has a regular expansion in odd powers of M. It is thus

convenient to express the equation of state in terms of another scaling variable $z \propto x^{-\beta}$ because

$$z \propto Mt^{-\beta}. \tag{21}$$

The equation of state then takes the form

$$H \propto t^{\beta\delta} F(z), \tag{22}$$

where the relation between exponents $\gamma = \beta(\delta - 1)$ has been used. Substituting into eq.(20) $x = x_0 z^{-1/\beta}$ (the constant x_0 takes care of the normalization of z) and expanding in ε one finds at order ε^2 for the function (22)

$$F(z) = \tilde{F}_0(z) + \varepsilon \tilde{F}_1(z) + \varepsilon^2 \tilde{F}_2(z) + O(\varepsilon^3),$$

with

$$\begin{aligned} \tilde{F}_0 &= z + \tfrac{1}{6} z^3 \\ \tilde{F}_1 &= \tfrac{1}{12}(-z^3 + \tilde{L}(2z + z^3)) \\ \tilde{F}_2 &= \tfrac{1}{1296}(-50z^3 + \tilde{L}(100z - 4z^3) + \tilde{L}^2(18z + 27z^3)) \end{aligned} \tag{23}$$

and $\tilde{L} = \log(1 + z^2/2)$.

Within the framework of the formal ε-expansion one can easily pass from one expansion to the other. Still a matching problem arises if one wants to use the ε-expansion to determine the equation of state for $d = 3$, i.e. $\varepsilon = 1$. One is thus naturally led to look for a uniform representation valid in both limits. Josephson–Schofield parametric representation [15] has this property.

4. PARAMETRIC REPRESENTATION OF THE EQUATION OF STATE

One parametrizes the equation of state in terms of two new variables R and θ, setting:

$$\begin{cases} M &= m_0 R^\beta \theta, \\ t &= R(1 - \theta^2), \\ H &= h_0 R^{\beta\delta} h(\theta), \end{cases} \tag{24}$$

where h_0, m_0 are two normalization constants and h_0 can be chosen such that

$$h(\theta) = \theta + O\left(\theta^3\right).$$

This parametrization also corresponds in terms of the scaling variables x of eq. (20) or z from eq.(21) to set

$$z = \rho\theta/(1 - \theta^2)^\beta, \quad \theta > 0, \tag{25}$$

$$x = x_0 \rho^{-1/\beta} \left(1 - \theta^2\right) \theta^{-1/\beta}, \tag{26}$$

where ρ is some other positive constant.

Then the function $h(\theta)$ is an odd function of θ which from Griffith's analyticity is regular near $\theta = 1$, which is x small, and near $\theta = 0$ which is x large. It vanishes for $\theta = \theta_0$ which corresponds to the coexistence curve $H = 0$, $T < T_c$.

Note that the mapping (25) is not invertible for values of θ such that $z'(\theta) = 0$. The derivative vanishes for $\theta = 1/\sqrt{(1-2\beta)} \approx 1.69$. One verifies that this value is reasonably larger than θ_0, the largest possible value of θ.

Finally it is useful for later purpose to write more explicitly the relation between the function $F(z)$ of eq. (22) and the function $h(\theta)$:

$$h(\theta) = \rho^{-1} \left(1 - \theta^2\right)^{\beta\delta} F\left(z(\theta)\right). \tag{27}$$

Expanding both functions

$$F(z) = z + \frac{1}{6} z^3 + \sum_{l=2} F_{2l+1} z^{2l+1}, \tag{28}$$

$$h(\theta)/\theta = 1 + \sum_{l=1} h_{2l+1} \theta^{2l}, \tag{29}$$

one finds the relations

$$\begin{aligned} h_3 &= \tfrac{1}{6}\rho^2 - \gamma \\ h_5 &= \tfrac{1}{2}\gamma(\gamma-1) + \tfrac{1}{6}(2\beta-\gamma)\rho^2 + F_5\rho^4 \\ h_7 &= \tfrac{1}{6}\gamma(\gamma-1)(\gamma-2) + \tfrac{1}{12}(2\beta-\gamma)(2\beta-\gamma+1)\rho^2 \\ &\quad + (4\beta-\gamma)F_5\rho^4 + F_7\rho^6 \end{aligned} \tag{30}$$

From the parametric representation of the equation of state it is then possible to derive a representation for the singular part of the free energy per unit volume as well as various universal ratios of amplitudes.

Parametric representation and ε-expansion. Up to order ε^2 the constant m_0 (or ρ) can be chosen in such a way that the function $h(\theta)$ reduces to:

$$h(\theta) = \theta\left(1 - \frac{2}{3}\theta^2\right) + O\left(\varepsilon^2\right). \tag{31}$$

The simple model in which $h(\theta)$ is approximated by a cubic odd function of θ is called the linear parametric model. At order ε^2 the linear parametric model is exact, but at order ε^3 the introduction of a term proportional to θ^5 becomes necessary [14,16]. One finds:

$$h(\theta) = \theta(1 + h_3\theta^2 + h_5\theta^4) + O\left(\varepsilon^4\right), \tag{32}$$

with

$$h_3 = -\frac{2}{3}\left(1 + \frac{\varepsilon^2}{12}\right), \quad h_5 = \frac{\varepsilon^3}{27}\left(\zeta(3) - \frac{1}{2}\lambda - \frac{1}{4}\right), \tag{33}$$

where λ is the constant

$$\lambda = \frac{1}{3}\psi'(1/3) - \frac{2}{9}\pi^2 = 1.17195361934\ldots \tag{34}$$

The function $h(\theta)$ vanishes on the coexistence curve for $\theta = \theta_0$:

$$\theta_0^2 = \frac{3}{2}\left(1 - \frac{\varepsilon^2}{12}\right) + O(\varepsilon^3). \tag{35}$$

Note that h_3 and thus θ_0 are determined only up to order ε^2. It follows

$$\rho^2 = 6(\gamma + h_3) = 2\left(1 + \frac{1}{2}\varepsilon + \frac{7}{108}\varepsilon^2\right) = 3.13 \pm 0.13,$$

because h_3 is determined only up to order ε^2,

Remark. In the more general $O(N)$ case, the parametric representation also automatically generates equations of state with satisfy the required regularity properties, and thus leads to uniform approximations. However for $N > 1$ the function $h(\theta)$ still has a singularity on the coexistence curve, due to the presence of Goldstone modes in the ordered phase and has therefore a more complicated form. The nature of this singularity can be obtained from the study of the non-linear σ-model. It is not clear whether a simple polynomial approximation would be useful. For $N = 1$ instead, one expects at most an essential singularity on the coexistence curve, due to barrier penetration, which is much weaker and non-perturbative in the small ε- or small g-expansion.

5. THE PERTURBATIVE EXPANSION AT FIXED DIMENSION THREE

We now discuss the calculations based on perturbative expansion at fixed $d = 3$ dimension. Five loop series for the renormalized effective potential of the ϕ_3^4 theory have been first reported in Bagnuls *et al.* [17], but the printed tables contain some serious misprints. These have been noticed by Halfkann and Dohm who have published corrected values [18]. We do not consider the case $N \neq 1$. Because the perturbative calculations are much more difficult, due to the appearance of two lengths, the correlation length along the applied field and transverse to it, the series are too short.

5.1 General remarks

The general framework again is the massive theory renormalized at zero momentum. The correlation functions $\Gamma_R^{(n)}$ of the renormalized field $\phi_R = \phi/\sqrt{Z}$ are fixed by the normalization conditions (8). Eventually the renormalized coupling constant g has to be set to its IR fixed point value g^*. The mass parameter m is proportional to the physical mass, or inverse correlation length, of the high temperature phase. It behaves for $t \propto T - T_c \to 0_+$ as $m \propto t^\nu$, where ν is the correlation length exponent.

From the conditions (8) it follows that the free energy F expressed in terms of the "renormalized" magnetization φ, i.e. the expectation value of the renormalized field $\varphi = \langle \phi_R \rangle$, has a small φ expansion of the form (in d dimensions)

$$\mathcal{F}(\varphi) = \mathcal{F}(0) + \frac{1}{2}m^2\varphi^2 + \frac{1}{4!}m^{4-d}g\varphi^4 + O\left(\varphi^6\right). \tag{36}$$

It is important to remember that the field renormalization Z is singular at g^*. For $g \to g^*$ it behaves like (see also eq. (9))

$$Z \propto (g^* - g)^{\eta/\omega} \propto (m/\Lambda)^\eta,$$

where the exponent η characterizes the field anomalous dimension. It follows that $\varphi \propto M/m^{\eta/2}$.

It is convenient to introduce the rescaled variable z

$$z = \varphi m^{(2-d)/2}\sqrt{g} \propto M/m^{(d-2+\eta)/2}, \tag{37}$$

and set

$$\mathcal{F}(\varphi) - \mathcal{F}(0) = \frac{m^d}{g} V(z, g). \tag{38}$$

Taking into account the definitions of section (3), as well as the definitions (21),(22) we conclude that $z \propto M t^{-\beta}$ and that the equation of state is related to the derivative F of the reduced effective potential V with respect to z

$$F(z, g) = \frac{\partial V(z, g)}{\partial z}. \tag{39}$$

by

$$H \propto t^{\beta\delta} F(z),$$

Ising symmetry implies that F is an odd function of z

$$F(z, g) = z + \frac{1}{6} z^3 + \sum_{l=2} F_{2l+1}(g) z^{2l+1}. \tag{40}$$

5.2 The problem of the low temperature phase

To determine the equation of state in the whole physical range, or universal ratios of amplitudes, a new problem arises. In this framework it is more difficult to calculate physical quantities in the ordered phase because the theory is parametrized in terms of the disordered phase correlation length $\xi = m^{-1} \propto (T - T_c)^{-\nu}$ which is singular at T_c (as well as all correlation functions normalized as in 8). Let us consider the perturbative expansion of the scaling equation of state (39). For example at one-loop order for $d = 3$ the function $F(z, g)$ is given by :

$$\begin{aligned} F(z, g) &= z + \tfrac{1}{6} z^3 - \tfrac{1}{8\pi} g z \left[(1 + z^2/2)^{1/2} - 1 - z^2/4 \right] \\ &= z + \tfrac{1}{6} z^3 + \tfrac{1}{256\pi} g z^5 - \tfrac{1}{2^{13}\pi} g z^7 + O(z^9), \end{aligned} \tag{41}$$

where the subtractions, due to the mass and coupling normalizations, are determined by the conditions (8). This expression is adequate for the description of the disordered phase, but all terms in the loopwise expansion become singular when t goes to zero for fixed magnetization. Indeed this corresponds to the limit $m \to 0$, at φ fixed and thus $z \to \infty$ as seen in eq. (21). In this limit one knows from eq. (17) that the equation of state becomes

$$H(M, t = 0) \propto M^\delta \;\Rightarrow\; F(z) \propto z^\delta. \tag{42}$$

In the case of the ε-expansion the scaling relations (and thus the limiting behaviour (42) are exactly satisfied order by order. Moreover the change to the variable $x \propto z^{-1/\beta}$ (more appropriate for the regime $t \to 0$) gives an expression for

$f(x) \propto F(x^{-\beta})x^{\beta\delta}$ that is explicitly regular in $x = 0$ (Griffith's analyticity): the singular powers of log x induced by the change of variables cancel non trivially at each order, leaving only regular corrections.

The situation changes when one deals with the perturbation theory in $d = 3$ dimensions: scaling is not satisfied for generic values of g but only at g^*. Consequently scaling properties are not satisfied order by order in g. In particular the change to the Widom function $f(x)$ will introduce singular terms in log x that violate Griffith's analyticity. An analogous problem arises if one first sums the series at $g = g^*$ before changing to the variable x. In this case the singular contributions (in the form of powers of x) do not cancel, as a result of unavoidable numerical summation errors.

Several approaches can be used to deal with the problem of continuation to the ordered phase. A rather powerful method, motivated by the results obtained within the ε-expansion scheme, is based on the parametric representation.

The parametric representation. Order dependent mapping (ODM). The problem that one faces is the following: to reach the ordered region $t < 0$ one must cross the point $z = \infty$. However we know from Griffith's analyticity that $F(z)z^{-\delta}$ is regular in the variable $z^{-1/\beta}$. This property is naturally satisfied within the parametric representation. One thus introduces an new variable θ and an auxiliary function $h(\theta)$ defined as in (25),(27) : in this way the exact function $h(\theta)$ will be regular near $\theta = 1$ (i.e. $z = \infty$) and up to the coexistence curve. However, the approximate $h(\theta)$ that one obtains by summing perturbation theory at fixed dimension, is still not regular. The singular terms generated by the mapping eq. (25) at $\theta = 1$ do not cancel exactly due to summation errors. The last step is to Taylor expand the approximate expression of $h(\theta)$ around $\theta = 0$ and to truncate the expansion, enforcing in this way regularity. A question then arises, to which order in θ should one expand? Since the coefficients of the θ expansion are in one to one correspondence with the coefficients of the small z expansion of the function $F(z, g^*)$, the maximal power of θ in $h(\theta)$, should be equal to the maximal power of z whose coefficient can be determined with reasonable accuracy. Indeed although the small z expansion of $F(z)$ at each finite loop order in g contains an infinite number of terms, the evaluation of the coefficients of the higher powers of z is increasingly difficult. The reasons are twofold:

(i) The number of terms of the series in g required to get an accurate estimate of F_l increases with l (see section (6.1)).

(ii) At any finite order in g the function $F(z)$ has spurious singularities in the complex z plane (see e.g. eq. (41), $z^2 = -2$) that dominate the behaviour of the coefficients F_l for l large.

In view of these difficulties one has to ensure the fastest possible convergence of the small θ expansion. For this purpose one can use the freedom in the choice of the arbitrary parameter ρ in eq. (25) : one determines ρ by minimizing the last term in the truncated small θ expansion, thus increasing the importance of small powers of θ which are more accurately calculated. This is nothing but the application to this particular example of the series summation method based on ODM [19].

This strategy applied to the available data, leads at leading order for $h(\theta)$ to a polynomial of degree 5, whose coefficients are given by the relations (30) :

$$h(\theta) = \theta[1 + h_3(\rho)\theta^2 + h_5(\rho)\theta^4]. \tag{43}$$

For the range of admissible values for F_5 the coefficient h_5 of θ^5 given by eq. (30) has no real zero in ρ. It has a minimum instead

$$\rho^2 = \rho_5^2 = \frac{1}{12F_5}(\gamma - 2\beta). \tag{44}$$

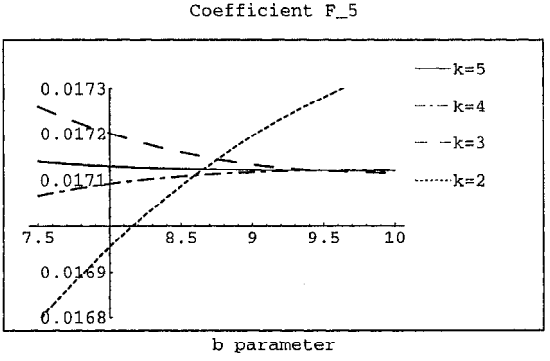


Figure 1. The summed coefficient F_5 as a function of the Borel–Leroy parameter b for successive orders k.

Substituting this value of ρ into expression (43) one obtains the first approximation for $h(\theta)$. At next order one looks for a minimum ρ_7 of $|h_7(\rho)|$. One finds a polynomial either of degree 5 in θ, when h_7 has a real zero, or of degree 7 when it has only a minimum.

It is not possible to go beyond $h_9(\rho)$ because already F_9 is too poorly determined.

Finally amplitude ratios will involve ratios of functions of θ at $\theta = 0$ and at $\theta = \theta_0$ where θ_0 is the zero of $h(\theta)$ closest to the origin.

6. NUMERICAL RESULTS

6.1 The small field expansion

The first problem thus is to determine the first coefficients F_{2l+1} of the small z expansion of the function $F(z)$ as accurately as possible, at the I.R. fixed point g^*. Exactly the same method as for exponents is used, i.e. Borel–Leroy transformation and conformal mapping. As one can anticipate the asymptotic regime sets in later when l increases, and thus the efficiency of the summation correspondingly decreases.

In figure 1 the behaviour of F_5 in terms of the Borel–Leroy parameter b is displayed. Increasing flatness of the curves when k increases, i.e. increasing insensitivity to the parameter b, presumably indicates convergence of the method. Table 2 contains the results of [1] together with other published estimates of the coefficients of the small z expansion of $F(z)$.

6.2 Parametric representation

One then determines by the ODM method the parameter ρ and the function $h(\theta)$, as explained in section (5.2). One obtains successive approximations in the form of polynomials of increasing degree for $h(\theta)$. Note that one here has a simple test of the

Table 2
The effective potential.

	g^*	F_5	F_7	F_9
ε–exp.	28.	0.0176 ± 0.0004	(4.5 ± 0.3) × 10^{-4}	– (3.2 ± 0.2) × 10^{-5}
$d = 3$	23.70 ± 0.05	0.01711 ± 0.00006	(4.8 ± 0.5) × 10^{-4}	– (7. ± 5.) × 10^{-5}
$d = 3$ [21]	23.73	.01727	.0010	
HT [22]	23.72 ± 1.49	0.0205 ± 0.0052		
HT [20]	24.45 ± 0.15	.017974 ± .00015		
HT [23]	23.69	.0166	.00055	.00001
MC [24]	23.3 ± 0.5	0.0227 ± 0.0026		
MC [25]	24.5 ± .2	0.027 ± 0.002	0.00236 ± .00040	
ERG [26]	28.9	0.016	4.3 × 10^{-4}	
ERG [27]	20.7 ± 0.2	0.0173 ± 0.0001	(5.0 ± 0.2) × 10^{-4}	– (4. ± 2.) × 10^{-5}

relevance of the ODM method. Indeed, once $h(\theta)$ is determined, assuming the values of the critical exponents γ and β, one can recover a function $F(z)$ which has an expansion to all orders in z. As a result one obtains a prediction for the coefficients F_{2l+1} which have not yet been taken into account to determine $h(\theta)$. The relative difference between the predicted values and the ones directly calculated gives an idea about the accuracy of the ODM method. Indeed from the values $F_5 = 0.01711$, $\gamma = 1.2398$, $\beta = 0.3256$, one obtains

$$F_7 = 4.83 \times 10^{-4}, \quad F_9 = -3.2 \times 10^{-5}, \quad F_{11} = 1.4 \times 10^{-7} \cdots .$$

Note that the value for F_7 is quite close to the central value one finds by direct series summation, while the value for F_9 is within the errors. This result gives us confidence in the method. It also shows that the value of F_9 obtained by direct summation contains little new information, it provides only a consistency check. Therefore the simplest representation of the equation of state, consistent with all data, is given by

$$h(\theta)/\theta = 1 - 0.76201(36)\,\theta^2 + 8.04(11) \times 10^{-3}\,\theta^4, \tag{45}$$

(errors on the last digits in parentheses) that is obtained from $\rho^2 = 2.8667$ (fixed according to eq. (44)). This expression of $h(\theta)$ has a zero at

$$\theta_0 = 1.154, \tag{46}$$

to which corresponds the value of the complex root z_0 of $F(z)$, $z_0 = 2.800 \times e^{-i\pi\beta}$. The coefficient of θ^7 in eq. (45) is smaller than 10^{-3}. Note that for the largest value of θ^2 which corresponds to θ_0^2, the θ^4 term is still a small correction. Finally the corresponding values for the ε-expansion are $h_3 = -0.72$, $h_5 = 0.013$. These values are reasonably consistent, because a small change in h_3 can be cancelled to a large extent by a correlated change in ρ.

The Widom scaling function $f(x)$, eq. (17), can then easily be obtained numerically from $h(\theta)$ and compared with other determinations. The main disagreement with other predictions comes from the region $x \to \infty$, i.e. from the small magnetization region, where the predictions of the present method should be specially reliable.

Table 3
Amplitude ratios.

	A^+/A^-	C^+/C^-	R_c	R_χ
ε – exp. (*a*)	0.524 ± 0.010	4.9		1.67
ε – exp. (*b*)	0.547 ± 0.021	4.70 ± 0.10	0.0585 ± 0.0020	1.649 ± 0.021
$d = 3$ fixed (*a*)	0.541 ± 0.014	4.77 ± 0.30	0.0594 ± 0.001	1.7
$d = 3$ fixed (*b*)	0.536 ± 0.019	4.80 ± 0.10	0.0575 ± 0.0020	1.670 ± 0.019
HTseries	0.523 ± 0.009	4.95 ± 0.15	0.0581 ± 0.0010	1.75
binary mixt.	0.56 ± 0.02	4.3 ± 0.3	0.050 ± 0.015	1.75 ± 0.30
liqu. – vap.	0.48 – 0.53	4.8–5.2	0.047 ± 0.010	1.69 ± 0.14
magn. systems	0.49–0.54	4.9 ± 0.5		

6.3 Amplitude ratios

Various amplitude ratios can then be derived from $h(\theta)$ and the values of the critical exponents determined from longer series. Table 3 contains a comparison of four amplitude ratios as obtained from RG, lattice calculations and experiments on binary mixtures, liquid–vapour, uniaxial magnetic systems. Results for ε-expansion (*a*) are taken from [28] and [16] (direct Padé summation of each corresponding series), while (*b*) are the results of [1], obtained by first summing $h(\theta)$ and then computing ratios. The results $d = 3$ fixed dimension (*a*) are taken from [17] and refers to direct summation up to $O(g^5)$ while $d = 3$ (*b*) are the results obtained by the method of [1], with the values of [12]. High Temperature results are taken from [29] (R_χ from [30]). Experimental data are extracted from [31], to which we refer for more results and references. One notes the overall consistency of the results obtained by different methods.

7. CONCLUDING REMARKS

Within the framework of renormalized quantum field theory and renormalization group, the presently available series allow, after proper summation, to determine with satisfactory accuracy critical exponents for the N-vector model and the complete scaling equation of state for 3D Ising-like ($N = 1$) systems. In the latter example additional technical tools, beyomd Borel summation methods, are required in which the parametric representation plays a central role. From the equation of state new estimates of some amplitude ratios have been deduced which seem reasonably consistent with all other available data.

Clearly a similar strategy could be applied to other quantities in a magnetic field, in the scaling region. Note also that an extension of the ε-expansion of the equation of state for $N = 1$ to order ε^4 or even better ε^5, that does not seem an unrealistic goal, would significantly improve the ε-expansion estimates and would therefore be quite useful. Finally the present approach could be extended to systems in the universality class of the $(\phi^2)^2_3$ field theory for $N \neq 1$, provided expansions of the renormalized effective potential at high enough order can be generated.

REFERENCES

1. R. Guida and J. *Zinn-Justin, Nucl. Phys.* B489 [FS] (1997) 626, (preprint hep-th/9610223).
2. For an early review see K.G. Wilson and J. Kogut, *Phys. Rep.* 12C (1974) 75.
3. E. Brézin, J.C. Le Guillou and J. Zinn-Justin, in *Phase Transitions and Critical Phenomena* vol. 6, C. Domb and M.S. Green eds. (Academic Press, London 1976).
4. K.G. Wilson and M.E. Fisher, *Phys. Rev. Lett.* 28 (1972) 240.
5. G. Parisi, *Cargèse Lectures 1973,* published in *J. Stat. Phys.* 23 (1980) 49.
6. A still partially relevant reference is *Phase Transitions* vol. B72, M. Lévy, J.C. Le Guillou and J. Zinn-Justin eds. Proceedings of the Cargèse Summer Institute 1980 (Plenum, New York 1982).
7. J.C. Le Guillou and J. Zinn-Justin, *Phys. Rev. Lett.* 39 (1977) 95; *Phys. Rev.* B21 (1980) 3976.
8. B.G. Nickel, D.I.Meiron, G.B. Baker, Univ. of Guelph Report 1977.
9. J.P. Eckmann, J. Magnen and R. Sénéor, *Commun. Math. Phys.* 39 (1975) 251
   J.S. Feldman and K. Osterwalder, *Ann. Phys. (NY)* 97 (1976) 80
   J. Magnen and R. Sénéor, *Commun. Math. Phys.* 56 (1977) 237
   J.-P. Eckmann and H. Epstein, *Commun. Math. Phys.* 68 (1979) 245.
10. J.A. Lipa, D.R. Swanson, J. Nissen, T.C.P. Chui and U.E. Israelson, *Phys. Rev. Lett.* 76 (1996) 944.
11. C.F. Baillie, R. Gupta, K.A. Hawick and G.S. Pawley, *Phys. Rev.* B45 (1992) 10438, and references therein
    R. Gupta and P. Tamayo, *Critical exponents of the 3D Ising model,* LA UR-96-93 preprint, cond-mat/9601048
    B. Li, N. Madras and A.D. Sokal, *J. Stat. Phys.* 80 (1995) 661
    S. Caracciolo, M.S. Causo and A. Pelissetto, preprint cond-mat/9703250.
12. R. Guida and J. Zinn-Justin, Saclay preprint T97/040 (1997), in preparation.
13. G.M. Avdeeva and A.A. Migdal, *JETP Lett.* 16 (1972) 178
    E. Brézin, D.J. Wallace and K.G. Wilson, *Phys. Rev. Lett.* 29 (1972) 591; *Phys. Rev.* B7 (1973) 232.
14. D.J. Wallace and R.P.K. Zia, *J. Phys. C: Solid State Phys.* 7 (1974) 3480.
15. P. Schofield, *Phys. Rev. Lett.* 23 (1969) 109
    P. Schofield, J.D. Litster and J.T. Ho, *Phys. Rev. Lett.* 23 (1969) 1098
    B.D. Josephson, *J. Phys. C: Solid State Phys.* 2 (1969) 1113.
16. J.F. Nicoll and P.C. Albright, *Phys. Rev.* B31 (1985) 4576.
17. C. Bagnuls, C. Bervillier, D.I. Meiron and B.G. Nickel, *Phys. Rev.* B35 (1987) 3585.
18. F.J. Halfkann and V. Dohm, *Z. Phys.* B89 (1992) 79.
19. R. Seznec and J. Zinn-Justin, *J. Math. Phys.* 20 (1979) 1398
    J.C. Le Guillou and J. Zinn-Justin, *Ann. Phys, (NY)* 147 (1983) 57
    R. Guida, K. Konishi and H. Suzuki, Ann. Phys. *(NY)* 241 (1995) 152; 249 (1996) 109.
20. S. Zinn, S.-N. Lai, and M.E. Fisher, *Phys. Rev. E* 54 (1996) 1176.
21. A.I. Sokolov, V.A. Ul'kov and E.V. Orlov, Proceedings of the conference *Renormalization Group 96,* Dubna 1996, data obtained from Padé Borel summation up to order g^3. See also A.I. Sokolov, *Fizika Tverdogo Tela (Solid State Physics)* 38(1996) 640.
22. T. Reisz,*Phys. Lett.* B360 (1995) 77.
23. P. Butera and M. Comi, *Phys. Rev.* E55 (1997) 6391.
24. M.M. Tsypin,*Phys. Rev. Lett.* 73 (1994) 2015.
25. J-K Kim and D.P. Landau, hep-lat/9608072.
26. N. Tetradis and C. Wetterich, *Nucl. Phys.* B422 (1994) 541.
27. T. R. Morris, *Nucl.Phys.* B495 (1997) 477.
28. C. Bervillier, *Phys. Rev.* B34 (1986) 8141.
29. A.J. Liu and M.E. Fisher, *Physica* A156 (1989) 35.
30. A. Aharony and P.C. Hohenberg, *Phys. Rev.* B13 (1976) 3081; *Physica* 86-88B (1977) 611.
31. V.Privman, P.C. Hohenberg, A. Aharony, *Universal Critical Point Amplitude Relations,* in Phase Transitions and Critical Phenomena vol. 14, C. Domb and J.L. Lebowitz eds., (Academic Press 1991).

COLLECTIVE DYNAMICS OF A DOMAIN WALL — AN OUTLINE

H. Arodź

Institute of Physics, Jagellonian University,
Reymonta 4, 30-059 Cracow, Poland

We shall consider domain walls in a relativistic field-theoretical model defined by the following Lagrangian

$$\mathcal{L} = -\frac{1}{2}\eta_{\mu\nu}\partial^{\mu}\Phi\partial^{\nu}\Phi - \frac{\lambda}{2}(\Phi^2 - \frac{M^2}{4\lambda})^2, \tag{1}$$

where Φ is a single real scalar field, $(\eta_{\mu\nu})$=diag(-1,1,1,1) is the metric in Minkowski space-time, and λ, M are positive parameters.

Euler-Lagrange equation corresponding to Lagrangian (1) has the particular time-independent solution

$$\Phi(x^{\mu}) = \Phi_0 \tanh(\frac{x^3}{2l_0}) \tag{2}$$

which describes a static, planar domain wall stretched along the $x^3 = 0$ plane. Here $\Phi_0 = M/(2\sqrt{\lambda})$ denotes one of the two vacuum values of the field Φ — the other one is equal to $-\Phi_0$. The parameter M can be identified with the mass of the scalar particle related to the field Φ. The corresponding Compton length $l_0 = M^{-1}$ gives the physical length scale in the model. Energy density for the planar domain wall (2) is exponentially localised in a vicinity of the $x^3 = 0$ plane. The transverse width of the domain wall is of the order $2l_0$.

The issue is time evolution of a non-planar domain wall. Such domain walls can be infinite, consider, e.g., locally deformed planar domain wall or a cylindrical domain wall. They can also be finite closed, e.g., like a sphere or a torus. We shall restrict our considerations to a single large, smooth domain wall. Such a domain wall is defined by a set of conditions which provide Lorentz and reparametrisation invariant formulation of the heuristic requirement that $l_0^2 R_{1,2} << 1$ at each point of the domain wall, where R_1, R_2 denote local main curvature radia in a local rest frame of the considered infinitesimal piece of the domain wall. In this case it is possible to develop the presented below perturbative approach to the dynamics of the domain wall. In other cases, e.g. when $l_0/R_i \sim 1$ at certain points of the domain wall, the only practical tool is numerical analysis.

New Developments in Quantum Field Theory
Edited by Damgaard and Jurkiewicz, Plenum Press, New York, 1998

Physical idea underlying the expansion in the width is rather simple. The domain wall is a stable solitonic object. One expects that each piece of the large, smooth domain wall in the local co-moving reference frame does not differ much from the planar domain wall. The differences, which are due to curvature of the domain wall, can be calculated perturbatively.

The tricky point is to find a convenient reference frame co-moving with the domain wall. From the seminal papers [1,2], which were devoted to dynamics of vortices but this does not make an essential difference here, we know that the co-moving reference frame should be based on a relativistic membrane. In the leading approximation[1,2] the membrane is of Nambu-Goto type, and at each instant of time it coincides with the surface on which the scalar field Φ vanishes. In the following we will call this surface the core of the domain wall. For instance, in the case of the static domain wall (2) the core is given by the plane $x^3 = 0$. When calculating corrections to the leading approximation one can still adhere to this identification of the membrane with the core, see, e.g., papers [3,4,5]. The price for this is that the membrane is no longer of the Nambu-Goto type. Effective action for such membrane contains higher derivatives with respect to time, and probably is nonlocal [6]. The reason for these unpleasant features is that the points at which the scalar field vanishes do not constitute a physical object (the physical object is the domain wall itself and not the core), and therefore the core, being a purely mathematical construct, can have strange from the physical point of view equation of motion. In papers [7,8] we have proposed to use a membrane which is of Nambu-Goto type to all orders. This membrane coincides with the core only at the initial instant of time.

Our description of the dynamics of the domain wall [7,8] involves the Nambu-Goto membrane and certain additional 2+1-dimensional fields defined on the membrane. They obey nonlinear equations of motion which are second order partial differential equations, and they describe time evolution of the domain wall completely. Transverse profile of the domain wall, that is dependence of the scalar field Φ on a variable changing in the direction perpendicular to the domain wall, is uniquely and explicitly given once the evolution of the Nambu-Goto membrane and of the 2+1-dimensional fields is found. The dependence on the transverse variable is calculated perturbatively, and one should observe that it is the case of a singular perturbation. Our approach to constructing the perturbative expansion follows a method used in condensed matter physics[9]. Because of the explicit dependence on the transverse variable the dynamics of the domain wall has been reduced to dynamics of the collective degrees of freedom given by the Nambu-Goto membrane and the fields on it. In the following we shall present basic steps of the approach developed in papers [7,8].

Step 1. The co-moving coordinate system.

We introduce a surface S co-moving with the domain wall, that is *the co-moving membrane.* It does not have to coincide with the core, except at the initial instant of time. The world-volume of S, denoted by Σ, is parametrised as follows

$$\Sigma \ni (Y^\mu)(u^a) = (\tau, Y^i(u^a)). \tag{3}$$

We use the notation $(u^a)_{a=0,1,2} = (\tau, \sigma^1, \sigma^2)$, where τ coincides with the laboratory frame time x^0, while σ^1, σ^2 parametrise S at each instant of time. The index $i = 1,2,3$ refers to the spatial components of the four-vector. The points of the co-moving membrane S at the instant τ_0 are given by $(Y^i)(\tau_0, \sigma^1, \sigma^2)$. The coordinate system $(\tau, \sigma^1, \sigma^2, \xi)$ co-moving with the domain wall is defined by the formula

$$x^\mu = Y^\mu(u^a) + \xi\, n^\mu(u^a), \tag{4}$$

where x^μ are Cartesian (laboratory frame) coordinates in Minkowski space-time, and (n^μ) is a normalised space-like four-vector orthogonal to Σ in the covariant sense,

$$n_\mu(u^a)Y^\mu_{,a}(u^a) = 0, \quad n_\mu n^\mu = 1,$$

where $Y^\mu_{,a} \equiv \partial Y^\mu/\partial u^a$. The three four-vectors $Y_{,a}$ are tangent to Σ. ξ is the transverse variable. Definition (4) implies that ξ and u^a are Lorentz scalars. In the co-moving coordinates the co-moving membrane is described by the condition $\xi = 0$. For points lying on S the parameter τ coincides with the laboratory time x^0, but for $\xi \neq 0$ in general τ is not equal to x^0.

The extrinsic curvature coefficients K_{ab} and induced metrics g_{ab} on Σ are defined by the following formulas:

$$K_{ab} = n_\mu Y^\mu_{,ab}, \quad g_{ab} = Y^\mu_{,a} Y_{\mu,b},$$

where $a, b = 0, 1, 2$. The covariant metric tensor in the new coordinates has the following form

$$[G_{\alpha\beta}] = \begin{bmatrix} G_{ab} & 0 \\ 0 & 1 \end{bmatrix},$$

where $\alpha, \beta = 0, 1, 2, 3$; $\alpha = 3$ corresponds to the ξ coordinate; and

$$G_{ab} = N_{ac} g^{cd} N_{db}, \quad N_{ac} = g_{ac} - \xi K_{ac}.$$

Thus, $G_{\xi\xi} = 1$, $G_{\xi a} = 0$. Straightforward computation gives

$$\sqrt{-G} = \sqrt{-g}\, h(\xi, u^a),$$

where as usual $g = det[g_{ab}]$, $G = det[G_{\alpha\beta}]$, and

$$h(\xi, u^a) = 1 - \xi K^a_a + \frac{1}{2}\xi^2 (K^a_a K^b_b - K^b_a K^a_b) - \frac{1}{3}\xi^3 K^a_b K^b_c K^c_a.$$

For raising and lowering the latin indices of the extrinsic curvature coefficients we use the induced metric tensors g^{ab}, g_{ab}.

The inverse metric tensor $G^{\alpha\beta}$ is given by the formula

$$[G^{\alpha\beta}] = \begin{bmatrix} G^{ab} & 0 \\ 0 & 1 \end{bmatrix},$$

where

$$G^{ab} = (N^{-1})^{ac} g_{cd} (N^{-1})^{db}.$$

N^{-1} is just the matrix inverse to $[N_{ab}]$. It has the upper indices by definition.

In general, the coordinates (u^a, ξ) are defined locally, in a vicinity of the world-volume Σ of the membrane. Roughly speaking, the allowed range of the ξ coordinate is determined by the smaller of the two main curvature radia of the membrane in a local rest frame. We assume that this curvature radius is sufficiently large so that on the outside of the region of validity of the co-moving coordinates there are only exponential tails of the domain wall, that is that the field ϕ is exponentially close to one of the two vacuum solutions.

Step 2. Field equation in the co-moving coordinates.
It is convenient to rescale the field Φ and the coordinate ξ,

$$\Phi(x^\mu) = \frac{M}{2\sqrt{\lambda}}\phi(s, u^a), \quad \xi = \frac{2}{M}s,$$

where ϕ and s are dimensionless. We also extract from the scalar field its component living on the co-moving membrane and treat it separately from the remaining part of the scalar field. To this end we write the identity

$$\phi(s, u^a) = B(u^a)\psi(s) + \chi(s, u^a), \tag{5}$$

where

$$B(u^a) \stackrel{df}{=} \phi(0, u^a) \tag{6}$$

is the component of the scalar field living on the co-moving membrane, and

$$\chi \stackrel{df}{=} \phi(s, u^a) - B(u^a)\psi_0(s) \tag{7}$$

is the remaining part. The auxiliary, *fixed* function $\psi_0(s)$ depends on the variable s only. It is smooth, concentrated around $s = 0$, and

$$\psi_0(0) = 1. \tag{8}$$

It follows that

$$\chi(0, u^a) = 0. \tag{9}$$

The best choice for $\psi_0(s)$ is given by formula [8]

$$\psi_0(s) = \frac{1}{\cosh^2(s)}.$$

Next we derive Euler-Lagrange equations by taking independent variations of $B(u^a)$ and χ. The variation $\delta\chi$ has to respect condition (9), hence

$$\delta\chi(0, u^a) = 0.$$

Because of this condition, variation of the action functional

$$\mathcal{S} = \frac{2}{M}\int ds d^3u \sqrt{-g}h(s, u^a)\mathcal{L}$$

with respect to χ gives Euler-Lagrange equation in the regions $s < 0$ and $s > 0$. It has the following form

$$\frac{2}{M^2}\frac{1}{\sqrt{-g}}\partial_a[\sqrt{-g}hG^{ab}\partial_b(B\psi_0 + \chi)] \tag{10}$$
$$+\frac{1}{2}\partial_s[h\partial_s(B\psi_0 + \chi)] + h(B\psi_0 + \chi)[1 - (B\psi_0 + \chi)^2] = 0.$$

At $s = 0$ there is no Euler-Lagrange equation corresponding to the variation $\delta\chi$. Instead, we have the condition (9). Equation (10) should be solved in the both regions separately, with (9) regarded as a part of boundary conditions for χ. To complete the boundary conditions we also specify the behaviour of χ for $|\xi|$ much larger than the characteristic length l_0, that is for $|s| >> 1$: we shall seek a solution such that χ is exponentially close to +1 for $s >> 1$, while for $s << -1$ it is exponentially close to – 1.

At this stage of considerations Eq.(10) should not be extrapolated to $s = 0$. For example, the l.h.s. of it could have a $\delta(s)$-type singularity. It would occur if χ was smooth for $s > 0$ and for $s < 0$ but had a spike at $s = 0$.

In addition to Eq.(10) we also have the Euler-Lagrange equation corresponding to variations of $B(u^a)$. This equation has the following form

$$\frac{2}{M^2}\int ds\,\frac{1}{\sqrt{-g}}\partial_a\left[\sqrt{-g}hG^{ab}\partial_b(B\psi_0+\chi)\right]\psi_0 \qquad (11)$$
$$-\frac{1}{2}\int ds\,h\partial_s\psi_0\partial_s(B\psi_0+\chi)$$
$$+\int ds\,h\psi_0(B\psi_0+\chi)\left[1-(B\psi_0+\chi)^2\right]=0.$$

Here and in the following we use $\int ds$ as a shorthand for the definite integral $\int_{-\infty}^{+\infty} ds$.

Using equations (10) *and* (11) one can prove that χ does not have the spike at $s = 0$. It follows that Eq. (10) is obeyed by χ also at $s = 0$, and that now Eq.(11) can be obtained by multiplying Eq.(10) by $\psi_0(s)$ and integrating over s. Therefore we may concentrate on solving Eq. (10).

Step 3. Expansion in the width.

First we solve Eq.(10) in the leading approximation obtained by putting $1/M = 0$. The equation is then reduced to

$$\frac{1}{2}\partial_s^2\phi^{(0)}+\phi^{(0)}[1-(\phi^{(0)})^2]=0, \qquad (12)$$

where

$$\phi^{(0)}=B^{(0)}\psi_0+\chi^{(0)}.$$

Equation (12) does not contain derivatives with respect to time, in spite of the fact that it is supposed to approximate the evolution equation (10). This annoying fact is due to the singular character of the perturbation given in Eq.(10) by the terms proportional to positive powers of $1/M$. We shall see that time evolution is obtained indirectly, from consistency conditions.

Equation (12) has the following particular, well-known solution

$$\phi^{(0)} = \tanh s. \qquad (13)$$

This solution together with conditions (8), (9) gives

$$B^{(0)}=0, \quad \chi^{(0)}=\tanh s. \qquad (14)$$

Notice that $\phi^{(0)}$ has the same form as the planar domain wall (2) — in this way we realise the idea that in the co-moving reference frame the domain wall does not differ much from the planar domain wall.

The solution (13) in the co-moving coordinates does not determine the field ϕ in the laboratory frame because we do not know yet the position of the co-moving membrane with respect to the laboratory frame. Equations (10), (11) yield an equation for the co-moving membrane, otherwise they would not form the complete set of evolution equations for the field Φ. In fact, we shall see that the first order terms in Eq.(10) imply Nambu-Goto equation for the membrane.

The expansion in the width has the form

$$\chi(s,u^a)=\tanh s+\frac{1}{M}\chi^{(1)}(s,u^a)+\frac{1}{M^2}\chi^{(2)}(s,u^a)+\frac{1}{M^3}\chi^{(3)}(s,u^a)+..., \qquad (15)$$

$$B(u^a)=\frac{1}{M}B^{(1)}(u^a)+\frac{1}{M^2}B^{(2)}(u^a)+\frac{1}{M^3}B^{(3)}(u^a)+..., \qquad (16)$$

where we have taken into account the zeroth order results (14). The expansion parameter is $1/M$ and not $1/M^2$ because $1/M$ in the first power appears in h and G^{ab} functions after passing to the s variable. In order to obey the condition (9), and to ensure the proper asymptotics of χ at large $|s|$ we assume that for $n \geq 1$

$$\chi^{(n)}(0, u^a) = 0, \quad \lim_{s \to \pm\infty} \chi^{(n)} = 0. \tag{17}$$

Inserting the perturbative Ansatz (15, 16) in Eqs.(10, 11), expanding the l.h.s.'s of them in powers of $1/M$, and equating to zero coefficients in front of the powers of 1/M we obtain a sequence of linear, inhomogeneous equations for $\chi^{(n)}(s, u^a), B^{(n)}(u^a)$ with $n \geq 1$.

In particular, Eq.(10) expanded in the powers of $1/M$ gives equations of the type

$$\hat{L}\chi^{(n)} = f^{(n)}, \tag{18}$$

where the source term $f^{(n)}$ is determined by the lower order terms in χ and B, and

$$\hat{L} \stackrel{df}{=} \frac{1}{2}\partial_s^2 + 1 - 3(\chi^{(0)})^2.$$

Explicit solution of Eq.(18) is given by the formula

$$\chi^{(n)}(s) = 2\psi_1(s) \int_{-\infty}^{s} dx\, \psi_0(x) f^{(n)}(x) - 2\psi_0(s) \int_0^s dx\, \psi_1(x) f^{(n)}(x). \tag{19}$$

This solution obeys the boundary conditions (17).

Obviously, we assume that all proportional to positive powers of $1/M$ terms in Eqs.(10, 11) are small. For this it is not sufficient that the extrinsic curvatures are small, that is that $l_0 K_b^a \ll 1$. We have also to assume that the derivatives $\chi_{,a}^{(n)}, B_{,a}^{(n)}$ are not proportional to M. It is not the case, for example, if χ and B contain modes oscillating with a frequency $\sim M$. They would give positive powers of M upon differentiation with respect to u^a. If such oscillating components were present the counting of powers of $1/M$ would no longer be so straightforward as we have assumed. This assumption excludes radiation modes as well as massive excitations of the domain wall. Therefore, the approximate solution we obtain gives what we may call *the basic curved domain wall.* To obtain more general domain wall solutions one would have to generalize appropriately the approximation scheme. Actually, the fact that such particular radiationless, unexcited curved domain wall exists is a prediction coming from the $1/M$ expansion. The expansion yields domain walls of concrete transverse profile — the dependence on s is explicit in the approximate solution we construct even at the initial instant of time. We may choose the initial position and velocity of points of the membrane but the dependence of the scalar field on the variable s at the initial time is given by formulas (19). This unique profile is characteristic for the basic curved domain wall.

Step 4. The consistency conditions.

The first order terms in Eq.(10) give the following equation

$$\hat{L}\chi^{(1)} = K_a^a \partial_s \chi^{(0)}, \tag{20}$$

where $\chi^{(0)}$ is given by the second of formulas (14).

The most important point in our approach is the observation that operator $\hat{L}$ has a zero-mode, that is the normalizable solution

$$\psi_0(s) = \frac{1}{\cosh^2 s}$$

of the homogeneous equation

$$\hat{L}\psi_0 = 0. \tag{21}$$

Notice that $\psi_0 = \partial_s \chi^{(0)}$ — this means that the zero-mode ψ_0 is related to the translational invariance of Eq.(12) under $s \to s$ + const. The presence of the zero-mode implies the consistency conditions. For example, let us multiply Eq.(20) by ψ_0 and integrate over s. It is easy to see that $\int \psi_0 \hat{L}\chi^{(1)}$ vanishes because of (21), and we obtain the following condition

$$K_a^a \int ds\ \psi_0(s)\partial_s \chi^{(0)}(s) = 0$$

which is equivalent to

$$K_a^a = 0. \tag{22}$$

Eq.(22) coincides with the well-known Nambu-Goto equation. It determines the motion of the co-moving membrane, that is the functions $Y^i(u^a)$, i=l,2,3, once initial data are fixed. When we know these functions we can calculate the extrinsic curvature coefficients K_{ab} and the metric g_{ab}. Review of properties of relativistic Nambu-Goto membranes can be found in, e.g., [9].

Due to Nambu-Goto equation (22) the r.h.s. of Eq.(20) vanishes and the resulting homogeneous equation

$$\hat{L}\chi^{(1)} = 0$$

with the boundary conditions (17) has only the trivial solution

$$\chi^{(1)} = 0.$$

Notice that vanishing $\chi^{(1)}$ does not mean that the first order correction to the total field ϕ also vanishes — there is the first order contribution equal to $B^{(1)}\psi_0/M$. It does not vanish on the co-moving membrane that is at $s = 0$.

Analogous reasoning can give nontrivial consistency conditions also for equations (18) with $n > 1$. For some n, e.g. $n = 2$ we obtain only the trivial identity $0 = 0$.

Equation (11) expanded in powers of $1/M$ gives equations for $B^{(n)}$ coinciding with the consistency conditions. This follows from the fact that both Eq.(11) and the consistency conditions are obtained by multiplying Eq.(10) by the zero-mode ψ_0 and next integrating over s. Euler-Lagrange equation (11) can be regarded as generating equation for the consistency conditions.

Equations (10), (11) in the first order do not give any restriction on the function $B^{(1)}$. Equation for $B^{(1)}$ follows from the third order terms in Eq.(11):

$$\frac{1}{\sqrt{-g}}\partial_a(\sqrt{-g}g^{ab}\partial_b B^{(1)}) + (\frac{\pi^2}{4} - 1)K_b^a K_a^b B^{(1)} + \frac{9}{35}(B^{(1)})^3 = (\frac{\pi^2}{6} - 1)K_b^a K_c^b K_a^c, \tag{23}$$

This situation is typical for singular perturbation theories of which the $1/M$ expansion is an example — higher order equations imply restrictions (the consistency conditions) for the lower order contributions [10].

Step 5. Initial data.

At this point we have the complete set of equations determining the evolution of the domain wall in the $1/M$ expansion. Each of Eqs.(10), (11), (22) describes different aspect of the dynamics of the curved domain wall. Expanded in the positive powers of $1/M$ Eq.(10) determines dependence of χ on s. Because the term $B\psi_0$ in formula (5) has explicit dependence on s, we may say that Eq. (10) for χ fixes the transverse profile of the domain wall.

Equation (11) determines the $B^{(n)}(u^a)$ functions, which can be regarded as (2+1)-dimensional scalar fields defined on Σ and having nontrivial nonlinear dynamics. The

extrinsic curvature K_{ab} of Σ acts as an external source for these fields. The fields $B^{(n)}$ can propagate along Σ. One may regard this effect as *causal* propagation of deformations which are introduced by the extrinsic curvature.

Finally, Nambu-Goto equation (22) for the co-moving membrane determines the evolution of the shape of the domain wall.

Equation (22) for the co-moving membrane and equations for $B^{(n)}$ obtained from the consistency conditions are of the evolution type — we have to specify initial data for them, otherwise their solutions are not unique. Equations (18) for the perturbative contributions $\chi^{(n)}$ are of different type — in order to ensure uniqueness of their solution it is sufficient to adopt the boundary conditions (17). The initial data for $B(u^a)$ and $Y^i(u^a)$ follow from initial data for the original field ϕ. From such data for ϕ we know the initial position and velocity of the core. We assume that at the initial instant τ_0 the co-moving membrane and the core have the same position and velocity. Hence,

initial data for the membrane = initial data for the core.

Using formula (5) one can show that then

$$B^{(n)}(\tau_0, \sigma^1, \sigma^2) = 0, \quad \partial_\tau B^{(n)}(\tau_0, \sigma^1, \sigma^2) = 0. \tag{24}$$

In order to find the domain wall solution one should first solve the collective dynamics, that is to compute evolution of the co-moving membrane and of the $B^{(n)}$ fields. The profile χ of the domain wall is found in the next step from formulas (15, 19). In our perturbative scheme the profile of the domain wall can not be chosen arbitrarily even at the initial time — it is fixed uniquely once the initial data for the membrane and for the B field are given. Evolution of the core can be determined afterwards, from the explicit expression for the scalar field ϕ[8].

* * *

In paper [8] the perturbative solution has been constructed up to the fourth order.

In paper [11] second order perturbative solutions for cylindrical and spherical domain walls have been compared with numerical solutions of the Euler-Lagrange equation. The results of the comparison are quite encouraging.

In paper [12] analogous approach has been applied to a vortex in the Abelian Higgs model.

The perturbative approach to dynamics of the domain walls we have sketched can be generalised to models involving several fields. Also the requirement of relativistic invariance can be dropped out.

Acknowledgment

It is a pleasure to thank the organizers for nice and stimulating atmosphere during the Workshop.

REFERENCES

1. H.B.Nielsen and P.Olesen, Nucl.Phys. B61, 45 (1973).
2. D.Förster, Nucl.Phys. B81, 84 (1974).
3. R.Gregory, Phys.Lett. B206, 199 (1988).
4. S.M.Barr and D.Hochberg, Phys.Rev. D39, 2308 (1989).
5. B.Carter and R.Gregory, Phys. Rev. D51, 5839 (1995).

6. H.Arodź and P.Węgrzyn, Phys.Lett. B291, 251 (1992).
7. H. Arodź, Nucl. Phys. B450, 174 (1995).
8. H. Arodź, Nucl. Phys. B, in press (hep-th/9703168).
9. B. Carter, in "Formation and Interactions of Topological Defects", p.303. A.-Ch. Davis and R. Brandenberger (Eds.). Plenum Press, New York and London, 1995.
10. N.G.van Kampen, Stochastic Processes in Physics and Chemistry. North-Holland Publ.Comp., Amsterdam, 1987. Chapt.8,§7.
11. J. Karkowski and Z. Świerczyński, Acta Phys. Pol. B 30, 234 (1996).
12. H. Arodź, Nucl. Phys. B450, 189 (1995).

This page intentionally left blank

PATH SPACE FORMULATION OF THE BFV THEOREM

K. Bering*

M.I.T. Center for Theoretical Physics
Bldg. 6-304A, 77 Massachusetts Avenue,
Cambridge, MA, 02139 USA

INTRODUCTION

In the mid-seventies, Batalin, Fradkin and Vilkovisky[2] developed a BRST formulation of an arbitrary Hamiltonian gauge theory with reducible first class constraints and possible open gauge algebra, nowadays known as BFV-BRST quantization[3]. The formulation made extensive use of Grassmann odd objects and symmetries, which was new at that time. Most notably are the nilpotent BRST charge Ω and the gauge fermion ψ. The BFV Theorem, which is the subject of this talk, states that the partition function $\mathcal{Z}$ is independent of the choice of ψ.

The path integral was treated in many respects at a formal level. First of all, the usual problems with giving precise meaning to the path integral, ordering problems, etc. Secondly, the considerations didn't take into account global obstructions. Let us here just mention a few: A) Assumption of a globally defined gauge choice ψ (no Gribov problems). B) Assumption of a globally defined symplectic potential ϑ. Whereas the former types of problems clearly are principle, the latter type of problems are of a more solvable, but technical nature.

What motivated the present work was a concern about the stability of time locality. It is a well-known fact that time plays a special rôle in Hamiltonian theories. One assumes that all virtual paths in the path integral are parametrized by the same global time parameter. It is therefore possible to speak of *equal*-time Poisson brackets. Moreover, we assume no interaction between two different times. In other words, the symplectic two-form $\omega(t) = \omega(z(t), t)$ is a *function* of the coordinates taken in the very same time. One says that the phase space Poisson structure is ultra-local in time. In particular, the symplectic structure does not depend on the past nor the future. Nevertheless, in the standard proof[2] for the BFV Theorem one performs an infinitesimal flow generated by a bosonic BRST-type vector field

$$X(t) = \mu X_{\Omega(t)} , \tag{1}$$

where $\mu \sim \int_0^T dt\ \delta_{\text{ext}}\psi(t)$ is a functional. The flow therefore carries information about the past and the future of the path. In a consistent geometric formulation such flows

* E-mail: bering@nbivms.nbi.dk .

New Developments in Quantum Field Theory
Edited by Damgaard and Jurkiewicz, Plenum Press, New York, 1998

violates the ultra-local ansatz for the two-form $\omega(t) = \omega(z(t), t)$. The BRST variation (1) is transporting well-defined ultra-local theories into doubtful non-local theories. Naively, one would have expected that the relevant BRST transformation would map ultra-local theories in ultra-local theories. Our analysis below shows that the non-locality arising in the canonical measure factor, the Pfaffian, remarkably can be recast into an ultra-local form, which turns out to be the gauge-fixing term in the action. This solves our original posed question, and it shows that the formalism is consistent.

However, as a spin-off, we gain further geometric understanding of the gauge fixing flow in the path phase space. We also find an anti-Lie-bracket, that turns the set of gauge fermions into an infinite-dimensional algebra. This algebra is interesting in its own right from a pure mathematical point of view.

A PATH SPACE APPROACH

Our basic philosophy is to consider the collection $P\Gamma$ of all paths as an infinite-dimensional manifold, *i.e.* a geometric object that does not depend on the specific choice of coordinates. For instance, given a volume form, the partition function is an integral over this manifold weighted with the Boltzmann factor. Let us emphasize that both the Boltzmann factor and the volume form are scalar objects, that does not depend on the choice of coordinates. In this picture the BRST transformation (1) is a non-Hamiltonian (and non-local) vector field on the manifold $P\Gamma$, *i.e.* it is not a symplectic or equivalently a canonical transformation. So the symplectic structure is affected by the BRST vector field (1) and in general it becomes non-local.

Let us hence consider a wider class of theories than the ultra-local ones. To include non-local theories, we have to lift the construction from phase space Γ to a *path* phase space $P\Gamma$, thereby giving room for non-trivial Poisson bracket between different times. In general, non-local theories are senseless, but we shall see that we can give meaning to a restricted class of non-local theories.

As advocated above, acting with a BRST type transformation (1) change the symplectic two-form depending on the gauge-fixing Ψ. Let us introduce a gauge-fixed symplectic two-form directly in the path space $P\Gamma = \prod_{t\in[0,T]}\Gamma$:

$$\begin{aligned}\omega_\Psi = \omega_\Psi(\gamma) &= \tfrac{1}{2}\int_\gamma\int_\gamma dt\, dt'\, dz^A(t)\,\omega_{\Psi AB}(t,t')\,\wedge\, dz^B(t') \\ &= \tfrac{1}{2}\int_\gamma\int_\gamma dt\, dt'\,\omega_{\Psi AB}(t,t')\, dz^B(t')\,\wedge\, dz^A(t)(-1)^{\epsilon_A+1}\ . \end{aligned} \tag{2}$$

In other words, the symplectic two-form ω_Ψ is modified as an accumulated result of acting with the BRST type of transformations (1). We shall derive the explicit formula later. When the gauge fermion is turned off, $\Psi = 0$, the gauge-fixed symplectic structure coincide with the original ultra-local symplectic structure. In detail, the ultra-local form is

$$\omega = \omega(\gamma) = \tfrac{1}{2}\int_\gamma\int_\gamma dt\, dt'\, dz^A(t)\,\omega_{AB}(z(t),t)\,\delta(t-t')\,\wedge\, dz^B(t')\ . \tag{3}$$

The inverse $\omega_\Psi{}^{AB}(t, t')$ gives rise to a gauge-fixed path space Poisson bracket

$$\{F,G\}_\Psi = \{F,G\}_\Psi(\gamma) = \int_\gamma\int_\gamma dt\, dt'\, F\frac{\overleftarrow{\delta^r}}{\delta z^A(t)}\omega_\Psi{}^{AB}(t,t')\frac{\overrightarrow{\delta^l}}{\delta z^B(t')}G\ , \tag{4}$$

where $F, G : P\Gamma \to \mathcal{C}$. For generic Ψ the elements $\omega_{\Psi AB}(t,t')$ are functionals of the path γ. More precisely, we allow for the following dependence: A) It could be a *functional*

of the full path γ, B) it could be a function of the two path values $\gamma(t)$ and $\gamma(t')$ and finally C) it could depend explicitly on the times t and t'. In symbols,

$$\omega_{\psi AB}(t,t') \;=\; \omega_{\psi AB}(\gamma,\gamma(t),\gamma(t'),t,t') \;. \tag{5}$$

In the same manner, instead of looking at classical observables $f:\Gamma\to\mathbb{C}$, let us consider classical observables $F:P\Gamma\to\mathbb{C}$ that are functions on the *path* space $P\Gamma$. The usual ultra-local classical observables F are of the form

$$F(\gamma) \;=\; \int_\gamma dt\; f(\gamma(t),t) \qquad \text{or} \qquad F(\gamma) \;=\; f(\gamma(t_1),t_1) \tag{6}$$

for some function $f:\Gamma\times[0,T]\to\mathbb{C}$ and $t_1\in[0,T]$. Examples are the BRST-improved Hamiltonian $H_{BRST}(z(t),t)$ and the BRST charge $\Omega(z(t),t)$. The path space versions read

$$\begin{aligned} H_{BRST} = H_{BRST}(\gamma) \;&=\; \int_\gamma dt\; H_{BRST}(\gamma(t),t)\;, \\ \Omega = \Omega(\gamma) \;&=\; \int_\gamma dt\; \Omega(\gamma(t),t)\;, \\ \{\Omega,\Omega\} = 0\;, \qquad & \{\Omega,H_{BRST}\} \;=\; 0\;. \end{aligned} \tag{7}$$

The classical observables of the form (6) are formally the fundamental building blocks for all classical observables, *i.e.* a classical observable F in path space is a formal powerserie of the type (6). As a particular example we allow that the gauge fermion $\psi:\mathcal{P}\Gamma\to\mathbb{C}$ is a powerserie of these building blocks.

AN ANTI-LIE-BRACKET

Before describing explicitly how gauge fermions ψ act on the path space manifold, let us analyze the set of gauge fermions in more detail. First of all, it turns out that we can define a Grassmann-odd bracket structure among the gauge fermions. We have chosen here to start with this construction to emphasize its fundamental nature. Later we shall then see how this anti-algebra can be realized in the algebra of vector fields as BRST-like vector fields. When exponentiating to the corresponding groups we realize the corresponding BRST flow or – what turn out to be the same – the gauge-fixing flow.

The Poisson bracket and the BRST charge give rise to an anti-Lie-bracket in the space $C^\infty(\mathcal{P}\Gamma)$ of "classical" observables in the following manner:

$$\begin{aligned} (F,G) &= \tfrac{i}{\hbar}(-1)^{\epsilon_F}\{\Omega,FG\} \;=\; \tfrac{i}{\hbar}\{FG,\Omega\}(-1)^{\epsilon_G+1} \\ &= -(-1)^{(\epsilon_F+1)(\epsilon_G+1)}(G,F)\;. \end{aligned} \tag{8}$$

The anti-Lie-bracket is of odd Grassmann parity and it has ghost number +1. It satisfies the correct symmetry property, the Jacobi identity, but not the Poisson property.†

†A note on supersymmetry. Although we shall not need it in the following let us point out a relationship between a super Poisson bracket and an anti-bracket. Consider a manifold equipped with a "super-symmetry", by which we shall simply mean a Grassmann-odd vector field X_{SUSY}. A natural correspondence between super Poisson bracket and anti-bracket is

$$X_{\text{SUSY}}\,[\{F,G\}] \;=\; (X_{\text{SUSY}}[F],X_{\text{SUSY}}[G])\;, \tag{9}$$

or

$$X_{\text{SUSY}}\,[(F,G)] \;=\; \{X_{\text{SUSY}}[F],X_{\text{SUSY}}[G]\}\;. \tag{10}$$

Group of Gauge Fermions

The space G of gauge fermions $\psi : \mathcal{P}\Gamma \to \mathbb{C}$ with ghost number –1 equipped with this anti-Lie-bracket is an infinite dimensional Lie-algebra. The corresponding group G is identified with precisely the same space of ghost number –1 functions $\Psi : \mathcal{P}\Gamma \to \mathbb{C}$, so $G = G$. The group G is endowed with an associative product $\diamond : G \times G \to G$,

$$\Psi_1 \diamond \Psi_2 \;=\; \Psi_1 + \Psi_2 + \tfrac{i}{\hbar}\Psi_1\{\Omega, \Psi_2\} \;, \tag{11}$$

and an inversion map $\tau : G \to G$,

$$\tau(\Psi) \;=\; -\frac{\Psi}{1 + \frac{i}{\hbar}\{\Omega, \Psi\}} \;=\; -\frac{\Psi}{1+B} \;=\; -\Psi\, e^{-b} \;. \tag{12}$$

Here we have for convenience introduced

$$b \equiv \tfrac{i}{\hbar}\{\Omega, \psi\} = \ln(1+B) \;, \qquad B \equiv \tfrac{i}{\hbar}\{\Omega, \Psi\} = e^b - 1 \;. \tag{13}$$

The algebra elements are in general denoted by a lowercase ψ and the group element are denoted by uppercase Ψ. They are connected via the *bijective* exponential map $\mathrm{Exp} : G \to G$:

$$\Psi = \mathrm{Exp}(\psi) = \psi\, e(b) \;, \qquad \psi = \mathrm{Ln}(\Psi) = \Psi\, l(B) \;, \tag{14}$$

where e (and l) are

$$e(b) \;=\; \int_0^1 d\alpha\, e^{\alpha b} \;=\; \frac{e^b - 1}{b} \;=\; \frac{B}{b} \;=\; \sum_{n=1}^{\infty} \frac{b^{n-1}}{n!} \;, \tag{15}$$

$$l(B) \;=\; \frac{\ln(1+B)}{B} \;=\; \frac{b}{B} \;=\; \sum_{n=1}^{\infty} \frac{(-B)^{n-1}}{n} \;. \tag{16}$$

The neutral element of the Lie group is $\Psi = 0$:

$$\Psi \diamond 0 = \Psi = 0 \diamond \Psi \;, \qquad \mathrm{Exp}(0) = 0 \;. \tag{17}$$

The exponential map satisfies the Baker-Campbell-Hausdorff formula:

$$\begin{aligned} \mathrm{BCH}(u\psi_1, u\psi_2) &= \mathrm{Ln}(\mathrm{Exp}(u\psi_1) \diamond \mathrm{Exp}(u\psi_2)) \\ &= u\psi_1 + u\psi_1 + \tfrac{1}{2}(u\psi_1, u\psi_2) + \mathcal{O}(u^3) \;, \\ \mathrm{Exp}(-\psi) &= \tau(\mathrm{Exp}(\psi)) \;. \end{aligned} \tag{18}$$

This is easiest to prove in an algebra/group representation – a so-called realization that we describe in the next Section.

Group of Reparametrizations

Consider the infinite dimensional Lie group Aut($P\Gamma$) of invertible "reparametrizations"[‡] $\sigma : \mathcal{P}\Gamma \to \mathcal{P}\Gamma$. The corresponding infinite dimensional Lie algebra Lie(Aut ($P\Gamma$))

[‡] Notation: We shall use the word *reparametrization* as synonymous with an *automorphism, i.e.* a diffeomorphism that starts and ends on the same manifold. Thus a reparametrization in our terminology is a purely geometric notion, that does not depend on the local choice of coordinates, despite what the name may perhaps indicate. If we want to change local coordinates, we instead speak of a *coordinate transformation.* One may say that a *reparametrization* is an active operation, while a *coordinate transformation* is a passive operation.

is identified with the set of bosonic vector fields $X : C^\infty(P\Gamma) \to C^\infty(P\Gamma)$ equipped with the usual Lie bracket

$$[X, Y][F] \;=\; X[Y[F]] - Y[X[F]] \;. \tag{19}$$

The exponential map $\mathrm{Exp} : \mathrm{Lie}(\mathrm{Aut}(P\Gamma)) \to \mathrm{Aut}(P\Gamma)$ gives rise to an "action" conventionally denoted by a dot

$$\cdot \;: \mathrm{Lie}(\mathrm{Aut}(P\Gamma)) \times P\Gamma \;\to\; P\Gamma \;. \tag{20}$$

The definition and main properties are

$$X.\gamma = (\mathrm{Exp}(X))\gamma \;, \qquad 0.\gamma = \gamma \;, \qquad X.(Y.\gamma) = (\mathrm{BCH}(X, Y)).\gamma \;\; . \tag{21}$$

THE GAUGE-FIXING FLOW

We can now construct a Lie algebra homomorphism $X : G \to \mathrm{Lie}(\mathrm{Aut}(P\Gamma))$ that takes a gauge fermion ψ to a bosonic BRST-type vector field X^Ψ by

$$X : \psi \mapsto X^\psi \equiv \tfrac{i}{\hbar}\psi X_\Omega \;, \qquad [X^{\psi_1}, X^{\psi_2}] = X^{(\psi_1,\psi_2)} \;. \tag{22}$$

The corresponding Lie group homomorphism $\sigma \equiv \mathrm{Exp} \circ X \circ \mathrm{Ln} : G \to \mathrm{Aut}(P\Gamma)$ reads

$$\begin{aligned} \sigma : \;\; \Psi \mapsto \sigma_\Psi \;&=\; \mathrm{id}_{\mathcal{P}\Gamma} + \tfrac{i}{\hbar}\Psi X_\Omega \;, \\ \sigma_{\Psi_1 \diamond \Psi_2} \;&=\; \sigma_{\Psi_1} \circ \sigma_{\Psi_2} \;, \\ \sigma_{\tau(\Psi)} \;&=\; (\sigma_\Psi)^{-1} \;. \end{aligned} \tag{23}$$

As a very important consequence we have a gauge-fixing left group action

$$\diamond \;:\; G \times \mathcal{P}\Gamma \to \mathcal{P}\Gamma \tag{24}$$

directly on the path space:

$$\begin{aligned} \gamma_\Psi \;=\; \Psi \diamond \gamma = \sigma_\Psi(\gamma) \;&=\; (\mathrm{id}_{\mathcal{P}\Gamma} + \tfrac{i}{\hbar}\Psi X_\Omega)\,\gamma \;, \\ 0 \diamond \gamma &= \gamma \;, \\ \Psi_1 \diamond (\Psi_2 \diamond \gamma) &= (\Psi_1 \diamond \Psi_2) \diamond \gamma \;. \end{aligned} \tag{25}$$

This is the sought-for geometric gauge-fixing. The gauge-fixing can be understood at the level of paths as a modification of the individual paths!

Gauge-Fixed Quantities

The gauge-fixed classical observables F_ψ are defined via the pull-back[4] of the gauge-fixing map:

$$\begin{aligned} F_\psi(\gamma) = (\sigma_\psi^* F)(\gamma) \;&=\; F(\sigma_\psi(\gamma)) \\ &= F(\gamma) + \tfrac{i}{\hbar}\Psi(\gamma)\;\{\Omega(\gamma), F(\gamma)\} \\ &= e^{\frac{i}{\hbar}\psi(\gamma)\{\Omega(\gamma),\cdot\}} F(\gamma) \;. \end{aligned} \tag{26}$$

The BRST-improved Hamiltonian H_{BRST} and the BRST charge Ω are not changed by the gauge fixing flow, because of (7). The gauge-fixed symplectic two-form and the symplectic potential are remarkably simple, but clearly non-local:

$$\omega_\Psi \;=\; \sigma_\Psi^* \omega \;=\; \omega + \tfrac{i}{\hbar} d\Psi \;\wedge\; d\Omega \;, \tag{27}$$

$$\vartheta_\Psi \;=\; \sigma_\Psi^* \vartheta \;=\; \vartheta + \tfrac{i}{\hbar}\Psi\; d\Omega + \tfrac{i}{\hbar} d[\Psi\; i_{X_\Omega}\vartheta] \;. \tag{28}$$

THE BFV PATH INTEGRAL

We shall in this section review the BFV Theorem in the path space formulation, by inspecting the various parts of the path integral:

$$\mathcal{Z}_\psi \;=\; \int \Omega_{\rm vol}\; e^{\frac{i}{\hbar}(i_T\vartheta - H_{BRST} - \{\Omega,\psi\})} \;. \tag{29}$$

Here we have denoted the volume form by $\Omega_{\rm vol} = Dz\ {\rm Pf}\ (\omega..)$. We have already argued that the BRST improved Hamiltonian H_{BRST} is not modified by the gauge fixing flow. Below we shall analyze the kinetic part and the measure part.

The Kinetic Part

It is handy to introduce a vector field T that calculates the difference between the total time derivative (along the path) and the explicit time derivative:

$$T \;\equiv\; \frac{d}{dt} - \frac{\partial}{\partial t} \;=\; \int_\gamma dt\ \dot\gamma^A(t) \frac{\overrightarrow{\delta}}{\delta z^A(t)} \quad : \quad C^\infty(\mathcal{P}\Gamma) \to C^\infty(\mathcal{P}\Gamma) \;. \tag{30}$$

Then the kinetic term reads

$$i_T\vartheta(\gamma) = \int_\gamma dt\ \vartheta_A(t)\dot\gamma^A(t) \;, \qquad \vartheta(\gamma) = \int_\gamma dt\ \vartheta_A(t) dz^A(t) \;. \tag{31}$$

One should only sum over paths in the path integral so that the kinetic term for these paths is left invariant under the gauge-fixing flow. To achieve this conclusion, the path should satisfy the following BRST boundary condition

$$(i_{X_\Omega}\vartheta)(\gamma)|_{t=0} + \Omega(\gamma)|_{t=0} \;=\; (i_{X_\Omega}\vartheta)(\gamma)|_{t=T} + \Omega(\gamma)|_{t=T} \;. \tag{32}$$

One may check that this condition is stabile under the gauge-fixing flow.

The Measure Part

The gauge-fixed volume form can remarkably be rewritten by use of formula (27) as

$$\sigma_\psi^*\Omega_{\rm vol} \;=\; \Omega_{\rm vol}\ \frac{1}{1+B} \;=\; \Omega_{\rm vol}\ e^{-\frac{i}{\hbar}\{\Omega,\psi\}} \;. \tag{33}$$

We thus conclude one of the main points of this talk: The gauge-fixing term in the BFV action emerge from the change in the measure under the gauge-fixing flow. In fact, we have realized the gauge-fixing as an internal reparametrization of the path integral:

$$\mathcal{Z}_\psi \;=\; \int \sigma_\psi^* \left(\Omega_{\rm vol}\ e^{\frac{i}{\hbar}(i_T\vartheta - H_{BRST})}\right) \;. \tag{34}$$

We emphasize that the gauge-fixing term is not introduced by hand in the action, but it arises as a consequence of the gauge-fixing flow. Thus the path integral cannot depend on the gauge fermion and we can conclude the BFV Theorem. Here the path space integration is restricted by the BRST boundary condition (32).

Acknowledgments

The Author would like to thank I.A. Batalin, P.H. Damgaard, A.M. Semikhatov, A.P. Nersessian and O.M. Khudaverdian for fruitful discussions. The work is partially supported by Uppsala University and Nordita.

REFERENCES

1. C. Becchi, A. Rouet and R. Stora, Comm. Math. Phys. **42** (1975) 127; Ann. Phys.(N.Y.) **98** (1976) 287;
   I.V. Tyutin, *Gauge Invariance in Field Theory and in Statistical Physics in the Operator Formalism*, Lebedev preprint FIAN No. 39 (1975), unpublished.
2. E.S. Fradkin and G.A. Vilkovisky, Phys. Lett. **B55** (1975) 224.
   I.A. Batalin and G.A. Vilkovisky, Phys. Lett. **B69** (1977) 309.
   E.S. Fradkin and T.E. Fradkina, Phys. Lett. **B72** (1978) 343.
   I.A. Batalin and E.S. Fradkin, Phys. Lett. **B122** (1983) 157.
3. M. Henneaux and C. Teitelboim, *Quantization of Gauge Systems*,
   Princeton University Press, 1992; M. Henneaux, Phys. Rep. **126** (1985) 1.
4. M. Nakahara, *Geometry, Topology and Physics*, (IOP Publishing Ltd, Bristol) 1990.

This page intentionally left blank

SURPLUS ANOMALY AND RANDOM GEOMETRIES

Piotr Bialas[1], Zdzislaw Burda[2], Desmond Johnston[3]

[1]Institute of Comp. Science, Jagellonian University,
Nawojki 11, 30-072 Kraków, Poland
[2]Institute of Physics, Jagellonian University,
Reymonta 4, 30-059 Kraków, Poland
[3]Dept. of Mathematics, Heriot–Watt University,
Edinburgh EH14 4AS, Scotland

Four dimensional simplicial gravity has been studied by means of Monte Carlo simulations for some time[1] and an extensive numerical documentation of the properties of the model has been gathered. The main outcome of the studies is that the model undergoes a discontinuous phase transition[2] between the elongated and the crumpled phase when one changes the coupling to curvature. In the crumpled phase there are singular vertices in the system having orders growing extensively with the volume of the system[3]. The Hausdorff dimension of the simplicial manifolds is infinite in this phase. The elongated phase has a Hausdorff dimension equal to two. This phase has all properties of a branched-polymer phase such as, for example, the form of the puncture-puncture correlator[4].

We have postulated[5] that this behaviour is a manifestation of the constrained-mean-field scenario as realised in the Branched Polymer[6] (BP) or Balls-in-Boxes model[7]. Those models share all the features of 4D simplicial gravity except that they exhibit a continuous phase transition. Here we show that the transition is discontinuous when one considers a slightly modified ensemble[8], which in fact corresponds more closely to the one used in the simulations of simplicial gravity.

The partition function of the Branched Polymer model can be mapped onto the partition function of the Balls-in-Boxes model[7] :

$$Z(M,N) = \sum_{q_1,\ldots,q_M} p(q_1)\cdots p(q_M)\delta(q_1+\cdots+q_M-N) \qquad (1)$$

which describes weighted partitions of N balls in M boxes. The numbers of balls in two different boxes are independent of each other. The independence is weakly broken by the global constraint in the delta function which prevents factorization. The partition functions of the Branched Polymer and Balls-in-Boxes models are equivalent when one restricts the minimal number of balls in box to one and sets $N = 2M - 2$. The number

New Developments in Quantum Field Theory
Edited by Damgaard and Jurkiewicz, Plenum Press, New York, 1998

of balls q corresponds in terms of the branched polymers to the number of branches emerging from vertex. The two conditions correspond to the fact that each vertex has at least one branch and that the total number of branches of a tree–graph is fixed by the Euler relation.

The Balls-in-Boxes model can be solved[7]. One is interested in the limit of infinite number of boxes and fixed density of balls per box : $M \to \infty$ and $\rho = N/M = const.$ In this limit the partition function can be expressed in terms of the free energy density per box $f(\rho)$ which depends only on the density ρ :

$$Z(M, N) = e^{Mf(\rho)+\ldots} \tag{2}$$

where the dots are finite size corrections that disappear in the limit of infinite M. By introducing the integral representation of the Kronecker delta function one finds by the steepest descent method that

$$f(\rho) = \mu_*(\rho)\rho + K(\mu_*(\rho)) \tag{3}$$

where K is a generating function given by

$$K(\mu) = \log \sum_{q=1}^{\infty} p(q)e^{-\mu q} \tag{4}$$

and $\mu_*(\rho)$ is a solution of the saddle point equation :

$$\rho + K'(\mu_*) = 0 \tag{5}$$

The phase structure of the model depends on the choice of the weights $p(q)$ and through them on the properties of the generating function $K(\mu)$. In particular if the $p(q)$ vanish for all q above a certain value then $K(\mu)$ is analytic function for all μ and in this case the saddle point solution holds for all ρ. The most interesting case is in fact when the weights are such that the series defining the generating function $K(\mu)$ has a finite radius of convergence. Denote the value of μ which corresponds to the radius of convergence of the series (4) by μ_{cr} and the value of ρ at which the solution μ_* of the saddle point equation hits μ_{cr} by ρ_{cr}. In general μ_* (5) need not reach the critical value μ_{cr} for finite ρ. It does when the derivative K' is finite when $\mu \to \mu_{cr}^+$. Positivity of the $p(q)$'s implies that K is a monotonic function of μ and further that μ_* is a monotonic function of ρ. When ρ approaches ρ_{cr} from below μ_* approaches μ_{cr} from above. When ρ is larger than ρ_{cr}, μ sticks to the critical value μ_{cr} and the free energy is linear in ρ :

$$f(\rho) = \mu_{cr}\rho + \kappa_{cr} \tag{6}$$

where $\kappa_{cr} = K(\mu_{cr})$. The change of regimes $\rho < \rho_{cr}$ (3) to $\rho \geq \rho_{cr}$ (6) corresponds to the phase transition. To see what happens in the system at the transition it is convenient to consider the dressed one-box probability. It is defined as a probability that a particular box contains q balls. One can find $\pi(q)$ by fixing q in one box and summing over all possible partitions. This corresponds to summing over all partitions of the remaining $N - q$ balls distributed in the remaining $M - 1$ boxes :

$$\pi(q) = p(q)\frac{Z(M-1, N-q)}{Z(M, N)} \tag{7}$$

In the large M limit the saddle point equation gives :

$$\pi(q) = \begin{cases} e^{-K(\mu_*)}p(q)e^{-\mu_* q} & \text{for} \quad \rho < \rho_{cr} \\ e^{-\kappa_{cr}}p(q)e^{-\mu_{cr}q} & \text{for} \quad \rho \geq \rho_{cr} \end{cases} \tag{8}$$

The approach of the dressed probability to the limiting form (8) is not uniform. It turns out that the moments of the distribution (8) are biased. In particular the average :

$$\langle q \rangle = \sum_q q\pi(q) = \begin{cases} \rho & \text{for} \quad \rho \leq \rho_{cr} \\ \rho_{cr} & \text{for} \quad \rho > \rho_{cr} \end{cases} \tag{9}$$

does not give ρ for $\rho > \rho_{cr}$ as it should. One can easily correct for this by adding an anomalous term to π for finite M in the phase above ρ_{cr} :

$$\pi_M(q) = \pi(q) + \frac{1}{M}\delta(q - M(\rho - \rho_{cr})) \tag{10}$$

The anomalous term represented here by the delta function introduces an additional peak with the height $1/M$ at the position which moves proportionally to M. The peak vanishes when one sends M to infinity with q fixed. On the other hand, if one calculates the average first for each finite M and then sends it to infinity one obtains :

$$\langle q \rangle_M = \sum q\pi_M(q) = \rho \tag{11}$$

We call this additional term the *surplus anomaly*. In effect, $M - 1$ boxes keep the critical form of the distribution and one box takes over the surplus of balls. One can easily check by performing directly finite size computations that this situation is indeed realized in the model. In fact a finite size analysis shows some additional effects. For finite M the peak is smeared and one has to go to sufficiently large M to see the peak depart from the rest of distribution. The effects are secondary and do not bias the main conclusion. In figure 1 we show the position of the peak as a function of ρ and in the lowest box in figure 3 as a function of M for fixed ρ. In both cases the dependence is linear as expected. For large M the height $1/M$ of the peak is proportional to the area under the peak and has an interpretation as the probability that we pick a box with $M(\rho - \rho_{cr})$ balls. The appearance of the surplus anomaly is a condensation similar to the Bose-Einstein condensation[9] and Kac-Berlin spherical model[10]. We call the phase $\rho < \rho_{cr}$ fluid and $\rho > \rho_{cr}$ condensed. The system may enter the condensed phase either by changing density or by modifying the weights p. For instance for the one parameter family of weights :

$$p(q) = q^{-\beta} \qquad \text{for } q \geq 1 \tag{12}$$

one has the following phase diagram which is shown in figure 2. The critical line $\rho = \rho_{cr}(\beta)$ where the condensation occurs is given by :

$$\rho = \frac{\zeta(\beta - 1)}{\zeta(\beta)} \tag{13}$$

The fluid phase is below the critical line and the condensed one above. One can change the phase either varying β for fixed ρ as in case of branched polymers where $\rho = 2$, or by keeping β and varying the density ρ. With the crossing of the critical line one can associate a singularity of the free energy density :

$$\partial_\rho f \sim \begin{cases} \Delta\rho^{1/(\beta-2)} & \text{for} \quad 3 > \beta > 2 \\ \Delta\rho^{\beta-2} & \text{for} \quad \beta > 3 \end{cases} \tag{14}$$

where $\Delta\rho = \rho_{cr} - \rho \to 0^+$, and similarly with respect to the change of the parameter β. The singularity determines the order of the transition. The transition in ρ (and β)

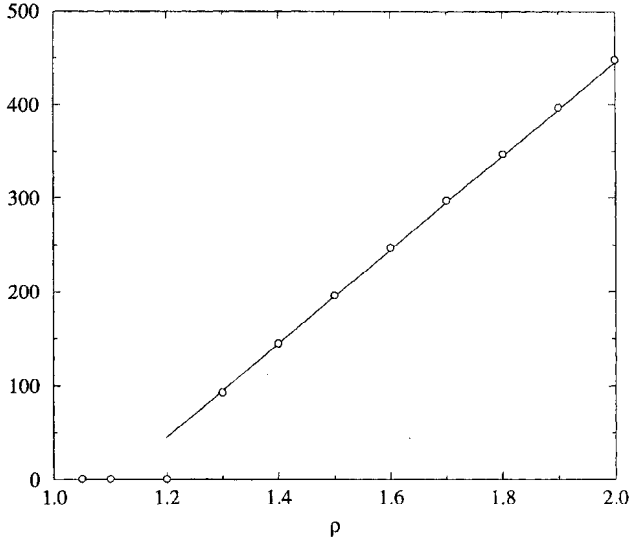


Figure 1. Position of peak as the function of ρ. The line is $M(\rho - \rho_{cr})$. ($p(q) = q^{-4}$, 500 boxes)

is continuous. On the critical line the system behaves very specifically as was first observed in the case of branched polymers[6]. For instance the susceptibility exponent γ can be tuned in the range 0, 1/2 and the Hausdorff dimension in the range 2, ∞. This is a new phase which is called *marginal*[12]. In the language of the balls-in-boxes model the critical line corresponds to a system lying between the fluid and condensed phase. The condensation is not yet fully developed so the number of balls in the singular box is not yet a fixed fraction of all balls but fluctuates in an extensive range. In figure 3 we compare the dressed probability distributions in the fluid, marginal and condensed phase. The long tail in the marginal case corresponds to the singular box with the singularity smeared in a wide range whose extension grows with the system size. This is the main difference with the condensed phase where the tail shrinks to a narrow peak with a well determined position. It is also interesting to consider ensembles with varying density. The simplest candidate is an ensemble with a given chemical potential coupled to the total number of particles :

$$Z(M, \mu) = \sum_N Z(M, N) e^{-\mu N} \tag{15}$$

This can be treated as an open ensemble in contact with a reservoir of balls. One immediately sees that it leads to a totally decoupled system and the free energy is exactly

$$f(\mu) = K(\mu) \tag{16}$$

This corresponds to M copies of the urn-model[11]. For $\mu \geq \mu_{cr}$ the average number of balls in urn diverges.

More natural in the context of simplicial gravity is the model where we keep the number of balls fixed and vary the number of boxes[8]. Indeed if we assume that box with q balls correspond to a vertex shared by the q vertices[5], then the analog of the

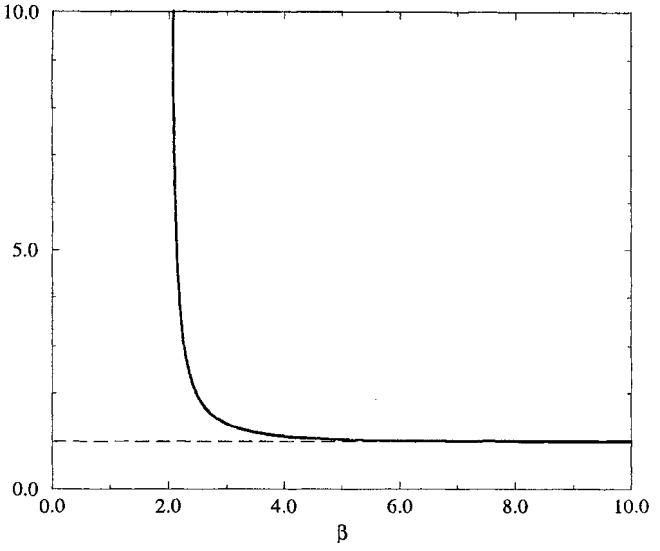


Figure 2. Phase diagram for Balls-in-Boxes model. ($p(q) = q^{-\beta}$)

simplicial gravity partition function would be:

$$Z(\kappa, N) = \sum_{M} Z(M, N) e^{\kappa M} \tag{17}$$

It is now more natural to consider curvature $r = 1/\rho$ instead of density. The partition function (17) can be rewritten as:

$$Z(\kappa, N) \approx N \int_{1/N}^{1} dr e^{N(f(r)+\kappa r)} \tag{18}$$

where $f(r)$ is the free energy per ball and is given by:

$$f(r) = \begin{cases} rK(\mu_*(r)) + \mu_*(r) & \text{for} \quad r > r_{cr} \\ r\kappa_{cr} + \mu_{cr} & \text{for} \quad r \leq r_{cr} \end{cases} \tag{19}$$

where $\mu(r)$ is a solution of the saddle point equation (5) expressed in terms of r rather then ρ. The saddle point equation for the integral (18) is:

$$\kappa + f'(r_*) = 0 \tag{20}$$

For $\kappa > \kappa_{cr}$ this equation reduces to $\kappa = K(\mu_{sp}(r_*))$ which has a unique solution for r_*. The value of r_* (the centre of the gaussian distribution) is the average curvature in the limit $N \to \infty$. This situation continues as long as $\kappa > \kappa_{cr}$.

For $\kappa < \kappa_{cr}$ the saddle point equation (20) has no solution and therefore the integrand is not gaussian anymore. It is a monotonic function of r. In particular, for $r < r_{cr}$ it is exponential $\exp N(\kappa - \kappa_{cr})r$ and for large N only this exponential part matters in the integral (18) and gives $\langle r \rangle \sim 1/N$.

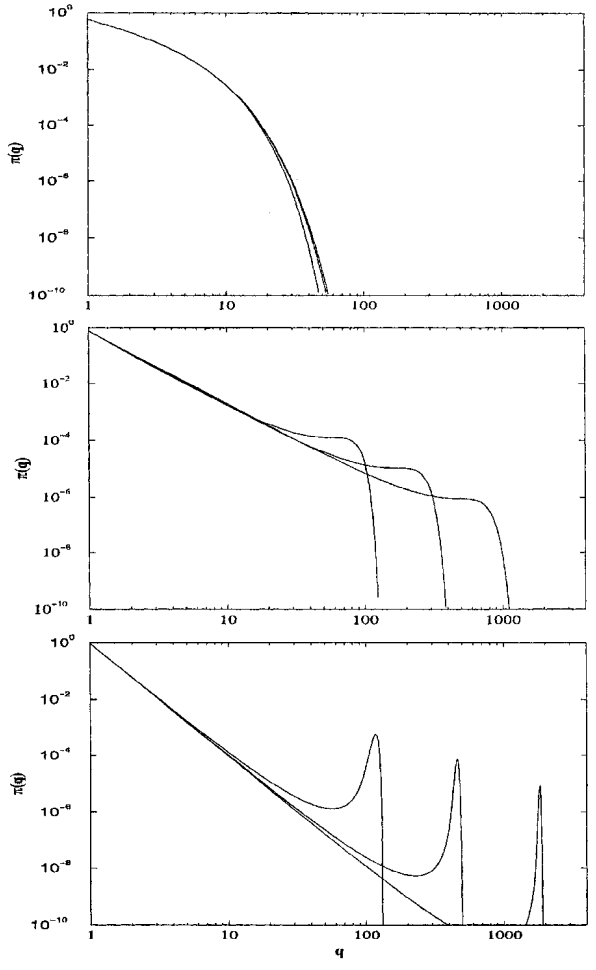


Figure 3. The $\pi(q)$ distribution in the fluid, marginal and condensed phases for systems with 128, 512 and 2048 boxes. ($\rho = 2$)

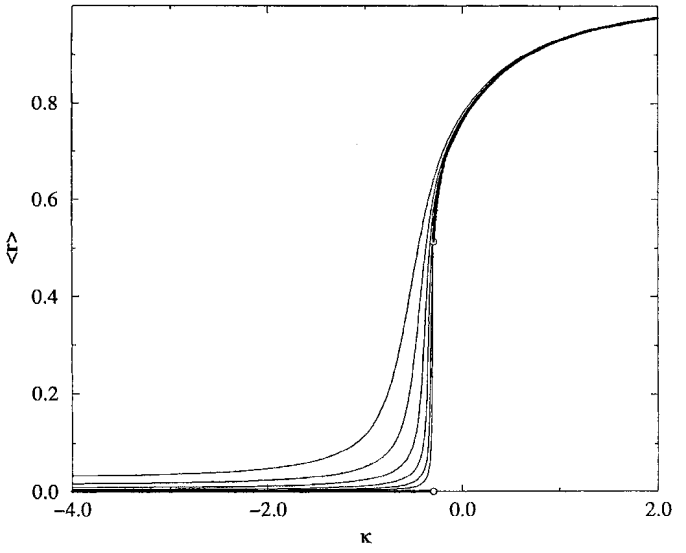


Figure 4. The average curvature as a function of κ. The heavy line corresponds to thermodynamical limit. (β = 5/2)

At the critical point $\kappa = \kappa_{cr}$, the equation (20) is fulfilled by all r between 0 and r_{cr} as can be seen from the second line of (19). Therefore r stays undetermined by the saddle point equation. In this case, to fix the shape of the integrand one has to consider systems for finite N.

The average curvature $\langle r \rangle$ is shown as a function of κ in figure 4. The bold line is a limiting curve for $N = \infty$. For $\kappa > \kappa_{cr}$ it is the solution of the saddle point equation (20). It stops at r_{cr} and falls to zero.

Apart from the limiting curve we show in the figure curvatures for a few finite values of N. The results are obtained by an improved version of the recursive technique[7]. In the gaussian phase $\kappa > \kappa_{cr}$ the curves lie close to each other indicating that the finite size effects are small there. For large κ the curves approach asymptotically an upper kinematic bound which in this case is $r = 1$ and corresponds to one ball in box. In the kinematic phase $\kappa < \kappa_{cr}$ finite size effects are stronger reflecting the size dependence of the lower kinematic bound $1/N$. This is similar to the results of numerical simulations of simplicial gravity but there the kinematical bound is $\frac{1}{\sqrt{N}}$ [13].

In the neighbourhood of the critical point the curves are steepest. This part of the curves corresponds to the pseudocritical region where the two phases coexist. One expects a double peak histograms of r : one peak near the maximum of the gaussian phase and the other near the kinematic limit $1/N$. In figure 5 we show the distributions of r for two different sizes N. One sees the coexistence of two phases. We tuned the value of κ in such a way as to have comparable heights of both the peaks. The spread of the range of κ where both the peaks are seen simultaneously in the distribution decreases with the size : $\delta\kappa \sim 1/N$ and reduces to one point for infinite N. The similar two peak structure has been observed in simplicial gravity for systems with 32000 and 64000 simplices[2].

To summarize shortly, the Balls-in-Boxes model describes well basic features of

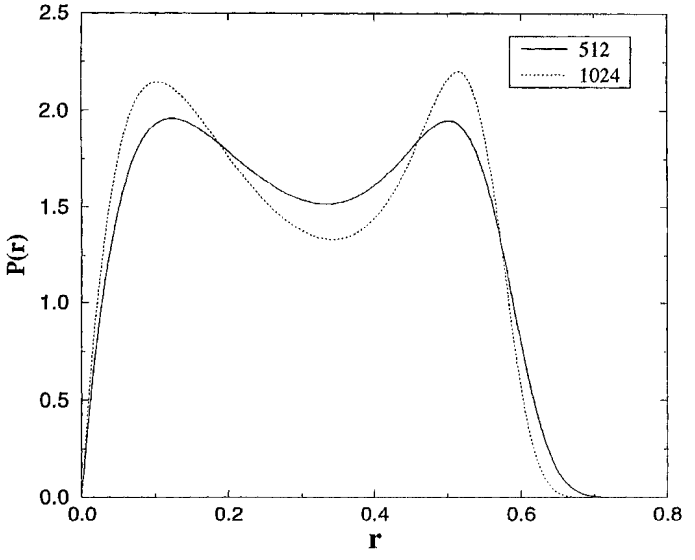


Figure 5. Distribution of r near the phase transition for systems with 512 ($\kappa = -0.32184$) and 1024 balls ($\kappa = -0.31910$) ($\beta = 5/2$)

simplicial gravity simulations such as the appearance of the singular vertices and the mother universe[5] in terms of the surplus anomaly. The phase transition associated with the appearance of the surplus anomaly is discontinuous in the ensemble with varying curvature, as is the one observed in the simulations of simplicial gravity. Recently an argument[14] was given that the bare weights $p(q)$ of vertex orders in simplicial gravity can be indeed approximated by the form $p(q) \sim e^{\kappa q} q^{-\beta}$. This completes the mean-field scenario. An open question, which is currently being studied, is whether one can somehow modify the integration measure or the action to obtain some nontrivial phase of higher dimensional gravity like the Liouville phase in two dimensions.

This work was partially supported by the KBN grant 2P03B04412.

REFERENCES

1. J. Ambjørn, D. Boulatov, A. Krzywicki and S. Varsted, Phys. Lett. **B276** (1992) 432;
   B. Brugmann and E. Marinari, Phys. Rev Lett. **70** (1993) 1908; J. Math Phys. **36** (1995) 6340.
   S. Catterall, J. Kogut and R. Renken, Phys. Lett. **B328** (1994) 277;
   J. Ambjørn and S. Varsted, Phys. Lett. **B266** (1991) 285;
   J. Ambjørn, Z. Burda, J. Jurkiewicz and C.F. Kristjansen, Acta Phys. Pol. **B23** (1992) 991.
   J. Ambjørn and J. Jurkiewicz, Phys. Lett. **B278** (1992) 42.
2. P. Bialas, Z. Burda, A. Krzywicki and B. Petersson, Nucl. Phys. **B472,** (1996) 293.
   B. V. de Bakker Phys. Lett. **B389** (1996) 238.
3. T. Hotta, T. Izubuchi and J. Nishimura, Nucl. Phys. **B** (Proc. Suppl.), **47** (1995) 609.
   S. Catterall, J. Kogut, R. Renken and G. Thorleifsson , Nucl.Phys. **B468** (1996) 263.
4. J. Ambjørn and J. Jurkiewicz, Nucl. Phys. **B451** (1995) 643.
5. P. Bialas, Z. Burda, B. Petersson and J. Tabaczek Nucl. Phys. **B495** (1997) 463.
6. P. Bialas and Z. Burda, Phys. Lett. **B384** (1996) 75.
7. P. Bialas, Z. Burda and D. Johnston Nucl. Phys. **B493** (1997) 505.

8. P. Bialas and Z. Burda *"Collapse of 4D random geometries"* hep-lat/9707028
9. P. Bialas, Z. Burda and D. Johnston, talk given at the Lattice 97 Conference, to appear in the Conference Proceedings.
10. T. Berlin and M. Kac, Phys. Rev. **86** (1952) 821.
11. G. Roepstorff and L. Schulman, J. Stat. Phys **34** (1984) 35,
B. Gaveau and L Schulman, J. Phys. A **20** (1987) 2865.
12. J. Jurkiewicz and A. Krzywicki Phys. Lett. **B392** (1997) 291.
13. S. Bilke, Z. Burda and B. Petersson, Phys. Lett. **B395** (1997) 4.
14. S. Catterall, J. Kogut and R. Renken, talk given at Los Alamos workshop "New directions in simplicial quantum gravity" Santa Fe 28 June – 8 September 1997.

This page intentionally left blank

TOPOLOGICAL CONTENTS OF 3D SEIBERG-WITTEN THEORY

Boguslaw Broda

Department of Theoretical Physics
University of Lódź
Pomorska 149/153
PL–90-236 Lódź
Poland

INTRODUCTION

In three dimensions (3D), we have the two interesting topological quantum field theories of "cohomological" type: $SU(2)$ non-abelian topological gauge theory of flat connection and 3D version of the (topological) Seiberg-Witten (SW) theory. The former is a 3D twisted $SU(2)$ pure gauge $N = 4$ SUSY theory or a 3D version of the Donaldson-Witten (DW) theory, and by "definition" it describes the Casson invariant, which appropriately counts the number of flat $SU(2)$ connections.[1] The latter is a 3D twisted version of abelian $N = 4$ SUSY theory with a matter hypermultiplet.[2] Also that theory should describe an interesting non-trivial topological invariant of 3D manifolds pertaining to SW invariant, which was conjectured to be equivalent to the Casson invariant (and thus to the former theory), and also to topological torsion.[3] The first conjecture is physically strongly motivated by the fact that the both theories can be derived from 4D $SU(2)$ pure gauge $N = 2$ SUSY theory corresponding via twist to DW theory.[4] That equivalence would be a 3D counterpart of the equivalence of 4D DW and SW theories. The latter being a "low-energy version" of the former. In this paper we present a qualitative physical scenario supporting the conjecture that topological contents of 3D SW theory is basically equivalent to an abelian version of the Casson invariant (of an auxiliary space). In turn, the abelian Casson invariant will be shown to be equivalent to the Alexander "polynomial" of the manifold.[5] In fact, as follows from mathematical literature the Alexander invariant is connected to the (non-abelian) Casson invariant as well as to topological torsion.[6]

PHYSICAL SCENARIO

Our physical inspiration is coming from an intuitive idea of Taubes and Witten[7] concerning "superconducting phenomena" in 4D SW theory. In mathematics, its classical version gives rise to the, so-called, "vanishing theorems".[2] Using physical arguments (Higgs mechanism, superconductivity, infrared regime, duality) and a geometric-

New Developments in Quantum Field Theory
Edited by Damgaard and Jurkiewicz, Plenum Press, New York, 1998

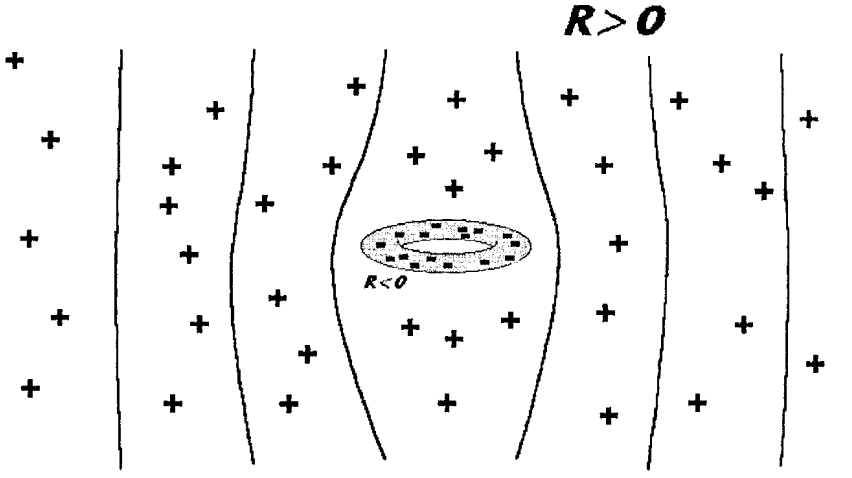


Figure 1. All fields become expelled from the (superconducting) ring of negative scalar curvature.

topological construction (scalar curvature distribution compatible with surgery), we propose a topological interpretation of 3D SW theory in terms of the abelian Casson invariant. Further algebraic reasoning shows equivalence of that invariant to the Alexander "polynomial". The scenario involves several simple steps. Our starting point is a 3D version (a dimensional compactification) of the original SW theory. Observing that the scalar curvature R plays the role of a mass-squared parameter for the monopole field (in the original 4D case as well as in our 3D one) we can use that observation to control the theory in low-energy limit. In the generic case of an arbitrary closed connected 3D manifold M^3, there are regions M^3_+ of positive scalar curvature, $R > 0$, and regions M^3_- of negative scalar curvature, $R < 0$ (we discard the regions of zero scalar curvature, $R = 0$). In fact, we will be able to assume that "almost" the whole manifold M^3 is of positive scalar curvature. Roughly speaking, we have a 3D manifold of positive scalar curvature with "bubbles" of negative curvature in the form of (in general, knotted) "superconducting" circuits/rings. In M^3_+, the matter is massive, and (super)electromagnetostatics is a long-range interaction, whereas in M^3_- we deal with the Higgs mechanism and the fields become short-range. In other words, due to the Meissner effect electromagnetic fields are expelled from the superconducting regions M^3_- (see, Fig. 1). In low-energy limit, we obtain pure (super)electromagnetostatic fields living only in M^3_+. In topological sector, due to abelianity of the theory, the fields could detect the first cohomology of M^3_+ (rather a trivial quantity). But happily, that is not the case as the theory is ill-defined in original variables, or at least not reliable because the coupling constant diverges. The trick of Seiberg and Witten to dualize the variables improves the situation. However, there is some subtlety in the case of non-simply-connected manifolds. We know from classical electrodynamics that, for example, the magnetostatic potential lives on the infinite cyclic covering space $\widetilde{M^3}$ of the original manifold M^3. Therefore, we begin to measure the first (co)homology of $\widetilde{M^3_+}$ (the Alexander invariant), rather than trivial one of M^3_+. Strictly speaking, we count the number of flat abelian connections (on $\widetilde{M^3}$), an abelian version of the Casson invariant, which we next "explicitly calculate" obtaining the Alexander "polynomial". What is left to show is a purely geometric-topological relation between the distribution

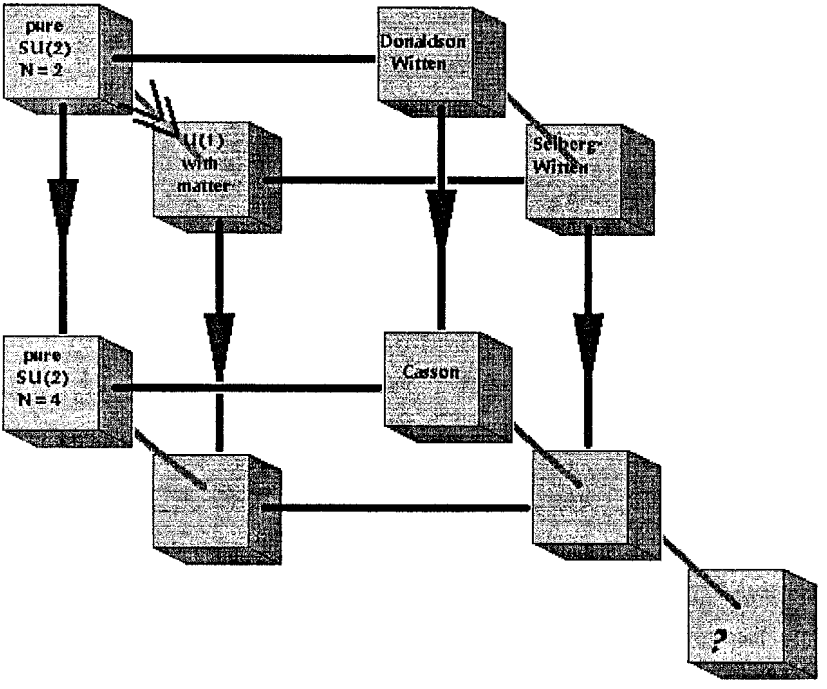


Figure 2. Web of mutually related quantum field theories in 3D and 4D.

of the scalar curvature R of M^3 and a topological description of M^3. It appears that it is possible to establish a necessary correspondence via surgery.

It is very instructive and suggestive to graphically present a web of interesting mutually related topological field theories in various dimensions (3 and 4 in our case) (see, Fig. 2, for a "three-dimensional" illustration): on the left-hand side we have "physical" theories, and on the right-hand side, the corresponding twisted (topological) ones; at the top, the dimension $D = 4$, whereas at the bottom, $D = 3$; in front, we have primary, microscopic theories, behind their low-energy limits (effective ones); solid black lines denote twists; solid black lines with arrows—dimensional reduction/compactification; a grey line with arrows—renormalization group procedure; other grey lines represent implicit or suspected connections between theories. The theory of interest in this paper could stand somewhere in the place of the box with the question mark. Since dimensional compactification/reduction and renormalization group procedures are not commutative, a further low-energy limit we perform should make sense (therefore, some objects at the bottom of Fig. 2 should be a bit fuzzy).

SUPERCONDUCTIVITY AND DUALITY

There are two equivalent possibilities to reach 3D version of SW theory. Either we can dimensionally compactify 4D SW theory or we can twist 3D (dimensionally compactified) abelian $N = 4$ SUSY theory coupled to monopole field. The "classical"

action of 4D SW theory, which we could assume as our starting point, looks as follows[2]

$$S = \int_X d^4x\sqrt{g}\left(\frac{1}{2}|F^+|^2 + g^{\mu\nu}D_\mu M^A D_\nu \overline{M}_A + \frac{1}{2}|M|^4 + \frac{1}{4}R|M|^2\right), \qquad (1)$$

where $F(= dA)$ is the strength of electromagnetic field, and M is a monopole field. The monopole field M is initially massless but in curved space the response of that field to gravitational interactions yields the scalar curvature R which effectively plays the role of the mass-squared parameter. Upon dimensional compactification on $X = S^1 \times Y$, where S^1 is a circle (of radius 1, for simplicity), we have the following decomposition of the $U(1)$ gauge field

$$A_\mu \longrightarrow (\phi, A_i).$$

Besides ordinary gauge transformations of A_i field, there is a residual gauge transformation of ϕ field,

$$\phi \to \phi + 1, \qquad (2)$$

which follows from the formula

$$A_0 \longrightarrow A_0 - i\partial_0 e^{ix_0}.$$

Thus, ϕ field assumes values in S^1 rather than in R^1.

As we have already mentioned, in generic case, the 3D manifold M^3 consists of the three kinds of regions: (1) M^3_+ with $R > 0$; (2) M^3_- with $R < 0$, and (3) M^3_0 with $R = 0$. The third kind is marginal (lower-dimensional), and we will ignore it. In M^3_+, the theory is in the Coulomb phase—effective theory contains only (free) pure (electro)magnetic long-range interactions—the matter fields are effectively massive and decouple from the theory in the infrared regime. In M^3_-, the theory is in the Higgs phase, and there are no long-range interactions at all. In other words (SUSY) (electro)magnetic interactions are repelled from M^3_-. In solid-state terminology, we have superconducting probes M^3_- put in M^3. What makes the analogy even closer and more suggestive is the fact that the dimension (= 3) is physical and the shape of the probes is "realistic": they are (possibly, knotted) circuits/rings, as will be explained later on. The difference lies in supersymmetric extension and in our limitation to topological sector. Thus, in the infrared limit we have only pure SUSY electromagnetic interactions which penetrate M^3_+. In twisted, topological sector it could naively correspond to measuring the first cohomology $H^1(M^3_+)$, not a very exciting invariant.

In 3D the coupling constant e is of positive dimension (dim $e = m^{1/2}$), and therefore it diverges in the infrared limit. But there is a well-known procedure to rescue the theory, namely we can pass to dual variables inverting the coupling constant e, making the theory more convergent. The dualization we propose is not the standard 3D one, but it is "four-dimensional".[8] It means that one of the components of the scalar part of the supermultiplet is dualized tensorially, i.e. to a vector, as it were an electrostatic potential of a 4D theory. It is sensible because we can treat our 3D theory as a time-independent 4D one since we think in terms of compactification on $X(= S^1 \times Y)$. Such a dualization does "almost nothing" to the variables, the scalar and vector potentials become simply exchanged

$$A_i \longrightarrow \chi,$$
$$\phi \longrightarrow V_i.$$

"Almost nothing" means that although the dual potentials have the same tensorial structure as the original ones, they are multivalued. On a simply-connected manifold nothing really topologically interesting happens. In general case however, since the dual

potentials are multivalued functions, they begin to live on an infinite cyclic covering space of a given 3D manifold. That way, in topological sector, the theory switched from measuring the first cohomology of M^3_+ to measuring the first cohomology of its infinite cyclic cover, $\widetilde{\mathcal{M}^3_+}$. In other words, upon twisting, we have an abelian "cohomological" theory on $\widetilde{\mathcal{M}^3_+}$. It should be stressed that the conclusion that our theory lives on a covering space as a result of the multivaluedness of fields can be true only in topological sector as only this sector uses "constant" configurations for which the conclusion is satisfied. Since M^3_+ has a boundary it would be rather difficult, if not impossible, to explicitly construct a local action for both the supersymmetric and even twisted version of the theory. But it is not necessary, since we are able to implicitly identify the theory (the fields are still known explicitly) in topological sector. Namely, it is a theory of abelian flat connection on $\widetilde{\mathcal{M}^3_+}$. The partition function Z of our theory should "ordinarily" (without sign, due to abelianity) count the number of (inequivalent) $U(1)$-representations of the fundamental group π_1 of $\widetilde{\mathcal{M}^3_+}$

$$Z = \#\mathrm{Rep}\left(\pi_1\left(\widetilde{\mathcal{M}^3_+}\right) \longrightarrow U(1)\right). \tag{3}$$

Eq. (3) defines an abelian analog of the Casson invariant. Our temporary conclusion says that the partition function of 3D SW theory describes the abelian Casson invariant Z of $\widetilde{\mathcal{M}^3_+}$.

SHAPE OF THE SUPERCONDUCTING PROBES

We have already suggested that the superconducting probes M^3_- are of the form of (in general, knotted) circuits/rings. More precisely, they are thicken knots (links). Mathematically, their relation to M^3 is the most natural, standard one—via the, so-called, surgery procedure— M^3 should be considered as a manifold obtained according to surgery instructions encoded in the knot/link, i.e. via cutting out (thicken) knots and pasting them back in a different way.

For our purposes we need a more or less obvious statement which, in a sense, is complementary to the Gromov-Lawson-Schoen-Yau (GLSY) theorem.[9] The GLSY theorem concerns, roughly speaking, higher-dimensional situation, and it says that carefully performing surgeries on a manifold of positive scalar curvature we do not spoil the positivity property. In our (lower-dimensional) case the GLSY theorem happily fails in generic case, and we expect something opposite, i.e. knots can introduce regions of negative scalar curvature. To see it (see, Fig. 3), let us consider a (straight) cylinder instead of a (curved) thicken knot for a moment. Cutting it vertically with a plane, i.e. discarding the third dimension, we obtain a circle on the plane. Adding the "fourth" dimension as a third one we can attach a "pipe" (a dimensionally reduced handle) to the plane along the circle. After smoothing out, the attaching area has obviously negative curvature. Cartesian product with the "third" (temporarily discarded) dimension does not change the sign of the scalar curvature. In reality, a knot is curved rather than straight but since we are interested in topology we can arrange the curvature of the attaching area large enough in comparison with the curvature of the knot itself.

CALCULATING THE PARTITION FUNCTION

Finally, we would like to know what is the relation between Z and some, possibly known topological invariant of M^3. To this end, we will "explicitly calculate" the

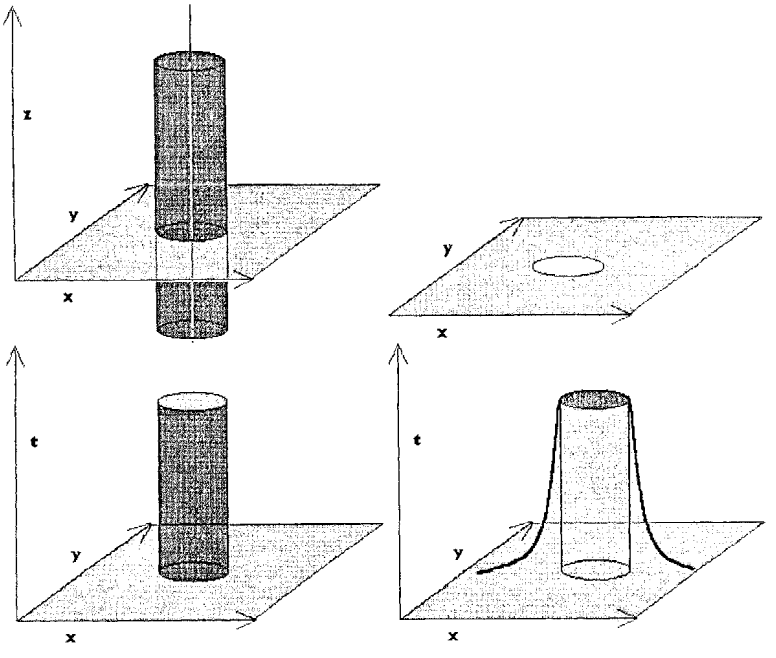


Figure 3. Scalar curvature compatible with surgery.

number of Rep $\left(\pi_1\left(\widetilde{\mathcal{M}}^3_+\right) \longrightarrow U(1)\right)$. First of all, since $U(1)$ is an abelian group a non-abelian part of π_1 drops out, and effectively we deal with an equivalent but simpler expression,

$$Z = \#\mathrm{Rep}\left(H_1\left(\widetilde{\mathcal{M}}^3_+\right) \longrightarrow U(1)\right), \tag{4}$$

where H_1 is an integer-valued first homology group. We claim that Z is the Alexander "polynomial" A understood as

$$A = \det A_{kl}, \tag{5}$$

where

$$\sum_l A_{kl}\alpha_l = 0, \qquad k, l = 1, \ldots, N, \tag{6}$$

for α_l—some homology basis. Eq. (6) represents the homology we are interested in, i.e. $H_1\left(\widetilde{\mathcal{M}}^3_+\right)$ (a mathematical fact). Now, we should pass to a $U(1)$ representation of (6). If

$$\alpha_l \longrightarrow e^{2\pi i\omega_l},$$

then for LHS of (6) we have

$$\sum_l A_{kl}\alpha_l \longrightarrow \prod_l \exp\left(2\pi i A_{kl}\omega_l\right) = \exp\left(2\pi i \sum_l A_{kl}\omega_l\right),$$

where $0 \leq \omega_1, \ldots, \omega_N < 1$, for uniqueness. RHS of Eq. (6) is now

$$0 \longrightarrow 1 = \exp(2\pi i m_k), \qquad m_k \in Z.$$

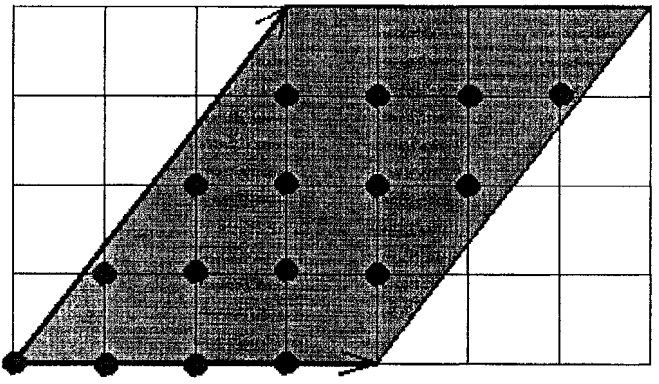


Figure 4. Partition function as the number of dots.

A new, $U(1)$ version of Eq. (6) is then

$$\sum_l A_{kl}\omega_l = m_k. \tag{7}$$

Z is the number of solutions of Eq. (7), i.e. the number of different m_k. m_k are integer-valued points in a parallelepiped spanned by the "base vectors" $A_{k(l)}$. One can observe that the number of m's is equal to the volume of that parallelepiped (see, Fig. 4), which in turn, is equal to the determinant A of A_{kl}.

CONCLUSIONS

We have found a physical scenario relating 3D SW theory to an abelian version of the Casson invariant Z of the infinite cyclic covering space $\widetilde{\mathcal{M}}^3_+$ of $\mathcal{M}^3_+$. In physical language, the space M^3_+ emerges as a complement of superconducting probes. Whereas in mathematical language, M^3_+ emerges as a complement of a thicken knot used to construct the manifold M^3 via surgery. In low-energy limit, the both languages match due to the Higgs mechanism generated by the scalar curvature R. Moreover, the partition function Z has been shown to be equivalent to the Alexander "polynomial" of M^3, which in principle, is related to topological torsion.[6]

ACKNOWLEDGMENTS

I am greatly indebted to the organizers of the Workshop, Prof. Poul H. Damgaard and Prof. Jerzy Jurkiewicz, for their kind invitation to Zakopane. The paper has been supported by the KBN grant 2P03B09410.

REFERENCES

1. M. Blau and G. Thompson, N=2 topological gauge theory, the Euler characteristic of moduli spaces, and the Casson invariant, *Commun. Math. Phys.* 152:41 (1993).
2. E. Witten, Monopoles and four-manifolds, *Math, Res. Lett.* 1:769 (1994).
3. G. Meng and C. Taubes, SW=Milnor torsion, *Math. Res. Lett.* 3:661 (1996). M.Hutchings and Y. Lee, Circle-valued Morse theory, Reidemeister torsion, and Seiberg-Witten invariants

of 3-manifolds, *preprint* dg-ga/9612004; Circle-valued Morse theory and Reidemaister torsion, *preprint* dg-ga/9706012.
4. E. Witten, Topological quantum field theory, *Commun. Math. Phys.* 117:353 (1988).
5. C. Frohman and A. Nicas, The Alexander polynomial via topological quantum field theory, *in:* "Differential Geometry, Global Analysis, and Topology, Canadian Math. Soc. Conf. Proc.", Vo1.12, Amer. Math. Sot., Providence, RI (1992) pp. 27-40.
6. V. Turaev, Reidemeister torsion in knot theory, *Russian Math. Surveys* 41:119 (1986).
7. E. Witten, talk at the "Trieste Conference on S-Duality and Mirror Symmetry", Trieste (1995) Italy.
8. N. Seiberg and E. Witten, Electric-magnetic duality, monopole condensation, and confinement in $N = 2$ supersymmetric Yang-Mills theory, *Nucl. Phys. B* 426:19 (1994).
9. M. Gromov and B. Lawson, The classification of simply connected manifolds of positive scalar curvature, *Ann. Math.* 111:423 (1980). R. Schoen and S.-T, Yau, On the structure of manifolds with positive scalar curvature, *Manuscripta Math.* 28:159 (1979).

FREE STRINGS IN NON-CRITICAL DIMENSIONS

Marcin Daszkiewicz, Zbigniew Hasiewicz, and Zbigniew Jaskólski

Insitute of Theoretical Physics
University of Wroclaw
pl. Maxa Borna 9
50-206 Wroclaw, Poland

INTRODUCTION

From the perspective of the today second string revolution the low energy hadron physics might seem to be much less exciting area for application of string theory then it was 25 years ago[1]. Still experimental and theoretical evidences for string behaviour of the low energy QCD are strong and compelling. To a good approximation hadrons can be grouped in linear Regge trajectories with an approximately universal slope and the Regge phenomenology provides surprisingly good explanation for many low energy scattering processes[2]. Also several theoretical attempts to go beyond the perturbative QCD: the lattice strong coupling expansion, the large N_c expansion, the flux tube picture, or the loop equations, seem to point at some effective QCD string model.

Whatever the final understanding of the QCD string would be it seems that the correct answer should include a consistent quantum model of 1-dim extended relativistic system in 4-dim Minkowski space. One can expect in particular that the spectrum of asymptotic states in such theory should be described by a non-critical free string model. This provides a physical motivation for analysing free strings in subcritical dimensions. A consistent free non-critical string is of course far from solution of the problem. For example, in the case of the Nambu-Goto string which can be constructed beyond the critical dimension using covariant formulation the real obstacle is related to the violation of unitarity at the one loop level of the perturbation expansion. Nevertheless, the classification of free string models can teach us what are the possible starting points for an interacting theory or at least what are the kinematical requirements such a theory should satisfy.

The aim of this seminar is to provide a brief review of consistent free string models in physical dimensions. The presentation is based on the recent results obtained in Ref.3,4. For the sake of simplicity we shall restrict ourselves to open bosonic strings, but the general picture of models and their relations is essentially the same for the closed bosonic strings and fermionic counterparts of the models considered[5].

All but one free string quantum models discussed below can be obtained by the covariant quantization of one of the two classical string models. One of them is the

New Developments in Quantum Field Theory
Edited by Damgaard and Jurkiewicz, Plenum Press, New York, 1998

well-known Nambu-Goto string which can be also reformulated in terms of the Brink, Di Vecchia, Howe, Polyakov (BDHP) quadratic action. The resulting quantum models allowed by the old no-ghost theorem[6,7] are briefly described in Section 2. The second classical string model which we call the massive string is described by the BDHP action supplemented by the Liouville action (with vanishing cosmological constants) for an extra dimensionless scalar world-sheet field[3]. This model contains two more functional degrees of freedom then the classical Nambu-Goto string. The family of quantum models obtained by the covariant quantization of the massive string is described in Section 3. It contains in particular the quantum massive string constructed long time ago by Chodos and Thorn[8].

As it was mentioned above the problem of the critical dimension for the Nambu-Goto string in the covariant formulation is related to the unitarity of loop amplitudes. In the unitary light-cone formulation the critical dimension shows up already in the free theory which is covariant only for $D = 26$[9]. The light-cone quantum model can be however modified by introducing longitudinal excitations in an appropriate way such that the Poincare covariance is restored. Such construction was recently proposed in Ref.4 and is briefly described in Section 4. In Section 5 some conclusions are presented. All free string models discussed in this note are schematically presented on Fig.1.

NAMBU-GOTO STRING

Introducing the world-sheet metric g_{ab} as an extra variable one can reformulate the classical Nambu-Goto string in terms of the BDHP quadratic action

$$S[x] = -\frac{\alpha}{2\pi}\int_M \sqrt{-g}\, d^2z\, g^{ab}\partial_a x^\mu \partial_b x^\nu \eta_{\mu\nu} \quad .$$

In the confromal gauge the system is described by the equations of motions

$$-\ddot{x}^\mu + x''^\mu = 0 \quad ,$$

the constraints

$$\frac{\alpha}{2\pi}\left(\dot{x}^2 + x'^2\right) = 0 \quad , \quad \frac{\alpha}{\pi}\dot{x}x' = 0 \quad ,$$

and the boundary conditions $x'^\mu(\tau,0) = x'^\mu(\tau,\pi) = 0$.

In the phase space holomorphic variables

$$\{\alpha_0^\mu, x_0^\nu\} = \frac{1}{\sqrt{\alpha}}\eta^{\mu\nu} \quad , \quad \{\alpha_m^\mu, \alpha_n^\nu\} = im\eta^{\mu\nu}\delta_{m,-n} \quad ,$$

the constraints can be written in the compact form

$$L_k = \frac{1}{2}\sum_{-\infty}^{+\infty} \alpha_{-n}\cdot\alpha_{k+n} = 0 \quad ,$$

and form the algebra of first class constraints

$$\{L_m, L_n\} = i(m-n)L_{n+m} \quad .$$

The covariant quantization consist in representing the commutation relations

$$[P_\mu, x_0^\nu] = -i\delta_\mu^\nu \quad , \quad [\alpha_m^\mu, \alpha_n^\nu] = m\eta^{\mu\nu}\delta_{m,-n}$$

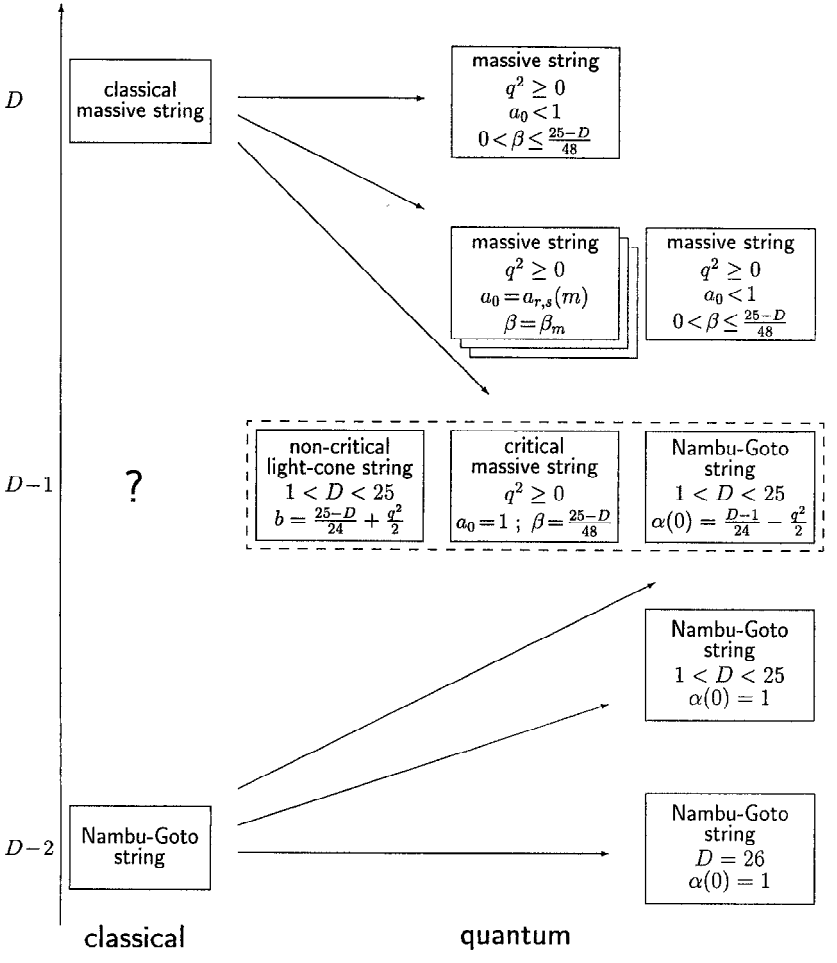


Figure 1. Classification of non-critical free string models. The number of "functional" degrees of freedom varies along the vertical axis. The arrows symbolize the covariant method of quantization. The quantum models in the dashed box are equivalent to each other. The covariant classical counterpart of the resulting model with the right number of degrees of freedom is not known.

in a pseudo-Hilbert space H[9]. Due to the normal ordering the quantum constraints are of the second class

$$[L_n, L_m] = (n-m)L_{n+m} + \frac{1}{12}D(n^3 - n)\delta_{n,-m}$$

and therefore only half of them can be imposed as conditions for the physical states

$$L_n|\psi\rangle = 0 \quad , \quad n \geq 0 \quad , \quad (L_0 - \alpha(0))|\psi\rangle = 0 \quad .$$

One gets a consistent quantum theory if and only if there are no physical states with negative norm. The necessary and sufficient conditions for the dimension D of the target space and for the intercept parameter $\alpha(0)$ are given by the no-ghost theorem proven long time ago by Goddard and Thorn[6], and by Brower[7].

The space $H_{ph} \subset H$ of physical states in the covariantly quantized Nambu-Goto string is ghost free if and only if $D = 26$, and $\alpha(0) = 1$, or $D < 26$, and $\alpha(0) \leq 1$.

The number of degrees of freedom of the quantum theory depends on the structure of the zero norm physical states. In the case of $D = 26$, $\alpha(0) = 1$ the space H_0 of null physical states is largest possible. In this case the number of degrees of freedom is the same as in the classical model. This is the only case when the covariant quantization yields the same result as the quantization in the light-cone gauge[9]. For $D < 26$, $\alpha(0) < 1$ there are no null states and the quantum model contains one "functional" degree of freedom more then the classical theory. For non-critcal string with the intercept $\alpha(0) = 1$ one has an infinite dimensional subspace of null states, but not large enough to cancel one "functional" degree of freedom. This situation is schematicaly presented on Fig.1 where the lower part of the diagram describes models obtained by the covariant quantization of the Nambu-Goto string.

MASSIVE STRING

The action for the massive string is a simple modification of the BDHP action by the Liouville action for additional dimensionless world-sheet scalar φ with vanishing cosmological constants:

$$\begin{aligned} S[g, \varphi, x] &= -\frac{\alpha}{2\pi}\int_M \sqrt{-g}\, d^2z\, g^{ab}\partial_a x^\mu \partial_b x^\nu \eta_{\mu\nu} \\ &\quad -\frac{\beta}{2\pi}\int_M \sqrt{-g}\, d^2z\, \left(g^{ab}\partial_a\varphi\partial_b\varphi + 2R_g\varphi\right) \quad . \end{aligned} \tag{1}$$

The functional above regarded as a two-dimensional conformal field theory action describes a special case of the induced (Liouville) gravity coupled to the conformal matter with the central charge $c = d$. This system has been extensively studied some time ago[10], both as a conformal field theory input of the Polyakov formulation of interacting string theory[11] (this application being restricted by the so called $c = 1$ barrier) and as a 2-dimensinal toy model of dilaton gravity for analysing black hole physics[12].

The action (1) can be also regarded as a world-sheet action for a relativistic one dimensional extended object. From this point of view it was first considered by Marnelius[13] in one of the first attempts to clarify the relation between non-critical string and the Liouville theory. The classical and quantum free string model determined by

(1) was recently analysed in Ref.3. Because of the properties of its quantum spectrum we shall call it the massive string.

In the conformal gauge the massive string is described by the equations of motion

$$-\ddot{x}^\mu + x''^\mu = 0 \quad , \quad -\ddot{\varphi} + \varphi'' = 0 \quad ,$$
$$\frac{\alpha}{2\pi}\left(\dot{x}^2 + x'^2\right) + \frac{\beta}{2\pi}\left(\dot{\varphi}^2 + \varphi'^2\right) - 2\frac{\beta}{\pi}\varphi'' = 0 \quad ,$$

the constraints

$$\frac{\alpha}{\pi}\dot{x}x' + \frac{\beta}{\pi}\dot{\varphi}\varphi' - 2\frac{\beta}{\pi}\dot{\varphi}' = 0 \quad , \quad \int_0^\pi d\sigma\dot{\varphi} = 0 \quad ,$$

and the boundary conditions $x'^\mu(\tau,0) = x'^\mu(\tau,\pi) = 0$, $\varphi'(\tau,0) = \varphi'(\tau,\pi) = 0$. Let us note that due to the gauge symmetry of the model there are no dynamical degrees of freedom in the metric sector.

In the phase space holomorphic variables

$$\{\alpha_0^\mu, x_0^\nu\} = \frac{1}{\sqrt{\alpha}}\eta^{\mu\nu} \quad , \quad \{\alpha_m^\mu, \alpha_n^\nu\} = im\eta^{\mu\nu}\delta_{m,-n} \quad ,$$
$$\{\omega_0, \varphi_0\} = \frac{1}{\pi} \quad , \quad \{\beta_m, \beta_n\} = im\delta_{m,-n} \quad ,$$

the constraints take the form

$$L_n = \tfrac{1}{2}\sum_{m=-\infty}^{+\infty} \alpha_{-m}\cdot\alpha_{n+m} + \tfrac{1}{2}\sum_{m=-\infty}^{+\infty} \beta_{-m}\beta_{n+m} + 2\sqrt{\beta}ik\beta_k + 2\beta\delta_{n,0} = 0 \quad , \quad \beta_0 = 0 \ ,$$

and satisfy the Poisson bracket algebra

$$\{L_m, L_n\} = i(m-n)L_{n+m} - 4i\beta(m^3 - m) \quad , \quad \{L_m, \beta_0\} = 0 \ .$$

In the classical Nambu-Goto string the first class constraints remove 2 out of the initial D "functional" degrees of freedom in the x-sector (within BDHP formulation the metric sector decouples as in the massive string case). In contrast to the Nambu-Goto (or BDHP) string the classical constraints are of the second class. In consequence the classical massive string has D "functional" degrees of freedom rather then $D-1$ which one might expect from the modification of the BDHP action by adding the Liouville sector with 1 "functional" degree.

The covariant quantization of the system follows the standard lines mentioned in the previous section. Due to the normal ordering the central extension of the algebra of constraints gets modified

$$[L_n, L_m] = (n-m)L_{n+m} + \frac{1}{12}(D + 1 + 48\beta)(n^3 - n)\delta_{n,-m} \quad , \quad [L_n, \beta_0] = 0 \ .$$

The subspace of physical states $H_{\rm ph} \subset H$ is defined by

$$L_n|\psi\rangle = 0 \quad , \quad n \geq 0 \ ; \quad (L_0 - a_0)|\psi\rangle = 0 \ ; \quad (\beta_0 - q)|\psi\rangle = 0 \quad ,$$

where the classical constraint $\beta_0 = 0$ has been replaced by the quantum condition $(\beta_0 - q)|\psi\rangle = 0$ with an arbitrary real parameter q.

The explicit construction of physical states in terms of appropriately modified DDF operators was given in Ref.3. It leads to the following no-ghost theorem.

For each real q the space of physical states in the massive string model is ghost free if and only if one of the following two conditions is satisfied:

$$a_0 \leq 1 \quad , \quad 0 < \beta \leq \frac{24 - D}{48} \quad ;$$

or

$$\beta = \beta_m \quad , \quad a_0 = a_{rs}(m) \quad for \quad m = 2, 3, \ldots; \quad 1 \leq r \leq m-1; \quad 1 \leq s \leq r \quad ;$$

where $\beta_m \equiv \frac{24 - D}{48} + \frac{1}{8m(m+1)}$, $a_{rs}(m) \equiv 1 - \frac{((m+1)r - ms)^2 - 1}{4m(m+1)}$.

In the case of $a_0 \leq 1$, $0 < \beta \leq \frac{24-D}{48}$ there are no null states and the number of degrees of freedom in the quantum theory is the same as in the classical one. For $0 < \beta \leq \frac{24-D}{48}$ and $a_0 = 1$ there is an infinite dimensional family of null states. Also for quantum models from the discrete series $a_0 = a_{rs}(m)$, $\beta = \beta_m$ the space H_0 of null states is nontrivial. The most interesting quantum model corresponds to $\beta = \beta_2 = \frac{25-D}{48}$ and $a_0 = a_{11}(2) = 1$. In this case the space H_0 is largest possible and one has $D-1$ "functional" degrees of freedom in the quantum theory. In order to emphasize this special structure we call this model the critical massive string. It coincides with the quantum model constructed long time ago with the use of the Farlie realization of the Virasoro algebra by Chodos and Thorn[8] . All quantum massive string models allowed by the no-ghost theorem are presented on the upper part of the diagram on Fig.1.

NON-CRITICAL LIGHT-CONE STRING

An explicitly unitary formulation of the perturbation series of the interacting theory requires a formulation of the free string in terms of independent physical degrees of freedom and their physical time evoultion. The quantization of the Nabu-Goto string in the light-cone gauge provides such formulation only for the critical dimension $D = 26$, and the intercept $\alpha(0)=1$. Beyond the critical dimension the Poincare algebra develops anomalous terms and this formulation breaks down[9]. One can however try to improve the model by adding longitudinal excitation in an appropriate way. Some hints for such construction can be obtained by analysing the time dependence of physical states of the covariantly quantized Nambu-Goto string in non-critical dimensions. The non-critical light-cone string briefly presented below was recently constructed in Ref.4. A similar construction motivated by the Liouville theory was also analysed by Marnelius in the context of non-critical Polyakov string[14].

One starts with the choice of a light-cone frame in the flat D-dimensional Minkowski target space. It consists of two light-like vectors k, k' satisfying $k \cdot k' = -1$, and an orthonormal basis $\{e_i\}_{i=1}^{D-2}$ of transverse vectors orthogonal to both k and k'. We introduce the operators

$$P^+ = k \cdot P \quad , \quad x^- = k' \cdot x \quad , \quad P^i = e^i \cdot P \quad , \quad x^i = e^i \cdot x \quad , \quad i = 1, .., D-2 \quad ,$$

corresponding to the light-cone components of the string total momentum and position and satisfying the standard commutation relations

$$[P^i, x^j] = i\delta^{ij} \quad , \quad [P^+, x^-] = -i \quad .$$

For each set of eigenvalues $p^+, \overline{p} = \sum p^i e^i$ of the total momentum component operators we define the Fock space $\mathcal{F}^{\mathrm{T}}(p^+, \overline{p})$ of transverse string excitations generated by the creation operators

$$[\alpha_m^i, \alpha_n^j] = m\delta^{ij}\delta_{m,-n} \quad , \quad (\alpha_m^i)^\dagger = \alpha_{-m}^i \quad , \quad m, n \in \mathbb{Z} \quad ,$$

out of the unique vacuum state $\Omega(p^+,\bar{p})$ satisfying

$$P^i\Omega(p^+,\bar{p}) = p^i\Omega(p^+,\bar{p}) \;\;,\;\; P^+\Omega(p^+,\bar{p}) = p^+\Omega(p^+,\bar{p}) \;\;,\;\; \alpha^i_m\Omega(p^+,\bar{p}) = 0 \;\;,\;\; m > 0 \;.$$

In order to describe the longitudinal string excitations we introduce the Verma module $V^{\mathrm{L}}(b)$ of the Virasoro algebra

$$[L^{\mathrm{L}}_n, L^{\mathrm{L}}_m] = (n-m)L^{\mathrm{L}}_{n+m} + \tfrac{c}{12}(n^3-n)\delta_{m,-n}$$

with the highest wight state

$$L^{\mathrm{L}}_0\,\Omega^{\mathrm{L}}(b) = b\,\Omega(b) \;\;,\;\; L^{\mathrm{L}}_n\,\Omega^{\mathrm{L}}(b) \;=\; 0 \;\;,\;\; n > 0 \;\;.$$

For value of the central charge c within the range $1 < c < 25$ and for $b > 0$ the hermicity properties of the generators:

$$L^{\mathrm{L}\dagger}_n = L^{\mathrm{L}}_{-n} \;\;,\;\; n \in \mathbb{Z} \;\;,$$

determine a positively defined non-degenerate inner product inducing a Hilbert space structure on $V^{\mathrm{L}}(b)$. For $b = 0$ this inner product acquires null directions and for $b < 0$ one gets negative norm states in $V^{\mathrm{L}}(b)$. The full space of states in the non-critical light-cone string model is defined as the direct integral of Hilbert spaces

$$H_{\mathrm{lc}} = \int_{\mathbb{R}\setminus\{0\}} \frac{dp_+}{|p^+|} \int_{\mathbb{R}^{d-2}} d^{d-2}\bar{p}\; \mathcal{F}^{\mathrm{T}}(p^+,\bar{p}) \otimes \mathcal{V}^{\mathrm{L}}(b)$$

In order to complete the construction one has to introduce a unitary realization of the Poincare algebra on H_{lc}. For that purpose we define the transverse Virasoro generators

$$L^{\mathrm{T}}_n = \tfrac{1}{2}\sum_{m=-\infty}^{+\infty} :\bar{\alpha}_{-m}\cdot\bar{\alpha}_{n+m}: \;\;,$$

where $\alpha^\mu_0 = \frac{1}{\sqrt{\alpha}}P^\mu$, and the dimensionful parameter α related to the conventional Regge slope α' by $\alpha = \frac{1}{2\alpha'}$. The generators of translations in the longitudinal and in the transverse directions are given by the operators P^+ and P^i, $i = 1,\ldots, D-2$, respectively. The generator of translation in the x^+-direction is defined by

$$P^- = \frac{\alpha}{P^+}(L^{\mathrm{T}}_0 + L^{\mathrm{L}}_0 - a_0) \;\;.$$

Within the light-cone formulation the x^+ coordinate is regarded as an evolution parameter. In consequence P^- plays the role of the Hamiltonian and the Schrödinger equation reads

$$i\frac{\partial}{\partial x^+}\Psi = P^-\Psi \;\;.$$

The generators of Lorentz rotations are defined by

$$\begin{aligned}
M^{ij}_{\mathrm{lc}} &= P^i x^j - P^j x^i + i\sum_{n\geq 1}\frac{1}{n}(\alpha^i_{-n}\alpha^j_n - \alpha^j_{-n}\alpha^i_n) \;\;,\\
M^{i+}_{\mathrm{lc}} &= P^+ x^i \;\;,\\
M^{+-}_{\mathrm{lc}} &= \tfrac{1}{2}(P^+x^- + x^-P^+) \;\;,\\
M^{i-}_{\mathrm{lc}} &= \tfrac{1}{2}(x^iP^- + P^-x^i) - P^ix^- - i\frac{\sqrt{\alpha}}{P^+}\sum_{n\geq 1}\frac{1}{n}\left(\alpha^i_{-n}(L^{\mathrm{T}}_n + L^{\mathrm{L}}_n) - (L^{\mathrm{T}}_{-n} + L^{\mathrm{L}}_{-n})\alpha^i_n\right).
\end{aligned}$$

They are all self-adjoint operators. The algebra of P^+, P^i, P^-, and $M^{\mu\nu}$closes to the Lie algebra of Poincare group if and only if the central charge c of the Virasoro algebra generating the "longitudinal" Verma module $V^L(b)$ and a_0 entering the definition of the Hamiltonian of the system take the critical values c=26 – D , a_0 = 1. Indeed, the only anomalous terms appear in the commutators:

$$\left[M^{i-}_{\rm lc} , M^{j-}_{\rm lc} \right] = -\left(2 - \tfrac{D-2+c}{12}\right) \frac{\alpha}{P^{+2}} \sum_{n>0} n(\alpha^i_{-n}\alpha^j_n - \alpha^j_{-n}\alpha^i_n)$$
$$-\left(\tfrac{D-2+c}{12} - 2a_0\right) \frac{\alpha}{P^{+2}} \sum_{n>0} \frac{1}{n}(\alpha^i_{-n}\alpha^j_n - \alpha^j_{-n}\alpha^i_n)$$

The model contains one free parameter b entering the mass shell condition $M^2 = 2P^+P^- - \vec{P}^2 = 2\alpha(N + b - 1)$. As it was mentioned above the only restriction on b is $b \geq 0$.

CONCLUSIONS

A common feature of all free non-critical string quantum models considered above and schematically presented on Fig.1 is presence of longitudinal excitations. Indeed all these models have more degrees of freedom than $D - 2$ transverse excitations characteristic for the classical Nambu-Goto string and its quantum counterpart in the critical dimension. On general physical grounds one can expect that the number of longitudinal degrees of freedom is the same as in each transverse direction. It means in our slightly informal terminology that one should expect $D - 1$ "functional" degrees of freedom. There are three models with this property: the non-critical light-cone string, the critical massive string, and the covariantly quantized non-critcal Nambu-Goto string. It turns out that in spite of different origins all these models are just different descriptions of the same quantum theory which we call the critical massive string. The proof of this fact based on the DDF operators technique was recently given in Ref.4. It should be also stressed that there exists yet another way of constructing this model. It was shown by explicit calculation of the string propagator that the Polyakov path over bordered surfaces leads in the range of dimensions $1 < D < 25$ to the critical massive string15.

So many equivalent descriptions of the critical massive string makes it a very interesting object to study. In particular one can calculate the spin content of the model using the light-cone description and the Fock space realization of the Verma module of longitudinal excitations. This realization opens new possibilities for analysing the interacting theory. This is in fact the first formulation in which one can construct and analyse the joining-splitting interactions of non-critical strings with longitudinal excitations. Whether or not there exists a consistent interacting theory is still an open problem. The fact that the old no-go arguments concern only dual model constructions or strings with only transverse excitations makes this question especially fascinating.

Let us finally mention very peculiar feature of the critical massive string. It seems that there is no covariant classical counterpart of the model with the right number of degrees of freedom. It can be obtained either by anti-anomalous quantization of the classical massive string or by the anomalous quantization of the Nambu-Goto or BDHP string. On the other hand the classical counterpart of the light-cone formulation of the model can be easily find but it is not covariant because of anomalous terms present in the Poisson bracket algebra of Poincare generators.

ACKNOWLEDGMENTS

This work is supported in part by the Polish Committee of Scientific Research (Grant Nr PB 1337/PO3/97/12).

REFERENCES

1. "Dual Theory," M. Jacob, ed., North-Holland, Amsterdam - Oxford (1974).
2. S. N. Ganguli and D. P. Roy, *Phys.Rep.* 67:201 (1980).
3. Z. Hasiewicz and Z. Jaskólski, *Nucl.Phys.* B464:85 (1996).
4. M. Daszkiewicz, Z. Hasiewicz, and Z.Jaskólski, "Non-critical light-cone string", IFT UWr-912-97 (hep-th/9706180), (1997)
5. Z. Hasiewicz, Z.Jaskólski, and A.Ostrowski, "No-ghost theorem for massive fermionic string", in preparation.
6. P. Goddard and C. B. Thorn, *Phys.Lett.* 40B:235 (1972).
7. R. C. Brower, *Phys.Rev.* D6:1655 (1972),.
8. A. Chodos and C. B. Thorn, *Nucl.Phys.* B72:509 (1974).
9. P. Goddard, J. Goldstone, C. Rebbi, and C. B. Thorn, *Nucl.Phys.* B56:109 (1973).
10. P. Ginsparg and G. Moore, "Lectures on 2D Gravity and 2D String Theory", YCTP-P23-92 (LA-UR-92-3479; hep-th/9304011), (1992/93).
11. A. Polyakov, *Mod.Phys.Lett.* A2:893 (1987) 893;
    V. Knizhnik, A. Polyakov, and A. Zomolodchikov, *Mod.Phys.Lett.* A3:819 (1988);
    F. David, *Mod.Phys.Lett.* A3:1651 (1988);
    J. Distler and K. Kawai, *Nucl.Phys.* B321:509 (1989).
12. C. G. Callan, S. B. Giddings, J. A. Harvey, and A. Strominger, *Phys.Rev.* D45:Rl005 (1992).
13. R. Marnelius, *Nucl.Phys.* B211:14 (1983).
14. R. Marnelius, *Phys.Lett.* 172B:337 (1986).
15. Z. Jaskólski and K. A. Meissner, *Nucl.Phys.* B428:331 (1994).

This page intentionally left blank

SEIBERG-WITTEN THEORY, INTEGRABLE SYSTEMS AND D-BRANES

A. Marshakov,[1]

[1]Theory Department
P.N.Lebedev Physics Institute
Leninsky prospect 53
Moscow, 117924, Russia
and
ITEP
Moscow 117259, Russia

INTRODUCTION

A lot of ideas has appeared recently in theoretical physics due to developments in what can be called nonperturbative string theory or *M*-theory *. The basic concept is that the physically interesting quantum field theories (QFT's) could be considered as various vacua of *M*-theory and the stringy symmetries or dualities may relate spectra and correlation functions in one QFT with those in another QFT; then typical string duality $R \leftrightarrow \frac{\alpha'}{R}$ allows in principle to relate perturbative regime in one model with the nonperturbative in another.

This general idea at the moment was put to more solid ground only for the case of complex backgrounds in string theory (≡ supersymmetric (SUSY) quantum field theories). In such theories physical data of the model (masses, couplings, etc) can be considered as functions on *moduli spaces* of complex manifolds and the duality symmetry can be regarded as action of a modular group. Useful information can be extracted by powerful machienery of complex geometry. Despite some progress achieved along these lines in the popular scheme of string compactifications on the Calabi-Yau manifolds (see, for example[1] and references therein), rigorous statements about the net result of nonperturbative string theory can be made till now only when nontrivial complex geometry can be effectively reduced to the geometry of one-dimensional complex (≡ two-dimensional real) manifolds – complex curves Σ. One of the most interesting examples of exact statements in this field is the Seiberg-Witten anzatz for the Coulomb phase of $N = 2$ SUSY Yang-Mills theories in four dimensions[2].

*There is no fixed terminology yet in this field – sometimes the term *M*-theory is applied in more "narrow" sense – to the theory of membranes, M(atrix) models etc. In this note we will use the term *M*-theory in wide sense – identify it with the hypothetical (11-dimensional) nonperturbative "string" theory.

New Developments in Quantum Field Theory
Edited by Damgaard and Jurkiewicz, Plenum Press, New York, 1998

In this note I will explain once more how the Seiberg-Witten (SW) anzatz arises from the brane configurations in *M*-theory along the lines of[3, 4]. Following the approach of [5, 6, 8], the language of integrable systems will be used for the formulations of the exact results in $N = 2$ SUSY four-dimensional gauge theories. I will try to pay attention to the subtleties of the exact formulation of the results in these terms and demonstrate how some of them are governed by the Diaconescu-Hanany-Witten-Witten (DHWW) construction.

SEIBERG-WITTEN ANZATZ: INTEGRABLE SYSTEMS

For the $N = 2$ SUSY gauge theory the SW anzatz can be *formulated* in the following way. The $N = 2$ SUSY vector multiplet has necessarily (complex) scalars with the potential $V(\phi) = \mathrm{Tr}[\phi, \phi^\dagger]^2$ whose minima (after factorization over the gauge group) correspond to the diagonal ($[\phi, \phi^\dagger] = 0$), constant and (in the theory with $SU(N_c)$ gauge group) traceless matrices. Their invariants

$$\det(\lambda - \phi) \equiv P_{N_c}(\lambda) = \sum_{k=0}^{N_c} s_{N_c-k}\lambda^k \tag{1}$$

(the total number of algebraically independent ones is rank $SU(N_c) = N_c - 1$) parameterize the moduli space of the theory. Due to the Higgs effect the off-diagonal part of the gauge field $\mathbf{A}_\mu$ becomes massive, since

$$[\phi, \mathbf{A}_\mu]_{ij} = (\phi_i - \phi_j)\mathbf{A}_\mu^{ij} \tag{2}$$

while the diagonal part, as it follows from (2) remains massless, i.e. the gauge group $G = SU(N_c)$ breaks down to $U(1)^{\mathrm{rank}G} = U(1)^{N_c-1}$.

The effective abelian theory is formulated in terms of a finite-dimensional integrable system: the spectral curve Σ defined over the genus-dimensional subspace of the full moduli space, e.g.

$$\Lambda^{N_c}\left(w + \frac{1}{w}\right) = 2P_{N_c}(\lambda) \tag{3}$$

for the pure $SU(N_c)$ gauge theory [†]; and the generating differential

$$dS = \lambda\frac{dw}{w} \tag{4}$$

whose basic property is that its derivatives over $N_c - 1$ moduli give rise to holomorphic differentials. The data (Σ, dS) with such properties are exactly the definition of the integrable system in the sense of [7] (see [8] and references therein for details). The period matrix of Σ $T_{ij}(\mathbf{a})$ as a function of the action variables

$$\mathbf{a} = \oint_{\mathbf{A}} dS \qquad \mathbf{a}^D = \oint_{\mathbf{B}} dS$$

$$T_{ij} = \frac{\partial a_i^D}{\partial a_j} \tag{5}$$

gives the set of coupling constants in the effective abelian $U(1)^{N_c-1}$ theory while action variables themselves are identified with the masses of the BPS states $M^2 \sim |\mathbf{n}\mathbf{a} + \mathbf{m}\mathbf{a}_D|^2$ with the $(\mathbf{n}, \mathbf{m})$ "electric" and "magnetic" charges.

[†] The genus of the curve (3) is $N_c - 1$, i.e. exactly equal to the number of independent parameters of the polynomial $P_{N_c}(\lambda)$ (1).

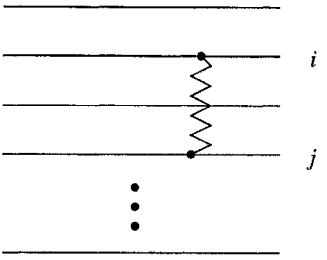


Figure 1. Open strings, stretched between D-branes induce the interaction via non-Abelian gauge fields $\mathbf{A}^{ij}$.

SPECTRAL CURVE AS TOPOLOGICALLY NONTRIVIAL PART OF M-THEORY 5-BRANE WORLD VOLUME

Let us show now that the spectral curve and generating differential can naturally arise from brane configurations [3].

- First step is to obtain a gauge group $SU(N_c)$ broken down to $U(1)^{N_c-1}$. The most elegant way of doing this in string theory is to introduce D-branes into type II string theory – the submanifolds in target space where strings can have their ends. N_c parallel D-branes would correspond exactly to what we need now since string strechted between i-th and j-th brane $(i, j = 1, \ldots, N_c)$ (see Fig.1) will have a vector field $\mathbf{A}^{ij}$ in its spectrum such that mass of this vector field is proportional to the length of the string, i.e. to the distance between i-th and j-th branes. This the $U(1)^{N_c-1}$ massless factor will come from strings having both ends on the same D-brane while the $\mathbf{A}^{ij}$ fields with $i \neq j$ will acquire "Higgs" masses (2) where scalars vev's are us usual proportional to the "transverse" co-ordinate of the D-brane $\phi \sim \frac{\sqrt{\vec{x}_\perp^2}}{\alpha'}$.

- Next step is that from 10-dimensional type II string theory ($\mathbf{A} = \| \mathbf{A}^{ij} \|$ is 10-dimensional gauge field in string picture) one wants to get 4-dimensional one. A natural way to reduce the number of space-time dimensions is to restrict ourselves to the effective theory on D-brane world volume. The world volume of the Dirichlet p-brane is p + 1-dimensional, so naively in order to get 4-dimensional theory one should consider D3-branes. This scenario is quite possible and realized in another context; however to get the SW anzatz it is better to use another option, the DHWW brane configuration with the N_c parallel D4-branes streched between two vertical walls (see Fig.2), so that the naive 5-dimensional D4-world-volume theory is macroscopically (in the light sector) 4-dimensional by conventional Kaluza-Klein argument for a system compactified on a circle or put into a box.

  The role of vertical walls should be played by 5-branes [3], this follows from the β-function considerations: the logariphmic behaivior of the macroscopic coupling constant can be ensured in the first approximation if corresponding co-ordinate (x^6) has logariphmic behaivior as a function of "transverse" direction, i.e. satisfy *two*-dimensional Laplace equation. The effective space is two-dimensional if parallel D-branes are streched between the 5-branes.

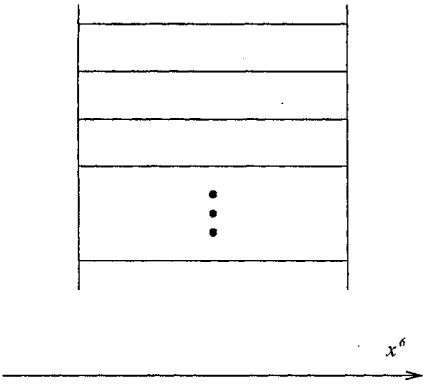


Figure 2. The 4-branes restricted by 5-branes to the finite volume (in horisontal x^6-direction) give rise to macroscopically 4-dimensional theory.

- The obtained picture of 4 and 5 branes in 10 dimensions is of course very rough and true in (semi)classical approximation. In particular it is naively singular at the points where 4-branes meet 5-branes. These singularities were resolved by Witten [3] who suggested to put the whole picture into 11-dimensional target space of M-theory with compact 11-th dimension and to consider D4-branes as M-theory 5-branes compactified to 11-th dimension. Thus the picture in Fig.2 becomes similiar rather to the surface of "swedish ladder" [‡] and apart of macroscopic directions $x^0, \ldots, x^3$ looks like (non-compact) Riemann surface with rather special properties (see Fig.3).

- In other words, one gets a 5-brane parameterized by $(x^0, x^1, x^2, x^3, x^6, x^9)$, which leaving aside four flat dimensions (x^0, x^1, x^2, x^3) along these lines ends up to N_c cylinders $R \times S^1$ embedded into the target space along, say, (x^6, x^9) dimensions (using notation $z = x^6 + ix^9$ for the corresponding complex co-ordinate). Different cylinders have different positions in the space $V^\perp = (x^4, x^5, x^7, x^8)$. Moreover the cylinders are all glued together (see Fig.3). The "effective" two-dimensional subspace of $V^\perp$ we will describe it in terms of the complex coordinate $\lambda = x^4 + ix^5$.

  Introducing coordinate $w = e^z$ to describe a cylinder, we see that the system of non-interacting branes (Fig. 1) is given by z-independent equation

$$P_{N_c}(\lambda) = \prod_{\alpha=1}^{N_c} (\lambda - \lambda_\alpha) = 0, \tag{6}$$

  while their bound state (Fig.3) is described rather by the complex curve Σ_{N_c} (3) or:

$$\Lambda^{N_c} \cosh z = P_{N_c}(\lambda) \tag{7}$$

  In the weak-coupling limit $\Lambda \to 0$ (i.e. $\frac{1}{g^2} \sim \log \Lambda \to \infty$) one comes back to disjoint branes (6) [§]. Thus we finally got a 5-brane of topology $R^3 \times \sum_{N_c}$

[‡]I am grateful to V.Kazakov for this not quite exact but illuminating comparison.

[§]Eqs.(3), (7) and Fig.3 decribe a hyperelliptic curve – a double covering of a punctured Riemann sphere,

$$y^2 = \frac{\Lambda^{2N_c}}{4}\left(w - \frac{1}{w}\right)^2 = P_{N_c}^2(\lambda) - \Lambda^{2N_c} \tag{8}$$

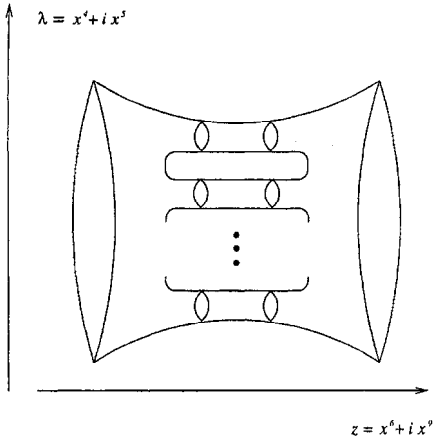


Figure 3. The brane configuration, represented as a result of "blowing up" Fig.2 – the ladder turns into hyperelliptic Riemann surface being at the same time N_c-fold covering of the horisontal cylinder.

embedded into a subspace $R^6 \times S^1$ (spanned by $x^1, \ldots, x^6, x^9$) of the full target space. The periodic coordinate is

$$x^9 = \arg P_{N_c}(\lambda) = \mathrm{Im} \log P_{N_c}(\lambda) = \sum_{\alpha=1}^{N_c} \arg(\lambda - \lambda_\alpha) \tag{9}$$

INTEGRABLE EQUATIONS FROM BRANE PICTURE

The arguments of the previous section show that the nontrivial part of the 5-brane world-volume looks rather similiar to the spectral curves arising in the exact formulas of the SW anzatz. The way to justify this proposed in [4] was based on parallels with the theory of integrable systems.

The integrable equations in this context arise as reductions on "unvisible" dimensions of the equations of motion (better the "square-root" of the: the BPS-like conditions) of the world-volume theory. In [9] Diaconescu using this idea obtained the Nahm equations. In [4] it was demonstrated that the simplest way for getting algebraic equation for the topologically nontrivial part of brane configurations may be searched among the Hitchin systems [10].

The Hitchin system on elliptic curve

$$y^2 = (x - e_1)(x - e_2)(x - e_3)$$

$$x = \wp(z) \quad y = 2\wp'(z) \quad dz \sim 2\frac{dx}{y} \tag{10}$$

with p marked points $z_1, \ldots, z_p$ can be defined by [11, 12] $(i, j = 1, \ldots, N)$

$$\bar{\partial}\Phi_{ij} + (a_i - a_j)\Phi_{ij} = \sum_{\alpha=1}^{p} J_{ij}^{(\alpha)} \delta(z - z_\alpha) \tag{11}$$

so that the solution has the form ($a_{ij} \equiv a_i - a_j$)

$$\Phi_{ij}(z) = \delta_{ij}\left(p_i + \sum_\alpha J_{ii}^{(\alpha)} \partial \log\theta(z - z_\alpha|\tau)\right) +$$

$$+ (1-\delta_{ij})\, e^{a_{ij}(z-\bar{z})} \sum_\alpha J_{ij}^{(\alpha)} \frac{\theta(z - z_\alpha + \frac{\mathrm{Im}\tau}{\pi} a_{ij})\theta'(0)}{\theta(z-z_\alpha)\theta(\frac{\mathrm{Im}\tau}{\pi} a_{ij})} \tag{12}$$

The exponential (nonholomorphic) part can be removed by a gauge transformation

$$\Phi_{ij}(z) \to (U^{-1}\Phi U)_{ij}(z) \tag{13}$$

with $U_{ij} = e^{a_{ij}\bar{z}}$.

The additional conditions to the matrices $J_{ij}^{(\alpha)}$ are

$$\sum_{\alpha=1}^{p} J_{ii}^{(\alpha)} = 0 \tag{14}$$

having the clear meaning that the sum of all residues of a function Φ_{ii} is equal to zero, and

$$\mathrm{Tr} J^{(\alpha)} = m_\alpha \tag{15}$$

with m_α = const being some parameters ("masses") of a theory. The spectral curve equation becomes

$$\mathcal{P}(\lambda; z) \equiv \det_{N\times N}(\lambda - \Phi(z)) = \lambda^N + \sum_{k=1}^{N} \lambda^{N-k} f_k(z) = 0 \tag{16}$$

where $f_k(z) \equiv f_k(x,y)$ are some functions (in general with k poles) on the elliptic curve (10). If, however, $J^{(\alpha)}$ are restricted by

$$\mathrm{rank} J^{(\alpha)} \le l \qquad l < N \tag{17}$$

the functions $f_k(z)$ will have poles at $z_1, \ldots, z_p$ of the order not bigger than l. The generating differential, as usual, should be

$$dS = \lambda dz \tag{18}$$

and its residues in the marked points $(z_\alpha, \lambda^{(i)}(z_\alpha))$ (different i correspond to the choice of different sheets of the covering surface) are related with the mass parameters (15) by

$$m_\alpha = \mathrm{res}_{z_\alpha} \lambda dz \equiv \sum_{i=1}^{N} \mathrm{res}_{\pi_{(i)}^{-1} z_\alpha} \lambda^{(i)}(z) dz = \mathrm{res}_{z_\alpha} \mathrm{Tr}\Phi dz \tag{19}$$

It is easy to see that the general form of the curve (16) coincides with the general curves proposed in [3] at least for $l = 1$, i.e. sourses of rank 1 (for the "rational" case the torus should be degenerated into a cylinder).

In [4] the Toda-chain spectral curve (3), (7) has been derived from the $SU(N_c)$ Hitchin system on torus with one marked point $p = 1$ in the double-scaling limit. Of course it is possible to write down the Hitchin equations directly on the bare cylinder with *trivial* gauge connection

$$\bar{\partial}_v \tilde{\mathcal{L}}^{TC}(v) = -\left(e^{-\boldsymbol{\alpha}_0\boldsymbol{\phi}} E_{\boldsymbol{\alpha}_0} + \sum_{\text{simple } \alpha} e^{\boldsymbol{\alpha}\boldsymbol{\phi}} E_{-\boldsymbol{\alpha}}\right)\delta(P_\infty) +$$

$$+\left(e^{-\alpha_0\phi}E_{-\alpha_0}+\sum_{\text{simple }\alpha}e^{\alpha\phi}E_\alpha\right)\delta(P_0) \tag{20}$$

where they can be easily solved giving rise to

$$\tilde{\mathcal{L}}^{TC}(v)=U^{-1}\mathcal{L}^{TC}(w)U=$$

$$=\mathbf{pH}+v^{-1}\left(e^{-\alpha_0\phi}E_{-\alpha_0}+\sum_{\text{simple }\alpha}e^{\alpha\phi}E_\alpha\right)+v\left(e^{-\alpha_0\phi}E_{\alpha_0}+\sum_{\text{simple }\alpha}e^{\alpha\phi}E_{-\alpha}\right) \tag{21}$$

GENERATING DIFFERENTIAL

Let us now turn to more subtle point and discuss how the auxiliary spectral Riemann surface is embedded into 11-dimensional target space. Partially that has been illustrated already above when the explicit formulas relating 11-dimensional co-ordinates x^I with the internal co-ordinates on the surface λ or z by $\lambda = x^4 + ix^5$ and $z = x^6 + ix^9$ were presented. In this section I will demonstrate that this embedding is in fact governed by the generating differential and its variations.

Already looking at Fig.3 it is clear that the corresponding Riemann surface is not compact; it means that the metric should have singularities at "infinities". Indeed, the metric in the target-space is flat (in case of absense of matter) $ds^2 \sim \sum_I \left(dx^I\right)^2$ and it means that the area of the surface is measured by

$$\Omega_\Sigma \sim \delta z \wedge \delta\bar{z} + \delta\lambda \wedge \delta\bar{\lambda} \tag{22}$$

Since the BPS massive spectrum in the theory is determined by the states corresponding to the 2-branes wrapped over the nontrivial cycles of "internal" complex manifold – in our case the Seiberg-Witten curve – the BPS masses should be proportional to the area of this surface [13]. This area is measured by another (holomorphic) two form

$$\Omega \sim \delta\lambda \wedge \delta z \tag{23}$$

which is directly related to the variation of generating differential or to the symplectic form of the corresponding integrable system.

The variation of generating differential over distinguished subfamily of moduli gives rise to holomorphic differentials

$$\delta_{a_i} dS \sim d\omega_i \tag{24}$$

The derivative over moduli (24) is taken after some connection is chosen – for example under condition that some function is covariantly constant. For the differential (4), (18) the canonical procedure implies that the covariantly constant function is $z = \log w$ so that

$$\delta dS|_{z=\text{const}} = \delta\lambda dz = -\sum_i \delta a_i \frac{\mathcal{P}_{a_i}}{\mathcal{P}_\lambda} dz \equiv \sum_i \delta a_i d\omega_i \tag{25}$$

¿From the point of view of M-theory one considers an effective theory on 5-brane world-volume with the co-ordinates $(x^0, \ldots, x^3, x^6, x^9)$. It means that when studying the 4-dimensional effective theory on "horisontal cylinders" on Fig.3 one should take the variation of λ which has the sense of vev of some (Higgs) field keeping fixed the world-volume coordinates $x^6 = \text{Re}z$ and $x^9 = \text{Im}z$. Physically this corresponds to the fact that we are taking the variation (24), (25) over the vev's of scalar fields only – which play the role of physical moduli in the system.

In principle, this procedure can be correctly defined if one notices that the differential dS possesses double zeroes ¶; and it is the action of g singular vector fields $L_{-2}^{(i)}$ at these points that gives rise to the distinguished subfamily of co-ordinates on the moduli space. More detailed discussion of the properties of generating differential is beyond the scope of this note (see the last ref. of [8] for some details).

ACKNOWLEDGMENTS

I am grateful to S.Kharchev, I.Krichever, A.Levin, A.Mironov, N.Nekrasov and especially to A.Morozov for valuable discussions and to T.Sultanov for the help in preparing figures.

The work was partially supported by the Cariplo Foundation, Institute of Physics of Cambridge University and RFBR grant 96-02-16117.

REFERENCES

1. see for example
S.Katz, P.Mayr and C.Vafa, hep-th/9706110 and references therein.
2. N.Seiberg and E.Witten, Nucl.Phys. B426 (1994) 19, hep-th/9407087; Nucl.Phys. B431 (1994) 484, hep-th/9408099; hep-th/9607163;
A.Klemm, W.Lerche, S.Theisen and S.Yankielowicz, Phys.Lett. B344 (1995) 169, hep-th/9411048;
P.Argyres and A.Farragi, Phys.Rev.Lett. 74 (1995) 3931, hep-th/9411057;
A.Hanany and Y.Oz, Nucl.Phys. B452 (1995) 283, hep-th/9505074.
3. E.Witten, hep-th/9703166.
4. A.Marshakov, M.Martellini and A.Morozov, hep-th/9706050.
5. A.Gorsky, I.Krichever, A.Marshakov, A.Mironov and A.Morozov, Phys.Lett. B355 (1995) 466, hep-th/9505035.
6. E.Martinec and N.Warner, Nucl.Phys. B459 (1996) 97, hep-th/9509161;
T.Nakatsu and K.Takasaki, Mod.Phys.Lett. A11 (1996) 157-168, hep-th/9509162;
R.Donagi and E.Witten, Nucl.Phys. B460 (1996) 299, hep-th/9510101;
E.Martinec, Phys.Lett. B367 (1996) 91, hep-th/9510204;
A.Gorsky and A.Marshakov, Phys.Lett. B375 (1996) 127, hep-th/9510224;
E.Martinec and N.Warner, hep-th/9511052;
H.Itoyama and A.Morozov, Nucl.Phys. B477 (1996) 855, hep-th/9511126;
R.Donagi, alg-geom/9705010.
7. B.Dubrovin, I.Krichever and S.Novikov, *Integrable systems - I, Sovremennye problemy matematiki (VINITI), Dynamical systems - 4* (1985) 179.
8. A.Marshakov, hep-th/9607159/9608161/9610242/9702187.
9. D.-E.Diaconescu, hep-th/9608163.
10. N.Hitchin, Duke.Math.J. 54 (1987) 91.
11. N.Nekrasov, hep-th/9503157.
12. A.Levin and M.Olshanetsky, alg-geom/9706010.
13. A.Fayyazuddin and M.Spalinski, hep-th/9706087;
M.Hennigson and P.Yi, heoth/9707251;
A.Mikhailov, hep-th/9708068.

¶ More strictly, since only the periods of dS are "observable" there is always a "canonical" representative in the universality class of the differentials with fixed periods $dS_1 \cong dS_2$ with double zeroes at certain points.

MICROSCOPIC UNIVERSALITY IN RANDOM MATRIX MODELS OF QCD

Shinsuke M. Nishigaki *

Niels Bohr Institute
University of Copenhagen
Blegdamsvej 17
DK-2100 Copenhagen Ø
Denmark

INTRODUCTION

Random Matrix Theory (RMT) has been a useful laboratory for simulating hamiltonians of statistical systems with random matrix elements[1], due possibly to impurities scattered around the material. An unphysical yet simplest choice is to assume each of matrix elements of the hamiltonian M to be independently derived from the gaussian distribution,

$$Z = \int_{N\times N \text{ matrices}} dM\, \mathrm{e}^{-N \operatorname{tr} M^2}, \tag{1}$$

to which is associated Wigner's single-banded, semi-circle spectrum. This macroscopic spectrum is, however, by no means realistic; physical spectra are far more complicated. Namely they may even have multi-band structures, which in the side of RMT correspond to non-gaussian distributions. Thus, it is only universal quantities insensitive to a chosen matrix measure that may be justifiably extracted from RMT for physical systems.

Various quantities concerning microscopic spectral correlation in the bulk of the spectral band have long been believed universal, as identical results are derived from non-gaussian ensembles corresponding to classical orthogonal polynomials other than Hermite[2]. Here the term 'microscopic' refers to that the correlation is measured in the unit of the mean level spacing, which is of order $O(1/N)$ for ensembles of the type (1). This conjectured universality of the sine kernel

$$K_s(\lambda, \lambda') = \frac{\sin \pi(\lambda - \lambda')}{\pi(\lambda - \lambda')}, \tag{2}$$

*Address after Oct. 1997: Institute for Theoretical Physics, University of California, Santa Barbara, CA 93106, USA.

New Developments in Quantum Field Theory
Edited by Damgaard and Jurkiewicz, Plenum Press, New York, 1998

which comprises all spectral correlators

$$\rho_s(\lambda_1, \cdots, \lambda_p) = \left\langle \prod_{a=1}^{p} \frac{1}{N} \mathrm{tr}\, \delta(\lambda_a - M) \right\rangle = \det_{1 \le a,b \le p} K_s(\lambda_a, \lambda_b), \tag{3}$$

is proved[3, 4] for unitary invariant ensembles with generic single-trace potentials $e^{-N \mathrm{tr} V(M)}$.

RMT has recently extended its range of applicability toward QCD[5]. There the Dirac operator $i \not{D} = \gamma_\mu (i \partial_\mu + A_\mu)$, having gauge fields as random elements, is regarded as a random hamiltonian. Schematically, the Euclidean QCD partition function†

$$\begin{aligned} Z_{\mathrm{QCD}} &= \int [dA_\mu \, d\bar{\psi}_f \, d\psi_f] \, e^{-\frac{1}{g^2} \int \mathrm{tr} F^2 - \int \bar{\psi}_f (i \not{D} + m_f) \psi_f} \\ &= \int [dA_\mu] \, e^{-\frac{1}{g^2} \int \mathrm{tr} F^2} \prod_f \mathrm{Det} \begin{pmatrix} m_f & -i\sigma_\mu (A_\mu + i\partial_\mu) \\ i\sigma_\mu (A_\mu + i\partial_\mu) & m_f \end{pmatrix} \end{aligned} \tag{4}$$

is transformed by the change of integration variables

$$A_\mu \mapsto i \not{D}_A \equiv M = \begin{pmatrix} 0 & W^\dagger \\ W & 0 \end{pmatrix} \tag{5}$$

into a RMT

$$Z_{\mathrm{RMT}} = \int_{N \times N \text{ matrices}} d\mu(W) \prod_f \det \begin{pmatrix} m_f & W^\dagger \\ W & m_f \end{pmatrix}. \tag{6}$$

Here N stands for the size of the Dirac operator i.e. the volume of the spacetime, and $d\mu(W)$ a measure invariant under

$$W \to U W V^\dagger \quad (U, V \in \mathrm{U}(N) \text{ for } W \text{ complex}). \tag{7}$$

Novel features of this RMT in addition to the conventional model (1) are:

- Presence of the fermion determinant. For small m_f, it expels Dirac eigenvalues of order $O(m_f)$ from the origin.
- Chiral structure of the random matrix. Since $\{M, \gamma_5\} = 0$ each eigenvalue λ is accompanied by its mirror image $-\lambda$. The Coulomb repulsion between these pairs prevents them from populating around the origin.

Among the quantities calculable from this type of RMTs, of particular interest are the distribution and correlation of soft eigenvalues of the Dirac operator as their accumulation is responsible for the spontaneous breaking of the chiral symmetry via[6]

$$\Sigma = |\langle \bar{\psi} \psi \rangle| = \frac{\pi \rho(0)}{N}. \tag{8}$$

Due to the above two features, correlations of soft eigenvalues are expected to deviate from the universal sine law (2). Verbaarschot and his collaborators[5] have calculated correlations of eigenvalues of order $\lambda \sim O(1/N)$ from the gaussian ensemble

$$d\mu(W) = dW \, e^{-N\Sigma \mathrm{tr} W^\dagger W}, \tag{9}$$

appealing to the conjecture of universality. These analytic predictions have been compared first to the numerical data from the simulation of the instanton liquid model[7], and more recently to that of the lattice gauge theory[8, 9]. The agreements are impressive.

† For simplicity, we consider only the topologically trivial sector.

In this article I shall review the proof of universality of the microscopic spectral correlations of the model (6) for $m_f = 0$ and with single-trace potential measures‡, in order to justify partly the use of gaussian ensemble (9) and to corroborate the above mentioned agreements. Rather than to follow closely the original proof[13, 14], I shall disguise it with the conventional approach of **Q**, **P** operators [15] and with the method employed in a recently proposed alternative proof[16, 17], in a hope that borrowing notions from quantum mechanics may add the proof some pedagogical flavor. At the end of the article I shall address a problem associated with multi-criticality.

ORTHOGONAL POLYNOMIAL METHOD

We start from recalling basic technology of random matrices: the orthogonal polynomial method. We consider the chiral unitary ensemble (χUE)

$$Z_{\chi\mathrm{UE}} = \int_{2N\times 2N\ \mathrm{hermite}} dM\, \mathrm{e}^{-N\,\mathrm{tr}\,V(M^2)}\, |\det M|^{\alpha}, \qquad M = \begin{pmatrix} 0 & W^{\dagger} \\ W & 0 \end{pmatrix} \tag{10}$$

capturing the global symmetries of $N_c \geq 3$ QCD$_4$ with $N_f = \alpha$ massless fundamental fermions, as well as the unitary ensemble (UE) without chiral structure

$$Z_{\mathrm{UE}} = \int_{N\times N\ \mathrm{hermite}} dM\, \mathrm{e}^{-N\,\mathrm{tr}\,V(M^2)}\, |\det M|^{\alpha} \tag{11}$$

modeling QCD$_3$ [18]. We allow non-integer $\alpha > -1$, for they can be treated on the same footing as integer cases. Both ensembles allow eigenvalue representations[19]

$$Z_{\chi\mathrm{UE}} = \int_0^{\infty} \prod_{i=1}^{N} \left(d\lambda_i\, \lambda_i^{\alpha}\, \mathrm{e}^{-NV(\lambda_i)} \right) \left| \det_{ij} \lambda_j^{i-1} \right|^2, \qquad (\lambda_i : \text{eigenvalues of } W^{\dagger} W) \tag{12}$$

$$Z_{\mathrm{UE}} = \int_{-\infty}^{\infty} \prod_{i=1}^{N} \left(d\lambda_i\, |\lambda_i|^{\alpha}\, \mathrm{e}^{-NV(\lambda_i^2)} \right) \left| \det_{ij} \lambda_j^{i-1} \right|^2. \qquad (\lambda_i : \text{eigenvalues of } M) \tag{13}$$

Inside the integrals, $\det \lambda_j^{i-1}$ may be replaced (up to an irrelevant constant) with polynomials $\det P_{i-1}(\lambda_j)$ orthonormal with respect to the measure $d\lambda\, |\lambda|^{\alpha} \mathrm{e}^{-NV(\lambda)}$:

$$\delta_{nm} = \int d\lambda\, |\lambda|^{\alpha} \mathrm{e}^{-NV(\lambda)}\, P_n(\lambda)\, P_m(\lambda) = \int d\lambda\, \psi_n(\lambda)\, \psi_m(\lambda). \tag{14}$$

Here we have introduced one-particle wave functions of a free fermion (at the n-th state)

$$\psi_n(\lambda) = |\lambda|^{\alpha/2} \mathrm{e}^{-NV/2}\, P_n(\lambda). \tag{15}$$

The fermionic nature comes from the Vandermonde determinant. The wave functions at first N levels comprise the Dirac sea, i.e. the N-particle ground-state wave function

$$\Psi_N(\{\lambda\}) = \det_{1\leq n,m\leq N} \psi_{n-1}(\lambda_m). \tag{16}$$

Using this ground-state wave function, the vacuum expectation value of an unitary invariant observable is written as

$$\langle \mathcal{O} \rangle = \int d^N\lambda\, \Psi_N(\{\lambda\})\, \mathcal{O}\, \Psi_N(\{\lambda\}). \tag{17}$$

‡ Microscopic universality for the gaussian potential plus certain terms which breaks the invariance (7) is proved in refs.10, 11. See also ref.12 for a perturbative treatment of a quartic potential.

Namely, spectral correlators are related to ψ_n by

$$\rho_N(\lambda_1,\cdots,\lambda_p) \equiv \left\langle \prod_{a=1}^{p} \frac{1}{N} \mathrm{tr}\, \delta(\lambda_a - M) \right\rangle \overset{\text{Wick's theorem}}{=} \det_{1\le a,b\le p} K_N(\lambda_a,\lambda_b) \tag{18}$$

where K_N is the projector to the Dirac sea:

$$\begin{aligned} K_N(\lambda,\lambda') &= \sum_{n=0}^{N-1} \psi_n(\lambda)\psi_n(\lambda') \\ &= q_N \frac{\psi_N(\lambda)\psi_{N-1}(\lambda') - \psi_{N-1}(\lambda)\psi_N(\lambda')}{\lambda-\lambda'}. \end{aligned} \tag{19}$$

Use is made of Christoffel-Dalboux formula in the last line, and q_N is a constant defined immediately below.

We note that the orthogonality relation (14) for the χUE case $(\int_0^\infty d\lambda)$ can absorbed into the UE case $(\int_{-\infty}^{\infty} d\lambda)$ by the change of variables $\lambda \to \sqrt{\lambda}$ accompanied by the redefinition $\alpha \to \frac{\alpha-1}{2}$ (or $\alpha \to \frac{\alpha+1}{2}$) and $P_{2n}(\sqrt{\lambda}) \to P_n(\lambda)$ (or $\sqrt{\lambda}\, P_{2n+1}(\sqrt{\lambda}) \to P_n(\lambda)$). Thus in the following we need only to consider the UE case, where orthogonal polynomials have definite parities.

Now let us for a while concentrate on $\alpha = 0$, where the logarithmic component of the potential is absent. In this case one may represent the Heisenberg algebra $[\frac{d}{d\lambda}, \lambda] = 1$ by matrices **Q** and **P** acting on the Hilbert space of orthonormal wave functions[15],

$$\lambda\,\psi_n = q_{n+1}\psi_{n+1} + q_n\psi_{n-1} \equiv \sum_m \mathbf{Q}_{nm}\,\psi_m, \tag{20}$$

$$\frac{d}{d\lambda}\psi_n = -\frac{N}{2}V'\,\psi_n + (\text{linear comb. of } \psi_{n-1},\psi_{n-3},\cdots) \equiv \sum_m \mathbf{P}_{nm}\,\psi_m. \tag{21}$$

(Anti-)self adjointness of multiplication and differentiation operators inherit to the (anti-)symmetry of the representing matrices,

$$\int d\lambda\, (\lambda\,\psi_n)\,\psi_m = \int d\lambda\,\psi_n\,(\lambda\,\psi_m) \qquad \Rightarrow \mathbf{Q}_{nm} = \mathbf{Q}_{mn}, \tag{22}$$

$$\int d\lambda \left(\frac{d}{d\lambda}\psi_n\right)\psi_m = -\int d\lambda\,\psi_n \left(\frac{d}{d\lambda}\psi_m\right) \Rightarrow \mathbf{P}_{nm} = -\mathbf{P}_{mn}. \tag{23}$$

Although eq.(21) and the antisymmetry (23) are enough to determine the **P** matrix as

$$\mathbf{P}_{nm} = \mp\frac{N}{2}\,V'(\mathbf{Q})_{nm} \quad (n \underset{>}{<} m), \tag{24}$$

we shall not use this fact except for fixing a integration constant in the sequel.

At this stage we consider the situation when the coefficient q_n become single-valued in the limit $n,\, m \simeq \mathrm{N} \gg 1$,

$$q_n \sim q_{n\pm1} \sim q = O(1), \tag{25}$$

a crucial assumption which is violated in matrix ensembles with multi-band spectra[20]. Then the matrix elements of **Q**, **P** for $n,\, m = N + O(1)$ take the forms:

$$\mathbf{Q}_{nm} = \begin{pmatrix} \ddots & \ddots & & & & \\ \ddots & \ddots & \ddots & & & \\ & \ddots & 0 & q & & \\ & & q & 0 & q & \\ & & & q & 0 & q \\ & & & & q & \ddots \end{pmatrix}, \quad \mathbf{P}_{nm} = N \begin{pmatrix} \ddots & \ddots & \ddots & \ddots & 0 & \ddots \\ \ddots & \ddots & \ddots & 0 & -p_2 & 0 \\ \ddots & \ddots & 0 & -p_1 & 0 & -p_2 \\ \ddots & \ddots & p_1 & 0 & -p_1 & 0 \\ \ddots & p_2 & 0 & p_1 & 0 & -p_1 \\ \ddots & 0 & p_2 & 0 & p_1 & \ddots \end{pmatrix}. \tag{26}$$

A matrix of the above form of **P** can always be written as

$$\mathbf{P} = \frac{N}{2} \sum_{k \geq 0} a_k \, \mathbf{Q}^{2k} \mathbf{\Lambda} \equiv \frac{N}{2} A(\mathbf{Q}) \mathbf{\Lambda}, \tag{27}$$

where the Λ matrix stands for the antisymmetrizer

$$\mathbf{\Lambda}_{nm} = \begin{pmatrix} \ddots & \ddots & & & & \\ \ddots & \ddots & \ddots & & & \\ & \ddots & 0 & -1 & & \\ & & 1 & 0 & -1 & \\ & & & 1 & 0 & -1 \\ & & & & 1 & \ddots \end{pmatrix}. \tag{28}$$

When $\alpha \neq 0$, $\frac{d}{d\lambda}$ can not be represented on the space of orthogonal polynomials as it involves a term proportional to $1/\lambda$. However one may verify[17] that the operator $\frac{d}{d\lambda} - \frac{(-)^{\hat{n}}\alpha}{\lambda}$ is indeed represented by the same **P** matrix as in (27). Here $\hat{n}$ is the number operator $\hat{n}\psi_n = n\psi_n$, and the existence of the signature $(-)^{\hat{n}}$ preserves the antisymmetry of **P**,

$$\psi_{2m+1} \left((-)^{\hat{n}} \psi_{2n} \right) = - \left((-)^{\hat{n}} \psi_{2m+1} \right) \psi_{2n}. \tag{29}$$

In the next section eqs.(26) and (27) are utilized to prove the microscopic universality.

PROOF OF UNIVERSALITY

Here we re-exhibit the asymptotic forms of recursion relations (26), (27) satisfied by the orthonormal wave functions ($\alpha = 0$):

$$\lambda \psi_n = q \left(\psi_{n+1} + \psi_{n-1} \right), \tag{30}$$

$$\frac{d}{d\lambda} \psi_n = \frac{NA(\lambda)}{2} \left(-\psi_{n+1} + \psi_{n-1} \right). \tag{31}$$

They are combined to yield unharmonic analogues of lowering and raising operators of a harmonic oscillator at highly excited levels $n \gg 1$:

$$\left(\frac{1}{NA(\lambda)} \frac{d}{d\lambda} + \frac{\lambda}{2q} \right) \psi_n = \psi_{n-1}, \tag{32}$$

$$\left(-\frac{1}{NA(\lambda)} \frac{d}{d\lambda} + \frac{\lambda}{2q} \right) \psi_n = \psi_{n+1}. \tag{33}$$

The requirement that two processes $\psi_n \overset{\psi_{n+1}}{\underset{\psi_{n-1}}{\diagup\!\diagdown}} \psi_n$ must commute amounts to the suppression of the commutator of the raising and lowering operators in the large-n and N limit,

$$\left[\frac{\lambda}{2q}, \frac{1}{NA(\lambda)} \frac{d}{d\lambda} \right] = O\left(\frac{1}{N^*}\right) \ll 1. \tag{34}$$

Thus we are lead to a second-order differential equation for ψ_n,

$$\left[\left(\frac{1}{NA(\lambda)} \frac{d}{d\lambda} \right)^2 + 1 - \left(\frac{\lambda}{2q} \right)^2 \right] \psi_n = 0. \tag{35}$$

As mentioned in the end of the last section, the only modification for $\alpha \neq 0$ case is the replacement

$$\frac{d}{d\lambda}\psi_n \to \left(\frac{d}{d\lambda} - \frac{(-)^n\alpha}{\lambda}\right)\psi_n. \tag{36}$$

The resulting equation takes the form

$$\left[\frac{1}{N^2}\frac{1}{A(\lambda)}\left(\frac{d}{d\lambda} + \frac{(-)^n\alpha}{\lambda}\right)\frac{1}{A(\lambda)}\left(\frac{d}{d\lambda} - \frac{(-)^n\alpha}{\lambda}\right) + 1 - \left(\frac{\lambda}{2q}\right)^2\right]\psi_n = 0. \tag{37}$$

Let us first take the macroscopic (large-N) limit:

$$N \to \infty, \quad \lambda : \text{fixed} \tag{38}$$

of (37). Since $1/N$ plays the role of $\hbar$ in this limit and thus eq.(37) is reduced to

$$\left[\frac{1}{N^2}\frac{1}{A(\lambda)^2}\frac{d^2}{d\lambda^2} + 1 - \left(\frac{\lambda}{2q}\right)^2\right]\psi_n = 0, \tag{39}$$

we obtain the WKB solution[3] [§]:

$$\psi_n(\lambda) \propto \cos\left(N\int_0^\lambda d\lambda\, A(\lambda)\sqrt{1 - \left(\frac{\lambda}{2q}\right)^2} + \frac{n\pi}{2}\right). \tag{40}$$

Substituting this solution to (19), we obtain the kernel in the large-N limit:

$$K(\lambda,\lambda') \equiv \lim_{N\to\infty} K_N(\lambda,\lambda') \propto \frac{\sin\left(N\int_{\lambda'}^{\lambda} d\lambda\, A(\lambda)\sqrt{1 - \left(\frac{\lambda}{2q}\right)^2}\right)}{\lambda - \lambda'}, \tag{41}$$

$$\rho(\lambda) = K(\lambda,\lambda) = \text{const.}A(\lambda)\sqrt{1 - \left(\frac{\lambda}{2q}\right)^2}. \tag{42}$$

The constant is not fixed within this approach, though it can fixed to be $1/\pi$ if we use the explicit form of the **P** operator (24). The meanings of q and $A(\lambda)$ become clear at this stage: $\pm 2q$ stand for the edges of the spectrum and $A(\lambda)$ the deviation of the spectral envelope from Wigner's semi-circle.

Now we proceed to take the microscopic limit:

$$N \to \infty, \ \lambda \to 0, \ z \equiv N\lambda : \text{fixed}. \tag{43}$$

We have already assumed the single-valuedness of the recursion coefficients q_n, which is known to be the sufficient condition for $A(0) = \pi\rho(0) > 0$. (In other words, we have assumed the chiral symmetry breaking as an input.) Then eq.(37) is reduced to

$$\left[\frac{1}{A(0)^2}\left(\frac{d}{dz} + \frac{(-)^n\alpha}{z}\right)\left(\frac{d}{dz} - \frac{(-)^n\alpha}{z}\right) + 1\right]\psi_n = 0. \tag{44}$$

Its solution which is regular at $z = 0$ is a Bessel function:

$$\psi_n \propto \sqrt{z}\, J_{\alpha - \frac{(-1)^n}{2}}(A(0)z). \tag{45}$$

[§] This process was initiated in the context of stochastic quantization of matrix models[21], although their Fokker-Planck equation differs from eq.(37). I thank C. F. Kristjansen for remarks on this point.

Substituting the above solution again into (19), we obtain universal forms of the microscopic kernels (called the Bessel kernels[22])

$$K_s(z,z') \equiv \lim_{N\to\infty} \frac{1}{N} K_N\left(\frac{z}{N}, \frac{z'}{N}\right), \tag{46}$$

$$\chi\text{UE}: \quad K_s(z,z') = \pi\rho(0)\sqrt{z\,z'}\frac{zJ_{\alpha+1}(2\pi\rho(0)z)J_\alpha(2\pi\rho(0)z') - (z \leftrightarrow z')}{z^2 - z'^2}, \tag{47}$$

$$\text{UE}: \quad K_s(z,z') = \frac{\pi\rho(0)}{2}\sqrt{z\,z'}\frac{J_{\alpha+\frac{1}{2}}(\pi\rho(0)z)J_{\alpha-\frac{1}{2}}(\pi\rho(0)z') - (z \leftrightarrow z')}{z - z'}. \tag{48}$$

and microscopic spectral densities $\rho_s(z) = K_s(z, z)$,

$$\chi\text{UE}: \quad \rho_s(z) = (\pi\rho(0))^2 |z| \left(J_\alpha^2 - J_{\alpha+1}J_{\alpha-1}\right)(2\pi\rho(0)z), \tag{49}$$

$$\text{UE}: \quad \rho_s(z) = \left(\frac{\pi\rho(0)}{2}\right)^2 z\left(J_{\alpha+\frac{1}{2}}^2 + J_{\alpha-\frac{1}{2}}^2 - J_{\alpha+\frac{1}{2}}J_{\alpha-\frac{3}{2}} - J_{\alpha-\frac{1}{2}}J_{\alpha+\frac{3}{2}}\right)(\pi\rho(0)z). \tag{50}$$

In the above, the integration constants are fixed by requiring $\lim_{z\to\infty} \rho_s(z) = \rho(0)$. When measured in the unit of mean level spacing $1/\rho(0)$, $K_s(z, z')\,dz'$ contains no free parameter. After this rescaling, the Bessel kernels (47) and (48) approach the sine kernel (2) in the limit $z, z' \to \infty$, $z - z' = O(1)$ as the repulsive effects due to the fermion determinant and the chiral structure become negligible.

DISCUSSIONS

In this article I have presented a proof of universality of microscopic correlations for RMT modeling $QCD_{3,4}$ within a single-trace potential class. The spectral correlation function is shown to be insensitive to the choice of matrix measures as long as the macroscopic spectrum is supported on a single interval.

Universality of the Bessel kernel is valid even beyond large-N matrix models. Namely, the zero-dimensional reduction of the SU(N_f) σ-model of pions also yields sum rules[23] which are identical to the moments of (49). Together with the numerical agreements mentioned in the introduction, it is very convincing that this wide range of universality encompasses QCD_4.

On the other hand, one may address a question: whether RMT can describe chiral symmetric phases of QCD such as Higgs phase or the large N_f case. In these cases, accumulation of the soft eigenvalues is not strong enough to form a chiral condensate, although the theory is still interacting. There is no satisfactory answer to it so far, yet one may check the existence of universality within RMT. If the microscopic universality in the vicinity of the origin is broken for random matrix ensembles with $\rho(0) \propto |\langle\bar{\psi}\psi\rangle| = 0$, we may not expect to extract informations from RMT. In view of the above, we are lead to analyze a RMT with a double-well potential such as

$$Z = \int_{N\times N \text{ matrices}} dM\, \mathrm{e}^{-N\,\mathrm{tr}(-M^2+gM^4)}, \quad g = \frac{1}{4} \tag{51}$$

whose coupling is tuned so as the macroscopic spectrum $\rho(\lambda)$ to have a double zero at $\lambda = 0$ [24]. According to (42) it corresponds to $A(\lambda)$ of the form

$$A(\lambda) \sim \lambda^2 + O(\lambda^4). \tag{52}$$

Then eq.(37) suggests a possible scaling limit:

$$N \to \infty,\ \lambda \to 0,\ z \equiv N^{1/3}\lambda : \text{fixed}. \tag{53}$$

However, the whole procedure must be reconsidered since the recursion coefficients q_n become double-valued in the large-n and N limit. A working hypothesis is to assume the following asymptotic form[25]

$$q_n = q_c + N^{-1/3}(-)^n f(t) + N^{-2/3} g(t) + O(N^{-1}), \quad t \equiv N^{2/3}\left(1 - \frac{n}{N}\right). \tag{54}$$

Then the function $A(\lambda)$ in the limit (53) is modified by a constant term involving $f'(0)$ and $g(0)$, which are the same order $O(N^{-2/3})$ as λ^2 under the limit (53),

$$A(\lambda) = q_n(-2q_c^2 + q_n^2 + q_{n+1}^2 + \lambda^2) \sim N^{-2/3}(2g(0) - (-)^n f'(0) + z^2). \tag{55}$$

The numerical values of $f'(0)$ and $g(0)$ are determined by solving Painlevé II equation (the microscopic limit of the Heisenberg algebra [**Q**, **P**] = 1) under an appropriate boundary condition[25]. We may also have g run toward $g_c = 1/4$ with

$$\gamma = N^{2/3}(g - g_c) : \text{fixed} \tag{56}$$

while retaining the macroscopic spectrum intact. In this case the functions $f(t)$ and $g(t)$ ought to be evaluated at $t = \gamma$ instead of $t = 0$. These numerical values enter the resulting second order linear differential equation for the wave function ψ_n (the counterpart of (44)). Thus the macroscopic spectral density $\rho(\lambda)$ is not sufficient to determine the microscopic correlation, but the latter depends upon the way how to approach the critical point. Microscopic universality holds in a weak sense for these cases, for Painlevé II equation is universally derived for a class of the critical potentials. Details of the analysis will appear elsewhere[26].

I thank M. Praszalowicz and the organizers of Workshop for their hospitality. Thanks are also due to P. H. Damgaard and G. Akemann for collaboration and discussions summarized in this article. The work of SMN is supported in part by JSPS Postdoctoral Fellowships for Research Abroad, and Nishina Memorial Foundation.

REFERENCES

1. M. L. Mehta, "Random Matrices", 2nd Ed. (Academic Press, New York, 1991). For a recent review, see: T. Guhr, A. Müller-Groeling, H. A. Weidenmüller, cond-mat/9707301.
2. G. Szegö, "Orthogonal Polynomials" (Am. Math. Soc., Providence, 1939).
3. E. Brézin, A. Zee, *Nucl. Phys.* B402:613 (1993).
4. G. Moore, *Prog. Theor. Phys. Suppl.* 102:255 (1990).
5. E. V. Shuryak, J. J. M. Verbaarschot, *Nucl. Phys.* A560:306 (1993). For an updated review, see: J. J. M. Verbaarschot, this volume.
6. T. Banks, A. Casher, *Nucl. Phys.* B169:103 (1980).
7. J. J. M. Verbaarschot, *Nucl. Phys.* B427:534 (1994).
8. J. J. M. Verbaarschot, *Phys. Lett.* B368:137 (1996).
9. M. E. Berbenni-Bitsch, S. Meyer, A. Schäfer, J. J. M. Verbaarschot, T. Wettig, hep-lat/9704018.
10. A. D. Jackson, M. K. Sener, J. J. M. Verbaarschot, *Nucl. Phys.* B479:707 (1996).
11. J. Jurkiewicz, M. A. Nowak, I. Zahed, *Nucl. Phys.* B478:605 (1996).
12. E. Brézin, S. Hikami, A. Zee, *Nucl. Phys.* B464:411 (1996).
13. S. Nishigaki, *Phys. Lett.* B387:139 (1996).
14. G. Akemann, P. H. Damgaard, U. Magnea, S. Nishigaki, *Nucl. Phys.* B487:721 (1997).
15. P. Di Francesco, P. Ginsparg, J. Zinn-Justin, *Phys. Rep.* 254:1 (1995).

16. E. Kanzieper, V. Freilikher, *Phys. Rev. Lett.* 78:3806 (1997).
17. E. Kanzieper, V. Freilikher, cond-mat/9704149.
18. J. J. M. Verbaarschot, I. Zahed, *Phys. Rev. Lett.* 73:2288 (1994).
19. T. R. Morris, *Nucl. Phys.* B356:703 (1991).
20. G. M. Cicuta, L. Molinari, E. Montaldi, *Mod. Phys. Lett.* A1:125 (1986).
21. J. Ambjørn, C. F. Kristjansen, *Int. J. Mod. Phys.* A8:1259 (1993).
22. T. Nagao, K. Slevin, *J. Math. Phys.* 34:2075 (1993); P. J. Forrester, *Nucl. Phys.* B402:709 (1993).
23. H. Leutwyler, A. Smilga, *Phys. Rev.* D46:5607 (1992).
24. Y. Shimamune, *Phys. Lett.* B108:407 (1982).
25. Č. Crnković, G. Moore, *Phys. Lett.* B257:322 (1991).
26. G. Akemann, P. H. Damgaard, S. M. Nishigaki, in preparation.

This page intentionally left blank

NEW DEVELOPMENTS IN NON-HERMITIAN RANDOM MATRIX MODELS *

Romuald A. Janik,[1] Maciej A. Nowak,[1,2,3] Gábor Papp [2,4] and Ismail Zahed[5]

[1] Department of Physics, Jagellonian University, 30-059 Kraków, Poland
[2] GSI, Planckstr. 1, D-64291 Darmstadt, Germany
[3] Institut für Kernphysik, TH Darmstadt, D-64289 Darmstadt, Germany
[4] Institute for Theoretical Physics, Eötvös University, Budapest, Hungary
[5] Department of Physics, SUNY, Stony Brook, NY 11794, USA

INTRODUCTION

Random matrix models provide an interesting framework for modeling a number of physical phenomena, with applications ranging from atomic physics to quantum gravity [1,2]. In recent years, non-hermitian random matrix models have become increasingly important in a number of quantum problems [3,4]. A variety of methods have been devised to calculate with random matrix models. Most prominent perhaps are the Schwinger-Dyson approach [2] and the supersymmetric method[5]. In the case of Non-hermitian Random Matrix Models (NHRMM) some of the standard techniques fail or are awkward.

In this talk we go over several new developments regarding the techniques[6,7] for a large class of non-hermitian matrix models with unitary randomness (complex random numbers). In particular, we discuss

(a) - A diagrammatic approach based on a $1/N$ expansion
(b) - A generalization of the addition theorem (R-transformation)
(c) - A conformal transformation on the position of pertinent singularities
(d) - A 'phase' analysis using appropriate partition functions
(e) - A number of two-point functions and the issue of universality.

Throughout, we will rely on two standard examples: a non-hermitian Gaussian random matrix model (Ginibre-Girko ensemble[8]), and a chiral Gaussian random matrix model in the presence of a constant non-hermitian part example will allow for a comparison of our methods to more conventional ones, the second ensemble will show the versatility of our approach to new problems with some emphasis on the physics issues. Further applications will be briefly mentioned.

* Talk presented by MAN at the NATO Workshop "New Developments in Quantum Field Theory", June 14-20, 1997, Zakopane, Poland.

New Developments in Quantum Field Theory
Edited by Damgaard and Jurkiewicz, Plenum Press, New York, 1998

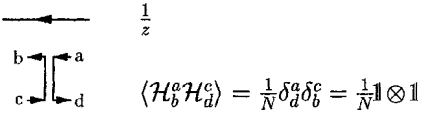


Figure 1. Large N "Feynman" rules for "quark" and "gluon" propagators.

DIAGRAMMATIC EXPANSION AND SPONTANEOUS BREAKDOWN OF HOLOMORPHY

The fundamental problem in random matrix theories is to find the distribution of eigenvalues in the large N (size of the matrix M) limit. According to standard arguments, the eigenvalue distribution is easy to reconstruct from the discontinuities of the Green's function

$$G(z) = \frac{1}{N}\left\langle \mathrm{Tr}\,\frac{1}{z-\mathcal{M}}\right\rangle \tag{1}$$

where averaging is done over the ensemble of $N \times N$ random matrices generated with probability

$$P(\mathcal{M}) = \frac{1}{Z}e^{-N\mathrm{Tr}V(\mathcal{M})}. \tag{2}$$

To illustrate our diagrammatic arguments let us first consider the well known case of a random hermitian ensemble with Gaussian distribution.

Hermitian diagrammatics

We use the diagrammatic notation introduced by [10], borrowing on the standard large N diagrammatics for QCD [11]. Consider the partition function

$$Z = <\det(z-H)> = \int d\psi d\psi^{\dagger} dH e^{-\mathcal{L}} e^{-\frac{N}{2}\mathrm{Tr}H^2} \tag{3}$$

with a "quark" Lagrangian L

$$\mathcal{L} = \bar{\psi}_a(z\mathbf{1}^b_a - H^b_a)\psi^b\,, \tag{4}$$

where H is a hermitian random matrix with Gaussian weight (the width of the Gaussian we set to 1). We will refer to ψ as a "quark" and to H as a "gluon". The "Feynman graphs" following from (4) allow only for the flow of "color" (no momentum), since (4) defines a field-theory in 0 + 0 dimensions. The names "quarks", "gluons", "color" *etc.* are used here in a figurative sense, without any connection to QCD. The "quark" and "gluon" propagators (double line notation) are shown in Fig. 1.

Introducing the irreducible self energy Σ, the Green's function reads

$$G(z) = \frac{1}{N}\mathrm{Tr}\,\frac{1}{z-\Sigma} = \frac{1}{z-\Sigma}\,. \tag{5}$$

In the large N limit the equation for the self energy Σ follows from resumming the rainbow diagrams of Fig. 2. All other diagrams (non-planar and "quark" loops) are subleading in the large N limit. The consistency equation ("Schwinger-Dyson" equation of Fig. 3) reads

$$\Sigma = G. \tag{6}$$

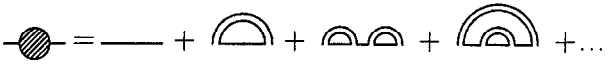


Figure 2. Diagrammatic expansion of the Green's function (1) for Gaussian randomness.

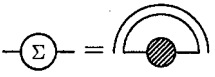


Figure 3. Schwinger-Dyson equation.

Equations (5) and (6) give immediately $G(z - G) = 1$, so the normalizable solution for the Green's function reads

$$G(z) = \frac{1}{2}(z - \sqrt{z^2 - 4}) \tag{7}$$

which, via the discontinuity (cut) leads to Wigner's semicircle for the distribution of the eigenvalues for hermitian random matrices

$$\nu(\lambda) = \frac{1}{2\pi}\sqrt{4 - \lambda^2}. \tag{8}$$

Non-hermitian diagrammatics

If we were to use non-hermitian matrices in the resolvent (1), then configuration by configuration, the resolvent displays poles that are scattered around $z = 0$ in the complex z-plane. In the large N limit, the poles accumulate in general on finite surfaces (for unitary matrices on circles), over which the resolvent is no longer holomorphic. The (spontaneous) breaking of holomorphic symmetry follows from the large N limit. As a result $\partial G/\partial \bar{z} \neq 0$ on the nonholomorphic surface, with a finite eigenvalue distribution. In this section we will set up the diagrammatic rules for investigating non-hermitian random matrix models. In addition to the "quarks" we introduce "conjugate quarks", defined by the 0 + 0 dimensional Lagrangian

$$\mathcal{L}_0 = \bar{\psi}(z - \mathcal{M})\psi + \bar{\phi}(\bar{z} - \mathcal{M}^\dagger)\phi. \tag{9}$$

For hermitian matrices, "quarks" ψ and "conjugate-quarks" ϕ decouple in the "thermodynamical" limit ($N \to \infty$). Their respective resolvents follow from (9) and do not 'talk' to each other. They are holomorphic (anti-holomorphic) functions modulo cuts on the real axis. For non-hermitian matrices, this is not the case in the large N limit. The spontaneous breaking of holomorphic symmetry in the large N limit may be probed in the z-plane by adding to (9)

$$\mathcal{L}_B = \lambda(\bar{\psi}\phi + \bar{\phi}\psi) \tag{10}$$

in the limit $\lambda \to 0$. The combination $L_0 + L_B$ will be used below as the non-hermitian analog of the Lagrangian (4).

From (9,10) we define the matrix-valued resolvent through

$$\hat{\mathcal{G}} = \begin{pmatrix} \mathcal{G}_{qq} & \mathcal{G}_{q\bar{q}} \\ \mathcal{G}_{\bar{q}q} & \mathcal{G}_{\bar{q}\bar{q}} \end{pmatrix} = \left\langle \begin{pmatrix} z - \mathcal{M} & \lambda \\ \lambda & \bar{z} - \mathcal{M}^\dagger \end{pmatrix}^{-1} \right\rangle \tag{11}$$

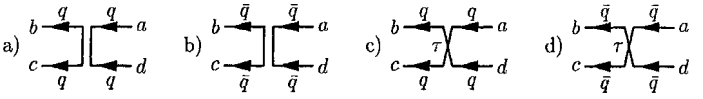


Figure 4. All "gluonic" amplitudes for complex Gaussian matrices.

where the limit $N \to \infty$ is understood before $\lambda \to 0$. The "quark" spectral density follows from Gauss law [13],

$$\nu(z,\bar{z}) = \frac{1}{\pi}\partial_{\bar{z}}\, G(z,\bar{z}) = \frac{1}{\pi N}\partial_{\bar{z}}\, \mathrm{Tr}_N\, \mathcal{G}_{qq} \tag{12}$$

which is the distribution of eigenvalues of M. For hermitian M, (12) is valued on the real axis. As $\lambda \to 0$, the block-structure decouples, and we are left with the original resolvent. For $z \to +i0$, the latter is just a measurement of the real eigenvalue distribution, as shown before in the case of Gaussian hermitian ensemble. For non-hermitian M, as $\lambda \to 0$, the block structure does not decouple, leading to a nonholomorphic resolvent for certain two-dimensional domains on the z-plane. Holomorphic separability of (9) is spontaneously broken in the large N limit. For more technical details we refer to the original work [7]. Similar construction has been proposed recently in [12].

Examples

Consider first the Ginibre-Girko ensemble, *i.e.* the case of complex matrices with the measure

$$< \ldots >= \int [d\mathcal{M}] \exp\left[-\frac{N}{(1-\tau^2)}\mathrm{tr}(\mathcal{M}\mathcal{M}^\dagger - \tau \mathrm{Re}\mathcal{M}\mathcal{M})\right] . \tag{13}$$

The "gluon" propagators read

$$\langle|\mathcal{M}_{ab}|^2\rangle = \frac{1}{N} \qquad \langle\mathcal{M}_{ab}\mathcal{M}_{ba}\rangle = \frac{\tau}{N} \tag{14}$$

corresponding to hermitian ($\tau = 1$), anti-hermitian ($\tau = -1$) or general complex ($\tau = 0$) matrix theory.

From (9,10) we note that there are two kinds of "quark" propagators ($1/z$ for "quarks" ψ and $1/\bar{z}$ for "conjugate-quarks" ϕ, where both can be incoming and outgoing). The relevant "gluonic" amplitudes correspond now to Fig. 4a–4d, where the (c,d) contribution corresponds to twisting the lines with a "penalty factor" τ.

The equation for the one particle irreducible (1PI) self-energy follows from Figs. 4–5 in the form

$$\begin{aligned}\begin{pmatrix} \Sigma_1 & \Sigma_2 \\ \Sigma_3 & \Sigma_4 \end{pmatrix} &= \frac{1}{N}\mathrm{tr}_N \begin{pmatrix} \mathcal{G}_{qq} & \mathcal{G}_{q\bar{q}} \\ \mathcal{G}_{\bar{q}q} & \mathcal{G}_{\bar{q}\bar{q}} \end{pmatrix} \circ \begin{pmatrix} \tau & 1 \\ 1 & \tau \end{pmatrix} \\ &= \frac{1}{N}\mathrm{tr}_N \begin{pmatrix} z-\Sigma_1 & \lambda-\Sigma_2 \\ \lambda-\Sigma_3 & \bar{z}-\Sigma_4 \end{pmatrix}^{-1} \circ \begin{pmatrix} \tau & 1 \\ 1 & \tau \end{pmatrix} .\end{aligned} \tag{15}$$

Here the trace is meant component-wise (block per block), and the argument of the trace is the dressed propagator. The operation $\circ$ is *not* a matrix multiplication, but a simple multiplication between the entries in the corresponding positions. Here tr_N is short for the trace on the $N \times N$ block-matrices.

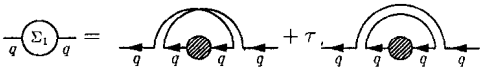


Figure 5. Self-energy equation for non-hermitian matrices.

Each of the entries has a diagrammatical interpretation, in analogy to the hermitian case. For example, the equality of the upper left corners of the matrices in (15) is represented diagrammatically in Fig. 5. The first graph on the r.h.s. in this figure does not influence the "quark-quark" interaction - it corresponds to the double line with a twist, hence non-planar, and therefore subleading. However, this twist could be compensated by the twisted part of the propagator coming from the second correlator (14), thereby explaining the factor τ in the upper left corner of (15). Other entries in (15) follow from Fig. 4 by inspection.

The "quark" one-point function is now

$$G(z,\bar{z}) = \frac{1}{N}\mathrm{tr}_N\, \mathcal{G}_{qq} = (\bar{z} - \Sigma_4)/det\,. \tag{16}$$

It follows that $\Sigma_2 = \Sigma_3$, with

$$det \cdot \Sigma_1 = \tau(\bar{z} - \Sigma_4) \tag{17}$$

$$det \cdot \Sigma_4 = \tau(z - \Sigma_1) \tag{18}$$

$$det \cdot \Sigma_2 = \Sigma_2 - \lambda\,, \tag{19}$$

where $det = (z - \Sigma_1)(\bar{z} - \Sigma_4) - (\lambda - \Sigma_2)^2$. Substituting $r = \Sigma_2 - \lambda$ in the last relation in (19) yields the equation

$$((z - \Sigma_1)(\bar{z} - \Sigma_4) - r^2)(r + \lambda) = r\,. \tag{20}$$

For $\lambda = 0$, the solution with $r = 0$ is holomorphic while that with $r \neq 0$ is nonholomorphic. In the holomorphic case, $\Sigma_1 (z - \Sigma_1) = \tau$, and the resolvent is simply

$$G(z) = \frac{z \mp \sqrt{z^2 - 4\tau}}{2\tau} \tag{21}$$

where the upper sign corresponds to the solution with the pertinent asymptotics. In the nonholomorphic case, $G(z, \bar{z}) = \bar{z} - \Sigma_4$, with

$$G(z,\bar{z}) = \frac{\bar{z} - \tau z}{1 - \tau^2} \tag{22}$$

in agreement with [13]. The boundary between the holomorphic and nonholomorphic solution follows from the condition $\Sigma_2 = 0$ imposed for the nonholomorphic solution, here this is equivalent to $|G(z,\bar{z})|^2 = |G(z)|^2 = 1$, that is

$$\frac{x^2}{(1+\tau)^2} + \frac{y^2}{(1-\tau)^2} = 1 \tag{23}$$

which is an ellipse in the complex plane. Inside (23) the solution is nonholomorphic and outside it is holomorphic. The case investigated by Ginibre [8] follows for $\tau = 0$. It is pedagogical to compare our method of solving this problem to the one coming from interpreting the Ginibre ensemble as a two-matrix ($H \sim H_1 + iH_2$) model.

As a second example, we consider the Chiral Random Matrix model, which got recently some attention as a schematic model for spontaneous breakdown of the chiral

symmetry. Here we consider for simplicity the Gaussian version of such model in the presence of a non-hermitian part, "chemical potential" μ, as suggested by Stephanov[9]. The non-hermitian character comes from the property of Dirac matrices in Euclidean space. The form of the determinant stems from the constant mode sector of the massless and chiral Dirac operator at finite chemical potential [14]. The corresponding partition function reads

$$Z = < \det(z - \mu\gamma - \mathcal{M}) > \tag{24}$$

where

$$\gamma = i\gamma_0 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \qquad \text{and} \qquad \mathcal{M} = \begin{pmatrix} 0 & H \\ H^\dagger & 0 \end{pmatrix}. \tag{25}$$

The only novel features come from the "chiral character" of the matrix M, *i.e.* the fact that it anticommutes with the $\gamma_5 \equiv \mathrm{diag}(\mathbf{1}_N, -\mathbf{1}_N)$. Due to this fact, the "gluon" propagator D inherits the block structure which in the tensor notation (see Fig. 1) reads

$$\mathcal{D} = \frac{1}{N}(\gamma_+ \otimes \gamma_- + \gamma_- \otimes \gamma_+) \tag{26}$$

with $\gamma_\pm = (\mathbf{1}_N \pm \gamma_5)/2$ and the bare "quark" propagator $1/z$ gets modified to $(z-\mu\gamma)^{-1}$. As a result, the 1PI self-energy equations in the planar approximation are given by

$$\begin{pmatrix} \Sigma_1 & \Sigma_2 \\ \Sigma_3 & \Sigma_4 \end{pmatrix} = \frac{1}{N}\mathrm{tr}_N \underbrace{\begin{pmatrix} z - \mu\gamma - \Sigma_1 & \lambda - \Sigma_2 \\ \lambda - \Sigma_3 & \bar{z} + \mu\gamma - \Sigma_4 \end{pmatrix}^{-1}}_{\hat{\mathcal{G}}} \circ \begin{pmatrix} \mathcal{D} & \mathcal{D} \\ \mathcal{D} & \mathcal{D} \end{pmatrix} \tag{27}$$

where D is the "gluon" propagator (26), and Σ_i are diagonal $2N \times 2N$ matrices. Inverting in (27) with respect to the "quark-conjugate quark" indices gives, after some elementary algebra, two kinds of solutions:

(i) A nonholomorphic solution ($\Sigma_2 = \Sigma_3 \neq 0$) ("quark-quark" resolvent)

$$G(z,\bar{z}) = \frac{1}{2N}\mathrm{tr}_N\, \mathcal{G}_{qq} = \frac{x}{2} - iy - \frac{1}{2}\frac{iy}{y^2 - \mu^2} \tag{28}$$

where $z = x + iy$, a result first derived in [9] using different arguments.

(ii) For $\Sigma_2 = \Sigma_3 = 0$ we recover the holomorphic solution [9,15], $\Sigma_1(z) = G(z)\mathbf{1}$, $\Sigma_4 = \Sigma_1^\dagger$, with $G(z)$ fulfilling the cubic Pastur-like equation

$$G^3(z) - 2zG^2(z) + (z^2 + \mu^2 + 1)G(z) - z = 0. \tag{29}$$

Note that in the case of this example the standard techniques of multi-matrix models do not apply.

ADDITION LAWS

The concept of addition law for hermitian ensembles was introduced in the seminal work by Voiculescu[16]. In brief, Voiculescu proposed the additive transformation (R transformation), which linearizes the convolution of non-commutative matrices, alike the logarithm of the Fourier transformation for the convolution of arbitrary functions. This method is an important shortcut to obtain the equations for the Green's functions for a sum of matrices, starting from the knowledge of the Green's functions of individual

ensembles of matrices. This formalism was reinterpreted diagrammatically by Zee[17], who introduced the concept of Blue's function. Let us consider the problem of finding the Green's function of a sum of two independent (free[16]) random matrices M_1 and M_2, provided we know the Green's functions of each of them separately. First, we note that the 1PI self-energy Σ can be always expressed as a function of G itself and *not* of z as usually done in the textbooks. For the Gaussian randomness, $\Sigma_H(G) = G$ (see (6)). Second, we note that the graphs contributing to the self-energy $\Sigma_{1+2}(G)$ split into two classes, belonging to $\Sigma_1(G)$ and $\Sigma_2(G)$, due to the independence of probabilities $P(M_1)$ and $P(M_2)$ and large N (planar) limit. Therefore

$$\Sigma_{1+2}(G) = \Sigma_1(G) + \Sigma_1(G). \tag{30}$$

Note that such a formula is not true if the energies are expressed as functions of z. Voiculescu R transformation is nothing but $R(u) \equiv \Sigma[G(u)]$. The addition (30) reads, for an arbitrary complex u, $R_{1+2}(u) = R_1(u) + R_2(u)$. The R operation forms and abelian group. The Blue's function, introduced by Zee[17], is simply

$$B(G) = \Sigma(G) + G^{-1}. \tag{31}$$

Therefore, using the identity $G(z) = (z - \Sigma)^{-1}$, we see that the Blue's function is the functional inverse of the Green's function

$$B[G(z)] = z \tag{32}$$

and the addition law for Blue's functions reads

$$B_{1+2}(z) = B_1(z) + B_2(z) - \frac{1}{z}. \tag{33}$$

The algorithm of addition is now surprisingly simple: Knowing G_1 and G_2, we find (32) B_1 and B_2. Then we find the sum B_{1+2} using (33), and finally, get the answer G_{1+2}, by reapplying (32). Note that the method treats on equal footing the Gaussian and non-Gaussian ensembles, provided that the measures P_1 and P_2 are independent (free).

The naive extension of this algorithm fails completely for the non-hermitian matrices. It is not *a priori* puzzling - the underlying mathematical reason for (33) is the holomorphy of the hermitian Green's function, not fulfilled for the case of NHRMM, as demonstrated in the previous section. However, since we managed to extend the diagrammatical analysis to the NHRMM, it is still possible to generalize the addition formula using the parallel diagrammatic reasoning like in the hermitian case. The generalization amounts to consider the matrix-valued Green's function (11). The generalized Blue's function [6,7] is now a matrix valued function of a 2×2 matrix variable defined by

$$\mathcal{B}(\mathcal{G}) = \mathcal{Z} = \begin{pmatrix} z & \lambda \\ \lambda & \bar{z} \end{pmatrix}. \tag{34}$$

where λ will be eventually set to zero. This is equivalent to the definition in terms of the self-energy matrix

$$\mathcal{B}(\mathcal{G}) = \Sigma + \mathcal{G}^{-1} \tag{35}$$

where Σ is a 2×2 self energy matrix expressed as a function of a matrix valued Green's function. The same diagrammatic reasoning as before leads to the addition formula for the self-energies and consequently for the addition law for generalized Blue's functions

$$\mathcal{Z} = \mathcal{B}_1(\mathcal{G}) + \mathcal{B}_2(\mathcal{G}) - \mathcal{G}^{-1}. \tag{36}$$

The power of the addition law for NHRMM stems from the fact that it treats Gaussian and non-Gaussian randomness on the same footing [18].

Example

Let us consider for simplicity complex random Gaussian matrices ($\tau = 0$), which we rewrite as the sum $H_1 + iH_2$, with H_1, H_2 hermitian. The generalized Blue's function for hermitian H_1 follows explicitly from

$$\Sigma^{(1)} \equiv \begin{pmatrix} \Sigma_1^{(1)} & \Sigma_2^{(1)} \\ \Sigma_3^{(1)} & \Sigma_4^{(1)} \end{pmatrix} = \hat{\mathcal{G}} \circ \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \hat{\mathcal{G}} \tag{37}$$

a matrix analog of (6). The generalized Blue's function for anti-hermitian iH_2 follows from

$$\Sigma^{(2)} = \hat{\mathcal{G}} \circ \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \tag{38}$$

where the entries reflect the antihermicity (set $\tau = -1$ in (15)). It is a straightforward exercise to check that the matrix equation (36) with the generalized Blue's functions B_1 and B_2, corresponding to (37) and (38), reproduces the two solutions indicated in (22) and (21) respectively, as well as the equation for the boundary (here the circle) separating them on the z plane.

CONFORMAL MAPPINGS

The existence of the nonholomorphic and as well holomorphic domains in the case of two solutions of NHRMM provides a powerful way to evaluate the supports for the level densities of NHRMM. The envelopes of these supports (supports form two-dimensional islands) can be derived very generally using a conformal transformation that maps the cuts of the hermitian ensemble onto the boundaries of its non-hermitian analog.

Let us consider the case where a Gaussian random and hermitian matrix H is added to an arbitrary matrix M. The addition law says

$$R_{H+M}(u) = R_H(u) + R_M(u) = u + R_M(u). \tag{39}$$

where we have used explicitly that for Gaussian $R_H(u) = u$. Now, if we were to note that in the *holomorphic* domain the R transformation for the Gaussian anti-hermitian ensemble is $R_{iH}(u) = -u$, we read*

$$R_{iH+M}(u) = R_{iH}(u) + R_M(u) = -u + R_M(u). \tag{40}$$

These two equations yield

$$B_{iH+M}(u) = B_{H+M}(u) - 2u \tag{41}$$

where we have used the relation $B(u) = R(u) + 1/u$. Substituting $u \to G_{H+M}(z)$ we can rewrite (41) as

$$B_{iH+M}[G_{H+M}(z)] = z - 2G_{H+M}(z). \tag{42}$$

Let w be a point in the complex plane for which $G_{iH+M}(w) = G_{H+M}(z)$. Then

$$w = z - 2G_{H+M}(z). \tag{43}$$

Equation (43) provides a conformal transformation mapping the *holomorphic* domain of the ensemble $H + M$ (*i.e.* the complex plane z minus cuts) onto the *holomorphic* domains of the ensemble $iH + M$, *i.e.* the complex plane w minus the "islands", defining in this way the support of the eigenvalues.

*Note that anti-hermitian Gaussian nullifies in the holomorphic domain the hermitian Gaussian in the sense of the group property of additive transformation R, i.e. $R_H + R_{iH} = 0$.

Examples

Consider the case of "summing" two Hermitian random Gaussian ensembles, *i.e.* consider the Hamiltonian $H = H_H + gH_H$, where g is some arbitrary coupling. The sum constitutes of course the Gaussian ensemble, and the spectrum follows from the properties of the R function $R_{H+gH}(z) = (1 + g^2)z$, or equivalently, Green's function

$$G(z) = \frac{1}{2(1+g^2)}[z - \sqrt{z^2 - 4(1+g^2)}] \tag{44}$$

i.e. the support of the eigenvalues forms the interval (cut) $\mathcal{I} = [-2\sqrt{1+g^2}, +2\sqrt{1+g^2}]$. According to (43), we can map the interval I onto the boundary delimiting the holomorphic domain of the *non-hermitian* ensemble $H = H_H + i\gamma H_H$, $(g \to i\gamma)$, that is

$$w = \frac{1}{1+g^2}[g^2 z + \sqrt{z^2 - 4(1+g^2)}] \tag{45}$$

with $z = t \pm i0$ and t in I. Equation (45) spans an ellipsis with axes $2/\sqrt{1+g^2}$ and $2g^2/\sqrt{1+g^2}$. For $g^2 = 1$ the ellipsis is just the Ginibre's circle.

A similar construction and an identical mapping (43) gives the support of the eigenvalues in the case of a schematic chiral Dirac operator with chemical potential. Let us first consider the case when $\mu = i\epsilon$, *i.e.* the case when the ensemble is hermitian. In this case, the resulting Green's function is known to fulfill the so-called Pastur equation (random Gaussian plus deterministic hermitian Hamiltonian E, here with $N/2$ levels ϵ and $N/2$ levels $-\epsilon$)

$$G(z) = \frac{1}{z - G(z) - \epsilon} + \frac{1}{z - G(z) + \epsilon} \tag{46}$$

encountered in many areas of physics[19]. This is exactly equation (29) with the formal replacement $\mu^2 \to -\epsilon^2$. At a particular value of the deterministic parameter $\epsilon = 1$, the single cut supporting the spectrum of the hermitian ensemble splits into two-arc support, manifesting therefore a structural change in the spectral properties, hence a "phase transition". The spectral properties of the non-hermitian model, with chemical potential μ, follow from the mapping (43), but with G_{H+M} replaced by an appropriate branch of the cubic Pastur equation (46). In particular, at the value $\mu^2 = 1$ the spectrum demonstrates the structural change - an island splits into two disconnected mirror islands (see Fig.6).

TWO-POINT FUNCTIONS

To probe the character of the correlations between the eigenvalues of non-hermitian random matrices, either on their holomorphic or nonholomorphic supports, it is relevant to investigate two-point functions. For example, a measure of the breaking of holomorphic symmetry in the eigenvalue distribution is given by the connected two-point function or correlator

$$N^2 G_c(z, \bar{z}) = \langle \left| \mathrm{tr} \frac{1}{z - \mathcal{M}} \right|^2 \rangle_c \tag{47}$$

where the z and $\bar{z}$ content of the averaging is probed simultaneously. The correlation function (47) will be shown to diverge precisely on the nonholomorphic support of the

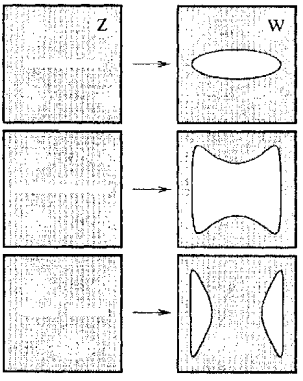


Figure 6. Conformal mappings for the case of Ginibre-Girko ensemble with $\tau = 0.5$ (upper), non-hermitian chiral ensemble with chemical potential $\mu^2 = 0.8$ (middle) and $\mu^2 = 1.2$ (lower). The shaded regions represent the holomorphic domains.

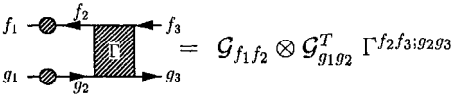


Figure 7. Two-point kernel with $f, g = q, \bar{q}$.

eigenvalue distribution, indicating an accumulation in the eigenvalue density. In the conventional language of "quarks" and "gluons" , (47) is just the correlation function between "quarks" and their "conjugates". A divergence in (47) in the z-plane reflects large fluctuations between the eigenvalues of the non-hermitian operators on finite z-supports, hence their "condensation".

It was shown in [20] and [21] that for hermitian matrices (with $\bar{z} \to w$) the fluctuations in connected and smoothened two-point functions satisfy the general lore of macroscopic universality. This means that all smoothened correlation functions are universal and could be classified by the support of the spectral densities, independently of the specifics of the random ensemble and genera in the topological expansion (see [22] for a recent discussion).

In the case of NHRMM the generalized two-point correlator reads [7]

$$N^2 G_c(z, w) = -\partial_z \partial_w \log \det(1 - \mathcal{G}_1 \otimes \mathcal{G}_2^T \Gamma) . \tag{48}$$

Here the logarithm is understood as a power series expansion. Equation (48) is valid for Gaussian ensembles and, in the general case, up to factorizable corrections in the sense of [10]. The operator $\mathcal{G}_1 \otimes \mathcal{G}_2^T \Gamma$ is a tensor product of 2×2 matrices (see Fig. 7). The kernel Γ includes the details of the "gluonic" interactions, depending on the particular measure. The tensor structure reflects the nonholomorphic solutions. The choice of "isospin" in the *e.g.* lower fermion line is done by choosing the appropriate derivative ∂_w for the "quark" and $\partial_{\bar{w}}$ for the "conjugate-quark".

For the holomorphic solutions of NHRMM, the structure is simpler, since the Green's functions are holomorphic in this case.

Examples

In the case of the Ginibre Girko ensemble, the two-point correlator in the *holomorphic domain* reads simply

$$C(z,\bar{z}) = -\partial_z \partial_{\bar{z}} \ln[1 - G(z)G(\bar{z})]\,. \tag{49}$$

Indeed, in the holomorphic domain instead of the matrix valued G we use $G(z)$ given by (21), and the kernel Γ reduces to the "quark-conjugate quark" coupling equal to 1 (upper-right corner of the last matrix in (15)). Note that the correlator (49) diverges on the line

$$1 - |G(z)|^2 = 0 \tag{50}$$

therefore the ellipse (23), confirming our statement.

For the nonholomorphic domain, the calculation is a bit more involved, due to the explicit matrix structure of G and Γ. The explicit form of the matrix-valued resolvent is

$$\mathcal{G} = \frac{1}{1-\tau^2}\begin{pmatrix} \bar{z} - \tau z & g_z \\ g_z & z - \tau\bar{z} \end{pmatrix} \tag{51}$$

with $g_z^2 = |z|^2(1-\tau)^2 - \tau(z+\bar{z})^2 - (1-\tau^2)^2$. One recognizes (22) as the upper left corner of (51). The explicit form of the kernel is

$$\Gamma = \mathrm{diag}(\tau, 1, 1, \tau) \tag{52}$$

which corresponds to all possible contributions from four graphs on Fig. 4. After some algebra, the determinant of 4 by 4 matrix $(1 - \mathcal{G}_1 \otimes \mathcal{G}_2^T\Gamma)$ turns out to be equal to $|z - w|^2$, giving the correlator

$$N^2 G_{qq}(z,w) \doteq -\frac{1}{(w-z)^2}\,. \tag{53}$$

In the non-hermitian chiral case the correlator in the *outside* (holomorphic) region is calculated using the same arguments as above. The only minor technical complication stems from the chiral (block) nature of the Green's functions. We skip the details published elsewhere [7, 15], showing only the final result. The determinant of $1 - \mathcal{G}_1 \otimes \mathcal{G}_2^T\Gamma$ gives

$$\frac{(D-\mu^2)^2 - |z-G|^4}{D^2} \tag{54}$$

where the holomorphic G is the appropriate branch of (29) and $D = |(z - G)/G|^2$.

The zero of the determinant in (48) occurs for $(D - \mu^2) = |z - G|^2$, that is

$$|z-G|^2(1-|G|^2) - \mu^2|G|^2 = 0 \tag{55}$$

as quoted in[15]. This is exactly the equation of the boundary separating the holomorphic and nonholomorphic solutions, obtained in examples before either as $\Sigma_2 = 0$ or as a result of conformal mapping.

In the hermitian case $\mu = 0$ (and $\bar{z} = w$), the determinant in (54) is simply $(1 - G^2(z)G^2(w))$ (chiral) as opposed to $(1 - G(z)G(w))$ (non-chiral). As a result, for $w = z$ and $\mu = 0$, (54) is

$$N^2 G(z,z) = \frac{1}{z^2(z^2-4)^2} \tag{56}$$

which coincides with (5.5) in [23].

Note that the form of (56) signals two kinds of *microscopic universalities*. The $1/N$ expansion breaks down at $z = \pm 2$ (endpoints of spectra) and $z = 0$ ("Goldstone" pole due to the "chiral" nature of the ensemble).

The divergence at $z = \pm 2$ points at the edge universal behavior of the spectral function (Airy oscillations) [24], the divergence at $z = 0$ signals the chiral microscopic universality [25]. Unfolding the spectra at these points allows to get the explicit universal kernels characterizing the fore mentioned universalities. In light of these remarks, it is tempting to speculate [7], that the divergences of the generalized correlators signal the onset some *new types* of microscopic universalities present in the NHRMM.

Note also that the relations (50) and (54) show that the two-point correlator depends solely on the one-point holomorphic Green's function, extending the macroscopic universality argument to NHRMM as well. What this means geometrically is an open question.

Before closing this example let us present for completeness the result for the chiral correlator in the nonholomorphic domain. The calculation is a bit tedious, due to the fact that in the nonholomorphic region "quarks" may turn into "conjugate-quarks" and vice-versa, with all "quark" species interacting with themselves, and species appear in chiral copies. Nevertheless, the final result for the determinant is remarkably simple:

$$\det(1-\mathcal{G}(z)\otimes\mathcal{G}^T(w)\Gamma) = |z-w|^2|z+w|^2\frac{(\mu^2-(\mu^2-y^2)(\mu^2-v^2))^2-v^2y^2}{\mu^4} \tag{57}$$

(where $y = \mathrm{Im}\, z$, $v = \mathrm{Im}\, w$), suggesting perhaps the possibility of further technical developments in the case of NHRMM.

PARTITION FUNCTIONS

We show now that the information carried by the one- and two-point functions is sufficient to specify the "thermodynamical" potential to order $O(1/N)$ in the entire z-plane modulo isolated singularities, as we now discuss. Similar ideas were used in different context in [26, 27, 28].

Let Z_N be the partition function in the presence of an external parameter z. In the $1/N$ approximation, the diagrammatic contributions to the partition function Z_N read

$$\log Z_N = NE_0 + E_1 + \mathcal{O}(\frac{1}{N}) \tag{58}$$

where E_0 is the contribution of the "quark" or "conjugate quark" loop in the planar approximation, and E_1 is contribution of the "quark-quark" loop, and so on, in the same approximation, We will restrict our attention to non-hermitian matrices with unitary randomness, in which case the non-planar corrections to E_0 are of order $1/N^2$. Hence, E_0 is determined by the one-point function and E_1 by the two-point functions.

For z such that (58) is real, continuous and nondecreasing function of the extensive parameters [29], $\log Z_N / N$ may be identified with the "pressure" of the random matrix model. As a result, the isolated singularities in the "pressure" are just the "phase" boundaries provided that the expansion is uniform. Below we give examples where the "phase" boundary is either mean-field-driven or fluctuation-driven. We have to distinguish two cases: holomorphic partition functions ("unquenched") and nonholomorphic partition functions, where the complex phases of the determinants are neglected.

Holomorphic Z

We consider the partition function

$$Z_N = \langle \det(z - \mathcal{M}) \rangle = \left\langle \int d\psi d\bar{\psi} e^{-\bar{\psi}(z-\mathcal{M})\psi} \right\rangle . \tag{59}$$

In contrast to the one- and two-point correlators discussed above, the determinant in (59) is not singular in the z-plane configuration by configuration. Hence, (59) is *a priori* holomorphic in z (minus isolated singularities).

The one- and two-point functions on their holomorphic support may be obtained from log Z_N by differentiation with respect to z. Therefore, from (58)

$$E_0 = \int^z dz' \; G(z') + const \tag{60}$$

or equivalently

$$E_0 = zG - \int dG \; z(G) + const \tag{61}$$

after integrating by part. Note that $z(G) = B(G)$ is just the Blue's function[17] of G. The constant in E_0 is fixed by the asymptotic behavior of (59), that is $Z_N \sim z^N$. The planar contribution to E_1 in (58) follows from the "quark-quark" wheel (two-point correlator). The final result for Z_N is

$$Z_N = e^{NE_0} \cdot \left(\left\{ \det(1 - \mathcal{G} \otimes \mathcal{G}^T \cdot \Gamma) \right\}^{-\frac{1}{2}} + \mathcal{O}\left(\frac{1}{N}\right) \right) . \tag{62}$$

Note that due to the power –1/2, the fluctuations have "bosonic" character, and are dwarfed by the "quark" contribution as $(1 : N)$ [30] . Both E_0 and E_1 are simple functions of the resolvent on a specific branch, as expected from generalized macroscopic universality.

We note that the partition function Z_N through (62) exhibits an essential singularity in $1/N$ as expected from thermodynamical arguments. Assuming that the expansion for log Z_N/N is uniform, then log Z_N/N follows from (62) using the holomorphic resolvent $G(z)$ for large z. The small z region follows by analytical continuation. However, since $G(z)$ is multi-valued (already the simple case of the Ginibre-Girko ensemble yields two branches for the resolvent in (21)), the analytical continuation is ambiguous. The ambiguity may be removed by identifying log $|Z_N|/N$ with some *generalized* "pressure" and taking $G(z)$ so that log $|Z_N|/N$ is maximum. As a result, V_N = log Z_N/N is piece-wise analytic in leading order in $1/N$ with "cusps" at

$$F^{(ij)}(x, y) \equiv V_N^{(i)}(x, y) - V_N^{(j)}(x, y) = 0 , \tag{63}$$

following the transition from branch i to branch j of G.

The character of the transition in the $1/N$ approximation can be highlighted by noting that for any finite N, the partition function (59) is a complex polynomial in z of degree N with random coefficients. In large N,

$$V_N = \frac{1}{N} \log |Z_N| = \frac{1}{2} \int dv \, d\bar{v} \, \varrho(v, \bar{v}) \, \log |z - v|^2 . \tag{64}$$

To leading order, the distribution of singularities along the "cusps" (63) is

$$\varrho(z, \bar{z}) = \frac{1}{2\pi} |\partial_z F|^2 \, \delta(F(z, \bar{z})) \tag{65}$$

which is normalized to 1 in the z-plane. Redefining the density of singularities by unit length along the curve $F(z,\bar{z}) = 0$, we may rewrite (65) as

$$\underline{\varrho(z,\bar{z})|_{F=0} = \frac{1}{2\pi}|\partial_z F| \equiv \frac{1}{4\pi}|G^{(i)} - G^{(j)}|\,.} \tag{66}$$

For $\varrho \neq 0$, the integrand in (64) is singular at $z = v$ which results into different forms for V_N, hence a cusp. For $\varrho = 0$, that is $\partial_z F = 0$, V_N is differentiable. For physical V_N (real and monotonically increasing), the points $\varrho = 0$ are multi-critical points. At these points all n-point ($n \geq 2$) functions diverge. This observation also holds for Ising models with complex external parameters[31]. Assuming macroscopic universality [20] for all n-points ($n \geq 2$), we conclude that $\partial_z F = 0$ means a branching point for the resolvents, hence $\partial_z G = \infty$ or $B'(G) = 0$[17]. For hermitian matrices, these conditions coincide with the end-points of the eigenvalue distributions[17, 32].

Examples

To illustrate the above concepts, consider again first the Ginibre-Girko ensemble. The resolvent in the holomorphic region satisfies (21), so

$$z = \tau G + \frac{1}{G}\,. \tag{67}$$

The integration (61) in E_0 is straightforward, and after fixing the asymptotic behavior we obtain

$$Z_N = G^{-N} e^{\frac{\tau}{2}NG^2} \left((1 - G^2(z)\tau)^{-\frac{1}{2}} + \mathcal{O}\left(\frac{1}{N}\right)\right)\,. \tag{68}$$

Here G is the solution of (67). The pre-exponent in (68) follows from (62) with the matrix G replaced by G and $\Gamma = \tau$, as seen in the "quark-quark" component of the vertex matrix in (15). Using (67) we observe that the pre-exponent diverges at two points in the z-plane, $z^2 = 4\tau$. At these points there is a "phase" change as we now show.

Given (68), the generalized "pressure" in leading order is

$$V_\pm = -\frac{1}{2}\log(G_\pm \overline{G}_\pm) + \frac{\tau}{4}(G_\pm^2 + \overline{G}_\pm^2) + \mathcal{O}\left(\frac{1}{N}\right)\,. \tag{69}$$

$V_\pm$ define two intersecting surfaces valued in the z-plane, for two branches $G_\pm$ of the solutions to (67). The parametric equation for the intersecting curve is

$$F(z,\bar{z}) = V_+ - V_- = 0. \tag{70}$$

As indicated above, V_N is piece-wise differentiable. Note that $F = 0$ on the cut along the real axis, $-2\sqrt{\tau} < z < +2\sqrt{\tau}$, and from (66) the density of singularities per unit length is

$$\varrho(z,\bar{z})|_{F=0} = \frac{1}{\pi}\frac{\sqrt{4\tau - z^2}}{2\tau}\,. \tag{71}$$

Along F, the density of singularities is semi-circle. The density (71) vanishes at the end-points $z = \pm 2\sqrt{\tau}$. This is easily seen to be the same as $\partial_z G = \infty$, or $dB(G)/dG = 0$ with $B(G) = \tau G + 1/G$. As noted above the term in bracket in Eq. (68) vanishes at these points, with a diverging "quark-quark" contribution. The transition is fluctuation-driven. These points may again signal the onset of scaling regions with possible universal microscopic behavior for non-hermitian random matrix models. This issue will be pursued elsewhere. At these points the $1/N$ expansion we have used breaks down.

Let us move now to the chiral non-hermitian ensemble. Elementary integration for this case leads to

$$Z_N(z,\mu) = e^{NE_0} \cdot \left(\left\{ D^{-2}[(D+\mu^2)^2 - (z-G)^4] \right\}^{-\frac{1}{2}} + \mathcal{O}\left(\frac{1}{N}\right) \right), \tag{72}$$

but now $D = (z-G)^2/G^2$, and

$$E_0(z,\zeta) = G^2 + \log\frac{z-G}{G} \tag{73}$$

with the appropriate branch of holomorphic G solution to (29).

Note that for $z = 0$ and $G^2 = -1 - \mu^2$, the pre-exponent in (72) diverges. It also diverges at $z = z_*$ which are the zeros of (66) for the present case (two zeros for small μ and four zeros for large μ), see Fig. 8. Again, at these points, the $1/N$ expansion breaks down marking the onset of scaling regions and the possibility of microscopic universality. The $z = 0$ divergence is just the notorious "Goldstone" mode in chiral models, illustrating the noncommutativity of $N \to \infty$ and $z \to 0$. The rest of the arguments follow easily from the preceding example, in agreement with the "thermodynamics" discussed in[30]. The analytical results for the nature and location of singularities of this example were confirmed by an extensive numerical analysis of Yang-Lee zeroes (up to 500 digits accuracy) in [33].

Nonholomorphic Z

The above analysis for the holomorphic thermodynamical potential may also be extended to nonholomorphic partition functions of the type

$$Z_N[z,\bar{z}] = \langle \det|z-\mathcal{M}|^2 \rangle = \left\langle \int d\psi d\bar{\psi} d\phi d\bar{\phi} e^{-\bar{\psi}(z-\mathcal{M})\psi - \bar{\phi}(\bar{z}-\mathcal{M}^\dagger)\phi} \right\rangle . \tag{74}$$

Note the important "quenching" of the phase of the determinant in comparison to the holomorphic case (59). As a consequence, the two-point correlators diverge rather on the one-dimensional boundary separating the phases (when approaching the boundary from the holomorphic domain) then at discrete points.

Similar reasoning as before leads to the explicit expression

$$Z_N[z,\bar{z}] = e^{NE_0} \cdot \left(\left\{ \det(1 - \mathcal{G} \otimes \bar{\mathcal{G}}^T \cdot \Gamma) \, |\det(1 - \mathcal{G} \otimes \mathcal{G}^T \cdot \Gamma)| \right\}^{-1} + \mathcal{O}\left(\frac{1}{N}\right) \right) \tag{75}$$

where E_0 comes from the solutions of

$$G(z,\bar{z}) = \frac{\partial E_0(z,\bar{z})}{\partial z} \qquad \bar{G}(z,\bar{z}) = \frac{\partial E_0(z,\bar{z})}{\partial \bar{z}} . \tag{76}$$

Note that again the contributions from the two-wheel diagrams are of the form $1/\sqrt{\det}$ and hence "bosonic" in character. The result could be easily guessed without performing the calculations: there are two contributions from the "quark-conjugate-quark" wheels (correlators) (square of the $1/\sqrt{\det}$ in first term in the curly bracket) one contribution from the "quark-quark" wheel and one contribution from the "conjugate-quark-conjugate-quark" wheel (represented together as a second (modulus) term in the curly bracket).

Again, the partition function Z_N has an essential singularity in $1/N$, but $\log Z_N/N$ does not. For any finite N, the latter diverges for

$$\det(1 - \mathcal{G} \otimes \bar{\mathcal{G}}^T \cdot \Gamma) = 0 \tag{77}$$

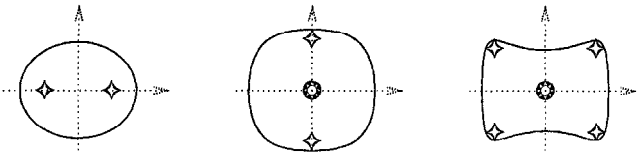


Figure 8. Critical lines (77) and critical points (78) for Ginibre-Girko ensemble with $\tau = 0.1$ (left) and non-hermitian chiral ensemble with small ($\mu^2 = 0.1$, middle) and large ($\mu^2 = 0.6$, right) value of chemical potential. The solid lines and the symbols represent the manifold on which the $1/N$ expansion breaks down and signal the location of possible microscopic universal behavior for NHRMM.

which defines the *line* of singularities, and for

$$\det(1 - \mathcal{G} \otimes \mathcal{G}^T \cdot \Gamma) = 0 \tag{78}$$

defining the *set of discrete points*, encountered in the case of the holomorphic partition function.

Examples

For the Ginibre-Girko example, the line of singularities (77) reads

$$1 - |G|^2 = 0\,. \tag{79}$$

The line of singularities (79) reproduces in this case the ellipse (23). The ellipse includes the points of the "phase" change (see (68)),

$$1 - \tau G^2(z) = 0 \tag{80}$$

i.e. the focal points $z^2 = 4\tau$, corresponding to (78), connected by the interval (70), *i. e.* $F = 0$.

In the case of the chiral non-hermitian random model the condition (77) reads [15]

$$D^{-2}[(D - \mu^2)^2 - |z - G|^4] = 0 \tag{81}$$

with $D = |(z - G)/G|^2$, therefore exactly the condition (55). This line represents the boundary between the holomorphic and nonholomorphic solutions. The set of discrete multi-critical points, corresponding to (78) is given by the condition

$$D^{-2}[(D + \mu^2)^2 - (z - G)^4] = 0 \tag{82}$$

but with $D = [(z - G)/G]^2$, in agreement with (72). Note the appearance of the modulus and the flip in the sign of μ when comparing the last two formulae. The explicit solution of (82) consists of two or four points, (depending on the value of μ). These points are Airy-type end-point singularities, with $z = 0$ being a distinct multi-critical point reflecting the chiral nature of the ensemble.

Figure 8 shows the critical lines and critical points corresponding to the conditions (77,78) for Ginibre-Girko and non-hermitian chiral ensembles. The end-points singularities are denoted by "NATO stars", and the chiral singularity – by "Zakopane sun".

The fact that the critical line in Fig. 8 b,c surrounds the multi-critical points of the unquenched partition function, explains the failure of quenched lattice calculations with finite baryonic potential. The nature of the chiral restoration is masked by unphysical fluctuations caused by neglecting the phase of the determinant. The critical line (81) exactly reproduces the shoreline along the islands of "mixed-condensate", in agreement with the replica method [9].

CONCLUSIONS

Most of the results presented here have already appeared in a number of papers [6, 7, 15] , and we refer the reader for further details. In our presentation, we tried to highlight the various relations between the hermitian and nonhermitian ensembles of random matrices, thereby enhancing the universality of some of the concepts. In particular, we described several ways for deriving the distribution of eigenvalues and their supports in the nonhermitian case either diagrammatically, or through the concepts of generalized Blue's functions and conformal mappings. We also indicated a number of approaches for the evaluation of smoothened (wide) correlators. The two canonical examples we have used throughout for illustration, allow us to see clearly when the standard methods work and when they do not. In the latter case, we have provided for the new alternatives.

In processing some of the examples, we have speculated on the possibility of a new type of macroscopic universality for NHRMM, as well as the onset of new forms of microscopic universality at pertinent critical points of the two-point correlators. This last issue we suspect, is of great interest to new developments in the context of "weak" non-hermiticity in random matrix models[34], and related applications[35]. We hope to return to it in the future.

Acknowledgments

This work was partially supported by the US DOE grant DE-FG-88ER40388, by the Polish Government Project (KBN) grant 2P03004412 and by Hungarian grants OTKA T022931 and FKFP0126/1997. MAN thanks the organizers of the Workshop for an opportunity to present this review and acknowledges interesting discussions with J. Ambjørn, J. Jurkiewicz, S. Nishigaki and J. Zinn-Justin.

REFERENCES

1. For a recent review, see T. Guhr, A. Müller-Gröling and H.A. Weidenmüller, *e-print* cond-mat/9707301.
2. P. Di Francesco, P. Ginsparg and J. Zinn-Justin, *Phys. Rept.* 254:1 (1995), and references therein.
3. Y. V. Fyodorov and H.-J. Sommers, *J. Math. Phys.* 38:1918 (1997), and references therein.
4. E. Brézin and A. Zee, *e-print* cond-mat/9708029, and references therein.
5. K.B. Efetov, *Adv. in Phys.* 32:53 (1983); J.J.M. Verbaarschot, H.A. Weidenmüller and M. R. Zirnbauer, *Phys. Rep.* 129:367 (1985).
6. R.A. Janik, M.A. Nowak, G. Papp, J. Wambach and I. Zahed, *Phys. Rev.* E55:4100 (1997)
7. R.A. Janik, M.A. Nowak, G. Papp and I. Zahed, *Nucl. Phys.* B501:603 (1997).
8. J. Ginibre, *J. Math. Phys.* 6:440 (1965); see also V.L. Girko, "Spectral Theory of Random Matrices" (in Russian), Nauka, Moscow, (1988).
9. M. Stephanov, *Phys. Rev. Lett.* 76:4472 (1996); M. Stephanov, *Nucl. Phys. Proc. Suppl.* 53:469 (1997).
10. E. Brézin and A. Zee, *Phys. Rev.* E49:2588 (1994); E. Brézin and A. Zee, *Nucl. Phys.* B453:531 (1995); E. Brézin, S. Hikami and A. Zee, *Phys. Rev.* E51:5442 (1995).
11. G. 't Hooft, *Nucl. Phys.* B75:464 (1974).
12. J. Feinberg and A. Zee, *e-print* cond-mat/9703087.
13. H.-J. Sommers, A. Crisanti, H. Sompolinsky and Y. Stein, *Phys. Rev. Lett.* 60:1895 (1988).
14. I. Barbour et al., *Nucl. Phys.* B275:296 (1996); C.T.H. Davies and E.G. Klepfish, *Phys. Lett.* B256:68 (1991); J.B. Kogut, M.P. Lombardo and D.K. Sinclair, *Phys. Rev.* D51:1282 (1995).
15. R.A. Janik, M.A. Nowak, G. Papp and I. Zahed, *Phys. Rev. Lett.* 77:4876 (1996).
16. D.V. Voiculescu, *Invent. Math.* 104:201 (1991); D.V. Voiculescu, K.J. Dykema and A. Nica, "Free Random Variables", Am. Math. Soc., Providence, RI, (1992); for new results see also A. Nica and R. Speicher, *Amer. J. Math.* 118:799 (1996) and references therein.
17. A. Zee, *Nucl. Phys.* B474:726 (1996).

18. J. Feinberg and A. Zee, *e-print* cond-mat/9704191.
19. L.A. Pastur, *Theor. Mat. Phys. (USSR)* 10:67 (1972); F. Wegner, *Phys. Rev*. B19:783 (1979); P. Neu and R. Speicher, *J. Phys.* A28:L79 (1995); E. Brézin, S. Hikami and A. Zee, *Phys. Rev.* E51:5442 (1995); A. D. Jackson and J.J.M. Verbaarschot, *Phys. Rev.* D53:7223 (1996); T. Wettig, A. Schäfer and H.A. Weidenmüller, *Phys. Lett.* B367:28 (1996); J. Jurkiewicz, M.A. Nowak and I. Zahed, *Nucl. Phys*. B478:605 (1996); M. Engelhardt, *Nucl. Phys.* B481:479 (1996).
20. J. Ambjørn, J. Jurkiewicz and Yu.M. Makeenko, *Phys. Lett.* B251:517 (1990).
21. E. Brézin and A. Zee, *Nucl. Phys.* B402:613 (1993).
22. G. Akemann and J. Ambjørn, *J. Phys.* A29:L555 (1996); G. Akemann, *Nucl. Phys.* B482:403 (1996); G. Akemann, *e-print* hep-th/9702005.
23. J. Ambjørn, C.F. Kristjansen and Yu.M. Makeenko, *Mod. Phys. Lett.* A7:3187 (1992).
24. M.J. Bowick and E. Brézin, *Phys. Lett.* B268:21 (1991).
25. E.V. Shuryak and J.J.M. Verbaarschot, *Nucl. Phys.* A560:306 (1993); J.J.M. Verbaarschot and I. Zahed, *Phys. Rev. Lett.* 70:3852 (1993); for recent results, see G. Akemann, P.H. Damgaard, U. Magnea and S. Nishigaki, *Nucl. Phys.* B487:721 (1997).
26. L. McLerran and A. Sen, *Phys. Rev.* D32:2795 (1985); T.H. Hansson and I. Zahed, *Phys. Lett.* B309:385 (1993); J.V. Steele, A. Subramanian and I. Zahed, *Nucl. Phys.* B452:545 (1995); J.V. Steele and I. Zahed, unpublished.
27. C. Itoi, *Nucl. Phys.* B493:651 (1997).
28. J. Ambjørn, L. Chekhov, C.F. Kristjansen and Yu. Makeenko, *Nucl. Phys.* B404:127 (1993).
29. K. Huang, "Statistical Mechanics", John Wiley, New York, (1987).
30. M. A. Nowak, M. Rho and I. Zahed, "Chiral Nuclear Dynamics", World Scientific, Singapore, (1996); R.A. Janik, M. A. Nowak, G. Papp and I. Zahed, *Acta Phys. Pol.* B27:3271 (1996); R.A. Janik, M. A. Nowak and I. Zahed, *Phys. Lett.* B392:155 (1997).
31. R. Shrock, private communication; G. Marchesini and R. Shrock, *Nucl. Phys.* B318:541 (1989); V. Matveev and R. Shrock, *J. Phys.* A28:1557 (1995); *Nucl. Phys. Proc. Suppl.* B42:776 (1995).
32. M.A. Nowak, G. Papp and I. Zahed, *Phys. Lett.* B389:137 (1996);
33. A. Halasz, A.D. Jackson and J.J.M. Verbaarschot, *Phys. Lett.* B395:293 (1997); A. Halasz,A.D. Jackson and J.J.M. Verbaarschot, *e-print* cond-mat/9703006.
34. Y.V. Fyodorov, B.A. Khoruzhenko and H.-J. Sommers, *Phys. Lett.* A226:46 (1997), and references therein; K. B. Efetov, *e-print* cond-mat/9702091.
35. N. Hatano and D.R. Nelson, *Phys. Rev. Lett.* 77:570 (1997); N. Hatano and D.R. Nelson, *e-print* cond-mat/9705290; J.T. Chalker and Z. Jane Wang, *e-print* cond-mat/9704198; J. Feinberg and A.Zee, *e-print* cond-mat/9706218; R.A. Janik, M.A. Nowak, G. Papp and I. Zahed, *e-print* cond-mat/9705098; P.W. Brouwer; P.G. Silvestrov and C.W.J. Beenakker, *e-print* cond-mat/9705186; I.Y. Goldsheid and B.A. Khoruzhenko, *e-print* cond-mat/9707230; see also[4].

POTENTIAL TOPOGRAPHY AND MASS GENERATION

Maxime Kudinov*, Enrique Moreno† and Peter Orland‡

Baruch College
The City University of New York
New York, NY 10010
USA

INTRODUCTION

We describe an approach to understanding exponential decay of correlation functions in asymptotically free theories. This approach is systematic; it does not start from any conjectured mechanism or picture. We begin by studying

- the metric on the space of configurations and
- the behavior of the potential-energy function on this space.

We begin by describing how these ideas fit in the framework of QCD, as discussed earlier[1]. We then consider the 1 + 1-dimensional $O(2)$ and $O(3)$ nonlinear sigma models and show that no gap exists in the former at weak coupling. In the $O(3)$ model a new kind of strong/weak-coupling duality is realized. We then briefly outline our proposals for understanding the spectrum.

THE YANG-MILLS METRIC

In the last few decades, there have been many serious attempts to understand QCD wave functionals on orbit space by isolating a fundamental region (the interior of the Gribov horizon)[2]. Instead we examine this space using "automorphic functions", i.e. gauge-invariant wave functionals[1, 3]. An interesting approach along similar lines for 2 + 1-dimensional gauge theories is that of reference[4].

The Hamiltonian of the D + 1-dimensional $SU(N)$ Yang-Mills theory in $A_0 = 0$ (temporal) gauge

$$H = \int_M d^D x \left[-\frac{e_0^2}{2} \, tr \, \frac{\delta^2}{\delta A_j(x)^2} + \frac{1}{4e_0^2} \, tr \, F_{jk}(x)^2 \right] .$$

* Also at: Queens College, The City University of New York, Flushing, NY 11367
† Also at: The City College of the City University of New York, New York, NY 10031
‡ Also at: The Graduate School and University Center, The City University of New York, New York, NY 10036

New Developments in Quantum Field Theory
Edited by Damgaard and Jurkiewicz, Plenum Press, New York, 1998

The allowed wave functionals Ψ satisfy the condition that if A and B are physically equivalent (same up to gauge transformations of C-S number zero) $\Psi[A] = \Psi[B]$.

This will be reformulated as a particle on a certain infinite-dimensional curved space on which there is a height function, namely the potential energy. We are interested ultimately in how geodesic motion (strong couping) is influenced by this height function (potential topography).

The lattice discussion will be used here to introduce the Yang-Mills metric. A more heuristic motivation following Feynman[5] was used in reference[1].

Consider a lattice gauge theory with D space and 1 time dimension. Label discrete points in Euclidean space-time by x and t, respectively. Let $\{U(t)\}$ denote the set of lattice gauge fields (in the fundamental representation of $SU(N)$) on links pointing in space directions at a particular time. Split the action S into a space-time plaquette term S_{st} and a space-space plaquette term S_{ss}, i.e. $S = S_{st} + S_{ss}$, where

$$S_{st} = -\sum_t \frac{1}{2e_0^2}\Re[\{U(t)\},\{U(t+1)\}]^2\,, \quad S_{ss} = -\sum_t \frac{1}{4e_0^2}\ell[\{U(t)\}]\,,$$

where

$$\Re[\{U(t)\},\{U(t+1)\}]^2 = \frac{1}{2}\sum_{x,i} tr[U_i(x,t)U_0(x+\hat{i},t)U_i(x+\hat{t},t)^\dagger U_0(x,t)^\dagger] + c.c.\,,$$

and

$$\ell[\{U(t)\}] = \sum_{x,i<j} tr[U_i(x,t)U_j(x+\hat{i},t)U_i(x+\hat{j},t)^\dagger U_j(x,t)^\dagger] + c.c.\,.$$

Now let's try to integrate out the links pointing in time direction, $U_0(x,t)$. As a first approximation to doing this, we can just solve their equation of motion. This says that $\Re[\{U(t)\},\{U(t+1)\}]^2$ is minimized with respect to these degrees of freedom. If we integrate them out explicitly the result is a product of a Bessel functions. Near the maximum of this product it has the form $\exp -\sum_t \rho[\{U(t)\},\{U(t+1)\}]$., where r is the *absolute* (not local) minimum

$$\rho[\{U(t)\},\{U(t+1)\}] = \min_{\{U_0(t)\}} \Re[\{U(t)\},\{U(t+1)\}]\,.$$

The quantity $\rho[\{U(t)\},\{U(t+1)\}]$ can be shown to be a metric on the space $\{U(t)\}$ modulo gauge transformations. Thus the kinetic term in the action by itself describes Brownian motion in this space.

In the continuum. the space of connections U is defined to contain only those gauge fields which are Lebesgue measurable, and square-integrable[1]. No distinction is made between gauge fields which are the same almost everywhere (U is a Hilbert space). Gauge transformations are $SU(N)$ valued functions $g(x)$ which are differentiable and for which $ig^{-1}\partial g \in U$. Any element of U is mapped into another element of U by such a gauge transformation.

The equivalence classes must actually be made larger in order to obtain a metric space M_D. Two vectors in U with representatives A and B will be said to be gauge-equivalent if there is a sequence of gauge transformations $g_1, g_2, \ldots$, such that

$$B = \lim_n A^{g_n}$$

in the usual metric of the Hilbert space U.

Let α and β be two physical configurations, with A a representative of α and B a representative of β. The distance function [5,6] is defined by

$$\rho[\alpha,\beta] = \inf_{h,f}\{\sqrt{\frac{1}{2}\int_M d^D x\ tr\ [A^h(x) - B^f(x)]^2}\}\ .$$

This is just the continuum version of the lattice metric defined above. The function $\rho[\cdot,\cdot]$ was shown to give a complete metric space on equivalence classes of gauge connections M_D.

There is a local metric on the space of connections. This turns out to be essentially that discussed some time ago [7,8]. The Laplacian actually contains several terms not found by these authors [1,9]. The geodesics of the space can be proved to be those conjectured by Babelon and Viallet [8]. With the metric tensor defined properly, there are no non-generic points as had been claimed (the orbit space is complete). We believe the problem some people found with non-generic points is related to the fact that they worked with connections in Sobolev space rather than those in U in which case there is no longer completenss using the metric $\rho[\cdot,\cdot]$.

The square of the infinitesmal distance in orbit space M_D due to a small displacement δA in U is

$$d\rho^2 = [\int_M d^D x \sum_{j=1}^{D}\sum_{a=1}^{N^2-1}][\int_M d^D y \sum_{k=1}^{D}\sum_{b-1}^{N^2-1}]\ G_{(x,j,a)\,(y,k,b)}\ \delta A^a_j(x)\delta A^b_k(y)\ ,$$

where the metric tensor is

$$G_{(x,j,a)\,(y,k,b)} = \delta_{j\,k}\delta_{a\,b}\delta^D(x-y) - (\mathcal{D}_j\ \mathcal{P}\frac{1}{\mathcal{D}^2}\ \mathcal{D}_k)_{a\,b}\ \delta^D(x-y)\ ,$$

and where in the Green's function $\mathcal{P}\frac{1}{\mathcal{D}^2}$, the principle value projects out the zero modes of D^2. Reducible connections[10] are a set of measure zero.

This metric tensor has zero eigenvalues; the dimension of the coordinate space U is larger than the dimension of the orbit space. One must either gauge fix (and deal with the Gribov problem by prescribing a fundamental domain) or develop methods for Riemannian geometry for metric tensors with zero modes. We follow the latter path. The Laplacian was first found by Schwinger on the basis of relativistic invariance and further discussed by Gawedzki [9]. It was constructed in reference[1] using a theory of tensors applicable when the dimension of coordinate space is greater than the dimension of the manifold. It is

$$\Delta\Psi[\alpha] = -\partial_Y(G^{Y\,U}\partial_U\Psi[\alpha]) + (\partial_Z G^Z_Y)G^{Y\,U}\partial_U\Psi[\alpha]\ ,$$

where capital Roman letters denote "indices" $X = (x, j, a)$ and $\partial_X \equiv \frac{\delta}{\delta A^a_j(x)}$.

The Yang-Mills Hamiltonian is

$$H = \frac{e^2}{2}\Delta + \frac{e^2\mathcal{R}}{12} + \int_M d^D x\ \frac{1}{4e^2}\ tr\ F_{j\,k}(x)^2\ ,$$

where R denotes the ultraviolet divergent scalar curvature.

STRUCTURE OF THE YANG-MILLS POTENTIAL ENERGY - RIVER VALLEYS AND GLUONS

The metric properties of the manifold M_D of configurations determine the spectrum of the kinetic term of the Hamiltonian. To understand the spectrum at weak coupling, the potential or magnetic energy must be examined.

A natural starting point is to try make a relief map of magnetic energy on M_D; in other words to investigate the magnetic topography [1].

Suppose that the manifold of physical space is very large. Make a rescaling of the coordinates and the connection $A \in U$ by a real factor s:

$$A_j(x) \longrightarrow sA_j(sx) .$$

A gauge-transformed connection A^h will transform the same way under a rescaling, provided $h(x)$ is redefined by

$$h(x) \longrightarrow h(sx) .$$

The distance of the point of orbit space α from an equivalence class of pure gauges, α_0, transforms as

$$\rho[\alpha, \alpha_0] \longrightarrow s^{\frac{2-D}{2}} \rho[\alpha, \alpha_0] .$$

Let $A \in U$ be a particular configuration of finite potential energy, for which the magnetic field $F_{jk}(x)$ decays rapidly to zero for x outside some finite bounded region, which will be called the **domain** of the magnetic field. By changing the size of the domain and the magnitude of the magnetic field, the distance from some given pure gauge can be made arbitrarily small (except when regularization effects become important) or large (except when volume effects become important).

The potential energy

$$U[\alpha] = \int_M d^D x \, \frac{1}{4e^2} \, tr \, F_{jk}^2(x) ,$$

transforms as

$$U[\alpha] \longrightarrow s^{4-D} U[\alpha]$$

and so for $D > 2$

$$U \sim \rho^{-2\frac{4-D}{D-2}} .$$

For $2 < D < 4$ the exponent is negative. Thus it is possible to have arbitrarily large U for arbitrarily small ρ.

For Abelian gauge theories, other rescalings can be considered;

$$A_j(x) \longrightarrow s^{\phi} A_j(sx) ,$$

where ϕ is any real number. By choosing ϕ satisfying $\frac{D-2}{2} < \phi < \frac{D}{2}$ it is always possible to make the potential energy arbitrarily small for small s, no matter what the dimension. The domain can be made large, while the field strength is made small; a quantum wave functional $\Phi[\alpha]$ whose amplitude is largest near this configuration is a long-wavelength photon. This quantum state must be orthogonal to the vacuum $\Psi_0[\alpha]$ because at least one of the two wave functionals is zero at any point in orbit space. This is why non-compact electrodynamics has no mass gap in any dimension. Our analysis seems to indicate that the same is true for Yang-Mills theory for dimension between 2+1 and 4+1.

Figure 1. illustrates the situation. Orbit space contains regions, which we call **river valleys** in which the potential energy vanishes in the thermodynamic limit. The configurations in the river valleys are not pure-gauge configurations α_0. The river valleys are preserved under scale transformations and are therefore finite-dimensional. In perturbation theory only the region near α_0 is explored. Perturbative gluons are oscillations along straight line extrapolations of these curves.

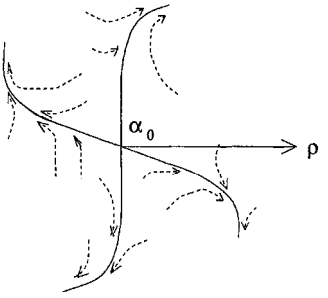


Figure 1. The topography of the Yang-Mills theory in four space-time dimensions. The dashed arrows represent directions of decreasing potential energy along scale transformations. The radius ρ is the distance from the zero potential configuration α_0. The solid curves (river valleys) are where the potential energy vanishes in the thermodynamic limit (only a finite subset is depicted).

OUR STRATEGY

In the light of the above discussion, how could the spectrum of QCD possibly contain anything other that massless particles?

Answer: The regions of small potential energy could have large electric (kinetic) energy by the uncertainty principle. The zero-point energy of the modes transverse to the river valleys must be added to the potential. In this way, the first excited state could have a finite gap above the ground state. There are examples of quantum-mechanical systems, including models motivated by Yang-Mills theory[11], for which this is true [12].

We should view the position along each river valley as a collective coordinate. We then integrate out all degrees of freedom except this coordinate. The resulting quantum-mechanical Hamiltonian will have eigenstates which correspond to the eigenstates of the field-theory Hamiltonian with zero total momentum. To insure consistency of the collective-coordinate approximation, we will must consider only small fluctuations around river valleys.

THE $O(n)$ NONLINEAR SIGMA MODEL

A great deal is known about the 1 + 1-dimensional sigma models. The phase transition in the $O(2)$ model is well understood[13]. By virtue of its integrability [14] the S-matrix [15] and spectrum [16] of the $O(3)$ model are also known. Unfortunately neither these methods nor the $1/n$-expansion [17] extend to gauge theories in higher dimensions.

This model will first be considered in $D + 1$ dimensions. Later we will specialize to $D = 1$. The field $s(x)$ (we are fixing time) with x on a D-dimensional lattice is a real n-tuple with $s^T(x)s(x) = 1$. The Hamiltonian is

$$H = \frac{e_0}{2} \sum_x L(x)^T L(x) - \frac{1}{2e_0} \sum_{<x,x'>} s(x)^T s(x') \ .$$

The fields s lie in equivalence classes:

$$\psi = \{Rs;\ R \in O(n)\}.$$

The definition of these equivalence classes isn't yet obviously right. Unlike the case of Yang-Mills theory, the equivalence class contains physically *different* configurations. We will worry about this issue later.

A natural metric on equivalence classes ψ, ϕ with $s \in \psi$, $f \in \phi$ is [5]

$$\rho[\phi,\psi]^2 = \inf_{R\in O(n)} \sum_x [Rf(x) - s(x)]^T [Rf(x) - s(x)] .$$

In the continuum this goes over to (up to factors of the lattice spacing)

$$\begin{aligned} \rho[\phi,\psi]^2 &= \inf_{R\in O(n)} \int d^D x \, [Rf(x) - s(x)]^T [Rf(x) - s(x)] \\ &= V - tr\sqrt{M^T M} , \end{aligned} \tag{1}$$

where V is now the volume of the space manifold and

$$M_{jk} = \int d^D x \, f_j(x) s_k(x) .$$

However, we no longer strictly have a metric without making certain restrictions on allowed spin configurations. Without such restrictions, different configurations are separated by a distance zero. This is a minor difficulty and will not trouble us.

THE SIGMA-MODEL RIVER VALLEYS

Let's denote the "pure gauge" configuration containg constant $s(x)$ by ψ_0.

Consider now the following problem for $D = 1$. For fixed $\rho[\psi,\psi_0]$ minimize the the potential energy

$$U(\psi) = \int_0^L dx \, s'(x)^2 , \quad s \in \psi .$$

subject to Neumann boundary conditions $s'(0) = s'(L) = 0$. Let's parametrize $s(x)$ using angles $\theta_1(x),..., \theta_{n-1}(x)$, by $s_1(x) = \sin\theta_1(x)... \sin\theta_{n-1}(x)$, ..., $s_n = \cos\theta_1(x)$, in the standard way.

The solution for the minima of $U(\psi)$ for fixed $\rho = \rho[\psi,\psi_0]$ (distance from the origin=pure gauge) is similar to that of a pendulum. We find that up to global rotations R there are minima labeled by an integer $N = 1, 2, ...$

$$\theta_1(x) = \pm\alpha_N(x,k) = \pm 2 \sin^{-1} k \, sn(\frac{2NK}{L}x - K) ,$$

$$\theta_2(x) = \cdots = \theta_{n-1}(x) = 0 ,$$

where $0 \le k \le 1$ is the modulus of the elliptic function $sn(u)$ and $K = K(k) = sn^{-1}(1)$ is the usual complete elliptic integral. The river valleys are nicely parametrized by the modulus k as shown in figure 2.

We find that

$$\rho(k)^2 = \begin{cases} 2L(1 - \frac{E}{K}) , & 0 \le k \le k^* \approx 0.82 \\ 2L\frac{E}{K} , & k^* \le k \le 1 \end{cases} , \tag{2}$$

where $E = E(k) = \int_0^1 dn^2(u)\, du$ is another standard elliptic integral. This function rises smoothly from 0 to L as k goes from 0 to k^*, then falls off to zero again as $k \to 1$. In fact, on the lattice the $k = 1$ solution is unphysical, because this solution has discontinuities in the continuum. Actually $k \le k_{max} \approx 1$ because of the regulator. A configuration along a river valley is maximally far from the origin at $k = k^*$.

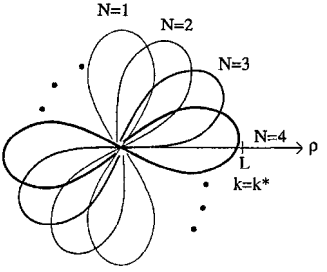


Figure 2. The river valleys for the $O(n)$ sigma model. The potential energy is nearly constant along the solid curves. As before, a straight line extrapolation along tangent vectors at the origin gives the spin wave approximation.

The potential energy function is

$$U(k) = \frac{32N^2K}{L}[E - (1 - k^2)K] \, .$$

For fixed volume L this diverges at $k = 1$, but, as mentioned earlier, this divergence is regularized by a lattice (or some other ultraviolet cut-off).

In the infinite volume limit for $k < k^*$, the potential is a constant; but the one-dimensional domain over which this is so has an infinite length ($=L$). If we view k as a collective variable, and ignore fluctuations in other degrees of freedom the gap is $O\left(\frac{1}{L}\right)$.

We note that the river valleys are not straight lines in configuration space. Their tangent vectors at k are

$$\beta_N(x,k) = \frac{\partial \alpha_N(x,k)}{\partial k} = \frac{2sn(u)dn(u) - Z(u)cn(u)}{1 - k^2} \, ,$$

where $u = 2NKx/L - K$ and $Z(u)$ is the Jacobi zeta function. The inner product of β_N and its derivative with respect to k is not zero; this means that the river valleys are curved. The tangent vector does not have unit length in our collective coordinates. It is convenient to define the unit tangent vector $\hat{\beta}_N(x,k) = \beta_N(x,k)/\sqrt{\int_0^L \beta_N(y,k)^2 \, dy}$.

COLLECTIVE COORDINATES

Up to now we've ignored the fact that a system with global symmetry has states which transform as some representation of that symmetry (For Yang-Mills theories, we have no such problem). For example if $n = 2$ our river valleys are not one-dimensional, but two-dimensional surfaces parametrized by θ_0 as well as k: $\theta(x) = \theta_0 \pm \alpha_N(x, k)$. In fact the river valleys of the $O(n)$ sigma model are really $n - 1$-dimensional manifolds. However this consideration is irrelevant if we are asking for only certain information. The degree of freedom corresponding to θ_0 is the very longest wavelength Goldstone mode. We can remove this mode if we are interested in mass spectra only and don't care about degeneracies of our states. For example, this can be done in the $O(3)$ model by adding a term $-\int dxdt \, \frac{\lambda}{V} s_1^2/(1 - s_2^2)$, where V is the space-time volume. Such a term is of no consequence in the thermodynamic limit but clearly keeps the river valleys one-dimensional, rather than three-dimensional manifolds. This won't matter as long as we aren't interested in the transformation properties or non-accidental degeneracies of our states.

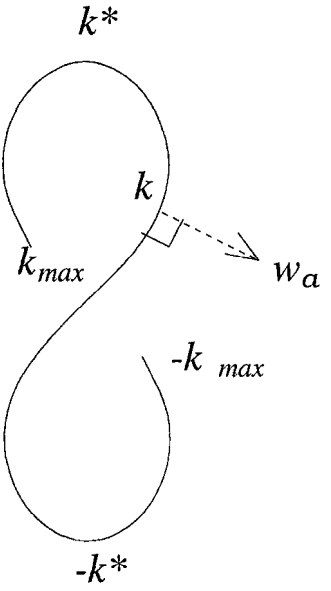


Figure 3. A depiction of the collective coordinate representation. The degree of freedom $k(t)$ parametrizes the river valley, while the $w_a(t)$'s parametrize normal displacements

The collective-coordinate representation[18] of $\theta_1(x)$ is

$$\theta_1(x,t) = \alpha_N(x,k(t)) + \sum_{a=1}^{\infty} w_a(t) T_N^a(x,k(t)) \ ,$$

where we have now taken the range of $k(t)$ to be $-k_{max} < k(t) < k_{max}$ and $\alpha_N(x,-k) \equiv -\alpha_N(x,\ k)$. The family of functions $T_N^a\ (x,\ k)$ satisfy

$$\int_0^L T_N^a(x,k) T_N^b(x,k)\ dx = \delta^{ab} \ , \quad \int_0^L T_N^a(x,k) \hat{\beta}_N(x,k)\ dx = 0 \ ,$$

and $T^{a\,\prime}(0) = T^{a\,\prime}(L) = 0$.

Physically, the collective coordinate k is the parameter along the N^{th} river valley, while the $w_a(x)$'s are coordinates of dispacements normal to the valley at the point described by k.

In order to proceed further, it is necessary to compute the Jacobian $\|J\|$ of the transformation to the collective coordinate system. The details of this computation will be presented elsewhere. The answer turns out to be

$$\|J\| = \sqrt{\int_0^L \beta_N(x,k)^2\, dx} - \sum_{a=1}^{\infty} w_a \int_0^L \frac{\partial \hat{\beta}_N(x,k)}{\partial k} T^a(x,k)\, dx \ .$$

NO GAP FOR THE $O(2)$ SIGMA MODEL

The behavior of the $O(2)$ sigma model at weak coupling is generally regarded as obvious [13]. From our perspective however, this model is nearly as hard to understand as all the others.

The functional measure on $\theta_1 = \theta$ is fairly simple. The fact that it is compact is responsible for vortices which are transitions between river valleys We will ignore these transitions for the following reason. A vortex can be regarded as a process through which a configuration evolves in time; the initial and final configurations are in the same equivalence class. Thus it is a loop in the metric space. The length of this loop can be computed and diverges in the thermodynamic limit. In the spirit of our approach, which is to consider only small fluctuations near river valleys, vortices can therefore be ignored. The details of this elementary calculation will be presented elsewhere.

The only nontrivial factor in the measure is the Jacobian to collective coordinates. The path integral

$$Z = \int \mathcal{D}\theta(x,t) \exp - \int dt \int_0^L dx \frac{1}{2e_0}(\partial_t\theta^2 + \partial_x\theta^2)$$

can be expanded to quadratic order in w_a:

$$Z \approx \int \mathcal{D}k(t)\mathcal{D}w_a(t)[\prod_t \|J(k(t), w_a(t))\|] \exp -\frac{1}{2e_0}\int dt\{[\int_0^L \beta_N(x,k)^2\, dx]\, \dot{k}^2$$

$$+ \sum_a \dot{w}_a^2 + \sum_{ab} w_a \Omega^{ab}(k) w_b + source\ terms\ for\ w_a\}\ ,$$

where the matrix $\Omega(k)$ is the projection of the operator $-\frac{d^2}{dx^2}$ onto the subspace of normalizable functions satisfying Neumann boundary conditions and which are orthogonal to β_n. In other words

$$\Omega^{ab}(k) = \int_0^L \frac{dT^a(x)}{dx}\frac{dT^b(x)}{dx}\, dx\ .$$

The source term can be shown to be unimportant in the limit of infinite time evolution.

Only the first term in $\|J\|$ should be included to one loop. This factor cancels out if a change of variable is made from $k(t)$ to the arc-length parameter $\gamma(t)$ defined by

$$\frac{d\gamma}{dk} = \sqrt{\int_0^L \beta_N(x,k)^2\, dx}\ .$$

We will write $k(t)$ as $k(\gamma(t))$ henceforth. For small k, $\gamma \approx \rho \approx k\sqrt{L}$.

The zero-point-energy contribution from the w_a's is simply half the sum of eigenfrequencies. The square of any one of these eigenfrequencies is an eigenvalue of $\Omega(k(\gamma))$. After integrating out these modes, we are left with

$$Z = \int \mathcal{D}\gamma(t) \exp - \int dt\ [\ \frac{\dot{\gamma}(t)^2}{2e_0} + \frac{1}{2} tr\sqrt{\Omega(k(\gamma(t)))} - \frac{1}{2} tr\sqrt{\Omega(0)}\ ]\ .$$

Computing the eigenvalues of $\Omega(k)$ is done in the following way. The operator

$$\omega(k) = (1 - |\hat{\beta}_N><\hat{\beta}_N|)\ (-\frac{d^2}{dx^2})\ (1 - |\hat{\beta}_N><\hat{\beta}_N|)$$

on the Hilbert space on $[0, L]$ with Neumann boundary conditions has the same eigenvalues as $\Omega(k)$. Here we use standard Dirac bra-ket notation for Hilbert space vectors. The eigenvalue problem for this operator is straightforward (we only present a quick and dirty derivation of the answer here).

Let $G(\lambda)$ be the inverse of $-d^2/dx^2 - \lambda$ on the Hilbert space. The operator $G(\lambda)$ isn't regular, since it has poles if $\lambda = n^2\pi^2/L^2$, but $G(\lambda)\sin\lambda$ is regular. Consider $G(\lambda)\sin\lambda|\hat{\beta}_N>$. If this is is orthogonal to $|\hat{\beta}_N>$, it must be an eigenvector of $w(k)$ with eigenvalue λ. By doing the analysis more carefully one can show that all the eigenvectors are of this form. The condition that λ be an eigenvalue of $\Omega(k)$ is therefore $\sin\lambda < \hat{\beta}_N|G(\lambda)|\hat{\beta}_N> = 0$ or

$$\int_0^L dx \int_0^{L-x} dy \ \cos\lambda x \cos\lambda y \ \hat{\beta}_N(x)\hat{\beta}_N(y) = 0 \ .$$

This function can be computed numerically, the zeros can be found, and the sum of the square roots obtained. A graph of the potential versus k (not γ) has vanishing slope at the origin and rises significantly only for $k \approx k^*$. One can therefore conclude that the gap to the first excited state is of order $1/L$.

Even without doing a very explicit calculation one can see that gap is impossible by a simple scaling argument. Each eigenvalue is directly proportional to $1/L^2$ and the sum of square roots of eigenvalues is finite (after making the subtraction at $k = 0$. Therefore, this sum must have the form

$$\frac{1}{2}\,tr\sqrt{\Omega(k(\gamma))} = \frac{1}{L}f(k(\gamma)) \ .$$

The only way any nontrivial γ dependence could emerge in the thermodynamic limit is if there is a term in $f(k)$ proportional to $1/k^2$, for small k. But then the result at large L would be

$$\frac{1}{2}\,tr\sqrt{\Omega(k(\gamma))} = \frac{C}{\gamma^2} \ ,$$

where C is a constant for $\gamma < \sqrt{L}$.. But this is a Calogero potential which has a continuous spectrum (there is no harmonic oscillator term).

THE $O(3)$ CASE AND THE LAMÉ EQUATION

The $O(3)$ model has two angles $\theta_1 = \theta$ and $\theta_2 = \phi$. Let us consider the lattice path integral

$$Z = [\prod_{x,t}\int_{-\pi/2}^{\pi/2} d\theta(x,t)\,|\sin\theta(x,t)|\,][\prod_{x,t}\int_{-\pi}^{\pi} d\phi(x,t)]$$

$$\times \exp -\sum_{x,t}\frac{1}{2e_0}\{[1-\cos(\theta(x,t+a)-\theta(x,t))]+[1-\cos(\theta(x+a,t)-\theta(x,t))]$$

$$+\sin^2\frac{\theta(x,t+a)+\theta(x,t)}{2}\,[1-\cos(\phi(x,t+a)-\phi(x,t))]$$

$$+\sin^2\frac{\theta(x+a,t)+\theta(x,t)}{2}\,[1-\cos(\phi(x+a,t)-\phi(x,t))]\} \ . \qquad (3)$$

Here x and t are discrete, namely integers times the lattice spacing a (which is assumed to be much smaller than L/N). To calculate the zero-point energy of the fluctuations to one loop, the quantities $\sin^2(\theta(x+a,\ t)+\theta(x,\ t))/2$ in the action can be

replaced by $\sin^2\alpha_N$. Similarly the measure factor $\prod_{x,t}\sin|\theta(x,t)|$ can be replaced by $\prod_{x,t}\sin|\alpha_N(x,k(t))|$. Notice that $\sin^2\alpha_N$ vanishes at $x_j = \frac{(2j-1)L}{2N}$, $j = 1,...,N$. This means that Neumann boundary conditions $\partial_x\phi = 0$ arise at these points; furthermore ϕ can be discontinuous at x_j. The degrees of freedom in the field ϕ do not communicate with one another across the line $x = x_j$. This breaking of the part of the action depending on ϕ into independent pieces in strips separated by the x_j is an artifact of the one-loop approximation. At higher loops, we can no longer assume the coefficients of the ϕ lattice derivatives vanish at these points.

As in the $O(2)$ case, vortex configurations of the θ field are of no importance at weak coupling. The contribution the fluctuations of this field give when integrated out is the same as before.

Consider next the integration over ϕ. If $\gamma << L$ (i.e. $|k| << 1$), then the coefficient of the ϕ lattice derivatives, namely $\sin^2\alpha_N(x, k(t))$ is small over most of spacetime for small k and large over most of spacetime for large k. This function is essentially a slowly-varying inverse coupling constant for ϕ. We therefore expect ϕ-field vortices to be important, and we cannot treat the integral as a Gaussian in ϕ. However, let us come back to this point a little later and see what happens if the Gaussian approximation is used for the ϕ integration.

If we assume that $k(t)$ is slowly varying and ignore its time derivatives, we can absorb the measure factor $[\Pi_{x,t}\sin\alpha_N(x, k(t))]$ into $D\phi(x, t)$, by defining $\Phi = \phi\sin\alpha_N$ and take the continuum limit, obtaining

$$\begin{aligned} Z &= \int \mathcal{D}\theta(x,t)\int \mathcal{D}\Phi(x,t)\exp -\int dt\int_0^L dx\,\frac{1}{2e_0}\{\partial_t\theta^2+\partial_x\theta^2 \\ &\quad + [\partial_t\Phi^2+\sin^2\alpha_N\,(\partial_x\frac{\Phi}{\sin\alpha_N})^2]\} \\ &= \int \mathcal{D}\theta(x,t)\int \mathcal{D}\Phi(x,t)\exp -\int dt\int_0^L dx\,\frac{1}{2e_0}[\partial_t\theta^2+\partial_x\theta^2 \\ &\quad + \Phi\,(-\partial_x^2+\frac{24N^2K^2}{L^2}k^2sn^2(u)-\frac{4N^2K^2}{L^2}(1+4k^2))\,\Phi]\,, \end{aligned} \tag{4}$$

supplemented by Neumann boundary conditions $\partial_x\phi = 0$ at $x = 0$, L and Dirichlet boundary conditions $\Phi = 0$ at the points x_j.

After making the transformation from θ to k, w_1, w_2,... and integrating out both the w_a's and Φ gives

$$\begin{aligned} Z &= \int \mathcal{D}\gamma(t)\exp -\int dt\,[\,\frac{\dot\gamma(t)^2}{2e_0}+\frac{1}{2}tr\sqrt{\Omega(k)}-\frac{1}{2}tr\sqrt{\Omega(0)} \\ &\quad + \frac{2NK}{L}tr\sqrt{\mathcal{L}(k)+(1+4k^2)}-\frac{N\pi}{L}tr\sqrt{\mathcal{L}(0)+1}\,] \end{aligned} \tag{5}$$

where Ω is defined as before and L is the Lamé operator

$$\mathcal{L}(k) = -\frac{d^2}{du^2}+m(m+1)k^2sn^2(u)\,,$$

where $m = 2$ in our case.

The eigenvalue problem $L\Lambda(u) = A\Lambda(u)$ is called the Lamé equation and was first solved in some generality by Hermite (see the book by Whittaker and Watson[19]).

A rough argument shows that the Gaussian approximation for ϕ does not yield a gap between the ground state and the first excited state. Large eigenvalues of the Lamé operator are well approximated by the eigenvalues of the Schrödinger operator

with zero potential. The fourth term in the exponent of (5) is then an ultraviolet-divergent expression of the form

$$S(\Lambda, L, k) = \frac{A}{L} \sum_{n=1}^{\Lambda L} \sqrt{n^2 + f(k)} + o(\frac{1}{L}) ,$$

where A is a constant and $f(k)$ is some function with no x-dependence. The reason the mode sum is cut off at ΛL is because that is the number of degrees of freedom in the problem (for example, on a lattice, where Λ is the inverse lattice spacing). For large ΛL, the sum becomes an integral which can be evaluated to be

$$S(\Lambda, L, k) = \frac{Af(k)^2}{2L}[\sinh^{-1} \frac{\Lambda L}{f(k)} + \frac{1}{2} \sinh(2 \sinh^{-1} \frac{\Lambda L}{f(k)}] + o(\frac{1}{L}) .$$

For large L all that remains is $S(\Lambda, L, k) \approx \frac{A\Lambda^2}{2}$ which has no k-dependence.

STRONG-WEAK COUPLING DUALITY

Let us now look once again at (3). We will make a Gaussian approximation for θ, which we know to be justified, but not ϕ (we can differentiate between θ and ϕ using the arguments in section 6). In order to find the effective action for γ, we need to find the contribution to the potential which is the free energy of the ϕ field with θ set equal to α_N, i.e.

$$W(\gamma) = - \lim_{T \to \infty} \frac{\log Z_\phi(\gamma)}{T} ,$$

where T is the time duration,

$$\begin{aligned} Z_\phi(\gamma) = & [\prod_{x,t} \int_{-\pi}^{\pi} d\phi(x,t) \beta(x,\gamma)^{\frac{1}{2}}] \\ & \times \exp - \sum_{x,t} \beta(x,\gamma) \{[1 - \cos(\phi(x, t+1) - \phi(x,t))] \\ & + [1 - \cos(\phi(x+1, t) - \phi(x,t))]\} , \end{aligned} \tag{6}$$

and

$$\beta(x,\gamma) = \frac{1}{2e_0} \sin^2 \alpha_N(x, k(\gamma)) = \frac{2k(\gamma)^2}{e_0} sn^2(u, k(\gamma))\, dn^2(u, k(\gamma)),$$

plays the role of inverse coupling constant.

What is very striking is that no matter how small the coupling e_o may be, the effective coupling in the "ϕ sector" is large for $\gamma = k^2\sqrt{L} << \sqrt{L}$. This is a kind of strong-coupling/weak-coupling duality. It tells us that to study $W(\gamma)$ at weak coupling, we need a strong-coupling expansion. If there is a minimum of $W(\gamma)$ for finite γ, compactness effects, i.e. *vortices* are responsible. How can this be reconciled with the philosophy of our approximation, namely that only configurations close to the river valleys may be considered? The answer is that, unlike the case of the $O(2)$ sigma model, vortices are *short* paths in configuration space, whose lengths are not divergent. This point will be discussed in a later publication.

We have not yet proved the existence of a gap from the ground state to the first excited state, but it seems clear how the proof should go. First, the strong-coupling expansion will yield the potential $W(\gamma)$ (we have already found this). Then it must be

checked that the gap does not disappear as $L \to \infty$. If this is so, the spatial correlation functions must automatically fall off exponentially; for if the wave function is localized at small k, the effective coupling of the ϕ-field must be strong. This is an important check of Lorentz invariance. Finally the dependence of the gap on e_0 must be checked for consistency with asymptotic freedom.

ACKNOWLEDGEMENTS

We are grateful to Paul Wiegmann for discussions. Several years ago, Michael Aizenman proposed a scheme for proving exponential decay in the $O(3)$ sigma model using vortices associated with a $O(2)$ subgroup. We thank him for describing his ideas to us. P.O. thanks the organizers of the workshop for the opportunity to present this work. The work of M.K. and P.O. was supported in part by PSC-CUNY grants, nos. 6-67438 and 6-68460. The work of E.M. and P.O. was supported in part by a CUNY Collborative Incentive Grant, no. 991999.

REFERENCES

1. P. Orland, Niels Bohr preprint NBI-HE-96-35, **hep-th 9607134.**
2. D. Zwanziger, *Nucl. Phys.* B209:336 (1982); Nucl. Phys. B412:657 (1994); P. Koller and P. van Baal, *Ann. Phys.* 174:299 (1987); *Nucl. Phys.* **B302** (1988) 1; P. van Baal, *Phys. Lett.* **224B** (1989) 397; P. van Baa1 and N. D. Hari-Das, *Nucl. Phys.* B385:185 (1992); P. van Baal and B. van den Heuvel, *Nucl. Phys.* B417:215 (1994); R. E. Cutkosky, *J. Math. Phys.* 25:939 (1984); R. E. Cutkosky and K. Wang, *Phys. Rev.* D37:3024 (1988); K. Fujikawa, *Nucl. Phys.* B468:355 (1996).
3. R. Friedberg, T. D. Lee, Y. Pang and H. C. Ren, *Ann. Phys.* 246:381 (1996).
4. D. Karabali and V. P. Nair, *Nucl. Phys.* B464:135 (1996); *Phys. Lett.* B379B:141 (1996); D. Karabali, C. Kim and V. P. Nair, CCNY preprint CCNY-HEP-97-5, hep-th/9705087 (1997).
5. R. P. Feynman, *Nucl. Phys.* B188:479 (1981).
6. M. F. Atiyah, N. J. Hitchin and I. M. Singer, *Proc. R. Soc. Lond.* A.362:425 (1978).
7. I. M. Singer, *Physica Scripta* 24: 817 (1980); P. K. Mitter and C. M. Viallet, *Commun. Math. Phys.* 79:457 (1981); *Phys. Lett.* 85B:256 (1979); M. S. Narasimhan and T. R. Ramadas, *Commun. Math. Phys.* 67:21 (1979); M. Asorey and P. K. Mitter, *Commun. Math. Phys.* 80:43 (1981).
8. O. Babelon and C. M. Viallet, *Commun. Math. Phys.* 81:515 (1981); *Phys. Lett.* 103B:45 (1981).
9. J. Schwinger, *Phys. Rev.* 125:1043 (1962); 127:324 (1962); K. Gawedzki, *Phys. Rev.* D26:3593 (1982).
10. J. Fuchs, M. G. Schmidt and C. Schweigert, *Nucl. Phys.* B426:107 (1994).
11. S. G. Matinian, G. K. Savvidy and N. G. Ter-Arutunian Savvidy, *J.E.T.P. Lett.* 34:590 (1981); 53:421 (1981).
12. B. Simon, *Ann. Phys.* 146:209 (1983).
13. V. J. Berezinskii, *J.E.T.P.* 32:493 (1971); J. M. Kosterlitz and D. V. Thouless, *J. Phys.* C6:1181 (1973); J. M. Kosterlitz, *J. Phys.* C7:1046 (1974);
14. K. Pohlmeyer, *Commun. Math. Phys.* 46:207 (1976); A. M. Polyakov, *Phys. Lett.* 72B:224 (1977); V. Zakharov and A. Mikhailov, *J.E.T.P.* 27:47 (1978); M. Lüscher and K. Pohlmeyer, *Nucl. Phys.* B137:46 (1978); E. Brezin, C. Itzykson, J. Zinn-Justin and J.-B. Zuber, *Phys. Lett.* 82B:442 (1979).
15. Al. B. Zamolodchikov and A. B. Zamolodchikov, *Ann. Phys.* 120:253 (1979).
16. A. M. Polyakov and P. B. Wiegmann, *Phys. Lett.* 131B:121 (1983); P. B. Wiegmann, *Phys. Lett.* 152B:209 (1985).
17. H. E. Stanley, *Phys. Rev.* 176:718 (1968)
18. J.-L. Gervais and B. Sakita, Nucl. Phys. B91:301 (1975).
19. E. T.Whittaker and G. N. Watson, "A Course of Modern Analysis," Cambridge University Press, Cambridge (1927).

This page intentionally left blank

PAST THE HIGHEST-WEIGHT, AND WHAT YOU CAN FIND THERE

A.M. Semikhatov

Lebedev Physics Institute, Moscow 117924, Russia

INTRODUCTION

In this talk, I discuss how the structure of modules over the $N = 2$ superconformal algebra in two dimensions can be described in a simpler way in terms of modules over the affine algebra $\widehat{s\ell}(2)$. The key statements have been proved in[1,2], but I will be much less formal in my discussion. To advertise the main result,

the $N = 2$ and affine-$sl(2)$ representations theories are "essentially equivalent".

Before giving this a more precise meaning, let us see whether this comes as a news:

– on the one hand, the two algebras ($\widehat{s\ell}(2)$ and $N = 2$) appear to have very little in common, since one is a rank-2 (bosonic) affine Lie algebra, while the other is a rank-3 *super*algebra that is *not* an affine Lie algebra;
– on the other hand, the appearance of the two algebras in CFT is often 'correlated', they 'share' parafermionic theories, etc.

The equivalence can be shown in the following general setting:

1. One considers an arbitrary complex level $k \in \mathbb{C} \setminus \{-2\}$ on the $\widehat{s\ell}(2)$ side.
2. On the $N = 2$ side, one considers the 'standard' representation category, which includes the Verma modules, the (unitary and non-unitary) irreducible representations, etc., along with their images under the spectral flow (*twists*).
3. *Modulo the spectral flow transform*, this $N = 2$ category is equivalent to the category of $\widehat{s\ell}(2)$ representations *of the type that has not been considered before* — the so-called RELAXED highest-weight-type representations (and their twists).
4. On the other hand, the standard highest-weight-type $\widehat{s\ell}(2)$ representations turn out to be related to a narrower category of (twisted) TOPOLOGICAL $N = 2$ Verma modules.

Thus, the objects that are quite standard on the $N = 2$ side can be described in the $\widehat{s\ell}(2)$ terms by introducing a new type of modules, while only a subclass of $N = 2$ representations corresponds to the usual $\widehat{s\ell}(2)$ representations. Of a crucial importance is, therefore, the distinction between two different types of 'Verma' modules over each algebra; for the $N = 2$ algebra, this distinction is masked (due to an effect that we are going to discuss), which has resulted in some confusion in the literature as regards the structure and properties of $N = 2$ modules. From the $\widehat{s\ell}(2)$ point of view, the distinction is much easier to see, and it can roughly be summarized by saying that in the relaxed Verma modules,

one goes past the highest-weight vector.

We now describe this in more detail.

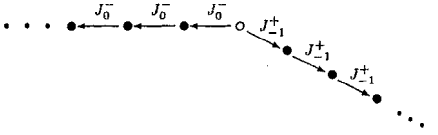
New Developments in Quantum Field Theory
Edited by Damgaard and Jurkiewicz, Plenum Press, New York, 1998

$\widehat{s\ell}(2)$ HIGHEST-WEIGHT REPRESENTATIONS

Let us begin with the $\widehat{s\ell}(2)$ algebra. We fix the level $k \neq -2$. Recall what one does when constructing a highest-weight-type module. The generators are broken into, roughly, two 'halves', one of which are declared annihilation operators with respect to a *highest-weight vector*, while the others *create* states, except for the 'Cartan' generator(s), whose eigenvalues simply 'label' the highest-weight vectors:

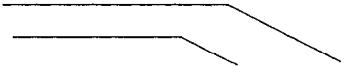
$$J^+_{\geq 0}\,|j,k\rangle_{s\ell(2)} = J^0_{\geq 1}\,|j,k\rangle_{s\ell(2)} = J^-_{\geq 1}\,|j,k\rangle_{s\ell(2)} = 0\,, \quad J^0_0\,|j,k\rangle_{s\ell(2)} = j\,|j,k\rangle_{s\ell(2)}\,, \quad (1)$$

where $j, k \in \mathbb{C}$. In the *Verma* module, by definition, there are no relations among the states produced by the creation operators from the highest-weight vector. The structure of $\widehat{s\ell}(2)$ Verma modules is conveniently encoded in the *extremal diagram*



(2)

The states shown in the diagram are *extremal* in the sense that they have boundary values of the (charge, level) bigrading; all of the other states of the module should be thought of as lying in the interior of the angle in the diagram. Finding a *submodule* in the Verma module can be (somewhat more schematically) represented as



(3)

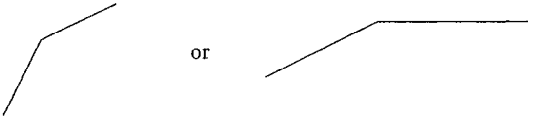
Whenever one considers quotients of Verma modules, the extremal diagrams become 'smaller', as some of the states are eliminated from the module. All of such extremal diagrams, therefore, satisfy the following criterion:

Any straight line going through any state intersects the boundary on at least one end,

which is formalized as follows: for any state $|X\rangle$ from the module,

$$\forall n \in \mathbb{Z}, \quad \exists N \in \mathbb{N}: \quad (\text{either} \quad (J^+_n)^N\,|X\rangle = 0 \quad \text{or} \quad (J^-_{-n})^N\,|X\rangle = 0) \qquad (4)$$

However, this criterion also selects the so-called *twisted* modules. As regards twisted *Verma* modules, their extremal diagrams are 'rotations' of the above, e.g.:



(5)

In more formal terms, a twisted Verma module $\mathfrak{M}_{j,k;\theta}$ is freely generated by $J^+_{\leq \theta-1}$, $J^-_{\leq -\theta}$, and $J^0_{\leq -1}$ from a *twisted highest-weight vector* $|j,\ k;\ \theta\rangle_{s\,l(2)}$ defined by the conditions

$$\begin{aligned} &J^+_{\geq \theta}\,|j,k;\theta\rangle_{s\ell(2)} = J^0_{\geq 1}\,|j,k;\theta\rangle_{s\ell(2)} = J^-_{\geq -\theta+1}\,|j,k;\theta\rangle_{s\ell(2)} = 0\,, \\ &\left(J^0_0 + \tfrac{k}{2}\theta\right)\,|j,k;\theta\rangle_{s\ell(2)} = j\,|j,k;\theta\rangle_{s\ell(2)}\,. \end{aligned} \qquad (6)$$

The mapping that underlies the construction of twisted modules is known as the spectral flow transform[4]

$$\mathcal{U}_\theta: \quad J^+_n \mapsto J^+_{n+\theta}\,, \qquad J^-_n \mapsto J^-_{n-\theta}\,, \qquad J^0_n \mapsto J^0_n + \tfrac{k}{2}\theta\delta_{n,0}\,, \qquad \theta \in \mathbb{Z}\,. \qquad (7)$$

The characteristic feature of extremal diagrams of $\widehat{s\ell}(2)$ Verma modules is the existence of an 'angle' that corresponds to the highest-weight vector. In a somewhat different context, this angle will reappear in the topological $N = 2$ Verma modules.

$N = 2$ ALGEBRA AND REPRESENTATIONS

The $N = 2$ superconformal algebra contains two fermionic currents, Q and G, in addition to the Virasoro generators L and the $U(1)$ current H. The commutation relations read as

$$\begin{aligned}
[\mathcal{L}_m, \mathcal{L}_n] &= (m-n)\mathcal{L}_{m+n}\,, & [\mathcal{H}_m, \mathcal{H}_n] &= \tfrac{\mathsf{C}}{3} m \delta_{m+n,0}\,,\\
[\mathcal{L}_m, \mathcal{G}_n] &= (m-n)\mathcal{G}_{m+n}\,, & [\mathcal{H}_m, \mathcal{G}_n] &= \mathcal{G}_{m+n}\,,\\
[\mathcal{L}_m, \mathcal{Q}_n] &= -n\mathcal{Q}_{m+n}\,, & [\mathcal{H}_m, \mathcal{Q}_n] &= -\mathcal{Q}_{m+n}\,, \qquad m,\ n \in \mathbb{Z}\,. \quad (8)\\
[\mathcal{L}_m, \mathcal{H}_n] &= -n\mathcal{H}_{m+n} + \tfrac{\mathsf{C}}{6}(m^2+m)\delta_{m+n,0}\,,\\
\{\mathcal{G}_m, \mathcal{Q}_n\} &= 2\mathcal{L}_{m+n} - 2n\mathcal{H}_{m+n} + \tfrac{\mathsf{C}}{3}(m^2+m)\delta_{m+n,0}\,,
\end{aligned}$$

The element C is central; in representations, we will not distinguish between C and its eigenvalue $\mathsf{c} \in \mathbb{C}$, which it will be convenient to parametrize as $\mathsf{c} = 3\frac{t-2}{t}$ with $t \in \mathbb{C}\setminus\{0\}$.

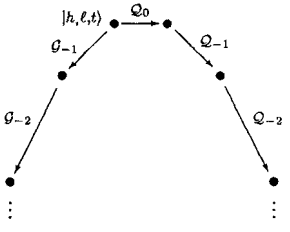
We now consider two types of 'Verma' modules over the $N = 2$ algebra, which we call the massive and the topological ones; the former are commonly considered as 'the' $N = 2$ Verma modules, while the latter are precisely those $N = 2$ modules that are in a good correspondence with the $\widehat{s\ell}(2)$ Verma modules from the previous section.

Massive $N = 2$ modules

A massive Verma module $\mathcal{U}_{h,\ell,t}$ is freely generated by the generators L_{-m}, H_{-m}, G_{-m}, $m \in \mathbb{N}$, and Q_{-m}, $m \in \mathbb{N}_0$ (with $\mathbb{N} = 1, 2, \ldots$ and $\mathbb{N}_0 = 0, 1, 2, \ldots$) from a *massive highest-weight vector* $|h, \ell, t\rangle$ satisfying the following set of highest-weight conditions:

$$\begin{aligned}
&\mathcal{Q}_{\geq 1}\,|h,\ell,t\rangle = \mathcal{G}_{\geq 0}\,|h,\ell,t\rangle = \mathcal{L}_{\geq 1}\,|h,\ell,t\rangle = \mathcal{H}_{\geq 1}\,|h,\ell,t\rangle = 0\,,\\
&\mathcal{H}_0\,|h,\ell,t;\theta\rangle = h\,|h,\ell,t\rangle\,, \qquad \mathcal{L}_0\,|h,\ell,t\rangle = \ell\,|h,\ell,t\rangle\,.
\end{aligned} \quad (9)$$

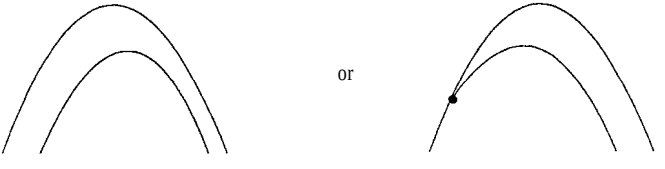
The extremal diagram of a massive Verma module reads



(10)

It has the shape of a parabola for the simple reason that, having acted on the highest-weight vector with, say, Q_0, applying the same operator once again gives identical zero, and 'the best one can do' to construct a state with the extremal (charge, level) bigrading is to act with the Q_{-1} mode, etc.

Now, in contrast to what we had with the $\widehat{s\ell}(2)$ Verma modules, there can be *two* different types of Verma submodules in $\mathcal{U}_{h,\ell,t}$. In the language of extremal diagrams, these look like (with the discrete parabolas replaced by the smooth ones for simplicity)



(11)

In the first case, we have a *massive* Verma *sub*module, all of the states on its extremal diagram (as well as those on the extremal diagram of the module itself) satisfying the annihilation conditions

$$\mathcal{Q}_{-\theta+m+1} \approx \mathcal{G}_{\theta+m} \approx \mathcal{L}_{m+1} \approx \mathcal{H}_{m+1} \approx 0\,,\ m \in \mathbb{N}_0 \tag{12}$$

for some θ (which ranges over the integers, from $-\infty$ in the left bottom end to $+\infty$ in the right end of the parabola). In the other case, on the contrary, there is a distinguished state, marked with a •, that satisfies the annihilation conditions

$$\mathcal{Q}_{-\underline{\theta}+m} \approx \mathcal{G}_{\underline{\theta}+m} \approx \mathcal{L}_{m+1} \approx \mathcal{H}_{m+1} \approx 0\,, \quad m \in \mathbb{N}_0 \tag{13}$$

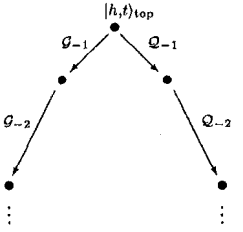
for some $\underline{\theta} \in \mathbb{Z}$. Such states will be referred to as (twisted) *topological highest-weight vectors*, and in the above context, as *topological singular vectors* (the $\underline{\theta} = 0$ case being to the 'untwisted' one). We now define these module systematically.

Topological $N = 2$ modules

A *twisted topological Verma module* is generated from a twisted *topological highest-weight vector* subjected to annihilation conditions (13), with the Cartan generators chosen to have the following eigenvalues:

$$\begin{aligned} (\mathcal{H}_0 + \tfrac{c}{3}\theta)\,|h,t;\theta\rangle_{\text{top}} &= h\,|h,t;\theta\rangle_{\text{top}}\,, \\ (\mathcal{L}_0 + \theta\mathcal{H}_0 + \tfrac{c}{6}(\theta^2+\theta))\,|h,t;\theta\rangle_{\text{top}} &= 0\,. \end{aligned} \tag{14}$$

The extremal diagram of a topological Verma module reads (in the $\theta = 0$ case for simplicity, with $|h,t\rangle_{\text{top}} \equiv |h,t;0\rangle_{\text{top}}$)



(15)

A characteristic feature of this extremal diagram is the existence of a 'cusp', i.e. a state that satisfies stronger highest-weight conditions than the other states in the diagram. As a result, the extremal diagram is narrower than that of a massive Verma module. When taking quotients, the extremal diagrams may only become smaller, which allows us to formulate a criterion that automatically singles out the *topological highest-weight-type* modules (the corresponding *O*-category [9]). For any $n \in \mathbb{Z}$, by the 'massive' parabola going through a state $|X\rangle$, we understand the set of states

$$\mathcal{Q}_{n-N} \ldots \mathcal{Q}_{n-1}\,\mathcal{Q}_n\,|X\rangle = 0\,, \quad \mathcal{G}_{-n-M} \ldots \mathcal{G}_{-n-2}\,\mathcal{G}_{-n-1}\,|X\rangle = 0\,, \qquad M, N \in \mathbb{N}. \tag{16}$$

Then, a module belongs to the twisted 'topological' *O*-category if, for any state,

any massive parabola intersects extremal diagram of the module on at least one end,

which, again, means simply that the states (16) become zero in at least one branch.

The massive $N = 2$ Verma modules do not satisfy this criterion. However, in the massive case as well, one can formulate a criterion that does not allow the modules to become too wide: for any element $|X\rangle$,

$$\forall n \in \mathbb{Z}, \quad \text{either } \ldots \mathcal{Q}_{n-2}\,\mathcal{Q}_{n-1}\,\mathcal{Q}_n\,|X\rangle = 0 \quad \text{or } \ldots \mathcal{G}_{-n-2}\,\mathcal{G}_{-n-1}\,\mathcal{G}_{-n}\,|X\rangle = 0\,. \tag{17}$$

We now address the problem of finding the $\widehat{s\ell}(2)$ counterpart of all these constructions. We first map the generators and then investigate the representations.

FROM $N = 2$ TO $\widehat{s\ell}(2)$

An operator construction

We now use an operator construction allowing us to build the $\widehat{s\ell}(2)$ currents out of the $N = 2$ generators and a free scalar with the operator product $\phi(z)\phi(w) = -\ln(z-w)$. As a necessary preparation, we 'pack' the modes of the $N = 2$ generators into the corresponding fields, $\mathcal{T}(z) = \sum_{n\in\mathbb{Z}} \mathcal{L}_n z^{-n-2}$, $\mathcal{G}(z) = \sum_{n\in\mathbb{Z}} \mathcal{G}_n z^{-n-2}$, $\mathcal{Q}(z) = \sum_{n\in\mathbb{Z}} \mathcal{Q}_n z^{-n-1}$, and $\mathcal{H}(z) = \sum_{n\in\mathbb{Z}} \mathcal{H}_n z^{-n-1}$, and similarly with the $s\ell(2)$ currents. We also define vertex operators $\psi = e^{\phi}$ and $\psi^* = e^{-\phi}$. Then, for $c \neq 3$,

$$J^+ = \mathcal{Q}\psi\,, \qquad J^- = \tfrac{3}{3-c}\,\mathcal{G}\psi^*\,, \qquad J^0 = -\tfrac{3}{3-c}\,\mathcal{H} + \tfrac{c}{3-c}\,\partial\phi \tag{18}$$

are the $\widehat{s\ell}(2)$ generators of level $k = \frac{2c}{3-c}$.*.

Relating the representations

The behaviour of *representations* under operator constructions of this sort can be quite complicated.† In our case, we take a topological Verma module $V_{h,t}$ and *tensor* it with the module Ξ of the free scalar. The latter module is defined as $\Xi = \oplus_{n\in\mathbb{Z}}\mathcal{H}_n$, where H_n is a Verma module with the highest-weight vector $|n\rangle_\phi$ such that

$$\phi_m|n\rangle_\phi = 0\,,\ m \geq 1\,, \quad \psi_m|n\rangle_\phi = 0\,,\ m \geq n+1\,, \quad \psi^*_m|n\rangle_\phi = 0\,,\ m \geq -n+2\,, \tag{19}$$

and $\phi_0|n\rangle_\phi = -n|n\rangle_\phi$. We then have the following Theorem[1], in which we also refer to a free scalar with signature –1, whose modes commute with the $\widehat{s\ell}(2)$ generators (18):

$$I^- = \sqrt{\tfrac{k+2}{2}}(\mathcal{H} - \partial\phi)\,. \tag{20}$$

The modes I^-_n generate a Heisenberg algebra. Then the module $\mathcal{H}^-_q$ is defined as a Verma module over this Heisenberg algebra with the highest-weight vector defined by $I^-_n|q\rangle^- = 0$, $n \geq 1$, and $I^-_0|q\rangle^- = q|q\rangle^-$. Now,

Theorem 1

1. *There is an isomorphism of* $\widehat{s\ell}(2)$ *representations*

$$\mathcal{V}_{h,t} \otimes \Xi \approx \bigoplus_{\theta\in\mathbb{Z}} \mathfrak{M}_{-\frac{t}{2}h,t-2;\theta} \otimes \mathcal{H}^-_{\sqrt{\frac{t}{2}}(h+\theta)} \tag{21}$$

   where on the LHS the $\widehat{s\ell}(2)$ *algebra acts by the generators (18), while on the RHS it acts on* $\mathfrak{M}_{-\frac{t}{2}h,t-2;\theta}$ *as on a twisted Verma module,*

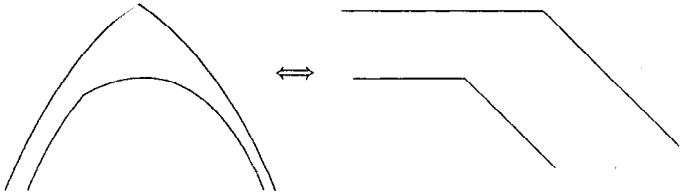
2. *A singular vector exists in* $V_{h,t}$ *if and only if a singular vector exists in one (hence, in all) of the modules* $\mathfrak{M}_{-\frac{t}{2}h,t-2;\theta}$, $\theta \in \mathbb{Z}$. *Whenever this is the case, moreover, the submodules associated with the singular vectors, in their own turn, satisfy an equation of the same type as (21).*

The statement regarding singular vectors appeared, in a rudimentary form, in[5]. The theorem means that, as regards the existence and the structure of submodules, the

* At the Conference, M. Halpern told me that such a mapping had been known to M. Peskin *et al.*, but I could not find the reference.

† Recall, for instance, how the $\widehat{s\ell}(2)$ Verma modules are rearranged under the Wakimoto bosonization[3] — Wakimoto modules more or less 'interpolate' between Verma and contragredient Verma modules.

topological N *equivalent'* to $\widehat{s\ell}(2)$ Verma modules. Twisted topological Verma submodu simultaneously with $s\ell(2)$ Verma submodule‡:



A common feature of $\widehat{s\ell}(2)$ and topological $N = 2$ Verma modules is that all of them are generated from a state that satisfies stronger annihilation conditions than the other states in the extremal diagram. What *is* somewhat unusual about this correspondence, though, is the fact that on the $\widehat{s\ell}(2)$ side such a 'cusp' state satisfies the same annihilation conditions as the highest-weight state of the module, whereas on the $N = 2$ side it satisfies *twisted* topological highest-weight conditions. Such, however, is the structure of the $N = 2$ algebra; ignoring this fact may mislead one to consider a submodule in a topological Verma module as generated from the *top-level* state in the extremal diagram of the submodule, in which case one could incorrectly conclude that one is dealing with a *massive* Verma submodule (which, as we saw in (16), is *not* the case).

Thus, to the well-known $\widehat{s\ell}(2)$ singular vectors $|\mathrm{MFF}(r,s,k)\rangle^{\pm}$, $r, s \in \mathbb{N}$, given by the construction of [6], there correspond the so-called *topological singular vectors* [7,8] $|E(r,s,t)\rangle^{\pm}$, $t = k+2$, which satisfy the twisted topological highest-weight conditions (13) with $\underline{\theta} = \mp r$ respectively. Topological singular vectors occur in the topological Verma module $V_{h,t}$ whenever $h = \mathsf{h}^{\pm}(r,s,t)$, $r, s \in \mathbb{N}$, where

$$\mathsf{h}^{-}(r,s,t) = \tfrac{r+1}{t} - s\,, \qquad \mathsf{h}^{+}(r,s,t) = -\tfrac{r-1}{t} + s - 1 \tag{22}$$

The explicit construction for $N = 2$ singular vectors can be found in [7,8].

The idea regarding the correspondence between submodules can be developed in the direction of category theory. Very roughly, a category is a collection of *objects*, some of which may be related by *morphisms*. Taking the objects to be all the topological $N = 2$ Verma modules, the morphisms would have to be the usual $N = 2$-homomorphisms. However, two *Verma* modules are related by a morphism only if one of the modules can be embedded into the other. We have just seen that such embeddings — i.e., the occurrence of *sub*modules — is 'synchronized' between the topological Verma modules over $N = 2$ and the $\widehat{s\ell}(2)$ Verma modules. In fact, there also exists a mapping (a *functor*) in the inverse direction, and one eventually concludes that the category *TOP* of topological $N = 2$ Verma modules is equivalent to the category *VER* of $\widehat{s\ell}(2)$ Verma modules. To be more precise, the appearance of the twist (the spectral flow transform) results in that this equivalence takes place only after one effectively factorizes over the spectral flows on either $N = 2$ and $\widehat{s\ell}(2)$ sides, see[1] for a rigorous statement. Anyway, an immediate consequence of this equivalence is that

Embedding diagrams of Verma modules are identical on the $N = 2$ and $\widehat{s\ell}(2)$ sides,

where we are so far restricted to *topological* Verma modules on the $N = 2$ side. Since the $\widehat{s\ell}(2)$ embedding diagrams are well-known, this spares us. the job of deriving them in a less friendly environment of the $N = 2$ algebra.

‡ in particular, the Ξ and $H^{-}_{\dots}$ modules in (21) are really 'auxiliary', since nothing can happen there that would violate the correspondence between submodules in topological $N = 2$ and $\widehat{s\ell}(2)$ Verma modules.

Where do the massive $N = 2$ modules go?

Having seen that the topological $N = 2$ modules are in a 'good' correspondence with $\widehat{s\ell}(2)$ modules, we may recall from (11) that this involves only a 'small' part of $N = 2$ Verma modules, whereas the massive $N = 2$ modules (the 'wide' ones) seem to have nowhere to go in the $\widehat{s\ell}(2)$ picture, since all of the capacities of the $\widehat{s\ell}(2)$ Verma modules are already used up to maintain the correspondence with the topological (the 'narrow') $N = 2$ Verma modules.

RELAXED $\widehat{s\ell}(2)$ VERMA MODULES

Solving the above problem requires introducing a new class of $\widehat{s\ell}(2)$ modules. These have a characteristic property that their extremal diagrams have *no 'cusps'*, which will be crucial for relating them to the massive $N = 2$ Verma modules (whose extremal diagrams have no cusps either). The recipe is to *relax* the annihilation conditions imposed on the highest-weight vector[§]:

$$\cancel{J_0^+ \, |j,k\rangle_{s\ell(2)} \; 0}$$

For $\theta \in \mathbb{Z}$, the twisted relaxed Verma module $\mathfrak{R}_{j,\Lambda,k;\theta}$ is generated from the state $|j,\Lambda,k;\theta\rangle_{s\ell(2)}$ that satisfies the annihilation conditions

$$J^+_{\geq\theta+1} \, |j,\Lambda,k;\theta\rangle_{s\ell(2)} = J^0_{\geq 1} \, |j,\Lambda,k;\theta\rangle_{s\ell(2)} = J^-_{\geq -\theta+1} \, |j,\Lambda,k;\theta\rangle_{s\ell(2)} = 0 \tag{23}$$

by a free action of the operators $J^+_{\leq\theta-1}$, $J^-_{\leq-\theta-1}$, and $J^0_{\leq-1}$, and by the action of operators J^+_θ and $J^-_{-\theta}$ subject to the constraint

$$\left(J^-_{-\theta}J^+_\theta + (k+2)\theta(j - \tfrac{k}{4}\theta)\right) \, |j,\Lambda,k;\theta\rangle_{s\ell(2)} = \Lambda \, |j,\Lambda,k;\theta\rangle_{s\ell(2)} \, . \tag{24}$$

In addition, the highest-weight state $|j,\Lambda,k;\theta\rangle_{s\ell(2)}$ satisfies

$$\left(J^0_0 + \tfrac{k}{2}\theta\right) \, |j,\Lambda,k;\theta\rangle_{s\ell(2)} = j \, |j,\Lambda,k;\theta\rangle_{s\ell(2)} \, . \tag{25}$$

Then, we can act on the highest-weight vector with both J^+_0 and J^-_0, thereby producing new states

$$|j,\Lambda,k;\theta|n\rangle_{s\ell(2)} = \begin{cases} (J^-_{-\theta})^{-n} \, |j,\Lambda,k;\theta\rangle_{s\ell(2)} \, , & n < 0 \, , \\ (J^+_\theta)^n \, |j,\Lambda,k;\theta\rangle_{s\ell(2)} \, , & n > 0 \, , \end{cases} \tag{26}$$

with $|j,\Lambda,k;\theta|0\rangle_{s\ell(2)} = |j,\Lambda,k;\theta\rangle_{s\ell(2)}$. As a result, the extremal diagram opens up to the straight angle; in the case of $\theta = 0$ it thus becomes

$$\cdots \xleftarrow{J^-_0} \bullet \xleftarrow{J^-_0} \bullet \xleftarrow{J^-_0} \star \xrightarrow{J^+_0} \bullet \xrightarrow{J^+_0} \bullet \cdots \tag{27}$$

where all of the other states from the module correspond to points below the line. The * state is the above $|j,\Lambda,k;\theta=0\rangle_{s\ell(2)}$. We also define $|j,\Lambda,k|n\rangle_{s\ell(2)} = |j,\Lambda,k;0|n\rangle_{s\ell(2)}$, then the norms of these extremal states are given by

$$\big\| |j,\Lambda,k|n\rangle_{s\ell(2)} \big\|^2 = \begin{cases} \prod_{i=0}^{-n-1}(\Lambda + 2(i+1)j - i(i+1)) \, , & n \leq -1 \, , \\ \prod_{i=0}^{n-1}(\Lambda - 2ij - i(i+1)) \, , & n \geq 1 \, . \end{cases} \tag{28}$$

Thus, as we move either right or left along the extremal diagram, the norm becomes negative eventually. The negative-norm states can be factored away if it happens that the norm of one of the extremal states is exactly zero. This is the case whenever $\Lambda = \Lambda_{\rm ch}(p,j) \equiv p(p+1)+2pj$, $p \in \mathbb{Z}$; then the factors in (28) become $(1+i+p)(2j+p-i)$ and $(p - i)\,(1 + i + 2j + p)$ respectively. The corresponding zero-norm state

[§] Yet the crossing out operation looked nicer in my transparencies.

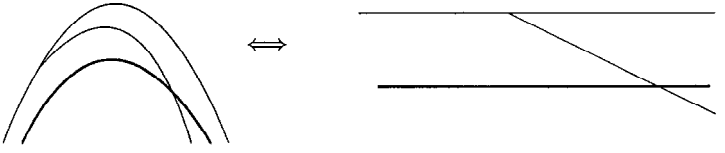
$$|C(p,j,k)\rangle_{s\ell(2)} = \begin{cases} (J_0^-)^{-p}\ |j,\Lambda_{\rm ch}(p,j),k\rangle_{s\ell(2)}\ , & p \le -1 \\ (J_0^+)^{p+1}\ |j,\Lambda_{\rm ch}(p,j),k\rangle_{s\ell(2)}\ , & p \ge 0 \end{cases} \tag{29}$$

then satisfies the Verma highest-weight conditions for $p \le -1$ and the twisted Verma highest-weight conditions with the twist parameter $\theta = 1$ for $p \ge 1$. Thus, it is a *singular vector*, which can be quotiened away along with a tail of negative-norm states. For historical reasons [17, 1], states (29) are called *charged* singular vectors.

Theorem 1 is now extended to a similar statement about the isomorphism of $\widehat{s\ell}(2)$ representations:

$$\mathcal{U}_{h,\ell,t} \otimes \Xi \approx \bigoplus_{\theta\in\mathbb{Z}} \mathfrak{R}_{-\frac{t}{2}h,t\ell,t-2;\theta} \otimes \mathcal{H}^-_{\sqrt{\frac{t}{2}}(h+\theta)} \tag{30}$$

where on the LHS the $\widehat{s\ell}(2)$ algebra acts by generators (18), while on the RHS it acts naturally on $\mathfrak{R}_{-\frac{t}{2}h,t\ell,t-2;n-\theta}$ as on a twisted relaxed Verma module (as follows from the notations, its parameters are $j = -\frac{t}{2}h$, $\Lambda = t\ell$, and $k = t - 2$). As in Theorem 1, the singular vectors appear in $U_{h,l,t}$ and $\mathfrak{R}_{-\frac{t}{2}h,t\ell,t-2;\theta}$ simultaneously, and, in addition, the same relation (30) holds for the corresponding submodules, which can be either massive/relaxed or topological/usual-Verma:



(for the topological/usual-Verma submodules, Eq. (21) holds instead of (30)). As a consequence,

embedding diagrams of massive $N = 2$ Verma modules are isomorphic to embedding diagrams of relaxed $\widehat{s\ell}(2)$ Verma modules,

see [10] for the classification of the embedding diagrams.

MASSIVE AND RELAXED VERMA MODULES IN BOSONIC STRING

According to the above equivalence statements, it is inessential in many respects whether one analyses $N = 2$ Verma modules or relaxed $\widehat{s\ell}(2)$ Verma modules. In this section, we show how the above constructions (extremal states, massive Verma modules, etc.) arise naturally in the bosonic string. In the noncritical bosonic string, one has the $N = 2$ algebra realized as in [15, 16]. Applying the mapping described in the previous section, one recovers the corresponding realization of $\widehat{s\ell}(2)$ currents found in [5].

Let us describe in more detail the $N = 2$ version of this construction. One starts with a matter theory represented by the energy-momentum tensor T with central charge $13 - 6/t - 6t$ and tensors it with the bc ghosts and a (free) Liouville scalar. The resulting $N = 2$ generators read as

$$\begin{aligned} \mathcal{T} &= T - t\partial\varphi\partial\varphi - (1+t)\partial^2\varphi - \partial bc - 2b\partial c\,, \quad \mathcal{H} = 2\partial\varphi + bc\,, \quad \mathcal{G} = b\,, \\ \mathcal{Q} &= -2b\partial cc - 2t\partial\varphi\partial\varphi c + 4\partial\varphi\partial c + 2Tc + (2-2t)\partial^2\varphi\, c + (1-\tfrac{2}{t})\partial^2 c\,. \end{aligned} \tag{31}$$

where the Liouville OPE is chosen in a non-canonical normalization $\partial\varphi(z)\,\partial\varphi(w) = -(1/2t)1/(z-w)^2$. The representation space is then constructed as described in the next subsection.

Constructing the representation

Each matter primary $|\Delta\rangle_{\rm m}$ of dimension Δ can be dressed into $N = 2$ primaries either as

$$\left|h, \Delta - \tfrac{1}{4}(2 - 2h - t + h^2 t), t; \theta\right\rangle_* = |\theta\rangle_{\rm gh} \otimes \left|e^{2t(-\frac{1}{2}-\frac{h}{2}-\frac{\theta}{t})\varphi}\right\rangle \otimes |\Delta\rangle_{\rm m} \tag{32}$$

or by replacing $h \mapsto \frac{2}{t} - h$ in the last formula (which does not change the $\mathcal{L}_0$ eigenvalue). Here,

$$|\theta\rangle_{\rm gh} = \begin{cases} b\,\partial b \,\ldots\, \partial^{-\theta-2} b\,|0\rangle_{\rm GH}\,, & \theta \le -2\,, \\ |0\rangle_{\rm GH}\,, & \theta = -1\,, \\ c\,\partial c \,\ldots\, \partial^{\theta} c\,|0\rangle_{\rm GH}\,, & \theta \ge 0\,, \end{cases} \tag{33}$$

are ghost vacua in different pictures[18]. Each of the states (32) satisfies the twisted massive highest-weight conditions with the twist parameter θ. Further, in the tensor product of the matter Verma module with the ghosts (and the Liouville), each of the states (32) comes together with an infinite number of extremal states obtained by tensoring the same matter primary with ghost vacua in different pictures:

$$|\alpha\rangle_{\rm gh} \otimes \left|e^{2t(-\frac{1}{2}-\frac{h}{2}-\frac{\theta}{t})\varphi}\right\rangle \otimes |\Delta\rangle_{\rm m} \quad \text{and} \quad |\alpha\rangle_{\rm gh} \otimes \left|e^{2t(-\frac{1}{2}-\frac{1}{2}+\frac{h}{2}-\frac{\theta}{t})\varphi}\right\rangle \otimes |\Delta\rangle_{\rm m}\,, \quad \alpha \in \mathbb{Z}\,. \tag{34}$$

We thus see that in this realization,

choosing the ghost picture corresponds to traveling over the extremal diagram.

The 'bosonization' (31) has the following effect: twisted *topological* highest-weight conditions (13) are satisfied whenever the dimension of a state (32) vanishes:

$$|h, 0, t; \theta\rangle_* = |h, t; \theta\rangle_{\rm top}\,. \tag{35}$$

We will thus call $|h,\ 0,\ t;\ \theta\rangle_*$ the *pseudomassive* (highest-weight) states.

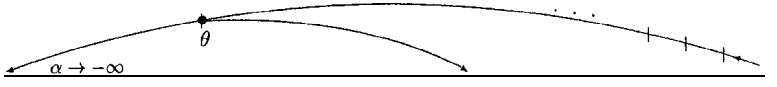
The *generalized DDK prescription* is that there be a twisted topological primary state among the extremal states (34). This is a condition on how the parameters in the tensor product of matter and Liouville are related: the matter dimension should be

$$\Delta(h, t) = \tfrac{1}{4}(2 - 2h - t + h^2 t)\,. \tag{36}$$

Then extremal states (34) become

$$|D(h, t, \theta, \alpha)\rangle \equiv |\alpha\rangle_{\rm gh} \otimes \left|e^{2t(-\frac{1}{2}-\frac{h}{2}-\frac{\theta}{t})\varphi}\right\rangle \otimes |\Delta(h, t)\rangle_{\rm m} = \left|h + \tfrac{2(\theta-\alpha)}{t}, \tfrac{(\theta-\alpha)(\alpha-\theta+1-ht)}{t}, t; \alpha\right\rangle_*\,,$$

As α runs over the integers, the $D(h,\ t,\ \theta,\ \alpha)$ states fill out an extremal diagram:



(37)

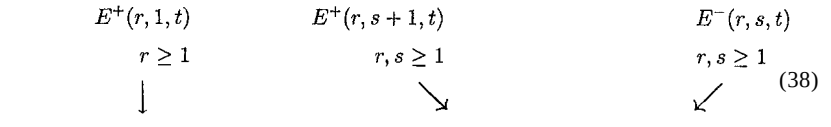
The states in the upper curve are freely generated from $+\infty$. The state at $\alpha = \theta$ is the twisted topological highest-weight state $|h,\ t;\ \theta\rangle_{\rm top}$, with a twisted topological Verma submodule being generated from it. We also have another extremal diagram $D'(h,\ t,\ \theta,\ \alpha) = D(\frac{2}{t} - h,\ t,\ \theta,\ \alpha)$.

Whenever $|D(h,\ t,\ \theta,\ \alpha_0))$ admits a singular vector for some $\alpha_0 \neq \theta$, each of the states $|D(h,\ t,\ \theta, \alpha\,)\rangle$, $\alpha \neq \theta$, admits a massive singular vector, while $|D(h, t, \theta, \theta)\rangle$ admits a topological singular vector. Then the states in the D- and D'-diagrams are given by

$$\begin{aligned} \left|D(\mathrm{h}^-(r, s, t), t, \theta, \alpha)\right\rangle &= |\alpha\rangle_{\rm gh} \otimes \left|e^{(-1-t-r+st-2\theta)\varphi}\right\rangle \otimes |\Delta_{r,s}(t)\rangle_{\rm m}\,, \\ \left|D(\mathrm{h}^+(r, s+1, t), t, \theta, \alpha\right\rangle &= |\alpha\rangle_{\rm gh} \otimes \left|e^{(-1-t+r-st-2\theta)\varphi}\right\rangle \otimes |\Delta_{r,s}(t)\rangle_{\rm m}\,, \end{aligned} \qquad r,\ s \ge 1\,,$$

with $\mathrm{h}^{\pm}(r,\ s,\ t)$ defined in (22). We see that the $|E(r, 1, t)\rangle^+$ topological singular vectors cannot be constructed out of the matter ones. In fact, the topological $N = 2$ singular

vectors evaluate in terms of the Virasoro algebra as follows:



(38)

Acting with the screening current

Whenever one uses an operator construction ('bosonization'), leading to some additional effects (e.g., vanishings) in representations (Eq. (35) in our case), one should expect the appearance of a screening current (cf., for instance, [19]). We, indeed, have a *fermionic screening current* of the form[20, 21]

$$F = b\, e^{t\varphi}\, \Psi_{12}\,, \tag{39}$$

where $\Psi_{1,2}$ is the '12' operator in the matter (Virasoro) sector. The $\Psi_{1,2}$ operator has two components that can be distinguished by picking out the following terms from the fusion relations:

$$\Psi_{12}^{\pm} * |\Delta(h,t)\rangle_{\mathrm{m}} \sim |\Delta(h \pm 1, t)\rangle_{\mathrm{m}}\,. \tag{40}$$

Here, the highest-weight states may be understood to be those in *Verma* modules over the Virasoro algebra. Using this, we now can construct the following action of the screening F on the pseudomassive modules over the $N = 2$ algebra (we omit the integral which makes the screening *charge* out of the current) :

$$F^{\pm} : |D(h,t,\theta,\alpha)\rangle \mapsto \begin{cases} |D(h \pm 1, t, \theta, \alpha - 1)\rangle\,, & \alpha \geq \theta + 1\,, \\ 0\,, & \text{otherwise} \end{cases} \tag{41}$$

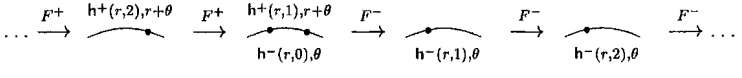
(which of the extremal states do, and which do not, vanish under the action of the screening, follows from a simple analysis of operator products). Then,

$$\begin{aligned} F^- &: \quad \left|D(\mathsf{h}^-(r,s,t),t,\theta,\alpha)\right\rangle \;\mapsto\; \left|D(\mathsf{h}^-(r,s+1,t),t,\theta,\alpha-1)\right\rangle, \\ F^+ &: \left|D(\mathsf{h}^+(r,s+1,t),t,\theta,\alpha)\right\rangle \;\mapsto\; \left|D(\mathsf{h}^+(r,s,t),t,\theta,\alpha-1)\right\rangle. \end{aligned} \qquad \alpha \geq \theta + 1\,. \tag{42}$$

Next, we observe that the identity

$$\left|D(\mathsf{h}^-(r,s,t),t,\theta_1,\alpha)\right\rangle = \left|D(\mathsf{h}^+(r,s+1,t),t,\theta_2,\alpha)\right\rangle, \quad \forall\alpha\,, \tag{43}$$

holds if and only if either $s = 0$, $r + \theta_1 = \theta_2$, or $t = \frac{r+\theta_1-\theta_2}{s}$. For generic t, we can use the $s = 0$ case in order to connect the two series of mappings (42) together. Omitting the t parameter, we label the extremal diagrams spanned out by the $|D(\mathsf{h}^{\pm}(r,s,t),t,\theta,\alpha)\rangle$ states by the corresponding $\mathsf{h}^{\pm}(r,s)$ and the value of θ that gives the position of the topological point. We then have the following mappings of modules with the extremal diagrams (37)



(44)

This sequence applies to *pseudomassive* modules, i.e., those in which the Virasoro part is taken to be *Verma* modules. We now investigate whether it is possible to go over

from (44) to a similar sequence for *quotient* modules. According to (38), the massive $N = 2$ singular vectors would be factored away in all of the terms starting with and after $\mathrm{h}^-(r, 1)$ as soon as the Virasoro singular vectors are factored away. This allows one to define the F^-mappings between the *irreducible* representations. The same is true for the modules before and including $\mathrm{h}^+(r, 2)$. In the middle term $\mathrm{h}^-(r, \mathrm{l})$, however, *there is no submodule* to factor over in the corresponding Virasoro Verma module. Yet, taking the composition F^-o F^+ allows us to make (44) into the exact sequence

$$\ldots \xrightarrow{F^+} \overline{\left|D(\mathrm{h}^+(r,3,t),t,\theta+r)\right\rangle} \xrightarrow{F^+} \overline{\left|D(\mathrm{h}^+(r,2,t),t,\theta+r)\right\rangle} \xrightarrow{F^-\mathrm{o}F^+} \overline{\left|D(\mathrm{h}^-(r,1,t),t,\theta)\right\rangle} \xrightarrow{F^-} \overline{\left|D(\mathrm{h}^-(r,2,t),t,\theta)\right\rangle} \xrightarrow{F^-} \ldots \qquad (45)$$

where the bars denote that the Virasoro singular vectors are declared to vanish (i.e., *irreducible* representations are taken in the matter sector in (32), (34), and similar formulae). It may be observed that this exact sequence is parallel to an exact sequence between representation of the quantum group $sl(2|1)q$ (for q not a root of unity in accordance with the above choice of generic t), which is not a coincidence[21]. The $sl(2|1)q$ embeddings are also performed by 'charged' singular vectors, with a due analogue of the F^- o F^+ composition in the center. The $sl(2|1)$ quantum group has long been suspected to govern the $\widehat{s\ell}(2)$ fusion rules [22], however the presently observed symmetry is only $osp(1|2)q$.

Acknowledgements. It is a pleasure to thank the Organizers of the Conference, first of all P.H. Damgaard and J. Jurkiewicz. I am grateful to B. Feigin and I. Tipunin for a fruitful collaboration, and to M. Halpern, I. Shchepochkina, V. Sirota, and V. Tolstoy for useful discussions.

REFERENCES

1. B.L. Feigin, A.M. Semikhatov, and I.Yu. Tipunin, Equivalence between chain categories of representations of affine $sl(2)$ and $N = 2$ superconformal algebras, hep-th/9701043.
2. A.M. Semikhatov and I.Yu. Tipunin, The structure of Verma modules over the $N = 2$ superconformal algebra, hep-th/9704111, to appear in *Commun. Math. Phys.*
3. B.L. Feigin and E.V. Frenkel, Representations of Affine Kac–Moody Algebras and Bosonization, in: *Physics and Mathematics of Strings, eds.* L. Brink, D. Friedan, and A.M. Polyakov.
4. K. Bardakçi and M.B. Halpern, *Phys. Rev. D*3 (1971) 2493.
5. A.M. Semikhatov, *Mod. Phys. Lett. A*9 (1994) 1867.
6. F.G. Malikov, B.L. Feigin, and D.B. Fuchs, *Funk. An. Prilozh.* 20, N2 (1986) 25.
7. A.M. Semikhatov and I.Yu. Tipunin, *Int. J. Mod. Phys. A*11 (1996) 4597.
8. A.M. Semikhatov and I.Yu. Tipunin, All Singular Vectors of the $N = 2$ Superconformal Algebra via the Algebraic Continuation Approach, hep-th/9604176.
9. V.G. Kač Infinite Dimensional Lie Algebras, *Cambridge University Press*, 1990.
10. A.M. Semikhatov and V.A. Sirota, Embedding diagrams of $N = 2$ and relaxed- $\widehat{s\ell}(2)$ Verma modules, *to appear*.
11. E.B. Kiritsis, *Int. J. Mod. Phys. A* (1988) 1871.
12. V.K. Dobrev, Phys. Lett. B186 (1987) 43.
13. Y. Matsuo, *Prog. Theor. Phys.* 77 (1987) 793.
14. M. Dörrzapf, *Commun. Math. Phys.* 180 (1996) 195.
15. B. Gato-Rivera and A.M. Semikhatov, *Phys. Lett. B*293 (1992) 72.
16. M. Bershadsky, W. Lerche, D. Nemeschansky, and N.P. Warner, *Nucl. Phys. B*401 (1993) 304.
17. W. Boucher, D. Friedan, and A. Kent, *Phys. Lett. B*172 (1986) 316.
18. D.H. Friedan, E.J. Martinec, and S.H. Shenker, *Nucl. Phys. B*271 (1986) 93.
19. G. Felder, *Nucl. Phys. B*317 215 (1989).
20. I. Tipunin, unpublished.
21. B.L. Feigin, A.M. Semikhatov, and I.Yu. Tipunin, *in preparation.*
22. B. Feigin and F. Malikov, Modular functor and representation theory of $\widehat{\mathfrak{sl}}_2$ at a rational level, q-alg/9511011.

This page intentionally left blank

THE SPECTRAL DIMENSION ON BRANCHED POLYMER ENSEMBLES

Thordur Jonsson,[1] and John F. Wheater [2*]

[1]Raunvisindastofnun Haskolans, University of Iceland
Dunhaga 3, 107 Reykjavik
Iceland
[2]Department of Physics, University of Oxford
Theoretical Physics,
1 Keble Road,
Oxford OX1 3NP

INTRODUCTION

The nature of the dimensionality of (euclidean) space-time in the ensemble of manifolds appearing in quantum gravity has recently attracted much attention[1-7]. On a smooth, almost flat (meaning without curvature singularities of any sort) manifold, all reasonable definitions of dimensionality will agree. However the ensemble of manifolds appearing in the path integral in quantum gravity contains many members which are neither smooth nor flat (in fact one expects the completely smooth and flat ones to have measure zero); various definitions of dimension can then lead to different numerical values because they probe different aspects of the global geometry. In this talk I considered the spectral dimension in models of two-dimensional quantum gravity coupled to conformal matter with central charge $c > 1$. Another talk at this workshop concentrated on the Hausdorff dimension in various cases[8]; as we shall see, in the branched polymer phase (which is neither smooth nor flat) these two dimensions are not the same.

Two-dimensional quantum gravity is defined in the canonical ensemble (CE) by the partition function

$$Z(V) = \int Dg\,\delta\left(\int d^2\xi\,\sqrt{g} - V\right)\exp\left(-S_{eff}[g]\right) \qquad (1)$$

where the functional integral runs over all physically inequivalent metrics g with the volume (area in 2d) constrained by the delta-function to be V. $S_{eff}[g]$ is the effective action obtained for the metric after integrating out all matter fields (which we assume are conformal). Expectation values of reparametrization invariant quantities in the CE

*Presented at the Workshop by J.F.Wheater

New Developments in Quantum Field Theory
Edited by Damgaard and Jurkiewicz, Plenum Press, New York, 1998

are given by

$$\langle \cdot \rangle_{CE} = \frac{1}{Z(V)} \int Dg\,(\cdot)\,\delta\left(\int d^2\xi\,\sqrt{g} - V\right) \exp\left(-S_{eff}[g]\right). \tag{2}$$

The grand canonical ensemble (or GCE) partition function is given by

$$\mathcal{Z}(\mu) = \int_0^\infty e^{-\mu V} Z(V)\, dV. \tag{3}$$

Often it is more convenient to calculate in the discretized formulation, as we do here; in this case the integral over metrics is replaced by a sum over triangulations (or dual graphs) and the volume V becomes the number of vertices N of the triangulation[9-11].

Some properties of this model are well understood, in particular for the $c < 1$ case, through matrix model calculations and the KPZ results. In this case the universe is locally two-dimensional in the normal mathematical sense. On the other hand for $c > 1$ there is an accumulation of evidence (much, but not all, of it numerical) that the space-time structure collapses to something that is locally less than two-dimensional; this is the branched polymer (BP) phase.

The Hausdorff dimension, d_h, of the BP phase can be computed analytically[12] (see also Ambjørn[8]) and it is found that $d_h = 2$. The spectral dimension, d_S, is defined by the behaviour of the diffusion equation. In the discretized formulation a random walker sets out on a graph G from a point i at time $t = 0$. Let the probability that he is back at i after t steps (each of which consists of a move from one vertex to a neighbouring vertex) be $P_G(i;\ t)$ then d_S is defined by

$$P_N(t) = \left\langle N^{-1} \sum_{i=1}^{N} P_g(i;t) \right\rangle_{CE} \sim \frac{1}{t^{d_S/2}}. \tag{4}$$

provided that $0 \ll t << N^{1/\Delta}$ for some exponent Δ so that discretization and finite size effects are avoided. The corresponding continuum quantity – the coincidence limit of the heat kernel – is in fact reparametrization invariant[13]. On the basis of extensive numerical simulations at different values of c, Ambjørn, Jurkiewicz and Watabiki[7] conjectured that $d_S = 2$ for $c < 1$ and that $d_h = 2d_S$ for all c (at least in unitary models). It is interesting to ask whether this is consistent with the large c phase being BP. If it were consistent then we should find that $d_S = 1$ for the BP phase.

CALCULATION OF d_S FOR BRANCHED POLYMERS

To calculate d_S for a pure BP ensemble it is much easier to work in the GCE where the sum over all polymers is straightforward than in the CE where the fixed N constraint makes the sum much more difficult. We therefore need to establish a relationship between GCE quantities and d_S which is more naturally defined in the CE. To explain this it is simplest to make the following ansatz for $P_N(t)$

$$P_N(t) = \frac{1}{N} + \frac{a}{t^{d_S/2}} \exp\left(\frac{t}{N^\Delta}\right). \tag{5}$$

This has the property that at very large t, when the walker can be anywhere on the finite sized system, the probabililty of finding him at a given point is simply N^{-1} whereas for $t << N^\Delta$ we get the power law behaviour $t^{-d_S/2}$. We should emphasize that this

ansatz is not *necessary* and we make it for pedagogical convenience; a derivation that does not depend on it is given in Jonsson and Wheater[14] which contains the full proofs of all the results given here. From (5) compute the generating function

$$\bar{P}_N(y) = \sum_{t=0}^{\infty} y^t P_N(t) \simeq \frac{1}{N}\frac{1}{1-y} + \frac{a}{(1-y+N^{-\Delta})^{1-d_S/2}}. \tag{6}$$

In the GCE the natural quantity to compute is

$$\mathcal{P}(z,y) = \sum_{G} z^{N_G} \sum_{t=0}^{\infty} y^t P_G(t) \tag{7}$$

where the sum runs over all graphs G in the ensemble, N_G is the number of points in G and $P_G(t)$ is the return probability on G. P is related to the CE quantity $\bar{P}_N(y)$ through

$$\mathcal{P}(z,y) = \sum_{N} z^N Z(N) \bar{P}_N(y). \tag{8}$$

We expect that for large N the CE partition function behaves as

$$Z(N) \sim N^{\gamma_{str}-2} z_0^{-N} \tag{9}$$

where z_0 is a constant and $\gamma_{str} \leq \frac{1}{2}$. Substituting (6) into (8) and dropping the pole term (which can always be computed exactly and therefore removed) to get $P'(z,y)$ we find

$$\mathcal{P}'(z,y) \sim \left(1-\frac{z}{z_0}\right)^{\beta} \Phi\left(\frac{1-y}{\left(1-\frac{z}{z_0}\right)^{\Delta}}\right) \tag{10}$$

where the prefactor exponent β is given by

$$\beta = 1 - \gamma_{str} + \Delta\left(\frac{d_S}{2} - 1\right) \tag{11}$$

and $\Phi(\upsilon)$ is analytic for $\upsilon \geq 0$ with $\Phi(0) = O(1)$. If we can compute the gap exponent Δ and the prefactor exponent β we can deduce d_S.

We will work with the simplest generic rooted branched polymer ensemble. A polymer is made by recursively stitching together two other polymers (its constituents) as shown in fig.1; the elementary polymer is a single link. If we let N_A be the number of external legs excluding the root then

$$N_{\mathrm{A}} = N_{\mathrm{B}} + N_{\mathrm{C}} \tag{12}$$

and the number of polymers with N external legs satisfies the recursion

$$\Omega_N = \sum_{M=1}^{N-1} \Omega_M \Omega_{N-M}, \qquad \Omega_1 = 1. \tag{13}$$

It is well known that this is solved by the Catalan numbers which for large N have the asymptotic behaviour

$$\Omega_N \sim \frac{4^N}{N^{\frac{3}{2}}} \tag{14}$$

from which we deduce that $z_{cr} = \frac{1}{4}$ and the well known result that $\gamma_{str} = \frac{1}{2}$.

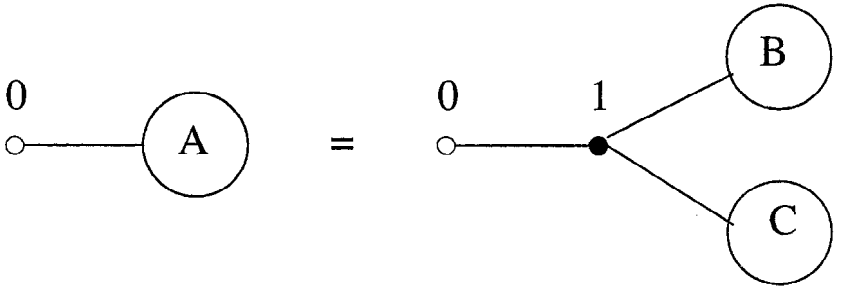


Figure 1. The fusing of polymers B and C to make a larger polymer A. The open circle is the root and the filled circle is the first vertex.

Now we want to compute the return probability for walks starting at the root (for BPs this is sufficiently general and we do not need to sum over all starting points). First let $P_{\mathrm{A}}^{1}(t)$ be the probability that the walker returns to the root *for the first time* after t steps, then $P_{\mathrm{A}}(t)$, the probability that after t steps the walker is at the root, is made up of walks which are returning for the first, second, third etc time so we have

$$P_{\mathrm{A}}(t) = \delta_{t,0} + P_{\mathrm{A}}^{1}(t) + \sum_{t_1+t_2=t} P_{\mathrm{A}}^{1}(t_1)P_{\mathrm{A}}^{1}(t_2) + \ldots + \sum_{t_1+t_2+\ldots+t_n=t} \prod_{i=1}^{n} P_{\mathrm{A}}^{1}(t_i) + \ldots . \quad (15)$$

This is a convolution so for the corresponding generating function we find

$$\begin{aligned} \bar{P}_{\mathrm{A}}^{1}(y) &= \sum_{t=0}^{\infty} y^{t/2} P_{\mathrm{A}}^{1}(t) \\ &= 1 + \bar{P}_{\mathrm{A}}^{1}(y) + (\bar{P}_{\mathrm{A}}^{1}(y))^2 + \ldots \\ &= \frac{1}{1-\bar{P}_{\mathrm{A}}^{1}(y)}. \end{aligned} \quad (16)$$

The factor $y^{t/2}$ is chosen because all returning walks on BPs have even t.

Next we relate the first return probability on A to that on its constituents B and C. A typical walk leaves the root and then goes out and back many times from the first vertex along B and C until finally returning to the root. Taking all these walks into account we get

$$\begin{aligned} P_{\mathrm{A}}^{1}(t) = {} & \frac{1}{3}\delta_{t,2} + \left(\frac{1}{3}\right)^2 \left(P_{\mathrm{B}}^{1}(t-2) + P_{\mathrm{C}}^{1}(t-2)\right) \\ & + \left(\frac{1}{3}\right)^3 \sum_{t_1+t_2=t-2} \left(P_{\mathrm{B}}^{1}(t_1) + P_{\mathrm{C}}^{1}(t_1)\right)\left(P_{\mathrm{B}}^{1}(t_2) + P_{\mathrm{C}}^{1}(t_2)\right) + \ldots \\ & \ldots + \left(\frac{1}{3}\right)^{n+1} \sum_{t_1+\ldots+t_n=t-2} \prod_{i=1}^{n} \left(P_{\mathrm{B}}^{1}(t_i) + P_{\mathrm{C}}^{1}(t_i)\right) + \ldots . \end{aligned} \quad (17)$$

Then calculating the generating function we find

$$\begin{aligned} \bar{P}_{\mathrm{A}}^{1}(y) = {} & \frac{1}{3}y + \left(\frac{1}{3}\right)^2 \left(\bar{P}_{\mathrm{B}}^{1}(y) + \bar{P}_{\mathrm{C}}^{1}(y)\right) + \left(\frac{1}{3}\right)^3 \left(\bar{P}_{\mathrm{B}}^{1}(y) + \bar{P}_{\mathrm{C}}^{1}(y)\right)^2 + \ldots \\ & \ldots + \left(\frac{1}{3}\right)^{n+1} \left(\bar{P}_{\mathrm{B}}^{1}(y) + \bar{P}_{\mathrm{C}}^{1}(y)\right)^n + \ldots \\ = {} & \frac{y}{3 - \left(\bar{P}_{\mathrm{B}}^{1}(y) + \bar{P}_{\mathrm{C}}^{1}(y)\right)}. \end{aligned} \quad (18)$$

Introducing a new quantity $h_A(y)$ defined by

$$h_A(y) = \frac{1}{1-y}\left(1 - \bar{P}_A^1(y)\right) \tag{19}$$

(despite appearances this is analytic at $y = 1$) we find that the return probability generating function in the GCE is given by

$$\mathcal{P}(z,y) = \frac{1}{1-y}\sum_B \frac{z^{N_B}}{h_B(y)} \tag{20}$$

with the recurrence relation

$$h_A(y) = \frac{1 + h_B(y) + h_C(y)}{1 + (1-y)\left(h_B(y) + h_C(y)\right)} \tag{21}$$

and for the elementary polymer $h_1(y) = 1$.

First we have to show that $P(z, y)$ has a simple pole at $y = 1$ which we can remove. To do this it is sufficient to note that h_B are non-decreasing functions and then to show that the sum in (20) is finite at $y = 1$. The recurrence (21) becomes

$$h_A(1) = 1 + h_B(1) + h_C(1) \tag{22}$$

with the initial condition for the elementary polymer $h_1(1) = 1$; but

$$N_A = N_B + N_C \tag{23}$$

so

$$h_A(1) = 2N_A - 1.$$

Now

$$\sum_B \frac{z^{N_B}}{2N_B - 1} \simeq \sum_N \frac{N^{-\frac{3}{2}} z^N}{2N - 1} \tag{25}$$

is finite at $z = z_{cr}$ so $\mathcal{P}(z, y)$ does indeed have a simple pole. We remove it by computing

$$\begin{aligned}\tilde{\mathcal{Q}}(z,y) &= -\frac{d}{dy}\left((1-y)\mathcal{Q}(z,y)\right) \\ &= \sum_{B\in\mathcal{B}} z^{N_B}\left(h_B(y)\right)^{-2}\frac{dh_B(y)}{dy}\end{aligned} \tag{26}$$

which does not have a pole at $y = 1$; $\tilde{\mathcal{Q}}$ has the same analytic structure as P' and in principal d_S can be determined from it. We proceed by computing the Taylor series of $\tilde{\mathcal{Q}}$ about $y = 1$; to show the technique we will compute the leading term in detail. From the recurrence (21) we find, differentiating once and setting $y = 1$, that

$$\begin{aligned}H^{(1)}(z) &= \sum_A z^{N_A} h'_A(1) \\ &= \sum_{B,C} z^{N_B+N_C}\left\{h'_B(1) + h'_C(1) + (1 + h_B(1) + h_C(1))(h_B(1) + h_C(1))\right\}.\end{aligned} \tag{27}$$

The first two terms on the r.h.s. of (27) reproduce the l.h.s. but with a factor of the GCE partition function. Rearranging we find

$$\begin{aligned}\sqrt{1-4z}\,H^{(1)}(z) = 2\sum_B z^{N_B}(2N_B - 1)^2 \sum_C z^{N_C} + 2\left(\sum_B z^{N_B}(2N_B - 1)\right)^2 \\ +2\sum_B z^{N_B}(2N_B - 1)\sum_C z^{N_C}.\end{aligned} \tag{28}$$

All the terms on the r.h.s. of (28) are calculable and we find that the leading behaviour (subleading terms correspond in the CE to pieces suppressed by inverse powers of N and vanish in the thermodynamic limit) as $z \uparrow z_{cr}$ is

$$H^{(1)}(z) \simeq \frac{1}{2}(1-4z)^{-2}. \tag{29}$$

Now

$$\left(2z\frac{\partial}{\partial z} - 1\right)^2 \tilde{\mathcal{Q}}(z,1) = H^{(1)}(z) \tag{30}$$

so integrating twice we get

$$\tilde{\mathcal{Q}}(z,1) \simeq -\frac{1}{8}\log(1-4z). \tag{31}$$

By iterating this kind of calculation we can show that

$$\sum_{\mathrm{B}} z^{N_B} \prod_{i=1}^{p} \frac{d^{n_i} h_{\mathrm{B}}(y)}{dy^{n_i}}\bigg|_{y=1} \simeq A^{(n_1,n_2,\ldots,n_p)}(1-4z)^{\frac{1}{2}-p-\frac{3}{2}\sum_{i=1}^{p} n_i} \tag{32}$$

where $A^{(n_1,n_2,\ldots,n_p)}$ is a constant[14]. This is enough to find the leading behaviour of

$$\left(\frac{\partial}{\partial y}\right)^n \tilde{\mathcal{Q}}(z,y)\bigg|_{y=1} \simeq (1-4z)^{-\frac{3}{2}n}. \tag{33}$$

It is in fact possible to find out enough about the coefficients $A^{(n_1,n_2,\ldots,n_p)}$ to be able to argue that the Taylor series for $\tilde{\mathcal{Q}}$ is Borel summable at $y = 1$.

From (31) we deduce that the prefactor exponent is $\beta = 0$ and from (33) that the gap exponent is $\Delta = \frac{3}{2}$. It follows that $d_S = \frac{4}{3}$.

DISCUSSION

This is a very interesting result. Firstly it shows that the conjecture that $d_h = 2d_S$ is false (it is also now known to be false for the non-unitary $c = -2$ theory[6]). However comparison with the data[7] shows that the $c = 5$ results are in very good agreement with $d_S = \frac{4}{3}$. Taken together with the excellent agreement with $\gamma_{str} = \frac{1}{2}$ and $d_h = 2$ which is also seen at $c = 5$ the evidence that the large c phase is BP is overwhelming.

The calculation presented here shows that $d_S = \frac{4}{3}$ for the simplest example of a generic branched polymer ensemble. By tracing the origin of the gap exponent Δ it is straightforward to see that the same structure will emerge for any other generic ensemble. However for the non-generic branched polymer ensembles which can be produced by allowing vertices of arbitrary order and weighting them in particular ways[16] it is possible to generate different values of γ_{str} and we expect that these ensembles will have $d_S \neq \frac{4}{3}$.

There is a close connection between our calculation and the spectral dimension on percolation clusters[17]. These authors originally conjectured that the spectral dimension on percolation clusters at criticality is $\frac{4}{3}$ independent of the embedding dimension D. High precision numerical calculations have long since shown that this is not correct for finite D (see Havlin[18] for a review). However for infinite D this result is believed to be exact and there are a number of scaling arguments and approximate calculations for it (see for example Leyvraz[19]). The BP ensemble is similar to the set of (bond) percolation clusters on a Cayley tree with the constraint that the bond from the root

to the first vertex is occupied. The difference is that the percolation clusters in general have vertices of order two as well; the ensemble can be rewritten as the BP ensemble with a sort of renormalized link[14]. Because of our interest in quantum gravity we defined d_S in terms of probabilities averaged over all clusters of size N rather than in terms of a particular "typical" cluster. The calculation picks out the leading non-analyticity as $y \to 1$ *after* taking the thermodynamic limit and therefore shows that in the thermodynamic limit a finite proportion of the T_N clusters of size N have $d_S = \frac{4}{3}$. Clusters with d_S *smaller* than $\frac{4}{3}$ are present in the ensemble but the number of such clusters is suppressed by a power of N relative to T_N and so the probability that they appear as an infinite percolation cluster is zero. It is in fact possible to rule out the appearance of clusters with d_S *larger* than $\frac{4}{3}$ so this calculation amounts to an exact calculation of $d_S = \frac{4}{3}$ for a typical cluster at criticality[14].

In some circumstances there is an extra scaling relation between Δ and d_S. On a fixed graph we have a transfer matrix T and

$$P_N(y) = \frac{1}{N} \sum_{i=1}^{N} \sum_{t=0}^{\infty} (T^t)_{ii} \tag{34}$$

letting $T = 1 - D$ where D is the (lattice) diffusion operator we get

$$P_N(y) = \frac{1}{N} \frac{1}{1-y} + \frac{1}{N} \sum_{j=1}^{N-1} \frac{1}{1-y+y\lambda_j} \tag{35}$$

where λ_j are the eigenvalues of D and we have separated the zero eigenvalue (note that it is a general theorem that $0 < \lambda_j \leq 1$). In the large N limit the non-pole part can be turned into an integral

$$\tilde{P}_N(y) = \int_{\lambda_1}^{\lambda_{N-1}} \frac{\rho(\mu)\, d\mu}{1-y+y\mu}. \tag{36}$$

Assuming that the density of states $\rho(\mu) \sim \mu^{d/2-1}$ we find that

$$\tilde{P}_N(y) \sim (y\lambda_1 + 1 - y)^{d/2-1} \tag{37}$$

ie $d = d_S$ and the behaviour at $y = 1$ is cut off by the lowest eigenvalue so we expect $\lambda_1 \sim N^{-\Delta}$. If the eigenvalues are non-degenerate it follows that $\mu \sim (jN^{-1})^{\Delta}$ and so $\rho \sim \mu^{1/\Delta - 1}$ and hence that $d_s = 2/\Delta$. This relation is obeyed both for the linear polymer (for which it is easy to show that $\Delta = 2$ and which of course has $d_S = 1$) and for the generic branched polymer. It should fail if there are degenerate eigenvalues – in this context the jth and kth eigenvalues are degenerate if $\lim_{N\to\infty}(\lambda_j - \lambda_k)/\lambda_j = 0$. So the extra scaling relation should not hold for systems which have some symmetry (eg rotational) which is restored at large distance scales but it probably does hold for objects like percolation clusters which do not.

It is interesting to ask whether these calculations can be extended to other systems, for example the pure gravity case. Unfortunately the method depends very much upon the branching nature of the BPS to obtain the recursion relation (21) between return probabilities on a polymer and those on its constituents. The graphs for pure gravity do not have this property so some other device will be needed to do the lattice calculation; of course in that case other techniques such as those of Liouville gravity are also available. There is however one potential simplification. It is tempting to suppose that $d_S = 2$ (as it appears to be for the $c = -2$ model[6]); in this case the prefactor exponent (11) becomes simply $1 - \gamma_{str}$. Running this backwards we see that if a calculation of the prefactor exponent were to give $1 - \gamma_{str}$ then it would not be necessary to compute the gap exponent in order to deduce that $d_S = 2$.

REFERENCES

1. H.Kawai, N.Kawamoto, T.Mogami and Y.Watabiki, Transfer matrix formalism for two-dimensional quantum gravity and fractal structures of space-time, Phys. Lett. B306 (1993) 19;
   M.Fukuma, N.Ishibashi, H.Kawai and M.Ninomiya, Two-dimensional quantum gravity in temporal gauge, Nucl. Phys. B427 (1994) 139.
2. J.Ambjørn and Y.Watabiki, Scaling in quantum gravity, Nucl. Phys. B445 (1995) 129.
3. S.Catterall, G.Thorleifsson, M.J.Bowick and V.John, Scaling and the fractal geometry of two-dimensional quantum gravity, Phys. Lett. B354 (1995) 58.
4. M. J.Bowick, V. John and G.Thorleifsson, The Hausdorff dimension of surfaces in two-dimensional quantum gravity coupled to Ising minimal matter, Phys. Lett. B403 (1997) 197.
5. J.Ambjørn and K.N.Anagnostopoulos, Quantum geometry of 2D gravity coupled to unitary matter, Nucl. Phys. B497 (1997) 445.
6. J.Ambjørn et al, The quantum space-time of c=-2 gravity, Niels Bohr Institute preprint NBI-HE-97-19, hep-lat/9706009.
7. J.Ambjørn, J.Jurkiewicz and Y.Watabiki, On the fractal structure of two-dimensional quantum gravity, Nucl. Phys. B454 (1995) 313.
8. J.Ambjørn, this volume.
9. J.Ambjørn, B.Durhuus and J.Fröhlich, Diseases of triangulated random surface models, and possible cures, Nucl. Phys. B 257 [FS14](1985) 433.
10. V.A.Kazakov, I.K.Kostov and A.A.Migdal, Critical properties of randomly triangulated planar random surfaces, Phys. Lett. 157B (1985) 295.
11. F.David, A model of random surfaces with non-trivial critical behaviour, Nucl. Phys. B 257 [FS14](1985) 543.
12. J.Ambjørn, B.Durhuus and T.Jonsson, Summing over all genera for $d > 1$: a toy model, Phys. Lett. B244 (1990) 403.
13. B.S. Dewitt, in *General Relativity, A Centenary Survey*, eds. S.W.Hawking and W.Israel, Cambridge University Press, 1979.
14. T.Jonsson and J.F.Wheater, The spectral dimension of the branched polymer phase of two-dimensional quantum gravity, Oxford University preprint OWTP-97-33-P.
15. F.David, A scenario for the $c > 1$ barrier in non-critical bosonic strings, Nucl. Phys. B487 (1997) 633.
16. P.Bialas and Z.Burda, Phase transition in fluctuating branched geometry, Phys. Lett. B384 (1996) 75.
17. S.Alexander and R.Orbach, Density of states on fractals: fractons, J.Physique Lettres 43 (1982) 625.
18. S.Havlin and D.Ben-Avraham, Diffusion in disordered media, Adv. Phys. 36 (1987) 695.
19. F.Leyvraz and H.E.Stanley, To what class of fractals does the Alexander-Orbach conjecture apply?, Phys. Rev. Lett. 51 (1983) 2048.

SOLVING THE BAXTER EQUATION IN HIGH ENERGY QCD

J. Wosiek

Institute of Physics
Jagellonian University
Reymonta 4
30-059 Cracow, Poland

INTRODUCTION

Quantitative description of the reggeization of QCD still remains a challenge for the Leading Logarithmic scheme and its extensions [1,2]. In the first approximation the problem naturally separates into sectors with fixed number n of the reggeized gluons propagating in the t channel. The lowest nontrivial case, $n = 2$, was solved in the classical papers by Balitskii, Kuraev, Fadin and Lipatov[3] who derived the simple expression for the intercept of the hard pomeron. Further progress for arbitrary n was achieved by Lipatov, Faddeev and Korchemsky [4,5,6] who have established exact equivalence with the one dimensional chain of n noncompact spins. Leading high energy behaviour of QCD amplitudes is given by the highest eigenvalue of the corresponding Heisenberg hamiltonian of n spins with nearest neighbour interaction. Moreover, by identifying enough constants of motion they were able to prove that this system is solvable for arbitrary n. The success of this, rather mathematical, approach was confirmed by rederiving the Lipatov et al. result in the $n = 2$ case [6,7]. However, the adopted procedure requires an analytic continuation from the integer values of the relevant conformal weight h (see later) because only for integer h they were able to diagonalize the two spin hamiltonian. The $n = 3$ case, which gives the lowest contribution to the odderon exchange, was also studied by Faddeev and Korchemsky, and Korchemsky[6,7] . Again, the spectrum of the system for integer h can be found for any finite $h = m$. However, the general expression for arbitrary m is not known, and consequently the analytical continuation to $h = 1/2$ is not available *.

We have developed [8,9], together with R. A. Janik, a new approach which a) works for arbitrary values of the conformal weight h, providing explicitly above continuation, and b) gives, for the first time, the analytic solution of the $n = 3$ case for arbitrary h and q_3 . For $n = 2$ our method reproduces again the *BFKL* result clarifying the problem of boundary conditions for arbitrary h. In this talk I will discuss some details of our $n = 2$ calculation emphasizing general features of this approach which are directly applicable to higher n cases. The results for $n = 3$ will be shortly summarized.

*The lowest state of the $n = 3$ hamiltonian is believed to occur at $h = 1/2$.

New Developments in Quantum Field Theory
Edited by Damgaard and Jurkiewicz, Plenum Press, New York, 1998

We rely on the formalism developped in Refs.[4, 5, 6, 7] and follow the conventions and notation of Refs.[6, 7].

THE POMERON CHANNEL

The Formalism

The intercept of the Pomeron trajectory is given by

$$\alpha_P(0) = 1 + \frac{\alpha_s N_c}{4\pi}\left(\epsilon_2(h) + \bar{\epsilon}_2(\bar{h})\right), \tag{1}$$

where ϵ_2 and $\bar{\epsilon}_2$ are respectively the largest eigenvalues of the $n = 2$ reggeon hamiltonian and its antiholomorphic counterpart [6, 7]. This system is equivalent to the misleadingly simple set of the two noncompact spins which for higher n generalizes to the one dimensional chain with nearest-neighbour interactions. Applying Bethe Ansatz to the latter one obtains in the $n = 2$ case

$$\epsilon_2 = i\left(\frac{\dot{Q}_2(-i)}{Q_2(-i)} - \frac{\dot{Q}_2(i)}{Q_2(i)}\right) - 4, \tag{2}$$

where $Q_2(\lambda)$ satisfies the following Baxter Equation

$$\hat{BE}(\lambda, q_2)Q_2 \equiv (\lambda + i)^2 Q_2(\lambda + i) + (\lambda - i)^2 Q_2(\lambda - i) - (2\lambda^2 + q_2)Q_2(\lambda) = 0, \tag{3}$$

q_2 is the eigenvalue of the square of the total spin of the system $\hat{q}_2$. It commutes with the hamiltonian and its spectrum is known from the symmetry considerations

$$q_2 = h(1-h), \quad h = \frac{1}{2}(1+m) - i\nu, \quad m \in Z, \nu \in R. \tag{4}$$

In order to solve the Baxter Equation (3) the following integral representation is customarily used, $K(z, \lambda) = z^{-i\lambda-1}(z-1)^{i\lambda-1}$

$$Q_2(\lambda) = \int_C K(z, \lambda)Q(z)\frac{dz}{2\pi i}. \tag{5}$$

Eq. (3) is equivalent to the simple hypergeometric equation for the transform $Q(z)$

$$\hat{HE}(z, q_2)Q \equiv \left[\frac{d}{dz}z(1-z)\frac{d}{dz} - q_2\right]Q(z) = 0, \tag{6}$$

provided the boundary term, which can be read off from the identity

$$\hat{BE}(\lambda, q_2)Q_2(\lambda) = \int_C \hat{HE}(z, q_2)Q(z)\frac{dz}{2\pi i} \tag{7}$$

$$z(1-z)[K'(z,\lambda)Q(z) - K(z,\lambda)Q'(z)]\Big|_{z_{start}}^{z_{end}}, \tag{8}$$

vanishes. The prime denotes the derivative over z and $z_{start(end)}$ are the start- and end-points of the contour C.

For arbitrary value of the conformal weight h, however, the singularity structure of the hypergeometric functions together with the nontrivial monodromy of the kernel $K(z, \lambda)$ precludes existence of the contour such that the boundary contributions cancel. Nevertheless, for integer $h = m$, the solution regular at $z = 0$ does not have a cut and consequently the simple contour encircling both $z = 0$ and $z = 1$ points guarantees

vanishing of the boundary terms. This observation was exploited in Refs.[6,7] leading to the elegant solution of the $n = 2$ problem for integer conformal weight. The BFKL formula resulted after the analytic continuation in h to $h = 1/2$. However, the case of noninteger h requires further insight. In particular the boundary conditions for $Q_2(\lambda)$ are not fully understood. For integer h, again, they can be deduced from the polynomial Bethe ansatz and are consistent with the above choice of the integration contour in Eq.(5). For arbitrary h, they are not available. It would be very instructive to investigate the so called functional Bethe ansatz in this connection.

The Method

We will present here a different approach . Instead of Eq.(5) we use the *double contour* representation (c.f. Fig.1) considered in Ref.[10],

$$Q_2(\lambda) = \int_{C_I} z^{-i\lambda-1}(z-1)^{i\lambda-1}Q_I(z)\frac{dz}{2\pi i} + \int_{C_{II}} z^{-i\lambda-1}(z-1)^{i\lambda-1}Q_{II}(z)\frac{dz}{2\pi i}, \tag{9}$$

expecting that additional of the second solution freedom would alleviate the problem of the choice of the contour.

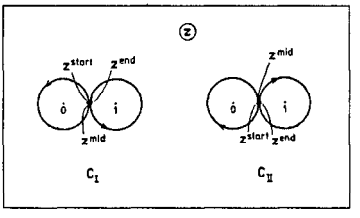


Figure 1. Integration contours used in Eq.(9). Start z^{start}, middle z^{mid}, and end z^{end} points coincide but they lie on the different sheets of the Riemann surface of the integrands.

We begin with the general solutions of Eq.(6) and then show how the initial freedom is restricted which leads to the unique solution. To this end we write the two fundamental sets of two, linearly independent solutions of Eq.(6)

$$\vec{u}(z) = (u_1(z), u_2(z)), \quad \vec{v}(z) = (v_1(z), v_2(z)), \tag{10}$$

around $z = 0$ and $z = 1$ respectively.

$$u_1(z) = F(h, 1-h, 1; z) = \sum_{n=0}^{\infty} f_n z^n \equiv F(z),$$
$$u_2(z) = -i\,\ln(z)F(z) + iG(z), \tag{11}$$

with $G(z)$ regular at $z = 0$

$$G(z) = \sum_{n=0}^{\infty} g_n z^n, \tag{12}$$

where $F(a, b, c; z)$ is the hypergeometric function, $f_n = h^{(n)}(1-h)^{(n)}/n!^2$, $g_n = f_n[2\psi(n+1) - \psi(n+h) - \psi(n+1-h)]$, $\psi(z)$ denotes the digamma function

and $\not{s} \equiv \sin(\pi h)/\pi$. The series in Eqs.(11) are convergent in the unit circle K_0 around $z = 0$. Similarly one can construct the $\vec{v}(z)$ solutions in the unit circle K_1 around $z = 1$. In fact, because of the symmetry of Eq.(6)under the transformation $z \to 1-z, q_2 \to q_2$, we take

$$v_1(z) = iu_1(1-z), \quad v_2(z) = -iu_2(1-z). \tag{13}$$

Usually only one basis is sufficient. However for higher n one can construct the bases only in the form of the series convergent in K_0 or K_1, and hence we will need both bases to integrate Eq.(9).

Since any solution is a linear combination of the fundamental solutions, we have in general

$$\begin{aligned} Q_I(z) &= au_1(z) + bu_2(z) \\ &\equiv A \cdot \vec{u}(z) = A \cdot \Omega\vec{v}(z), \\ Q_{II}(z) &= cu_1(z) + du_2(z) \\ &\equiv B \cdot \vec{u}(z) = B \cdot \Omega\vec{v}(z), \end{aligned} \tag{14}$$

with an obvious vector notation. The transition matrix Ω is defined by

$$\vec{u}(z) = \Omega\vec{v}(z), \tag{15}$$

and provides the analytic continuation of our solutions $Q(z)$ between K_0 and K_1. It plays an important role for higher n and its direct calculation for $n > 2$ is rather nontrivial. For the hypergeometric equation, and for the special choice of both bases, Eqs.(11,13), Ω is very simple. Due to the identity [11]

$$F(z) = -\not{s}\mathrm{Log}(1-z)F(1-z) + G(1-z), \tag{16}$$

$$\mathrm{Log}(z) = |z| + i\mathrm{Arg}(z), \ |\mathrm{Arg}(z)| < \pi, \tag{17}$$

$u_2(z) = iu_1(1–z)$, and (13) implies

$$\Omega = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{18}$$

Next, we introduce the monodromy matrix M_u, which describes the behaviour of the

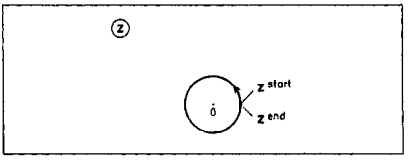


Figure 2. Closed contour used to define the monodromy matrix, Eq. (19). $z^{start} = z^{end}$, however they belong to the different sheets of the Riemann surface.

basis $\vec{u}$ in the vicinity of the branch point $z = 0$ (see Fig.2).

$$\vec{u}(z_{end}) = M_u\vec{u}(z_{start}), \quad M_u = \begin{pmatrix} 1 & 0 \\ 2\not{s} & 1 \end{pmatrix}, \tag{19}$$

and symmetrically ($z_{start} = 1/2 - i\epsilon, z_{end} = 1/2 + i\epsilon$) for the v basis. It is easy to see that $M_v = M_u^{-1}$.

We are now ready to write the condition for the cancellation of the boundary contributions in Eq.(9). With the choice of the contours C_I and C_{II} as shown in Fig.1, and with the apropriate choice of the branch of the kernel $K(z,\lambda)$ (see below) the boundary contributions cancel if

$$A^T M_I + B^T M_{II} = 0, \tag{20}$$

where the combined monodromy matrices † for the corresponding contours read

$$M_I = \Omega M_v \Omega^{-1} - M_u^{-1}, \quad M_{II} = \Omega M_v^{-1} \Omega^{-1} - M_u. \tag{21}$$

In terms of the coefficients, condition (20) reads simply

$$a = c, \quad b = d. \tag{22}$$

Hence the original freedom of four coefficients in Eqs.(14) was reduced to the two free parameters. In fact the energy of the system, Eq. (2), is insensitive to the absolute normalization, hence only the ratio

$$\rho = a/b, \tag{23}$$

remains relevant. This variable parametrizes all possible boundary conditions which are consistent with the cancellation of the end-point contributions in the sum (9). The role of remaining freedom is better seen when the explicit result for ϵ_2 is derived. Before integrating Eq.(9) consistent choice of the branches of the kernel $K(z,\lambda)$ and of $Q_K(z)$, $K = I, II$; must be made, as explained below.

Details of the Analytic Structure

General solutions $Q_K(z)$, $K = I, II$; of the hypergeometric equation (6) have a nontrivial monodromy at $z = 0,1$ and ∞. Therefore one has to supplement the generic formulas (11,13) with the chioce of cuts and corresponding branches of the multivalued functions $Q_K(z)$, as well as the kernel $K(z,\lambda)$. We define

$$\begin{aligned} u_1(z) &= F(z), \\ u_2^I(z) &= -i\,\pi\,\mathrm{llog}(z)F(z) + iG(z), \\ u_2^{II}(z) &= -i\,\pi\log(z)F(z) + iG(z), \quad |z| < 1, \end{aligned} \tag{24}$$

for the $\vec{u}$ basis, and

$$\begin{aligned} v_1(z) &= iF(1-z), \\ v_2^I(z) &= -\,\pi\log(1-z)F(1-z) + G(1-z), \\ v_2^{II}(z) &= -\,\pi\,\mathrm{llog}(1-z)F(1-z) + G(1-z), \quad |1-z| < 1, \end{aligned} \tag{25}$$

for the $\vec{v}$ basis. Different branches of $u_2(z)$ and $v_2(z)$, labelled by the superscript $K = I, II$; are chosen on the two contours C_I and C_{II} respectively. They are realized by the different choice of the branch of the logarithm according to the definitions

$$\log(z) = \begin{cases} \mathrm{Log}(z), & Im(z) > 0, \\ \mathrm{Log}(z) + 2\pi i, & Im(z) < 0, \end{cases} \tag{26}$$

$$\mathrm{llog}(z) = \begin{cases} \mathrm{Log}(z) - 2\pi i, & Im(z) > 0, \\ \mathrm{Log}(z), & Im(z) < 0, \end{cases} \tag{27}$$

†Reduced to the middle point z_m.

where Log(z) denotes the main branch of the logarithm, cf.(17). With the above choice of branches identity (16) implies

$$(u_1(z), u_2^I(z)) = (F(z), iF(1-z)) = (v_2^I(z), v_1(z)), \tag{28}$$

on the *lower* ($Im(z) < 0$) half of the contour C_I, and

$$(u_1(z), u_2^{II}(z)) = (F(z), iF(1-z)) = (v_2^{II}(z), v_1(z)), \tag{29}$$

on the *upper* half of C_{II}. Therefore the transition matrix Ω (18) provides required analytic continuation between the two bases along the central sections the contours C_K. In particular at the respective middle points $z_m^I = 1/2 - i\epsilon$ and $z_m^{II} = 1/2 + i\epsilon$ relation (15) holds provided corresponding branches $u_2^K(z)$ and v_2^K, $K = I, II$; are used. Explicit expressions for the transforms Q_K read

$$\begin{aligned} Q_I(z) &= \begin{cases} au_1(z) + bu_2^I(z), & \text{if } z \in C_I^{(0)}, \\ av_2^I(z) + bv_1(z), & \text{if } z \in C_I^{(1)}, \end{cases} \\ Q_{II}(z) &= \begin{cases} cu_1(z) + du_2^{II}(z), & \text{if } z \in C_{II}^{(0)}, \\ cv_2^{II}(z) + dv_1(z), & \text{if } z \in C_{II}^{(1)}. \end{cases} \end{aligned} \tag{30}$$

Where the relations (22) have not yet been used. $C_K^{(0)}$ and $C_K^{(1)}$ denote the unit circles around $z = 0$ and $z = 1$ which compose contours C_K.

Our conventions define the following analytic structure of the transforms $Q_K(z)$ Consider the complex z plane with the two cuts $|\mathrm{Arg}(z)| < \pi$ and $|\mathrm{Arg}(1-z)| < \pi$. On that plane the "main" branch of the fundamental solution $(F(z), iF(1-z))$ is defined. At the points $z = 1/2$ and $z = 3/2$, where the contours C_K cross these cuts, $Q_K(z)$ are continuous due to the appropriate choice of log's and llog's in Eqs.(24,25). In another words, our contours lie on the different Riemann sheets, which coincide with the the "main" sheet for lower (upper) part of $C_I(C_{II})$ and then, at $z = 1/2$ and $z = 3/2$, they join smoothly to another (different for different contours) sheets guaranteeing continuity of $Q_K(z)$. It follows from Eqs.(28,29) that, with the choice of the coefficients as in Eq. (30), $Q_K(z)$ are continuous along the contours C_K at the respective middle points z_m^K, $K = I, II$. Other choices of branches are also possible, however the continuity of the solutions along the contours must be assured, by appropriate adjustment of the coefficients in expansions (30).

Finally we discuss the choice of the apropriate branch of the kernel $K(z, \lambda)$. As mentioned above the Baxter Equation (3) is equivalent to the differential equation (6) provided the boundary term present in the identity (8) vanishes. This term contains the end point values of the kernel and of the transform $Q(z)$ together with their derivatives. Moreover, in the two-contour case, the end point contributions from both contours add. By the suitable choice of the branch of the kernel $K(z, \lambda)$ we can decouple its end-point values from the boundary term thus simplyfing the analysis. We choose

$$K_I(z,\lambda) = \exp[(-i\lambda - 1)\log(z) + (i\lambda - 1)\log(z-1)], \text{ if } z \in C_I \tag{31}$$

$$K_{II}(z,\lambda) = \exp[(-i\lambda - 1)\mathrm{llog}(z) + (i\lambda - 1)\log(z-1)], \text{ if } z \in C_{II} \tag{32}$$

With this definition the end-point values of the kernel, and its derivative, are the same for both contours, i.e.

$$K_I(z_{start}, \lambda) = K_I(z_{end}, \lambda) = K_{II}(z_{start}, \lambda) = K_{II}(z_{end}, \lambda), \tag{33}$$

$$K_I'(z_{start}, \lambda) = K_I'(z_{end}, \lambda) = K_{II}'(z_{start}, \lambda) = K_{II}'(z_{end}, \lambda), \tag{34}$$

With this choice the vanishing of the boundary term is equivalent to the condition

$$Q_I(z) + Q_{II}(z)|_{z_{start}}^{z_{end}} = 0,$$
$$Q'_I(z) + Q'_{II}(z)|_{z_{start}}^{z_{end}} = 0. \tag{35}$$

These conditions are satisfied if the algebraic relations (20) among coefficients a, b, c, d are fulfilled.

The Integrals

Analytic calculation of the function $Q_2(\mathrm{l})$ is not avaiable for arbitrary l. However at $\lambda = \pm i$, needed in Eq.(2), the kernel $K(z, \lambda)$ simplifies, and in the consequence the integrals can be readily performed. For illustration we will calculate explicitly $Q(i)$. Substituting Eqs.(30) with (22) and $K_I(z,i) = K_{II}(z,i) = (1-z)^{-2}$ into (9) one obtains

$$Q_2(i) = \int_{C_I^{(0)}} (au_1 + bu_2^I)K + \int_{C_I^{(1)}} (bv_1 + av_2^I)K \tag{36}$$
$$+ \int_{C_{II}^{(0)}} (au_1 + bu_2^{II})K + \int_{C_{II}^{(1)}} (bv_1 + av_2^{II})K. \tag{37}$$

where the measure $dz/2\pi i$ is understood. Since contours C_I and C_{II} run in the opposite directions, integrals of the single valued solutions u_1 and v_1 cancell out giving

$$Q_2(i) = b \int_{C^{(0)}} (u_2^I - u_2^{II})K + a \int_{C^{(1)}} (v_2^I - v_2^{II})K. \tag{38}$$

The differences of two branches are simple, cf. Eqs.(24,25)

$$Q_2(i) = ib \oint_{C^{(0)}} F(z) \frac{dz}{(1-z)^2} - a \oint_{C^{(1)}} F(1-z) \frac{dz}{(1-z)^2}. \tag{39}$$

The first integral vanishes since the integrand is analytic inside $C^{(0)}$, while the second is given by the residue at $z = 1$ resulting in the final expression

$$Q_2(i) = 2ia \sin(\pi h) f_1, \tag{40}$$

with $f_1 = h(1-h)$. Similar steps at $\mathrm{l} = -i$ give

$$Q_2(-i) = -2b \sin(\pi h) f_1. \tag{41}$$

Calculation of the derivatives of the $Q(\lambda)$ is more tedious but proceeds analogously. We obtain

$$\dot{Q}_2(i) = -2a \sin(\pi h)(1 + g_1) - 2\pi i(b + a \sin(\pi h)) f_1, \tag{42}$$
$$\dot{Q}_2(-i) = 2\pi(a + b \sin(\pi h)) f_1 + 2ib \sin(\pi h)(1 + g_1), \tag{43}$$

where $g_1 = f_1(2\psi(2) - \psi(1+h) - \psi(2-h))$.

Results in the Pomeron Channel

Substitutining results (40-43) into (2) gives after some algebra

$$\epsilon_2(h) = 4\psi(1) - 2\psi(h) - 2\psi(1-h)$$
$$- \frac{i\pi}{s}(\rho - \rho^{-1}). \tag{44}$$

with ρ given by Eq.(23).

It is instructive to compare this expression with the original hamiltonian of the two spins [7]

$$\hat{\mathcal{H}}_2 = 4\psi(1) - 2\psi(-\hat{J}_{12}) - 2\psi(1+\hat{J}_{12}). \tag{45}$$

where the eigenvalues of $\hat{J}_{12}$ are equal to $-h$ c.f. Eq.(4). It is now evident that the choice

$$r = \pm 1, \tag{46}$$

gives the correct spectrum of energies. We emphasize, however, that the additional information was required to fix the remaining freedom. This is different in the $n = 3$ case (see below). It is important to note that the above choice is independent of h which *a priori* is not guaranteed.

Substituting Eq.(44), with (46), in Eq.(l), and setting $h = \bar{h} = 1/2$, we reproduce the BFKL formula

$$\alpha_P(0) = 1 + \frac{\alpha_s N}{\pi} 4\mathrm{Log}2. \tag{47}$$

This was also obtained in Ref.[7] after analytic continuation of their result from integer values of h. The difference between both approaches is best seen by comparing Eq. (44) with Eq.(6.31) of Ref.[7]. It follows from the form of the hamiltonian, Eq.(45), that the complete holomorphic eigenenergy $\in_2(h)$ is singular also at positive integer h. This is true for our result, Eq.(44), while that of Ref.[7] is finite, In fact their approach gives only the *real* part of $\in_2$. This is sufficient to reproduce the intercept (1), however does not give the correct analytic structure in the individual holomorphic sectors.

We emphasize that above calculation was organized from the point of view of the generalization to higher number of reggeized gluons. Such elements as the analytic structure of the solutions, series expansions of both bases, the transition matrix, choice of the branches and integration term by term, are the same when solution of higher hamiltonians is attempted.

THE ODDERON CHANNEL, N=3

For $n = 3$ we have carried out this procedure explicitly [8]. The complete set of linearly independent solutions of the corresponding third order differential equation was constructed. The transition matrix between the $\vec{u}$ and $\vec{v}$ bases was also obtained. Since in this case there is no simple identity connecting linearly independent solutions, the Ω matrix is nontrivial. Remarkably it turns out that the condition for cancellation of the end-point contributions in the double integral representation determines *uniquely* the final solution of the Baxter equation. Existing arbitrariness in both transform $Q_{I/II}(z)$ is irrelevant. Consequently we have obtained the holomorphic (and antiholomorphic) energies as the analytic function of the two relevant parameters h and q_3. The new variable q_3 is the eigenvalue of the second, commuting with hamiltonian, observable $\hat{q}_3$. Our formula reproduces exactly known values of $\in_3(h, q_3)$ in the polynomial case. It also agrees with the asymptotic calculations of Korchemsky [12] performed in the limit $h \to \infty$, $q_3/h^3 = const.$, see Fig.3.

The spectrum of $\hat{q}_3$ is not known spite of many interesting attempts [5, 13, 14, 15.] We have therefore mapped numerically the analytic structure of $\in_3(1/2, q_3)$ in the complex q_3 plane. It turns out he holomorphic energy has a series of poles at imaginary q_3 ‡.The intercept of the odderon trajectory is smaller than one for almost all values

‡ Our definition of q_3 is the same as in Ref.[7]

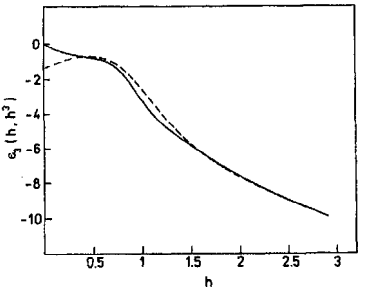


Figure 3. Comparison of the exact result for the three regeon energy (solid line) with the asymptotic formula (dashed line) of Ref.[12].

of q_3 including all $q_3 \in R$. However in the vicinities of the poles it can be arbitrarily large. Therefore any further conclusion about the numerical value of the $\alpha_o(0)$ depends crucially on the spectrum of q_3.

Acknowledgements

We would like to thank L. N. Lipatov and G. P. Korchemsky for interesting discussions. This work is supported by the Polish Committee for Scientific Research under the grants no PB 2P03B19609 and PB 2P03B04412.

REFERENCES

1. H. T. Nieh and Y. -P. Yao, *Phys. Rev. Lett.* 32:1074 (1974); *Phys. Rev.* D13:1082 (1976). B. McCoy and T.T.Wu, *Phys. Rev.* D12:3257 (1975).
2. J. Kwieciński and M. Praszalowicz, *Phys. Lett.* B94:413 (1980); J. Bartels, *Nucl. Phys.* B175:365 (1980); H. Cheng, J. Dickinson, C. Y. Lo and K. Olaussen, *Phys. Rev.* D23:534 (1981).
3. E. A. Kuraev, L. N. Lipatov and V. S. Fadin, *Phys. Lett.* B60:50 (1975); *Sov. Phys. JETP* 44:443 (1976); ibid: 45:199 (1977); Ya. Ya. Balitskii and L. N. Lipatov, *Sov. J. Nucl. Phys.* 28:822 (1978).
4. L. N. Lipatov, Padova preprint DFPD/93/TH/70 *unpublished;* L. N. Lipatov, *Sov. Phys. JETP Lett.* 59:571 (1994).
5. L. N. Lipatov, *Phys. Lett.* B309:394 (1993).
6. L. D. Faddeev and G. P. Korchemsky, *Phys. Lett.* B342:311 (1994).
7. G. P. Korchemsky, *Nucl. Phys.* B443:255 (1995).
8. J. Wosiek and R. A. Janik, Jagellonian University preprint TPJU-21/96, hep-th/9610208.
9. J. Wosiek and R. A. Janik, in: "Proceedings of the ICHEP96 Conference", Z. Ajduk and A.K. Wróblewski, ed., World Scientific (1997); hep-th/9611025.
10. R. A. Janik, *Acta Phys. Polon.* B27:1819 (1996).
11. M. Abramowitz and I. Stegun, eds., "Handbook of Mathematical Functions", Dover Pub., New York (1972).
12. G. P. Korchemsky, *Nucl. Phys.* B462:333 (1996).
13. P. Gauron, L. N. Lipatov and B. Nicolescu, *Phys. Lett.* B304:334 (1993); *Zeit. Phys.* C63:253 (1994).
14. R. A. Janik, *Phys. Lett.* B371:293 (1996); hep-th/9511210.
15. G. P. Korchemsky, Orsay preprint LPTHE-ORSAY-96-76, hep-th/9609123.

This page intentionally left blank

LIST OF PARTICIPANTS

1. J. Ambjorn,
   Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen, Denmark;
   e-mail : ambjorn@nbivax.nbi.dk

2. H. Arodź.,
   Jagellonian University, Institute of Physics, Reymonta 4,
   PL-30-059 Krakow, Poland;
   e-mail : ufarodz@thrisc.if.uj.edu.pl

3. K. Bering,
   Uppsala University, Inst. f. Theor. Phys., P.O. BOX 803,
   S-751 08 Uppsala, Sweden;
   e-mail : bering@nbivms.nbi.dk

4. P. Bialas,
   Jagellonian University, Institute of Comp. Sciences, Reymonta 4,
   PL-30-059 Krakow, Poland;
   e-mail : pbialas@thrisc.if.uj.edu.pl

5. B. Broda,
   University of Łódź, Dep. of Theor. Phys., Pomorska 149/153,
   PL–90-236 Łódź, Poland;
   e-mail : bobroda@mvii.uni.lodz.pl

6. B. Brzezicka,
   The Institute of Nuclear Physics, Radzikowskiego 152, PL-31-342 Krakow, Poland;
   e-mail : brzezicka@bron.ifj.edu.pl

7. Z. Burda,
   Jagellonian University, Institute of Physics, Reymonta 4,
   PL-30-059 Krakow, Poland;
   e-mail : burda@thrisc.if.uj.edu.pl

8. V. Călian,
   University of Craiova, Dept. of Phys., 13 A.I. Cuza, Craiova-1100, Romania;
   e-mail : violetax@radiotel.ro

9. P.H. Damgaard,
   Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen, Denmark;
   e-mail : phdamg@hetws5.nbi.dk

10. K. Gawedzki,
IHES, 91440 Bures-sur-Yvette, France;
e-mail : bongo@ihes.fr

11. J. Greensite,
San Francisco State University, Phys. and Astronomy Dept.,
San Francisco, CA 94132, USA;
e-mail : greensit@stars.sfsu.edu

12. L. Hadasz,
Jagellonian University, Institute of Physics, Reymonte 4, PL-30-059 Krakow, Poland;
e-mail : hadasz@thp1.if.uj.edu.pl

13. P. Haagensen,
Massachusetts Institute of Technology, Center for Theor. Phys.,
Cambridge, MA 02139, USA;
e-mail : haagense@ctpa01.mit.edu

14. M. Halpern,
University of California, Dept. of Phys., Berkeley, CA 94720, USA;
e-mail : halpern@physics. berkeley.edu

15. R. Janik,
Jagellonian University, Institute of Physics, Reymonte 4, PL-30-059 Krakow, Poland;
e-mail : ufrjanik@jetta.if.uj.edu.pl

16. Z. Jaskólski,
Wroclaw University, Inst. of Theor. Phys.,
pl. Maksa Borna 9, PL-50-204 Wroclaw, Poland;
e-mail : jaskolsk@ift.uni.wroc.pl

17. T. Jonsson,
University of Iceland, Dunhaga 3, 107 Reykjavik, Iceland;
e-mail : thjons@raunvis.hi.is

18. J. Jurkiewicz,
Jagellonian University, Institute of Physics, Reymonta 4,
PL-30-059 Krakow, Poland;
e-mail : jurkiewi@thp1.if.uj.edu.pl, jurkiewicz@nbi.dk

19. V. Kazakov,
Ecole Normale Superiure, Lab. de Phys. Theor.,
24 rue Lhomond, 75005 Paris, France;
e-mail : kazakov@physique.ens.fr

20. J. Kijowski,
Center for Theoretical Physics, Polish Academy of Sciences,
Aleja Lotnikow 32/46, PL-02-668 Warszawa, Poland;
e-mail : kijowski@ifpan.edu.pl

21. A. Marshakov,
Theory Dept., P.N. Lebedev Phys. Inst., Russian Academy of Sciences,
53 Leninsky prosp., 117924 Moscow, Russia;
e-mail : mars@lpi.ac.ru

22. K. Meissner,
Th. Div., CERN, CH-1211 Geneve 23, Switzerland;
e-mail : krzysztof.meissner@cern.ch

23. T. Morris,
University of Southampton, Phys Dept., Highfield, Southampton, SO17 1BJ, United Kingdom;
e-mail : trmorris@hep.ph.soton.ac.uk

24. S. Nishigaki,
Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen, Denmark;
e-mail : shinsuke@alf.nbi.dk

25. M.A. Nowak,
GSI, Planckstr. 1, D-64291 Darmstadt, Germany;
e-mail : nowak@tpri6n.gsi.de

26. P. Olesen,
Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen, Denmark;
e-mail : polesen@nbi.dk

27. P. Orland,
City Univ. of New York, Baruch College, 17 Lexington Ave,
New York, NY 10010, USA;
e-mail : orland@gursey. baruch.cuny.edu

28. M. Praszalowicz,
Jagellonian University, Institute of Physics, Reymonta 4,
PL-30-059 Krakow, Poland;
e-mail : michal@thrisc.if.uj.edu.pl

29. A. Semikhatov,
Theory Dept., P.N. Lebedev Phys. Inst., Russian Academy of Sciences,
53 Leninsky prosp., 117924 Moscow, Russia; Tel.:
e-mail : asemikha@td.lpi.ac.ru

30. A. Staruszkiewicz,
Jagellonian University, Institute of Physics, Reymonta 4,
PL-30-059 Krakow, Poland;
no e-mail

31. J. Verbaarschot,
SUNY at Stony Brook, Dept. of Phys., Stony Brook, NY 11794, USA;
e-mail : verbaarschot@nuclear.physics.sunysb.edu

32. P. Wegrzyn,
Jagellonian University, Institute of Physics, Reymonta 4,
PL-30-059 Krakow, Poland;
wegrzyn@thp1.if.uj.edu.pl

33. J. Wheater,
University of Oxford, Theor. Phys., 1 Keble Road, Oxford, OX1 3NP, United Kingdom;
e-mail : j.wheater1@physics.oxford.ac.uk

34. J. Wosiek,
Jagellonian University, Institute of Physics, Reymonta 4, PL-30-059 Krakow, Poland;
wosiek@thp1.if.uj.edu.pl

35. J. Zinn-Justin,
Saclay, Dept. de Phys. Theor., Ormes des Merisiers, F-91191 Gif-sur-Yvette, France;
e-mail : zinn@spht.saclay.cea.fr

INDEX

Affine Lie algebra, 81

Baierlein–Sharp–Wheeler action, 35
Banks–Casher relation, 191
Baxter equation, 349
Bethe ansatz, 99
BFV theorem, 243
Bianchi identites, 70, 84
Borel transformation, 221
BPS state, 139
Branched polymer, 127, 251, 341
BRST transformation, 244

Callan–Symanzik β-function, 220
Center dominance, 47, 49
Central charge, 88, 89
Chemical potential μ, 209
Chiral condensate,188
Confinement, 47
Conformal field theory, 79
Conformal mapping, 221, 304
Cosmological constant, 43
Creutz ratio, 51, 54, 55
Critical exponent, 217
C-theorem, 156

D-branes, 279
Dirac spectra, 187
Domain wall, 233
Dp-brane, 129, 137
Duality, 65, 201 263
Dynamo effect, 160

Early universe, 160
Effective action, 148
Electric charge, 182, 184
Euclidean signature, 39
Euler number, 177

Fractal dimension, 1, 6, 13
Friedman–Robertson–Walker universe, 44

Gauge invariance, 116
Geodesic distance, 6, 17, 36
Green–Schwartz action, 134, 144

Hartle–Hawking Wave-functional, 5, 7, 11
Hausdorff dimension, 5, 13, 342
Highest-weight representation, 330
Highest-weight vector, 329
Hilbert space, 115, 122, 124
Hirota equation, 97
Holomorphy, 298
Hopf equation, 109

IKKT matrix model, 135, 136
Inflation, 161
Instanton, 136
Ising model, 154, 222

Killing metric, 84
Kolmogorov theory, 25, 27
Kraichnan model, 26, 27

Lagrangian trajectories, 28
Lamé equation, 324
Lapse function, 35
Lattice QCD, 204
Light cone, 179, 274
Liouville theory, 15
Lorentz group, 180
Lorentzian signature, 39
Lyapunov exponent, 30

M theory, 128
Magnetohydrodynamics,162
Matrix model, 97, 104, 106, 127, 171
Membrane, 133, 234
Microscopic spectral density, 207
Minisuperspace, 41
Monopole, 59, 61

Nambu–Goto string, 270
Navier–Stokes equation, 24, 159
NBI matrix model, 140

Odderon, 356
Orthogonal polynomial method, 289

Partition function, 187, 192, 251, 265, 308
Passive advection, 26

P-brane, 129, 138
Penner model, 177
Poisson bracket, 173, 273
Pomeron, 355
Primordial magnetic field, 159, 160
Proper-time gauge, 37

Quantum electrodynamics, 113
Quantum gravity, 33

Random matrix theory, 188, 198, 287, 297
Renormalization group, 65, 147

Scaling relations, 3
Schild action, 135
Seiberg–Witten theory, 261, 279
Shell model, 165
Sigma model, 67, 79 ,83, 87, 319
Silk damping, 169
Simplicial gravity, 251
Spectral dimension, 20, 341
Spin boundaries, 16
Stress tensor, 85, 87
String, 269
Sum Rules, 193
Superconducting cosmic string, 161
Superselection, 121
Superspace, 36
Superstring, 127, 133, 143, 171, 175
SUSY transformation, 135

Toda-chain, 284
Topography, 315
Transfer-matrix, 99
Triangulation, 8
Turbulence, 23
Two-dimensional quantum gravity, 2, 341

Verma module, 332
Vielbein, 83
Virasoro master equation, 79
Vortex, 59, 240

Weyl anomaly, 72
Weyl spinor, 115
Wheeler–DeWitt equation, 38, 42
Wilson loop, 49, 51 ,53

Young tableaux, 100, 104