Invited Review Papers

Random State Technology

Fengpin Jin[1], Dennis Willsch[1], Madita Willsch[1], Hannes Lagemann[1,2],
Kristel Michielsen[1,2], and Hans De Raedt[1,3]*

[1]*Institute for Advanced Simulation, Jülich Supercomputing Centre, Forschungszentrum Jülich, Jülich 52425, Germany*
[2]*RWTH Aachen University, Aachen 52056, Germany*
[3]*Zernike Institute for Advanced Materials, University of Groningen, Groningen 9747 AG, The Netherlands*

We review and extend, in a self-contained way, the mathematical foundations of numerical simulation methods that are based on the use of random states. The power and versatility of this simulation technology is illustrated by calculations of physically relevant properties such as the density of states of large single particle systems, the specific heat, current–current correlations, density–density correlations, and electron spin resonance spectra of many-body systems. We explore a new field of applications of the random state technology by showing that it can be used to analyze numerical simulations and experiments that aim to realize quantum supremacy on a noisy intermediate-scale quantum processor. Additionally, we show that concepts of the random state technology prove useful in quantum information theory.

1. Introduction

Random numbers are ubiquitous in computational and theoretical physics. A prime example is random matrix theory, whose application in physics emerged when Eugene Wigner introduced the notion of random matrices to model the nuclei of heavy atoms, explaining the level spacing in the spectra of the nuclei.[1] In the meantime, random matrix theory has found applications in many other fields such as quantum chaos,[2] quantum optics,[3] and quantum computing,[4] as well as in other disciplines of engineering and finance.[5] In quantum statistical mechanics, a way to study eigenstate thermalization is to introduce a small perturbation in the form of a random matrix.[6] The resulting model is assumed to be applicable to physical situations, in particular to the study of equilibration of local observables.[7,8]

In the past two decades, the concept of the so-called canonical typicality has been developed.[9–15] In essence, the idea is that the reduced density matrix of a system is a thermal state if the state of the whole system is a pure state of the overwhelming majority of wave functions, that is a random state on the subspace defined by a small energy interval containing the energy of the microcanonical ensemble. Similar ideas appear in the works from Bocchieri and Loinger,[16] and Tasaki,[17] who did not emphasize the role of the random state though. Gemmer et al. studied the process of equilibration of a two-level system coupled to a special many-level bath and derived a rate equation for the two levels by Hilbert space averaging.[18] The derivation implicitly uses the notion of the bath being in a pure random state. The idea of the thermal pure quantum state formulation of statistical mechanics was given in Refs. 19 and 20, some elements of which were already contained in S. Lloyd's Phd thesis in 1988.[21] This thesis was barely noticed in the community of typicality until the relevant chapter was made available on arXiv in 2013.[21]

In the numerical-simulation arena, (pseudo-)random numbers are key to all Monte Carlo methods.[22–24] The focus of this paper is on the theory and application of a numerical method for computing properties of quantum systems, based on the use of a random state, which in essence is just a collection of random amplitudes defining the state vector of the quantum system. We coin this numerical tool to calculate various static and dynamic properties of quantum systems as **random state technology**. In the literature, this technology is also referred to as quantum dynamical typicality.[25–28]

Calculating traces of square matrices is one of the basic computational problems in quantum statistical physics. If calculating each matrix element $\langle i|X|i\rangle$ for $i = 1, \ldots, D$ takes $\mathcal{O}(D)$ arithmetic operations, computing $\mathbf{Tr}\, X = \sum_{i=1}^{D} \langle i|X|i\rangle$ requires $\mathcal{O}(D^2)$ arithmetic operations. In general, D is the dimension of the Hilbert space containing the state vectors describing the quantum system. D grows exponentially with the system size; for instance, for a system of n qubits, $D = 2^n$. Random state technology makes calculations feasible that would otherwise be out of reach with present-day supercomputers because it changes the operation count from $\mathcal{O}(D^2)$ to $\mathcal{O}(D)$, a significant reduction if D is large.

To the best of our knowledge, the basic idea of the random state technology is due to Alben et al., who used a random phase vector to reduce the numerical work of calculating the density of states (DOS) for a particle moving on a lattice.[29] Similar ideas have been used in the quantum transfer matrix Monte Carlo Method,[30] the estimation of the eigenvalue spectrum (by calculating the moments of the distribution of eigenvalues),[31] the calculation of the DOS and other properties of the Anderson model and of the spin-1/2 Heisenberg model,[32–34] and in the calculation of the DOS[35–37] and linear response functions.[38–41] A rigorous justification of the working principle of the random state technology and applications to the calculation of the eigenvalues of very large matrices, the DOS, and static properties of spin models were given by Hams and De Raedt.[42]

One area of applications for which the random state technology proves very useful is the study of decoherence, thermalization, and dissipation of a quantum system interacting with a bath, that is an open quantum system.[43] A simple but approximate method to treat the dynamics of an open quantum system is to use a master equation for the reduced density matrix of the quantum system.[43] A direct, approximation-free simulation of a relatively large system (quantum system + bath) becomes possible[44–55] if we use random state technology to represent the bath state by a

©2021 The Author(s)

Table I. Overview of the combination of probability densities $p(\boldsymbol{\xi})$ and amplitudes $c_j(\boldsymbol{\xi})$ that we consider in this paper. Columns four to six give the moments that appear in Eqs. (3) and (4). Columns seven and eight give the exact expressions (see Appendix A for the proof) of the average and the variance of $D\langle\Phi|X|\Phi\rangle/\langle\Phi|\Phi\rangle$, respectively.

Case	$p(\boldsymbol{\xi})$	$c_j(\boldsymbol{\xi})$	$E[|c|^2]$	$E[|c|^2|\widehat{c}|^2]$	$E[|c|^4]$	$E\left[\dfrac{D\langle\Phi|X|\Phi\rangle}{\langle\Phi|\Phi\rangle}\right]$	$\text{Var}\left[\dfrac{D\langle\Phi|X|\Phi\rangle}{\langle\Phi|\Phi\rangle}\right]$
A	$\prod_{j=1}^{D}\frac{1}{\pi}e^{-|\xi_j|^2}$	$\frac{\xi_j}{\sqrt{|\xi_1|^2+\cdots+|\xi_D|^2}}$	$\frac{1}{D}$	$\frac{1}{D(D+1)}$	$\frac{2}{D(D+1)}$	$\mathbf{Tr}\,X$	$\frac{D\,\mathbf{Tr}\,XX^\dagger-|\mathbf{Tr}\,X|^2}{D+1}$
B	$\prod_{j=1}^{D}\frac{1}{\pi}e^{-|\xi_j|^2}$	$\frac{\xi_j}{\sqrt{D}}$	$\frac{1}{D}$	$\frac{1}{D^2}$	$\frac{2}{D^2}$	$\mathbf{Tr}\,X$	$\frac{D\,\mathbf{Tr}\,XX^\dagger-|\mathbf{Tr}\,X|^2}{D+1}$
C	$\left(\frac{1}{2\pi}\right)^D$	$\frac{1}{\sqrt{D}}e^{ia_j}$	$\frac{1}{D}$	$\frac{1}{D^2}$	$\frac{1}{D^2}$	$\mathbf{Tr}\,X$	$\mathbf{Tr}\,XX^\dagger-\sum_{i=1}^{D}|\langle i|X|i\rangle|^2$

projected random state, the so-called thermal pure quantum state.[19,20] Another important class of applications is calculation of various static and dynamic properties, such as the DOS, the specific heat, the diffusion constant, the optical conductivity, etc., for systems such as the spin-1/2 Heisenberg model,[27,56–61] the Hubbard model,[62,63] tight-binding models of 2D materials,[64–67] etc.

For the efficient calculation of dynamic properties, the random state technology has to be supplemented by an efficient method for computing the result of applying the time-evolution operator to the state vector. Methods most often used are Suzuki–Trotter product formula algorithms,[68,69] the Chebyshev polynomial algorithm,[70–72] a Lanczos-iteration based method,[73,74] and the Runge–Kutta method.[27,75]

The paper is structured as follows. In Sect. 2 we present an extensive, rigorous analysis of the random state technology using basic tools of calculus only. We derive general results for the mean and variance for different random states, and use Markov's lemma to derive bounds. Then, we discuss various applications of the random state technology such as the calculation of the DOS (Sect. 3), the specific heat (Sect. 4), current–current and density–density correlations, and electron spin resonance (ESR) spectra (Sect. 5). Sections 6 and 7 apply the ideas of the random state technology to analyze numerical simulations and experiments that aim to establish quantum supremacy on a noisy intermediate-scale quantum processor[76] and generalize a statement in quantum information theory.[77,78] Section 8 summarizes the material presented in this paper.

2. Theoretical Background

The underlying idea of the random state approach is that accurate approximations to $\mathbf{Tr}\,X$ may be obtained by computing $\langle\Phi|X|\Phi\rangle$ with the pure state $|\Phi\rangle$ defined by

$$|\Phi\rangle = \sum_{j=1}^{D} c_j|j\rangle, \tag{1}$$

where $\{|j\rangle \mid j=1,\ldots,D\}$ denotes (any) complete set of orthonormal basis states spanning the D-dimensional Hilbert space and where the complex-valued amplitudes $\{c_j = c_j(\boldsymbol{\xi}); \mid j=1,\ldots,D\}$ are functions of complex-valued random variables $\boldsymbol{\xi}=(\xi_1,\ldots,\xi_D)$, i.e., $\xi_j = a_j + ib_j$ with real-valued a_j and b_j, distributed according to a probability density $p(\boldsymbol{\xi})$. We denote the expectation and variance of a random variable $z=z(\boldsymbol{\xi})$ with respect to $p(\boldsymbol{\xi})$ by $E[z]$ and $\text{Var}[z]=E[z^2]-E^2[z]$, respectively. By construction, we have $E[1]=1$ and $\text{Var}[1]=0$.

We focus on the case where $p(\boldsymbol{\xi}) = p(a_1,b_1,\ldots,a_D,b_D)$ is a symmetric function of the a's and b's, meaning that its

value does not change if we interchange its arguments. Furthermore, we impose the constraint that $p(\boldsymbol{\xi}) = p(a_1,b_1,\ldots,a_D,b_D)$ is an even function of each of the a's and b's and that both the real and imaginary part of $c_j(\boldsymbol{\xi}) = c_j(a_1,b_1,\ldots,a_D,b_D)$ are odd functions of each of the a's and b's. Exploiting these symmetries, it is easy to see that

$$E[c_i^*]=E[c_j]=E[c_ic_j]=E[c_i^*c_j^*]=0, \tag{2}$$

$$E[c_i^*c_j]=\delta_{i,j}E[|c_i|^2]=\delta_{i,j}E[|c|^2], \tag{3}$$

$$\begin{aligned}
E[c_i^*c_jc_kc_l^*] &= (1-\delta_{i,l})[\delta_{i,j}\delta_{k,l}+\delta_{i,k}\delta_{j,l}]E[|c_i|^2|c_l|^2] \\
&\quad + \delta_{i,l}\delta_{i,j}\delta_{k,l}E[|c_i|^4] \\
&= (1-\delta_{i,l})[\delta_{i,j}\delta_{k,l}+\delta_{i,k}\delta_{j,l}]E[|c|^2|\widehat{c}|^2] \\
&\quad + \delta_{i,l}\delta_{i,j}\delta_{k,l}E[|c|^4], \tag{4}
\end{aligned}$$

where we used the symmetries of $p(\boldsymbol{\xi})$ to drop subscripts in the argument of $E[.]$ and introduced $\widehat{c}$ to keep track of the fact that $|c|^2$ and $|\widehat{c}|^2$ correspond to two different random variables ξ_j and ξ_k with $j\neq k$.

In this paper, we limit the discussion to the random states listed in Table I. We refer to states generated by the methods implied by Case A and B as Gaussian random states and those generated by the method implied by Case C as random phase states.[79] Salient features of these random states are that in practice they are easy to generate and that exact results for their statistical properties can be derived. In Appendix D, we present two different algorithms for Case A and show that the Gaussian random state is drawn from a uniform distribution on the D-dimensional sphere (Haar measure). The Gaussian random state generated by Case B does not have this property by itself; however, as we normalize the state vector in most applications of quantum theory, Case B is very similar to Case A. Generating the random phase state is close to trivial.

In the following, we first prove general results using the symmetries of the probability density only. Then, if no further progress can be made, we use Table I to obtain explicit expressions for the Cases A, B, and C. Using Eq. (3), we find

$$E[\langle\Phi|X|\Phi\rangle] = \sum_{i,j=1}^{D} E[c_i^*c_j]\langle i|X|j\rangle = E[|c|^2]\,\mathbf{Tr}\,X. \tag{5}$$

Evaluating Eq. (5) for the special case $X=\mathbb{1}$ yields $E[|c|^2]=E[\langle\Phi|\Phi\rangle]/D$. Using this to eliminate $E[|c|^2]$, we have

$$\mathbf{Tr}\,X = \frac{DE[\langle\Phi|X|\Phi\rangle]}{E[\langle\Phi|\Phi\rangle]}, \tag{6}$$

which shows that the $\mathcal{O}(D^2)$ computational effort of evaluating the trace (see above) is replaced by the $\mathcal{O}(D)$

effort of evaluating the matrix elements in Eq. (6) and the estimation of the expectation values $E[.]$.

A further, significant reduction of work is possible if we can obtain accurate estimates of the averages in Eq. (6) by using only one "realization" of the random vector $\boldsymbol{\xi}$. Therefore, in practice, we generate a random vector $\boldsymbol{\xi}$, construct the pure state $|\Phi\rangle$, compute $D\langle\Phi|X|\Phi\rangle/\langle\Phi|\Phi\rangle$ and expect that

$$\mathbf{Tr}\,X \approx \frac{D\langle\Phi|X|\Phi\rangle}{\langle\Phi|\Phi\rangle}. \tag{7}$$

Markov's lemma[80] (see below) tells us that if the variance of $D\langle\Phi|X|\Phi\rangle/\langle\Phi|\Phi\rangle$ vanishes, the probability that $D\langle\Phi|X|\Phi\rangle/\langle\Phi|\Phi\rangle$ is equal to $\mathbf{Tr}\,X$ approaches one. Therefore, in order to assess the usefulness of the "one random state" approximation, we have to compute the variance of $D\langle\Phi|X|\Phi\rangle/\langle\Phi|\Phi\rangle$.

In Appendix A, we derive the exact expressions for the average and the variance of the right-hand side of Eq. (7). These expressions are listed in the seventh and eight column of Table I, respectively. From column seven, it follows directly that on average $D\langle\Phi|X|\Phi\rangle/\langle\Phi|\Phi\rangle$ is equal to $\mathbf{Tr}\,X$. Note that in Case A and B, the variance does not depend on the choice of the basis, whereas in Case C, it does. Therefore, in Case C, a proper choice (depending on X) of the basis may help to reduce the variance.[79]

At this stage of the discussion, X can be any matrix with entries of any size. In order to have a numerically meaningful measure of the statistical fluctuations we define the relative variance rVar by

$$
\begin{aligned}
\mathrm{rVar}\left[\frac{D\langle\Phi|X|\Phi\rangle}{\langle\Phi|\Phi\rangle}\right] &= \frac{\mathrm{Var}\left[\dfrac{D\langle\Phi|X|\Phi\rangle}{\langle\Phi|\Phi\rangle}\right]}{\left(\mathrm{E}\left[\dfrac{D\langle\Phi|X|\Phi\rangle}{\langle\Phi|\Phi\rangle}\right]\right)^2} \\
&= \frac{\mathrm{Var}\left[\dfrac{D\langle\Phi|X|\Phi\rangle}{\langle\Phi|\Phi\rangle}\right]}{|\mathbf{Tr}\,X|^2}.
\end{aligned} \tag{8}
$$

Clearly, Eq. (8) only makes sense if $E[X]$ is not a small number. If it is, we should take $\mathrm{Var}[X]$ as a measure for the absolute statistical error. Let us consider the physically relevant case where $X = X^\dagger$ with real eigenvalues λ_j, $j = 1, \ldots, D$. Then, for Cases A and B we have

$$
\begin{aligned}
\mathrm{rVar}\left[\frac{D\langle\Phi|X|\Phi\rangle}{\langle\Phi|\Phi\rangle}\right] &= \frac{D\,\mathbf{Tr}\,XX^\dagger - |\mathbf{Tr}\,X|^2}{(D+1)|\mathbf{Tr}\,X|^2} \\
&= \frac{D^{-1}\displaystyle\sum_{j=1}^{D}\lambda_j^2 - \left(D^{-1}\displaystyle\sum_{j=1}^{D}\lambda_j\right)^2}{(D+1)\left(D^{-1}\displaystyle\sum_{j=1}^{D}\lambda_j\right)^2} \\
&= \frac{1}{(D+1)}\frac{\overline{\lambda^2} - \overline{\lambda}^2}{\overline{\lambda}^2},
\end{aligned} \tag{9}
$$

and for Case C we have

$$\mathrm{rVar}\left[\frac{D\langle\Phi|X|\Phi\rangle}{\langle\Phi|\Phi\rangle}\right] \leq \frac{\mathbf{Tr}\,XX^\dagger}{|\mathbf{Tr}\,X|^2}$$

$$
\begin{aligned}
&= \frac{D^{-1}\displaystyle\sum_{j=1}^{D}\lambda_j^2}{D\left(D^{-1}\displaystyle\sum_{j=1}^{D}\lambda_j\right)^2} \\
&= \frac{1}{D}\frac{\overline{\lambda^2}}{\overline{\lambda}^2},
\end{aligned} \tag{10}
$$

where $\overline{\lambda^k} \equiv D^{-1}\sum_{j=1}^{D}\lambda_j^k$ denotes the average with respect to the index j. The last factor in Eq. (9) is the relative mean square deviation of the eigenvalues of X. If we assume that this number does not increase faster than $D + 1$, a reasonable assumption for physically relevant problems, then $\mathrm{rVar}[D\langle\Phi|X|\Phi\rangle/\langle\Phi|\Phi\rangle]$ vanishes as $D \to \infty$. Similarly, if $\overline{\lambda^2}/\overline{\lambda}^2$ does not increase faster than D, then the left-hand side of Eq. (10) vanishes as $D \to \infty$.

The vanishing of the relative variances as $D \to \infty$ is the key idea behind any successful, practical application of the random state technology. Indeed, once it has been shown that relative variances vanish as $D \to \infty$, for large D (which is a very large number in most cases of interest) it suffices to use only one realization of a random state $|\Phi\rangle$ to obtain an accurate estimate of $\mathbf{Tr}\,X$, implying that the computational burden has been reduced by a factor D. This can be shown by the following argument. Appealing to a variant of Markov's inequality (Chebyshev's inequality)[80]

$$P(|A - \langle A\rangle| \geq \epsilon) \leq \frac{\mathrm{Var}[A]}{\epsilon^2}, \quad \epsilon > 0, \tag{11}$$

we see that if the variance of A is sufficiently small, the probability that the value of this random variable deviates from its average value will also be small. We use this basic result in the form

$$
\begin{aligned}
P&\left(\left|\frac{D\langle\Phi|X|\Phi\rangle}{\langle\Phi|\Phi\rangle} - \mathbf{Tr}\,X\right| \geq \epsilon|\mathbf{Tr}\,X|\right) \\
&\leq \frac{\mathrm{Var}[D\langle\Phi|X|\Phi\rangle/\langle\Phi|\Phi\rangle]}{\epsilon^2|\mathbf{Tr}\,X|^2} \\
&= \epsilon^{-2}\,\mathrm{rVar}\left[\frac{D\langle\Phi|X|\Phi\rangle}{\langle\Phi|\Phi\rangle}\right], \quad \epsilon > 0.
\end{aligned} \tag{12}
$$

This shows that if $(D\,\mathbf{Tr}\,XX^\dagger - |\mathbf{Tr}\,X|)/|\mathbf{Tr}\,X|^2$ (Case A and B) or $\mathbf{Tr}\,XX^\dagger/|\mathbf{Tr}\,X|^2$ (Case C) do not increase faster than D, the probability that the right-hand side of Eq. (7), computed from one realization of a random state, deviates from its expectation value $\mathbf{Tr}\,X$, vanishes as D increases.

It is instructive to estimate the expected error of using Eq. (7) in a different, somewhat more general manner. Let us multiply both sides of Eq. (7) by $\langle\Phi|\Phi\rangle$ and define

$$\widetilde{X} = D\langle\Phi|X|\Phi\rangle - \langle\Phi|\Phi\rangle\,\mathbf{Tr}\,X. \tag{13}$$

Obviously, Eq. (6) implies $E[\widetilde{X}] = 0$. The mean square deviation is given by

$$\mathrm{Var}[\widetilde{X}] = \mathrm{E}[|D\langle\Phi|X|\Phi\rangle - \langle\Phi|\Phi\rangle\,\mathbf{Tr}\,X|^2]. \tag{14}$$

By straightforward application of Eqs. (2)–(4) we find

$$\mathrm{E}[|\langle\Phi|X|\Phi\rangle|^2]$$

$$= \sum_{i,j=1}^{D}\sum_{k,l=1}^{D}\mathrm{E}[c_i^* c_j c_k c_l^*]\langle i|X|j\rangle\langle k|X|l\rangle^*$$

J. Phys. Soc. Jpn.
Downloaded from journals.jps.jp by Forschungszentrum Juelich GmbH on 01/19/21

J. Phys. Soc. Jpn. **90**, 012001 (2021)   **Invited Review Papers**   F. Jin et al.

$$= \mathrm{E}[|c|^2|\widehat{c}|^2] \sum_{i,j=1}^{D} \sum_{k,l=1}^{D} [\delta_{i,j}\delta_{k,l} + \delta_{i,k}\delta_{j,l}] \langle i|X|j\rangle \langle k|X|l\rangle^*$$

$$+ (\mathrm{E}[|c|^4] - 2\mathrm{E}[|c|^2|\widehat{c}|^2]) \sum_{i=1}^{D} \langle i|X|i\rangle \langle i|X|i\rangle^*$$

$$= \mathrm{E}[|c|^2|\widehat{c}|^2](\mathbf{Tr}\, XX^\dagger + |\mathbf{Tr}\, X|^2)$$

$$+ (\mathrm{E}[|c|^4] - 2\mathrm{E}[|c|^2|\widehat{c}|^2]) \sum_{i=1}^{D} |\langle i|X|i\rangle|^2 \qquad (15)$$

$$\mathrm{E}[\langle \Phi|X|\Phi\rangle \langle \Phi|\Phi\rangle^*]$$

$$= \sum_{i,j=1}^{D} \sum_{k,l=1}^{D} \mathrm{E}[c_i^* c_j c_l c_k^*] \langle i|X|j\rangle \langle k|l\rangle^*$$

$$= \{(D-1)\mathrm{E}[|c|^2|\widehat{c}|^2] + \mathrm{E}[|c|^4]\}\, \mathbf{Tr}\, X, \qquad (16)$$

and therefore

$$\mathrm{Var}[\widetilde{X}] = D^2 \mathrm{E}[|\langle \Phi|X|\Phi\rangle|^2] + \mathrm{E}[|\langle \Phi|\Phi\rangle|^2]|\mathbf{Tr}\, X|^2$$

$$- 2D\Re(\mathrm{E}[\langle \Phi|X|\Phi\rangle \langle \Phi|\Phi\rangle^*]\, \mathbf{Tr}\, X)$$

$$= D^2 \mathrm{E}[|c|^2|\widehat{c}|^2](\mathbf{Tr}\, XX^\dagger + |\mathbf{Tr}\, X|^2)$$

$$+ D^2(\mathrm{E}[|c|^4] - 2\mathrm{E}[|c|^2|\widehat{c}|^2]) \sum_{i=1}^{D} |\langle i|X|i\rangle|^2$$

$$+ D(D-1)\mathrm{E}[|c|^2|\widehat{c}|^2]|\mathbf{Tr}\, X|^2 + D\mathrm{E}[|c|^4]|\mathbf{Tr}\, X|^2$$

$$- 2D(D-1)\mathrm{E}[|c|^2|\widehat{c}|^2]|\mathbf{Tr}\, X|^2$$

$$- 2D\mathrm{E}[|c|^4]|\mathbf{Tr}\, X|^2$$

$$= D^2 \mathrm{E}[|c|^2|\widehat{c}|^2]\, \mathbf{Tr}\, XX^\dagger$$

$$+ D(\mathrm{E}[|c|^2|\widehat{c}|^2] - \mathrm{E}[|c|^4])|\mathbf{Tr}\, X|^2$$

$$+ D^2(\mathrm{E}[|c|^4] - 2\mathrm{E}[|c|^2|\widehat{c}|^2]) \sum_{i=1}^{D} |\langle i|X|i\rangle|^2$$

$$= \begin{cases} (D\,\mathbf{Tr}\, XX^\dagger - |\mathbf{Tr}\, X|^2)/(D+1), & \text{Case A} \\ (D\,\mathbf{Tr}\, XX^\dagger - |\mathbf{Tr}\, X|^2)/D, & \text{Case B} \\ \mathbf{Tr}\, XX^\dagger - \sum_{i=1}^{D} |\langle i|X|i\rangle|^2, & \text{Case C} \end{cases}, \qquad (17)$$

where the last line follows by substituting the corresponding values of $\mathrm{E}[|c|^2]$, $\mathrm{E}[|c|^2|\widehat{c}|^2]$, and $\mathrm{E}[|c|^4]$, given in Table I. Except for Case B where there is an irrelevant difference ($D+1$ instead of D) in the denominator, the final expressions given by Eq. (17) agree with the exact expressions for the variance given in Table I. Therefore, the expressions in Eq. (17) for the mean square deviation $\mathrm{Var}[\widetilde{X}]$ lead to the same conclusion as the one obtained for the exact expressions listed in Table I.

3. Application: Density of States

To our knowledge, the first application of the random state technology was for the calculation of the DOS of an alloy.[29] In this section, we present a discussion of this kind of application from a more general perspective. An example implementation that can be used to reproduce the figures presented in this section and to compute the DOS of many other lattices is available online at Ref. 81. In the following, we set $\hbar = 1$ for convenience.

Given a state $|\Psi\rangle$, the local density of states $\mathrm{LDOS}_\Psi(\omega)$ of a system described by a Hamiltonian H is defined by

$$\mathrm{LDOS}_\Psi(\omega) = \langle \Psi|\delta(\omega - H)|\Psi\rangle$$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{i\omega t} \langle \Psi|e^{-itH}|\Psi\rangle \, dt$$

$$\approx \frac{1}{2\pi} \int_{-T}^{+T} e^{i\omega t} \langle \Psi|e^{-itH}|\Psi\rangle \, dt, \qquad (18)$$

where $[-T, T]$ denotes the relevant time interval (see below for a criterion on how to choose T given H). In practice, T is limited by the amount of computational work it takes to calculate the matrix element in Eq. (18). If $|\Psi\rangle$ is taken to be a random state $|\Phi\rangle$, applying Eq. (5) yields,

$$\mathrm{E}[\mathrm{LDOS}_\Psi(\omega)] = \frac{1}{2\pi} \mathrm{E}[|c|^2] \int_{-\infty}^{+\infty} e^{i\omega t}\, \mathbf{Tr}\, e^{-itH} \, dt$$

$$= \mathrm{E}[|c|^2] \sum_{n=1}^{D} \delta(\omega - E_n)$$

$$= D\mathrm{E}[|c|^2]\mathrm{DOS}(\omega), \qquad (19)$$

where $\mathrm{DOS}(\omega) = \sum_{n=1}^{D} \delta(\omega - E_n)/D$ is, by definition, the density of states. For numerical work, problems with large D are demanding, but by using the random state technology we can reduce the computational burden by writing

$$\mathrm{DOS}(\omega) \approx \frac{1}{2\pi D\mathrm{E}[|c|^2]} \int_{-T}^{+T} e^{i\omega t} \langle \Phi|e^{-itH}|\Phi\rangle \, dt. \qquad (20)$$

In practice, we perform the integration over time in Eq. (18) by means of the Discrete Fourier Transform (DFT). To this end we rewrite Eq. (18) as

$$\mathrm{LDOS}_\Psi(k\pi/T) = \frac{T}{2\pi N} \sum_{j=-N}^{N-1} e^{\pi ijk/N} \langle \Psi|e^{-ij\tau H}|\Psi\rangle$$

$$= \frac{T}{2\pi N} \mathrm{DFT}[\langle \Psi|e^{-ij\tau H}|\Psi\rangle], \qquad (21)$$

$$\mathrm{DOS}(k\pi/T) = \frac{T}{2\pi N D\mathrm{E}[|c|^2]} \sum_{j=-N}^{N-1} e^{\pi ijk/N} \langle \Phi|e^{-ij\tau H}|\Phi\rangle$$

$$= \frac{T}{2\pi N D\mathrm{E}[|c|^2]} \mathrm{DFT}[\langle \Phi|e^{-ij\tau H}|\Phi\rangle], \qquad (22)$$

where $\tau = T/N$ is the time step at which we sample the function $\langle \Psi|e^{-itH}|\Psi\rangle$. Accordingly, we obtain $\mathrm{DOS}(\omega)$ at frequencies $\omega = k\pi/T$ for $k = -N, \dots, N-1$. As before, we see that the random state technology reduces the amount of computational work from $\mathcal{O}(D^2)$ to $\mathcal{O}(D)$.

In practice, the most efficient method to compute $\langle \Psi|e^{-itH}|\Psi\rangle$ for $t = \tau, 2\tau, \dots, N\tau$ is to solve the time-dependent Schrödinger equation (TDSE)

$$i\frac{\partial}{\partial t}|\phi\rangle = H|\phi\rangle, \qquad (23)$$

using a suitable method such as the Suzuki–Trotter product formula algorithm,[68,69] the Chebyshev polynomial algorithm,[70–72] a Lanczos-iteration based method,[73,74] or the fourth-order Runge–Kutta method.[27,75] The values of $\langle \Psi|e^{-itH}|\Psi\rangle$ for $t = -\tau, -2\tau, \dots, -N\tau$ are obtained by using $\langle \Psi|e^{+itH}|\Psi\rangle = \langle \Psi|e^{-itH}|\Psi\rangle^*$.

In numerical work, the matrix H is bounded. Denoting by $\|H\|$ the maximum absolute value of the eigenvalues of H (i.e., the spectral norm or 2-norm of H), the function $\mathrm{LDOS}_\Psi(\omega) = 0$ if $|\omega| > \|H\|$, that is this function is band
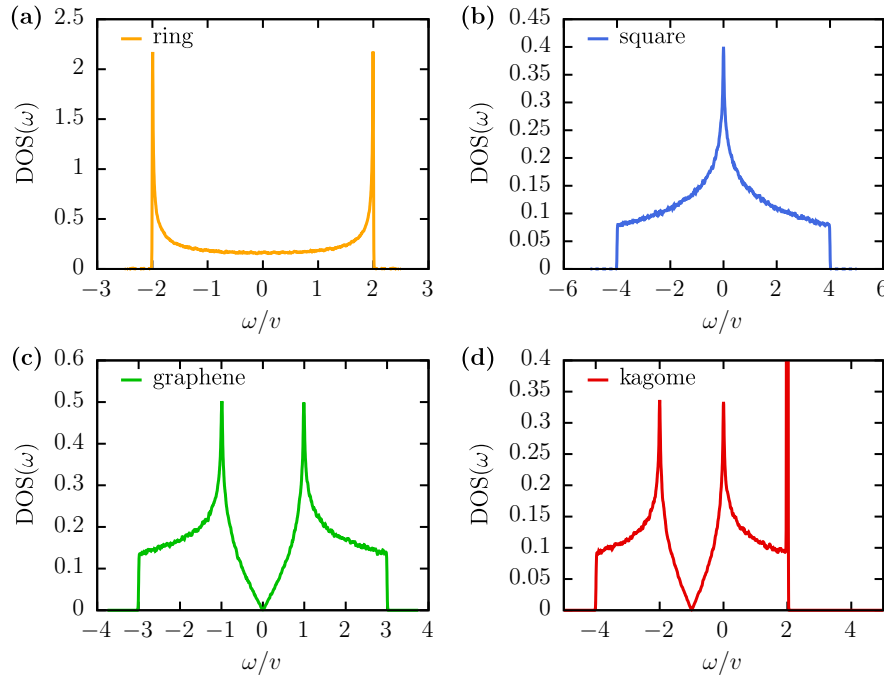
Fig. 1. (Color online) DOS as a function of the frequency in units of the hopping constant $v_{mn} = v$ [see Eq. (24)]. (a) One-dimensional chain with $D = 1000000$ sites; (b) two-dimensional square lattice with $D = 1000000$ sites; (c) graphene lattice with $D = 1080000$ sites; (d) kagome lattice with $D = 1008200$ sites. All lattices have periodic boundary conditions. In all cases, $DOS(\omega)$ has been obtained with the random state technology Eq. (22), where $N = 1000$ and $T = N\tau$ with $\tau = 0.8\pi/\|H\|_1$, and where $\|H\|_1 = 2v$ for the ring, $\|H\|_1 = 4v$ for the square and kagome lattices, and $\|H\|_1 = 3v$ for the graphene lattice. The TDSE (23) is solved by using the second-order real-space product-formula algorithm[68] with a time step $\tau/5$ [see Eq. (25)]. The source code to reproduce this figure can be downloaded at Ref. 81.

limited. Then Nyquist's sampling theorem tells us that we should sample with a time step $\tau < \pi/\|H\|$ in order to cover the full range of eigenvalues and avoid aliasing.[24] As $\tau = T/N$, this means that the time interval should be chosen such that $T < N\pi/\|H\|$. Eigenvalues that differ by less than π/T cannot be resolved.

In summary, the number of samples $2N$ determines the resolving power of the method, and the time interval $2T$ has to satisfy $T < N\pi/\|H\|$ to cover the full range of eigenvalues. For many problems of interest, $\|H\|$ is not known, but it can be replaced by a bound such as $\|H\| \leq \|H\|_1 = \max_{1 \leq j \leq D} \sum_{i=1}^{D} |H_{i,j}|$ (i.e., the maximum absolute column sum, which is closely related to the concept of the Gershgorin disk[82]) which is easy to compute.

We illustrate the application of the random state-based technique by computing the DOS for a single particle hopping on a one-dimensional ring, a square, a graphene, and a kagome lattice. The Hamiltonian is given by

$$H = \sum_{\langle m,n \rangle} v_{mn}(a_m^+ a_n + \text{h.c.}) + \sum_m w_m a_m^+ a_m, \qquad (24)$$

where a_m^+ (a_n) is the fermion creation (annihilation) operator of a particle at site m (n), v_{mn} are the hopping integrals, and w_m is an on-site potential. The sum over $\langle m, n \rangle$ is over all nearest-neighbor bonds of the lattice with $m < n$. Note that because of the restriction to a one-particle problem, it does not matter whether we use fermion or boson operators.

Being a one-particle problem, we can compute all physical properties if we can diagonalize the quadratic form defined by the matrix $\mathbf{V}$ with elements $v_{mn} + \delta_{mn} w_m$. In some cases (see below), $\mathbf{V}$ can be diagonalized analytically for any size or shape of the lattice. If that is not possible, we can diagonalize $\mathbf{V}$ numerically if its dimension is not too large, in practice limiting this approach to matrices with a linear dimension in the range 10000 to 100000. The random state technology can handle (much) larger matrices.[42]

The numerical procedure consists of three steps (see also the demonstration available at Ref. 81):
1. Generate a Gaussian random state $|\Phi\rangle = \sum_m c_m a_m^+ |0\rangle$ and a copy: $|\Psi\rangle = |\Phi\rangle$.
2. Use the first-order real-space Suzuki–Trotter formula[68]

$$e^{-i\tilde{\tau}H} \approx U_1(\tilde{\tau})$$
$$= \prod_{\langle m,n \rangle} \exp[-i\tilde{\tau}v_{mn}(a_m^+ a_n + \text{h.c.})]$$
$$\times \prod_m \exp[-i\tilde{\tau}w_m a_m^+ a_m], \qquad (25)$$

to construct the second-order approximation $U_2(\tilde{\tau}) = U_1^T(\tilde{\tau}/2)U_1(\tilde{\tau}/2)$. The order in which the matrix exponentials in Eq. (25) appear can be chosen freely but once chosen, this order has to be kept fixed. Solve the TDSE by repeating $|\Psi(t + \tau)\rangle \leftarrow (U_2(\tilde{\tau}))^l |\Psi(t)\rangle$, where $\tilde{\tau} = \tau/l$ and with $l \geq 1$ controlling the accuracy of the product formula approximation. During the time stepping, at each time $t = \tau, 2\tau, \ldots, N\tau$, store the values $f(t) = \langle \Phi | \Psi \rangle$ and $f(-t) = f(t)^*$.
3. Compute the DFT of $f(t)g(t)$ where $g(t)$ is a Gaussian window function.

In Fig. 1, we present the (well-known) results for the DOS of the four mentioned lattices with $v_{mn} = v$ and $w_m = 0$, as obtained by the above procedure.

For problems for which all eigenvalues of $\mathbf{V}$ are known in analytical or numerical form, such as the examples for a particle moving on one of the lattices considered above, there
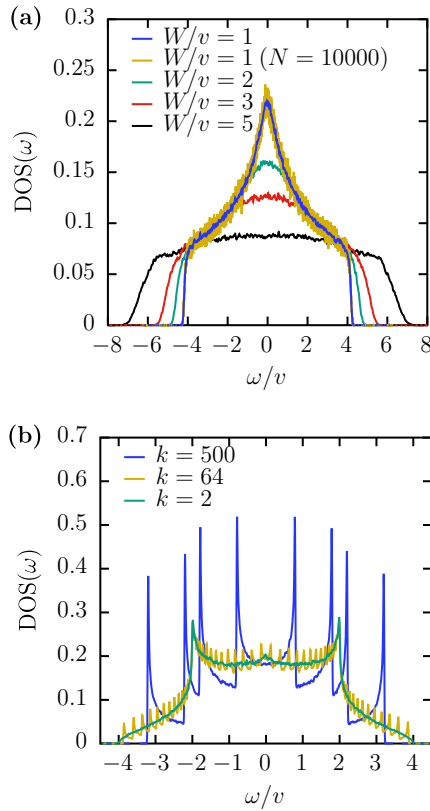
Fig. 2. (Color online) DOS for square lattices of dimension $D = L \times L = 1000000$ for which all the eigenvalues of $\mathbf{V}$ cannot be computed analytically or by numerical diagonalization. (a) Anderson model with random energies w_m drawn uniformly from $[-W, W]$; (b) sinusoidal bond model with anisotropic bond strengths $v_x = v \sin^2(\pi k x/2L)$ and $v_y = v$ where $x = 1, \ldots, L - 1$ enumerates the horizontal bonds of the square lattice (the case $k = 2$ corresponds to strong bonds in the center and weak bonds at the boundaries and the case $k = 500$ corresponds to alternating weak and strong bonds with a period of four sites). In all cases, $\text{DOS}(\omega)$ has been obtained with the random state technology Eq. (22), where $N = 1000$ (or $N = 10000$ if stated in the legend) and $T = N\tau$ with $\tau = 0.8\pi/\|H\|_1$, where $\|H\|_1 = 4v + |W|$ for the Anderson model and $\|H\|_1 = 4v$ for the sinusoidal bond model. The TDSE (23) is solved by using the second-order real-space product-formula algorithm[68] with a time step $\tau/5$. The source code to reproduce this figure can be downloaded at Ref. 81.

exist much more efficient methods to compute the (L)DOS than the one based on solving the TDSE. For instance, simply counting the number of times that the energies fall in bins $[E, E + \Delta E]$ gives the DOS in the form of a histogram. However, the method based on random states in combination with the solution of the TDSE excels when the calculation of **all** the eigenvalues of H is prohibitive. We illustrate this point

by Fig. 2 where we present DOS results for the two-dimensional Anderson model of localization and a square lattice with rather exotic anisotropic hopping integrals.

4. Application: Quantum Statistical Physics

An important class of problems for which the random state technology can be put to good use is the calculation of thermal equilibrium averages of observables of a quantum system, described in terms of its Hamiltonian $H = H^\dagger$. If $Y = Y^\dagger$ represents the matrix of such an observable, the task is to compute

$$\langle Y \rangle = \frac{\mathbf{Tr}\, e^{-\beta H} Y}{\mathbf{Tr}\, e^{-\beta H}} = \frac{\mathbf{Tr}\, e^{-\beta H/2} Y e^{-\beta H/2}}{\mathbf{Tr}\, e^{-\beta H}}, \qquad (26)$$

where β denotes the inverse temperature in units of $1/k_B$, and we used the invariance of the trace under cyclic permutation ($\mathbf{Tr}\, AB = \mathbf{Tr}\, BA$) to bring $\langle Y \rangle$ into a form that is most suited for the application of the random state technology. According to the general recipe laid out in Sect. 2, the idea is to replace the calculation of the traces by the calculation of only one matrix element.

The first step in the procedure is to generate a random state $|\Phi\rangle$, by using for instance one of the methods given in Appendix D. Next, we compute the so-called random thermal state defined by

$$|\Phi_\beta\rangle \equiv e^{-\beta H/2} |\Phi\rangle, \qquad (27)$$

whereby it is implicitly understood that one has an efficient and accurate algorithm to compute $e^{-\beta H/2} |\Phi\rangle$. In practice, the Chebyshev polynomial representation of $e^{-\beta H/2}$ can be used to compute, with close to optimal efficiency, $e^{-\beta H/2} |\Phi\rangle$ to almost machine precision.[38,70,71,83] The final computational step is then to estimate the thermal equilibrium expectation value according to

$$\langle Y \rangle = \frac{\mathbf{Tr}\, e^{-\beta H/2} Y e^{-\beta H/2}}{\mathbf{Tr}\, e^{-\beta H}} \approx \frac{\langle \Phi_\beta | Y | \Phi_\beta \rangle}{\langle \Phi_\beta | \Phi_\beta \rangle} = \frac{\langle \Phi | X | \Phi \rangle}{\langle \Phi | Z | \Phi \rangle}, \quad (28)$$

where $X = e^{-\beta H/2} Y e^{-\beta H/2}$ and $Z = e^{-\beta H}$. As before, in order to show that the replacement of the traces by single matrix elements makes sense, one has to show that the variance of the right-hand side of Eq. (28) is small. Because of the presence of Z in the denominator, we are unable to calculate the average or the variance of $\langle \Phi | X | \Phi \rangle / \langle \Phi | Z | \Phi \rangle$ exactly, except if $\beta = 0$ in which case the problem reduces to the one treated in Sect. 2. Therefore, we have to resort to the approximate treatment based on Eqs. (B·1) and (B·2). To calculate the first three terms in Eqs. (B·1) and (B·2), we only need

$$\mathrm{E}[\langle \Phi | X | \Phi \rangle] = \mathrm{E}[|c|^2] \mathbf{Tr}\, X = \mathrm{E}[|c|^2] \mathbf{Tr}\, e^{-\beta H} Y = \mathrm{E}[|c|^2] \mathbf{Tr}\, ZY$$

$$\mathrm{E}[\langle \Phi | Z | \Phi \rangle] = \mathrm{E}[|c|^2] \mathbf{Tr}\, Z$$

$$\mathrm{E}[\langle \Phi | X | \Phi \rangle^2] = \mathrm{E}[|c|^2 |\widehat{c}|^2](\mathbf{Tr}(ZY)^2 + (\mathbf{Tr}\, ZY)^2) + (\mathrm{E}[|c|^4] - 2\mathrm{E}[|c|^2 |\widehat{c}|^2]) \sum_{i=1}^{D} \langle i | X | i \rangle^2$$

$$\mathrm{E}[\langle \Phi | Z | \Phi \rangle^2] = \mathrm{E}[|c|^2 |\widehat{c}|^2](\mathbf{Tr}\, Z^2 + (\mathbf{Tr}\, Z)^2) + (\mathrm{E}[|c|^4] - 2\mathrm{E}[|c|^2 |\widehat{c}|^2]) \sum_{i=1}^{D} \langle i | Z | i \rangle^2$$

$$\mathrm{E}[\langle \Phi | X | \Phi \rangle \langle \Phi | Z | \Phi \rangle] = \mathrm{E}[|c|^2 |\widehat{c}|^2](\mathbf{Tr}\, ZY\, \mathbf{Tr}\, Z + \mathbf{Tr}\, Z^2 Y) + (\mathrm{E}[|c|^4] - 2\mathrm{E}[|c|^2 |\widehat{c}|^2]) \sum_{i=1}^{D} \langle i | X | i \rangle \langle i | Z | i \rangle. \qquad (29)$$

J. Phys. Soc. Jpn.
Downloaded from journals.jps.jp by Forschungszentrum Juelich GmbH on 01/19/21

J. Phys. Soc. Jpn. **90**, 012001 (2021)   **Invited Review Papers**   F. Jin et al.

It is now straightforward to use the expressions for the moments given in Table I and obtain the expressions for the average and variance of $\langle\Phi|X|\Phi\rangle/\langle\Phi|Z|\Phi\rangle$. For example, in Case B we find

$$
\begin{aligned}
\mathrm{E}\left[\frac{\langle\Phi|X|\Phi\rangle}{\langle\Phi|Z|\Phi\rangle}\right] &\approx \frac{\mathbf{Tr}\,ZY}{\mathbf{Tr}\,Z} - \frac{\mathbf{Tr}\,Z^2Y}{(\mathbf{Tr}\,Z)^2} + \frac{\mathbf{Tr}\,ZY}{\mathbf{Tr}\,Z}\frac{\mathbf{Tr}\,Z^2}{(\mathbf{Tr}\,Z)^2} \\
&= \frac{\mathbf{Tr}\,ZY}{\mathbf{Tr}\,Z} + \frac{\mathbf{Tr}\,Z^2}{(\mathbf{Tr}\,Z)^2}\left\{\frac{\mathbf{Tr}\,ZY}{\mathbf{Tr}\,Z} - \frac{\mathbf{Tr}\,Z^2Y}{\mathbf{Tr}\,Z^2}\right\},
\end{aligned}
\tag{30}
$$

and

$$
\mathrm{Var}\left[\frac{\langle\Phi|X|\Phi\rangle}{\langle\Phi|Z|\Phi\rangle}\right] \approx \frac{\mathbf{Tr}\,Z^2}{(\mathbf{Tr}\,Z)^2}\left\{\frac{\mathbf{Tr}(ZY)^2}{\mathbf{Tr}\,Z^2} - 2\frac{\mathbf{Tr}\,ZY}{\mathbf{Tr}\,Z}\frac{\mathbf{Tr}\,Z^2Y}{\mathbf{Tr}\,Z^2} + \left(\frac{\mathbf{Tr}\,ZY}{\mathbf{Tr}\,Z}\right)^2\right\}.
\tag{31}
$$

The absolute values of the contributions in the curly brackets are readily shown to be bounded by $2\|Y\|$ [Eq. (30)] and $4\|Y\|^2$ [Eq. (31)]. In practice, we are only interested in observables Y for which $\|Y\| = \mathcal{O}(\log D)$. Therefore, the magnitude of correction [second term of Eq. (30)] to and the variance of the thermal average of Y are primarily determined by the factor $\mathbf{Tr}\,Z^2/(\mathbf{Tr}\,Z)^2 = 1/D$. As $\mathbf{Tr}\,Z = \mathbf{Tr}\,e^{-\beta H} = e^{-\beta F(\beta)}$ where $F(\beta)$ is the free energy, we have, in general,

$$
\frac{1}{D} \leq \frac{\mathbf{Tr}\,Z^2}{(\mathbf{Tr}\,Z)^2} = e^{-2\beta(F(2\beta)-F(\beta))} \leq 1,
\tag{32}
$$

where the first inequality follows from applying the Schwarz inequality to the inner product $\mathbf{Tr}\,A^\dagger B$. The second inequality follows by writing $\mathbf{Tr}\,Z = \sum_{i=1}^{D}\exp(-\beta E_i)$, $\mathbf{Tr}\,Z^2 = \sum_{i=1}^{D}\exp(-2\beta E_i)$ and noting that

$$
\frac{\mathbf{Tr}\,Z^2}{(\mathbf{Tr}\,Z)^2} = \frac{1}{1 + \left(\sum_{i\neq j}^{D}\exp(-\beta(E_i+E_j))\right)\Big/\left(\sum_{i=1}^{D}\exp(-2\beta E_i)\right)} \leq 1.
\tag{33}
$$

From the left-hand side of Eqs. (32) and (33), it follows immediately that $F(2\beta) \geq F(\beta)$. This is in concert with the second law of thermodynamics. Furthermore, since the free energy is an extensive quantity, $\mathbf{Tr}\,Z^2/(\mathbf{Tr}\,Z)^2$ vanishes exponentially with increasing system size. If we assume that $\|Y\|$ is at most proportional to the system size then, using the fact that the terms in the curly brackets are bounded and their prefactors vanish exponentially with increasing system size, the correction to the average and the variance itself vanish in the same manner.

As $F(2\beta) \approx F(\beta) \approx E_0$ for $\beta \to \infty$, Eq. (32) suggests that the efficiency of the random state approach is lost when $\beta \to \infty$. Indeed, when $\beta \to \infty$, the random thermal state turns into the ground state $|0\rangle$ of H, that is $\lim_{\beta\to\infty}|\Phi_\beta\rangle \to |0\rangle$, but we cannot recommend calculating the ground state by projection with $e^{-\beta H/2}$ because for large β, the efficiency of this projection is inferior to that of the Lanczos method, for instance.

It is instructive to illustrate these general features through a very simple example. To this end, we compute $e^{-2\beta(F(2\beta)-F(\beta))} = \mathbf{Tr}\,Z^2/(\mathbf{Tr}\,Z)^2$ for a system of N non-interacting spins, described by the Hamiltonian $H = -\Omega\sum_{i=1}^{N}\sigma^z$ where σ^z is the z-component of the Pauli-spin matrices, having eigenvalues ± 1. A simple calculation yields

$$
\frac{\mathbf{Tr}\,Z^2}{(\mathbf{Tr}\,Z)^2} = \left(\frac{1 + \tanh^2\beta\Omega}{2}\right)^N,
\tag{34}
$$

which shows that $e^{-2\beta(F(2\beta)-F(\beta))}$ smoothly increases from $D^{-1} = 2^{-N}$ to one when β increases from zero to infinity. For $\beta < \infty$, Eq. (34) vanishes exponentially with increasing N.

In summary, the factor $\mathbf{Tr}\,Z^2/(\mathbf{Tr}\,Z)^2$ determines the rate at which the variance Eq. (31) vanishes with the system size but this factor can, depending on the temperature $1/\beta$, vary from $1/D$ to close to one. In the latter case, the random state technology loses its efficiency.

4.1 Specific heat

In this section, we use the random state technology to calculate the specific heat of frustrated and non-frustrated spin-1/2, nearest-neighbor, antiferromagnetic Heisenberg models. The Hamiltonian is defined by

$$
H = -J\sum_{\langle i,j\rangle} S_i \cdot S_j,
\tag{35}
$$

where $S_i = (S_i^x, S_i^y, S_i^z)$ is the spin-1/2 operator at the i-th site, $\langle i,j\rangle$ refers to the nearest-neighbor sites on the lattice, and $J = -1$ is the antiferromagnetic coupling. The specific heat is calculated from the fluctuation of the energy, i.e., as $C = \frac{1}{N}\beta^2(\langle H\rangle^2 - \langle H^2\rangle)$, where N is the number of lattice sites.

Numerically, there are only two important operations involved. The first is the preparation of the random thermal state [projection on a random state, Eq. (27)] and the second is the operation $H|\Phi_\beta\rangle$. There are several ways to do the projection, e.g., by the power method,[84] a product formula algorithm,[68,69] or the Chebyshev polynomial algorithm.[70–72] These methods only need storage for a few state vectors, and the numerical errors are well under control. Another method, widely used in the solid-state physics community, is the finite-temperature Lanczos method.[73] This method uses the eigenvalues and eigenstates obtained from the standard Lanczos method to calculate the matrix element of the operator Y. It requires storage of wave functions proportional to the Lanczos iteration order M and reproduces the results of the high-temperature expansion up to order M. According to the theory given in the previous section, the errors due to the use of the random state are bounded (simply replace Y by H or H^2) and vanish as $D \to \infty$.

Figure 3 shows the results for the specific heat for three different lattices, namely, the square, triangular, and kagome lattice. For all cases, the lattice size $N = 36$ and periodic boundaries are used. For both triangular and kagome lattices,
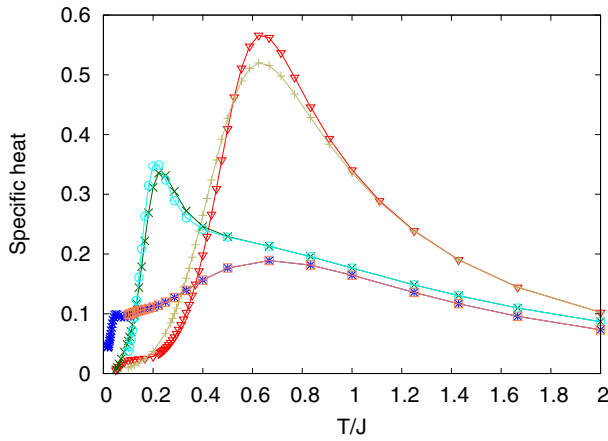
Fig. 3. (Color online) Specific heat for the square (triangles and pluses), triangular (crosses and circles), and kagome (stars and squares) lattice with $N = 36$ spins. For each lattice, we show results obtained by using two different seeds to generate the initial random state.

the shape of the lattice is a rhombus. Within the range of temperatures covered, we clearly see that for the square and triangular lattice, there is only one peak in the specific heat, located around $T/J = 0.6$ and 0.2, respectively. The specific heat for the kagome lattice shows multiple peaks. The first peak is around $T/J = 0.6$. The second peak is around $T/J = 0.1$, shoulder-like and greatly reduced compared to results obtained for system sizes less than $N = 36$ (data not shown).

Our results are consistent with those obtained from exact diagonalization,[85] finite-temperature Lanczos method (FTLM),[86,87] and the transfer-matrix quantum Monte Carlo methods.[88] Shimokawa and Kawamura used the random state technology to calculate the finite-temperature properties of the antiferromagnetic Heisenberg model on the kagome lattice up to size 36.[89] Sugiura and Shimizu developed the same method and demonstrated its use by calculating the specific heat on the same model up to size 30.[20] At even lower temperatures, a third peak appears, see Ref. 86, in which the finite-temperature Lanczos method was used to calculate the thermal properties for a kagome lattice with up to $N = 42$ spins.

As the temperature decreases further, the factor $\mathbf{Tr}\, Z^2/(\mathbf{Tr}\, Z)^2$ approaches one. The corresponding loss of statistical accuracy can only be compensated for by averaging over different realizations of the random states,[42] see also Appendix C. As this paper focuses on the basic principles of the random state technology, we do not pursue the aspect of averaging over different realizations of the random states here and therefore, we do not calculate the specific heat for $T/J < 0.01$. Instead, as an illustration, we also include in Fig. 3 the results obtained by using different random states, for temperatures down to $T/J = 0.1$. Figure 3 shows that the data for the kagome lattice, obtained using two different initial random states, agree with each other up to $T/J = 0.1$, while the results for the other two lattices only agree up to temperatures of about $T/J = 1$. This can be understood as follows. At sufficiently low temperatures, we have $\mathbf{Tr}\, Z^2/(\mathbf{Tr}\, Z)^2 \approx (1 + e^{-2\beta(E_1 - E_0)})/(1 + e^{-\beta(E_1 - E_0)})^2$ where $E_1 - E_0$ is the gap between the ground state and the first excited state. The gaps for square, triangular, and kagome lattice with

$N = 36$ sites are $E_1 - E_0 = 0.29$, 0.34, and 0.01, respectively. Therefore, for fixed N, the temperature at which $\mathbf{Tr}\, Z^2/(\mathbf{Tr}\, Z)^2$ approaches one (and the variance may become large) is about 30 times higher for the square and triangular lattice than for the kagome lattice. In other words, if we want to compute temperature-dependent averages with approximately the same accuracy and $\exp[-\beta(E_1 - E_0)]$ becomes small, it is necessary to average over more than one realization of the random state.

For extremely low temperatures, the specific heat can be calculated through the exact low-lying eigenvalues which can be obtained by several methods, such as Lanczos-based algorithms[90] and the Sakurai–Sugiura method.[91] For moderately low temperatures, Morita and Tohyama proposed two algorithms improved from FTLM by utilizing the exact low-lying eigenvalues and eigenvectors.[92] The first so-called replaced FTLM algorithm replaces the energies by the exact low-lying energies in the FTLM. The second so-called orthogonalized FTLM is to start the FTLM by random states orthogonal to all the exact low-lying eigenvectors. The effectiveness of these algorithms is illustrated by results for the specific heat and structure factor for Kitaev–Heisenberg models on kagome and triangular lattices for systems with up to $N = 36$ spins. The second algorithm can be easily adapted to the projection method used to calculate the specific heat in the present paper.

5. Application: Quantum Dynamics

Random state technology can be used for the calculation of the expectation value of time- or frequency dependent observables as well. In this section, we illustrate this fact by showing calculations of the current–current correlation, density–density correlations, and ESR spectrum for spin-1/2 models. For examples of the application of the random state technology to the calculations of the dynamic properties of 2D materials see Refs. 64 and 65.

5.1 Current–current correlation

In this subsection we focus on the current–current autocorrelation function defined by

$$C(t) = \frac{\mathbf{Tr}\, e^{-\beta H} j(t) j(0)}{\mathbf{Tr}\, e^{-\beta H}}, \qquad (36)$$

where j is the (problem-dependent) current and $j(t) = e^{itH} j e^{-itH}$. If we set $Y = j(t) j(0)$, it is obvious that the random state technology can be applied to calculate the current–current correlation, Eq. (31) guaranteeing that the variance of $C(t)$ vanishes exponentially with increasing system size. In numerical work, we introduce two auxiliary states

$$|\phi(t)\rangle = e^{-iHt} e^{-\beta H/2} |\Phi\rangle, \qquad (37)$$

$$|\varphi(t)\rangle = e^{-iHt} j e^{-\beta H/2} |\Phi\rangle, \qquad (38)$$

and obtain the autocorrelation function $C(t)$ by computing

$$C(t) = \frac{\langle \phi(t) | j | \varphi(t) \rangle}{\langle \Phi | e^{-\beta H} | \Phi \rangle}. \qquad (39)$$

As already mentioned, the action of the time evolution operator e^{-iHt} on a state vector can be computed by algorithms such as the fourth-order Runge–Kutta scheme,[27,75] the Suzuki–Trotter product formula algorithm,[68,69] and the Chebyshev polynomial algorithm.[70–72]

J. Phys. Soc. Jpn.
Downloaded from journals.jps.jp by Forschungszentrum Juelich GmbH on 01/19/21

J. Phys. Soc. Jpn. **90**, 012001 (2021)   **Invited Review Papers**   F. Jin et al.

We take as an example the 1D spin-1/2 Heisenberg XXZ ring defined by

$$H = -J \sum_i S_i^x S_{i+1}^x + S_i^y S_{i+1}^y + \Delta S_i^z S_{i+1}^z, \qquad (40)$$

where $J = -1$ sets the energy scale and $\Delta = 1.5$ is the anisotropy. From the lattice version of the continuity equation, it follows that the spin current operator j is given by

$$j = -J \sum_i S_i^x S_{i+1}^y - S_i^y S_{i+1}^x. \qquad (41)$$

Spin transport properties can be obtained by Fourier transform of the current–current correlation function.

Figure 4 shows the results of the current–current autocorrelation function for two system sizes $N = 37, 38$ obtained by using the random state technology. The Suzuki–Trotter product formula algorithm[68,69] was used to compute the time evolution of the two state vectors. For simplicity, we choose $\beta = 0$. The data for the two different system sizes follow the same curve up to $tJ \approx 15$ and then start to deviate from each other. This suggests that the current–current autocorrelation function shows strong finite size effects, except for relatively short times. The data presented here are in concert with earlier results.[27,93,94] A way to alleviate this effect is to apply the idea of numerical linked cluster expansion, which allows the accurate estimation of $C(t)$ for much longer times.[95,96]

5.2 Density–density correlation

The density–density correlation is intimately related to the current–current correlation via the continuity equation but is of interest by itself. It is defined by

$$C(t) = \frac{\mathbf{Tr}\, e^{-\beta H} n(t) n(0)}{\mathbf{Tr}\, e^{-\beta H}}, \qquad (42)$$

where n is the density operator and $n(t) = e^{itH} n(0) e^{-itH}$. We only have to set $Y = n(t)n(0)$ to see that also in the case of the density–density correlation function, the random state technology can safely be applied. Numerically we can calculate this quantity by exactly the same technique as the one used for the current–current correlation, namely from the time evolution of two projected state vectors. For $\beta = 0$, we can eliminate one of the time evolutions. For a local density operator, it is always possible to find its square root, either analytically or if necessary numerically. Then, it is straightforward to show that

$$C(t) = \frac{\mathbf{Tr}\, \sqrt{n(0)}\, n(t) \sqrt{n(0)}}{D}$$
$$\approx \langle \Phi | \sqrt{n(0)}\, n(t) \sqrt{n(0)} | \Phi \rangle$$
$$= \langle \psi(t) | n(0) | \psi(t) \rangle, \qquad (43)$$

where $|\psi(t)\rangle = e^{-itH} \sqrt{n(0)} |\Phi\rangle$ is an unnormalized state vector. Note that if the density operator has negative eigenvalues, the square root of this operator is obtained by shifting the eigenvalues by the lowest eigenvalue, and the above formula should be changed accordingly.

We illustrate the method by considering once more the Heisenberg XXZ model Eq. (40). As the density operator, we take the local occupation number $n_l = S_l^z + 1/2$. Note that
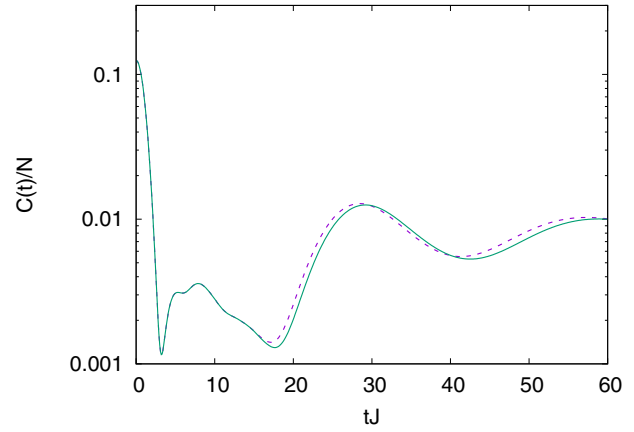

Fig. 4. (Color online) Current–current autocorrelation function normalized by the lattice size for the 1D XXZ model with $\Delta = 1.5$ for system sizes $N = 37$ (dashed line) and $N = 38$ (solid line).
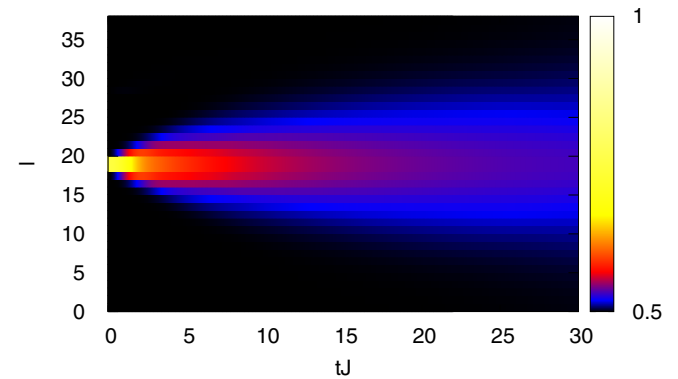

Fig. 5. (Color online) Density plot for the time evolution of the local occupation number p_l for the 1D XXZ model [see Eq. (40)] with $\Delta = 1.5$ and $N = 39$.

$\sqrt{n_l} = n_l$ here. The density–density correlation function at $\beta = 0$ is given by

$$C(t) = \frac{\mathbf{Tr}\, n_l(t) n_{L/2}}{D} = \frac{\mathbf{Tr}\, n_{L/2} n_l(t) n_{L/2}}{D}$$
$$\approx \langle \Phi | n_{L/2} n_l(t) n_{L/2} | \Phi \rangle = \langle \psi(t) | n_l | \psi(t) \rangle = p_l / 2, \qquad (44)$$

where $|\psi(t)\rangle = e^{-itH} n_{L/2} |\Phi\rangle$, $\langle \psi(t)|\psi(t)\rangle = 1/2$, and p_l is the expectation value of n_l (after proper normalization of the state vector). As only one time evolution of the state vector $|\psi(t)\rangle$ is needed, using the same resources as in the case of the current–current correlations allows us to simulate a system with one extra spin, that is the 1D XXZ model with $N = 39$ spins.

Figure 5 shows the results p_l. The initial state produces a peak in the center of the chain, i.e., $p_{20} = 1$. As a function of time, this peak then spreads over the neighbors and the density profile shows a tendency to become stationary. Within the maximum time $tJ = 30$ shown, the profile does not reach the boundary yet, indicating that up to this time finite size effects are not yet relevant. It is clear that the width of the profile develops slowly, which, in fact, fits well to a square-root increase after an initial linear increase for short time scales $tJ \leq 1$. The data provides clear evidence for diffusion in the integrable XXZ model with large anisotropy $\Delta > 1$. Our results agree with results obtained from time-dependent density matrix renormalization group calcula-

tions.[97] Further detailed discussions about diffusion and data for other models can be found in Refs. [58], [60], [61], and [63].

5.3 Electron spin resonance

According to linear response theory, the ESR signal is related to the Fourier transform of the autocorrelation function.[98] The ESR signal averaged over a period $2T$ is proportional to[98]

$$C(\omega) \equiv \frac{1}{2T} \int_{-T}^{T} \frac{\mathbf{Tr}\, e^{-\beta H} C(t)}{\mathbf{Tr}\, e^{-\beta H}} \cos \omega t\, dt, \qquad (45)$$

where

$$C(t) = \frac{1}{2} \{ e^{itH} M^x e^{-itH}, M^x \}, \qquad (46)$$

and M^x is the x-component of the total magnetization. We set $X(\omega) = e^{-\beta H/2} Y(\omega) e^{-\beta H/2}$ and $Y(\omega) = \frac{1}{2T} \int_{-T}^{T} C(t) \cos \omega t\, dt$. Because $Y(\omega) = Y^\dagger(\omega)$ and $X(\omega) = X^\dagger(\omega)$, we can use Eq. (31) to show that the variance of $C(\omega, T)$ vanishes exponentially with increasing system size.

Again, we take the XXZ model as an example, except that now we add an extra term to Eq. (40), given by $H_m = -h \sum_i S_i^z$, to represent the external magnetic field. For simplicity, we do not include the dipole–dipole interaction term.[99] The parameters are taken to be $J = -1$, $\Delta = 0.84$, and $h = 5$ and we adopt open boundary conditions.

Figure 6 presents the results for the ESR spectra obtained from simulating the given XXZ model up to times $tJ = 4096$ for $N = 30$ and $tJ = 2048$ for $N = 32$ and 34, respectively. As the scale of the x-axis of Fig. 6 indicates, this particular calculation requires a high resolution in the frequency domain. Accordingly, we need simulation data for a very large time interval, effectively limiting the system sizes we can study within a reasonable amount of elapsed time. Therefore, we computed data for system sizes $N = 30, 32, 34$ and $\beta = 0$ only.

As Fig. 6 shows, the ESR line shape clearly displays a four-peak structure, which fits well to a sum of four Lorentzians.[100] The separation between the two central peaks seems to decrease as N increases from $N = 30$ to 34. This may suggest that the central double-peak structure disappears in the thermodynamic limit. However, on the basis of numerical data, it is difficult to draw a definite conclusion about the vanishing of the central double-peak structure. In fact, this is a subtle issue and we refer the interested reader to Refs. [101]–[103] for additional data on the size and temperature dependence of the ESR line shape.

6. Application: Quantum Supremacy

Recently, Gaussian random states (see Table I, Case A) found new applications in the field of quantum information processing. In this rather long section, we scrutinize the possibility of letting a universal quantum computer generate Gaussian random states. We address the conceptual differences of realizing such states on a conventional digital computer and a hypothetical, universal quantum computer. We also discuss the difficulties that arise in designing and testing quantum circuits which are tailored to generate states that exhibit features characteristic of Gaussian random states.

A demonstration that a programmable quantum device can perform a (not necessarily useful) task which is out of reach
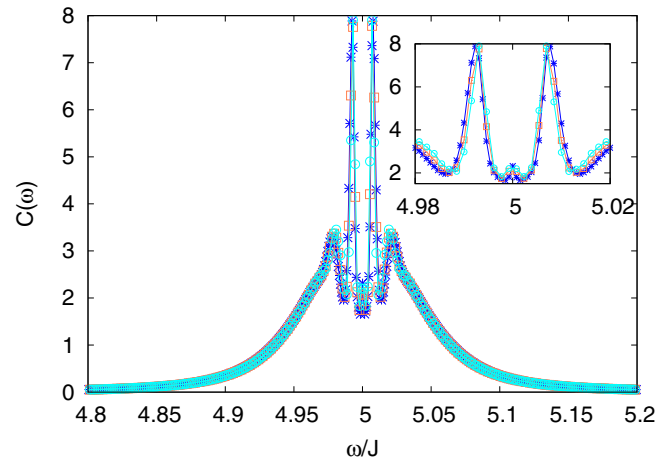

Fig. 6. (Color online) ESR spectrum [see Eq. (45)] for $N = 30$ (stars), $N = 32$ (squares), $N = 34$ (circles), and $\beta = 0$ for the 1D XXZ model given by Eq. (40) with an additional term $H_m = -h \sum_i S_i^z$ to represent the external magnetic field. The inset shows the spectrum $C(\omega)$ for $4.98 \le \omega/J \le 5.02$.

for present-day and near-future supercomputers is called quantum supremacy.[4,76,104] Assessing the potential of such a quantum device involves performing a series of established tests, so-called benchmarks. In this section and in Sect. 7, the term "benchmarking" refers to measuring the quality of a quantum information processor as a computing device by comparing experimental data produced by the processor with results obtained from (a computer simulation on a digital computer of) the quantum theoretical model for the device. More specifically, this section discusses benchmarking of the Sycamore superconducting processor[76] using the cross entropy[4] (see Sect. 6.3) as a measure. In Sect. 7, we present theoretical results related to the theory of randomized benchmarking,[105,106] which involves averaging the fidelity[77] of experiments performed with different, randomly chosen, quantum circuits.

We start by giving an outline of the content of this section. In Sect. 6.1, we analyze the general problem of sampling from a probability distribution that is specified through the amplitudes of a pure quantum state such as the Gaussian random state of Table I, Case A. We compare the resources that are needed to perform this sampling on a digital computer and on the theoretical model of a quantum computer. We also address the complexity of constructing and validating quantum gate circuits that are designed to generate states that exhibit features characteristic of Gaussian random states.

Section 6.2 discusses general aspects of the recent quantum supremacy experiment with the Sycamore superconducting processor.[76] Section 6.3 introduces the measure, the cross entropy, that is used to discriminate between outcomes that are distributed uniformly over all possibilities and outcomes reflecting the quantum gate sequence that the device was instructed to carry out.

In Sect. 6.4, we use the maximum entropy method and the knowledge of the measured cross entropy to establish a relation between the theoretically expected probability distribution and the unknown probability distribution that describes the observed outcomes.

In Sect. 6.5, we adopt the Gaussian random state (see Table I, Case A) as the model of the theoretically expected probability distribution. We show that, given a numerical estimate of the cross entropy, the maximum entropy method[107] yields a unique solution for the probability distribution that describes the observed outcomes. This solution indeed allows us to discriminate between a device that generates outcomes reflecting the applied quantum gate sequence and a device that produces uniformly distributed outcomes.

Finally, in Sect. 6.6, we combine supercomputer simulations with the experimental data produced by the Sycamore superconducting processor[76] to assess the claim that quantum supremacy has been demonstrated.

6.1 Using a quantum computer to sample from a random state

Up to this date, applications of the random state technology employ digital computers to solve problems. Obviously, for this purpose, we need to be able to store at least one pure state $|\Phi\rangle = \sum_{j=0}^{D-1} c_j |j\rangle$ in the computer's memory. This means that the computer should have enough memory to store the D complex numbers c_j. A supercomputer with 1 PiB (2^{50} bytes) of random access memory (RAM) can store 2^{46} complex double-precision coefficients c_j. Translating this into how many qubits (or $S = 1/2$ spins) we can represent with such a large amount of RAM, we find the disappointingly low number $L = 46$. Every time we add a qubit to the system, we have to double the amount of memory, just to be able to store the state $|\Phi\rangle$. The required amount of memory $D = 2^L$ grows exponentially with the number of qubits L. Clearly, the cost of such a digital computer, its power consumption, and its speed of operation severely limit the number of quantum objects one is able to study by application of the otherwise very efficient random state technology.

Random states $|\Phi\rangle$ that are drawn from the uniform distribution on the $2D$-dimensional unit sphere, Case A of Table I, are not only very convenient for many applications of the random state technology but, as shown in Sect. 7 and will become clear later in this section, they are important for the theory of randomized benchmarking of real quantum processors too. Therefore, in this section, we limit the discussion to random states that are drawn from the uniform distribution on the $2D$-dimensional unit sphere and, to simplify the writing, will refer to such states as random states in the following.

Applications of the random state technology generically involve the calculation of

$$\langle X \rangle = \langle \Phi | X | \Phi \rangle, \tag{47}$$

that is the expectation value of a matrix X for a random state $|\Phi\rangle$. Let us simplify the discussion by considering matrices X which are diagonal with respect to the (computational) basis states $\{|0\rangle, \ldots, |D-1\rangle\}$, $X|j\rangle = x_j|j\rangle$. Using this, Eq. (47) becomes

$$\langle X \rangle = \sum_{j=0}^{D-1} x_j |c_j|^2 \approx \frac{1}{m} \sum_{j \in \mathcal{J}} x_j, \tag{48}$$

were $\mathcal{J}$ is a set of $m \ll D$ states drawn with probability $|c_j|^2$ and we have used the law of large numbers[80] to approximate

the exact expressions by a sum of m instead of D terms. The problem of computing $\langle X \rangle$ has now been replaced by the problem of sampling integers j with probabilities $|c_j|^2$ (recall $\sum_{j=0}^{D-1} |c_j|^2 = 1$). The generic way for performing this sampling on a digital computer is to compute the cumulative probability distribution $P(j) = \sum_{i=0}^{j} |c_i|^2$ for all $j = 0, \ldots, D-1$, generate a uniform pseudo-random number $0 < r < 1$, and output the largest value of j for which $P(j) \leq r$.

Let us now ask ourselves how difficult, in terms of memory and computation time, it is to compute and store the cumulative probability distribution $P(j)$. It may not be too far off to assume that the total amount of storage available in the world at this time is of the order of 10 EiB ($\approx 2^{63}$ bytes, although it is not easy to find a reliable number). For a value of D of that order, the calculation of $P(j)$ needs to be performed with 8-byte floating point arithmetic. This implies that with 10 EiB of storage, we may be able to store 2^{59} values of the $P(j)$'s, the corresponding quantum system is described in terms of 59 qubits.

Using the algorithm described in Appendix D.2, we can generate $c_0, c_1, \ldots, c_{D-1}$ independently. Thus, to fill the table of $P(j)$'s, we can simply generate c_j and set $P(j) = P(j-1) + |c_j|^2$ for j running from 0 to $D-1$ [with $P(-1) = 0$]. This takes of the order of D floating-point operations. To reduce the elapsed time, we may distribute this calculation over many processors. Suppose that we can really get exclusive access to the huge amount of storage that our application needs, can we then sample from the cumulative distribution $P(j)$? Disregarding the fact that this storage is distributed over the whole world and access times to memory locations may be relatively long, we probably can because for a given random number r, we can find j by binary search, that is in a number of steps that is proportional to L, not to $D = 2^L$. The upshot of these considerations is that the limit on the value of D of the random state from which we would like to sample is dictated by the amount of available memory.

Can a gate-based quantum computer do better than a digital computer in terms of problem size? According to the mathematical model of gate-based quantum computing, the state of an L-qubit quantum computer is described by a linear superposition of the $D = 2^L$ computational basis states.[108] As the state of the quantum computer changes, all $D = 2^L$ coefficients change in parallel.[108] Thus, the mathematical model of a quantum computer provides us with a machine that exhibits an unprecedented level of intrinsic parallelism and a tremendous amount of memory (with each additional qubit, we double the amount of coefficients describing the state). The relevant question then is "can we exploit these features in practice?".

Returning to the conceptually simple task of sampling from a random state, with a quantum computer at our disposal we would proceed as follows:

1. Design a gate circuit C, that is a sequence of unitary operations,[108] that changes the state $|j = 0\rangle$ into the random state $|\Phi\rangle = \sum_{j=0}^{D-1} c_j |j\rangle$.
2. Apply C to the initial state $|j = 0\rangle$.
3. With the quantum computer in the state $|\Phi\rangle = \sum_{j=0}^{D-1} c_j |j\rangle$, perform a measurement of all the qubits. This measurement yields values $q_k = 0, 1$ for each of the $k = 0, \ldots, L-1$ qubits. The sequence of values

$q_{L-1} \cdots q_0$ is called a bitstring and is equivalent to the integer $j = \sum_{k=0}^{L-1} 2^k q_k$, corresponding to the index of the state $|j\rangle \in \{|0\rangle, \ldots, |D-1\rangle\}$. By Born's rule, the probability for this event is given by $|c_j|^2$.

4. Store the m values of j, obtained by m repetitions of steps 2 and 3, to form the set $\mathcal{M}$.

5. Compute $m^{-1} \sum_{j \in \mathcal{M}} x_j$ to obtain the desired approximation to $\langle X \rangle$.

Note that designing such circuits C is by no means trivial, as C should act on all D coefficients to create the random state. To the best of our knowledge, there is no rigorous proof that such circuits can be constructed with a depth polynomial in the number of qubits. Instead, Boixo et al. constructed so-called random circuits from single- and two-qubit gates.[4] For those circuits which can be simulated numerically, they demonstrated that they have a depth which grows polynomially with the number of qubits and produce states that yield the Porter–Thomas distribution, a direct consequence of the state being a Gaussian random state.[4] We make the plausible assumption that the length/depth of the circuit C and the time it takes to measure all qubits are of order L, not of order D. Then in the mathematical realm, the gate-based quantum computer can sample from a random state of dimension D, for values of D which are out of reach for a digital computer. Memory is not an issue.

Although the mathematical model holds great promise for very fast and very large computation, the technological hurdles to build a quantum information processor that realizes (part of) these promises seem enormous. In fact, there still is a wide gap between the mathematical model and its physical realizations. For instance, experiments with publicly accessible quantum processors show that they are not yet capable of reliably performing simple computational tasks such as adding two small integers.[109] Moreover, the performance of the current generation of gate-based quantum information processors is adversely affected by various sources of noise. For this reason, they are often referred to as Noisy Intermediate-Scale Quantum (NISQ) devices. The prospects of building a fully error-corrected quantum processor with NISQ technology are rather dim. Therefore, in order for quantum information processing to become a practical reality on short terms, it is necessary to

(i) characterize the performance of NISQ devices by a well-defined procedure

(ii) develop algorithms for solving practically relevant, nontrivial problems using these error-prone NISQ devices.

The recent quantum supremacy experiment[76] targets point (i), not point (ii).

6.2 Quantum supremacy: General aspects

As mentioned earlier, current NISQ devices have difficulties to perform, e.g., simple arithmetic operations. However, because of their noisy operation, they excel at producing "random" output. The latter feature may be put to good use to demonstrate that a NISQ device can perform a computational task which is prohibitive for state-of-the-art digital (super)-computers. Such a demonstration is coined "quantum supremacy". Quoting Boixo et al.: "we propose the task of sampling from the output distribution of random quantum

circuits as a demonstration of quantum supremacy".[4] The output distribution Boixo et al. have in mind corresponds to a realization of a state drawn from the uniform distribution on the $2D$-dimensional unit sphere, Case A of Table I, that is a random state. The Case A distribution is exceptional among multidimensional distributions because for any D, many of its properties can be studied analytically through closed form expressions. For instance, one can show [see the derivation of Eq. (A·11)] that the exact expression for probability density $p(z)$ of "the probability $z = |c_j|^2$ to observe the bitstring j" is given by $p(z) = (D-1)(1-z)^{D-2}$ [see Eq. (A·11) in Appendix A]. For large D, $p(z) \approx e^{-Dz}$, a result which is often referred to as the Porter–Thomas law.[110]

Referring to item 1 in Sect. 6.1, the first step is to design circuits C that perform the desired task. To this end, Boixo et al. use what they call a "random circuit", denoted by R in the following, These R's are specifically designed to generate states that are close to random states. Heuristics and simulation are used to guide the construction of the R's.[4,76,111] Boixo et al. have constructed a large number of such R's.[4,76]

The quantum supremacy demonstration[4,76] has four different components:

(a) Design of a random circuit R, operating on L qubits, for which the output distribution is known through simulation of the same (or approximately the same) circuit on a digital computer.

(b) A real quantum processor that can execute the random circuit R and can produce a set of measured bitstrings $\mathcal{M}$.

(c) A measure for the relation between the set of observed bitstrings and the random circuit executed on the ideal quantum computer.

(d) An error model to account for the NISQ character of the quantum processor and a procedure to extrapolate the results to values of L that cannot be simulated by a state-of-the-art digital computer.

Regarding (b), recall that simulating an L-qubit circuit C on a digital computer requires resources of the order D. At the time of writing, this prohibits the simulation (with an accuracy of 10 digits or better) of a universal quantum computer with $L > 45$.[112] Therefore, merely executing C on a real quantum processor having say, $L = 53$ qubits, is a trivial kind of "quantum supremacy". One has to demonstrate that the output distribution is indeed the one expected for the circuit C. In principle, this requires taking of the order of $S \times D$ samples where S is the number of repetitions (say of the order of $S = 10000$) required to estimate each $|c_j|^2$ with sufficient statistical accuracy. Obviously, for large L (even for $L = 45$), this sampling task is prohibitive, even with a noise-free quantum computer. Therefore, in order to proceed, one has to be less demanding and feel content with showing that the measured set of bitstrings $\mathcal{M}$ complies with the hypothesis that these bitstrings are samples drawn from the distribution corresponding to R. This then requires the specification of one or more measures for the correlation, as mentioned in item (c), which is the topic of the following subsection. Item (d) is essential for the interpretation of the data produced in the quantum supremacy experiment[76] but is outside the scope of this paper.

6.3 Cross entropy

Consider an ideal quantum computer with L qubits executing a gate circuit U represented by a unitary matrix U. This quantum computer generates states labeled by $j = 0, \ldots, D-1$ ($D = 2^L$) with probabilities $p_U(j)$. Assume that we have a physical realization of the quantum computer. As discussed above, currently only NISQ devices are able to execute the gate circuit U. This physical device produces bitstrings j with a-priori unknown probabilities denoted by $p_V(j)$. We denote the collection of bitstrings (represented by integers) generated by this device by $\mathcal{J} = \{j_1, \ldots, j_m\}$ where each $j_k \in \{0, \ldots, D-1\}$. The key question is now

> To what extent are the probability distributions p_U and p_V similar?

If we had enough data to build a histogram that approximates p_V and we knew p_U, we could resort to standard statistical tests for comparing the two distributions.[24] However, even for small L (say $L = 20$, $D = 2^L = 1048576$), a large number of bitstrings (say $m \gg 100D$) is required in order to estimate the histogram with some confidence, rendering this approach useless for practical purposes if L is large. Boixo et al.[4] proposed to circumvent this "sampling problem" by using the cross entropy,

$$C(V, U) = -\sum_{j=0}^{D-1} p_V(j) \log p_U(j), \qquad (49)$$

as a measure for the difference between the observed (p_V) and expected (p_U) distribution of bitstrings. Note that the cross entropy $C(V, U)$ is closely related to the Kullback–Leibler divergence $D(V, U) = \sum_{j=0}^{D-1} p_V(j) \log[p_V(j)/p_U(j)]$, i.e., $C(V, U) = S(V) + D(V, U)$, where $S(V) = -\sum_{j=0}^{D-1} p_V(j) \log p_V(j)$ is the Shannon entropy. By invoking the central limit theorem, Boixo et al. use the sample average

$$c_U = -\frac{1}{m} \sum_{j \in \mathcal{J}} \log p_U(j), \qquad (50)$$

of the elements in the set $\mathcal{J}$ (i.e., the m bitstrings generated by the device) to approximate the sum over all j in Eq. (49). This circumvents the problem of not being able, in practice, to collect enough data to estimate all the $p_V(j)$'s. For completeness, we mention that the recent quantum supremacy demonstration used, in addition to Eq. (49), also the "linearized" cross entropy $L(V, U) = \sum_{j=0}^{D-1} p_V(j)(D p_U(j) - 1) \approx (1/m) \sum_{j \in \mathcal{J}} (D p_U(j) - 1)$.[76] For the purpose of discussing the ideas behind these recent quantum supremacy experiments, it does not seem to matter if one uses $C(V, U)$ or $L(V, U)$.[76] Therefore, in the remainder of this section, we only consider the cross entropy $C(V, U)$.

For a fixed p_U, c_U is a positive number, obtained by computing the cross entropy Eq. (50) using a data set of bitstrings $\mathcal{J}$, generated by a device. The key question can now be reformulated as

> Given p_U and the numerical value $c_U > 0$, what can be said about the unknown distribution p_V?

In order to make a statement, one has to adopt a model. In the next subsection we appeal to the principle of maximum entropy, a recipe to obtain a probability distribution with specified averages of functions of variables (and nothing else) and which otherwise is as uniform as possible.[107]

6.4 Maximum entropy method

We follow Jaynes (Ref. [107], Sect. 11.6) and search for the extrema of

$$F = \sum_{j=0}^{D-1} \left[-p_V(j) \log p_V(j) + \lambda \left(p_V(j) - \frac{1}{D} \right) \right.$$
$$\left. + \mu \left(p_V(j) \log p_U(j) + \frac{c_U}{D} \right) \right], \qquad (51)$$

where the Lagrange multipliers λ and μ account for the constraints $\sum_{j=1}^{N} p_V(j) = 1$ and $c_U = -(1/m) \sum_{j \in \mathcal{J}} \log p_U(j)$, respectively. There is only one extremum, namely the maximum of F.[107] Hence, we solve the problem of maximizing the first term in Eq. (51) which is the entropy (or Shannon information), subject to the named constraints. The solution for the maximum is given by $p_V(j) = \exp(\lambda + \mu - 1 + \mu \log p_U(j))$. Using the named constraints we find $\exp(1 - \lambda - \mu) = \sum_{j=0}^{D-1} p_U^\mu(j)$ and

$$p_V(j) = \frac{e^{\mu \log p_U(j)}}{\displaystyle\sum_{j=0}^{D-1} e^{\mu \log p_U(j)}} = \frac{p_U^\mu(j)}{\displaystyle\sum_{j=0}^{D-1} p_U^\mu(j)}. \qquad (52)$$

We can find μ by solving

$$-\sum_{j=0}^{D-1} p_U^\mu(j) \log p_U(j) = c_U \sum_{j=0}^{D-1} p_U^\mu(j). \qquad (53)$$

In summary, if we combine the knowledge that the cross entropy $C(V, U) = c_U$ with the principle of maximum entropy, there is a definitive relation between the known probabilities $p_U(j)$ and the unknown probabilities $p_V(j)$ from which the samples $\mathcal{J}$ are drawn. Given the value of c_U and probabilities $p_U(j)$'s the maximally non-committal probability distribution is given by Eq. (52) with μ being the solution of Eq. (53). Phrased differently, there are many more probabilities to be assigned than there are constraints (the normalization and the value of c_U) and the maximum entropy principle yields the broadest distribution that is compatible with these constraints. Therefore Eq. (52) may be viewed as a minimally prejudiced assignment complying with the named constraints.

Assume that an experiment with a device yields a value of c_U for which the solution of Eq. (53) yields $\mu = 1$, Then, from Eq. (52), it follows that $p_V(j) = p_U(j)$. In other words, excluding all other knowledge we might have, the principle of maximum entropy suggests (but not guarantees) that the device is working properly, meaning that it generates bitstrings according to the probabilities $p_U(j)$ that correspond to the circuit U. Similarly, if the device produces a value of c_U for which the solution of Eq. (53) yields $\mu = 0$, the principle of maximum entropy suggests (but not guarantees) that $p_V(j) = 1/D$, that is the states are sampled from a uniform superposition of basis states (see Case C in Table I).

Up to this point, the discussion has been entirely general, independent of a particular choice of the gate sequence U. Clearly, in combination with the maximum entropy principle, the cross entropy estimate Eq. (50) is a useful measure for the correspondence between the observed bitstrings $\mathcal{J}$ produced by a quantum device and the ideal gate sequence U that one would like this device to carry out.

6.5 Quantum supremacy: Theory

A key advantage of using the cross entropy is that it can be estimated without having to perform a prohibitively large number of measurements. To be useful, the only requirement is that $p_U(j)$ is known, through simulation on a digital computer[76,112,113] or through a model that can be treated analytically. For an arbitrary sequence U, memory requirements limit the calculation of the $p_U(j)$'s to less than 50 qubits ($D < 2^{50}$) on current supercomputers.[76,112] To demonstrate that a quantum processor can perform a task which cannot be performed by a digital computer, one has to resort to circuits U for which the cumulative probability distribution cannot be determined analytically and is not amenable to computation by a digital computer.

It is instructive to first consider the case in which the circuit E generates a set of bitstrings $\mathcal{E}$ for the uniform superposition state by applying a Hadamard gate to each qubit,[108] that is Case C of Table I (with all a_j's equal to zero) for which $p_E(j) = 1/D$ and, from Eq. (50), $c_E = \log D$. Then Eq. (53) is an identity for arbitrary μ and Eq. (52) yields $p_V(j) = p_E(j) = 1/D$. Thus, assuming the device to work properly, the circuit E is expected to generate m bitstrings $\mathcal{E}$ for which $c_E = \log D$ (up to statistical fluctuations).

Next, we consider somewhat less trivial states, namely those that are distributed uniformly on the $2D$-dimensional unit sphere. These are the states A of Table I that are at the heart of the random state technology. In the recent quantum supremacy experiment,[76] the quantum processor executes a so-called random circuit R and produces sets of bitstrings $\mathcal{R}$ for each instance of the circuit R. The basic idea is that the state produced by R is an instance of the uniform distribution on the $2D$-dimensional unit sphere. Each sample $\mathcal{R}$ of m bitstrings carries a "fingerprint" of the particular circuit R that produced these bitstrings. In symbols, for a particular random circuit R, the state of the ideal quantum computer executing R reads $|\Psi_R\rangle = \sum_{j=0}^{D-1} z_j |j\rangle$ where $(z_1, \ldots, z_{D-1})$ is a sample drawn from a uniform distribution over the $2D$-dimensional unit sphere. However, if we assume that the state generated by such a circuit is indeed an instance of the uniform random state $|\Phi\rangle$, we can use random state technology to explore its features analytically.

In the remainder of this subsection, we consider the case that the quantum processor executes the circuit R without producing any error and that the resulting state seems to have been drawn from the uniform distribution over the $2D$-

dimensional unit sphere. The first issue to address is whether, for the application to cross-entropy benchmarking, we can obtain accurate estimates from one realization of the state $|\Phi\rangle$ if D is large. It follows by calculation (see Appendix A) that

$$\text{Var}\left[\sum_{j=0}^{D-1} p_R(j) \log p_R(j)\right] = \frac{\pi^2 - 9}{3D} + \mathcal{O}\left(\frac{1}{D^2}\right), \quad (54)$$

such that also in the case of random sampling from $|\Phi\rangle$, we can expect an accurate estimate of the entropy from one realization (if D is large). Similarly, it follows that for large D, the variance of the difference between entropy and its estimate computed from the set $\mathcal{R}$ having J elements is given by

$$\text{E}\left[\left(\frac{D}{J}\sum_{j\in\mathcal{R}} p_R(j)\log p_R(j) - \sum_{j=0}^{D-1} p_R(j)\log p_R(j)\right)^2\right]$$

$$= \frac{1 - J/D}{J}\left[\frac{\pi^2 - 9}{3} + (\log D + \gamma - 2)^2\right], \quad (55)$$

where $\gamma \approx 0.577$ is Euler's constant. As $\log D \approx 0.69L$, Eq. (55) shows that if $J \gg L^2$, any subset $\mathcal{R}$ will yield an estimate of the entropy which, on average, will be an accurate approximation to the entropy $-\sum_{j=1}^{D} p_R(j) \log p_R(j)$.

Finally, we consider the solution of the maximum entropy problem Eq. (53). According to Eq. (54), if D is large, we may replace the calculation based on one realization of the random state by the average over all equivalent random states. This amounts to performing integrals over multi-dimensional Gaussians, which is straightforward. It then follows from Eq. (A·12) that we have to find the value of μ that satisfies the equation

$$c_R = \log D - \text{PolyGamma}(0, \mu + 1), \quad c > 0. \quad (56)$$

The function $\text{PolyGamma}(0, \mu + 1)$ is the logarithmic derivative of the Gamma function, that is $\text{PolyGamma}(0, \mu + 1) = \Gamma'(\mu + 1)/\Gamma(\mu + 1)$. This function is monotonically increasing function for $\mu > -1$ with a divergence at -1, negative for $-1 < \mu < 0.46163$, and positive for $\mu > 0.46164$. For large μ, we have $\text{PolyGamma}(0, \mu + 1) = \log \mu + 1/2\mu + \mathcal{O}(1/\mu^2)$. Therefore, Eq. (56) has a unique solution for μ.

Instead of solving Eq. (56) for μ, it is easier to compute, for a chosen value of μ, the difference between the value of $c_E = \log D$ of the uniform distribution $p_E(j) = 1/D$ and the theoretical value c_R given by Eq. (56). We find

$$c_E + \gamma - c_R = \begin{cases} 0, & \mu = 0 \leftarrow \text{uniform distribution over all } D \text{ states} \\ 1, & \mu = 1 \leftarrow \text{uniform distribution on the } 2D\text{-dimensional unit sphere.} \\ 3/2, & \mu = 2 \end{cases} \quad (57)$$

In the next subsection, we use the analytical results of this subsection to interpret the experimental results of a recent quantum supremacy experiment.[76]

6.6 Quantum supremacy: Experiment and simulation

In this subsection, we scrutinize the results for the cross entropy, obtained by combining the bitstrings $\mathcal{M}$ produced by the 53-qubit Sycamore superconducting processor[76] and probabilities $p_R(j)$ calculated with the universal quantum computer simulator JUQCS-E.[112–114]

For a given quantum circuit R, designed to generate a random state, JUQCS-E[112] executes R and computes the probability distribution $p_R(j)$ for each quantum state $j \in \{0, \ldots, D-1\}$. JUQCS-E also computes the cumulative distribution function $P_R(k) = \sum_{j=0}^{k} p_R(j)$ and samples states from this distribution, yielding a set of bitstrings $\mathcal{R}$. All these calculations are numerically exact (up to at least 10 digits). A feature of JUQCS-E, not documented in Ref. 112, allows the user to specify a set $\mathcal{M}$ of m bitstrings for which JUQCS-E calculates $p_R(j)$ for all $j \in \mathcal{M}$. The latter feature allows us to

compute the estimate $-(1/m)\sum_{j\in\mathcal{M}}\log p_R(j)$ for the cross entropy.

Following the methodology for cross-entropy benchmarking,[4,76] the quantities of interest are

$$\alpha_{R,R} \equiv c_E + \gamma - c_R = \log D + \gamma + \sum_{j=0}^{D-1} p_R(j)\log p_R(j), \quad (58)$$

$$\alpha_{R,\mathcal{X}} \equiv \log D + \gamma + \frac{1}{m}\sum_{j\in\mathcal{X}}\log p_R(j), \quad (59)$$

where $\mathcal{X} = \mathcal{R}, \mathcal{M}, \mathcal{E}$, and $\mathcal{R}$ and $\mathcal{M}$ are sets of bitstrings generated by JUQCS-E and by the Sycamore processor,[76] respectively. $\mathcal{E}$ is a set of bitstrings sampled from the uniform distribution $[p_E(j) = 1/D]$. From the theoretical model presented in Sect. 6.5, it follows that $\alpha_{R,\mathcal{M}} \approx 0$ if a quantum processor produces bitstrings that are distributed uniformly. The larger the value of $\alpha_{R,\mathcal{M}}$, the more likely it is that the quantum processor has been sampling from the correct distribution. In this sense, the uniform distribution corresponding to $\alpha_{R,\mathcal{M}} \approx 0$ provides a baseline for the assessment of the quality of a NISQ device.

If m is sufficiently large ($m = 500000$ for the experimental data sets[76]), we may expect that $\alpha_{R,\mathcal{R}} \approx \alpha_{R,R}$. If the circuit R produces a genuine random state, averaging over all such R's yields $\langle\alpha_{R,R}\rangle_R = 1$, see Eq. (57). Note that the last term in Eq. (58) [Eq. (59)] is equal to minus the cross entropy $C(R,R)$ [$C(R,\mathcal{X})$].

In Table II, we present results for the α's defined by Eqs. (58) and (59). Most of these results were included in the original report on the quantum supremacy experiment.[76] Note that if we use Eq. (55) to estimate the standard deviation of $\alpha_{R,\mathcal{X}}$ we find that for $m = 500000$ samples, the standard deviation is about 0.0014. The second and third column show that the results produced by JUQCS-E are in excellent agreement with the theoretical prediction Eq. (57) for $\mu = 1$. Recall that the latter has been obtained by averaging over all states distributed uniformly over the $2D$-dimensional unit sphere whereas the former is obtained from a simulation with a single instance of the random circuits which have been "engineered"[76] to generate instances of such a state. The fact that $\alpha_{R,\mathcal{R}} \approx 1$ may suggest, but is not a proof, that the circuit R, executed by JUQCS-E, produces a random state of the type A (see Table I). In the fourth column we present the results obtained by using the sets of bitstrings $\mathcal{M}$ measured in the recent quantum supremacy experiment.[76] The fifth column shows two results for $\alpha_{R,\mathcal{M}}$, obtained by using the JUQCS-E data for $p_R(j)$, computed for circuit 39[b], and the experimental data generated by circuits 39[c] and 39[d], respectively. The α's obtained by replacing the measured bitstrings $\mathcal{M}$ by bitstrings $\mathcal{E}$ sampled from a uniform distribution are given in the sixth column of Table II.

The numerical results presented in Table II can be summarized as follows:

(i) For $L = 39, 42, 43$, the experimental bitstrings yield values of $\alpha_{R,\mathcal{M}}$ in the range 0.018–0.038 (column four). The value of $\alpha_{R,\mathcal{M}}$ decreases exponentially as the number of qubits L increases (analysis not shown, see Ref. 76).

(ii) The values of $|\alpha_{R,\mathcal{E}}|$ (column six) are at least one order of magnitude smaller than those of $|\alpha_{R,\mathcal{M}}|$ (column four).

Table II. Results for the α's (directly related to the cross entropies) defined by Eqs. (58) and (59). The probabilities $p_R(j)$ for the circuit R have been obtained by JUQCS-E.[112] The corresponding sets of $m = 500000$ bitstrings $\mathcal{M}$ have been obtained from experiments.[76] In the first column, the letter in brackets identifies the instance of the different random circuits R used in the experiments.[76] The results obtained by using the circuit 39[b] and the measured data generated by the circuit indicated in the corresponding row are listed as $\alpha_{39[b],\mathcal{M}}$. R: pseudo-random circuit; $\mathcal{R}$: sampled states, obtained by executing R on the simulator JUQCS-E; $\mathcal{M}$: sampled states, produced by the Sycamore processor executing R;[76] $\mathcal{E}$: states sampled from a uniform distribution.

qubits[circuit ID]	$\alpha_{R,R}$	$\alpha_{R,\mathcal{R}}$	$\alpha_{R,\mathcal{M}}$	$\alpha_{39[b],\mathcal{M}}$	$\alpha_{R,\mathcal{E}}$
30[a]	1.0000	0.9997	0.0708		0.0026
39[a]	1.0000	0.9992	0.0281		−0.0003
39[b]	1.0000	1.0002	0.0350		0.0006
39[c]	1.0000	0.9996	0.0351	−0.0013	0.0034
39[d]	1.0000	0.9999	0.0375	−0.0007	0.0036
42[a]	1.0000	0.9998	0.0287		−0.0024
42[b]	1.0000	1.0027	0.0254		0.0014
43[a]	1.0000	1.0013	0.0182		−0.0010

(iii) Replacing bitstrings produced by the experiment implementing circuit 39[b] by bitstrings $\mathcal{M}$ generated by an experiment performing a different circuit R = 39[c], 39[d] yields values of $|\alpha_{39[b],\mathcal{M}}|$ (column five). They are at least one order of magnitude smaller than those of $|\alpha_{R,\mathcal{M}}|$ (column four).

On the basis of (i) alone, it seems that there is little evidence in support of the hypothesis that the states of the set $\mathcal{M}$ have been sampled from the distribution that is characteristic for the circuit R. In fact, one might even be tempted to conclude that the Sycamore processor executing circuit R samples bitstrings from a uniform distribution. However, even though the numerical values of $\alpha_{R,\mathcal{M}}$ are small, they are still significantly larger than the values of $\alpha_{R,\mathcal{E}}$ obtained by sampling from the uniform distribution, see (ii).

We can make this tentative conclusion more quantitative by using the model derived by the maximum entropy method and by performing a hypothesis test. First, given a value of c_R, the solution of Eq. (56) is unique. Therefore, $\alpha_{R,\mathcal{M}} \approx 0$ implies that $\mu \approx 0$. From Eq. (52) it then follows that $p_V(j) \approx 1/D$. Second, following Jaynes (Ref. 107, Sect. 9.11), we consider the expression

$$\psi_X = \sum_{j=0}^{D-1} n_j \log\left(\frac{n_j}{mp_X(j)}\right) \geq 0, \quad (60)$$

where n_j is the number of times a bitstring j has been observed in m repetitions of the experiment and we have omitted constants that are irrelevant for the present purpose. The larger the value of ψ_X, the less weight the hypothesis X has relative to any alternative hypothesis belonging to the same Bernoulli class (represented by probabilities for the bitstrings that are independent and stationary).[107] Thus, according to this hypothesis test, the data gives more weight to hypothesis R than to hypothesis E if $\psi_E > \psi_R$. From Eq. (60), we have $\psi_E - \psi_R \propto \alpha_{R,\mathcal{M}} - \gamma$. Therefore, we should opt for hypothesis R if $\alpha_{R,\mathcal{M}} > \gamma$, an inequality which is obviously not supported by the data (since $\alpha_{R,\mathcal{M}} \approx 0$). Note that this test does not rule out that there are other, better hypotheses than the two, E and R, that we have considered. At any rate, on the basis of the values of $\alpha_{R,\mathcal{M}}$ alone, we should conclude that the processor is more likely to execute E

J. Phys. Soc. Jpn.
Downloaded from journals.jps.jp by Forschungszentrum Juelich GmbH on 01/19/21

J. Phys. Soc. Jpn. **90**, 012001 (2021)   **Invited Review Papers**   F. Jin et al.

than R. However, this conclusion is premature because the preceding analysis focuses on the sampling and hypothesized distributions without taking into account what the circuit R actually does when it is executed by Sycamore processor.

According to (ii), feeding JUQCS-E with the bitstrings $\mathcal{E}$, generated on a digital computer using a uniform distribution, does not support the latter conclusion. Indeed, the values of $|\alpha_{R,\mathcal{E}}|$ (column six) are at least one order of magnitude smaller than those of $|\alpha_{R,\mathcal{M}}|$ (column four). Furthermore, from (iii) it follows that the $|\alpha_{R,\mathcal{M}'}|$'s of bitstrings $\mathcal{M}'$ produced by another circuit R′ are also much smaller than those obtained by using bitstrings $\mathcal{M}$ produced by circuit R. This observation strongly suggests that the bitstrings $\mathcal{M}$ produced by the Sycamore processor carry a definite (albeit weak) signature of the circuit R that was used to generate bitstrings $\mathcal{M}$. An important missing element in the above analysis is mentioned in (d) above in Sect. 6.2, namely the fact that the NISQ processor is error-prone. In fact, these errors are substantial.[76] For a detailed discussion of the analysis that incorporates an error model, see Ref. [76].

To refute a claim that quantum supremacy has been demonstrated, the relevant question is then "for each R, can we construct an approximation that uses only a (non-exponential) fraction of the D-dimensional Hilbert space and yields values of $\alpha_{R,\mathcal{M}}$ similar to those shown in Table II. A recent paper suggests that the answer to this question may be affirmative although the results presented in that paper do not refute quantum supremacy yet.[115] By restricting state vectors of the quantum computer to a class of matrix-product states, the paper shows that it is possible to generate states that carry signatures (expressed through fidelities rather than cross entropies) of the generating circuit with quality similar to those observed in the quantum supremacy experiments.[76] However, using an approximation based on matrix-product states allows computation on a digital computer at a cost that does not increase exponentially with the number of qubits.[115] Clearly, the last word about a quantum processor surpassing the power of a digital computer for a specific computational task has not been said yet.

7. Application: Quantum Information Theory

It has become common practice to characterize NISQ devices through what is called randomized benchmarking.[105,106] In essence, randomized benchmarking characterizes a quantum processor in terms of averages over many different (random) instances of sequences of gates.

In this section, we use elements of the random state technology to address some theoretical aspects of characterizing NISQ devices through randomized benchmarking. The result will be a generalization of a well-known formula for the average fidelity[77,78] to non-trace-preserving quantum operations.

Let ρ and σ denote the density matrices describing the state of the mathematical idealization of a quantum processor and of a real device, respectively. The fidelity defined by[116]

$$F(\rho, \sigma) = \left(\mathbf{Tr} \sqrt{\sqrt{\rho}\sigma\sqrt{\rho}} \right)^2 \qquad (61)$$

is a measure of the difference between the density matrices ρ and σ. In the case that $\rho = |\psi\rangle\langle\psi|$ is a pure state, we have $\rho = |\psi\rangle\langle\psi|$, $\sqrt{\rho} = \rho = \rho^2$, and the fidelity simplifies to the overlap, that is $F(\rho, \sigma) = \langle\psi|\sigma|\psi\rangle$.

A mathematically convenient way to discuss the result of a sequence of quantum gate operations applied to the initial state $|\psi\rangle$ is through the language of linear maps and quantum operations.[43,108] In short, a quantum operation transforms an initial density matrix ρ as $\rho' = \mathcal{E}(\rho) = \sum_\alpha E_\alpha \rho E_\alpha^\dagger$ where the E_α are the so-called operation elements (also known as Kraus operators; not to be confused with gate operations).[108] The number of different E_α's is at most D^2.[108] In general, $\rho \mapsto \mathcal{E}(\rho)$ defines a completely positive map[43,108] which need not be trace-preserving, meaning that $\mathbf{Tr} \sum_\alpha E_\alpha^\dagger E_\alpha$ may be less than one.[43,108] In the special case that $\mathcal{E}$ is trace-preserving, we have $\mathbf{Tr}\,\mathcal{E}(\rho) = \mathbf{Tr}\,\rho = 1$.

In this language, an ideal quantum gate operation corresponding to a unitary matrix U is described by the map $\mathcal{E}_{\mathrm{id}}(\rho) = U\rho U^\dagger$. We imagine that a real quantum device, prepared in a pure state $\rho = |\psi\rangle\langle\psi|$, actually carries out a slightly different operation denoted by the linear map $\mathcal{E}_{\mathrm{ac}}(\rho)$. We consider the quantum operation $\mathcal{E}(\rho) = \mathcal{E}_{\mathrm{ac}}(\mathcal{E}_{\mathrm{id}}^{-1}(\rho)) = \mathcal{E}_{\mathrm{ac}}(U^\dagger \rho U)$. If the device's implementation of U was perfect, the operation $\mathcal{E}$ would be equal to the identity operation. The fidelity of the operation is given by

$$F(\rho, \mathcal{E}(\rho)) = \langle\psi|\mathcal{E}(|\psi\rangle\langle\psi|)|\psi\rangle. \qquad (62)$$

Clearly, the fidelity in Eq. (62) is a numerical measure for how well a quantum processor performs the gate sequence represented by the unitary matrix U. If the fidelity is equal to one (or zero), the quantum processor is performing perfectly (or producing output that has no relation to U).

Instead of estimating the fidelities of NISQ devices for a collection of different states $|\psi\rangle$, randomized benchmarking is motivated by drawing a sample of $|\psi\rangle$'s from a uniform distribution on the unit sphere and computing the average over the fidelity in Eq. (62),

$$F_{\mathrm{avg}}(\mathcal{E}) = \int \langle\psi|\mathcal{E}(|\psi\rangle\langle\psi|)|\psi\rangle \, d\psi, \qquad (63)$$

where the integral in Eq. (63) is over all pure states or, equivalently, over all points on the surface of a unit sphere of dimension $2D$. These integrals can be evaluated by the same method that we have used before. Writing $|\psi\rangle = \sum_{j=1}^D c_i|j\rangle$ [see Eq. (1)], we have

$$F_{\mathrm{avg}}(\mathcal{E}) = \sum_\alpha \sum_{j,k=1}^D \sum_{l,m=1}^D \langle j|E_\alpha|k\rangle\langle l|E_\alpha^\dagger|m\rangle \int c_j^* c_k c_l^* c_m \, d\mathbf{c}$$

$$= \sum_\alpha \sum_{j,k=1}^D \sum_{l,m=1}^D \langle j|E_\alpha|k\rangle\langle l|E_\alpha^\dagger|m\rangle \mathrm{E}[c_j^* c_k c_l^* c_m], \quad (64)$$

where the integral is over all complex-valued c's such that $\sum_{j=1}^D c_j^* c_j = 1$. Using Eq. (4) and the results given in the first row of Table I, we have $\mathrm{E}[c_j^* c_k c_l^* c_m] = (\delta_{j,k}\delta_{l,m} + \delta_{j,m}\delta_{k,l})/D(D+1)$, and Eq. (64) becomes

$$F_{\mathrm{avg}}(\mathcal{E}) = \sum_\alpha \frac{|\mathbf{Tr}\,E_\alpha|^2 + \mathbf{Tr}\,E_\alpha^\dagger E_\alpha}{D(D+1)}$$

$$= \frac{D^2 F_{\mathrm{ent}}(\mathcal{E}) + \sum_\alpha \mathbf{Tr}\,E_\alpha^\dagger E_\alpha}{D(D+1)}, \qquad (65)$$

where $F_{\mathrm{ent}}(\mathcal{E}) = \sum_\alpha |\mathbf{Tr}\,E_\alpha|^2/D^2$ is the so-called entanglement fidelity.[108,117] In the case that $\mathcal{E}$ is trace preserving we have $\sum_\alpha \mathbf{Tr}\,E_\alpha^\dagger E_\alpha = D$ and Eq. (65) reduces to[118]

$$F_{\text{avg}}(\mathcal{E}) = \frac{DF_{\text{ent}}(\mathcal{E}) + 1}{D + 1}. \tag{66}$$

The relation Eq. (65) generalizes an earlier result[77,78] to the case of non-trace-preserving quantum operations and also yields simple expressions for the average and entanglement fidelity in terms of the operation elements E_α.

8. Final Remarks

As mentioned in the introduction, for several decades many scientists have used random states in their simulation work. In this paper, we have presented a systematic and rigorous analysis of the mathematical foundation of the random state technology as it is being used in numerical simulation. Applications of this technology in areas of quantum statistical physics, quantum dynamics, and quantum information processing have been given, with the primary aim to illustrate the power and versatility of the random state technology.

In essence, the random state technology as applied in the numerical simulation arena reduces the computational burden from $\mathcal{O}(D^2)$ to $\mathcal{O}(D)$, where D is the dimension of the Hilbert space used to describe the quantum system. For such applications, there are rigorous bounds on the errors and statistical fluctuations that result from the use of random states.

We have also shown how the random state technology can help to analyze numerical simulations and experiments that aim for establishing quantum supremacy on a NISQ processor, as well as to prove a statement for non-trace-preserving maps in quantum information theory.

In view of the computational power, flexibility and versatility of the random state technology and the fact that it is based on solid principles, we may expect to see many new applications in the near future.

Appendix A: Some Useful Integrals

We first consider the Gaussian random states listed in the first two rows of Table I. Since the expressions of interest involve averages of $\langle\Phi|X|\Phi\rangle/\langle\Phi|\Phi\rangle$ only, both choices for $c_j(\boldsymbol{\xi})$ yield the same expressions for $\langle\Phi|X|\Phi\rangle/\langle\Phi|\Phi\rangle$. In other words, the exact expressions to be derived apply to both choices of $c_j(\boldsymbol{\xi})$. Some of the integrals used below can also be found in Ref. 120.

Using the symmetries of the probability densities and the fact that both the real and imaginary part of $c_j(\boldsymbol{\xi}) = c_j(a_1, b_1, \ldots, a_D, b_D)$ are odd functions of each of the a's and b's, we have

$$\mathrm{E}\left[\frac{\langle\Phi|X|\Phi\rangle}{\langle\Phi|\Phi\rangle}\right] = \mathrm{E}\left[\frac{\sum_{i,j=1}^{D} c_i^* c_j \langle i|X|j\rangle}{\sum_{i=1}^{D} c_i^* c_i}\right]$$

$$= \mathrm{E}\left[\frac{c_1^* c_1}{\sum_{i=1}^{D} c_i^* c_i}\right] \mathbf{Tr}\, X$$

$$= \mathrm{E}\left[\frac{\xi_1^* \xi_1}{\sum_{i=1}^{D} \xi_i^* \xi_i}\right] \mathbf{Tr}\, X, \tag{A.1}$$

and

$$\mathrm{E}\left[\frac{\xi_1^* \xi_1}{\sum_{i=1}^{D} \xi_i^* \xi_i}\right] = \frac{1}{\pi^D} \int_{-\infty}^{+\infty} \frac{a_1^2 + b_1^2}{a_1^2 + b_1^2 + a_2^2 + b_2^2 + \cdots + a_D^2 + b_D^2} e^{-(a_1^2 + b_1^2 + a_2^2 + b_2^2 + \cdots + a_D^2 + b_D^2)} \, da_1\, db_1 \cdots da_D\, db_D$$

$$= \frac{1}{\pi^D} \int_{S_2} \int_{S_{2D-2}} \int_0^\infty \int_0^\infty \frac{r^2}{r^2 + R^2} e^{-(r^2 + R^2)} r R^{2D-3} \, dr\, dR\, d\Omega_2\, d\Omega_{2D-2}. \tag{A.2}$$

Using $\int_{S_p} d\Omega_p = 2\pi^{p/2}/\Gamma(p/2)$ and making the change of variables $r = x\cos\theta$ and $R = x\sin\theta$, we obtain

$$\mathrm{E}\left[\frac{\xi_1^* \xi_1}{\sum_{i=1}^{D} \xi_i^* \xi_i}\right] = \frac{4}{\Gamma(1)\Gamma(D-1)} \int_0^\infty \int_0^{\pi/2} x^{2D-1} e^{-x^2} \cos^3\theta \sin^{2D-3}\theta \, dx\, d\theta$$

$$= \frac{4}{\Gamma(1)\Gamma(D-1)} \frac{\Gamma(2)\Gamma(D-1)}{4\Gamma(D+1)} \int_0^\infty y^{D-1} e^{-y} \, dy = \frac{\Gamma(D)}{\Gamma(D+1)} = \frac{1}{D}, \tag{A.3}$$

where we used the identity $\int_0^{\pi/2} \cos^{2p-1}\theta \sin^{2q-1}\theta \, d\theta = \Gamma(p)\Gamma(q)/2\Gamma(p+q)$ and made the substitution $y = x^2$.

Similarly, we have

$$\mathrm{E}\left[\frac{(c_1^* c_1)^2}{\left(\sum_{i=1}^{D} c_i^* c_i\right)^2}\right] = \mathrm{E}\left[\frac{(\xi_1^* \xi_1)^2}{\left(\sum_{i=1}^{D} \xi_i^* \xi_i\right)^2}\right]$$

$$= \frac{1}{\pi^D} \int_{-\infty}^{+\infty} \frac{(a_1^2 + b_1^2)^2}{(a_1^2 + b_1^2 + a_2^2 + b_2^2 + \cdots + a_D^2 + b_D^2)^2} e^{-(a_1^2 + b_1^2 + a_2^2 + b_2^2 + \cdots + a_D^2 + b_D^2)^2} \, da_1\, db_1 \cdots da_D\, db_D$$

J. Phys. Soc. Jpn.
Downloaded from journals.jps.jp by Forschungszentrum Juelich GmbH on 01/19/21

J. Phys. Soc. Jpn. **90**, 012001 (2021)   **Invited Review Papers**   F. Jin et al.

$$= \frac{1}{\pi^D} \int_{S_2} \int_{S_{2D-2}} \int_0^\infty \int_0^\infty \frac{r^4}{(r^2+R^2)^2} e^{-(r^2+R^2)} r R^{2D-3} \, dr \, dR \, d\Omega_2 \, d\Omega_{2D-2}$$

$$= \frac{4}{\Gamma(1)\Gamma(D-1)} \int_0^\infty \int_0^{\pi/2} x^{2D-1} e^{-x^2} \cos^5 \theta \sin^{2D-3} \theta \, dx \, d\theta = \frac{2}{D(D+1)}, \tag{A.4}$$

and

$$\mathrm{E}\left[\frac{c_1^* c_1 c_2^* c_2}{\left(\sum_{i=1}^D c_i^* c_i \right)^2} \right] = \mathrm{E}\left[\frac{\xi_1^* \xi_1 \xi_2^* \xi_2}{\left(\sum_{i=1}^D \xi_i^* \xi_i \right)^2} \right]$$

$$= \frac{1}{\pi^D} \int_{-\infty}^{+\infty} \frac{(a_1^2+b_1^2)(a_2^2+b_2^2)}{(a_1^2+b_1^2+a_2^2+b_2^2+\cdots+a_D^2+b_D^2)^2} e^{-(a_1^2+b_1^2+a_2^2+b_2^2+\cdots+a_D^2+b_D^2)^2} \, da_1 \, db_1 \cdots da_D \, db_D$$

$$= \frac{1}{\Gamma^2(1)\Gamma(D-2)} \int_0^\infty \int_0^\infty \int_0^\infty \frac{r^2 s^2}{(r^2+s^2+R^2)^2} r s R^{2D-5} e^{-(r^2+s^2+R^2)} \, dx \, dy \, dz$$

$$= \frac{1}{\Gamma^2(1)\Gamma(D-2)} \int_0^{\pi/2} \int_0^{\pi/2} \int_0^\infty \cos^3 \phi \sin^3 \phi \sin^5 \theta \cos^{2D-5} \theta x^{2D-1} e^{-x^2} \, dx \, d\theta \, d\phi$$

$$= \frac{1}{D(D+1)}, \tag{A.5}$$

from which it follows that

$$\mathrm{E}\left[\left| \frac{D\langle\Phi|X|\Phi\rangle}{\langle\Phi|\Phi\rangle} \right|^2 \right] = \frac{\mathbf{Tr}\, XX^\dagger + |\mathbf{Tr}\, X|^2}{D(D+1)}. \tag{A.6}$$

Combining Eqs. (A·1), (A·3), and (A·6), we obtain the exact expressions

$$\mathrm{E}\left[\frac{D\langle\Phi|X|\Phi\rangle}{\langle\Phi|\Phi\rangle} \right] = \mathbf{Tr}\, X,$$

$$\mathrm{Var}\left[\frac{D\langle\Phi|X|\Phi\rangle}{\langle\Phi|\Phi\rangle} \right] = \frac{D\,\mathbf{Tr}\, XX^\dagger - |\mathbf{Tr}\, X|^2}{(D+1)}. \tag{A.7}$$

In the case of the random phase state (Case C in Table I), $\langle\Phi|\Phi\rangle = 1$ by construction. Then, from Eq. (5), it follows directly that $\mathrm{E}[D\langle\Phi|X|\Phi\rangle/\langle\Phi|\Phi\rangle] = \mathbf{Tr}\, X$. Using Eq. (17), $\langle\Phi|\Phi\rangle = 1$, and the expressions for the moments given in the third line of Table I we find

$$\mathrm{Var}\left[\frac{D\langle\Phi|X|\Phi\rangle}{\langle\Phi|\Phi\rangle} \right] = \mathbf{Tr}\, XX^\dagger - \sum_{j=1}^D |\langle j|X|j\rangle|^2 = \sum_{j,k=1}^D (1-\delta_{j,k}) |\langle j|X|k\rangle|^2. \tag{A.8}$$

Note that unlike in the case of the Gaussian random states, for the random phase state the variance depends on the choice of the basis states $\{|j\rangle | j = 1, \dots, D\}$.

Given a real number z, we may ask for the probability density $p(z|D)$ that the j-th basis state occurs with probability $p(j) = z$. For a given random state, we have $p(j) = (a_j^2 + b_j^2)/(a_1^2 + b_1^2 + \cdots + a_D^2 + b_D^2)$. By symmetry, all basis states are equivalent and it suffices to consider only $j = 1$, for instance. Because $0 \le p(j) \le 1$ we have $p(z|D) = 0$ if $z < 0$ or $z > 1$. For $0 \le z \le 1$ we can use the same tricks as before and by some elementary algebra we obtain

$$p(z|D) = \int \delta\left(z - \frac{a_j^2+b_j^2}{a_1^2+b_1^2+\cdots+a_D^2+b_D^2} \right) p(a_1, b_1, \dots, a_D, b_D) \, da_1 \cdots db_D$$

$$= \frac{1}{(2\pi)^D} \frac{2\pi \, 2\pi^{D-1}}{\Gamma(D-1)} \int_0^\infty \int_0^\infty \delta\left(z - \frac{r^2}{r^2+R^2} \right) e^{-r^2/2} e^{-R^2/2} r \, dr \, R^{2N-3} \, dR$$

$$= \frac{1}{(2\pi)^D} \frac{2\pi \, 2\pi^{D-1}}{4\Gamma(D-1)} \int_0^\infty \int_0^\infty \delta\left(z - \frac{x}{x+y} \right) y^{D-2} e^{-x/2} e^{-y/2} \, dx \, dy$$

$$= \frac{1}{2^D} \frac{1}{\Gamma(D-1)} \int_0^\infty \frac{y^{D-1}}{(1-z)^2} e^{-y/[2(1-z)]} \, dy$$

$$= \frac{1}{2^D} \frac{2^D \Gamma(D)}{\Gamma(D-1)} (1-z)^{D-2} = (D-1)(1-z)^{D-2}. \tag{A.9}$$

More generally, we can ask for the probability density that $p(j_1), \dots, p(j_k)$ (all j's different from each other) take the values $z_1, \dots, z_k$, respectively. A calculation similar to the one used to obtain Eq. (A·9) yields for $1 \le k < D$

$$p(z_1, z_2, \dots, z_k|D) = (D-1)(D-2)\cdots(D-k)(1 - z_1 - z_2 - \cdots - z_k)^{D-k-1}$$

$$\times \Theta(z_1) \cdots \Theta(z_k) \Theta(1 - z_1 - z_2 - \cdots - z_k). \tag{A.10}$$

Table B·I. Expressions for the various terms that appear in Eqs. (B·1) and (B·2), corresponding to the combination of probabilities $P(\xi)$ and amplitudes $c_j(\xi)$ listed in Table I. We use the shorthand $x = D\langle\Phi|X|\Phi\rangle$ and $y = \langle\Phi|\Phi\rangle$. The last two columns show the result of keeping the first three terms in the expansions Eqs. (B·1) and (B·2).

E[y]	E[x]	E[xy]	cov[x, y]	Var[y]	Var[x]	$E[\frac{x}{y}]$	$Var[\frac{x}{y}]$
1	**Tr** X	**Tr** X	0	0	$\frac{D\,\mathbf{Tr}\,XX^\dagger - \|\mathbf{Tr}\,X\|^2}{D+1}$	**Tr** X	$\frac{D\,\mathbf{Tr}\,XX^\dagger - \|\mathbf{Tr}\,X\|^2}{D+1}$
1	**Tr** X	$\frac{D+1}{D}\,\mathbf{Tr}\,X$	$\frac{1}{D}\,\mathbf{Tr}\,X$	$\frac{1}{D}$	**Tr** $XX^\dagger$	**Tr** X	$\frac{D\,\mathbf{Tr}\,XX^\dagger - \|\mathbf{Tr}\,X\|^2}{D}$
1	**Tr** X	**Tr** X	0	0	**Tr** $XX^\dagger - \sum_{i=1}^{D} |\langle i|X|i\rangle|^2$	**Tr** X	**Tr** $XX^\dagger - \sum_{i=1}^{D} |\langle i|X|i\rangle|^2$

The results of Eqs. (A·3), (A·4), and (A·5) are now readily obtained by calculating $\int_0^1 dz\, zp(z|D)$, $\int_0^1 dz\, z^2 p(z|D)$, and $\int_0^1 dz\, z_1 z_2 p(z_1, z_2|D)$, respectively.

If D is large and zD is not, we have

$$p(z|D) = (D-1)(1-z)^{D-2}$$

$$\approx D\left(1 - \frac{Dz}{D}\right)^D \approx e^{-Dz}, \tag{A·11}$$

which is known as Porter–Thomas distribution.[110] Note that for evaluating averages over the unit hypersphere there is no advantage of using the (approximate) Porter–Thomas distribution in place of the exact distribution Eq. (A·9).

With Eqs. (A·9) and (A·10), it is straightforward to compute the averages, denoted by $\langle.\rangle$, over all random states $|\Phi\rangle$. In general, we have $\langle F(p(j))\rangle = \int_0^1 dz\, F(z)p(z|D)$, $\langle F(p(j))G(p(k \neq j))\rangle = \int_0^1 \int_0^1 dz_1\, dz_2\, F(z_1)G(z_2)p(z_1 z_2|D)$, etc. Specifically we have

$$\frac{\sum_{j=0}^{D-1}\langle p^\mu(j)\log p(j)\rangle}{\sum_{j=0}^{D-1}\langle p^\mu(j)\rangle}$$

$$= \text{PolyGamma}(0, D+\mu) - \text{PolyGamma}(0, \mu+1)$$

$$\overset{D\to\infty}{\approx} \log D - \text{PolyGamma}(0, \mu+1)$$

$$= \begin{cases} \log D + \gamma, & \mu = 0 \\ \log D + \gamma - 1, & \mu = 1, \\ \log D + \gamma - 3/2, & \mu = 2 \end{cases} \tag{A·12}$$

where γ is Euler's constant. For large D, we have

$$\langle(p(j)\log p(j))^2\rangle - \langle p(j)\log p(j)\rangle^2$$

$$= \frac{\pi^2/3 + (\log D + \gamma)^2 - 4(\log D + \gamma) + 1}{D^2} + \mathcal{O}\left(\frac{1}{D^3}\right). \tag{A·13}$$

Appendix B: Approximate Treatment

The presence of the fraction in Eq. (7) makes it difficult to derive a generally valid expression for the average and the variance. However, generally valid approximations can be obtained through the use of the multivariate Taylor expansion for the average

$$E\left[\frac{x}{y}\right] \approx \frac{E[x]}{E[y]} - \frac{E[xy] - E[x]E[y]}{E^2[y]} + \frac{E[x]\text{Var}[y]}{E^3[y]} + \cdots, \tag{B·1}$$

and a similar expansion for the variance

$$\text{Var}\left[\frac{x}{y}\right] \approx \frac{\text{Var}[x]}{E^2[y]} - 2\frac{E[x](E[xy] - E[x]E[y])}{E^3[y]}$$

$$+ \frac{E^2[x]\text{Var}[y]}{E^4[y]} + \cdots. \tag{B·2}$$

where, x and y can be any functions of the random variables ξ.

In Sect. 2, we give explicit expression for the expectation values of x, y, etc., in terms of the moments that are listed in Table I. Table B·I summarizes the results of the calculations. It directly follows that up to the three terms of the expansions shown in Eq. (B·1), $E[x/y] = \mathbf{Tr}\,X$, for each of the choices listed in Table I and that the expressions for the variances agree with those given by Eq. (17).

Appendix C: Sampling over Random States

In the main text, we focus on the calculation of the trace by using only one random state. Here we consider the case that further averaging over $R > 1$ different random states is necessary to produce results with good statistics. This is necessary for small systems or at very low temperature.

We assume that samples of pairs of variables (x_i, y_i) for $i = 1, \ldots, R$ have been obtained from different independent realizations of a random variable ξ. Obviously, for $i \neq j$, $E[x_i y_j] = E[x_i]E[y_j] = E[x]E[y] = \mu_x \mu_y$ and the correlation vanishes. For $i = j$, $E[x_i y_i] \neq E[x]E[y] = \mu_x \mu_y$, i.e., there can be a non-zero correlation because x_i and y_i are obtained from the same realization of the random variable ξ. Furthermore, $E[x_i^2] = E[x^2]$ and $E[y_i^2] = E[y^2]$.

There are two ways to average over the R samples,[20,74] namely

$$M1 = \frac{1}{R}\sum_{i=1}^{R}\frac{x_i}{y_i}, \tag{C·1}$$

or

$$M2 = \frac{\sum_{i=1}^{R} x_i}{\sum_{i=1}^{R} y_i}. \tag{C·2}$$

Before we use Eqs. (B·1) and (B·2) to analyze the mean and variance for the two different ways of computing the average of the R samples, we first recall the well-known formulas for the ensemble mean and variance. We have

J. Phys. Soc. Jpn.
Downloaded from journals.jps.jp by Forschungszentrum Juelich GmbH on 01/19/21

J. Phys. Soc. Jpn. **90**, 012001 (2021) **Invited Review Papers** F. Jin et al.

$$E\left[\frac{1}{R}\sum_{i=1}^{R}x_i\right] = \frac{1}{R}\sum_{i=1}^{R}E[x_i] = \mu_x, \tag{C·3}$$

$$E\left[\frac{1}{R}\sum_{i=1}^{R}y_i\right] = \frac{1}{R}\sum_{i=1}^{R}E[y_i] = \mu_y, \tag{C·4}$$

$$Var\left[\frac{1}{R}\sum_{i=1}^{R}x_i\right] = E\left[\left(\frac{1}{R}\sum_{i=1}^{R}x_i\right)^2\right] - \mu_x^2 = E\left[\frac{1}{R^2}\sum_{i=1}^{R}x_i^2 + \frac{1}{R^2}\sum_{i\neq j}^{R}x_ix_j\right] - \mu_x^2$$

$$= \frac{1}{R^2}\sum_{i=1}^{R}E[x_i^2] + \frac{1}{R^2}\sum_{i\neq j}^{R}E[x_ix_j] - \mu_x^2 = \frac{1}{R}E[x^2] + \frac{R-1}{R}E^2[x] - \mu_x^2$$

$$= \frac{E[x^2] - \mu_x^2}{R} = \frac{Var[x]}{R}, \tag{C·5}$$

$$Var\left[\frac{1}{R}\sum_{i=1}^{R}y_i\right] = \frac{Var[y]}{R}, \tag{C·6}$$

$$Cov\left[\frac{1}{R}\sum_{i=1}^{R}x_i\frac{1}{R}\sum_{j=1}^{R}y_j\right] = E\left[\frac{1}{R}\sum_{i=1}^{R}x_i\frac{1}{R}\sum_{j=1}^{R}y_j\right] - E\left[\frac{1}{R}\sum_{i=1}^{R}x_i\right]E\left[\frac{1}{R}\sum_{j=1}^{R}y_j\right]$$

$$= E\left[\frac{1}{R^2}\sum_{i=1}^{R}x_iy_i + \frac{1}{R^2}\sum_{i\neq j}^{R}x_iy_j\right] - \mu_x\mu_y = \frac{1}{R}E[xy] + \frac{R-1}{R}E[x]E[y] - \mu_x\mu_y$$

$$= \frac{Cov[xy]}{R}. \tag{C·7}$$

Therefore, for the first way of averaging Eq. (C·1), we have

$$E[M1] = E\left[\frac{x}{y}\right] \approx \frac{E[x]}{E[y]} - \frac{E[xy] - E[x]E[y]}{E^2[y]} + \frac{E[x]Var[y]}{E^3[y]} + \cdots,$$

$$Var[M1] = \frac{Var\left[\frac{x}{y}\right]}{R}, \tag{C·8}$$

whereas for the second way of averaging Eq. (C·2), we obtain

$$E[M2] = E\left[\frac{\frac{1}{R}\sum_{i=1}^{R}x_i}{\frac{1}{R}\sum_{i=1}^{R}y_i}\right] \approx \frac{E[x]}{E[y]} - \frac{1}{R}\frac{E[xy] - E[x]E[y]}{E^2[y]} + \frac{1}{R}\frac{E[x]Var[y]}{E^3[y]} + \cdots,$$

$$Var[M2] = Var\left[\frac{\frac{1}{R}\sum_{i=1}^{R}x_i}{\frac{1}{R}\sum_{i=1}^{R}y_i}\right]$$

$$\approx \frac{1}{R}\frac{Var[x]}{E^2[y]} - 2\frac{1}{R}\frac{E[x](E[xy] - E[x]E[y])}{E^3[y]} + \frac{1}{R}\frac{E^2[x]Var[y]}{E^4[y]} + \cdots = \frac{Var\left[\frac{x}{y}\right]}{R}. \tag{C·9}$$

Comparing Eqs. (C·8) and (C·9), it is clear that the variances show the same $1/R$ dependence (as usual for independent samples). However, the presence of the factor $1/R$ in the second and third term of the expression for the mean Eq. (C·9) obtained using Eq. (C·2) implies that these correction terms vanish as $R \to \infty$, unlike in the case the mean Eq. (C·8) is computed according to Eq. (C·1). In other words, if possible, we should use Eq. (C·2) to compute the average of independent realizations of the random states.

Appendix D: Numerical Methods for Uniformly Picking Points from the Unit Sphere

D.1 Muller's method

Muller's method[121] for generating a vector **x** that is distributed uniformly on the $2D$-dimensional unit sphere consists of the following steps:

1. Using the Box–Muller method[24] to generate D pairs of real-valued pseudo-random numbers with a normal

(Gaussian) distribution and use these pairs to assign values to the elements of $\mathbf{x}$.

2. Compute the Euclidean norm $\|\mathbf{x}\|$ of $\mathbf{x}$ and replace $\mathbf{x}$ by $\mathbf{x}/\|\mathbf{x}\|$.

We now show that this procedure indeed generates points that are distributed uniformly over the $2D$-dimensional unit sphere.

Specifying an unnormalized vector in the D-dimensional Hilbert space requires $2D$ real numbers. In this subsection, we simplify the notation by introducing the symbol $d = 2D$. We start from the normalized Gaussian distribution

$$p(x_1, \ldots, x_d) = \frac{1}{\pi^{d/2}} e^{-x_1^2 - \cdots - x_d^2}, \tag{D·1}$$

and write it in spherical coordinates

$$
\begin{aligned}
x_1 &= r \cos\theta_1 \sin\theta_2 \sin\theta_3 \cdots \sin\theta_{d-1} \\
x_2 &= r \sin\theta_1 \sin\theta_2 \sin\theta_3 \cdots \sin\theta_{d-1} \\
x_3 &= \quad\quad r \cos\theta_2 \sin\theta_3 \cdots \sin\theta_{d-1} \\
x_4 &= \quad\quad\quad\quad r \cos\theta_3 \cdots \sin\theta_{d-1} \\
\cdots &= \cdots \\
x_{d-1} &= \quad\quad\quad\quad r \cos\theta_{d-2} \sin\theta_{d-1} \\
x_d &= \quad\quad\quad\quad\quad r \cos\theta_{d-1}, \tag{D·2}
\end{aligned}
$$

where $0 \le \theta_1 \le 2\pi$, $0 \le \theta_j \le \pi$ for $j > 1$, and the condition $x_1^2 + \cdots + x_d^2 = r^2$ is automatically satisfied. The Jacobian $|J|$ of this transformation reads[122]

$$|J| = r^{d-1} \sin^{d-2}\theta_{d-1} \sin^{d-3}\theta_{d-2} \cdots \sin\theta_2, \tag{D·3}$$

and we have

$$
\begin{aligned}
&p(r, \theta_1, \ldots, \theta_d) \\
&= \frac{1}{\pi^{d/2}} r^{d-1} e^{-r^2} \sin^{d-2}\theta_{d-1} \sin^{d-3}\theta_{d-2} \cdots \sin\theta_2. \tag{D·4}
\end{aligned}
$$

Using the identity

$$\int_0^\infty r^{d-1} e^{-r^2} \, dr = \frac{\Gamma(d/2)}{2}, \tag{D·5}$$

and $\int p(x_1, \ldots, x_d) dx_1 \cdots dx_d = 1$, we can write Eq. (D·4) as

$$p(r, \theta_1, \ldots, \theta_d) = p(r)p(\theta_1, \ldots, \theta_d), \tag{D·6}$$

where

$$p(r) = \frac{2}{\Gamma(d/2)} r^{d-1} e^{-r^2/2}, \tag{D·7}$$

and

$$
\begin{aligned}
&p(\theta_1, \ldots, \theta_n) \\
&= \frac{\Gamma(d/2)}{2\pi^{d/2}} \sin^{d-2}\theta_{d-1} \sin^{d-3}\theta_{d-2} \cdots \sin\theta_2. \tag{D·8}
\end{aligned}
$$

In these spherical coordinates, the infinitesimal volume element of the unit hypersphere reads

$$\Omega_d = \sin^{d-2}\theta_{d-1} \sin^{d-3}\theta_{d-2} \cdots \sin\theta_2 \, d\theta_{d-1} \cdots d\theta_2 \, d\theta_1. \tag{D·9}$$

As $p(x_1, \ldots, x_d)$ is a properly normalized density and $\int_0^\infty p(r) = 1$, if follows from Eq. (D·6) that the integral of $p(\theta_1, \ldots, \theta_d)$ over all angles must be equal to one. Integrating Eq. (D·8) over the surface of the d-dimensional unit hypersphere S_d we find that $\int_{S_d} d\Omega_d = 2\pi^{d/2}/\Gamma(d/2)$.

From Eqs. (D·6) and (D·8) it immediately follows that r and the θ's are all independent random numbers. Denoting $\|x\|^2 = x_1^2 + \cdots + x_d^2$, the probability density for picking a point $x = (x_1, \ldots, x_d)$ on a hypersphere of radius R is

$$
\begin{aligned}
\widehat{p}(R) &= \int \delta(\|x\| - R) p(x_1, \ldots, x_d) \, dx_1 \cdots dx_d \\
&= \int \delta(\|x\| - R) p(r) p(\theta_1, \ldots, \theta_d) \\
&= \int \delta(r - R) p(r) = \frac{2\pi^{d/2}}{\Gamma(d/2)} R^{d-1} e^{-d^2/2}, \tag{D·10}
\end{aligned}
$$

showing that because of the independence of r and the θ's, $\widehat{p}(R)$ is independent of the direction of the vector $x = (x_1, \ldots, x_n)$. Therefore, the distribution of points $x_i/\|x\|$ is uniform over the sphere.

D.2 An alternative to Muller's method

The following is based on material presented in Ref. 123, pages 111 and 182. In contrast to representation Eq. (D·2) where we use $2D$ spherical coordinates to encode pairs of real numbers that determine the amplitude of a basis state, the representation that we adopt in this section uses $D - 1$ spherical coordinates to encode the square root of the probabilities of the basis states and another D random numbers $\{\nu_1, \ldots, \nu_D\}$ to encode the phases. We represent the coefficients of the state vector by $\{x_1 e^{i\phi_1}, \ldots, x_D e^{i\phi_D}\}$ where for all $1 \le j \le D$, $0 \le \phi_j \le 2\pi$, $x_j \ge 0$ and $x_1^2 + \cdots + x_D^2 = r^2$. We will set $r = 1$ later. The x_j's and ϕ_j are called octant coordinates.[123] In terms of the spherical coordinates, we have

$$
\begin{aligned}
x_1 &= r \cos\theta_1 \sin\theta_2 \sin\theta_3 \cdots \sin\theta_{D-1} \\
x_2 &= r \sin\theta_1 \sin\theta_2 \sin\theta_3 \cdots \sin\theta_{D-1} \\
x_3 &= \quad\quad r \cos\theta_2 \sin\theta_3 \cdots \sin\theta_{D-1} \\
x_4 &= \quad\quad\quad\quad r \cos\theta_3 \cdots \sin\theta_{D-1} \\
\cdots &= \cdots \\
x_{D-1} &= \quad\quad\quad\quad r \cos\theta_{D-2} \sin\theta_{D-1} \\
x_D &= \quad\quad\quad\quad\quad r \cos\theta_{D-1}, \tag{D·11}
\end{aligned}
$$

where $0 \le \theta_j \le \pi/2$ and the conditions $0 \le x_j$ and $x_1^2 + \cdots + x_D^2 = r^2$ are automatically satisfied.

We can find the volume element of the $2D$-dimensional sphere in these coordinates as follows. First consider a single complex number $z = u + iv = x e^{i\phi}$ and compute the Jacobian for the transformation $(u, v) \to (x, \phi)$ to find that $du \, dv = x \, dx \, d\phi$. For D complex variables, the volume element in Cartesian coordinates is $du_1 \, dv_1 \cdots du_D \, dv_D = x_1 \cdots x_D \, dx_1 \cdots dx_D \, d\phi_1 \cdots d\phi_D$. Next, we use the spherical coordinates (restricted to the first octant) for $x_1, \ldots, x_D$, see Eq. (D·11). The Jacobian of this transformation is given by Eq. (D·3) with $n = D$. Collecting all sines and cosines we find

$$
\begin{aligned}
d\Omega_D &= r^{2D-1} \prod_{k=1}^{D-1} \cos\theta_k (\sin\theta_k)^{2k-1} \, d\theta_k \, d\phi_k \, d\phi_D \\
&= r^{2D-1} \prod_{k=1}^{D-1} y_k^{2k-1} \, dy_k \, d\phi_k \, d\phi_D \tag{D·12}
\end{aligned}
$$

where $y_k = \sin\theta_k$. Integrating Eq. (D·14) over $0 \le r \le R$, $0 \le y_k \le 1$, and $0 \le \phi_k \le 2\pi$ we find

J. Phys. Soc. Jpn.
Downloaded from journals.jps.jp by Forschungszentrum Juelich GmbH on 01/19/21

J. Phys. Soc. Jpn. **90**, 012001 (2021)   **Invited Review Papers**   F. Jin et al.

$$V_D = \int d\Omega_D = \frac{R^{2D}}{2D}\frac{(2\pi)^D}{2^{D-1}(D-1)!} = \frac{\pi^D R^{2D}}{D!}, \qquad \text{(D·13)}$$

which is the volume of a $2D$-dimensional sphere of radius R, as expected.

The probability density for the random variables $\{y_1, \ldots, y_{D-1}, \phi_1, \ldots, \phi_D\}$ reads

$$p(y_1, \ldots, y_{D-1}, \phi_1, \ldots, \phi_D) = \frac{1}{(2\pi)^D}\prod_{k=1}^{D-1}(2k)y_k^{2k-1}. \quad \text{(D·14)}$$

From Eq. (D·14), it is easy to find the probability density for any of the x's. For instance, the probability that x_D^2 is less than z is given by

$$
\begin{aligned}
P(x_D^2 < z) &= P(1 - y_{D-1}^2 < z) \\
&= \left(\prod_{k=1}^{D-2}\int_0^1 (2k)y_k^{2k-1}\,dy_k\right)\int_{\sqrt{1-z}}^1 2(D-1)y_{D-1}^{2D-3}\,dy_{D-1} \\
&= (D-1)\int_{1-z}^1 x^{D-2}\,dx, \qquad \text{(D·15)}
\end{aligned}
$$

from which, by differentiation with respect to z, we find $p(z|D) = (D-1)(1-z)^{D-2}$, see Eq. (A·9).

Unlike with the standard spherical coordinates, see Eq. (D·8), the expression Eq. (D·14) allows us to sample the angles θ_k's, or equivalently, the y_k's independently. The probability density and probability for y_k are given by

$$p(y_k) = 2ky^{2k-1}$$
$$P(y_k \le Y_k) = 2k\int_0^{Y_k} y^{2k-1}\,dy_k = Y_k^{2k}, \qquad \text{(D·16)}$$

respectively, from which it follows that in order to generate a random number Y_k with the correct distribution, we simply have to generate a uniform random variable r_k and put $Y_k = r_k^{1/2k}$. In practice, k can be very large and numerically, we should use $Y_k^2 = \exp(\log(r)/k) \approx 1 - \log(r)/k$ if $\log(r)/k$ is very small. For numerical purposes, we find that Muller's method is more convenient, in particular if D is large.

---

*Corresponding author. E-mail: deraedthans@gmail.com

1) E. P. Wigner, Ann. Math. **62**, 548 (1955).
2) O. Bohigas, M. J. Giannoni, and C. Schmit, Phys. Rev. Lett. **52**, 1 (1984).
3) N. J. Russell, L. Chakhmakhchyan, J. L. O'Brien, and A. Laing, New J. Phys. **19**, 033007 (2017).
4) S. Boixo, S. V. Isakov, R. Smelyanskiy, V. N. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, Nat. Phys. **14**, 595 (2018).
5) A. Edelman and N. R. Rao, Acta Numer. **14**, 233 (2005).
6) J. M. Deutsch, Phys. Rev. A **43**, 2046 (1991).
7) M. Srednicki, Phys. Rev. E **50**, 888 (1994).
8) M. Rigol, V. Dunjko, and M. Olshanii, Nature **452**, 854 (2008).
9) S. Goldstein, J. L. Lebowitz, R. Tumulka, and N. Zanghì, Phys. Rev. Lett. **96**, 050403 (2006).
10) S. Goldstein, J. L. Lebowitz, C. Mastrodonato, R. Tumulka, and N. Zanghì, Phys. Rev. E **81**, 011109 (2010).
11) S. Goldstein, T. Hara, and H. Tasaki, New J. Phys. **17**, 045002 (2015).
12) S. Popescu, A. J. Short, and A. Winter, Nat. Phys. **2**, 754 (2006).
13) P. Reimann, Phys. Rev. Lett. **99**, 160404 (2007).
14) P. Reimann, New J. Phys. **12**, 055027 (2010).
15) P. Reimann, Nat. Commun. **7**, 10821 (2016).
16) P. Bocchieri and A. Loinger, Phys. Rev. **114**, 948 (1959).
17) H. Tasaki, Phys. Rev. Lett. **80**, 1373 (1998).
18) J. Gemmer, M. Michel, and G. Mahler, *Quantum Thermodynamics: Emergence of Thermodynamic Behavior Within Composite Quantum Systems* (Springer, Berlin/Heidelberg, 2004).
19) S. Sugiura and A. Shimizu, Phys. Rev. Lett. **108**, 240401 (2012).
20) S. Sugiura and A. Shimizu, Phys. Rev. Lett. **111**, 010401 (2013).
21) S. Lloyd, arXiv:1307.0378.
22) J. H. Hammersley and D. C. Handscomb, *Monte Carlo Methods* (Wiley, New York, 1964).
23) D. P. Landau and K. Binder, *A Guide to Monte Carlo Simulation in Statistical Physics* (Cambridge University Press, Cambridge, U.K., 2000).
24) W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes* (Cambridge University Press, Cambridge, U.K., 2003).
25) J. Gemmer and G. Mahler, Eur. Phys. J. B **31**, 249 (2003).
26) C. Bartsch and J. Gemmer, Phys. Rev. Lett. **102**, 110403 (2009).
27) R. Steinigeweg, J. Gemmer, and W. Brenig, Phys. Rev. Lett. **112**, 120601 (2014).
28) R. Steinigeweg, J. Herbrych, F. Pollmann, and W. Brenig, Phys. Rev. B **94**, 180401 (2016).
29) R. Alben, M. Blume, H. Krakauer, and L. Schwartz, Phys. Rev. B **12**, 4090 (1975).
30) M. Imada and M. Takahashi, J. Phys. Soc. Jpn. **55**, 3354 (1986).
31) *Maximum Entropy and Bayesian Methods*, ed. J. Skilling (Springer, Cambridge, U.K., 1988) p. 455.
32) H. De Raedt and P. de Vries, Z. Phys. B **77**, 243 (1989).
33) P. de Vries and H. De Raedt, Phys. Rev. B **47**, 7929 (1993).
34) K. Fukamachi and H. Nishimori, Phys. Rev. B **49**, 651 (1994).
35) D. A. Drabold and O. F. Sankey, Phys. Rev. Lett. **70**, 3631 (1993).
36) R. N. Silver and H. Röder, Int. J. Mod. Phys. C **05**, 735 (1994).
37) R. N. Silver and H. Röder, Phys. Rev. E **56**, 4822 (1997).
38) T. Iitaka, S. Nomura, H. Hirayama, X. Zhao, Y. Aoyagi, and T. Sugano, Phys. Rev. E **56**, 1222 (1997).
39) S. Nomura, T. Iitaka, X. Zhao, T. Sugano, and Y. Aoyagi, Phys. Rev. B **56**, R4348 (1997).
40) S. Nomura, T. Iitaka, X. Zhao, T. Sugano, and Y. Aoyagi, Phys. Rev. B **59**, 10309 (1999).
41) H. Nishida, R. Fujiuchi, K. Sugimoto, and Y. Ohta, J. Phys. Soc. Jpn. **89**, 023702 (2020).
42) A. Hams and H. De Raedt, Phys. Rev. E **62**, 4365 (2000).
43) H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford University Press, Oxford, U.K., 2002).
44) S. Yuan, M. I. Katsnelson, and H. De Raedt, JETP Lett. **84**, 99 (2006).
45) S. Yuan, M. I. Katsnelson, and H. De Raedt, Phys. Rev. B **77**, 184301 (2008).
46) S. Yuan, M. I. Katsnelson, and H. De Raedt, J. Phys. Soc. Jpn. **78**, 094003 (2009).
47) S. Yuan, J. Comput. Theor. Nanosci. **8**, 889 (2011).
48) F. Jin, H. De Raedt, S. Yuan, M. I. Katsnelson, S. Miyashita, and K. Michielsen, J. Phys. Soc. Jpn. **79**, 124005 (2010).
49) H. De Raedt, B. Barbara, S. Miyashita, K. Michielsen, S. Bertaina, and S. Gambarelli, Phys. Rev. B **85**, 014408 (2012).
50) M. A. Novotny, F. Jin, S. Yuan, S. Miyashita, H. De Raedt, and K. Michielsen, Phys. Rev. A **93**, 032110 (2016).
51) P. Zhao, H. De Raedt, S. Miyashita, F. Jin, and K. Michielsen, Phys. Rev. E **94**, 022126 (2016).
52) H. De Raedt, F. Jin, M. I. Katsnelson, and K. Michielsen, Phys. Rev. E **96**, 053306 (2017).
53) D. Gelman and R. Kosloff, Chem. Phys. Lett. **381**, 129 (2003).
54) D. Gelman, C. P. Koch, and R. Kosloff, J. Chem. Phys. **121**, 661 (2004).
55) G. Katz, D. Gelman, M. A. Ratner, and R. Kosloff, J. Chem. Phys. **129**, 034108 (2008).
56) T. Monnai and A. Sugita, J. Phys. Soc. Jpn. **83**, 094001 (2014).
57) R. Steinigeweg, J. Herbrych, X. Zotos, and W. Brenig, Phys. Rev. Lett. **116**, 017202 (2016).
58) R. Steinigeweg, F. Jin, D. Schmidtke, H. De Raedt, K. Michielsen, and J. Gemmer, Phys. Rev. B **95**, 035155 (2017).
59) Y. Yamaji, T. Suzuki, and M. Kawamura, arXiv:1802.02854.
60) J. Richter, F. Jin, H. De Raedt, K. Michielsen, J. Gemmer, and R. Steinigeweg, Phys. Rev. B **97**, 174430 (2018).
61) J. Richter, F. Jin, L. Knipschild, J. Herbrych, H. De Raedt, K. Michielsen, J. Gemmer, and R. Steinigeweg, Phys. Rev. B **99**, 144422 (2019).
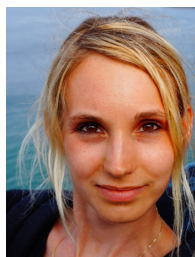
62) F. Jin, R. Steinigeweg, F. Heidrich-Meisner, K. Michielsen, and H. De Raedt, Phys. Rev. B **92**, 205103 (2015).

63) R. Steinigeweg, F. Jin, H. De Raedt, K. Michielsen, and J. Gemmer, Phys. Rev. E **96**, 020105 (2017).

64) S. Yuan, H. De Raedt, and M. I. Katsnelson, Phys. Rev. B **82**, 115448 (2010).

65) S. Yuan, H. De Raedt, and M. I. Katsnelson, Phys. Rev. B **82**, 235409 (2010).

66) S. Yuan, R. Roldán, M. I. Katsnelson, and F. Guinea, Phys. Rev. B **90**, 041402 (2014).

67) F. Jin, R. Roldán, M. I. Katsnelson, and S. Yuan, Phys. Rev. B **92**, 115440 (2015).

68) H. De Raedt, Comput. Phys. Rep. **7**, 1 (1987).

69) H. De Raedt and K. Michielsen, in *Handbook of Theoretical and Computational Nanotechnology*, ed. M. Rieth and W. Schommers (American Scientific Publishers, Los Angeles, 2006) p. 2.

70) H. Tal-Ezer and R. Kosloff, J. Chem. Phys. **81**, 3967 (1984).

71) V. V. Dobrovitski and H. De Raedt, Phys. Rev. E **67**, 056702 (2003).

72) A. Weiße, G. Wellein, A. Alvermann, and H. Fehske, Rev. Mod. Phys. **78**, 275 (2006).

73) P. Prelovšek, in *The Physics of Correlated Insulators, Metals, and Superconductors Modeling and Simulation*, ed. E. Pavarini, E. Koch, R. Scalettar, and R. Martin (Forschungszentrum Jülich, Jülich, 2017).

74) J. Schnack, J. Richter, and R. Steinigeweg, Phys. Rev. Res. **2**, 013186 (2020).

75) T. A. Elsayed and B. V. Fine, Phys. Rev. Lett. **110**, 070404 (2013).

76) Google AI Quantum and Collaborators, Nature **574**, 505 (2019).

77) M. A. Nielsen, Phys. Lett. A **303**, 249 (2002).

78) A. Gilchrist, N. K. Langford, and M. A. Nielsen, Phys. Rev. A **71**, 062310 (2005).

79) T. Iitaka and T. Ebisuzaki, Phys. Rev. E **69**, 057701 (2004).

80) G. R. Grimmet and D. R. Stirzaker, *Probability and Random Processes* (Clarendon Press, Oxford, U.K., 2001).

81) Web [https://jugit.fz-juelich.de/qip/random-state-technology] (2020).

82) S. Gershgorin, Bulletin de l'Académie des Sciences de l'URSS. Classe des sciences mathématiques et na **6**, 749 (1931).

83) C. Leforestier, R. H. Bisseling, C. Cerjan, M. D. Feit, R. Friesner, A. Guldberg, A. Hammerich, G. Jolicard, W. Karrlein, H.-D. Meyer, N. Lipkin, O. Roncero, and R. Kosloff, J. Comput. Phys. **94**, 59 (1991).

84) J. H. Wilkinson, *The Algebraic Eigenvalue Problem* (Clarendon Press, Oxford, U.K., 1965).

85) N. Elstner and A. P. Young, Phys. Rev. B **50**, 6871 (1994).

86) J. Schnack, J. Schulenburg, and J. Richter, Phys. Rev. B **98**, 094423 (2018).

87) P. Prelovšek and J. Kokalj, Phys. Rev. B **98**, 035107 (2018).

88) T. Nakamura and S. Miyashita, Phys. Rev. B **52**, 9174 (1995).

89) T. Shimokawa and H. Kawamura, J. Phys. Soc. Jpn. **85**, 113702 (2016).

90) J. K. Cullum and R. A. Willoughby, *Lanczos Algorithms for Large Symmetric Eigenvalue Computations* (Society for Industrial and Applied Mathematics, Philadelphia, USA, 2002).

91) T. Sakurai and H. Sugiura, J. Comput. Appl. Math. **159**, 119 (2003), 6th Japan-China Joint Seminar on Numerical Mathematics; In Search for the Frontier of Computational and Applied Mathematics toward the 21stCentury.

92) K. Morita and T. Tohyama, Phys. Rev. Res. **2**, 013205 (2020).

93) C. Karrasch, J. H. Bardarson, and J. E. Moore, Phys. Rev. Lett. **108**, 227206 (2012).

94) C. Karrasch, J. Hauschild, S. Langer, and F. Heidrich-Meisner, Phys. Rev. B **87**, 245128 (2013).

95) J. Richter and R. Steinigeweg, Phys. Rev. B **99**, 094419 (2019).

96) J. Richter, F. Jin, L. Knipschild, H. De Raedt, K. Michielsen, J. Gemmer, and R. Steinigeweg, arXiv:1906.09268.

97) C. Karrasch, J. E. Moore, and F. Heidrich-Meisner, Phys. Rev. B **89**, 075139 (2014).

98) R. Kubo, J. Phys. Soc. Jpn. **12**, 570 (1957).

99) S. Miyashita, T. Yoshino, and A. Ogasahara, J. Phys. Soc. Jpn. **68**, 655 (1999).

100) S. El Shawish, O. Cépas, and S. Miyashita, Phys. Rev. B **81**, 224421 (2010).

101) M. Oshikawa and I. Affleck, Phys. Rev. Lett. **82**, 5136 (1999).

102) H. Ikeuchi, H. De Raedt, S. Bertaina, and S. Miyashita, Phys. Rev. B **92**, 214431 (2015).

103) H. Ikeuchi, H. De Raedt, S. Bertaina, and S. Miyashita, Phys. Rev. B **95**, 024402 (2017).

104) J. Preskill, Quantum **2**, 79 (2018).

105) J. Emerson, R. Alicki, and K. Życzkowski, J. Opt. B **7**, S347 (2005).

106) E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Phys. Rev. A **77**, 012307 (2008).

107) E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, Cambridge, U.K., 2003).

108) M. Nielsen and I. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, U.K., 2000).

109) K. Michielsen, M. Nocon, D. Willsch, F. Jin, Th. Lippert, and H. De Raedt, Comput. Phys. Commun. **220**, 44 (2017).

110) C. Porter and R. Thomas, Phys. Rev. **104**, 483 (1956).

111) A. Zlokapa, S. Boixo, and D. Lidar, arXiv:2005.02464.

112) H. De Raedt, F. Jin, D. Willsch, M. Willsch, N. Yoshioka, N. Ito, S. Yuan, and K. Michielsen, Comput. Phys. Commun. **237**, 47 (2019).

113) K. De Raedt, K. Michielsen, H. De Raedt, B. Trieu, G. Arnold, M. Richter, Th. Lippert, H. Watanabe, and N. Ito, Comput. Phys. Commun. **176**, 121 (2007).

114) D. Willsch, H. Lagemann, M. Willsch, F. Jin, H. De Raedt, and K. Michielsen, in *NIC Symposium 2020*, ed. M. Müller, K. Binder, and A. Trautmann (John von Neumann Institute for Computing, Jülich, 2020) NIC Series, Vol. 50, p. 255.

115) Y. Zhou, E. M. Stoudenmire, and X. Waintal, arXiv:quant-ph/2002.07730.

116) R. Jozsa, J. Mod. Opt. **41**, 2315 (1994).

117) B. Schumacher, Phys. Rev. A **54**, 2614 (1996).

118) M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. A **60**, 1888 (1999).

119) D. Krause and P. Thörnig, J. Large-Scale Res. Facil. **4**, A132 (2018).

120) N. Ullah, Nucl. Phys. **58**, 65 (1964).

121) M. E. Muller, Commun. ACM **2** [4], 19 (1959).

122) K. Miller, *Multidimensional Gaussian Distributions* (Wiley, New York, 1964).

123) I. Bengtsson and K. Życzkowski, *Geometry of Quantum States* (Cambridge University Press, Cambridge, U.K., 2006).

Fengping Jin (ORCID: 0000-0003-3476-524X), born in Zhejiang (P. R. China), studied physics at Zhejiang University, China, and received his Ph.D. in computational physics at the University of Groningen, the Netherlands, in 2011. Since then, he is a member of the group of quantum information processing at Jülich Supercomputing Centre, Forschungszentrum Jülich, Germany. His expertise is in the simulation of classical and quantum dynamics. His current research interests focus on quantum computation, quantum annealing, and decoherence and equilibration of quantum systems in general.

Dennis Willsch (ORCID: 0000-0003-3855-5100), born in Cologne (Germany), received the Ph.D. degree in physics from RWTH Aachen University (Germany) in 2020 for simulating the time evolution of transmon quantum computers on digital supercomputers. He is a postdoctoral researcher in the Quantum Information Processing group at the Jülich Supercomputing Centre, Forschungszentrum Jülich (Germany). His current research interests include modelling quantum technology experiments, simulating the time evolution of quantum information processors, and benchmarking existing quantum computers and quantum annealers.

Madita Willsch (ORCID: 0000-0002-2351-3162), born in Leverkusen (Germany), received her Ph.D. in physics from RWTH Aachen University (Germany) in 2020 for simulating the real-time dynamics of superconducting flux qubits during quantum annealing. Currently, she is a postdoctoral researcher in the Quantum Information Processing group at the Jülich Supercomputing Centre, Forschungszentrum Jülich (Germany). Her research interests include quantum computation, quantum annealing and simulations of quantum systems.

Hannes Lagenmann (ORCID: 0000-0002-4233-1874), born in Lingen (Ems, Germany), is currently a doctoral candidate (Ph.D. student) at the Institute for Advanced Simulation, Jülich Supercomputing Centre, Forschungszentrum Jülich (Germany). His research focuses on simulating superconducting, flux-tunable transmon qubits that are being used in various experimental, gate-based quantum computer architectures.

Kristel Michielsen (ORCID: 0000-0003-1444-4262), born in Antwerp (Belgium), received her Ph.D. from the University of Groningen (Netherlands), for work on the simulation of strongly correlated electron systems in 1993. Since 2009 she is group leader of the research group Quantum Information Processing at the Jülich Supercomputing Centre, Forschungszentrum Jülich (Germany) and is also Professor of Quantum Information Processing at RWTH Aachen University (Germany). Her current research interests include quantum computation, quantum annealing, quantum statistical physics, event-based simulation methods of quantum phenomena, logical inference approach to quantum mechanics and computational electrodynamics.

In 1976, **Hans De Raedt** (ORCID: 0000-0001-8461-4015), born in Antwerp (Belgium), received the Ph.D. degree from the University of Antwerp, Belgium, for work on magnetism in one dimension. Since 1990 he is (emer.) Professor of Computational Physics at the Zernike Institute for Advanced Materials, University of Groningen (Netherlands). Currently he is also a visiting scientist at the Institute for Advanced Simulation, Jülich Supercomputing Centre, Forschungszentrum Jülich (Germany). After contributing to the early development of quantum Monte Carlo methods, his interest shifted towards the simulation of physical phenomena in the time domain, including computational electrodynamics with emphasis on applications to optics, the simulation of physical models for quantum computers, the dynamic properties of nanoscale magnets and time-dependent quantum phenomena in general. His recent work focuses on foundational issues of quantum physics, in particular on the development of event-based simulation methods of quantum phenomena and on establishing a bridge between objective knowledge gathered through simulation or experiment and their description in terms of concepts.