Stochastic Matrix Product States

Kristan Temme and Frank Verstraete

Faculty of Physics, University of Vienna, Boltzmanngasse 5, A-1090 Vienna, Austria

(Dated: October 26, 2018)

The concept of stochastic matrix product states is introduced and a natural form for the states is derived. This allows to define the analogue of Schmidt coefficients for steady states of non-equilibrium stochastic processes. We discuss a new measure for correlations which is analogous to the entanglement entropy, the entropy cost S_C, and show that this measure quantifies the bond dimension needed to represent a steady state as a matrix product state. We illustrate these concepts on the hand of the asymmetric exclusion process.

This paper explores parallels between the many-body description of non-equilibrium steady states of classical stochastic processes and ground states of strongly correlated quantum many-body systems. Classical non-equilibrium steady states are typically much richer than their equilibrium counterparts and can exhibit interesting behavior such as the presence of a current, non-equilibrium phase transitions and entire phases with a diverging correlation length [1, 2]. There are strong parallels between the many-body description of non-equilibrium classical stochastic spin systems [3], and it turns out that matrix product states (MPS) play a very important role in both fields [4, 5].

The theory of entanglement, developed in the context of quantum information theory, has recently proven to yield valuable new insights in the nature of the wavefunctions arising in such strongly correlated quantum many-body systems. One of its main virtues is that it allows to quantify the amount of quantum correlations in a many-body system in terms of so–called area laws [6–8] and relate this to the structure of the underlying wavefunctions representing the ground states of quantum spin Hamiltonians [9, 10]. The ground states of quantum spin chains is well described by the class of MPS, since for local Hamiltonians the associated Schmidt coefficients of the ground state decay very fast. A MPS approximation with finite bond dimension can be obtained by setting the smallest Schmidt coefficients σ_I equal to zero. This approximation is justified whenever an area law is satisfied [9]. If we define $p_i = \sigma_i^2$, then those p_i form a probability distribution. The entropy of this distribution quantifies the dimension of the matrices that has to be chosen in the MPS approximation such as to obtain a given fidelity. For such pure states, this entropy has multiple operational meanings, ranging from the amount of distillable Bell states by local operations to the maximum amount of secret information that can be sent in a cryptographic setting [11]. For mixed states and in the presence of classical correlations the situation is more complicated. A lot of research in quantum information has concentrated on resolving the relationships between classical and quantum correlations in mixed quantum states. A measure

of particular importance to our approach is the entanglement of purification [12]. It quantifies both quantum and classical correlations in terms of the number of maximally entangled states needed to asymptotically generate the state.

The main topic of this paper is to explore the classical non-equilibrium analogous of the quantum notions of entanglement entropy, area laws and the density matrix renormalization group which justify the use of MPS in the quantum setting. The main technical difficulty arises from the fact that classical probability distributions are normalized in the L_1 norm ($\sum_i |p_i| = 1$), while quantum states are normalized in the L_2 norm ($\sum_i |\psi_i|^2 = 1$). This can partly be overcome by working with a subclass of MPS where all matrices only contain non-negative entries; we call this subclass stochastic matrix product states (sMPS). The concept of mutual information, defined for classical bipartite distributions p_{AB} where A and B will represent the variables or spins on both halves of a chain, plays a role analogous to the entanglement entropy: $I(A : B) = \sum_{AB} p_{AB} \log_2 \left(\frac{p_{AB}}{p_A p_B} \right)$. It immediately gives an upper bound to the error made when approximating p_{AB} by a product of its marginals, since $\|p_{AB} - p_A p_B\|_1^2 \leq 2\ln(2)I(A : B)$ [13]. Just as in the quantum case, one would expect that the global non-equilibrium steady state probability distribution of the stochastic process can be represented as a stochastic matrix product state (sMPS) with small bond dimension if this mutual information is small. However, more subtle measures are needed in the case of stochastic processes, and we will introduce the notion of entropy cost to quantify the bond dimension needed for the corresponding sMPS.

Definition 1 *The entropy cost S_C for a bipartite probability distribution $P(x, y)$ is given by:*

$$S_C = \min_{p_\lambda, P_A, P_B} S(\{p_\lambda\}) \qquad (1)$$

where $S(\{p_\lambda\}) = -\sum_\lambda p_\lambda \log_2 (p_\lambda)$ is the Shannon information of $\{p_\lambda\}$, and where the optimization is over all probability distributions p_λ and over all conditional probabilities P_A and P_B for which $P(x, y) =$

$\sum_\lambda P_A(x|\lambda)P_B(y|\lambda)p_\lambda$.

The entropy cost bears a lot of resemblances to the notion of common information introduced by Wyner in the context of cryptography and classical information theory [14], and this entropy cost serves as an upper bound to the common information. The entropy cost can be thought of as the classical analogue of the entanglement of purification, and the probability distribution p_λ plays a role analogous to the Schmidt coefficients in the quantum case.

To give an example, consider the equilibrium system defined by the classical Ising model $H = \sum_{i=1}^{N-1} s_i s_{i+1}$, with $s_i = \pm 1$. The equilibrium distribution $p(\{s_i\})_I = \exp(-\beta H)/Z$ can be written in terms of a very simple MPS with $D = 2$ [15]; the entropy cost S_C can then easily be calculated exactly in the case $N = 2$ and is given by

$$
\begin{aligned}
S_C = - \big(&\mathrm{e}^{-\beta}\cosh(\beta)\log_2\left(\mathrm{e}^{-\beta}\cosh(\beta)\right) + \\
&\mathrm{e}^{-\beta}\sinh(\beta)\log_2\left(\mathrm{e}^{-\beta}\sinh(\beta)\right)\big).
\end{aligned} \quad (2)
$$

As expected, this function monotonously increases from 0 to 1, i.e. from the paramagnetic without correlations to the ferromagnetic region with strong correlations. A different measure that was recently investigated is the shared information [16]. It has been shown that it obeys an area law for several non-critical stochastic models and that critical behavior can be identified by logarithmic corrections.

Let us next define a D-dimensional sMPS describing a classical probability distribution of N-spins each of dimension d; obviously, those sMPS were already extensively used in the literature, and we are just formalizing the definition here.

Definition 2 *A stochastic matrix product state (sMPS) is given by:*

$$
|p_D\rangle = \sum_{i_1,\ldots,i_N=1}^{d} \langle L\,|\,B_{i_1}^1 \ldots B_{i_N}^N\,|\,R\rangle\,|\,i_1 \ldots i_N\rangle, \quad (3)
$$

where we only consider real matrices that are $D_k \times D_{k+1}$ dimensional, with $D_k \leq D$, and additionally fulfill the requirement $[B_{i_k}^k]_\delta^\gamma \geq 0$ for every element individually. This ensures that all the weights of the distribution are positive after contraction. The left and right vector $\langle L|$ and $|R\rangle$ are also elementwise positive and can be absorbed into the matrices $B_{i_1}^1$ and $B_{i_N}^N$, which corresponds to choosing $D_1 = D_N = 1$. Furthermore, we require $|p_D\rangle$ to be normalized in the L_1 norm, $\|\,|p_D\rangle\,\|_1 = 1$.

Every multipartite probability distribution of a chain of discrete variables can obviously be written in the form (3) if we allow for a sufficiently large matrix dimension $D^{\max}$, i.e. exponential in the number of sites.

Let us next show how the entropy cost can be calculated. Upon inserting a partition of unity $\mathbb{I} = \sum_{\lambda=1}^{D_k} |\lambda\rangle\langle\lambda|$ in (3), we can write $|p\rangle =$

$\sum_\lambda \sum_{\{i_n\}} \langle L\,|\,B_{i_1}^1 \ldots B_{i_k}^k\,|\,\lambda\rangle \langle\lambda\,|\,B_{i_{k+1}}^{k+1} \ldots B_{i_N}^N\,|\,R\rangle\,|\,\{i_n\}\rangle$.
Now

$$
p_\lambda = \langle L\,|\,\prod_{n=1}^{k} C^{[n]}\,|\,\lambda\rangle \langle\lambda\,|\,\prod_{n=k+1}^{N} C^{[n]}\,|\,R\rangle. \quad (4)
$$

defines a new probability distribution if we chose $C^{[n]} = \sum_{i_n} B_{i_n}^{[n]}$ (i.e. the transfer matrix). This allows us to rewrite the MPS as

$$
|p\rangle = \sum_{\lambda=1}^{D_k} \sum_{i_1 \ldots i_N} P_A(\{i_n\}_{n\in A}|\lambda)p_\lambda P_B(\{i_n\}_{n\in B}|\lambda) \quad (5)
$$

where

$$
\begin{aligned}
P_A(\{i_n\}_{n\in A}|\lambda) &= \frac{\langle L\,|\,B_{i_1}^1 \ldots B_{i_k}^k\,|\,\lambda\rangle}{\langle L\,|\,\prod_{l=1}^{k} C^{[l]}\,|\,\lambda\rangle} \\
P_B(\{i_n\}_{n\in B}|\lambda) &= \frac{\langle\lambda\,|\,B_{i_{k+1}}^{k+1} \ldots B_{i_N}^N\,|\,R\rangle}{\langle\lambda\,|\,\prod_{l=k+1}^{N} C^{[l]}\,|\,R\rangle}.
\end{aligned} \quad (6)
$$

The probability distribution $\{p_\lambda\}$ sums up to one due to the normalization of (3), and $P_A(\{i_n\}|\lambda)$ and $P_B(\{i_n\}|\lambda)$ can be interpreted as information channels. Note that there are several partitions of unity that will give rise to valid P_A, P_B and p_λ, and therefore the decomposition (5) is not unique. It is probably a NP-hard problem to find the optimal decomposition that minimizes the entropy cost, and in practice we will therefore rely on the construction that was just given for finding upper bounds to it. It is also easy to find lower bounds to the entropy cost:

Lemma 1 *For a given distribution $|p\rangle$ the mutual information $I(A : B)$ is bounded by the entropy cost S_C: $I(A : B) \leq S_C$.*

Proof: By virtue of (5) we can focus on calculating the mutual information $I(\lambda : \mu)$ of the distribution $P(\lambda, \mu) = p_\lambda \delta_{\lambda,\mu}$, since the full distribution can be read as $P_{AB} = \sum_{\lambda,\mu} P_A(\{i_A\}|\lambda)P_B(\{i_B\}|\mu)P(\lambda, \mu)$. This corresponds to a source that generates two outputs, which are then transformed by the channels P_A and P_B. The data-processing inequality [13] guarantees that the mutual information of the processed source $I(A : B) = I(p_A(P) : p_B(P)) \leq I(\lambda : \mu)$ is smaller than the mutual information of the source itself which is equal to its entropy $I(\lambda : \mu) = S(\{p_\lambda\})$. The decomposition $[P_A(\{i_A\}|\lambda), P_B(\{i_B\}|\lambda), p(\lambda)]$ is not unique, therefore the bound is improved by taking the minimum over all decompositions. $\square$

The decomposition of the sMPS as given in (5) suggests the existence of the following (non-unique) normal form:

$$
|p\rangle = \sum_{i_1 \ldots i_N} A_{i_1}^{[1]} P^{[1]} A_{i_2}^{[2]} \ldots P^{[N-1]} A_{i_N}^{[N]}\,|\,i_1 \ldots i_N\rangle. \quad (7)
$$

Here the matrices $P^{[k]}$ represent diagonal matrices with probabilities $\{p_{\lambda_k}\}$ sorted in decreasing order. The matrices $C^{[k]} = \sum_{i_k} A^{[k]}$, where $[A_{i_k}^{[k]}]_\delta^\gamma \geq 0$, combined

with the $P^{[k]}$'s behave as stochastic matrices. That is, $P^{k-1}C^{[k]} = S^k$ and $C^k P^k = S_k^T$, where S^k and S_k denote different stochastic matrices.

To see that every MPS distribution can be written this way consider the following scenario: We start by introducing the first bipartitioning between the first two sites. After the necessary normalization we proceed to the next site and perform the same procedure renormalizing the resulting matrices by the total contraction of the two halves of the chain. Proceeding along the chain results in the desired form.
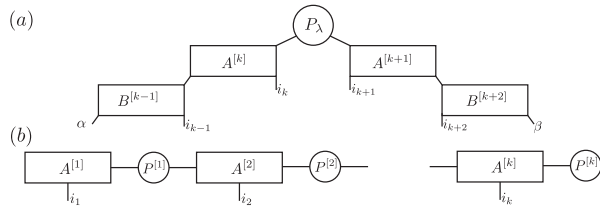

FIG. 1: Pictorial representation of the natural sMPS decomposition: Image (a) can be seen as the graphical representation of eqn. (5). From a given source P_λ, the correlations are distributed via the two channels on the left and right. The normalizing factor is included in the A's. In (b) the analogy to the quantum MPS becomes evident for the decomposition as given in eqn. (7).The probabilities in the matrices $P^{[k]}$ are the analogous of the singular values which arise upon a Schmidt decomposition of the quantum state [17].

This representation (7) enables us to give a good estimate on the error measured in the L_1 norm which is made upon truncating the dimension of the source space, i.e. neglecting probabilities smaller than a given value along each bipartition:

Lemma 2 *For every multipartite distribution* $|p\rangle$ *there exists a MPS* $|p_D\rangle$ *of the form (7) with dimension D, such that*

$$\||p\rangle - |p_D\rangle\|_1 \le 2\sum_{k=1}^{N-1} \epsilon_k(D),$$

where $\epsilon_k(D) = \sum_{\lambda=D+1}^{D_k^{max}} p_\lambda^{[k]}$.

Proof: We can always write $|p\rangle$ as a distribution of the form (7) with a $D_k^{\max} = d^N$. We now introduce another MPS $|p_D\rangle$ in natural form with a bond dimension of D. Let $|p_D\rangle = |p_D^*\rangle / \||p_D^*\rangle\|_1$, where $|p_D^*\rangle$ is the unnormalized probability distribution which arises from neglecting along each cut all the probabilities $\{p_{\lambda_k}\}_{D+1}^{D_k^{\max}}$. We write $|p_D^*\rangle = \sum_{\{i_k\}} A_{i_1}^{[1]} P^{*[1]} \ldots A_{i_N}^{[N]} |\{i_k\}_{k=1}^N\rangle$. Note, that if $\||p\rangle - |p_D^*\rangle\|_1 \le \epsilon$, then $\||p\rangle - |p_D\rangle\|_1 \le 2\epsilon$ due to the triangle inequality. Since $|p_D^*\rangle$ arose by only neglecting positive numbers, we may write $\||p\rangle - |p_D^*\rangle\|_1 = C^{[1]} P^1 \ldots C^{[N]} - C^{[1]} P^{*1} \ldots C^{[N]} = (\langle l^{N-1}| - \langle l^{N-1}|^*)C^{[N]} = \|(\langle l^{N-1}| - \langle l^{N-1}|^*)\|_1$. Here we have used the fact that all summands are

positive and we defined $\langle l^k| = C^{[1]} P^{[1]} \ldots P^{[k]}$ as well as $\langle l^k|^* = C^{[1]} P^{*[1]} \ldots P^{*[k]}$. The difference $\|\langle l^1| - \langle l^1|^*\|_1 = \|\langle 1\ldots1| (P^{[1]} - P^{*[1]})\|_1 = \sum_{\alpha_1=D+1}^{D_1^{\max}} p_\lambda$ is simply given by $\epsilon_1(D)$. Note that due to (7) $\langle l^{k-1}|C^{[k]} = \langle 1\ldots1|$. Proceeding to calculate the difference for other k we find that $\|\langle l^{[k]}| - \langle l^{*[k]}|\|_1 \le \|(\langle l^{[k-1]}| - \langle l^{*[k-1]}|) C^{[k]} P^{[k]}\|_1 + \|\langle l^{*[k-1]}|C^{[k]} (P^{[k]} - P^{*[k]})\|_1 \le \sum_{n=1}^{k-1} \epsilon_n(D) + \|\langle 1\ldots1| (P^{[k]} - P^{*[k]})\|_1$. The last summand corresponds exactly to $\sum_{\alpha_k=D+1}^{D_k^{\max}} p_{\alpha_k} = \epsilon_k(D)$, which completes the proof. $\square$

We have therefore proven that there exists an efficient parameterization of the steady state in terms of a sMPS with low bond dimension if there exists a parameterization of this steady state for which the entropy cost with respect to all bipartite cuts is small: if this is the case, then $\sum_{k=1}^{N-1} \epsilon_k(D)$ can be made small by following the arguments outlined in [9]. This is the analogue of the quantum case for which the existence of an area law implies the existence of an efficient representation in terms of MPS. Note however that the classical statement is a bit weaker, as the same normal form has to be used with respect to all bipartite cuts, and there is no guarantee that the same parameterization is optimal for all of bipartitions.

To make the investigations concrete, we consider the non-equilibrium steady state of the asymmetric exclusion process (ASEP) [18]. This process is modelled by a chain of sites labeled $k = 1\ldots N$ occupiable by hardcore particles, i.e. classical spins $i_k \in \{0,1\}$. The particles are only allowed to hop to the right, and this only if the next site is empty. To drive the system, particles at the left are injected with a given rate α and removed on the right with a rate β. This process exhibits three phases determined by the inflow α and outflow β. As was shown in [18] the steady state $|p\rangle$ of the corresponding master equation can be found exactly in terms of a MPS, albeit one for which the matrices are infinite dimensional and not necessarily positive.

The unnormalized steady state solution of the ASEP master equation [18] is given in MPS form (3) with site independent matrices $B_0^1 = \ldots = B_0^N = \mathcal{E}$ and $B_1^1 = \ldots = B_1^N = \mathcal{D}$ and boundary vectors $\langle L|, |R\rangle$, where all those satisfy the algebraic constraints $\mathcal{DE} = \mathcal{E} + \mathcal{D}$, $\langle L|\mathcal{E} = \frac{1}{\alpha}\langle L|$ and $\mathcal{D}|R\rangle = \frac{1}{\beta}|R\rangle$. Except for the special case when $\alpha + \beta = 1$, the representations obeying those constraints are infinite dimensional. As the total occupation number of particles is limited by N, we will construct truncated representations of a given dimension $D = N + 1$ that still reproduce the exact solution for a chain of length N; the entropy cost can then immediately be upper bounded by the logarithm of the system size, just as in the case of critical quantum spin chains. These representations are of the form

$\mathcal{E} = \sum_{i,j=0}^{1} [B]_{i,j} |i\rangle\langle j| + \sum_{n=2}^{N} |n\rangle\langle n| + |n\rangle\langle n-1|$ and $\mathcal{D} = \sum_{i,j=0}^{1} [A]_{i,j} |i\rangle\langle j| + \sum_{n=2}^{N} |n\rangle\langle n| + |n-1\rangle\langle n|$. With the left and right vector $\langle L| = \sum_{n=0}^{1} w_n \langle n|$ and $|R\rangle = \sum_{n=0}^{1} v_n |n\rangle$. Here A and B are 2-dimensional matrices fulfilling $AB + \sigma^- \sigma^+ = A + B$ (here σ^+ and σ^- are the Pauli raising and lowering operators), and the 2-component vectors v and w must be chosen to be eigenvectors: $Av = (1/\beta)v$, $wB = (1/\alpha)w$.

The construction given in (5) gives us a way of bounding the entropy cost as a function of the parameters α, β and N. We varied over all A, B, v, w satisfying the algebraic relations, and found that the minimum is obtained for 3 different solutions depending on the parameter range of α and β. The solution is depicted in Fig 2. If $\alpha + \beta \leq 1$ and $\beta \leq \alpha$, then $A = 1/\beta|0\rangle\langle 0| + |1\rangle\langle 1|$, $B = 1/(1-\beta)|0\rangle\langle 0| + b|1\rangle\langle 0| + 1/\alpha|1\rangle\langle 1|$, $w = (\alpha(1-\beta) \quad b)$ and $v^T = (1 \quad 0)$ with $b = \sqrt{1 - \alpha - \beta}$. However, if $\alpha + \beta \leq 1$ but $\beta \geq \alpha$, the optimal solution can be obtained from the previous one by the replacements, $\tilde{A} = B^T(\alpha \rightleftarrows \beta)$, $\tilde{B} = A^T(\alpha \rightleftarrows \beta)$, and $\tilde{w} = v^T$, $\tilde{v} = w^T(\alpha \rightleftarrows \beta)$. For $\alpha + \beta \geq 1$, the optimal solution is given by $A = 1/\beta|0\rangle\langle 0| + a|0\rangle\langle 1| + |1\rangle\langle 1|$, $B = 1/\alpha|0\rangle\langle 0| + a|1\rangle\langle 0| + |1\rangle\langle 1|$, $v^T = w = (1 \quad 0)$ and $a = \sqrt{1/\alpha + 1/\beta - 1/\alpha\beta}$.
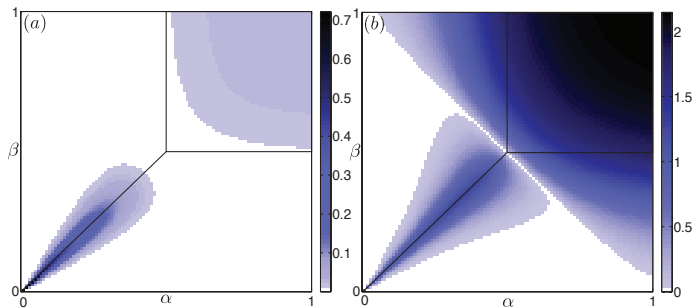

FIG. 2: (color online) Figure (a) The mutual information of a chain length $N = 20$ for different inflow parameters α and β has been calculated numerically. Figure (b) entropy cost for a representation with $D^{\max} = 21$. Note the different scales of the two plots. The resulting plots clearly reflect the underlying phase diagram of the ASEP [18], both for the mutual information and the approximate entropy cost. The entropy cost as well as the actual mutual information drop to zero along the mean-field line $\alpha + \beta = 1$. The mutual information is consistently low throughout the diagram, explaining why mean-field approaches have given such good results [1].

Conclusions: We have revisited the notion of stochastic matrix product states, and showed that a low bond dimension suffices to efficiently parameterize steady states of non-equilibrium distribution if the entropy cost in the system is low. This opens up the interesting question to characterize the conditions under which such steady states have a low entropy cost. It would be interesting to see to what extent this relates to the gap of the corresponding stochastic process. This also opens up novel ways for constructing numerical renormalization
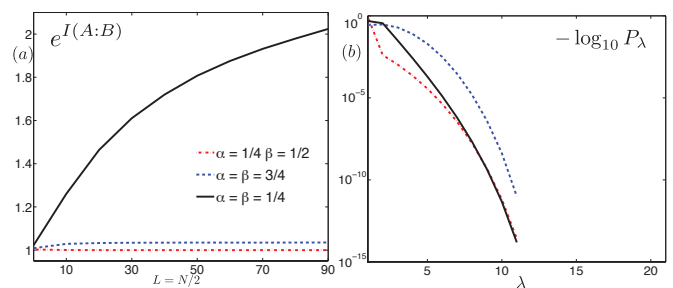

FIG. 3: (color online) Plot (a): Monte Carlo simulations of the mutual information for different α and β values. The system size was varied from $N = 2 \ldots 180$. The chain was cut in the middle, the length L of the individual blocks are given by half of the total system size. The plot shows the exponential of the mutual information $\exp I(A : B)$. The simulations suggest that the mutual information grows only sub-logarithmically along the coexistence line $\alpha = \beta$, $\alpha + \beta \leq 1$; Plot (b): The logarithm of the Probability distribution $\{p_\lambda\}$ plotted for different values of α and β. The dimension of the matrices $\mathcal{E}$ and $\mathcal{D}$ is $D^{\max} = 21$, corresponding to a chain of length $N = 20$. The distributions decay super-exponentially. For all $\lambda \geq 11$, $p_\lambda = 0$.

group methods for simulating non-equilibrium systems in the line of the MPS algorithms for quantum spin chains [5, 19, 20].

[1] B. Derrida, et al. *J. Stat. Phys.* **69**, 667-687 (1992)
[2] B. Derrida, M. R. Evans,*J. Phys. I France* **3**, 311 (1993)
[3] B. U. Felderhof,*Rep. Math. Phys.* **1**, 215234 (1971)
[4] R. A. Blythe, M. R. Evans,*J. Phys. A Math. Theor.* **40**, R333-R441 (2007)
[5] F. Verstraete, et al. *Adv. Phys.* **57**,143 (2008).
[6] J. Eisert, et al. *Rev. Mod. Phys.* **82**, 277 (2010).
[7] G. Vidal, et al. *Phys. Rev. Lett.* **90**, 227902 (2003)
[8] P. Calabrese, J. Cardy, *J. Stat. Mech.* P06002 (2004)
[9] F. Verstraete and J. I. Cirac, *Phys. Rev. B* **73**, 094423 (2006)
[10] M. B. Hastings, *J. Stat. Mech.* P08024 (2007)
[11] A. Winter, *Proc. IEEE (2005) Symposium*, 2270-2274
[12] B. Terhal et al, *J. Math. Phys.* **43**, 4286-4298 (2002)
[13] T. M. Cover and J. A. Thomas,*Elements of Information Theory* (Wiley) (1991)
[14] A. D. Wyner,*IEEE Information Theory*, Vol. IT-21, No. 2, (1975)
[15] F. Verstraete, et al. *Phys. Rev. Lett.* **96**, 220601 (2006)
[16] F.C. Alcaraz, et al. *Phys. Rev. E* **80** 030102(R) (2009)
[17] G. Vidal,*Phys. Rev. Lett.* **91**, 147902 (2003)
[18] B. Derrida, et al. *J. Phys. A: Math. Gen.* **26**, 1493 (1993)
[19] S. R. White, Phys. Rev. Lett. **69**, 2863 (1992)
[20] U. Schollwöck, Rev. Mod. Phys. **77**, 259 (2005)