Matrix Product States, Random Matrix Theory and the Principle of Maximum Entropy

Benôit Collins¹, ², Carlos E. González-Guillén³, ⁵ and David Pérez-García⁴, ⁵

¹Département de Mathématique et Statistique, Université d’Ottawa, K1N6N5 Ottawa, Canada.
²CNRS, Institut Camille Jordan Université Lyon 1, 69622 Villeurbanne, France.
³Departamento de Matemáticas, E.T.S.I. Industriales, Universidad Politécnica de Madrid, 28006 Madrid, Spain.
⁴Departamento Análisis Matemático, Universidad Complutense de Madrid, 28040 Madrid, Spain.
⁵IMI, Universidad Complutense de Madrid, 28040 Madrid, Spain.

Abstract Using random matrix techniques and the theory of Matrix Product States we show that reduced density matrices of quantum spin chains have generically maximum entropy.

Jaynes’ principle of maximum entropy [Jay57a, Jay57b] gives a pretty satisfactory solution to the old problem of dealing with prior information in probability theory. Generalizing the old principle of indifference of Laplace, it briefly states that among all possible probability distributions compatible with our prior information, the best choice is the one which maximizes the Shannon entropy. Apart from its important applications on decision theory, since its origin it has succeeded in giving a very useful information-theoretical view of statistical mechanics, both classical [Jay57a] and quantum [Jay57b] –where the function to maximize is the von-Neumann entropy. As an easy illustration, given the average energy of a quantum system as prior information, the density matrix which maximizes entropy is exactly the thermal state associated to that particular energy. This spirit has been recently recovered with great success in [PSW06] and further developed in a number of ways in [BCH+11, Cra11, LPSW09, LPSW10].

To which extent the principle of maximum entropy can be extended to more and more general situations has been a very active and controversary field in the last half-century. For instance, very recently a series of theoretical and experimental works [CFM⁺08, FCM⁺08, TCF⁺11] seem to validate the principle in relaxation processes of quantum systems when focusing on a particular small subsystem –which, as argued in Figure 1, is the most relevant situation.

Another, even older, principle to assign prior probabilities in physical problems relies on the symmetries of the problem (see [Jay68] for a discussion). For instance, if one wants to incorporate in the problem some invariance, i.e. independence of the reference frame, this already reduces the class of prior probability distributions available. Indeed, if one has enough symmetries –they form a compact group– there is indeed a unique way of defining a prior distribution compatible with the symmetries –the Haar measure– and the problem is solved.

But what if one wants to incorporate to the problem some less standard knowledge? For instance, that the interactions in our model are local and homogeneous and that we work at zero temperature, but not any assumption on the particular interactions in the model itself. Note the difference with Jaynes’ approach where the particular Hamiltonian of the model is known. Is there any way of incorporating this information to the problem? Which is then the right prior probability? Is it related to maximizing some entropy? Since this type of assumptions are natural and widely accepted, solving these questions could be of upmost importance in quantum condensed-matter problems. In this paper we attack (and to some extent solve) them in the particular case of 1D spin systems.

To do that we will take advantage of the recent developments in the understanding of quantum spin chains, where it is nowadays widely well justified, both numerically [Whi92] and analytically [Has07], that their ground states are exactly represented by the set of Matrix Product States (MPS) with polynomial bond dimension. We will concentrate in the situation of a chain with boundary effects in exponentially small regions of size $b$ at both ends, homogeneity in the bulk and experimental access to an exponentially small central region of size $l$ (see Figure [1]). Tracing out the boundary terms leads to a bulk state given by

\[ \rho = \sum_{i_{b+1} \cdots i_{N-b}, j_{b+1} \cdots j_{N-b}} \text{tr}(L A_{i_{b+1}} \cdots A_{i_{N-b}} R A_{j_{N-b}} \cdots A_{j_{b+1}})|i_{b+1} \cdots i_{N-b}\rangle\langle j_{b+1} \cdots j_{N-b}|, \]
Figure 1. In this work we consider a chain of \(N\) sites, with homogeneous interactions in the bulk and boundary effects in exponentially small regions of size \(b\) at the borders. We assume that the experimentally accessible region (and hence the region we are interested in) is an exponentially small region of size \(l\) in the center of the chain.

where all \(A_i, L \geq 0\) and \(R \geq 0\) are \(D \times D\) matrices with \(D = \text{poly}(N)\). This will be our starting point, that is, the prior information can be understood as restricting the bulk-states of our system as having the form \([1]\).

Now, it is also known from the general theory of MPS \cite{PGVWC06} that this set has a natural (over)parametrization by the group \(U(dD)\), via the map \(U \mapsto A_i = \langle 0|U|i\rangle\). Being \(U(dD)\) a unitary group, one can use the symmetry-based assignment of prior distributions to sample from the Haar measure. Similarly, the fact that the map \(X \mapsto \sum_j A_j X A_j^\dagger\) is trace-preserving leads to consider \(\text{tr}(R) = 1, \|L\|_\infty \leq 1\), giving us natural ways of sampling also the boundary conditions (see below). One can therefore ask about which is then the generic reduced density matrix \(\rho_l\) of \(l \ll N\) sites. Note that, by the above comments, this is nothing but asking about generic observations of 1D quantum systems. This idea has been already exploited for the non-translational invariant case in \cite{GdOZ10}. The aim of the present work is to show that \(\rho_l\) has generically maximum entropy:

**Theorem 0.1.** Let \(\rho_l\) be taken at random from the ensemble introduced with \(D \geq N^5\). Then \(\|\rho_l/\text{tr} \rho_l - 1/d^d\|_\infty \leq (d^d - 1)\sqrt{dO(D^{-1/10})}\) except with probability exponentially small in \(D\).

Note that, since the accessible region \(l\) is exponentially smaller than the system size, the bound can be made arbitrary small while keeping the size of the matrices \(D\) polynomial in the system size.

To prove the theorem, we will rely on recent developments of random matrix theory, in particular on the graphical Weingarten calculus provided in \cite{CN10}, and on a novel estimate of the Weingarten function.

The paper is organized as follows. First we introduce the Matrix Product State formalism, then we introduce the map acting only in \(A\) and creating \(l\) sites, where one should understand the map acting only in \(A\) and creating \(l\) sites, and then we introduce basic results of the concentration of measure phenomenon. Finally, in section \([4]\) we prove **Theorem 0.1** using the tools already introduced together with a novel asymptotic bound of the Weingarten function.

1. **Random Matrix Product States**

In this section we just fix the notation, for a detailed exposition see \cite{PGVWC06}. Let \(\dim(H_A) = D\) and \(\dim(H_B) = d\), our initial state \(\rho\) given by Equation \([1]\) can be expressed by means of the map \(\mathcal{E}(X) : B(H_A) \to B(H_A \otimes H_B)\) given by \(\mathcal{E}(X) = \sum_{i,j} A_i X A_j^\dagger \otimes |i\rangle\langle j|\), simply as

\[
\rho = \text{tr}_A[L\mathcal{E}^n(R)],
\]

where one should understand the map acting only in \(A\) and creating the systems \(B\) in order from 1 to \(n\).

In the rest of this work we will be interested in the reduced state of the \(l\) consecutive central sites of the chain, where \(l << n = 2t + l\), that, up to normalization, will be described as

\[
\rho_l = \text{tr}_{A,B_1 \ldots B_l,B_{l+1} \ldots B_n}[L\mathcal{E}^n(R)].
\]

The general boundary conditions \(L\) and \(R\) come from tracing out the boundary sites as described in figure \([4]\). MPS theory leads to consider them belonging respectively to the sets \(\mathcal{L} = \{L \geq 0 : \|L\|_\infty \leq 1, L \in M_D\}\) and \(\mathcal{R} = \{R \geq 0 : \text{tr}(R) = 1, R \in M_D\}\), where \(\|\cdot\|_\infty\) means the usual operator norm. Diagonalizing
by \( \lambda \) in \([CN10]\). We will follow the standard notation of representation theory of symmetric groups. We denote refer to \([Col03]\). Here we just describe its main ingredients to focus on the graphical calculus introduced indicated and the ancillary system \( A \) from right to left living the other sites invariant.

where the systems of the sites \( B \) in the symmetric group \( \sigma \) corresponding Schur function, see \([Ful97]\). If \( \sigma \) can be written as a product of transpositions, \#\( \sigma \) is the number of cycles in \( \sigma \) and both quantities are related by the formula \(|\sigma| = p - \#\sigma\).

**Definition** The Weingarten function \( Wg(n, \sigma) \) takes as inputs a dimension parameter \( n \) and a permutation \( \sigma \) in the symmetric group \( S_p \) and is given by

\[
Wg(n, \sigma) = \frac{1}{p!^2} \sum_{\lambda \vdash p} \frac{(\chi^\lambda(1))^2 \chi^\lambda(\sigma)}{s_{\lambda,n}(1)}.
\]

Its importance relies on the following theorem from \([Col03]\), which tells us that the average of a monomial over the unitary group can be computed in terms of sums of Weingarten functions.
Theorem 2.1. Let \( n \) be a positive integer and \( i = (i_1, \ldots, i_p), j = (j_1, \ldots, j_p) \) be \( p \)-tuples of positive integers from \( 1, 2, \ldots, n \). Then
\[
\int_{U(n)} U_{i_1 j_1} \cdots U_{i_p j_p} U_{i'_1 j'_1} \cdots U_{i'_p j'_p} dU = \sum_{\sigma, \tau \in S_p} \delta_{i_1 \sigma(i_1)} \cdots \delta_{i_p \sigma(i_p)} \delta_{j_1 \tau(j_1)} \cdots \delta_{j_p \tau(j_p)} Wg(n, \tau \sigma^{-1}).
\]

In [CN10] the authors introduce a graphical paradigm in order to simplify the computation of the average of polynomials over the unitary group. Consider a polynomial \( P \) of degree \( p \) in \( U \) and \( \overline{U} \) then
\[
E_U(P(U)) = \sum_{\sigma, \tau \in S_p} C_{(\sigma, \tau)} Wg(n, \tau \sigma^{-1}),
\]
where the coefficients \( C_{(\sigma, \tau)} \) can be computed by the following procedure in figures \( 2 \) and \( 3 \). One has to enumerate the matrices \( U \) and \( \overline{U} \) respectively from 1 to \( p \) and for any two permutations \( \sigma, \tau \in S_p \) we delete the \( U \) and \( \overline{U} \) boxes and we connect the white square and circle in \( U_i \) with the white square and circle respectively in \( U_{\sigma(i)} \), and analogously with the black objects and the \( \tau \) permutation. Now, loops represent traces over the matrices involved in it. If there is no matrix involved in a loop then it represents the trace of the identity of the system. Finally, if there are paths that are not loops they translate into the contraction with the boundary conditions that appear in it. The number \( C_{(\sigma, \tau)} \) is just the product of the values of all the contractions. Note that, as drawn in figure \( 2 \), a monomial in \( U_{\uparrow} \) can be substituted by a monomial in \( \overline{U} \).

3. Measure of Concentration Phenomenon

In this section we introduce the basic results of the measure concentration phenomenon that we are going to use; for a detailed exposition see for example [MSS6, Led01].

Definition Let \((X, d)\) be a metric space with probability measure \( \mu \), its concentration function \( \alpha_{(X,d,\mu)}(r) \) is defined as
\[
\alpha_{(X,d,\mu)}(r) = \sup\{1 - \mu(A_r); A \subset X, \mu(A) \geq \frac{1}{2} \}, \quad r > 0
\]
where \( A_r = \{ x \in X; d(x, A) < r \} \) is the open \( r \)-neighbourhood of \( A \) (with respect to \( d \)).

This definition allows to prove directly that almost all the images of a Lipschitz function concentrate around the median, where the concentration factor is given by the concentration function. Nevertheless, we are interested in the concentration around the mean which is a consequence of the other, as one can bound the distance between the median and the mean depending on the concentration function.

Theorem 3.1 (Measure concentration phenomenon). Let \( F \) be a Lipschitz function on \((X, d)\), and \( \mu \) a probability measure on \((X, d)\), then
\[
\mu(\{ F > E_\mu(F) + r \}) \leq 2\alpha_\mu(r/\|F\|_{\text{Lip}}),
\]
\[
\mu(\{ F < E_\mu(F) - r \}) \leq 2\alpha_\mu(r/\|F\|_{\text{Lip}}),
\]
where \( E_\mu(F) \) is the mean of \( F \) with respect to \( \mu \) and \( \|F\|_{\text{Lip}} \) is the Lipschitz constant of \( F \).

When one is interested in the concentration properties of a family of spaces, what matters is the scaling of the concentration function depending on the parameter defining the family of spaces. Thus looking at the definition of concentration function one can prove that the concentration properties of two spaces behave at least as well as the worst one of the two.

Proposition 3.2. Let \( \mu, \nu \) two probability measures on metric spaces \((X, d)\) and \((Y, \delta)\) respectively. Then, if \( \mu \times \nu \) is the product measure in \( X \times Y \) equipped with the \( l^1 \)-metric, \( \alpha_{(X \times Y, d, \mu \times \nu)} \leq \alpha_{(X, d, \mu)} + \alpha_{(Y, \delta, \nu)} \). If we apply this proposition to the spaces we are interested in we have the following lemma.

Lemma 3.3. Let \( \mu \) be the Haar measure in \((U(D), d_2)\), the unitary group with the Hilbert-Schmidt distance. Let \( \nu \) be the Lebesgue measure in \(([0,1]^D, d_\infty)\), the hypercube with the maximum distance. Then, for any
$k \in \mathbb{N}$, the product space $(X,d_1)$, where $X = U(kD) \times U(D) \times U(D) \times [0,1]^D$ and $d_1$ is the $l_1$ distance of the product space, with the product probability measure $\eta = \mu \times \mu \times \mu \times \nu$ has concentration function

$$\alpha_{(X,\delta,\eta)}(r) \leq ce^{-Cr^2},$$

where $c$ and $C$ are universal constants.

**Remark** Note that in this lemma, that will be used to prove the concentration in theorem [4.7], we do not consider the space $S_1([0,1]^D)$, as in all the theorems below the result holds independently of $\Omega \in S_1([0,1]^D)$.

### 4. Proof of the main Theorem

Recall that we are considering the ensemble of MPS defined by the tuple $(U,L,R) = (U,V,W,\Lambda,\Omega)$ where $U$, $V$ and $W$ are distributed with respect to the Haar measure in the respective unitary group, $\Lambda$ is distributed according to the Lebesgue measure in $[0,1]^D$ and $\Omega$ according to any permutation invariant probability measure in $S_1([0,1]^D)$. To prove Theorem 0.1 we need to compute the mean and the Lipschitz constant of the trace normalized version of $f(\rho) = \text{tr}(\rho^2(U,L,R))$ over the introduced ensemble. The difficulty of this calculus comes from computing the mean of the function. To simplify our computations we will first compute the mean and the Lipschitz constant for both function $f(\rho)$ and its normalization function $g(\rho) = (\text{tr}(\rho(U,L,R))^2$ and then argue about the concentration of $\text{tr}(\rho^2_{\text{norm}}) = f(\rho)/g(\rho)$.

To compute the mean of $f(\rho)$ we first need to give a novel asymptotic bound of the Weingarten function.

**Theorem 4.1.** Let $p$, $n$ and $k$ be nonnegative integers such that $p^k \leq n$. Then there exists a constant $K$ depending only on $k$ such that for any $\sigma \in S_p$,

$$Wg(n,\sigma) \leq Kn^{-p-|\sigma|(1-2/k)}.$$

**Proof.** We recall, see [Ful97], that for any partition $\lambda \vdash p$ of the integer $p$,

$$s_{\lambda,n}(1) = \frac{\chi^{\lambda}(1)}{p!} \prod_{i=1}^{p} (n - \lambda_i)$$

where $\lambda_i$ is an integer in $\{0,\ldots,p-1\}$. Equivalently, the Weingarten function becomes

$$(3) \quad Wg(n,\sigma) = \frac{1}{np^p} \sum_{\lambda\vdash p} \frac{\chi^{\lambda}(1)\chi^{\lambda}(\sigma)}{\prod_{i=1}^{p} (1 - \lambda_i/n)}.$$

Consider the function

$$f_\lambda : z \rightarrow (\prod_{i=1}^{p} (1 - z\lambda_i))^{-1}$$

This function is holomorphic in a neighborhood of zero. Moreover $2 \leq k$, since $p^2 \leq n$, we have, for any $|z| \leq p^{-2}$,

$$|f_\lambda(z)| \leq e.$$

As a consequence, writing $f_\lambda(z) = \sum_{i \geq 0} a_{i,\lambda}z^i$, we obtain the Cauchy estimate

$$a_{i,\lambda} \leq ep^{2i}.$$

But equation (3) implies that

$$Wg(n,p,\sigma) = \frac{1}{np^p} \sum_{\lambda\vdash p} \chi^{\lambda}(1)\chi^{\lambda}(\sigma)(1 + \sum_{i \geq 1} a_{i,\lambda}n^{-i}).$$

Therefore the coefficient in $n^{-p-i}$ has norm smaller than $ep^{2i}$. But all coefficients are zero until $i = |\sigma|/k$ [Cca03], therefore

$$Wg(n,\sigma) \leq n^{-p-|\sigma|}e(1 + \frac{p^2}{n} + (\frac{p^2}{n})^2 + \ldots)p^{2|\sigma|}.$$

For $n \geq 2$, and since $p^k \leq n$, this implies $p^{2|\sigma|} \leq n^{2|\sigma|/k}$. Furthermore $(1 + \frac{p^2}{n} + (\frac{p^2}{n})^2 + \ldots)$ can be bounded by a universal constant (5, for example). The result follows.
To organize the computations and the reasoning used in the bound of the mean of $f(\rho)$ we prove the following two lemmas.

**Lemma 4.2.** Let $\alpha, \beta, \gamma \in S_n$,

a) The quantity $|\gamma^{-1}\alpha\gamma^{-1}\beta| + |\beta|$ is an even number.

b) If $\gamma^{-1}\alpha\gamma^{-1}\beta = c$ and $\gamma^{-1}\alpha\gamma^{-1}\beta = c$, then $\alpha^{-1}\alpha$ commutes with $\gamma$.

c) If $p = 2n+4$ and $\gamma = (2n+1, 1, 2, ..., n, 2n+3)(2n+2, n+1, n+2, ..., 2n, 2n+4)$ and $\alpha(2n+i) = 2n+i$ for $i = 1, ..., 4$. Then the function that takes $(\alpha, \beta)$ to $(g, h)$ with $g = \gamma^{-1}\alpha\gamma^{-1}$ and $h = \beta\alpha^{-1}$ is one to one.

**Proof.**

a) The result follows from the fact that the parity of $|\alpha\beta|$ is the same as the parity of $|\alpha| + |\beta|$.

b) If $\gamma^{-1}\alpha\gamma^{-1}\beta = c$ and $\gamma^{-1}\alpha\gamma^{-1}\beta = c$, then $\alpha\gamma^{-1} = \alpha'\gamma^{-1}$. Multiplying by the inverse of the right hand side we have that $\gamma^{-1}\alpha^{-1}\alpha^{-1}\alpha' = 1$ which happens if and only if $\alpha'\alpha$ commutes with $\gamma$.

c) If $\alpha$ is fixed, the change $h = \beta\alpha^{-1}$ is clearly one to one. Now, by b), the change $g = \gamma^{-1}\alpha\gamma^{-1}$ is one to one if and only if $\gamma^{-1}\alpha\gamma^{-1} = 1$ has only the trivial solution $\alpha = 1$, which can be easily check to be the case by the definition of $\gamma$ and the constraints in $\alpha$.

**Lemma 4.3.** Let $L \in L$ and $R \in R$ be given at random with respect to the measures introduced in Section 4. For any $\Omega \in S_1([0, 1]^D)$, we have that $E[\text{tr}(L)] = D/2$, $E[\text{tr}(L^2)] = D/4$, $E[\text{tr}(R)] = 1$, $E[\text{tr}(R^2)] \leq 1$, $E[\text{tr}(LR)] = 1/2$, $E[\text{tr}(LRR)] \leq 1/2$, $E[\text{tr}(LLR)] = 1/4$, $E[\text{tr}(LLRR)] \leq 1/4$ and $E[\text{tr}(LRLR)] \leq 1/4 + 1/4D$.

**Proof.** These averages are not difficult to compute directly, but they can also be computed using the graphical calculus described in section 2 as a warming up for the forthcoming computations. We compute the last one as example.

$$E[\text{tr}(LRLR)] = E[\text{tr}(LAV^\dagger W\Omega^\dagger V^\dagger AW^\dagger W)] = E[\text{tr}(LU^\dagger\Omega U^\dagger\Omega)]$$

where $\Lambda \in [0, 1]^D$, $\Omega \in S_1([0, 1]^D)$ and $U, V, W \in U(D)$ and the second equation follows by the invariance of the Haar measure. Using that $E_\Lambda[\text{tr}(\Lambda)] = D/2$, $E_\Lambda[\text{tr}(\Lambda^2)] = D/4$, $\tr(\Omega^2) \leq 1$ and $(\tr\Omega)^2 = 1$ together with the graphical calculus in figure 3 we get

$$E_{L,R}[\text{tr}(LRLR)] = E_{\Lambda,\Omega} \left[ \sum_{\alpha,\beta \in S_2} C_{(\alpha,\beta)} W \gamma(D, \alpha\beta^{-1}) \right] =$$

$$= E_{\Lambda,\Omega} \left[ (\tr\Lambda)^2 \tr(\Omega^2) + \tr(\Lambda)^2 (\tr\Omega)^2 \right] W \gamma(D, (1/2)) - (\tr\Lambda)^2 (\tr\Omega)^2 + (\tr\Lambda)^2 (\tr\Omega^2) W \gamma(D, (12)) \right]$$

$$\leq 1/4 + 1/4D.$$

**Theorem 4.4.** Let $\rho_1$ be taken at random from the ensemble introduced. Then, for any $\Omega \in S_1([0, 1]^D)$ and $U \in U(dD)$, we have that

$$E_{\Lambda,\Omega,V,W}(\tr(\Lambda, \Omega, U, V, W)) = 1/2$$
Theorem 4.5. Let $\rho_1$ be taken at random from the ensemble introduced and $D \geq n^5$. For any $\Omega \in S_1([0, 1]^D)$, then

$$\mathbb{E}(\text{tr}(\rho_1^2)) \leq \frac{1}{4d^4} + O(D^{-1/5}).$$

In order to apply the measure concentration phenomenon we only need to compute the Lipschitz constants of both functions are upper bounded by $4n + 10$.

Theorem 4.6 (Lipschitz constants). For any $\Omega \in S_1([0, 1]^D)$, let $f(U, V, W, \Lambda) = (\text{tr}(\rho_1(U, V, W, \Lambda, \Omega))^2$ and $g(U, V, W, \Lambda) = \text{tr}(\rho_1(U, V, W, \Lambda, \Omega)$ where

$$\rho_1(U, V, W, \Lambda, \Omega) = \text{tr}_{A,B_1 \cdots B_{n-1}, B_n}(V \Lambda V^\dagger U_{A,B_1} \cdots U_{A,B_{n}} W \Omega W^\dagger \otimes (|0\rangle\langle 0|)^{\otimes n}) U_{A,B_n}^\dagger \cdots U_{A,B_1}^\dagger.$$ 

Then the Lipschitz constants of both functions are upper bounded by $4n + 10$.

Now we can show which is the behavior of the $2$-Renyi entropy, or equivalently the purity of the normalized state $\rho_{\text{Norm}} = \rho_1/\text{tr}(\rho_1)$.

Theorem 4.7. Let $\rho_1$ be taken at random from the ensemble introduced with $D \geq n^5$. Then $\text{tr}(\rho_{\text{Norm}}^2) = \text{tr}(\rho_1^2)/(\text{tr}(\rho_1)^2)^2 = 1/d^4 + O(D^{-1/5})$ except with probability exponentially small in $D$.

Proof. Putting together the measure concentration phenomenon 3.1 the bounds on the Lipschitz constant 4.6 and the union bound, we have for all $\Omega \in S_1([0, 1]^D)$ that except with probability $c_1 e^{-c_2 D/n^2}$

$$\text{tr}(\rho_1^2) \leq \mathbb{E}(\text{tr}(\rho_1^2)) + \epsilon \quad \text{and} \quad \text{tr}(\rho_1^2) \geq \mathbb{E}(\text{tr}(\rho_1^2)^2) - \epsilon$$

both at the same time where $c_1$ and $c_2$ are universal constants. Thus, we can bound

$$\frac{\text{tr}(\rho_1^2)^2}{\text{tr}(\rho_1^2)^2} \leq \mathbb{E}(\text{tr}(\rho_1^2)^2) + \epsilon \leq \mathbb{E}(\text{tr}(\rho_1^2)) + \epsilon \leq \frac{1}{4d^4} + O(D^{-1/5}) = \frac{1}{d^4} + O(D^{-1/5}),$$

where the second inequality follows from Jensen’s inequality, the third inequality follows from Theorem 4.4 and 4.5 and in the last equality we have used that we can take $\epsilon = O(D^{-1/5})$. The result follows.

Finally we can easily prove our main theorem, which bounds the distance between the reduced density matrix of a generic random MPS and the completely mixed state.

Proof of theorem 0.4

$\rho_{\text{Norm}} = \rho_1/\text{tr}(\rho_1)$ is trace normalized, that is, its eigenvalues sum up to one. Thus, in order to have an eigenvalue of $\rho$ as far as possible from $1/d^4$, the distribution of eigenvalues optimizing this problem is the one that has one eigenvalue as small or big as possible and the rest all equal. In both cases the distance between this eigenvalue and $1/d^4$ is $(d^4 - 1) \sqrt{d} O(D^{-1/10})$. 

5. Conclusions

In this work we have shown how reduced density matrices of small subsystems of translational invariant random MPS have generically maximum entropy. This can be read as recovering Jayne’s principle of maximum entropy in the situation where the prior information to incorporate in the sampling procedure is the locality and homogeneity of the interactions. For that we have relied on the (well justified) fact that MPS are the right representation for ground states of one dimensional local Hamiltonians and in the natural way of sampling MPS based on the symmetry principle.

We acknowledge Ion Nechita for very insightful discussions during the preparation of this manuscript and Mittag-Leffler institute for the organization of the 2010 fall in Quantum Information Theory where this work was initiated. C.G.G. and D.P.G.’s research was supported by EU grant QUEVADIS and Spanish projects QUITEMAD and MTM2011-26912. B.C.’s research was supported by ANR GranMa, ANR Galoisint and NSERC grant RGPIN/341303-2007.

REFERENCES

[BCH+11] F. G. S. L. Brandão, P. ́Cwikli´ nski, M. Horodecki, P. Horodecki, J. Korbicz, and M. Mozrzymas. Convergence to equilibrium under a random hamiltonian. http://arxiv.org/abs/1108.2985, 2011.

[CFM+08] M. Cramer, A. Flesch, I. P. McCulloch, U. Schollwöck, and J. Eisert. Exploring local quantum many-body relaxation by atoms in optical superlattices. Phys. Rev. Lett., 101:063001, Aug 2008.

[CN10] Benoît Collins and Ion Nechita. Random quantum channels i: Graphical calculus and the bell state phenomenon. Communications in Mathematical Physics, 297(2):345–370, 2010.

[Col03] Benoît Collins. Moments and cumulants of polynomial random variables on unitarygroups, the itzykson-zuber integral, and free probability. International Mathematics Research Notices, 2003(17):953–982, 2003.

[Cra11] M. Cramer. Thermalization under randomized local Hamiltonians. http://arxiv.org/abs/1112.5295, 2011.

[FCM+08] A. Flesch, M. Cramer, I. P. McCulloch, U. Schollwöck, and J. Eisert. Probing local relaxation of cold atoms in optical superlattices. Phys. Rev. A, 78:033608, Sep 2008.

[Ful97] William Fulton. Young tableaux, volume 35 of London Mathematical Society Student Texts. Cambridge University Press, Cambridge, 1997. With applications to representation theory and geometry.

[GdOZ10] Silvano Garnerone, Thiago R. de Oliveira, and Paolo Zanardi. Typicality in random matrix product states. Physical Review A, 81(3), 03 2010.

[Has07] M. B. Hastings. Entropy and entanglement in quantum ground states. Physical Review B, 76(3), 07 2007.

[Jay57a] E. T. Jaynes. Information theory and statistical mechanics. Physical Review, 106(4), 05 1957.

[Jay57b] E. T. Jaynes. Information theory and statistical mechanics. ii. Physical Review, 108(2), 10 1957.

[Jay68] E.T. Jaynes. Prior probabilities. Systems Science and Cybernetics, IEEE Transactions on, 4(3):227 –241, sept. 1968.

[Led01] Michel Ledoux. The concentration of measure phenomenon, volume 89 of Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, 2001.

[LPSW09] Noah Linden, Sandu Popescu, Anthony J. Short, and Andreas Winter. Quantum mechanical evolution towards thermal equilibrium. Phys. Rev. E, 79:061103, Jun 2009.

[LPSW10] Noah Linden, Sandu Popescu, Anthony J Short, and Andreas Winter. On the speed of fluctuations around thermodynamic equilibrium. New Journal of Physics, 12(5):055021, 2010.

[MS86] V.D. Milman and G. Schechtman. Asymptotic theory of finite dimensional normed spaces. Lecture notes in mathematics. Springer, 1986.

[PGVWC06] D Perez-Garcia, F Verstraete, M M Wolf, and J I Cirac. Matrix product state representations. Quantum Inf Comput, 7(5 & 6):1–28, 2006.

[PSW06] Sandu Popescu, Anthony J. Short, and Andreas Winter. Entanglement and the foundations of statistical mechanics. Nat Phys, 2(11):754–758, 11 2006.

[TCF+11] S. Trotzky, Y.-A. Chen, A. Flesch, I. P. McCulloch, U. Schollwöck, J. Eisert, and I. Bloch. Probing the relaxation towards equilibrium in an isolated strongly correlated 1D Bose gas. http://arxiv.org/abs/1101.2659, 2011.

[Wei78] D. Weingarten. Asymptotic behavior of group integrals in the limit of infinite rank. Journal of Mathematical Physics, 19:99–1001, may 1978.

[Whi92] Steven R. White. Density matrix formulation for quantum renormalization groups. Phys. Rev. Lett., 69:2863–2866, Nov 1992.

APPENDIX A. PROOF OF THEOREM 4.5

We can take first the average in U and using equation [2] together with the bound of theorem [4.1] we have
The number of loops relating the circles is $\#\gamma^{-1}\alpha'\beta'^{-1} - 2 = 2n + 2 - |\gamma^{-1}\alpha'\beta'^{-1}|$, taking into account those where $L$ or $R$ appears. Note that we are subtracting 2 loops $(2n + 1)(2n + 2)$ that we have added when including the (non-existing) links of the permutation $\gamma$. The number of loops relating the squares is
\#\tau\alpha = n - |\tau\alpha|. All the loops are trivial and thus they correspond to the dimension of the system, except those where L or R appears, in which we will take averages. For \( \alpha = \beta = 1 \), we have that

\[
\mathbb{E}_{L,R}[C_{(1,1)}] = \mathbb{E}_{L,R}[tr(L)^2 tr(R)^2 D^{2n-2} d^{2n-l}] = 1/4D^{2n} d^{2n-l}.
\]

Taking averages in L and R and using the bounds from lemma 4.3, it can be shown, by inspection on all possible combinations of L and R in different loops, that it is enough to distinguish the following two cases:

1. \( \alpha, \beta \in A_{2n} \), where \( A_{2n} \) is the set of tuples where \( 2n+3 \) and \( 2n+4 \) are not in the same cycle of the permutation \( \gamma^{-1} \alpha' \gamma \alpha'^{-1} \beta \), that is, both R matrices appear in different loops. In this case we have that

\[
\mathbb{E}_{L,R}[C_{(\alpha,\beta)}] \leq 1/4D^{2n-\gamma^{-1} \alpha' \gamma \alpha'^{-1} \beta} d^{2n-|\tau\alpha|} \leq D^{2n-\gamma^{-1} \alpha' \gamma \alpha'^{-1} \beta} d^{2n}.
\]

Making the change of variables \( h = \beta\alpha^{-1} \), \( g = \gamma^{-1} \alpha' \gamma \alpha'^{-1} \beta \) that is proven to be one to one in lemma 4.2 and denoting \( h' = h(2n+1)(2n+2)(2n+3)(2n+4) \) we get that

\[
\sum_{(\alpha,\beta)\in A_{2n}} \mathbb{E}_{L,R}[C_{(\alpha,\beta)}] W g(Dd, \beta\alpha^{-1}) \leq K \sum_{(g,h)\in S_{2n+4} \times S_{2n}} D^{-|gh^{-1}|-(3/5)|h|} d^{-(3/5)|h|} =
\]

\[
= K \sum_{g \neq 1 \in S_{2n+4}} D^{-|g|} + K \sum_{h \neq 1, g \in S_{2n+4}} D^{-|gh^{-1}|-(3/5)|h|} \leq
\]

\[
\leq K \left( \sum_{|g|=1} (2n+4)(2n+3) 2D - |g| + \sum_{|g|=0} (2n+4)(2n+3) 2D |g| \sum_{|h|=1} (2n(2n-1)) 2D^{3/5} \right) \leq
\]

\[
\leq K \left( \frac{(2n+4)(2n+3)}{2D - (2n+4)(2n+3)} + \frac{2D}{2D - (2n+4)(2n+3)} \frac{2n(2n-1)}{2D^{3/5} - 2n(2n-1)} \right)
\]

In the first inequality we just upper bound the number of permutations with a given number of transpositions, the second is just a geometric sum. As \( D \geq n^5 \) we get further

\[
\sum_{(\alpha,\beta)\in A_{2n} \backslash \{(1,1)\}} \mathbb{E}_{L,R}[C_{(\alpha,\beta)}] W g(Dd, \beta\alpha^{-1}) \leq O(D^{-1/5})
\]

2. \( \alpha, \beta \in B_{2n} \), where \( B_{2n} \) is the set of tuples where \( 2n+3 \) and \( 2n+4 \) are in the same cycle of the permutation \( \gamma^{-1} \alpha' \gamma \alpha'^{-1} \beta \), that is, both R matrices appear in the same loop. In this case we have that

\[
\mathbb{E}_{L,R}[C_{(\alpha,\beta)}] \leq 1/4D^{2n+1-\gamma^{-1} \alpha' \gamma \alpha'^{-1} \beta} d^{2n-|\tau\alpha|} \leq D^{2n+1-\gamma^{-1} \alpha' \gamma \alpha'^{-1} \beta} d^{2n}.
\]

Applying the same change of variables and consider \( B'_{2n} \) the image of \( B_{2n} \) under the change of variables, we get

\[
\sum_{(\alpha,\beta)\in B_{2n}} \mathbb{E}_{L,R}[C_{(\alpha,\beta)}] W g(Dd, \beta\alpha^{-1}) \leq K \sum_{(g,h)\in B'_{2n}} D^{-1-|gh^{-1}|-(3/5)|h|} d^{-(3/5)|h|}.
\]

In order to bound this sum one has to proceed more carefully. The proof follows by bounding independently over the different cases where: \( h = 1 \), \( h = (2n+3,2n+4) \), \( h \) is a different transposition, and the rest of terms. For all these cases one has to take into account the properties of the elements in \( B'_{2n} \), that is, \( 2n+3 \) and \( 2n+4 \) belongs to the same cycle of \( gh^{-1} \) and the parity of \( |gh^{-1}| + |h| \) that is proven in lemma 4.2. Following this procedure one can prove that

\[
\sum_{(\alpha,\beta)\in B_{2n}} \mathbb{E}_{L,R}[C_{(\alpha,\beta)}] W g(Dd, \beta\alpha^{-1}) \leq O(D^{-1/5})
\]

The result follows joining the two cases and the case \( \alpha = \beta = 1 \)
Appendix B. Proof of Theorem 4.6

We use the notation $\| \cdot \|_p$ for the Schatten p-norm. For the first function we have

$$
\| f \|_{Lip} = \frac{|f(U, V, W, \Lambda, \Omega) - f(U', V', W', \Lambda', \Omega)|}{d((U, V, W, \Lambda), (U', V', W', \Lambda'))} \leq \frac{|(tr \rho_1(U, V, W, \Lambda, \Omega)) - (tr \rho_1(U', V', W', \Lambda'))|^2}{d((U, V, W, \Lambda), (U', V', W', \Lambda'))}
$$

$$
\leq 2 |tr(\rho(U, V, W, \Lambda, \Omega) - \rho(U', V', W', \Lambda'))| \leq 2 tr|\rho(U, V, W, \Lambda, \Omega) - \rho(U', V', W', \Lambda')| \leq \frac{2 |V^\Lambda V^\dagger U^{n}(W^\Lambda W^\dagger \otimes (|0\rangle \langle 0|)^{\otimes n})(U^\dagger)^n|}{\|U - U'\|_2 + \|V - V'\|_2 + \|W - W'\|_2 + \|\Lambda - \Lambda'\|_\infty},
$$

where we are using standard inequalities and for the sake of simplicity we are denoting by the same U unitaries that are acting on different systems and with the same letter V a unitary and its tensor product with the identity. Adding and substracting terms and applying the triangular inequality we get 2n+5 terms of the following form

$$
2 |V^* \Lambda^* V^\dagger U^{n}(W^* \Omega W^* \otimes (|0\rangle \langle 0|)^{\otimes n})(U^\dagger)^n| \leq \frac{2 |V^* \Lambda^* V^\dagger U^{n}(W^* \Omega W^* \otimes (|0\rangle \langle 0|)^{\otimes n})(U^\dagger)^n|}{\|U - U'\|_2 + \|V - V'\|_2 + \|W - W'\|_2 + \|\Lambda - \Lambda'\|_\infty},
$$

where any $X^*$ stands for $X$, $X'$ or $X - X'$ and in any term the latter only appears once. Then

$$
\| f \|_{Lip} \leq \frac{(4n + 10) |V^* \Lambda^* V^\dagger U^{n}(W^* \Omega W^* \otimes (|0\rangle \langle 0|)^{\otimes n})(U^\dagger)^n|}{\|U - U'\|_2 + \|V - V'\|_2 + \|W - W'\|_2 + \|\Lambda - \Lambda'\|_\infty},
$$

Applying the inequality $\|XY\|_1 \leq \|X\|_1 \|Y\|_\infty$ we get

$$
\| f \|_{Lip} \leq \frac{(4n + 10) |V^* \Lambda^* V^\dagger U^{n}(W^* \Omega W^* \otimes (|0\rangle \langle 0|)^{\otimes n})(U^\dagger)^n|}{\|U - U'\|_2 + \|V - V'\|_2 + \|W - W'\|_2 + \|\Lambda - \Lambda'\|_\infty}.
$$

Now, by the decomposition we have done, any term has only one norm in the numerator of the form of the ones in the denominator. The other norms in the numerator are trivially bounded by one. Thus we get

$$
\| f \|_{Lip} \leq 4n + 10
$$

For the second function we have

$$
\| g \|_{Lip} = \frac{|g(U, V, W, \Lambda, \Omega) - g(U', V', W', \Lambda', \Omega)|}{d((U, V, W, \Lambda), (U', V', W', \Lambda'))} \leq \frac{|tr(\rho_1^2(U, V, W, \Lambda, \Omega) - \rho_1^2(U', V', W', \Lambda', \Omega))|}{\|U - U'\|_2 + \|V - V'\|_2 + \|W - W'\|_2 + \|\Lambda - \Lambda'\|_\infty} \leq 4n + 10
$$

where the result follows using the same techniques.