Large Deviations of Extreme Eigenvalues of Random Matrices

David S. Dean[1] and Satya N. Majumdar[2]
[1] *Laboratoire de Physique Théorique (UMR 5152 du CNRS), Université Paul Sabatier, 118, route de Narbonne, 31062 Toulouse Cedex 4, France*
[2] *Laboratoire de Physique Théorique et Modèles Statistiques (UMR 8626 du CNRS), Université Paris-Sud, Bât. 100, 91405 Orsay Cedex, France*

We calculate analytically the probability of large deviations from its mean of the largest (smallest) eigenvalue of random matrices belonging to the Gaussian orthogonal, unitary and symplectic ensembles. In particular, we show that the probability that all the eigenvalues of an $(N \times N)$ random matrix are positive (negative) decreases for large N as $\sim \exp[-\beta\theta(0)N^2]$ where the parameter β characterizes the ensemble and the exponent $\theta(0) = (\ln 3)/4 = 0.274653\ldots$ is universal. We also calculate exactly the average density of states in matrices whose eigenvalues are restricted to be larger than a fixed number ζ, thus generalizing the celebrated Wigner semi-circle law. The density of states generically exhibits an inverse square-root singularity at ζ.

PACS numbers: 02.50.-r, 02.50.Sk, 02.10.Yn, 24.60.-k, 21.10.Ft

Studies of the statistics of the eigenvalues of random matrices have a long history going back to the seminal work of Wigner [1]. Since then, random matrices have found applications in multiple fields including nuclear physics, quantum chaos, disordered systems, string theory and number theory [2]. Three classes of matrices with Gaussian entries have played important roles [2]: $(N \times N)$ real symmetric (Gaussian Orthogonal Ensemble (GOE)), $(N \times N)$ complex Hermitian (Gaussian Unitary Ensemble (GUE)) and $(2N \times 2N)$ self-dual Hermitian matrices (Gaussian Symplectic Ensemble (GSE)). A central result in the theory of random matrices is the celebrated Wigner semi-circle law. It states that for large N and on an average, the N eigenvalues (suitably scaled) lie within a finite interval $\left[-\sqrt{2N}, \sqrt{2N}\right]$, often referred to as the Wigner 'sea'. Within this sea, the average density of states has a semi-circular form (see Fig. 1) that vanishes at the two edges $-\sqrt{2N}$ and $\sqrt{2N}$

$$\rho_{\rm sc}(\lambda, N) = \sqrt{\frac{2}{N\pi^2}}\left[1 - \frac{\lambda^2}{2N}\right]^{1/2}. \tag{1}$$

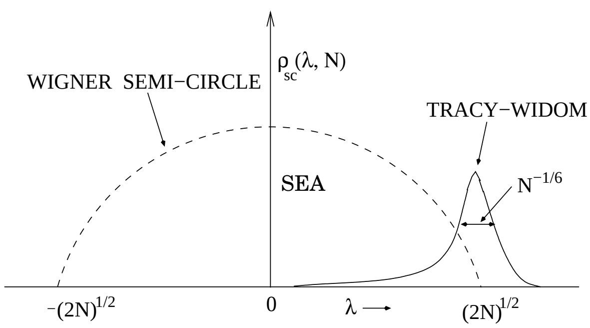

FIG. 1: The dashed line shows the semi-circular form of the average density of states. The largest eigenvalue is centered around its mean $\sqrt{2N}$ and fluctuates over a scale of width $N^{-1/6}$. The probability of fluctuations on this scale is described by the Tracy-Widom distribution (shown schematically).

Thus, the average of the maximum (minimum) eigenvalue is $\sqrt{2N}$ (-$\sqrt{2N}$). However, for finite but large N, the maximum eigenvalue fluctuates, around its mean $\sqrt{2N}$, from one sample to another. Relatively recently Tracy and Widom [3] proved that these fluctuations *typically* occur over a narrow scale of $\sim O(N^{-1/6})$ around the upper edge $\sqrt{2N}$ of the Wigner sea (see Fig. 1). More precisely, they showed [3] that asymptotically for large N, the scaling variable $\xi = \sqrt{2}\, N^{1/6}\,[\lambda_{\max} - \sqrt{2N}]$ has a limiting N-independent probability distribution, ${\rm Prob}[\xi \le x] = F_\beta(x)$ whose form depends on the value of the parameter $\beta = 1$, 2 and 4 characterizing respectively the GOE, GUE and GSE. The function $F_\beta(x)$, computed as a solution of a nonlinear differential equation [3], approaches to 1 as $x \to \infty$ and decays rapidly to zero as $x \to -\infty$. For example, for $\beta = 2$, $F_2(x)$ has the following tails [3],

$$\begin{aligned} F_2(x) &\to 1 - O\left(\exp[-4x^{3/2}/3]\right) \quad \text{as } x \to \infty \\ &\to \exp[-|x|^3/12] \quad \text{as } x \to -\infty. \end{aligned} \tag{2}$$

The probability density function dF_β/dx thus has highly asymmetric tails. The distribution of the minimum eigenvalue simply follows from the fact that ${\rm Prob}[\lambda_{\min} \ge \zeta] = {\rm Prob}[\lambda_{\max} \le -\zeta]$. Amazingly, the Tracy-Widom distribution has since emerged in a number of seemingly unrelated problems such as the longest increasing subsequence problem [4], directed polymers in $(1+1)$-dimensions [5], various $(1+1)$-dimensional growth models [6], a class of sequence alignment problems [7] and in finance [8]. Recently, it has been shown that the statistics of the largest eigenvalue is also of importance in population growth of organisms in fluctuating environments [9].

The Tracy-Widom distribution describes the probability of *typical and small* fluctuations of $\lambda_{\max}$ over a very narrow region of width $\sim O(N^{-1/6})$ around the mean $\langle\lambda_{\max}\rangle \approx \sqrt{2N}$. A natural question is how to describe the probability of *atypical and large* fluctuations of λ_{max} around its mean, say over a wider region of width $\sim O(N^{1/2})$? For example, what is the probability that all the eigenvalues of a random matrix are negative (or equivalently all are positive)? This is the same as the probability that $\lambda_{\max} \leq 0$ (or equivalently $\lambda_{\min} \geq 0$). Since $\langle \lambda_{\max} \rangle \approx \sqrt{2N}$, this requires the computation of the probability of an extremely rare event characterizing a large deviation of $\sim -O(N^{1/2})$ to the left of the mean. This question recently came up in the context of random landscape models of antropic principle based string theory [10, 11] as well as in quantum cosmology [12]. Here one is interested in the statistical properties of vacua associated with a random multifield potential, e.g., how many minima are there in a random string landscape? Similar questions also arise in disordered systems where one is interested in counting the number of local minima of a random Gaussian field [13]. In order to have a local minimum of the random landscape one needs to ensure that the eigenvalues of the associated random Hessian matrix are all positive. A related important question is: if one conditions all the eigenvalues to be positive, how does the average density of states get modified from the Wigner semi-circle form? In this Letter, we address these issues analytically.

It is useful to summarize our main results. In Ref. [11], it was shown numerically that the probability that all the eigenvalues of a $(N \times N)$ GOE matrix ($\beta = 1$) are positive (or equivalently all the eigenvalues are negative, i.e., $\lambda_{\max} \leq 0$) decreases rapidly with large N as $\mathrm{Prob}[\lambda_{\max} \leq 0] \sim \exp\left[-\theta(0)N^2\right]$. A crude approximate argument was provided for the exponent $\theta(0) \approx 1/4$ [11], along with numerical simulations. Here we show exactly that for all ensembles characterized by the parameter β,

$$\theta(0) = \beta \frac{\ln 3}{4} = (0.274653\ldots)\,\beta. \tag{3}$$

More generally we calculate the exact large deviation function associated with large fluctuations of $\sim -O(N^{1/2})$ of $\lambda_{\max}$ to the left of its mean value $\sqrt{2N}$. We show that for large N and for all ensembles

$$\mathrm{Prob}\left[\lambda_{\max} \leq t, N\right] \sim \exp\left[-\beta N^2 \Phi\left(\frac{\sqrt{2N}-t}{\sqrt{N}}\right)\right] \tag{4}$$

where $t \sim O(N^{1/2}) \leq \sqrt{2N}$ is located deep inside the Wigner sea. The large deviation function $\Phi(y)$ is zero for $y \leq 0$, but is nontrivial for $y > 0$ which we compute exactly. For *small* deviations to the left of the mean, taking the $y \to 0$ limit of $\Phi(y)$, we recover the left tail of the Tracy-Widom distribution as in Eq. (2). Thus our result for *large* deviations of $\sim -O(N^{1/2})$ to the left of the mean is complementary to the Tracy-Widom result for *small* fluctuations of $\sim -O(N^{-1/6})$ and the two solutions match smoothly. In the process, we also calculate exactly the modified average density of states when all the eigenvalues are constrained to be on the right of a barrier say at $\lambda = \zeta$, thus generalizing Wigner's semi-circle law.

Our starting point is the celebrated result due to Wigner for the joint probability density function (pdf) of the eigenvalues of a random $(N \times N)$ matrix [2]

$$P(\{\lambda_i\}) = B_N \exp\left[-\frac{\beta}{2}\left(\sum_{i=1}^{N} \lambda_i^2 - \sum_{i\neq j} \ln(|\lambda_i - \lambda_j|)\right)\right], \tag{5}$$

where B_N normalizes the pdf and $\beta = 1$, 2 and 4 correspond respectively to the GOE, GUE and GSE. The joint law allows one to interpret the eigenvalues as the positions of charged particles, repelling each other via a 2-d Coulomb potential (logarithmic); they are confined on a 1-d line and each is subject to an external harmonic potential. The parameter β that characterizes the type of ensemble can be interpreted as the inverse temperature. The average density of states $\rho_{\mathrm{sc}}(\lambda, N) = \sum_{i=1}^{N} \langle \delta(\lambda - \lambda_i) \rangle / N$ can be calculated [2] from the joint pdf in Eq. (5) and has the Wigner semi-circular form of Eq. (1). In the Coulomb gas language, this is the average equilibrium charge density.

Here we are interested in the probability $Q_N(\zeta)$ that all the eigenvalues are bigger than say ζ, i.e., the probability that all charges lie to the right of the barrier at ζ. Note that, due to the $\lambda \to -\lambda$ symmetry of the pdf in Eq. (5), this is also the probability that all eigenvalues are less than $-\zeta$, i.e., the probability that $\lambda_{\max} \leq -\zeta$. Let us first define the restricted partition function

$$Z_N(\zeta) = \int_{\lambda_i > \zeta}^{\infty} \prod_{i=1}^{N} d\lambda_i \; \exp\left[-\frac{\beta}{2}\left(\sum_{i=1}^{N} \lambda_i^2 - \sum_{i\neq j} \ln\left(|\lambda_i - \lambda_j|\right)\right)\right] \tag{6}$$

It then follows that

$$Q_N(\zeta) = \frac{Z_N(\zeta)}{Z_N(-\infty)}. \tag{7}$$

Let $\rho_N(\lambda) = \sum_{i=1}^{N} \delta(\lambda - \lambda_i)/N$ denote the spatial density of charges. Using standard techniques of functional integration we may express $Z_N(\zeta)$ as [14]

$$Z_N(\zeta) \propto \int \mathcal{D}[\rho_N] \exp\left[-\frac{\beta N}{2}\int_{\zeta}^{\infty} d\lambda\, \rho_N(\lambda)\lambda^2 + \frac{\beta N^2}{2}\int_{\zeta}^{\infty} d\lambda\, d\lambda'\, \rho_N(\lambda)\rho_N(\lambda') \ln\left(|\lambda - \lambda'|\right) - N\int_{\zeta}^{\infty} d\lambda\, \rho_N(\lambda) \ln\left(\rho_N(\lambda)\right)\right]. \tag{8}$$

where the first two terms represent the energy of the charges as in Eq. (6). The third term represents the entropy which has a mean field form due to the fact that all charges interact with each other via the long-range logarithmic potential. The charge density $\rho_N(\lambda)$ evidently satisfies the constraints: $\rho_N(\lambda) = 0$ for $\lambda < \zeta$ and $\int_{\zeta}^{\infty} d\lambda \rho_N(\lambda) = 1$.

Since we are interested in fluctuations of $\sim O(N^{1/2})$, it is convenient to work with the rescaled variables, $\lambda = \mu\sqrt{N}$ and

$\zeta = z\sqrt{N}$. It is reasonable to assume that the charge density scales as, $\rho_N(\lambda) = N^{-1/2} f\left(\lambda N^{-1/2}\right)$. The scaling function evidently satisfies the constraints:

$$\int_z^\infty d\mu f(\mu) = 1; \quad f(\mu) = 0 \text{ for } \mu < z. \tag{9}$$

Expressing the action in Eq. (8) in terms of rescaled charged density $f(\mu)$, one finds that the energy term scales as $\sim O(N^2)$ whereas the entropy term $\sim O(N)$ is subdominant for large N. For large N, the functional integration can be carried out using the method of steepest descent. This gives, as a function of rescaled variable $z = \zeta/\sqrt{N}$,

$$Z_N(z) \propto \exp\left[\beta N^2 S(z) + O(N)\right] \tag{10}$$

where $S(z) = \max_f \{\Sigma(f)\}$ and

$$\Sigma(f) = -\frac{1}{2}\int_z^\infty d\mu\, f(\mu)\mu^2 + \frac{1}{2}\int_z^\infty\int_z^\infty d\mu d\mu' f(\mu) f(\mu') \ln\left(|\mu - \mu'|\right). \tag{11}$$

The stationarity condition $\delta\Sigma(f)/\delta f = 0$ gives

$$\frac{\mu^2}{2} + C = \int_z^\infty d\mu'\, f(\mu') \ln\left(|\mu - \mu'|\right), \tag{12}$$

where C is a Lagrange multiplier enforcing the normalization of f in Eq. (9). Differentiating Eq. (12) with respect to μ gives

$$\mu = \mathcal{P}\int_z^\infty d\mu'\, f(\mu') \frac{1}{\mu - \mu'}, \tag{13}$$

where $\mathcal{P}$ indicates the Cauchy principle part. It is convenient to introduce a shift $\mu = z + x$ where $x \geq 0$ represents the distance from the barrier (to the right) at z. In terms of the variable x, Eq. (13) becomes an integral equation for the charge density

$$x + z = \mathcal{P}\int_0^\infty dx'\, f(x') \frac{1}{x - x'} \tag{14}$$

where the rhs represents a semi-infinite Hilbert transform. The real technical challenge is to invert this integral equation and obtain a closed form expression for the rescaled charge density $f(x)$. Fortunately this can be done [14]. We find that $f(x)$ is nonzero inside a finite box $x \in [0, L(z)]$ and vanishes outside this box. For $0 \leq x \leq L(z)$, the density is given exactly by

$$f(x) = \frac{1}{2\pi\sqrt{x}}\sqrt{L(z) - x}\,\left[L(z) + 2x + 2z\right]. \tag{15}$$

The length of the box $L(z)$ can be determined from the normalization condition in Eq. (9) and is given by

$$L(z) = \frac{2}{3}\left[\sqrt{z^2 + 6} - z\right]. \tag{16}$$

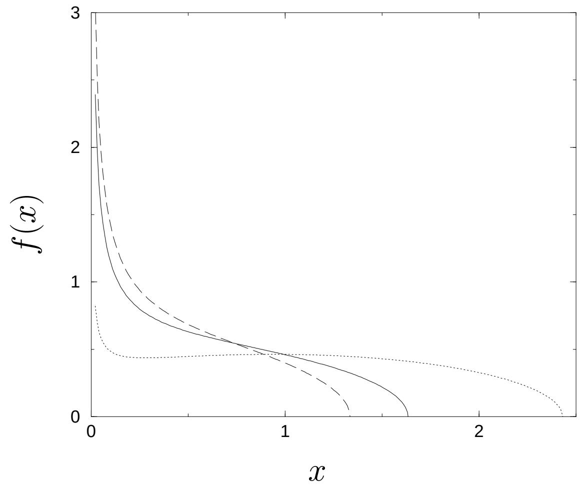

FIG. 2: The average density of states $f(x)$ plotted as a function of the shifted variable x for $z = -1$ (dotted line), $z = 0$ (solid line), and $z = 0.5$ (dashed line).

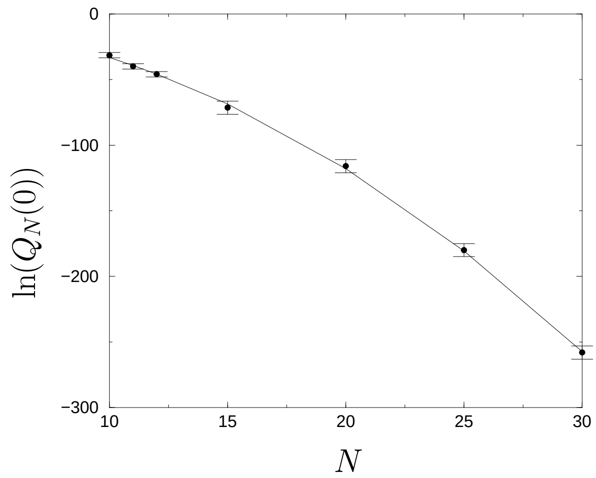

FIG. 3: Monte Carlo computation of $\ln(Q_N(0))$ points with error bars along with a quadratic fit (solid line).

Note that the charge density $f(x)$ depends on z, i.e. the location of the barrier. A plot of this density for several values of z are shown in Fig. 2.

A couple of remarks are in order: (i) the charge density $f(x)$ must be positive for all x including $x = 0$. As $x \to 0$, $f(x)$ diverges as $x^{-1/2}$. However in order that it remains positive, we need to ensure that the amplitude $L(z)+2z \geq 0$ at $x = 0$ in Eq. (15). This condition, using $L(z)$ from Eq. (16), requires $z \geq -\sqrt{2}$. Thus the results in Eqs. (15) and (16) are valid only for $z \geq -\sqrt{2}$. Indeed, this is expected because exactly at $z = -\sqrt{2}$, i.e., when the barrier is placed at the left edge of the Wigner sea, we recover from Eq. (15) the Wigner semi-circle law. For $z = -\sqrt{2}$, Eq. (16) gives $L = 2\sqrt{2}$ (the support of the semi-circle) and Eq. (15) gives $f(\mu) = \sqrt{2 - \mu^2}/\pi$ for $-\sqrt{2} \leq \mu \leq \sqrt{2}$. Thus, for any $z < -\sqrt{2}$, our exact solution indicates that the charge density remains

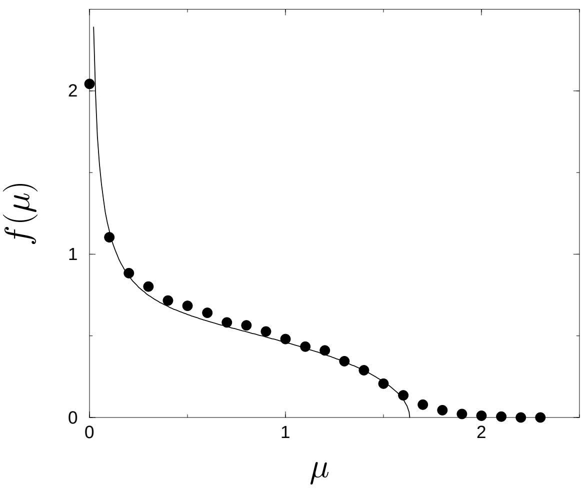

FIG. 4: The analytic large N formula for f with $z = 0$ (solid line) in Eq. (15) is compared to the numerically generated averaged histogram of (6×6) Gaussian matrices with positive eigenvalues. Despite the small size $N = 6$, the agreement is already fairly good, except near the large μ tail.

unchanged from the Wigner semi-circular form. Physically this means that if the wall is placed to the left of the lower edge of the Wigner sea, it has no effect on the charge distribution. (ii) The second remark is that the charge density $f(x)$ changes its shape in an interesting fashion as one changes the barrier location z (see Fig. 2). It turns out that for any $z > -\sqrt{2}$, the charges always accumulate near the barrier at $x = 0$ leading to a square-root divergence of $f(x) \sim x^{-1/2}$ as $x \to 0$. In particular, for $z = 0$, this accumulation of eigenvalues near $x = 0$ can be interpreted as the accumulation of massless modes in the context of a (stable) field theory, a fact that may be of relevance in anthropic principle based string theory.

Knowing $f(x)$ exactly, the Lagrange multiplier is determined by setting $\mu = z$ in Eq. (12). This gives, following a shift in the integral, $C = -z^2/2 + \int_0^\infty dx'\ f(x') \ln(x')$. The saddle point action can now be evaluated explicitly [14]

$$S(z) = -\frac{1}{216}\left[72z^2 - 2z^4 + (30z + 2z^3)\sqrt{6+z^2}\right.$$
$$\left. +\ 27\left(3 + \ln(1296) - 4\ln\left(-z + \sqrt{6+z^2}\right)\right)\right]. \quad (17)$$

The probability that all eigenvalues are to the right of $\zeta = z\sqrt{N}$ is then given by, to leading order in large N, using Eqs. (10) and (7)

$$Q_N\left(\zeta = z\sqrt{N}\right) \approx \exp\left[-\beta N^2 \theta(z)\right] \quad (18)$$

where $\theta(z) = S(-\sqrt{2}) - S(z)$ and $S(z)$ is given by Eq. (17). Note that we have used $S(-\infty) = S(-\sqrt{2})$ following remark (i) above. The result in Eq. (4) can then be derived by setting $t = -\zeta = -z\sqrt{N}$ and one finds the large deviation function for $y \geq 0$, $\Phi(y) = S(-\sqrt{2}) - S(-\sqrt{2} + y)$. For small y, $\Phi(y) \approx y^3/6\sqrt{2}$ and for large y, $\Phi(y) \approx y^2/2$. Thus for $\sqrt{2N} - t << \sqrt{N}$, using $\Phi(y) \approx y^3/6\sqrt{2}$ we get,

$$\mathrm{Prob}[\lambda_{\max} \leq t, N] \approx \exp\left[-\frac{\beta}{24}\left|\sqrt{2}\, N^{1/6}\,(t - \sqrt{2N})\right|^3\right] \quad (19)$$

which matches exactly with the left tail of the Tracy-Widom distribution for all β. For example, for $\beta = 2$ one can easily verify this by comparing Eqs. (19) and (2).

The probability that all eigenvalues are positive is obtained by setting $z = 0$ in Eq. (18) resulting in a remarkably simple and exact formula stated in Eq. (3). The fact that this probability decreases as rapidly as $\sim \exp[-\beta\theta(0)N^2]$ for large N and that there are significant $\sim O(N)$ corrections indicate that numerically it is extremely difficult to measure the exponent $\theta(0)$ accurately. An attempt was made in Ref. [11] using GOE ($\beta = 1$) matrices up to sizes of $N = 7$ to fit the probability with the form $\exp[-aN^\alpha]$ that yielded $\alpha \approx 2.00387$ and $a \approx 0.3291$. Clearly, the system sizes are too small to take this fit seriously. It turns out that instead it is easier to evaluate $Q_N(0)$ directly from Eq. (7) via a clever Monte Carlo method that allows us to go up to $N \sim 30$ [14]. In Fig. 3 we show a plot of $\ln(Q_N(0))$ measured using this Monte Carlo method (for $\beta = 1$) with a fit of the form $aN^2 + bN + c$. This fit yields $a \approx -0.2755$ which is in good agreement with the exact value of $\theta(0) = 0.274653..$ predicted here.

Another numerical check consists in computing the charge density $f(\mu)$ by direct sampling of Gaussian matrices and comparing it to the theoretical prediction in Eq. (15). Here, we are clearly restricted to small values of N. In Fig. 4, we compare the numerically computed $f(\mu)$ for $z = 0$ obtained from matrices of size (6×6) with the theoretical prediction. Despite the small value of N, the agreement is already fairly good.

We thank O. Bohigas for useful comments.

[1] E.P. Wigner, Proc. Cambridge Philos. Soc. **47**, 790 (1951).

[2] M.L. Mehta, Random Matrices, 2nd Edition, (Academic Press) (1991).

[3] C.A. Tracy and H. Widom, Commun. Math. Phys. **159**, 151 (1994); *ibid* **177**, 727 (1996).

[4] J. Baik, P. Deift, and K. Johansson, J. Am. Math. Soc. **12**, 1119 (1999).

[5] J. Baik and E.M. Rains, J. Stat. Phys. **100**, 523 (2000); K. Johansson, Commun. Mat. Phys. **209**, 437 (2000).

[6] M. Prähofer and H. Spohn, Phys. Rev. Lett. **84**, 4882 (2000); J. Gravner, C.A. Tracy, and H. Widom, J. Stat. Phys. **102**, 1085 (2001); S.N. Majumdar and S. Nechaev, Phys. Rev. E **69**, 011103 (2004); T. Imamura and T. Sasamoto, Nucl. Phys. B **699**, 503 (2004).

[7] S.N. Majumdar and S. Nechaev, Phys. Rev. E **72**, 020901(R) (2005).

[8] G. Biroli, J-P. Bouchaud, and M. Potters, cond-mat/0609070.

[9] E. Kussell and S. Leibler, Science, **309**, 2075 (2005).

[10] L. Susskind, arXiv:hep-th/0302219; M.R. Douglas, B. Shiffman, and S. Zelditch, Commu. Math. Phys. **252**, 325 (2004).

[11] A. Aazami and R. Easther, J. Cosmol. Astropart. Phys. JCAP03 013 (2006).
[12] L. Mersini-Houghton, Class. Quant. Grav. **22**, 3481 (2005).
[13] Y.V. Fyodorov Phys. Rev. Lett. **92**, 240601 (2004) ; *ibid* Acta Physica Polonica B, **36**, 2699 (2005).
[14] Details will be published elsewhere.