

Chapter 4

SUSY Statistical Mechanics and Random Band Matrices

Thomas Spencer

4.1 An Overview

The study of large random matrices in physics originated with the work of Eugene Wigner who used them to predict the energy level statistics of a large nucleus. He argued that because of the complex interactions taking place in the nucleus there should be a random matrix model with appropriate symmetries, whose eigenvalues would describe the energy level spacing statistics. Recent developments summarized in [19], give a rather complete description of the universality of eigenvalue spacings for the *mean field* Wigner matrices. Today, random matrices are studied in connection with many aspects of theoretical and mathematical physics. These include Anderson localization, number theory, generating functions for combinatorial enumeration, and low energy models of QCD. See [6] for an explanation of how random matrix theory is related to these topics and others.

The goal of these lectures is to present the basic ideas of *supersymmetric* (SUSY) statistical mechanics and its applications to the spectral theory of large Hermitian random matrices—especially *random band matrices*. There are many excellent reviews of SUSY and random matrices in the theoretical physics literature starting with the fundamental work of Efetov [17]. See for example [15, 18, 28, 31, 36, 49, 51]. This review will emphasize mathematical aspects of SUSY statistical mechanics and will try to explain ideas in their simplest form. We shall begin by first studying the average Green's function of $N \times N$ GUE matrices—Gaussian matrices whose distribution is invariant under unitary transformations. In this case the SUSY models can be expressed as integrals in just two real variables. The size of the matrix N appears only as a parameter. See Sects. 4.3–4.5.

The simple methods for GUE are then generalized to treat the more interesting case of *random band matrices*, RBM, for which much less is rigorously proved.

T. Spencer (✉)

School of Mathematics, Institute for Advanced Study, Einstein Drive, Princeton, NJ 08540, USA
e-mail: spencer@math.ias.edu

Random band matrices H_{ij} are indexed by vertices i, j of a lattice \mathbb{Z}^d . Their matrix elements are random variables which are 0 or small for $|i - j| \geq W$ and hence these matrices reflect the geometry of the lattice. The parameter W will be referred to as the width of the band. As we vary W , random band matrices approximately interpolate between mean field $N \times N$ GUE or Wigner type matrix models where $W = N$ and random Schrödinger matrices, $H = -\Delta + v_j$ on the lattice in which the randomness only appears in the potential v . Following Anderson, random Schrödinger matrices are used to model the dynamics of a quantum particle scattered by random impurities on a crystalline lattice. In Sect. 4.3 precise definitions of random band and random Schrödinger matrices are given and a qualitative relation between them is explained.

The key to learning about spectral properties of random matrices H lies in the study of averages of its Green's functions, $(E_\epsilon - H)^{-1}(j, k)$ where j, k belong to a lattice. We use the notation $A(j, k)$ to denote the matrix elements of the matrix A . The energy E typically lies inside the spectrum of H and $E_\epsilon = E - i\epsilon$ with $\epsilon > 0$. Efetov [17] showed that averages of products of Green's functions can be exactly expressed in terms of correlations of certain supersymmetric (SUSY) statistical mechanics ensembles. In SUSY statistical mechanics models, the spin or field at each lattice site has both real and Grassmann (anticommuting) components. See (4.20) below for the basic identity. These components appear symmetrically and thus the theory is called supersymmetric. Efetov's formalism builds on the foundational work of Wegner and Schäfer [41, 50] which used replicas instead of Grassmann variables. A brief review of Gaussian and Grassmann integration is given in Appendix 4.10. Although Grassmann variables are a key element of SUSY models, we shall frequently integrate them out so that the statistical mechanics involves only real integrals.

The first step in the SUSY approach to random band matrices is to define a SUSY statistical mechanics model whose correlations equal the averaged Green's functions. This is basically an algebraic step, but some analysis is needed to justify certain analytic continuations and changes of coordinates. This gives us a *dual representation* of the averaged Green's function in terms of SUSY statistical mechanics. In the case of Gaussian randomness, the average of the Green's function can be explicitly computed and produces a quartic action in the fields. The advantage of this dual representation is that many of the concepts of statistical mechanics such as phase transitions, saddle point analysis, cluster expansions and renormalization group methods [40], can then be used to analyze the behavior of Green's functions of random band matrices. Section 4.6 will review some results about phase transitions and symmetry breaking for classical statistical mechanics. This may help to give some perspective of how SUSY statistical mechanics is related to its classical counterpart.

The second step is to analyze correlations of the resulting SUSY statistical mechanics model. Typically these models have a formal non-compact symmetry. For large W , the functional integral is expected to be governed by a finite dimensional *saddle manifold*. This manifold is the orbit of a critical point under the symmetry group. The main mathematical challenge is to estimate the fluctuations about the

manifold. Most (SUSY) lattice models are difficult to analyze rigorously due to the lack of a spectral gap and the absence of a positive integrand. These lectures will focus on some special cases for which the analysis can be rigorously carried out.

4.1.1 Green's Functions

Let H be a random matrix indexed by vertices $j \in \mathbb{Z}^d$. We can think of H as a random Schrödinger matrix or an infinite random band matrix of fixed width acting on $\ell_2(\mathbb{Z}^d)$. To obtain information about time evolution and eigenvectors of H we define the average of the square modulus of the Green's function

$$\langle |G(E_\epsilon; 0, j)|^2 \rangle \equiv \langle |(E - i\epsilon - H)^{-1}(0, j)|^2 \rangle \equiv \Gamma(E_\epsilon, j) \quad (4.1)$$

where $j \in \mathbb{Z}^d$, $\epsilon > 0$, and $\langle \cdot \rangle$ denotes the average over the random variables of H . We wish to study $\Gamma(E_\epsilon, j)$ for very small $\epsilon > 0$. If for all small $\epsilon > 0$ we have

$$\Gamma(E_\epsilon, j) \leq C\epsilon^{-1}e^{-|j|/\ell} \quad (4.2)$$

then the eigenstates with eigenvalues near E are *localized*. This means the eigenstate ψ decays exponentially fast $|\psi_j|^2 \approx e^{-|j-c|/\ell}$, about some point c of the lattice with probability one. The length $\ell = \ell(E)$ is called the *localization length*.

Quantum diffusion near energy E corresponds to

$$\Gamma(E_\epsilon, j) \approx (-D(E)\Delta + \epsilon)^{-1}(0, j) \quad \text{or} \quad \hat{\Gamma}(E_\epsilon, p) \approx (D(E)p^2 + \epsilon)^{-1} \quad (4.3)$$

where $D(E)$ is the diffusion constant and Δ is the discrete lattice Laplacian and $|p|$ is small. Note that the right side is the Laplace transform of the heat kernel. When (4.3) holds with $D(E) \geq \delta > 0$, particles with energy near E are mobile and can conduct. The eigenstates ψ_j at these energies are *extended*. This means that if H is restricted to a large volume Λ , an $\ell_2(\Lambda)$ normalized eigenfunction ψ satisfies $|\psi_j|^2 \approx |\Lambda|^{-1}$ for all $j \in \Lambda$. A brief perturbative discussion of quantum diffusion is presented in Appendix 4.11. In the infinite volume limit, quantum diffusion implies absolutely continuous spectrum near E , whereas localization corresponds to dense point spectrum and the absence of conduction. Although the quantum diffusion in 3D is well studied in the physics literature, it remains a major open problem to establish it at a rigorous mathematical level.

The basic identity

$$\text{Im} \langle G(E_\epsilon; 0, 0) \rangle = \epsilon \sum_j \langle |G(E_\epsilon; 0, j)|^2 \rangle \quad (4.4)$$

is easily proved by applying the resolvent identity to $G - \tilde{G}$. It reflects conservation of probability or unitarity and is some times referred to as a Ward identity. In

Sect. 4.5 we shall see that the left side of (4.4) is proportional to the density of states, $\rho(E)$. Thus the identity $\sum_j \Gamma(E_\epsilon, j) \propto \epsilon^{-1} \rho(E)$ always holds and explains the factor ϵ^{-1} in (4.2).

In the language of SUSY statistical mechanics, localization roughly corresponds to high temperature in which distant spins become exponentially decorrelated, whereas quantum diffusion corresponds to an ordered phase in which spins are aligned over long distances. For random band matrices, the inverse temperature is

$$\beta \approx W^2 \rho(E)^2 \quad (4.5)$$

where $\rho(E)$ is the density of states, and W is the band width.

For a more intuitive time dependent picture, let $U(j, t)$ be the solution to the Schrödinger equation on the lattice

$$i \partial_t U(j, t) = H U(j, t). \quad (4.6)$$

Suppose that at time 0, $U(j, 0)$ is supported near the origin. For any unitary evolution we have conservation of probability: $\sum_j |U(j, t)|^2$ is constant in t . To measure the spread of U we define the average mean-square displacement by

$$R^2(t) = \langle \sum_j |U(j, t)|^2 |j|^2 \rangle. \quad (4.7)$$

Quantum diffusion corresponds to $R^2 \approx Dt$, whereas if all states are uniformly localized then the wave packet U does not spread as time gets large, $R^2 \leq \text{Const } \ell^2$. The Green's function at energy E , essentially projects the initial wave function U onto a spectral subspace about E . Time is roughly proportional to $t \approx \epsilon^{-1}$.

Note that if $H = -\Delta$ is the lattice Laplacian on \mathbb{Z}^d , and there no disorder, the motion is *ballistic*: $R^2 \propto t^2$. The presence of disorder, i.e. randomness, should change the character of the motion through multiple scatterings. It is expected that in the presence of spatially uniform disorder, the motion is never more than diffusive, $R^2(t) \leq Ct^{2\alpha}$ with $0 \leq \alpha \leq 1/2$ with possible logarithmic corrections in the presence of spin-orbit coupling. However, in two or more dimensions, this is still a conjecture, even for $\alpha < 1$. It is easy to show that for the lattice random Schrödinger operator we have $R^2(t) \leq Ct^2$ for any potential.

4.1.2 Symmetry and the 1D SUSY Sigma Model

Symmetries of statistical mechanics models play a key role in the macroscopic behavior of correlations. In Sect. 4.6 we present a review of phase transitions, symmetry breaking and Goldstone modes (slow modes) in classical statistical mechanics. The SUSY lattice field action which arises in the study of (4.1) is

invariant under a global hyperbolic $SU(1, 1|2)$ symmetry. As mentioned earlier, the spin or field has both real and Grassmann components. The symmetry $SU(1, 1|2)$ means that for the real components there exists a hyperbolic symmetry $U(1, 1)$ preserving $|z_1|^2 - |z_2|^2$ as discussed in Sect. 4.7. On the other hand, the Grassmann variables (reviewed in Appendix 4.10) are invariant under a compact $SU(2)$ symmetry. More details about these symmetries are in Sect. 4.9.

In both physics and mathematics, many statistical mechanics systems are studied in the *sigma model approximation*. In this approximation, spins take values in a symmetric space. The underlying symmetries and spatial structure of the original interaction are respected. The Ising model, the rotator, and the classical Heisenberg model are well known sigma models where the spin s_j lies in $\mathbb{R}^1, \mathbb{R}^2, \mathbb{R}^3$, respectively with the constraint $|s_j| = 1$. Thus they take values in the groups Z_2, S^1 and the symmetric space S^2 . One can also consider the case where the spin is matrix valued. In the Efetov sigma models, the fields are 4×4 matrices with both real and Grassmann entries. It is expected that sigma models capture the qualitative physics of more complicated models with the same symmetry, see Sect. 4.6.

In a one dimensional chain of length L , the SUSY sigma model governing conductance has a simple expression first found in [16]. The Grassmann variables can be traced out and the resulting model is a nearest neighbor spin model with *positive* weights given as follows. Let $S_j = (h_j, \sigma_j)$ denote spin vector with h_j and σ_j taking values in a hyperboloid and the sphere S^2 respectively. The Gibbs weight is then proportional to

$$\prod_{j=0}^L (h_j \cdot h_{j+1} + \sigma_j \cdot \sigma_{j+1}) e^{\beta(\sigma_j \cdot \sigma_{j+1} - h_j \cdot h_{j+1})}. \quad (4.8)$$

As in classical statistical mechanics, the parameter $\beta > 0$ is referred to as the inverse temperature. The hyperbolic components $h_j = (x_j, y_j, z_j)$ satisfy the constraint $z_j^2 - x_j^2 - y_j^2 = 1$. The Heisenberg spins σ are in \mathbb{R}^3 with $\sigma \cdot \sigma = 1$ and the dot product is Euclidean. On the other hand the dot product for the h spins is hyperbolic: $h \cdot h' = zz' - xx' - yy'$. It is very convenient to parameterize this hyperboloid with horospherical coordinates $s, t \in \mathbb{R}$:

$$z = \cosh t + s^2 e^t / 2, \quad x = \sinh t - s^2 e^t / 2, \quad y = s e^t. \quad (4.9)$$

The reader can easily check that $z_j^2 - x_j^2 - y_j^2 = 1$ is satisfied for all values of s and t . This parametrization plays an important role in later analysis of hyperbolic sigma models [14, 47]. The integration measure in σ is the uniform measure over the sphere and the measure over h_j has the density $\prod e^{t_j} ds_j dt_j$. At the end points of the chain we have set $s_0 = s_L = t_0 = t_L = 0$. Thus we have nearest neighbor hyperbolic spins (Boson–Boson sector) and Heisenberg spins (Fermion–Fermion sector) coupled via the Fermion–Boson determinant. In general, this coupling quite complicated. However in 1D it is given by $\prod_j (h_j \cdot h_{j+1} + \sigma_j \cdot \sigma_{j+1})$. As in (4.5), the inverse temperature $\beta \approx W^2 \rho(E)^2$ where $\rho(E)$ is the density of states, and W

is the band width. When $\beta \gg 1$, (4.8) shows that the spins tend to align over long distances.

By adapting the recent work of [12] it can be proved that the model given by (4.8) has a localization length proportional to β for all large β . More precisely it is shown that the conductance goes to zero exponentially fast in L , for $L \gg \beta$.

4.1.3 SUSY Sigma Models in 3D

Although the SUSY sigma models are widely used in physics to make detailed predictions about eigenfunctions, energy spacings and quantum transport, for disordered quantum systems, there is as yet no rigorous analysis of the $SU(1, 1|2)$ Efetov models in 2 or more dimensions. Even in one dimension, where the problem can be reduced to a transfer matrix, rigorous results are restricted to the sigma model mentioned above. A key difficulty arises from the fact that SUSY lattice field models cannot usually be treated using probabilistic methods due to the presence of Grassmann variables and strongly oscillatory factors.

In recent work with Disertori and Zirnbauer [12, 14] we have established a phase transition for a *simpler* SUSY hyperbolic sigma model in three dimensions. We shall call this model the $H^{2|2}$ model. The notation refers to the fact that the field takes values in hyperbolic 2 space augmented with 2 Grassmann variables to make it supersymmetric. This model, introduced by Zirnbauer [52], is expected to reflect the qualitative behavior, such as localization or diffusion, of random band matrices in any dimension. The great advantage of the $H^{2|2}$ model is that the Grassmann variables can be traced out producing a statistical mechanics model with *positive* but nonlocal weights. (The non locality arises from a determinant produced by integrating out the Grassmann fields.) This means that probabilistic tools can be applied. In fact we shall see that quantum localization and diffusion is closely related to *a random walk in a random environment*. This environment is highly correlated and has strong fluctuations in 1 and 2D.

In Sect. 4.8 we will describe the $H^{2|2}$ model and sketch a proof of a phase transition as $\beta(E) > 0$ goes from small to large values. Small values of β , high temperature, will correspond to localization—exponential decay of correlations and lack of conductance. In three dimensions, we shall see that large values of β correspond to quantum diffusion and extended states. Standard techniques for proving the existence of phase transitions, such as reflection positivity, do not seem to apply in this case. Instead, the proof of this transition relies crucially on Ward identities arising from SUSY symmetries of the model combined with estimates on random walk in a random environment. The simplest expression of these Ward identities is reflected by the fact that the partition function is identically one for all parameter values which appear in a SUSY fashion. Thus derivatives in these parameters vanish and yield identities. The SUSY $H^{2|2}$ model is nevertheless highly non trivial because physical observables are not SUSY and produce interesting and complicated correlations. Classical Ward identities will be discussed in Sects. 4.6 and 4.7. In Sect. 4.9 we give a very brief description of Efetov's sigma model.

4.2 Classical Models of Quantum Dynamics

In this section we digress to describe two classical models which have some connection with quantum dynamics. The first is called linearly edge reinforced random walk, ERRW. This history dependent walk favors moving along edges it has visited in the past. Diaconis [10] introduced this model and discovered that ERRW can be expressed as random walk in a highly correlated random environment. The statistical description of the random environment appears to be closely related but not identical to that of the SUSY, $H^{2|2}$ model described in Sect. 4.8. In both cases it is important to observe that the generator for the walk is not uniformly elliptic, making it possible to get non-diffusive dynamics. The second model, called the Manhattan model, arises from network models developed to understand the Quantum Hall transition. This model is equivalent to the behavior of a particle undergoing a random unitary evolution. The remarkable feature of this model is that after the randomness is integrated out, the complex phases along paths are canceled. Thus paths have positive weights and the motion has a probabilistic interpretation.

To define linearly edge reinforced random walk (ERRW), consider a discrete time walk on \mathbb{Z}^d starting at the origin and let $n(e, t)$ denote the number of times the walk has visited the edge e up to time t . Then the probability $P(v, v', t + 1)$ that the walk at vertex v will visit a neighboring edge $e = (v, v')$ at time $t + 1$ is given by

$$P(v, v', t + 1) = (\beta + n(e, t)) / S_\beta(v, t) \quad (4.10)$$

where S_β is the sum of $\beta + n(e', t)$ over all the edges e' touching v . The parameter β is analogous to β in (4.8) or to the β in the $\mathbb{H}^{2|2}$ model defined later. Note that if β is large, the reinforcement is weak and in the limit $\beta \rightarrow \infty$, (4.10) describes simple random walk.

In 1D and 1D strips, ERRW is *localized* for any value of $\beta > 0$ [35]. This means that the probability of finding an ERRW, $w(t)$, at a distance r from its initial position is exponentially small in r , thus

$$\text{Prob}[|w(t) - w(0)| \geq r] \leq C e^{-r/\ell}. \quad (4.11)$$

Merkel and Rolles [34] established this result by proving that conductance across an edge goes to zero exponentially fast with the distance of the edge to the origin. Note that in this model, (and in the $H^{(2|2)}$ model with $\epsilon_j = \delta_{j,0}$), the random environment is not translation invariant and the distribution of the local conductance depends on starting point of the walk. Localization is proved, using a variant of the Mermin–Wagner type argument and the localization length ℓ is proportional to $\beta|S|$ where $|S|$ is the width of the strip. Using similar methods, Merkel and Rolles also show that in 2D the local conductance goes to zero slowly away from 0. In 3D, there are no rigorous theorems for ERRW. Localization probably occurs for strong reinforcement, i.e., for β small. It is natural to conjecture that in 2D, ERRW is always exponentially localized for all values of reinforcement. On the Bethe lattice,

with probability p at the vertices, angled at $\pm 45^\circ$ with respect to the x axis. Orbits are obtained by imagining that a light beam light travels along lattice lines until it is reflected by a mirror. If $p = 1$ and the choice of angles is independent with probability $1/2$, the model is equivalent to *critical bond percolation*. Although the loops have finite length with probability one, their average length is *infinite*. For $0 < p < 1$ little is known rigorously but one expects that the average loop length to still be infinite.

Remark. Although the two models described above are quite special, they share one of the essential difficulties of the random Schrödinger models and RBM—they are highly non-Markovian. It is this feature that makes these models challenging to analyze. On the other hand, it is the memory or non-Markovian character which is partly responsible for localization. The problem of proving strictly subballistic dynamics, $R^2(t) \propto t^{2\alpha}$ with $\alpha < 1$, is also unresolved for these classical systems in dimension two or more.

4.3 Introduction Random Matrix Ensembles and SUSY

In this section we shall define various Gaussian matrix ensembles. The simplest of these is GUE—Gaussian Unitary ensemble. In this case the matrix entries H_{ij} are mean zero, complex, independent random variables for $i \leq j$ and $1 \leq i, j \leq N$. Since H has a Gaussian distribution it suffices to specify its covariance:

$$\langle H_{ij} \tilde{H}_{i'j'} \rangle = \langle H_{ij} H_{j'i'} \rangle = \delta(ii') \delta(jj') / N. \quad (4.12)$$

The average over the randomness or disorder is denoted by $\langle \cdot \rangle$ and \tilde{H} denotes the complex conjugate of H . The density for this ensemble is given by

$$1/Z_N e^{-N \operatorname{tr} H^2/2} \prod_{i < j} dH_{ii} \prod_{i < j} dH_{ij}^{\operatorname{Re}} dH_{ij}^{\operatorname{Im}}.$$

The factor of $1/Z_N$ ensures that the integral is 1. It is clear that H and $U^* H U$ have identical distributions for any fixed unitary matrix U . This invariance is a crucial feature in the classical analysis of such matrices via orthogonal polynomials. It also holds when $\operatorname{tr} H^2$ above is replaced by $\operatorname{tr} V(H)$ for polynomials V , which are bounded from below [8, 9, 37]. However, Wigner matrices, and random band matrices do not have unitarily invariant distributions and new methods are needed to obtain the desired spectral information. See [19, 46] for recent reviews of Wigner matrices and random band matrices.

To define random band matrices, RBM, with a Gaussian distribution, let i and j range over a periodic box $\Lambda \subset \mathbb{Z}^d$ of volume $|\Lambda|$ and set

$$\langle H_{ij} \tilde{H}_{i'j'} \rangle = \langle H_{ij} H_{j'i'} \rangle = \delta(ii') \delta(jj') J_{ij}. \quad (4.13)$$

Here J_{ij} is a symmetric function which is *small* for large $|i - j|$. We shall assume that for fixed i , $\sum_j J_{ij} = 1$. With this normalization, it is known that most of the spectrum of H is concentrated in the interval $[-2, 2]$ with high probability. One especially convenient choice of J is given by the lattice Green's function

$$J_{jk} = (-W^2 \Delta + 1)^{-1}(j, k) \quad (4.14)$$

where Δ is the lattice Laplacian on Λ with periodic boundary conditions

$$\Delta f(j) = \sum_{|j' - j|=1} (f(j') - f(j)).$$

Note that with this choice of J , $\sum_j J_{ij} = 1$ and the variance of the matrix elements is exponentially small when $|i - j| \gg W$. In fact, in one dimension $J_{ij} \approx e^{-|i-j|/W} / W$. Here $|i - j|$ is the natural distance in the periodic box. Hence W will be referred to as the width of the band. The advantage of (4.14) is that its *inverse*, which appears by duality in certain SUSY formulations, has only *nearest neighbor* matrix elements and W^2 is just a parameter, see (4.31) below.

The density for the Gaussian orthogonal ensemble, GOE, is also proportional to $\exp(-N \operatorname{tr} H^2 / 2)$ but the matrix H is real symmetric. The covariance is given by $\langle H_{ij}^2 \rangle = 1/N$ and

$$\langle H_{ij} H_{kl} \rangle = (2N)^{-1} \{ \delta(ik) \delta(jl) + \delta(il) \delta(jk) \}, \quad i \neq j.$$

The symmetric RBM are defined by introducing J_{ij} as in (4.13). Although the SUSY formalism can also be applied to the symmetric case, the formulas are a bit more complicated and we shall not discuss them.

4.3.1 Some Conjectures About RBM and Random Schrödinger

Let us now compare lattice random Schrödinger matrices (RS) on \mathbb{Z}^d given by

$$H_{RS} = -\Delta + \lambda v_j$$

and RBM (4.13) of width W on \mathbb{Z}^d . Above, v_j are assumed to be independent identically distributed Gaussian random variables of mean 0 and variance $\langle v_j^2 \rangle = 1$. The parameter $\lambda > 0$ measures the strength of the disorder. See [22, 45] for recent reviews of mathematical aspects of Anderson localization. Although RBM and RS look quite different, they are both local, that is their matrix elements j, k are small (or zero) if $|j - k|$ is large. The models are expected to have the same qualitative properties when $\lambda \approx W^{-1}$. For example, eigenvectors for RS are known to decay exponentially fast in one dimension with a localization length proportional to λ^{-2} .

On the other hand for 1D, RBM the localization length is known to be less than W^8 by work of Schenker [42], and is expected to be approximately W^2 . From the point of view of perturbation theory, if we compute the bare diffusion constant D_0 , obtained by summing ladder diagrams, then we have $D_0 \propto \lambda^{-2}$, W^2 for the RS and RBM respectively. See Appendix 4.11 for a review of perturbation theory for RS and RBM.

Localization is reasonably well understood for RS in one dimension for any disorder. Localization has also been proved on \mathbb{Z}^d for strong disorder or at energies where the density of states $\rho(E)$ is very small. See [45] for references to the mathematical literature. On the other hand a mathematical proof of quantum diffusion in 3D, as defined by (4.3) as $\epsilon \downarrow 0$ for RS or RBM for fixed W , remains a major open problem. In 2D it is conjectured that all states are localized with a localization length growing exponentially fast in W^2 or λ^{-2} for real symmetric matrices. Note the similarity of this conjecture with that for the Manhattan model defined in Sect. 4.2.

4.3.2 Conjectured Universality of GUE and GOE Models

GUE or GOE matrices are mean field and have no spatial or geometric structure. Nevertheless, the local eigenvalue statistics of these models are expected to match those of certain RBM for large N . For example, in 1D the RBM of the form (4.13), (4.14) should have the same local energy level statistics as GUE as $N \rightarrow \infty$, (modulo a simple rescaling), provided that $W^2 \gg N$ and energies lie in the bulk, $|E| \leq 2 - \delta$. In 3D universality is expected to hold for E in the bulk, and for fixed $W \gg 1$ independent of $N = |\Lambda| \rightarrow \infty$. A. Sodin [43] recently proved universality of local energy spacing about the spectral edge, $E \approx 2$ provided $W \gg N^{5/6}$. His result is sharp in the sense that for smaller values of W another distribution appears. Universality has been established for Wigner matrices at all energies [19].

In terms of SUSY statistical mechanics, universality of local energy correlations for Hermitian band matrices, can be intuitively understood as follows. For appropriate values of E , W and dimension, energy correlations are essentially governed by the saddle manifold which is the orbit of a single saddle point under the action of $SU(1, 1|2)$. This manifold only depends on the symmetry of the ensemble. See Sect. 4.3 of [36] for a more detailed (but not rigorous) explanation of universality and explicit corrections to it arising from finite volume effects. Universality fails when the fluctuations about the saddle manifold are too pronounced. This can happen if W is not large enough or when $d \leq 2$. In such cases localization may occur and the eigenvalue spacing will be Poisson. Note that in 1D localization effects should not appear unless $N \gg W^2$.

The local eigenvalue spacing statistics for a random Schrödinger matrix in a large box is expected to match those of GOE after simple rescaling, provided that one looks in a region of energies where diffusion should appear, i.e., (4.3) holds. RS

corresponds to GOE rather than GUE because it is real symmetric matrix and the corresponding saddle manifold is different from GUE. See [48] for some interesting results showing that in some special, limiting cases the local eigenvalue spacings for the random Schrödinger matrix has GOE statistics.

Remark. The local eigenvalue statistics of GUE appear to be mysteriously connected to the statistics of the spacings of zeros of the Riemann zeta function. See [39] and Keating's article in [6] for reviews of this conjecture and others concerning the relation of random matrix theory and number theory.

4.3.3 Green's Functions and Gaussian Integrals

For an $N \times N$ Hermitian matrix H , define the inverse matrix:

$$G(E_\epsilon) = (E_\epsilon - H)^{-1} \quad \text{where} \quad E_\epsilon = E - i\epsilon. \quad (4.15)$$

This a bounded matrix provided that E is real and $\epsilon > 0$. The Green's matrix is denoted $G(E_\epsilon)$, and $G(E_\epsilon; k, j)$, its matrix elements, will be called the Green's function.

Let $z = (z_1, z_2, \dots, z_N)$ with $z_j = x_j + iy_j$ denote an element of \mathbb{C}^N and define the quadratic form

$$[z; H z] = \sum_{i,j} \bar{z}_k H_{kj} z_j. \quad (4.16)$$

Then we can calculate the following Gaussian integrals:

$$\int e^{-i[z; (E_\epsilon - H)z]} D z = (-i)^N \det(E_\epsilon - H)^{-1} \quad \text{where} \quad D z \equiv \prod_j^N dx_j dy_j / \pi \quad (4.17)$$

and

$$\int e^{-i[z; (E_\epsilon - H)z]} z_k \bar{z}_j D z = (-i)^{N+1} \det(E_\epsilon - H)^{-1} G(E_\epsilon; k, j). \quad (4.18)$$

It is important to note that the integrals above are *convergent* provided that $\epsilon > 0$. The quadratic form $[z; (E - H)z]$ is real so its contribution only oscillates. The factor of $i = \sqrt{-1}$ in the exponent is needed because the matrix $E - H$ has an indefinite signature when E is in the spectrum of H .

There is a similar identity in which the complex commuting variables z are replaced by anticommuting Grassmann variables $\psi_j, \bar{\psi}_j, j = 1, 2 \dots N$. Let A be an $N \times N$ matrix and as in (4.16) define

$$[\psi; A \psi] = \sum \bar{\psi}_k A_{kj} \psi_j.$$

Then we have

$$\int e^{-[\psi; A \psi]} D\psi = \det A \quad \text{where} \quad D\psi \equiv \prod_j^N d\bar{\psi}_j d\psi_j. \quad (4.19)$$

See Appendix 4.10 for a brief review of Gaussian and Grassmann integration. The Grassmann integral is introduced so that we can cancel the unwanted determinant in (4.18) and so that we can perform the average over the randomness in H . Thus we obtain a SUSY representation for the Green's function:

$$G(E_\epsilon; k, j) = i \int e^{-i[z; (E_\epsilon - H)z]} e^{-i[\psi; (E_\epsilon - H)\psi]} z_k \bar{z}_j D z D \psi. \quad (4.20)$$

Equation (4.20) is the starting point for all SUSY formulas. Notice that if H has a Gaussian distribution, the expectation of (4.20) can be explicitly performed since H appears linearly. We obtain:

$$\langle G(E_\epsilon; k, j) \rangle = i \int e^{-iE_\epsilon([z; z] + [\psi; \psi])} e^{-\frac{1}{2} \langle \{[z; H z] + [\psi; H \psi]\}^2 \rangle} z_k \bar{z}_j D z D \psi. \quad (4.21)$$

The resulting lattice field model will be quartic in the z and ψ fields. Notice that if the observable $i z_k \bar{z}_j$ were absent from (4.21), then the determinants would cancel and the integral would be equal to 1 for all parameters. Thus in SUSY systems, the *partition function is identically 1*.

In a similar fashion we can obtain more complicated formulas for

$$\Gamma(E_\epsilon, j) = \langle G(E_\epsilon; 0, j) \bar{G}(E_\epsilon; 0, j) \rangle.$$

To do this we must introduce additional variables $w \in \mathbb{C}^N$ and independent Grassmann variables $\chi, \bar{\chi}$ to obtain the second factor, \bar{G} , which is the complex conjugate of G . See Sects. 4.7 and 4.9 below.

4.4 Averaging $\text{Det}(E_\epsilon - H)^{-1}$

Before working with Grassmann variables we shall first show how to calculate the average of the inverse determinant over the Gaussian disorder using only the complex variables z_j . Although this average has no physical significance, it will illustrate some basic mathematical ideas.

First consider the simplest case: H is an $N \times N$, GUE matrix. Let us apply (4.17) to calculate

$$(i)^{-N} \langle \det(E_\epsilon - H)^{-1} \rangle = \langle \int e^{-i[z; (E_\epsilon - H)z]} D z \rangle. \quad (4.22)$$

Since the expectation is Gaussian and the covariance is given by (4.12), the average over H can be expressed as follows:

$$\langle e^{-i[z; Hz]} \rangle_{GUE} = e^{-1/2 \langle [z; Hz]^2 \rangle} = e^{-\frac{1}{2N} [z; z]^2}. \quad (4.23)$$

The most direct way to estimate the remaining integral over the z variables is to introduce a new coordinate $r = [z, z] = \sum |z_j|^2$. Then we have

$$\langle \det(E_\epsilon - H)^{-1} \rangle = C_N \int_0^\infty e^{-\frac{1}{2N} r^2 - i E_\epsilon r} r^{N-1} dr$$

where C_N is an explicit constant related to the volume of the sphere in $2N$ dimensions. It is convenient to rescale $r \rightarrow Nr$ and obtain an integral of the form

$$\int_0^\infty e^{-N(r^2/2 - \ln r - i E_\epsilon r)} dr/r.$$

The method of steepest descent can now be applied. We deform the contour of integration over r so that it passes through the saddle point. The saddle point r_s is obtained by setting the derivative of the exponent to 0: $r_s - 1/r_s - i E_\epsilon = 0$. This is a quadratic equation with a solution $r_s = iE/2 + \sqrt{1 - (E/2)^2}$. The contour must be chosen so that the absolute value of the integrand is dominated by the saddle point.

Exercise. Stirling's formula:

$$N! = \int_0^\infty e^{-t} t^N dt \approx N^N e^{-N} \sqrt{2N\pi}.$$

To derive this let $t = Ns$ and expand to quadratic order about the saddle point $s = 1$. The square root arises from the identity $N \int e^{-Ns^2/2} ds = \sqrt{2N\pi}$. We shall see many SUSY calculations are related to generalizations of the gamma function.

Remark. For other ensembles, radial coordinates do not suffice and one must compute the Jacobian for the new collective variables. See the section on matrix polar coordinates in Appendix 4.10. Collective coordinates can also be set up with the help of the coarea formula.

An alternate way to compute $\langle \det(E_\epsilon - H)^{-1} \rangle$ uses the Hubbard–Stratonovich transform. In its simplest form, introduce a real auxiliary variable a to unravel the quartic expression in z as follows:

$$e^{-\frac{1}{2N} [z; z]^2} = \sqrt{N/2\pi} \int e^{-Na^2/2} e^{ia[z; z]} da. \quad (4.24)$$

If we apply (4.24) to (4.22) and (4.23), the z variables now appear quadratically and we can integrate over them. Since there are N independent z_j to integrate, we get a factor $(E_\epsilon - a)^{-N}$, hence:

$$\langle \det(E_\epsilon - H)^{-1} \rangle = i^N \sqrt{N/2\pi} \int e^{-Na^2/2} (E_\epsilon - a)^{-N} da. \quad (4.25)$$

Let

$$f(a) = N [a^2/2 + \ln(E_\epsilon - a)].$$

The saddle point a_s is obtained by setting $f'(a_s) = 0$. This gives a quadratic equation whose solution is

$$a_s = E_\epsilon/2 + i \sqrt{1 - (E_\epsilon/2)^2}. \quad (4.26)$$

We shift our contour of integration by $a \rightarrow a + a_s$. Note $|a_s| \approx 1$ and that we have chosen the $+$ sign in (4.26) so that the pole of $(E - i\epsilon - a)^{-N}$ has not been crossed. Along this contour one checks that for E satisfying $|E| \leq 1.8$, the maximum modulus of the integrand occurs at the saddle a_s . In particular, this deformation of contour avoids the small denominator $E_\epsilon - a$ occurring when $a \approx E$. For energies such that $1.8 \leq |E| \leq 2$ another contour must be used [11]. Note that the Hessian at the saddle is

$$f''(a_s) = N(1 - a_s^2) = N\{2(1 - (E/2)^2) - iE\sqrt{1 - (E/2)^2}\} \quad (4.27)$$

has a positive real part for $|E| < 2$. To complete our estimate of (4.25) for large N we calculate the quadratic correction to the saddle point:

$$\langle \det(E_\epsilon - H)^{-1} \rangle \approx i^N \sqrt{N/2\pi} e^{-Nf(a_s)} \int e^{-N/2 f''(a_s) a^2} da.$$

Higher order corrections to the saddle point are suppressed by powers of N^{-1} .

Remark. The imaginary part of right hand side of (4.25), (N even), is proportional to

$$e^{-E^2/2} H_N(E \sqrt{N/2})$$

where H_N is the N 'th Hermite polynomial. This can be seen by integrating by parts $N - 1$ times. Thus such identities point to a connection between the method of orthogonal polynomials and SUSY statistical mechanics. See [11] for more details.

4.4.1 Gaussian Band Matrices and Statistical Mechanics

Now let us consider the more general case when H is a finite Gaussian random band matrix indexed by lattice sites in a periodic box $\Lambda \subset \mathbb{Z}^d$ with covariance given by (4.13) and (4.14): $J_{jk} = (-W^2 \Delta + 1)^{-1}(j, k)$. Then we have

$$\langle e^{-i[z;Hz]} \rangle = e^{-1/2\langle [z;Hz]^2 \rangle} = e^{-1/2\sum |z_i|^2 J_{ij}|z_j|^2}. \quad (4.28)$$

In order to average over the z variables we introduce real Gaussian auxiliary fields a_j with 0 mean and covariance J_{ij} . Let $\langle \cdot \rangle_J$ be the corresponding expectation so that $\langle a_i a_j \rangle_J = J_{ij}$ and

$$e^{-1/2\sum_{ij}|z_i|^2 J_{ij}|z_j|^2} = \langle e^{i\sum a_j |z_j|^2} \rangle_J. \quad (4.29)$$

This is again the Hubbard–Stratonovich trick. We can now average over the z 's since they appear quadratically. By combining (4.22), (4.28) and (4.29)

$$\begin{aligned} \langle \det(E_\epsilon - H)^{-1} \rangle &= \langle \int e^{-i[z;(E_\epsilon - H)z]} D\mathbf{z} \rangle \\ &= \langle \int e^{-i[z;(E_\epsilon - a)z]} D\mathbf{z} \rangle_J = \langle \prod_{j \in \Lambda} (E_\epsilon - a_j)^{-1} \rangle_J. \end{aligned} \quad (4.30)$$

Since by (4.14), the Gaussian a_j fields have covariance $(-W^2\Delta + 1)^{-1}$, it follows that (4.30) is proportional to:

$$\int e^{-\frac{1}{2}\sum_j (W^2(\nabla a_j)^2 + a_j^2)} \prod_{j \in \Lambda} (E_\epsilon - a_j)^{-1} da_j. \quad (4.31)$$

The expression $(\nabla a_j)^2$ is the square of finite difference gradient at j . The large parameter W tends to make the a_j fields constant over a large range of lattice sites. The saddle point can be calculated as it was for GUE and we find that it is independent of the lattice site j and is again given by (4.26). We deform the contour of integration $a_j \rightarrow a_j + a_s$. We must restrict $|E| \leq 1.8$ as in the case of GUE so that the norm of the integrand along the shifted contour is maximized at the saddle. The Hessian at the saddle is given by

$$H_s'' = -W^2\Delta + 1 - a_s^2. \quad (4.32)$$

Since $\text{Re}(1 - a_s^2) > 0$, $[H'']^{-1}(j, k)$ decays exponentially fast for separations $|j - k| \gg W$.

Since (4.31) is an integral over many variables a_j , $j \in \Lambda$, its behavior is not so easy to estimate as (4.25). To control fluctuations away from the saddle one uses the exponential decay of $[H'']^{-1}$ and techniques of *statistical mechanics* such as the cluster or polymer expansion [40], about blocks of side W . This will show that the integral can be approximated by a product of nearly independent integrals over boxes of size W . It enables one to justify the asymptotic expansion for

$$|\Lambda|^{-1} \log \langle \det(E_\epsilon - H)^{-1} \rangle$$

in powers of W^{-1} . Moreover, one can prove that with respect to the expectation defined using (4.31), a_j, a_k are exponentially independent for $|j - k| \gg W$ uniformly in $\epsilon > 0$. The main reason this works is that the inverse of the Hessian at a_s has exponential decay and the corrections to the Hessian contribution are small when W is large. As in (4.25), corrections beyond quadratic order are small for large N . See [7, 13] for the SUSY version of these statements.

Remark. In one dimension, (4.31) shows that we can calculate the $\langle \det(E_\epsilon - H)^{-1} \rangle$ using a nearest neighbor transfer matrix. In particular if the lattice sites j belong to a circular chain of length L , there is an integral operator $T = T(a, a')$ such that $\langle \det(E_\epsilon - H)^{-1} \rangle = \text{Tr } T^L$.

Remark. For the lattice random Schrödinger matrix, $H = -\Delta + \lambda v$, in a finite box, with a Gaussian potential v of mean 0 and variance 1, it is easy to show that

$$\langle \det(E_\epsilon - H)^{-1} \rangle = \int e^{-i[z; (E_\epsilon + \Delta)z] - \frac{\lambda^2}{2} \sum_j |\tilde{z}_j z_j|^2} D\tilde{z}.$$

The complex variables z_j are indexed by vertices in a finite box contained in \mathbb{Z}^d . For small λ this is a highly oscillatory integral which is more difficult than (4.31) to analyze. Rigorous methods for controlling such integrals are still missing when λ is small and $d \geq 2$. Note that for RBM, the oscillations are much less pronounced because the off diagonal terms are also random. In this case the factor

$$\exp -\frac{1}{2} \left(\sum_j W^2 (\nabla a)_j^2 + a_j^2 \right)$$

dominates our integral for large W .

4.5 The Density of States for GUE and RBM

The *integrated* density of states for an $N \times N$ Hermitian matrix H is denoted $n(E) = \int^E d\rho(E')$ is the fraction of eigenvalues less than E and $\rho(E)$ denotes the density of states. The average of the density of states is given by the expression

$$\langle \rho_\epsilon(E) \rangle = \frac{1}{N} \text{tr} \langle \delta_\epsilon(H - E) \rangle = \frac{1}{N\pi} \text{tr } \text{Im} \langle G(E_\epsilon) \rangle \quad (4.33)$$

as $\epsilon \downarrow 0$. Here we are using the well known fact that

$$\delta_\epsilon(x - E) \equiv \frac{1}{\pi} \frac{\epsilon}{(x - E)^2 + \epsilon^2} = \frac{1}{\pi} \text{Im}(E_\epsilon - x)^{-1}$$

is an approximate delta function at E as $\epsilon \rightarrow 0$.

Remarks. The Wigner semicircle distribution asserts that the density of states of a GUE matrix is given by $\pi^{-1}\sqrt{1-(E/2)^2}$. Such results can be proved for many ensembles including RBM by *first fixing* ϵ in (4.33) and then letting N , and $W \rightarrow \infty$. Appendix 4.11 gives a simple formal derivation of this law using self-consistent perturbation theory. Note that the parameter ϵ is the scale at which we can resolve different energies. For a system of size N we would like to understand the number of eigenvalues in an energy window of width $N^{-\alpha}$ with $\alpha \approx 1$. Thus we would like to get estimates for $\epsilon \approx N^{-\alpha}$. Such estimates are difficult to obtain when $\alpha \approx 1$ but have recently been established for Wigner ensembles, see [20]. Using SUSY [7, 13] estimates on the density of states for a special class of Gaussian band matrices indexed by \mathbb{Z}^3 are proved *uniformly* in ϵ and the size of the box for fixed $W \geq W_0$. See the theorem in this section.

Note that the density of states is *not* a universal quantity. At the end of Appendix 4.11 we show that 2 coupled GUE $N \times N$ matrices has a density of states given by a cubic rather than quadratic equation.

We now present an identity for the average Green's function for GUE starting from (4.21). Note that by (4.12)

$$\begin{aligned} -\frac{1}{2} < ([z; Hz] + [\psi; H\psi])^2 >_{GUE} \\ &= -\frac{1}{2N} \{ [z; z]^2 - [\psi; \psi]^2 - 2[\psi; z][z; \psi] \}. \end{aligned} \quad (4.34)$$

Let us introduce another real auxiliary variable $b \in \mathbb{R}$ and apply the Hubbard–Stratonovich transform to decouple the Grassmann variables. As in (4.24) we use the identity

$$e^{[\psi; \psi]^2/2N} = \sqrt{N/2\pi} \int db e^{-Nb^2/2} e^{-b[\psi; \psi]}. \quad (4.35)$$

Now by combining (4.24) and (4.35) we see that the expressions are quadratic in z and ψ . Note that

$$\int e^{-\sum^N \bar{\psi}_j \psi_j (b + iE_\epsilon)} D\psi D\bar{\psi} = i^N (E_\epsilon - ib)^N.$$

The exponential series for the cross terms appearing in (4.34), $[\psi; z][z; \psi]$, terminates after one step because $[\psi; z]^2 = 0$. We compute its expectation with respect to the normalized Gaussian measure in ψ, z and get

$$\begin{aligned} < [\psi; z][z; \psi] > &= < \sum_{ij} z_i \bar{\psi}_i \bar{z}_j \psi_j > = \sum_j < \bar{\psi}_j \psi_j \bar{z}_j z_j > \\ &= -N(E_\epsilon - ib)^{-1} (E_\epsilon - a)^{-1}. \end{aligned}$$

Thus after integrating over both the z and ψ fields and obtain the expression:

$$\begin{aligned}
 < \frac{1}{N} \text{tr} G(E_\epsilon) > = N^{-1} \int [z; z] e^{-\frac{1}{2N} \{[z; z]^2 - [\psi; \psi]^2 - 2[\psi; z][z; \psi]\}} D z D \psi \\
 &= N/2\pi \int da db (E_\epsilon - a)^{-1} e^{-N(a^2 + b^2)/2} (E_\epsilon - i b)^N (E_\epsilon - a)^{-N} \\
 &\quad \times \left[1 - \frac{N+1}{N} (E_\epsilon - a)^{-1} (E_\epsilon - i b)^{-1} \right] \approx < (E_\epsilon - a)^{-1} >_{\text{SUSY}} .
 \end{aligned} \tag{4.36}$$

The first factor of $(E_\epsilon - a)^{-1}$ on the right hand side roughly corresponds to the trace of the Green's function. The partition function is

$$\begin{aligned}
 &N/2\pi \int da db e^{-N(a^2 + b^2)/2} (E_\epsilon - i b)^N (E_\epsilon - a)^{-N} [1 - (E_\epsilon - a)^{-1} (E_\epsilon - i b)^{-1}] \\
 &\equiv 1
 \end{aligned} \tag{4.37}$$

for all values of E , ϵ and N . This is due to the fact that if there is no observable the determinants cancel. The last factor in (4.37) arises from the cross terms and is called the Fermion–Boson (*FB*) contribution. It represents the coupling between the determinants. For band matrices it is useful to introduce auxiliary dual Grassmann variables (analogous to a and b) to treat this contribution. See [11, 13, 36].

The study of $\rho(E)$ reduces to the analysis about the saddle points of the integrand. As we have explained earlier, there is precisely one saddle point

$$a_s(E) = E_\epsilon/2 + i \sqrt{1 - (E_\epsilon/2)^2} \tag{4.38}$$

in the a field. However, the b field has two saddle points

$$i b_s = a_s, \quad \text{and} \quad i b'_s = E_\epsilon - a_s = \bar{a}_s.$$

Hence, both saddle points (a_s, b_s) and (a_s, b'_s) contribute to (4.36).

Let us briefly analyze the fluctuations about the saddles as we did in (4.26), (4.27). The first saddle gives the Wigner semicircle law. To see this note that the action at a_s, b_s takes the value 1 since $a_s^2 + b_s^2 = 0$ and $(E_\epsilon - i b_s)/(E_\epsilon - a_s) = 1$. The integral of quadratic fluctuations about the saddle,

$$N \int e^{-N(1 - a_s^2)(a^2 + b^2)} da db \tag{4.39}$$

is exactly canceled by the *FB* contribution at (a_s, b_s) . Thus to a high level of accuracy we can simply replace the observable in the SUSY expectation (4.36) by its value at the saddle. This gives us Wigner's semicircle law:

$$\begin{aligned}\rho(E) &= \frac{1}{\pi N} \text{Im } \text{tr} \langle G(E_\epsilon) \rangle \approx \pi^{-1} \text{Im} \langle (E_\epsilon - a)^{-1} \rangle_{\text{SUSY}} \\ &\approx \pi^{-1} \text{Im}(E_\epsilon - a_s)^{-1} = \pi^{-1} \sqrt{1 - (E/2)^2}.\end{aligned}\quad (4.40)$$

It is easy to check that the second saddle vanishes when inserted into the *FB* factor because $(E - a_s)(E - i b'_s) = \bar{a}_s a_s = 1$. Thus to leading order it does not contribute to the density of states and hence (4.40) holds. However, the second saddle will contribute a highly oscillatory contribution to the action proportional to

$$\frac{1}{N} \left(\frac{E_\epsilon - a'_s}{E_\epsilon - a_s} \right)^N e^{-N/2(a_s^2 - \bar{a}_s^2)}.\quad (4.41)$$

If we take derivatives in E , this contribution is no longer suppressed when $\epsilon \approx 1/N$. This result is not easily seen in perturbation theory. I believe this is a compelling example of the nonperturbative power of the SUSY method.

Remarks. If $\epsilon = 0$, then (4.41) has modulus $1/N$. We implicitly assume that the energy E is inside the bulk ie $|E| < 2$. Near the edge of the spectrum the Hessian at the saddle point vanishes and a more delicate analysis is called for. The density of states near $E = \pm 2$ then governed by an Airy function. We refer to Disertori's review of GUE [11], for more details.

4.5.1 Density of States for RBM

We conclude this section with a brief discussion of the average Green's function for RBM in 3D with the covariance J given by (4.14). The SUSY statistical mechanics for RBM is expressed in terms of a_j, b_j and Grassmann fields $\bar{\eta}_j, \eta_j$ with $j \in \Lambda \subset \mathbb{Z}^d$ and it has a local (nearest neighbor) weight

$$\begin{aligned}&\exp \left[-\frac{1}{2} \sum_j \{ W^2 (\nabla a_j)^2 + W^2 (\nabla b_j)^2 + a_j^2 + b_j^2 \} \right] \prod_j \frac{E_\epsilon - i b_j}{E_\epsilon - a_j} \\ &\times \exp - \sum_j \{ W^2 \nabla \bar{\eta}_j \nabla \eta_j + \bar{\eta}_j \eta_j (1 - (E_\epsilon - a_j)^{-1} (E_\epsilon - i b_j)^{-1}) \}.\end{aligned}$$

After the Grassmann variables have been integrated out we get:

$$\begin{aligned}&\exp \left[-\frac{1}{2} \sum_j \{ W^2 (\nabla a_j)^2 + W^2 (\nabla b_j)^2 + a_j^2 + b_j^2 \} \right] \prod_j \frac{E_\epsilon - i b_j}{E_\epsilon - a_j} \\ &\times \det \{ -W^2 \Delta + 1 - \delta_{ij} (E_\epsilon - a_j)^{-1} (E_\epsilon - i b_j)^{-1} \} \text{DaDb}.\end{aligned}\quad (4.42)$$

Note the similarity of the above weight with (4.31) and (4.37). The determinant is called the Fermion–Boson contribution. In one dimension, the DOS can be reduced to the analysis of a nearest neighbor transfer matrix. Large W keeps the fields a_j, b_j nearly constant. This helps to control fluctuations about saddle point.

Theorem ([13]). Let $d = 3$, J be given by (4.14) and $|E| \leq 1.8$. For $W \geq W_0$ the average $\langle G(E_\epsilon, j, j) \rangle$ for RBM is uniformly bounded in ϵ and Λ . It is approximately given by the Wigner distribution with corrections of order $1/W^2$. Moreover, we have

$$| \langle G(E_\epsilon; 0, x) G(E_\epsilon; x, 0) \rangle | \leq e^{-m|x|} \quad (4.43)$$

for $m \propto W^{-1}$.

The proof of this theorem follows the discussion after (4.31). We deform our contour of integration and prove that for $|E| \leq 1.8$, the dominant contribution comes from the saddle point a_s, b_s which is independent of j . However, as in GUE one must take into account the second saddle which complicates the analysis. Note that the Hessian at the saddle is still given by (4.32). Its inverse governs the decay in (4.43).

Remark. The first rigorous results of this type appear in [7] for the Wegner N orbital model. Their SUSY analysis is similar to that presented in [13].

Remark. As explained at the end of the previous section, one can derive a formula analogous to (4.42) for the random Schrödinger matrix in 2D or 3D, but we do not have good control over the SUSY statistical mechanics for small λ . A formal perturbative analysis of the average Green's function for the random Schrödinger matrix with small λ is given in Appendix 4.11. Note that for large λ , it is relatively easy to control the SUSY integral since in this case the SUSY measure is dominated by the product measure $\prod_j \exp[-\lambda^2(\bar{z}_j z_j + \bar{\psi}_j \psi_j)^2]$. As a final remark note that in the product of Green's functions appearing in (4.43), energies lie on the same side of the real axis. The expectation $\langle |G(E_\epsilon; 0, j)|^2 \rangle$ is quite different and for such cases hyperbolic symmetry emerges, see Sect. 4.7, and gapless modes may appear.

4.6 Statistical Mechanics, Sigma Models, and Goldstone Modes

In this section we shall review some classical sigma models and their lattice field theory counterparts. This may help to see how classical statistical mechanics is related to the SUSY sigma models which will be described later. The symmetry groups described in this section are compact whereas those of related the Efetov models are noncompact as we shall see in Sects. 4.7–4.9.

Consider a lattice model with a field $\phi(j) = (\phi^1(j), \dots, \phi^{m+1}(j))$ taking values in \mathbb{R}^{m+1} and whose quartic action is given by

$$A_\Lambda(\phi) = \sum_{j \in \Lambda} \{ |\nabla \phi(j)|^2 + \lambda(|\phi(j)|^2 - a)^2 + \epsilon \cdot \phi^1 \}. \quad (4.44)$$

As before ∇ denotes the nearest neighbor finite difference gradient and Λ is a large periodic box contained in \mathbb{Z}^d . The partition function takes the form

$$Z_\Lambda = \int e^{-\beta A_\Lambda(\phi)} \prod_{j \in \Lambda} d\phi_j.$$

To define its sigma model approximation, consider spins, $S_j \in S^m$ where $j \in \Lambda$. More precisely $S_j = (S_j^{(1)}, \dots, S_j^{(m+1)})$ with $S_j^2 = S_j \cdot S_j = 1$. For $m=0$, the spin takes values ± 1 and this is the Ising spin. The energy or action of a configuration $\{S_j\}$ is given by

$$A_{\tilde{\beta}, \Lambda}(S) = \tilde{\beta} \sum_{j \sim j' \in \Lambda} (S_j - S_{j'})^2 - \epsilon \sum_j S_j^{(1)}. \quad (4.45)$$

The Gibbs weight is proportional to $e^{-A(S)}$, the parameters $\tilde{\beta}, \epsilon \geq 0$ are proportional to the inverse temperature and the applied field respectively and $j \sim j'$ denote adjacent lattice sites. If $\epsilon > 0$ then the minimum energy configuration is unique and the all the spins point in the direction $(1, 0, \dots, 0)$.

If $\epsilon = 0$, A is invariant under a global rotation of the spins. The energy is minimized when all the spins are pointing in the same direction. This is the ground state of the system. When $m=1$ we have $O(2)$ symmetry and for $m=2$ the symmetry is $O(3)$ corresponding to the classical rotator (or X-Y) model and classical Heisenberg model respectively. When $\epsilon > 0$ the full symmetry is broken but in the case of $O(3)$ is broken to $O(2)$. The parameter ϵ in (4.45) also appears in SUSY models and is related to the imaginary part of E_ϵ in the Green's function.

Consider the classical scalar $|\phi|^4$ model, (4.44), with $\phi_j \in \mathbb{C}$ with a $U(1)$ symmetry. The idea behind the sigma approximation is that the complex field ϕ_j can be expressed in polar coordinates $r_j S_j$ where $|S_j| = 1$. The radial variables r are “fast” variables fluctuating about some effective radius r^* . If the quartic interaction has the form $\lambda(|\phi_j|^2 - 1)^2$ and $\lambda \gg 1, r^* \approx 1$. If we freeze $r_j = r^*$ we obtain (4.45) with $\tilde{\beta} \approx (r^*)^2 \beta$. The $|\phi|^4$ model is expected to have the same qualitative properties as the $O(2)$ sigma models. Moreover, at their respective phase transitions, the long distance asymptotics of correlations should be identical. This means that the two models have the same critical exponents although their critical temperatures will typically be different. This is a reflection of the principle universality for second order transitions. Its proof is a major mathematical challenge.

Remark. The role of large λ above will roughly be played by W in the SUSY models for random band matrices.

Now let us discuss properties of the sigma models with Gibbs expectation:

$$\langle \cdot \rangle_{\Lambda}(\bar{\beta}, \epsilon) = Z_{\Lambda}(\bar{\beta}, \epsilon)^{-1} \int \cdot e^{-A_{\bar{\beta}\Lambda}(S)} \prod_j d\mu(S_j). \quad (4.46)$$

The measure $d\mu(S_j)$ is the uniform measure on the sphere. The partition function Z is defined so that this defines a probability measure. Basic correlations are the spin–spin correlations $\langle S_0 \cdot S_x \rangle$ and the magnetization $\langle S_0^1 \rangle$. Let us assume that we have taken the limit in which $\Lambda \uparrow \mathbb{Z}^d$. First consider the case in which there is no applied field $\epsilon = 0$. In one dimension there is no phase transition and the spin–spin correlations decay exponentially fast for all non zero temperatures, (β finite),

$$0 \leq \langle S_0 \cdot S_x \rangle \leq C e^{-|x|/\ell} \quad (4.47)$$

where ℓ is the correlation length. For the models with continuous symmetry, $m = 1, 2$, ℓ is proportional to β . However, for the Ising model ℓ is exponentially large in β . At high temperature, β small, it is easy to show that in any dimension the spins are independent at long distances and (4.47) holds.

In dimension $d \geq 2$, the low temperature phase is interesting and more difficult to analyze. In 2D we know that the Ising model has long range order and a *spontaneous magnetization* M for $\beta > \beta_c$. This means that for large x

$$\langle S_0 \cdot S_x \rangle(\beta, \epsilon = 0) \rightarrow M^2(\beta) > 0 \text{ and } \lim_{\epsilon \downarrow 0} \langle S_0^{(1)} \rangle(\beta, \epsilon) = M. \quad (4.48)$$

Thus at low temperature spins align even at long distances. Note that the order of limits here is crucial. In (4.48) we have taken the infinite volume limit first and then sent $\epsilon \rightarrow 0$. For any finite box it is easy to show that second the limit is 0 because the symmetry of the spins is present. For this reason (4.48) is called symmetry breaking.

In two dimensions, the Mermin–Wagner Theorem states that a continuous symmetry cannot be broken. Hence, $M = 0$ for the $O(2)$ and $O(3)$ models. In classical XY model with $O(2)$ symmetry there is a Kosterlitz–Thouless transition [25, 32]: the magnetization, M , vanishes but the spin–spin correlation with $\epsilon = 0$ goes from an exponential decay for small β to a slow power law decay $\approx |x|^{-c/\beta}$ for large β .

A major conjecture of mathematical physics, first formulated by Polyakov, states that for the classical Heisenberg model in 2D, the spin–spin correlation always decays exponentially fast with correlation length $\ell \approx e^{\beta}$. This conjecture is related to Anderson localization expected in 2D. It is a remote cousin of confinement for non-Abelian gauge theory. The conjecture, can partly be understood by a renormalization group argument showing that the effective temperature increases under as we scale up distances. The positive curvature of sphere play a key role in this analysis.

In three dimensions it is known through the methods of reflection positivity [27], that there is long range order and continuous symmetry breaking at low temperature. These methods are rather special and do not apply to the SUSY models described here.

Continuous symmetry breaking is accompanied by Goldstone Bosons. This means that there is a very slow decay of correlations. This is most easily seen in the Fourier transform of the pair correlation. Consider a classical XY or Heisenberg model on \mathbb{Z}^d . The following theorem is a special case of the Goldstone–Nambu–Mermin–Wagner Theorem. Let M be defined as in (4.48).

Theorem. Let correlations be defined as above and suppose that the spontaneous magnetization $M > 0$. Then

$$\sum_x e^{-ip \cdot x} \langle S_0^{(2)} S_x^{(2)} \rangle(\beta, \epsilon) \geq CM^2(\beta p^2 + \epsilon M)^{-1}. \quad (4.49)$$

Thus for $p = 0$ the sum diverges as $\epsilon \downarrow 0$. This theorem applies to the case $m = 1, 2$ but not the Ising model. It is also established for lattice field theories with continuous symmetry.

Remark. In 2D, if the spontaneous magnetization were positive, then the integral over p of the right side of (4.49) diverges for small ϵ . On the other hand, the integral over p of the left side equals 1, since $S_0 \cdot S_0 = 1$. Thus M must vanish in 2D. In higher dimensions there is no contradiction since the integral of $|p|^{-2}$, $|p| \leq \pi$ is finite.

We shall derive a simple Ward identity for the $O(2)$ sigma model in d dimensions. In angular coordinates we have

$$A(\theta) = - \sum_{j \sim j'} \beta \cos(\theta_j - \theta_{j'}) - \epsilon \sum_j \cos(\theta_j). \quad (4.50)$$

Consider the integral in a finite volume: $\int \sin(\theta_0) e^{-A(\theta)} \prod_{j \in \Lambda} d\theta_j$. Make a simple change of variables $\theta_j \rightarrow \theta_j + \alpha$. The integral is of course independent of α . If we take the derivative in α and then set $\alpha = 0$ the resulting expression must vanish. This yields a simple Ward identity

$$M = \langle \cos(\theta_0) \rangle = \epsilon \sum_j \langle \sin(\theta_0) \sin(\theta_j) \rangle. \quad (4.51)$$

After dividing by ϵ we obtain the theorem for $p = 0$. Ward identities are often just infinitesimal reflections of symmetries in our system. In principle we can apply this procedure to any one parameter change of variables.

Remark. Note the similarity of (4.51) with the Ward identity (4.4). The density of states $\rho(E)$ plays the role of the spontaneous magnetization M . Moreover, the right side of (4.49) is analogous to (4.3). However, there is an important distinction:

$\rho(E)$ does not vanish in the high temperature or localized regions. Unlike the magnetization in classical models, $\rho(E)$ is *not* an order parameter. Indeed, $\rho(E)$ is not expected to reflect the transition between localized and extended states. So in this respect classical and SUSY models differ. For SUSY or RBM, the localized phase is reflected in a vanishing of the diffusion constant in (4.3), $D(E, \epsilon) \propto \epsilon$. Thus in regions of localization $< |G(E_\epsilon, 0, 0)|^2 >$ diverges as $\epsilon \rightarrow 0$ as we see in (4.1).

Proof of Theorem. Let $|\Lambda|$ denote the volume of the periodic box and set $D = |\Lambda|^{-1/2} \sum_j e^{-ip \cdot j} \frac{\partial}{\partial \theta_j}$ and $\hat{S}(p) = |\Lambda|^{-1/2} \sum_j e^{+ip \cdot j} \sin(\theta_j)$. By translation invariance, integration by parts, and the Schwarz inequality we have

$$\begin{aligned} M = < \cos(\theta_0) >_A = < D \hat{S}(p) >_A = < \hat{S}(p) D(A) >_A \\ \leq < |\hat{S}(p)|^2 >_A^{1/2} < \overline{D(A)} D(A) >_A^{1/2}. \end{aligned} \quad (4.52)$$

Since $< |\hat{S}(p)|^2 >$ equals the left side of (4.49) the theorem follows by integrating by parts once again,

$$\begin{aligned} < \overline{D(A)} D(A) >_A &= < D \overline{D(A)} >_A \\ &= |\Lambda|^{-1} \sum_{j \sim j'} < 2\beta \cos(\theta_j - \theta_{j'}) (1 - \cos(j - j')p) \\ &\quad + \epsilon \cos(\theta_j) >_A \\ &\leq C(\beta p^2 + \epsilon < \cos(\theta_0) >_A) \end{aligned} \quad (4.53)$$

which holds for small p . Here we have followed the exposition in [26].

4.7 Hyperbolic Symmetry

Let us analyze the average of $|\det(E_\epsilon - H)|^{-2}$ with H a GUE matrix. We study this average because it is the simplest way to illustrate the emergence of hyperbolic symmetry. This section analyses the so called Boson–Boson sector in the sigma model approximation. More complicated expressions involving Grassmann variables appear when analyzing the average of $|G(E_\epsilon; j, k)|^2$. This is touched upon in Sect. 4.9.

To begin, let us contrast $< G >$ and $< |G|^2 >$ for the trivial scalar case $N = 1$. In the first case we see that $\int e^{-H^2} (E_\epsilon - H)^{-1} dH$ is finite as $\epsilon \rightarrow 0$ by shifting the contour of integration off the real axis $H \rightarrow H + i\delta$ with $\delta > 0$ so that the pole is not crossed. On the other hand, if one considers the average of $|(E_\epsilon - H)|^{-2}$, we cannot deform the contour integral near E and the integral will diverge like ϵ^{-1} . This divergence will be reflected in the hyperbolic symmetry. Later in this section we shall see that in 3D this symmetry will be associated with gapless Goldstone modes. These modes were absent in Sects. 4.3–4.5.

Let $z, w \in C^N$. As in (4.17) we can write:

$$\begin{aligned} |\det(E_\epsilon - H)|^{-2} &= \det(E_\epsilon - H) \times \det(E_{-\epsilon} - H) \\ &= \int e^{-i[z; (E_\epsilon - H)z]} Dz \times \int e^{i[w; (E_{-\epsilon} - H)w]} Dw. \end{aligned} \quad (4.54)$$

Note that the two factors are complex conjugates of each other. The factor of i has been reversed in the w integral to guarantee its convergence. This sign change is responsible for the hyperbolic symmetry. The Gaussian average over H is

$$\langle e^{-i([z; Hz] - [w; Hw])} \rangle = e^{-1/2 \langle ([z; Hz] - [w; Hw])^2 \rangle}. \quad (4.55)$$

Note that

$$\langle ([z; Hz] - [w; Hw])^2 \rangle = \langle \left[\sum H_{kj} (\bar{z}_k z_j - \bar{w}_k z_j) \right]^2 \rangle. \quad (4.56)$$

For GUE the right side is computed using (4.12)

$$\langle ([z; Hz] - [w; Hw])^2 \rangle = 1/N ([z, z]^2 + [w, w]^2 - 2[z, w][w, z]). \quad (4.57)$$

Note that the hyperbolic signature is already evident in (4.55). Following Fyodorov [29], introduce the 2×2 non negative matrix:

$$M(z, w) = \begin{pmatrix} [z, z] & [z, w] \\ [w, z] & [w, w] \end{pmatrix} \quad (4.58)$$

and let

$$L = \text{diag}(1, -1). \quad (4.59)$$

Then we see that

$$\langle |\det(E_\epsilon - H)|^{-2} \rangle = \int e^{-\frac{1}{2N} \text{tr}(ML)^2 - i \text{Etr}(ML) + \epsilon \text{tr} M} Dz Dw. \quad (4.60)$$

For a positive 2×2 matrix P , consider the delta function $\delta(P - M(z, w))$ and integrate over z and w . It can be shown [29, 47], also see (4.116) in Appendix 4.10, that

$$\int \delta(P - M(z, w)) Dz Dw = (\det P)^{N-2}. \quad (4.61)$$

Assuming this holds we can now write the right side in terms of the new collective coordinate P

$$\langle |\det(E_\epsilon - H)|^{-2} \rangle = C_N \int_{P>0} e^{-\frac{1}{2N} \text{tr}(PL)^2} e^{-i \text{Etr}(PL) - \epsilon \text{tr} P} \det P^{N-2} dP. \quad (4.62)$$

After rescaling $P \rightarrow NP$ we have

$$\langle |\det(E_\epsilon - H)|^{-2} \rangle = C'_N \int_{P>0} e^{-N\{tr(PL)^2/2 + iEtr(PL) + \epsilon tr P\}} \det P^{N-2} dP. \quad (4.63)$$

In order to compute the integral we shall again change variables and integrate over PL . First note that for $P > 0$, PL has two real eigenvalues of opposite sign. This is because PL has the same eigenvalues as $P^{1/2}LP^{1/2}$ which is self adjoint with a negative determinant. Moreover, it can be shown that

$$PL = TDT^{-1} \quad (4.64)$$

where $D = \text{diag}(p_1, -p_2)$ with p_1, p_2 positive and T belongs to $U(1, 1)$. By definition $T \in U(1, 1)$ when

$$T^*LT = L \rightarrow T \in SU(1, 1). \quad (4.65)$$

The proof is similar to that for Hermitian matrices. We shall regard PL as our new integration variable.

Note that (4.63) can be written in terms of p_1, p_2 except for $\epsilon tr P$ which will involve the integral over $SU(1, 1)$. The p_1, p_2 are analogous to the radial variable r introduced in Sect. 4.6. Converting to the new coordinate system our measure becomes

$$(p_1 + p_2)^2 dp_1 dp_2 d\mu(T) \quad (4.66)$$

where $d\mu(T)$ the Haar measure on $U(1, 1)$. For large N , the p variables are approximately given by the complex saddle point

$$p_1 = -iE/2 + \rho(E), \quad -p_2 = -iE/2 - \rho(E), \quad \rho(E) \equiv \sqrt{1 - (E/2)^2}. \quad (4.67)$$

The p variables fluctuate only slightly about (4.67) while the T matrix ranges over the symmetric space $SU(1, 1)/U(1)$ and produces a singularity for small ϵ . With the p_1, p_2 fixed the only remaining integral is over $SU(1, 1)$. Thus from (4.64) we have:

$$Q \equiv PL \approx \rho(E) TLT^{-1} + iE/2.$$

If we set $\epsilon = \bar{\epsilon}/N$ and take the limit of large N we see that (4.63) is given by

$$\int e^{-\bar{\epsilon}\rho(E)tr(LS)} d\mu(T), \quad \text{where } S \equiv TLT^{-1}. \quad (4.68)$$

This is the basis for the sigma model. The second term iE above is independent of T so it is dropped in (4.68).

4.7.1 Random Band Case

The band version or Wegner's N orbital [41] version of such hyperbolic sigma models was studied in [47]. The physical justification of this sigma model and also the Efetov SUSY sigma model comes from a granular picture of matter. The idea is that a metal consists of grains (of size $N \gg 1$). Within the grains there is mean field type interaction and there is a weak coupling across the neighboring grains. As in (4.68) if the grain interactions are scaled properly and $N \rightarrow \infty$ we will obtain a sigma model.

For each lattice site $j \in \Lambda \subset \mathbb{Z}^d$ we define a new spin variable given by

$$S_j = T_j^{-1} L T_j \quad \text{and} \quad P_j L \approx \rho(E) S_j.$$

Note that $S_j^2 = 1$ and S_j naturally belongs to $SU(1, 1)/U(1)$. This symmetric space is isomorphic to the hyperbolic upper half plane. In the last equation we have used the form of the p_i given as above. The imaginary part of the p_1 and $-p_2$ are equal at the saddle so that T and T^{-1} cancel producing only a trivial contribution. The explicit dependence on E only appears through $\rho(E)$.

There is a similar picture for the square of the average determinant using Grassmann integration. We can integrate out the Grassmann fields and in the sigma model approximation obtain an integral over the symmetric space $SU(2)/U(1) = S^2$ —this gives the classical Heisenberg model.

The action of the hyperbolic sigma model, see [47], on the lattice is

$$A(S) = \beta \sum_{j \sim j'} \text{tr} S_j S_{j'} + \epsilon \sum_j \text{tr} L S_j. \quad (4.69)$$

The notation $j \sim j'$ denotes nearest neighbor vertices on the lattice. The Gibbs weight is proportional to $e^{-A(S)} d\mu(T)$. The integration over $SU(1, 1)$ is divergent unless $\epsilon > 0$. The last term above is a symmetry breaking term analogous to a magnetic field. For RBM, $\beta \approx W^2 \rho(E)^2$.

To parametrize the integral over $SU(1, 1)$ we use *horospherical* coordinates $(s_j, t_j) \in \mathbb{R}^2$ given by (4.9). In this coordinate system, the action takes the form:

$$A(s, t) = \beta \sum_{j \sim j'} [\cosh(t_j - t_{j'}) + \frac{1}{2}(s_j - s_{j'})^2 e^{(t_j + t_{j'})}] + \epsilon \sum_{j \in \Lambda} [\cosh(t_j) + \frac{1}{2}s^2 e^{t_j}]. \quad (4.70)$$

Equivalently, if h_j are the hyperbolic spins appearing in (4.8) note that we have

$$h_j \cdot h_{j'} = z_j z_{j'} - x_j x_{j'} - y_j y_{j'} = \cosh(t_j - t_{j'}) + \frac{1}{2}(s_j - s_{j'})^2 e^{(t_j + t_{j'})}.$$

The symmetry breaking term is just $\epsilon \sum z_j$. There is a symmetry $t_j \rightarrow t_j + \gamma$ and $s_j \rightarrow s_j e^{-\gamma}$ which leaves the action invariant when $\epsilon = 0$. Note that $A(s, t)$

is quadratic in s_j . Let us define the quadratic form associated to the s variables in (4.70):

$$[v; D_{\beta,\varepsilon}(t) v]_\Lambda = \beta \sum_{(i \sim j)} e^{t_i + t_j} (v_i - v_j)^2 + \varepsilon \sum_{k \in \Lambda} e^{t_k} v_k^2 \quad (4.71)$$

where v is a real vector, $v_j, j \in \Lambda$. Integration over the s fields produces $\det^{-1/2}(D_{\beta,\varepsilon}(t))$. The determinant is positive but non local in t . Thus the total *effective action* is given by

$$A(t) = \sum_{j \sim j'} \{\beta \cosh(t_j - t_{j'})\} + \frac{1}{2} \ln \det(D_{\beta,\varepsilon}(t)) + \varepsilon \sum_j \cosh(t_j). \quad (4.72)$$

The quadratic form $D_{\beta,\varepsilon}(t)$ will also appear in the SUSY sigma model defined later. Note that if $t = 0$, $D_{\beta,\varepsilon}(t) = -\beta\Delta + \varepsilon$. For $t \neq 0$, $D_{\beta,\varepsilon}(t)$ is a finite difference elliptic operator which is the generator of a *random walk in a random environment*. The factor $e^{(t_j + t_{j'})}$ is the conductance across the edge (j, j') . Note that since t_j are not bounded, $D_{\beta,\varepsilon}(t)$ is *not* uniformly elliptic.

Lemma. $\det(D_{\beta,\varepsilon}(t))$ is a log convex function of the t variables.

Following D. Brydges, this is a consequence of the matrix tree theorem which expresses the determinant of as a sum:

$$\sum_{\mathcal{F} \in \mathcal{R}} \prod_{j, j' \in \mathcal{F}} \beta e^{t_j + t_{j'}} \prod_{j \in \mathcal{R}} \varepsilon e^{t_j}$$

where the sum above ranges over spanning rooted forests on Λ . See [1] for a proof of such identities.

Thus the effective action A (4.72) is convex and in fact its Hessian $A''(t) \geq -\beta\Delta + \varepsilon$. The sigma model can now be analyzed using Brascamp–Lieb inequalities and a Ward identity. Its convexity will imply that this model does not have a phase transition in three dimensions.

Theorem (Brascamp–Lieb [5]). Let $A(t)$ be a real convex function of $t_j, j \in \Lambda$ and v_j be a real vector. If the Hessian of the action A is convex $A''(t) \geq K > 0$ then

$$\langle e^{([v; t] - \langle [v; t] \rangle)} \rangle_A \leq e^{\frac{1}{2} [v; K^{-1} v]}. \quad (4.73)$$

Here K is a positive matrix independent of t and $\langle \cdot \rangle_A$ denotes the expectation with respect to the normalized measure $Z_A^{-1} e^{-A(t)} D t$. Note if A is quadratic in t , (4.73) is an identity.

Note that in 3D the Brascamp–Lieb inequality only gives us estimates on functions of $([v; t] - \langle [v; t] \rangle)$. Below we obtain estimates on $\langle [v; t] \rangle$ by using Ward identities and translation invariance. Then the above theorem will yield a bound on moments of the local conductance $e^{(t_j + t_{j'})}$. These bounds on the

conductance imply that in three dimensions we have “diffusion” with respect to the measure defined via (4.70). See the discussion following (4.87).

Theorem ([47]). In the 3D hyperbolic sigma model described by (4.72), all moments of the form $\langle \cosh^p(t_0) \rangle$ are bounded for all β . The estimates are uniform in ϵ provided we first take the limit $\Lambda \rightarrow \mathbb{Z}^3$.

Proof. First note that if we let $v_j = p\delta_0(j)$ then in three dimensions since $K = -\beta\Delta + \epsilon$ we have $[v; K^{-1}v] \leq \text{Const}$ uniformly as $\epsilon \rightarrow 0$. To estimate the average $\langle t_0 \rangle$ we apply the Ward identity:

$$2 \langle \sinh(t_0) \rangle = \langle s_0^2 e^{t_0} \rangle = \langle D_{\beta, \epsilon}^{-1}(0, 0) e^{t_0} \rangle. \quad (4.74)$$

To obtain this identity we apply $t_j \rightarrow t_j + \gamma$ and $s_j \rightarrow s_j e^{-\gamma}$ to (4.70) then differentiate in γ and set $\gamma = 0$. We have also assumed translation invariance. Since the right side of (4.74) is positive, if we multiply the above identity by $e^{-\langle t_0 \rangle}$ we have

$$\langle e^{(t_0 - \langle t_0 \rangle)} \rangle = \langle e^{-t_0 - \langle t_0 \rangle} \rangle + e^{-\langle t_0 \rangle} \langle s_0^2 e^{t_0} \rangle \geq e^{-2\langle t_0 \rangle}$$

where the last bound follows from Jensen’s inequality. The left side is bounded by Brascamp–Lieb by taking $v_j = \delta_0(j)$. Thus we get an upper bound on $-\langle t_0 \rangle$ and on $\langle e^{-t_0} \rangle$. To get the upper bound on $\langle t_0 \rangle$ we use the inequality

$$e^{t_0} D_{\beta, \epsilon}^{-1}(0, 0) \leq e^{t_0} \beta \sum_{j \sim j'} e^{-t_j - t_{j'}} (G_0(0, j) - G_0(0, j'))^2 + O(\epsilon)$$

where $G_0 = (-\beta\Delta + \epsilon)^{-1}$. See (4.121) in Appendix 4.12 for a precise statement and proof of this inequality. The sum is convergent in 3D since the gradient of $G_0(0, j)$ decays like $|j|^{-2}$. Multiplying (4.74) by $e^{\langle t_0 \rangle}$ we get

$$\begin{aligned} \langle e^{(t_0 + \langle t_0 \rangle)} \rangle &\leq \beta \sum_{j \sim j'} \langle e^{-t_j - t_{j'}} e^{(t_0 + \langle t_0 \rangle)} \rangle (G_0(0, j) - G_0(0, j'))^2 \\ &\quad + \langle e^{-(t_0 - \langle t_0 \rangle)} \rangle + O(\epsilon). \end{aligned}$$

By translation invariance $\langle t_j \rangle = \langle t_0 \rangle$, the exponent on the right side has the form $([v; t] - \langle [v; t] \rangle)$ and thus the Brascamp–Lieb inequality gives us a bound on the right side and a bound on $\langle t_0 \rangle$ by Jensen’s inequality. Above we have ignored the $O(\epsilon)$ term. If we include it then we obtain an inequality of the form $e^{2\langle t_0 \rangle} \leq C + \epsilon e^{\langle t_0 \rangle}$ and the bound for $e^{\langle t_0 \rangle}$ still follows. Since we now have estimates on $\langle |t_j| \rangle$, the desired estimate on $\langle \cosh^p(t_j) \rangle$ follows from the Brascamp–Lieb inequality.

Remark. The effective action for the SUSY hyperbolic sigma model, $H^{2|2}$, described in Sect. 4.8, looks very much like the formula above except that the coefficient of $\ln \det$ is replaced by $-1/2$. For this case the action is *not convex* and so phase transitions are not excluded. In fact in 3D for small β there is localization

and $\langle e^{-t_0} \rangle$ will diverge as $\epsilon \rightarrow 0$. Note that the above results and those described in Sect. 4.8 rely heavily on the use of horospherical coordinates.

4.8 Phase Transition for a SUSY Hyperbolic Sigma Model

In this section we study a simpler version of the Efetov sigma models introduced by Zirnbauer [52]. This model is the $\mathbb{H}^{2|2}$ model mentioned in the introduction. This model is expected to qualitatively reflect the phenomenology of Anderson's tight binding model. The great advantage of this model is that the Grassmann degrees of freedom can be explicitly integrated out to produce a real effective action in bosonic variables. Thus probabilistic methods can be applied. In 3D we shall sketch the proof [12, 14] that this model has the analog of the *Anderson transition*. The analysis of the phase transition relies heavily on Ward identities and on the study of a random walk in a strongly correlated random environment.

In order to define the $\mathbb{H}^{2|2}$ sigma model, let u_j be a vector at each lattice point $j \in \Lambda \subset \mathbb{Z}^d$ with three bosonic components and two fermionic components

$$u_j = (z_j, x_j, y_j, \xi_j, \eta_j) ,$$

where ξ, η are odd elements and z, x, y are even elements of a real Grassmann algebra. The scalar product is defined by

$$(u, u') = -zz' + xx' + yy' + \xi\eta' - \eta\xi' , \quad (u, u) = -z^2 + x^2 + y^2 + 2\xi\eta$$

and the action is obtained by summing over nearest neighbors j, j'

$$\mathcal{A}[u] = \frac{1}{2} \sum_{(j, j') \in \Lambda} \beta(u_j - u_{j'}, u_j - u_{j'}) + \sum_{j \in \Lambda} \varepsilon_j (z_j - 1) . \quad (4.75)$$

The sigma model constraint, $(u_j, u_j) = -1$, is imposed so that the field lies on a SUSY hyperboloid, $\mathbb{H}^{2|2}$.

We choose the branch of the hyperboloid so that $z_j \geq 1$ for each j . It is very useful to parametrize this manifold in horospherical coordinates:

$$x = \sinh t - e^t \left(\frac{1}{2} s^2 + \bar{\psi} \psi \right) , \quad y = s e^t , \quad \xi = \bar{\psi} e^t , \quad \eta = \psi e^t ,$$

and

$$z = \cosh t + e^t \left(\frac{1}{2} s^2 + \bar{\psi} \psi \right)$$

where t and s are even elements and $\bar{\psi}, \psi$ are odd elements of a real Grassmann algebra.

In these coordinates, the sigma model action is given by

$$\begin{aligned} \mathcal{A}[t, s, \psi, \bar{\psi}] = & \sum_{(ij) \in \Lambda} \beta (\cosh(t_i - t_j) - 1) \\ & + \frac{1}{2} [s; D_{\beta, \epsilon} s] + [\bar{\psi}; D_{\beta, \epsilon} \psi] + \sum_{j \in \Lambda} \epsilon_j (\cosh t_j - 1). \end{aligned} \quad (4.76)$$

We define the corresponding expectation by $\langle \cdot \rangle = \langle \cdot \rangle_{\Lambda, \beta, \epsilon}$.

Note that the action is quadratic in the Grassmann and s variables. Here $D_{\beta, \epsilon} = D_{\beta, \epsilon}(t)$ is the generator of a *random walk in random environment*, given by the quadratic form

$$[v; D_{\beta, \epsilon}(t) v]_{\Lambda} \equiv \beta \sum_{(j \sim j')} e^{t_j + t_{j'}} (v_j - v_{j'})^2 + \sum_{k \in \Lambda} \epsilon_k e^{t_k} v_k^2 \quad (4.77)$$

as it is in (4.71). The weights, $e^{t_j + t_{j'}}$, are the *local conductances* across a nearest neighbor edge j, j' . The $\epsilon_j e^{t_j}$ term is a killing rate for the walk at j . For the random walk starting at 0 without killing, we take $\epsilon_0 = 1$ and $\epsilon_j = 0$ otherwise. For the random band matrices if we set $\epsilon_j = \epsilon$ then ϵ may be thought of as the imaginary part of the energy.

After integrating over the Grassmann variables $\psi, \bar{\psi}$ and the variables $s_j \in \mathbb{R}$ in (4.76) we get the effective bosonic field theory with action $\mathcal{E}_{\beta, \epsilon}(t)$ and partition function

$$Z_{\Lambda}(\beta, \epsilon) = \int e^{-\mathcal{E}_{\beta, \epsilon}(t)} \prod e^{-t_j} dt_j \equiv \int e^{-\beta \mathcal{L}(t)} \cdot [\det D_{\beta, \epsilon}(t)]^{1/2} \prod_j e^{-t_j} \frac{dt_j}{\sqrt{2\pi}}. \quad (4.78)$$

where

$$\mathcal{L}(t) = \sum_{j \sim j'} [\cosh(t_j - t_{j'}) - 1] + \sum_j \frac{\epsilon_j}{\beta} [(\cosh(t_j) - 1)]. \quad (4.79)$$

Note that the determinant is a positive but nonlocal functional of the t_j hence the effective action, $\mathcal{E} = \mathcal{L} - 1/2 \ln \det D_{\beta, \epsilon}$, is also nonlocal. The additional factor of e^{-t_j} in (4.78) arises from a Jacobian. Because of the internal supersymmetry, we know that for all values of β, ϵ the partition function

$$Z(\beta, \epsilon) \equiv 1. \quad (4.80)$$

This identity holds even if β is edge dependent.

The analog of the Green's function $\langle |G(E_{\epsilon}; 0, x)|^2 \rangle$ of the Anderson model is the average of the Green's function of $D_{\beta, \epsilon}$,

$$\langle s_0 e^{t_0} s_x e^{t_x} \rangle (\beta, \epsilon) = \langle e^{(t_0 + t_x)} D_{\beta, \epsilon}(t)^{-1}(0, x) \rangle (\beta, \epsilon) \equiv \mathcal{G}_{\beta, \epsilon}(0, x) \quad (4.81)$$

where the expectation is with respect to the SUSY statistical mechanics weight defined above. The parameter $\beta = \beta(E)$ is roughly the bare conductance across an edge and we shall usually set $\varepsilon = \varepsilon_j$ for all j . In addition to the identity $Z(\beta, \varepsilon) \equiv 1$, there are additional Ward identities

$$\langle e^{t_j} \rangle \equiv 1, \quad \varepsilon \sum_x \mathcal{G}_{\beta, \varepsilon}(0, x) = 1 \quad (4.82)$$

which hold for all values of β and ε . The second identity above corresponds to the Ward identity (4.4).

Note that if $|t_j| \leq \text{Const}$, then the conductances $e^{t_j + t_{j'}}$ are uniformly bounded from above and below and

$$D_{\beta, \varepsilon}(t)^{-1}(0, x) \approx (-\beta\Delta + \varepsilon)^{-1}(0, x)$$

is the diffusion propagator. Thus localization can only occur due to large deviations of the t field.

An alternative Schrödinger like representation of (4.81) is given by

$$\mathcal{G}_{\beta, \varepsilon}(0, x) = \langle \tilde{D}_{\beta, \varepsilon}^{-1}(t)(0, x) \rangle \quad (4.83)$$

where

$$e^{-t} D_{\beta, \varepsilon}(t) e^{-t} \equiv \tilde{D}_{\beta, \varepsilon}(t) = -\beta\Delta + \beta V(t) + \varepsilon e^{-t}, \quad (4.84)$$

and $V(t)$ is a diagonal matrix (or “potential”) given by

$$V_{jj}(t) = \sum_{i: |i-j|=1} (e^{t_i - t_j} - 1).$$

In this representation, the potential is highly correlated and $\tilde{D} \geq 0$ as a quadratic form.

Some insight into the transition for the $\mathbb{H}^{2|2}$ model can be obtained by finding the configuration $t_j = t^*$ which minimizes the effective action $\mathcal{E}_{\beta, \varepsilon}(t)$ appearing in (4.78). It is shown in [14] that this configuration is unique and does not depend on j . For large β , it is given by

$$1\text{D}: \quad \varepsilon e^{-t^*} \simeq \beta^{-1}, \quad 2\text{D}: \quad \varepsilon e^{-t^*} \simeq e^{-\beta}, \quad 3\text{D}: \quad t^* \simeq 0. \quad (4.85)$$

Note that in one and two dimensions, t^* depends sensitively on ε and that negative values of t_j are favored as $\varepsilon \rightarrow 0$. This means that at t^* a mass εe^{-t^*} in (4.84) appears even as $\varepsilon \rightarrow 0$. Another interpretation is that the classical conductance $e^{t_j + t_{j'}}$ should be small in *some sense*. This is a somewhat subtle point. Due to large deviations of the t field in 1D and 2D, $\langle e^{t_j + t_{j'}} \rangle$ is expected to diverge, whereas $\langle e^{t_j/2} \rangle$ should become small as $\varepsilon \rightarrow 0$. One way to adapt

the saddle approximation so that it is sensitive to different observables is to include the observable when computing the saddle point. For example, when taking the expectation of $e^{p t_0}$, the saddle is only slightly changed when $p = 1/2$ but for $p = 2$ it will give a divergent contribution when there is localization.

When β is small, $\varepsilon e^{-t^*} \simeq 1$ in any dimension. Thus the saddle point t^* suggests localization occurs in both 1D and 2D for all β and in 3D for small β . In 2D, this agrees with the predictions of localization by Abrahams, Anderson, Licciardello and Ramakrishnan [2] at any nonzero disorder. Although the saddle point analysis has some appeal, it is not easy to estimate large deviations away from t^* in one and two dimensions. In 3D, large deviations away from $t^* = 0$ are controlled for large β . See the discussion below.

The main theorem established in [14] states that in 3D, fluctuations around $t^* = 0$ are rare. Let $G_0 = (-\beta\Delta + \epsilon)^{-1}$ be the Green's function for the Laplacian.

Theorem 14 *If $d \geq 3$, and the volume $\Lambda \rightarrow \mathbb{Z}^d$, there is a $\bar{\beta} \geq 0$ such that for $\beta \geq \bar{\beta}$ then for all j ,*

$$\langle \cosh^8(t_j) \rangle \leq \text{Const} \quad (4.86)$$

where the constant is uniform in ϵ . This implies diffusion in a quadratic form sense: Let \mathcal{G} be given by (4.81) or (4.83). There is a constant C so that we have the quadratic form bound

$$\frac{1}{C} [f; G_0 f] \leq \sum_{x,y} \mathcal{G}_{\beta,\epsilon}(x, y) f(x) f(y) \leq C [f; G_0 f], \quad (4.87)$$

where $f(x)$ is nonnegative function.

Remarks. A weaker version of the lower bound in (4.87) appears in [14]. Yves Capdeboscq (private communication) showed (4.87) follows directly from the (4.86). This proof is explained in Appendix 4.12. The power 8 can be increased by making β larger. One expects pointwise diffusive bounds on $\mathcal{G}_{\beta,\epsilon}(x, y)$ to hold. However, in order to prove this one needs to show that the set $(j : |t_j| \leq M)$ percolates in a strong sense for some large M . This is expected to be true but has not yet been mathematically established partly because of the strong correlations in the t field.

The next theorem establishes localization for small β in any dimension. See [12].

Theorem 12 *Let $\varepsilon_x > 0$, $\varepsilon_y > 0$ and $\sum_{j \in \Lambda} \varepsilon_j \leq 1$. Then for all $0 < \beta < \beta_c$ (β_c defined below) the correlation function $\mathcal{G}_{\beta,\epsilon}(x, y)$, (4.83), decays exponentially with the distance $|x - y|$. More precisely:*

$$\mathcal{G}_{\beta,\epsilon}(x, y) = \langle \tilde{D}_{\beta,\epsilon}^{-1}(t)(x, y) \rangle \leq C_0 \left(\varepsilon_x^{-1} + \varepsilon_y^{-1} \right) \left[I_\beta e^{\beta(c_d-1)} c_d \right]^{|x-y|} \quad (4.88)$$

where $c_d = 2d - 1$, C_0 is a constant and

$$I_\beta = \sqrt{\beta} \int_{-\infty}^{\infty} \frac{dt}{\sqrt{2\pi}} e^{-\beta(\cosh t - 1)}.$$

Finally β_c is defined so that:

$$[I_\beta e^{\beta(c_d-1)} c_d] < [I_{\beta_c} e^{\beta_c(c_d-1)} c_d] = 1 \quad \forall \beta < \beta_c.$$

These estimates hold uniformly in the volume. Compare (4.88) with (4.2).

Remarks. The first proof of localization for the $\mathbb{H}^{2|2}$ model in 1D was given by Zirnbauer in [52]. Note that in 1D, $c_d - 1 = 0$ and exponential decay holds for all $\beta < \infty$ and the localization length is bounded by β when β is large. One expects that for 1D strips of width $|S|$ and β large, the localization length is proportional to $\beta|S|$. However, this has not yet been proved. The divergence in ε^{-1} is compatible with the Ward identity (4.4) and is a signal of localization.

4.8.1 Role of Ward Identities in the Proof

The proof of Theorems 3 and 4 rely heavily on Ward identities arising from internal SUSY. These are described below. For Theorem 14 we use Ward identities to bound fluctuations of the t field by getting bounds in 3D on $\langle \cosh^m(t_i - t_j) \rangle$. This is done by induction on the distance $|i - j|$. Once these bounds are established we use ϵ to get bounds for $\langle \cosh^p t \rangle$. For Theorem 12 we use the fact that for any region Λ , the partition function $Z_\Lambda = 1$.

If an integrable function S of the variables $x, y, z, \psi, \bar{\psi}$ is supersymmetric, i.e., it is invariant under transformations preserving

$$x_i x_j + y_i y_j + \bar{\psi}_i \psi_j - \psi_i \bar{\psi}_j$$

then $\int S = S(0)$. In horospherical coordinates the function S_{ij} given by

$$S_{ij} = B_{ij} + e^{t_i + t_j} (\bar{\psi}_i - \bar{\psi}_j)(\psi_i - \psi_j) \quad (4.89)$$

where

$$B_{ij} = \cosh(t_i - t_j) + \frac{1}{2} e^{t_i + t_j} (s_i - s_j)^2 \quad (4.90)$$

is supersymmetric. If i and j are nearest neighbors, $S_{ij} - 1$ is a term in the action \mathcal{A} given in (4.76) and it follows that the partition function $Z_\Lambda(\beta, \epsilon) \equiv 1$. More generally for each m we have

$$\langle S_{ij}^m \rangle_{\beta, \epsilon} = \langle B_{ij}^m [1 - m B_{ij}^{-1} e^{t_i + t_j} (\bar{\psi}_i - \bar{\psi}_j)(\psi_i - \psi_j)] \rangle_{\beta, \epsilon} \equiv 1. \quad (4.91)$$

Here we have used that $S_{ij}^m e^{-\mathcal{A}}$ is integrable for $\epsilon > 0$. The integration over the Grassmann variables in (4.91) is explicitly given by

$$G_{ij} = \frac{e^{t_i + t_j}}{B_{ij}} [(\delta_i - \delta_j); D_{\beta, \epsilon}(t)^{-1}(\delta_i - \delta_j)]_A \quad (4.92)$$

since the action is quadratic in $\bar{\psi}, \psi$. Thus we have the identity

$$< B_{ij}^m (1 - m G_{ij}) > \equiv 1. \quad (4.93)$$

Note that $0 \leq \cosh^m(t_i - t_j) \leq B_{ij}^m$. From the definition of $D_{\beta, \epsilon}$ given in (4.77) we see that for large β , G in (4.92) is typically proportional to $1/\beta$ in 3D. However, there are rare configurations of $t_k \ll -1$ with k on a surface separating i and j for which G_{ij} can diverge as $\epsilon \rightarrow 0$. In 2D, G_{ij} grows logarithmically in $|i - j|$ as $\epsilon \rightarrow 0$ even in the ideal case $t \equiv 0$.

Our induction starts with the fact that if i, j are nearest neighbors then it is easy to show that G_{ij} is less than β^{-1} for all t configurations. This is because of the factor $e^{t_i + t_j}$ in (4.92).

If $|i - j| > 1$ and $m G_{ij} \leq 1/2$, then (4.93) implies that

$$0 \leq < \cosh^m(t_i - t_j) > \leq < B_{ij}^m > \leq 2.$$

It is not difficult to show that one can get bounds on G_{ij} depending only on the t fields in a double cone with vertices at i and j . In fact for k far way from i, j , the dependence of G_{ij} on t_k is mild. Nevertheless, there is still no uniform bound on G_{ij} due to t fluctuations. We must use induction on $|i - j|$ and SUSY characteristic functions of the form $\chi\{S_{ij} \leq r\}$, to prove that configurations with $1/2 \leq m G_{ij}$, with $\beta \gg m \gg 1$, are rare for large β in 3D. We combine these estimates with the elementary inequality $B_{ik} B_{kj} \geq 2 B_{ij}$ for $i, j, k \in \mathbb{Z}^d$ to go to the next scale. See [14] for details.

The proof of the localized phase is technically simpler than the proof of Theorem 14. Nevertheless, it is of some interest because it shows that $\mathbb{H}^{2|2}$ sigma model reflects the localized as well as the extended states phase in 3D. The main idea relies on the following lemma. Let M be an invertible matrix indexed by sites of Λ and let γ denote a self avoiding path starting at i and ending at j . Let M_{ij}^{-1} be matrix elements of the inverse and let M_{γ^c} be the matrix obtained from M by striking out all rows and columns indexed by the vertices covered by γ .

Lemma. Let M and M_{γ^c} be as above, then

$$M_{ij}^{-1} \det M = \sum_{\gamma: i, j} [(-M_{i j_1})(-M_{j_1 j_2}) \cdots (-M_{j_m j})] \det M_{\gamma^c}$$

where the sum ranges over all self-avoiding paths γ connecting i and j , $\gamma_{i, j} = (i, j_1, \dots, j_m, j)$, with $m \geq 0$.

Apply this lemma to

$$M = e^{-t} D_{\beta,\varepsilon}(t) e^{-t} \equiv \tilde{D}_{\beta,\varepsilon}(t) = -\beta \Delta + V(t) + \varepsilon e^{-t} \quad (4.94)$$

and notice that for this M , for all non-zero contributions, γ are nearest neighbor self-avoiding paths and that each step contributes a factor of β . To prove the lemma, write the determinant of M as a sum over permutations of $j \in \Lambda$ indexing M . Each permutation can be expressed as a product of disjoint cycles covering Λ . Let E_{ij} denote the elementary matrix which is 0 except at ij place where it is 1. The derivative in s of $\det(M + sE_{ji})$ at $s = 0$ equals $M_{ij}^{-1} \det M$ and selects the self-avoiding path in the cycle containing j and i . The other loops contribute to $\det M_{\gamma^c}$. By (4.83) and (4.94) we have

$$\mathcal{G}_{\beta,\varepsilon}(x, y) = \langle M_{xy}^{-1} \rangle = \int e^{-\beta \mathcal{L}(t)} M_{xy}^{-1} [\det M]^{1/2} \prod_j \frac{dt_j}{\sqrt{2\pi}}. \quad (4.95)$$

Note the factors of e^{-t_j} appearing in (4.78) have been absorbed into the determinant. Now write

$$M_{xy}^{-1} [\det M]^{1/2} = \sqrt{M_{xy}^{-1}} \sqrt{M_{xy}^{-1} \det M}.$$

The first factor on the right hand side is bounded by $\epsilon_x^{-1/2} e^{t_x/2} + \epsilon_y^{-1/2} e^{t_y/2}$. For the second factor, we use the lemma. Let $\mathcal{L} = \mathcal{L}_\gamma + \mathcal{L}_{\gamma^c} + \mathcal{L}_{\gamma,\gamma^c}$ where \mathcal{L}_γ denotes the restriction of \mathcal{L} to γ . Then by supersymmetry,

$$\int e^{-\beta \mathcal{L}_{\gamma^c}} [\det M_{\gamma^c}]^{1/2} \prod_j \frac{dt_j}{\sqrt{2\pi}} \equiv 1$$

we can bound

$$0 \leq \mathcal{G}_{\beta,\varepsilon}(x, y) \leq \sum_{\gamma_{xy}} \sqrt{\beta}^{|\gamma_{xy}|} \int e^{-\beta \mathcal{L}_\gamma + \beta \mathcal{L}_{\gamma,\gamma^c}} [\epsilon_x^{-1/2} e^{t_x/2} + \epsilon_y^{-1/2} e^{t_y/2}] \prod_j \frac{dt_j}{\sqrt{2\pi}}$$

where $|\gamma_{xy}|$ is the length of the self-avoiding path from x to y . The proof of Theorem 12 follows from the fact that the integral along γ is one dimensional and can be estimated as a product. See [12] for further details.

4.9 Efetov's Sigma Model

In this final section we present a very brief description of the Efetov sigma model for Hermitian matrices, following [17, 18, 36, 51]. This sigma model is the basis for most random band matrix calculations in theoretical physics. Unfortunately, the

mathematical analysis of this model is still limited to 1D. Even in this case our analysis is far from complete. The ideas of this section are quite similar to those in Sect. 4.7 leading to (4.68) and (4.69) except that we now include Grassmann variables.

In order to obtain information about averages of the form (4.1) we introduce a field with four components

$$\Phi_j = (z_j, w_j, \psi_j, \chi_j) \quad (4.96)$$

where z, w are complex fields and ψ, χ are Grassmann fields. Let $L = \text{diag}(1, -1, 1, 1)$, and $\Lambda = \text{diag}(1, -1, 1, -1)$. For a Hermitian matrix H , define the action

$$A(E, \epsilon) = \Phi^* \cdot L \{i(H - E) + \epsilon \Lambda\} \Phi. \quad (4.97)$$

Note that the signature of L is chosen so that the z and w variables appear as complex conjugates of each other as they do in (7.1). Then we have the identity:

$$|(E - i\epsilon - H)^{-1}(0, j)|^2 = \int z_0 \bar{z}_j w_0 \bar{w}_j e^{-A(E, \epsilon)} D\Phi \quad (4.98)$$

where

$$D\Phi \equiv Dz Dw D\psi D\chi.$$

Without the observable $z_0 \bar{z}_j w_0 \bar{w}_j$, $\int e^{-A} = 1$. The integral over Gaussian H can be calculated as in (4.55) and will produce a quartic interaction in Φ . However, now the Hubbard–Stratonovich transformation, which is usually used in the Bosonic sector, involves a subtle analytic continuation first worked out in [41], see also [28, 51].

Now let us define a matrix of the form

$$M = \begin{pmatrix} [BB] & [BF] \\ [FB] & [FF] \end{pmatrix} \quad (4.99)$$

where each block is a 2×2 matrix. M will be called a *supermatrix* if the diagonal blocks, BB and FF are made of commuting (even elements) variables while the off diagonal blocks FB and BF consist of odd elements in the Grassmann algebra. Define the supertrace

$$\text{Str}(M) \equiv \text{Tr}([BB] - [FF]).$$

Note that for supermatrices A and B we have $\text{Str}(AB) = \text{Str}(BA)$. We define the adjoint of a supermatrix M by

$$M^\dagger = \begin{pmatrix} [BB]^* & [FB]^* \\ -[BF]^* & [FF]^* \end{pmatrix}.$$

The symbol $*$ denotes the usual transpose followed by conjugation. For Grassmann variables we have $\overline{\psi_a \psi_b} = \bar{\psi}_a \bar{\psi}_b$. But $\bar{\bar{\psi}} = -\psi$ so that \dagger is an involution and

$$\Phi_1^\dagger(M\Phi_2) = (M^\dagger\Phi_1)^\dagger\Phi_2.$$

For $\epsilon = 0$ the action is invariant under the action of matrices T in $SU(1; 1|2)$ which satisfy:

$$T^\dagger L T = L. \quad (4.100)$$

As in (4.68) and (4.69) the spin or field of the sigma model is given by matrices

$$S_j = T_j^{-1} \Lambda T_j \quad (4.101)$$

as T ranges over $SU(1; 1|2)$. It is the orbit of a critical point, which is proportional to Λ under the action of $SU(1, 1|2)$. Thus the matrix S ranges over a supersymmetric space $U(1, 1|2)/(U(1|1) \times U(1, 1))$. A general discussion of more general supersymmetric spaces appears in [53]. The SUSY sigma model has a Gibbs density defined by

$$\exp\{-\beta \text{Str} \sum_{j \sim j'} (S_j - S_{j'})^2 - \epsilon \text{Str} \sum_j \Lambda S_j\}. \quad (4.102)$$

In a one dimensional chain of length L with $\epsilon = 0$ except at the end points, the Grassmann variables can be explicitly integrated over producing the formula (4.8). In higher dimensions the action (4.102) is harder to analyze because of the Fermion–Boson coupling of the BB sector (hyperbolic sigma model) and the FF sector (Heisenberg model). An explicit parametrization of the 4×4 matrices S_j and integration measure is given in [17, 18, 36]. Fluctuations about the saddle should produce massless modes—Goldstone Bosons in 3D.

4.10 Appendix A: Gaussian and Grassmann Integration

Let $z = x + iy$ with $x, y \in \mathbb{R}$. Let $dz = dxdy/\pi$ and suppose $\text{Re } a > 0$, $a \in \mathbb{C}$. Then

$$\int e^{-a\bar{z}z} dz = \pi^{-1} \iint e^{-ar^2} r dr d\theta = a^{-1}. \quad (4.103)$$

Also

$$\frac{1}{\sqrt{2\pi}} \int e^{-ax^2/2} dx = a^{-1/2}.$$

In the multi-dimensional case let $z = (z_1, z_2, \dots, z_n)$, $z^* = \bar{z}^t$. For an $n \times n$ matrix A with $\text{Re } A > 0$ as a quadratic form

$$\int e^{-[z; Az]} D\bar{z} = (\det A)^{-1} \quad \text{where} \quad D\bar{z} = \prod_1^n dx_i dy_i / \pi. \quad (4.104)$$

Recall the notation $[z; Az] \equiv \sum \bar{z}_j A_{ij} z_j = z^* A z$. To prove the formula, first note that it holds in the Hermitian case by diagonalization. Since both sides of the

equation are analytic in the matrix entries, and agree for the Hermitian case, the result follows by uniqueness of analytic continuation. The expectation with respect to A is defined by

$$\langle \cdot \rangle_A \equiv \det(A) \int e^{-z^* A z} \cdot D z$$

and from integration by parts, the pair correlation is given by

$$\langle z_j \bar{z}_k \rangle_A = A_{jk}^{-1}. \quad (4.105)$$

Note that $\langle z_j \bar{z}_k \rangle_A = \langle \bar{z}_j z_k \rangle_A = 0$. This is because the integral is invariant under the global transform $z \rightarrow e^{i\phi} z$, $\bar{z} \rightarrow e^{-i\phi} \bar{z}$. The generating function is given by

$$\langle e^{z^* v + w^* z} \rangle_A = e^{w^* A^{-1} v} = e^{[w; A^{-1} v]}.$$

This identity follows by changing variables: $z \rightarrow z - A^{-1} v$ and $\bar{z} \rightarrow \bar{z} - (A^t)^{-1} \bar{w}$.

For real variables $x = (x_1, \dots, x_n)$ if A is symmetric and positive

$$\int e^{-[x; Ax]/2} D x = (\det A)^{-1/2} \quad \text{where} \quad D x = \prod_i dx_i / \sqrt{2\pi}. \quad (4.106)$$

Its generating function $\langle e^{[x; y]} \rangle_A = e^{[y; A^{-1} y]/2}$, is obtained by completing the square.

There are similar formulas for integration over $N \times N$ matrices:

$$\int e^{-N \text{Tr} H^2/2} e^{i \text{Tr} M H} D H = e^{-\text{Tr} M^2/2N} \int e^{-N \text{Tr} H^2/2} D H. \quad (4.107)$$

For the case of band matrices the generating function is

$$\langle e^{i \text{Tr} H M} \rangle = e^{-\langle (\text{tr} H M)^2 \rangle / 2} = e^{-1/2 \sum J_{ij} M_{ji} M_{ij}}. \quad (4.108)$$

4.10.1 Grassmann Integration

Grassmann variables $\psi_i, \bar{\psi}_j$ are anti-commuting variables $1 \leq i, j \leq N$ satisfying $\psi_j^2 = \bar{\psi}_j^2 = 0$, and $\psi_j \psi_i = -\psi_i \psi_j$. Also

$$\psi_j \psi_i = -\psi_i \psi_j, \quad \bar{\psi}_j \bar{\psi}_i = -\bar{\psi}_i \bar{\psi}_j. \quad (4.109)$$

The $\bar{\psi}_j$ is simply convenient notation for another independent family of Grassmann variables which anti-commute with ψ . Even monomials in the Grassmann variables and complex numbers commute with Grassmann variables. The polynomials in these variables form a Z_2 graded algebra, with the even and odd monomials belonging to the even and odd gradings respectively. One way to think about the

Grassmann variables is to let $\psi_j = dx_j$ and $\bar{\psi}_j = dy_j$ and consider the product as the wedge product in the theory of differential forms.

The Grassmann integral, defined below, plays an important role in many parts of physics. It is an extremely efficient and useful notation for the analysis of interacting Fermi systems, Ising models (Free Fermions), and SUSY. Although most of the time we shall eliminate the Grassmann variables by integrating them out, they are nevertheless an essential tool for obtaining the identities we shall analyze. See [1, 4, 24, 36, 40] for more details about Grassmann integration.

We define integration with respect to

$$D\psi \equiv \prod_j^N d\bar{\psi}_j d\psi_j \quad (4.110)$$

as follows. For $N = 1$

$$\int (a\psi_1\bar{\psi}_1 + b\psi_1 + c\bar{\psi}_1 + d) D\psi = a.$$

The general rule is that the integral of a polynomial in $2N$ variables with respect to $D\psi$ equals the coefficient of the top monomial of degree $2N$ ordered as $\prod^N \psi_j \bar{\psi}_j$. Note that since the factors in the product are even, their order does not matter. Any element of the Grassmann algebra can be expressed as a polynomial and the top monomial can always be rearranged using the anti-commutation rules so that it coincides with $\prod^N \psi_j \bar{\psi}_j$.

To differentiate a Grassmann monomial, use the rule $\frac{\partial}{\partial \psi_j} \psi_k = \delta_{jk}$. The derivative anti-commutes with other Grassmann variables. We have

$$\frac{\partial}{\partial \psi_j} \psi_j \prod_{k \neq j} \psi_k = \prod_{k \neq j} \psi_k.$$

To differentiate a general monomial in ψ_j , use linearity and the anti-commutation relations so that it is of the above form. If ψ_j is not a factor then the derivative is 0.

For any $N \times N$ matrix A we have the following analog of Gaussian integration

$$\int e^{-[\psi; A\psi]} D\psi = \det A \quad \text{where} \quad [\psi; A\psi] = \sum \bar{\psi}_i A_{ij} \psi_j. \quad (4.111)$$

Moreover,

$$\langle \psi_i \bar{\psi}_j \rangle \equiv \det A^{-1} \int \psi_i \bar{\psi}_j e^{-[\psi; A\psi]} D\psi = A_{ij}^{-1}. \quad (4.112)$$

More generally for a polynomial F in $\psi, \bar{\psi}$ we can integrate by parts to obtain

$$\int \psi_i F e^{-[\psi; A\psi]} D\psi = \sum_j A_{ij}^{-1} \int \left\{ \frac{\partial}{\partial \bar{\psi}_j} F \right\} e^{-[\psi; A\psi]} D\psi. \quad (4.113)$$

To prove this formula we use

$$\sum_j A_{ij}^{-1} \int \frac{\partial}{\partial \bar{\psi}_j} (e^{-[\psi; A\psi]} F) D\psi = 0.$$

Let us establish (4.111) in the simplest case:

$$\int e^{-a\bar{\psi}_1\psi_1} D\psi = \int (1 - a\bar{\psi}_1\psi_1) D\psi = a \int \psi_1 \bar{\psi}_1 D\psi = a.$$

Exercise. Show that if A is a 2×2 matrix, (4.111) holds.

To prove the general case note that the exponential can be written as a product $\prod_i (1 - \sum_j A_{ij} \bar{\psi}_i \psi_j)$ and we look at the terms:

$$\sum A_{1j_1} A_{2j_2} \dots A_{N,j_N} \int \bar{\psi}_1 \psi_{j_1} \bar{\psi}_2 \psi_{j_2} \dots \bar{\psi}_N \psi_{j_N} D\psi.$$

The j_i are distinct and hence are a permutation of $1 \dots N$. The integral then is the sign of the permutation and thus we obtain the determinant. The generating function is given by

$$\langle e^{\bar{\rho}'\psi + \bar{\psi}'\rho} \rangle = \det A^{-1} \int e^{-[\psi; A\psi]} e^{\bar{\rho}'\psi + \bar{\psi}'\rho} D\psi = e^{\bar{\rho}' A^{-1} \rho} \quad (4.114)$$

where $\rho, \bar{\rho}$ are independent Grassmann variables.

4.10.2 Polar Coordinates for Grassmann and Bosonic Matrices

Grassmann integrals can also be expressed as integrals over unitary matrices. Let $d\mu(U)$ denote the Haar measure on $U(m)$. Given a family of Grassman variables $\bar{\psi}_j^\alpha, \psi_j^\alpha$ with $1 \leq \alpha \leq N$ and $1 \leq j \leq m$, define the matrix $M_{ij} = \psi_i \cdot \bar{\psi}_j$, where the dot product is the sum over α . Then for a smooth function F of M we have

$$\int F(M) D\psi = C_{N,m} \int F(U) \det(U)^{-N} d\mu(U) \quad (4.115)$$

where $U \in U(m)$. For example for $m = 1$

$$\int e^{a\psi \cdot \bar{\psi}} D\psi = a^N = \frac{N!}{2\pi} \int \exp(ae^{i\theta}) e^{-iN\theta} d\theta.$$

It suffices to check (4.115) when $F = e^{tr Q M}$ for a general matrix Q since any function can be approximated by linear combinations of such exponentials. From (4.111) the left side is equal to $\det(Q)^N$. These expressions agree since

$$\int e^{tr Q U} \det(U)^{-N} d\mu(U) \propto \det(Q)^N.$$

To prove this statement first suppose that $Q = e^{i\tau H}$ is a unitary matrix. Our identity now follows by the invariance of Haar measure. Since both sides are analytic in t and agree for t real, they agree for all matrices Q by taking t complex.

For bosonic fields the analog of the above relation is given by the generalized gamma function or Ingham–Siegel identity, see [29]:

$$\int_{P>0} e^{-tr Q P} \det(P)^{N-m} dP \propto \det(Q)^{-N}.$$

Here P and Q denote positive Hermitian $m \times m$ matrices and dP denotes the flat measure restricted positive matrices. We must assume that $N \geq m$. Hence, if we set $M_{ij} = \bar{z}_i \cdot z_j$ we have

$$\int F(M) Dz = C'_{N,m} \int_{P>0} F(P) \det(P)^{N-m} dP. \quad (4.116)$$

Note that $\det(P)^{-m} dP$ is proportional to the measure $d\mu(P)$ which is invariant under the transformation $P \rightarrow g^* P g$, for $g \in GL_m(\mathbb{C})$. For example dt/t is invariant under $t \rightarrow at$ for $a > 0$.

We may think of U and P as the polar or collective variable for the Grassmann and Bosonic variables respectively and their eigenvalues could be referred to as the radial coordinates.

Let $\Phi = (z, \psi)$ denote a column vector of with n bosonic components, z_j , and n Grassmann components, ψ_j . Define Φ^* to be the transpose conjugate of Φ . If M denotes a supermatrix with $n \times n$ blocks of the form (4.99), then the standard SUSY integration formula

$$\int e^{-\Phi^* M \Phi} Dz D\psi = S \text{Det}^{-1}(M)$$

holds, where the superdeterminant is given by

$$S \text{Det}(M) = \det^{-1}([FF]) \det([BB] - [BF][FF]^{-1}[FB]).$$

This formula may be derived by first integrating out the ψ variables using (4.114) and then integrating the z variables. An equivalent formula given by

$$S \text{Det}(M) = \det([BB]) \det^{-1}([FF] - [FB][BB]^{-1}[BF])$$

is obtained by reversing the order of integration. If A and B are supermatrices then $SDet(AB) = SDet(A) SDet(B)$ and $\ln SDet A = Str A$.

Now consider a supermatrix with both Grassmann and bosonic components. For example, in the simplest case, ($m = 1$):

$$M = \begin{pmatrix} z \cdot \bar{z} & z \cdot \bar{\psi} \\ \psi \cdot \bar{z} & \psi \cdot \bar{\psi} \end{pmatrix}.$$

then for suitably regular F , the SUSY generalization of the above formulas is given by

$$\int F(M) D z D \psi \propto \int F(Q) SDet(Q)^N D Q.$$

Here Q is a 2×2 supermatrix and $m = 1$. For $m \geq 1$ of Q has the form

$$Q = \begin{pmatrix} P & \bar{\chi} \\ \chi & U \end{pmatrix}$$

with blocks of size $m \times m$ and

$$DQ \equiv dP d\mu(U) det^m(U) D\bar{\chi} D\chi$$

where $d\mu$ Haar measure on $U(m)$, and dP is the flat measure on positive Hermitian matrices. As a concrete example let us evaluate the integral using the above formula for $m = 1$ in the special case

$$\int e^{-a\bar{z} \cdot z + b\psi \cdot \bar{\psi}} D z D \psi = b^N a^{-N} = \int e^{-ap + b e^{i\theta}} SDet^N(Q) dp d\theta e^{i\theta} d\bar{\chi} d\chi$$

where

$$SDet^N(Q) = p^N e^{-iN\theta} (1 - N p^{-1} \bar{\chi} e^{-i\theta} \chi).$$

Note that in the contribution to the integral, only the second term of $SDet$ above contributes. It is of top degree equal to $N p^{N-1} e^{-i(N+1)\theta} \bar{\chi} \chi$.

Remarks. The above integral identity is a simple example of *superbozonization*, See [33] for a more precise formulation and proof as well as references.

4.11 Appendix B: Formal Perturbation Theory

In this appendix we will explain some formal perturbation calculations for the average Green's function of the random Schrödinger matrix when the strength of the disorder λ is small. For a rigorous treatment of perturbation theory, see for example [22]. At present it is not known how to rigorously justify this perturbation

theory when the imaginary part of the energy, ϵ is smaller than λ^4 . In fact the best rigorous estimates require $\epsilon \geq \lambda^p$, $p \leq 2 + \delta$, $\delta > 0$. We shall explain some of the challenges in justifying the perturbation theory. Similar perturbative calculations can be done for random band matrices and in special cases these can be justified using SUSY as explained in Sect. 4.5.

Let us write the Green's function for the random Schrödinger matrix as

$$G(E_\epsilon; j, k) = [E_\epsilon + \Delta - \lambda^2 \sigma(E_\epsilon) - \lambda v + \lambda^2 \sigma(E_\epsilon)]^{-1}(j, k).$$

We assume that the potential v_j are independent Gaussian random variables of 0 mean and variance 1. Note that we have added and subtracted a constant $\lambda^2 \sigma(E_\epsilon)$ and we shall now expand around

$$G_0(E_\epsilon; j, k) \equiv [E_\epsilon + \Delta - \lambda^2 \sigma(E_\epsilon)]^{-1}.$$

The first order term in λ vanishes since v has mean 0. The basic idea is to choose the constant $\sigma(E_\epsilon)$ so that after averaging, the second order term in λ vanishes:

$$\lambda^2 \langle G_0 v G_0 v G_0 - G_0 \sigma G_0 \rangle = 0 \quad \text{hence} \quad G_0(E_\epsilon; j, j) = \sigma(E_\epsilon) \quad (4.117)$$

holds for all j . This gives us an equation for $\sigma(E_\epsilon)$ which is the leading contribution to the self-energy. In field theory this is just self-consistent Wick ordering. When E lies in the spectrum of $-\Delta$, the imaginary part of $\sigma(E_\epsilon)$ does not vanish even as $\epsilon \rightarrow 0$. In fact, to leading order $\text{Im} \sigma(E)$ is proportional to the density of states for $-\Delta$. See (4.33). Thus to second order in perturbation theory we have $\langle G(E_\epsilon; j, k) \rangle \approx G_0(E_\epsilon; j, k)$. Note that since the imaginary part of $\sigma(E_\epsilon) > 0$, $G_0(E_\epsilon; j, k)$ will decay exponentially fast in $\lambda^2 |j - k|$ uniformly for $\epsilon > 0$.

Exponential decay is also believed to hold for $\langle G(E_\epsilon; j, k) \rangle$ in all dimensions but has only been proved in one dimension. This decay is *not* related to localization and should also hold at energies where extended states are expected to occur. If λ is large and if v has a Gaussian distribution then it is easy to show that in any dimension the average Green's function decays exponentially fast. One way to see this is to make a shift in the contour of integration $v_j \rightarrow v_j + i\delta$. Since λ is large it will produce a large diagonal term. We get a convergent random walk expansion for $\langle G \rangle$ by expanding in the off diagonal matrix elements.

Remark. Note that by the Ward identity (4.4) and (4.117) we have

$$\sum_j |G_0(E_\epsilon; 0, j)|^2 (\lambda^2 \text{Im} \sigma(E_\epsilon) + \epsilon) = \text{Im} \sigma(E_\epsilon). \quad (4.118)$$

We now explain a problem in justifying perturbation theory. First note that one could proceed by defining higher order corrections to the self-energy, $\lambda^4 \sigma_4(E_\epsilon, p)$, where p is the Fourier dual of the variable j so that the fourth order perturbation

contribution vanishes. In this case σ acts as a convolution. However, this perturbative scheme is not convergent and must terminate at some order. We wish to estimate the remainder. This is where the real difficulty appears to lie. Consider the remainder term written in operator form:

$$< [G_0(E_\epsilon) \lambda v]^n G(E_\epsilon) > .$$

For brevity of notation we have omitted the contributions of the self-energy $\lambda^2 \sigma$ which should also appear in this expression. Suppose we use the Schwarz inequality to separate $[G_0(E_\epsilon) \lambda v]^n$ from G . Then we have to estimate the expression:

$$< [G_0(E_\epsilon) \lambda v]^n \cdot [\bar{G}_0(E_\epsilon) \lambda v]^n > .$$

Here \bar{G}_0 denotes the complex conjugate of G_0 . The high powers of λ may suggest that this is a small term. However, this is not so. If we write the above in matrix form we have

$$\begin{aligned} \lambda^{2n} < \sum_{j,k} [G_0(E_\epsilon; 0, j_1) v_{j_1} G_0(E_\epsilon; j_1, j_2) v_{j_2} \dots v_{j_{n-1}} G_0(E_\epsilon; j_{n-1}, j_n) v_{j_n} \\ \times \bar{G}_0(E_\epsilon; 0, k_1) v_{k_1} \bar{G}_0(E_\epsilon; k_1, k_2) v_{k_2} \dots v_{k_{n-1}} \bar{G}_0(E_\epsilon; k_{n-1}, j_n) v_{k_n}] > . \end{aligned} \quad (4.119)$$

When computing the average over v the indices $\{j, k\}$ must be paired otherwise the expectation vanishes since $< v_j > = 0$ for each j . There are $n!$ such pairings, each of which gives rise to a graph. The pairing of any adjacent j_i, j_{i+1} or k_i, k_{i+1} is canceled by the self-energy contributions which we have omitted. We shall next discuss some other simple pairings.

After canceling the self-energy, the leading contribution to (4.119) should be given by *ladder graphs* of order n . These are a special class of graphs obtained by setting $j_i = k_i$ and summing over the vertices. Assuming G_0 is translation invariant, this sum is approximately given by:

$$[\lambda^2 \sum_j |G_0(E_\epsilon; 0, j)|^2]^n = \left[\frac{\lambda^2 \text{Im} \sigma(E_\epsilon)}{\lambda^2 \text{Im} \sigma(E_\epsilon) + \epsilon} \right]^n = \left[1 + \frac{\epsilon}{\lambda^2 \text{Im} \sigma(E_\epsilon)} \right]^{-n}.$$

The right hand side of this equation is obtained using (4.118). Note that the right hand side of this equation is not small if $\epsilon \leq \lambda^p$ unless $n \gg \lambda^{-(p-2)}$. Although contributions obtained from other pairings give smaller contributions, there are about $n!$ terms of size λ^{2n} . Hence we must require $n \ll \lambda^{-2}$ so that the number of graphs is offset by their size $n! \lambda^{2n} \ll 1$. Thus, $p < 4$ and this naive method of estimating $G(E_\epsilon)$ can only work for $\epsilon \gg \lambda^4$.

4.11.1 Leading Contribution to Diffusion

In three dimensions, the leading contribution to $\langle |G(E_\epsilon; 0, x)|^2 \rangle$ is expected to be given by the sum of ladder graphs described above - denoted L . Let

$$K(E_\epsilon; p) = \sum_j e^{i \cdot j \cdot p} |G_0(E_\epsilon; 0, j)|^2.$$

Note that $K(p)$ is analytic for small values of p^2 since $Im\sigma > 0$. Then the sum of the ladders is given for $p \approx 0$ by

$$\begin{aligned} \hat{L}(p) &= K(p)(1 - \lambda^2 K(p))^{-1} \approx \frac{K(0)}{1 - \lambda^2 K(0) - \frac{1}{2}\lambda^2 K''(0)p^2} \\ &\approx \frac{Im\sigma}{\lambda^4 Im\sigma K''(0) p^2/2 + \epsilon}. \end{aligned}$$

In the last approximation we have used the Ward identity (4.118) to get

$$1 - \lambda^2 K(0) = \frac{\epsilon}{\lambda^2 Im\sigma + \epsilon} \quad \text{so} \quad K(0) \approx \lambda^{-2}.$$

Thus the sum of ladder graphs produces quantum diffusion as defined by (4.3) with $D_0 = \lambda^4 Im\sigma \hat{K}''(0)/2$. Since $K''(0) \propto \lambda^{-6}$, we see that D_0 is proportional to λ^{-2} for small λ . We refer to the work of Erdős, Salmhofer and Yau [21, 22], where such estimates were proved in 3D for $\epsilon \approx \lambda^{2+\delta}$, $\delta > 0$. At higher orders of perturbation theory, graphs in 3D must be grouped together to exhibit cancellations using Ward identities so that the bare diffusion propagator $1/p^2$ does not produce bogus divergences of the form $\int (1/p^2)^m$ for $m \geq 2$. These higher powers are offset by a vanishing at $p=0$ of the vertex joining such propagators. In both 1D and 2D this perturbation theory breaks down at higher orders. There is a divergent sum of crossed ladder graphs obtained by setting $j_i = k_{n-i}$ above. The sum of such graphs produces an integral of the form

$$\int_{|p| \leq 1} (p^2 + \epsilon)^{-1} d^2 p$$

which diverges logarithmically in 2D as $\epsilon \rightarrow 0$. This divergence is closely related to localization which is expected for small disorder in 2D [2].

4.11.2 Perturbation Theory for Random Band Matrices

To conclude this appendix, we briefly show how to adapt the perturbative methods described above to the case of random band matrices. Let H be a Gaussian random

band matrix with covariance as in (4.14). Let

$$G(E_\epsilon) = (E_\epsilon - \lambda^2 \sigma - \lambda H + \lambda^2 \sigma)^{-1}.$$

Here the parameter λ is a book keeping device and it will be set equal to 1. Define $G_0(E_\epsilon) = [E_\epsilon - \lambda^2 \sigma(E_\epsilon)]^{-1}$ and perturb $G(E_\epsilon)$ about G_0 . Note that in this case G_0 is a scalar multiple of the identity. Proceeding as above we require that to second order perturbation theory in λ vanishes. From (4.13) and (4.14) we have

$$\sum_k \langle H_{jk} H_{kj} \rangle = \sum_k J_{jk} = 1$$

and we obtain the equation

$$G_0 = [E_\epsilon - \lambda^2 \sigma(E_\epsilon)]^{-1} = \sigma(E_\epsilon).$$

This is a quadratic equation for σ and when we set $\lambda = 1$ and take the imaginary part we recover Wigner's density of states. For a band matrix of width W the corrections at λ^4 to the Wigner semicircle law is formally of order $1/W^2$. If apply the Schwarz inequality as above we will encounter similar difficulties. In fact it is known that $\epsilon \gg W^{-2/5}$ is needed to ensure convergence of perturbation theory. However, by using different methods, Erdős et al. [23] and Sodin [44] establish control of the DOS to scale $\epsilon \approx W^{-1}$. For 3D Gaussian band matrices with covariance given by (4.14) these difficulties are avoided with nonperturbative SUSY methods [7, 13] and ϵ may be sent to zero for fixed large W . The calculation of bare diffusion constant explained above for random Schrödinger matrices can be applied to RBM. In this case $D_0 \approx W^2$.

We now illustrate the perturbation theory for the two site N -orbital model:

$$M = \begin{pmatrix} \lambda A & I \\ I & \lambda B \end{pmatrix}$$

where A and B are independent $N \times N$ GUE matrices and I is the $N \times N$ identity matrix. Define

$$G_0^{-1} = \begin{pmatrix} E_\epsilon - \lambda^2 \sigma & I \\ I & E_\epsilon - \lambda^2 \sigma \end{pmatrix}.$$

By requiring second order perturbation theory to vanish as above we obtain a cubic equation for the self-energy $\sigma = (E_\epsilon - \lambda^2 \sigma)[(E_\epsilon - \lambda^2 \sigma)^2 - 1]^{-1}$. This equation is analogous to (4.117). The imaginary part of $\sigma(E)$ is proportional to the density of states for M for large N . To calculate finer properties of the density of states one can apply the SUSY methods of Sects. 4.3–4.5 and it can be shown that the main saddle point will coincide the self-energy σ . Note that the density of states for M no longer looks like the semicircle law, in fact it is peaked near ± 1 when λ is small.

Nevertheless, the local eigenvalue spacing distribution is expected to be equal to that of GUE after a local rescaling.

Remark. One of the major challenges in the analysis of random band or random Schrödinger matrices is to find a variant of perturbation theory which allows one to analyze smaller values of ϵ as a function of W or λ . For certain Gaussian band matrices SUSY gives much better estimates but the class of models to which it applies is relatively narrow. Perhaps there is a method which combines perturbative and SUSY ideas and emulates both of them.

4.12 Appendix C: Bounds on Green's Functions of Divergence Form

In this appendix we will show how to obtain upper and lower bounds on (4.87)

$$\langle [fe'; [D_{\beta,\epsilon}(t)]^{-1} fe'] \rangle = \sum_{x,y} \mathcal{G}_{\beta,\epsilon}(x, y) f(x) f(y)$$

in terms of $[f; G_0 f] = [f; (\beta \nabla^* \nabla + \epsilon)^{-1} f]$ for $f(j) \geq 0$. Recall $[\cdot; \cdot]$ is the scalar product and that $D_{\beta,\epsilon}(t)$ is given by (4.71) or (4.77). We shall suppose that $d \geq 3$ and that $\langle \cosh^8(t_j) \rangle$ is bounded as in Theorem 14 of Sect. 4.8. For brevity let

$$G_t(i, j) = [D_{\beta,\epsilon}(t)]^{-1}(i, j).$$

We first prove the lower bound. For any two real vectors X and Y we have the inequality

$$X \cdot X \geq 2X \cdot Y - Y \cdot Y.$$

Let

$$X_1(j, \alpha) = \sum_k e^{(t_j + t_j + e_\alpha)/2} \nabla_\alpha G_t(j, k) f(k) e^{t_k}$$

where e_α is the unit vector in the α direction. Define

$$X_2(j) = \sqrt{\epsilon} e^{t_j/2} \sum_k G_t(j, k) f(k) e^{t_k}.$$

If we set $X = (X_1, X_2)$ we see that $\langle X \cdot X \rangle = [f; \mathcal{G}_{\beta,\epsilon} f]$. We shall define Y to be proportional to X with G_t replaced by G_0 ,

$$Y_1(j, \alpha) = a e^{-2t_0} \sum_k e^{(t_j + t_j + e_\alpha)/2} \nabla_\alpha G_0(j, k) f(k) e^{t_k}$$

and

$$Y_2(j) = a e^{-2t_0} \sqrt{\epsilon} e^{t_j/2} \sum_k G_0(j, k) f(k) e^{t_k}.$$

The constant a is chosen so that the error term $Y \cdot Y$ is small. By integrating by parts we have

$$X \cdot Y = a [f e^t; G_0 f e^t] e^{-2t_0}.$$

Since $\langle e^{t_k + t_j - 2t_0} \rangle \geq 1$ by Jensen's inequality and translation invariance, we get the desired lower bound on $\langle X \cdot Y \rangle$ in terms of $a [f; G_0 f]$. The error term

$$\langle Y_1 \cdot Y_1 \rangle = a^2 \langle [\nabla G_0 f e^t; e^{t_j + t_{j'}} \nabla G_0 f e^t] e^{-4t_0} \rangle \leq C a^2 [f; G_0 f].$$

The last inequality follows from our bounds on $\langle \cosh^4(t_j - t_k) \rangle$ which can be bounded by $\langle \cosh^8 t_j \rangle$. Note that we also need to use $f \geq 0$ and

$$\sum_{ijk} f(i) |\nabla G_0(i, j)| |\nabla G_0(j, k)| f(k) \leq \text{Const} [f; G_0 f] \quad (4.120)$$

which holds in three dimensions.

$$\langle Y_2 \cdot Y_2 \rangle = a^2 \epsilon \langle [e^t G_0 e^t f; G_0 e^t f] e^{-4t_0} \rangle \leq C a^2 [f; G_0 f].$$

In the last inequality we bound $\langle e^{-t_0} \rangle$ as well as $\langle \cosh^4(t_j - t_k) \rangle$. The parameter a is chosen small so that $\langle X \cdot Y \rangle$ is the dominant term.

The upper bound is a standard estimate obtained as follows. Let $L_0 = \beta \nabla^* \nabla + \epsilon$ and $G_0 = L_0^{-1}$. Then

$$[f e^t; G_t f e^t] = [L_0 G_0 f e^t; G_t f e^t].$$

Integrating by parts we see that the right side equals

$$[\sqrt{\beta} e^{-(t_j + t_{j'})/2} \nabla G_0 f e^t; \sqrt{\beta} e^{(t_j + t_{j'})/2} \nabla G_t f e^t] + [\sqrt{\epsilon} e^{-t/2} G_0 f e^t; \sqrt{\epsilon} e^{t/2} G_t f e^t].$$

By the Schwarz inequality we get

$$\begin{aligned} & [f e^t; G_t f e^t] \\ & \leq [f e^t; G_t f e^t]^{1/2} \left(\beta \sum_j |\nabla_j (G_0 f e^t)|^2 e^{-(t_j + t_{j'})} + \epsilon \sum_j |(G_0 f e^t)(j)|^2 e^{-t_j} \right)^{1/2} \end{aligned}$$

Therefore

$$[fe^t; G_t fe^t] \leq \beta \sum_j |\nabla_j (G_0 e^t f)|^2 e^{-(t_j + t_{j'})} + \varepsilon \sum_j |(G_0 e^t f)(j)|^2 e^{-t_j}. \quad (4.121)$$

The proof of the upper bound now follows using (4.120). Note that in both inequalities we need to assume that $f \geq 0$.

The desired upper bound on the expectation (4.121) in terms of G_0 now follows from bounds on $\langle \cosh^8(t_j) \rangle$ and (4.120). Note that if the factor of e^t times f were not present we would not need to require $f \leq 0$ and to use (4.120).

Acknowledgements I wish to thank my collaborators Margherita Disertori and Martin Zirnbauer for making this review on supersymmetry possible. Thanks also to Margherita Disertori and Sasha Sodin for numerous helpful comments and corrections on early versions of this review and to Yves Capdeboscq for improving the lower bound in (4.87). I am most grateful to the organizers of the C.I.M.E. summer school, Alessandro Giuliani, Vieri Mastropietro, and Jacob Yngvason for inviting me to give these lectures and for creating a stimulating mathematical environment. Finally, I thank Abdelmalek Abdesselam for many helpful questions during the course of these lectures.

References

1. A. Abdesselam, Grassmann-Berezin calculus and theorems of the matrix-tree type. *Adv. Appl. Math.* **33**, 51–70 (2004). <http://arxiv.org/abs/math/0306396>
2. E. Abrahams, P.W. Anderson, D.C. Licciardello, T.V. Ramakrishnan, Scaling theory of localization: Absence of quantum diffusion in two dimensions. *Phys. Rev. Lett.* **42**, 673 (1979)
3. E.J. Beamd, A.L. Owczarek, J. Cardy, Quantum and classical localisation and the Manhattan lattice. *J. Phys. A Math. Gen.* **36**, 10251 (2003) [arXiv:cond-mat/0210359]
4. F.A. Berezin, *Introduction to Superanalysis* (Reidel, Dordrecht, 1987)
5. H. Brascamp, E. Lieb, On extensions of the Brunn-Minkowski and Prekopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *J. Func. Anal.* **22**, 366–389 (1976)
6. E. Brezin, V. Kazakov, D. Serban, P. Wiegman, A. Zabrodin, in *Applications of Random Matrices to Physics*. Nato Science Series, vol. 221 (Springer, Berlin, 2006)
7. F. Constantinescu, G. Felder, K. Gawedzki, A. Kupiainen, Analyticity of density of states in a gauge-invariant model for disordered electronic systems. *J. Stat. Phys.* **48**, 365 (1987)
8. P. Deift, *Orthogonal Polynomials, and Random Matrices: A Riemann-Hilbert Approach* (CIMS, New York University, New York, 1999)
9. P. Deift, T. Kriecherbauer, K.T.-K. McLaughlin, S. Venakides, X. Zhou, Uniform asymptotics for polynomials orthogonal with respect to varying exponential weights and applications to universality questions in random matrix theory. *Comm. Pure Appl. Math.* **52**, 1335–1425 (1999)
10. P. Diaconis, in *Recent Progress on de Finetti's Notions of Exchangeability*. Bayesian Statistics, vol. 3 (Oxford University Press, New York, 1988), pp. 111–125
11. M. Disertori, Density of states for GUE through supersymmetric approach. *Rev. Math. Phys.* **16**, 1191–1225 (2004)
12. M. Disertori, T. Spencer, Anderson localization for a SUSY sigma model. *Comm. Math. Phys.* **300**, 659–671 (2010). <http://arxiv.org/abs/0910.3325>
13. M. Disertori, H. Pinson T. Spencer, Density of states for random band matrices. *Comm. Math. Phys.* **232**, 83–124 (2002). <http://arxiv.org/abs/math-ph/0111047>
14. M. Disertori, T. Spencer, M.R. Zirnbauer, Quasi-diffusion in a 3D supersymmetric hyperbolic sigma model. *Comm. Math Phys.* **300**, 435 (2010). <http://arxiv.org/abs/0901.1652>

15. K.B. Efetov, in *Anderson Localization and Supersymmetry*, ed. by E. Abrahams. 50 Years of Anderson Localization (World Scientific, Singapore, 2010). <http://arxiv.org/abs/1002.2632>
16. K.B. Efetov, Minimum metallic conductivity in the theory of localization. JETP Lett. **40**, 738 (1984)
17. K.B. Efetov, Supersymmetry and theory of disordered metals. Adv. Phys. **32**, 874 (1983)
18. K.B. Efetov, *Supersymmetry in Disorder and Chaos* (Cambridge University Press, Cambridge, 1997)
19. L. Erdős, Universality of Wigner random matrices: A survey of recent results. Russian Math. Surveys. **66**(3) 507–626 (2011). <http://arxiv.org/abs/1004.0861>
20. L. Erdős, B. Schlein, H-T. Yau, Local semicircle law and complete delocalization for Wigner random matrices. Comm. Math. Phys. **287**, 641–655 (2009)
21. L. Erdős, M. Salmhofer, H-T. Yau, Quantum diffusion of the random Schrödinger evolution in the scaling limit. Acta Math. **200**, 211 (2008)
22. L. Erdős, in Les Houches Lectures: *Quantum Theory from Small to Large Scales*, Lecture notes on quantum Brownian motion, vol. 95, <http://arxiv.org/abs/1009.0843>
23. L. Erdős, H.T. Yau, J. Yin, Bulk universality for generalized Wigner matrices. to appear in Prob. Th. and Rel.Fields. [arxiv:1001.3453]
24. J. Feldman, H. Knörrer, E. Trubowitz, in *Fermionic Functional Integrals and the Renormalization Group*. CRM Monograph Series (AMS, Providence, 2002)
25. J. Fröhlich, T. Spencer, The Kosterlitz Thouless transition. Comm. Math. Phys. **81**, 527 (1981)
26. J. Fröhlich, T. Spencer, On the statistical mechanics of classical Coulomb and dipole gases. J. Stat. Phys. **24**, 617–701 (1981)
27. J. Fröhlich, B. Simon, T. Spencer, Infrared bounds, phase transitions and continuous symmetry breaking. Comm. Math. Phys. **50**, 79 (1976)
28. Y.V. Fyodorov, in *Basic Features of Efetov's SUSY*, ed. by E. Akkermans et al. Mesoscopic Quantum Physics (Les Houches, France, 1994)
29. Y.V. Fyodorov, Negative moments of characteristic polynomials of random matrices. Nucl. Phys. B **621**, 643–674 (2002). <http://arXiv.org/abs/math-ph/0106006>
30. I.A. Gruzberg, A.W.W. Ludwig, N. Read, Exact exponents for the spin quantum Hall transition. Phys. Rev. Lett. **82**, 4524–4527 (1999)
31. T. Guhr, in *Supersymmetry in Random Matrix Theory*. The Oxford Handbook of Random Matrix Theory (Oxford University Press, Oxford, 2010). <http://arXiv:1005.0979v1>
32. J.M. Kosterlitz, D.J. Thouless, Ordering, metastability and phase transitions in two-dimensional systems. J. Phys. C **6**, 1181 (1973)
33. P. Littelmann, H.-J. Sommers, M.R. Zirnbauer, Superbosonization of invariant random matrix ensembles. Comm. Math. Phys. **283**, 343–395 (2008)
34. F. Merkl, S. Rolles, in *Linearly Edge-Reinforced Random Walks*. Dynamics and Stochastics. Lecture Notes-Monograph Series, vol. 48 (2006), pp. 66–77
35. F. Merkl, S. Rolles, Edge-reinforced random walk on one-dimensional periodic graphs. Probab. Theor. Relat. Field **145**, 323 (2009)
36. A.D. Mirlin, in *Statistics of Energy Levels*, ed. by G.Casati, I. Guarneri, U. Smilansky. New Directions in Quantum Chaos. Proceedings of the International School of Physics “Enrico Fermi”, Course CXLIII (IOS Press, Amsterdam, 2000), pp. 223–298. <http://arXiv.org/abs/cond-mat/0006421>
37. L. Pastur, M. Shcherbina, Universality of the local eigenvalue statistics for a class of unitary invariant matrix ensembles. J. Stat. Phys. **86**, 109–147 (1997)
38. R. Pemantle, Phase transition in reinforced random walk and RWRE on trees. Ann. Probab. **16**, 1229–1241 (1988)
39. Z. Rudnick, P. Sarnak, Zeros of principal L-functions and random-matrix theory. Duke Math. J. **81**, 269–322 (1996)
40. M. Salmhofer, *Renormalization: An Introduction* (Springer, Berlin, 1999)
41. L. Schäfer, F. Wegner, Disordered system with n orbitals per site: Lagrange formulation, hyperbolic symmetry, and Goldstone modes. Z. Phys. B **38**, 113–126 (1980)

42. J. Schenker, Eigenvector localization for random band matrices with power law band width. *Comm. Math. Phys.* **290**, 1065–1097 (2009)
43. A. Sodin, The spectral edge of some random band matrices. *Ann. Math.* **172**, 2223 (2010)
44. A. Sodin, An estimate on the average spectral measure for random band matrices, *J. Stat. Phys.* **144**, 46–59 (2011). <http://arxiv.org/1101.4413>
45. T. Spencer, in *Mathematical Aspects of Anderson Localization*, ed. by E. Abrahams. 50 Years of Anderson Localization (World Scientific, Singapore, 2010)
46. T. Spencer, in *Random Band and Sparse Matrices*. The Oxford Handbook of Random Matrix Theory, eds. by G. Akemann, J. Baik, and P. Di Francesco. (Oxford University Press, Oxford, 2010)
47. T. Spencer, M.R. Zirnbauer, Spontaneous symmetry breaking of a hyperbolic sigma model in three dimensions. *Comm. Math. Phys.* **252**, 167–187 (2004). <http://arXiv.org/abs/math-ph/0410032>
48. B. Valko, B. Virag, Random Schrödinger operators on long boxes, noise explosion and the GOE, <http://arXiv.org/abs/math-ph/09120097>
49. J.J.M. Verbaarschot, H.A. Weidenmüller, M.R. Zirnbauer, *Phys. Rep.* **129**, 367–438 (1985)
50. F. Wegner, The Mobility edge problem: Continuous symmetry and a Conjecture. *Z. Phys. B* **35**, 207–210 (1979)
51. M.R. Zirnbauer, The Supersymmetry method of random matrix theory, [http://arXiv.org/abs/math-ph/0404057\(2004\)](http://arXiv.org/abs/math-ph/0404057(2004))
52. M.R. Zirnbauer, Fourier analysis on a hyperbolic supermanifold with constant curvature. *Comm. Math. Phys.* **141**, 503–522 (1991)
53. M.R. Zirnbauer, Riemannian symmetric superspaces and their origin in random-matrix theory. *J. Math. Phys.* **37**, 4986–5018 (1996)

Quantum Many Body Systems

Cetraro, Italy 2010, Editors: Alessandro Giuliani, Vieri

Mastropietro, Jakob Yngvason

Rivasseau, V.; Seiringer, R.; Solovej, J.P.; Spencer, Th. -

Giuliani, A.; Mastropietro, V.; Yngvason, J. (Eds.)

2012, XIII, 180 p. 11 illus., 1 illus. in color., Softcover

ISBN: 978-3-642-29510-2