Quantum algorithms for the computation of quantum thermal averages at work

Riccardo Aiudi,[1,*] Claudio Bonanno,[2,†] Claudio Bonati,[3,‡] Giuseppe Clemente,[4,3,§] Massimo D'Elia,[3,‖] Lorenzo Maio,[3,5,¶] Davide Rossini,[3,**] Salvatore Tirone,[6,††] and Kevin Zambello[3,‡‡]

[1]*Dipartimento di Fisica, Università di Parma, Parco Area delle Scienze 7 /A, 43100 Parma, Italy*
[2]*Instituto de Física Teórica UAM-CSIC, c/ Nicolás Cabrera 13-15, Universidad Autónoma de Madrid, Cantoblanco, E-28049 Madrid, Spain*
[3]*Dipartimento di Fisica dell'Università di Pisa and INFN—Sezione di Pisa, Largo Pontecorvo 3, I-56127 Pisa, Italy*
[4]*Deutsches Elektronen-Synchrotron (DESY), Platanenallee 6, 15738 Zeuthen, Germany*
[5]*Aix Marseille Univ, Université de Toulon, CNRS, CPT, Marseille, France*
[6]*Scuola Normale Superiore, I-56126 Pisa, Italy*

(Received 29 August 2023; accepted 29 November 2023; published 26 December 2023)

Recently, a variety of quantum algorithms have been devised to estimate thermal averages on a genuine quantum processor. In this paper, we consider the practical implementation of the so-called Quantum-Quantum Metropolis Algorithm. As a test bed for this purpose, we simulate a basic system of three frustrated quantum spins and discuss its systematics, also in comparison with the Quantum Metropolis Sampling algorithm.

DOI: 10.1103/PhysRevD.108.114509

I. INTRODUCTION

The advent of quantum computation is expected to lead to breakthroughs in various fields of computational science [1–3]. In fact, it may disclose novel pathways to the solution of notable unsolved questions lying at the basis of classically intractable problems, from quantum chemistry to condensed-matter and high-energy physics [4,5]. One such example deals with the physics of fundamental interactions, in particular, when considering the strongly coupled regime, which is not treatable by perturbative analytical tools. Classical computational schemes, based on a discretized path-integral formulation, are indeed known to face hard, yet unsolved difficulties. This happens, for instance, when considering real-time processes and non-equilibrium physics, or even in the equilibrium case when the path-integral measure is not positive defined (as it happens at finite baryon density), a fact which prevents the application of classical Monte Carlo algorithms. Such an algorithmic obstruction is, for example, the main reason for our incomplete knowledge of the QCD phase diagram and of the physics of strongly interacting matter at finite density, which is required for the investigation of compact astrophysical objects [6,7].

In the case of equilibrium physics, one needs to devise quantum algorithms capable of efficiently exploring the Gibbs ensemble of the target quantum system. At present, the availability of quantum resources adequate for the numerical investigation of systems of direct physical interest, such as QCD, is still far from being achieved. Nevertheless, a variety of candidate quantum algorithms have already been proposed, either by directly computing observables on the (mixed) thermal state [8–12] or by preparing an ensemble of pure states, sampled with the proper thermal distribution [13–25]. At the present stage, it is thus important to investigate how the practical implementation of such algorithms works in simplified models, in order to better understand their systematics and pave the way to future and more realistic applications.

In Ref. [26], some of the present authors have already focused on the so-called Quantum Metropolis Sampling (QMS) algorithm [13], applied to a simple frustrated system made up of three quantum spins, which presents a sign problem when formulated in the path-integral approach. The QMS algorithm is based on a quantum Metropolis step, by which one can implement a quantum Markov chain across the Hamiltonian eigenstates of the system, which is capable of correctly sampling, after a proper thermalization time, the quantum eigenstates with

*riccardo.aiudi@unipr.it
†claudio.bonanno@csic.es
‡claudio.bonati@unipi.it
§giuseppe.clemente@desy.de
‖massimo.delia@unipi.it
¶lorenzo.maio@cpt.univ-mrs.fr
**davide.rossini@unipi.it
††salvatore.tirone@sns.it
‡‡kevin.zambello@pi.infn.it

the desired ensemble probability. In this paper, we also consider the alternative Quantum-Quantum Metropolis Algorithm (Q^2MA) [14], which is based on a quite different strategy. In a few words, the idea is to search for a pure quantum state, the so-called coherent encoding of the thermal state (CETS), with the property that a measurement of the Hamiltonian on such a state returns a given eigenstate with the correct Gibbs distribution. The search follows a Grover-like quantum approach, hence the double "Quantum" in the name of the algorithm, which therefore, at least in principle, promises a more effective quantum advantage.

A fundamental ingredient for both approaches is the Quantum Phase Estimation (QPE) algorithm [27–30], which allows one to estimate the energy eigenvalues once the Hamiltonian has been properly encoded in the quantum computer. However, the overall strategy used in the two methods is quite different: In the QMS algorithm the distribution is sampled via the Metropolis step, while in the Q^2MA algorithm it is encoded in the superposition amplitudes of a pure state, which contains information about the finite-temperature density matrix of the system, once the auxiliary registers are traced out.

We are not aware of practical implementations and benchmarks of the Q^2MA algorithm in concrete examples, and the main purpose of the present study is to fill this gap by considering the same target system as in Ref. [26]. As we will discuss in more detail later, there are some aspects of the algorithm which make its practical implementation nontrivial.

Another issue regards the numerical efficiency, measured in terms of the number of quantum gates needed to reach a given uncertainty on the final determination of the quantum thermal averages, for which we present a preliminary comparison between the two algorithms. In carrying out such an analysis, particular attention is given to both statistical and systematic contributions to the final error budget.

The paper is organized as follows. In Sec. II, we first review the basic features of the QMS and the Q^2MA algorithms, and then comment on the possible sources of systematical errors. In Sec. III, we introduce the specific quantum spin system tested here and all the metrics used to benchmark the systematical errors. In Sec. IV, we present the results of our investigation, first assuming an exact encoding of the energy levels and then relaxing such constraint. Finally, our conclusions are drawn in Sec. V.

II. ALGORITHMS

The main focus of the present study is on the Q^2MA algorithm since a practical implementation of the QMS algorithm has already been presented and discussed in Ref. [26]. However, we find it useful to give a brief overview of both algorithms, in order to clarify their computational requirements and to better identify the possible sources of systematical errors in both cases. A summary of the resources and gate counts is reported in the Appendix for both algorithms.

A. Quantum Metropolis Sampling

The QMS algorithm follows quite closely the scheme of the classical Metropolis algorithm (see Refs. [13,26] for further details): A Markov chain is built in such a way that, at each step, one gets an eigenstate $|\varphi_k\rangle$ of the Hamiltonian of the system under study, with eigenvalue E_k. This is selected with a probability given (asymptotically after thermalization) by its Boltzmann weight $e^{-\beta E_k}/Z(\beta)$, where $\beta = 1/(k_B T)$ is the inverse temperature and $Z(\beta)$ is the partition function.

Four registers are required: a register encoding the state of the system (denoted by the subscript 1), two energy registers encoding the energies before (denoted by 2) and after (denoted by 3) the Metropolis step, if accepted, and finally a single-qubit acceptance register (denoted by 4). Thus, the layout of the quantum state needed by the QMS algorithm is the following:

$$|\mathrm{acc}\rangle_4 |E_{\mathrm{new}}\rangle_3 |E_{\mathrm{old}}\rangle_2 |\mathrm{state}\rangle_1. \tag{1}$$

The first step of the Markov chain is the initialization of the system state in register 1 to an arbitrary eigenstate of the Hamiltonian (and of register 2 to the corresponding eigenvalue). If the quantum state of the QMS algorithm has been prepared in the initial state $|0\rangle_4|0\rangle_3|0\rangle_2|0\rangle_1$, this can be realized by using a QPE between registers 1 and 2, followed by a measure of register 2:

$$\begin{aligned} |0\rangle_2|0\rangle_1 &\xrightarrow{\mathrm{QPE}_{1,2}} \sum_{k'} \alpha_{k'} |E_{k'}\rangle_2 |\varphi_{k'}\rangle_1 \\ &\xrightarrow{\mathrm{Meas}_2} |E_k\rangle_2 |\varphi_k\rangle_1. \end{aligned} \tag{2}$$

Registers 3 and 4 are unmodified in this initial step.

A single step of the Markov chain involves an appropriate generalization of the Metropolis accept/reject algorithm [31] to the quantum case. In order to update the state, we apply to register 1 a unitary operator C randomly selected from a set $\mathcal{C}$; this set has to be large enough to ensure mixing between all eigenstates (ergodicity) and that if $A \in \mathcal{C}$ then also $A^{-1} \in \mathcal{C}$ (reversibility). Apart from these general requirements, there is still much freedom in the choice of the operators entering the set $\mathcal{C}$, freedom that can eventually be used to optimize the algorithm. Thus

$$|\varphi_k\rangle_1 \xrightarrow{C \in \mathcal{C}} \sum_p x^{(C)}_{k,p} |\varphi_p\rangle_1. \tag{3}$$

At this point, we perform a second QPE between the register of the system (labeled as 1) and the new energy register (labeled as 3):

$$\sum_p x_{k,p}^{(C)}|0\rangle_4|0\rangle_3|E_k\rangle_2|\varphi_p\rangle_1 \xrightarrow{\mathrm{QPE}_{1,3}} \sum_p x_{k,p}^{(C)}|0\rangle_4|E_p\rangle_3|E_k\rangle_2|\varphi_p\rangle_1. \quad (4)$$

To introduce the information about the Boltzmann weights, one applies an oracle operator G which reads out the difference between the two energy registers and acts on the acceptance qubit as follows:

$$|0\rangle_4|E_p\rangle_3|E_k\rangle_2|\psi_p\rangle_1 \xrightarrow{G} \left(\sqrt{f_{p,k}}|1\rangle_4+\sqrt{1-f_{p,k}}|0\rangle_4\right)\otimes|E_p\rangle_3|E_k\rangle_2|\psi_p\rangle_1, \quad (5)$$

where

$$f_{p,k}=\min\left(1,e^{-\beta(E_p-E_k)}\right) \quad (6)$$

is the usual Metropolis acceptance probability. At this stage, one performs a measurement in the acceptance register 4, which can only have two outcomes: the value 1 with probability $\sum_p |x_{k,p}^{(C)}|^2 f_{p,k}$ and the value 0 with the complementary probability. The first case corresponds to the "accepted" move, and the resulting state will be a superposition of eigenstates of the form

$$\sum_p x_{k,p}^{(C)}\sqrt{f_{p,k}}|1\rangle_4|E_p\rangle_3|E_k\rangle_2|\varphi_p\rangle_1. \quad (7)$$

The new configuration of the Markov chain (and thus the corresponding eigenstate in register 1) is obtained by measuring the new energy register 3. From this configuration, the update procedure can be iterated by applying a new randomly chosen unitary operator from set $\mathcal{C}$. If, on the other hand, the outcome of the measure on the acceptance register is 0, one needs to revert the system to an eigenstate with the same energy as the one that was previously present.[1] This reversal operation can be performed by applying the previous sequence of unitary operators (i.e., QPE and C) in reverse, followed by a measurement on the energy register until the energy measured matches E_k (see Refs. [13,26] for further details).

The above description of the QMS follows the original paper [13], where two different energy registers were used. We should, however, stress that it is possible to use just a single energy register: Since E_{old} is measured only at the beginning of each MC step, its value can be stored in a classical register to be used later by the oracle G in Eq. (5), thus halving the number of qubits needed to represent the energies.

[1]There is no need for the system state to be reverted exactly to the same state that was present before the application of the unitary operator C because this kind of rejection can be considered as a microcanonical step in the classical Metropolis algorithm.

B. Quantum-Quantum Metropolis Algorithm

The goal of the Q^{2}MA [14] is to build a CETS containing all the information on the Gibbs ensemble in the entanglement between the quantum registers. Its explicit form can be written as

$$|\alpha_0(\beta)\rangle=\sum_i \sqrt{e^{-\beta E_i}/Z(\beta)}\,|\varphi_i\rangle|\tilde{\varphi}_i\rangle|0\rangle, \quad (8)$$

where $|\varphi_i\rangle$ denotes, as in the previous section, the Hamiltonian eigenstate with eigenvalue E_i, $|\tilde{\varphi}_i\rangle$ is its complex conjugate copy, while $|0\rangle$ is an ancillary register needed for an operation analogous to the one of Eq. (5). There are two reasons for the presence of the complex conjugate copy of the system state: building the system's density matrix and computing energy differences through a QPE routine. The reason for the suffix 0 of α_0 will become clear in the following. The state $|\alpha_0(\beta)\rangle$ is (apart from the irrelevant ancilla $|0\rangle$) the purified form of the system's density matrix

$$\rho(\beta)=\frac{1}{Z(\beta)}\sum_i e^{-\beta E_i}|\varphi_i\rangle\langle\varphi_i|. \quad (9)$$

The core of the Q^{2}MA algorithm resides in the construction of the so-called generalized Szegedy operator W, which is described in Ref. [14]. A fundamental ingredient for this construction is the so-called "kick" operator K, which is a unitary operator that is symmetric in the computational basis. The matrix elements of K between eigenstates of the quantum Hamiltonian correspond to an *a priori* selection probability that, together with the Metropolis filter, can be used to sample the energy eigenstates by a Markov chain which has CETS as the invariant distribution (corresponding to the eigenvalue 1 of the Markov chain stochastic matrix). This procedure is very general, however; until an operator K is specified for a given quantum Hamiltonian, it cannot be shown that the corresponding Markov chain is ergodic and thus that the CETS is the *unique* eigenvector with eigenvalue 1 of the Szegedy operator. Let us assume for the moment that this is the case; we will come back to this point at the end of this section.

At this point, it is important to recall what the main conceptual differences are between the QMS and the Q^{2}MA algorithm, which have already been illustrated in Ref. [14]. The QMS is a quantum algorithm in the sense that its purpose is to exploit a quantum computing device to implement a Markov chain within the Hilbert space of a given quantum system. Apart from this, the main conceptual scheme is that of classical Markov chains, with additional limitations related to the no-cloning theorem.

On the other hand, the conceptual scheme of the Q^{2}MA is typical of quantum searching algorithms attaining a quadratic advantage, the searched state being the CETS, hence the double "quantum" in the name. In fact, it is not by chance that the Szegedy operator W is built by means of a Grover-like reflection algorithm, which requires the computation of the eigenvalues and eigenstates of the Hamiltonian, performed by means of a QPE, as in the QMS. Moreover, an operation analogous to the one in Eq. (5) is required, for which a single-qubit dedicated register has to be used.

Thus, the layout of the quantum state needed by Q^{2}MA is the following:

$$|w\rangle_5|\mathrm{acc}\rangle_4|\Delta E\rangle_3|\mathrm{state}\rangle_2|\widetilde{\mathrm{state}}\rangle_1, \tag{10}$$

where registers 2 and 1 are used to store the system state and its complex conjugate, respectively; register 3 is used in the QPE of the energy differences, which can be directly computed thanks to the presence of the complex conjugate copy of the system; and register 4 is used by an oracle analogous to that in Eq. (5).

Once the generalized Szegedy operator (obviously depending on β) is constructed, it can be used as the time evolution of a QPE, storing the phases in register 5, on which a classical measure is finally performed. If the outcome of this measure is 0 (which corresponds to the eigenvalue 1 of W), the input state is projected onto the CETS $|\alpha_0(\beta)\rangle$ (up to systematical errors due, e.g., to the finite number of qubits adopted in the QPE) and can thus be used to estimate observables. If, on the contrary, the measure returns a nonvanishing result, the state has to be rejected, and one should restart the algorithm.

If the Q^{2}MA procedure is successful, the registers $|w\rangle_5$ and $|\mathrm{acc}\rangle_4$ in the final state are both equal to $|0\rangle$, and also $|\Delta E\rangle_3$ is in $|0\rangle$ since in the construction of the Szegedy operator W, both a QPE in energy and its inverse are applied. The entire procedure can be formally thought of as the application of the projector

$$\Pi(\beta) = |\alpha_0(\beta)\rangle\langle\alpha_0(\beta)| \otimes |0\rangle\langle 0|_5 + \left[\sum_{k\neq 0}|\alpha_k(\beta)\rangle\langle\alpha_k(\beta)|\right] \otimes [\mathbb{1}_5 - |0\rangle\langle 0|_5] \tag{11}$$

to the initial system state, where the $|\alpha_k\rangle$ are the eigenstates of the Szegedy operator W (and $|\alpha_0\rangle$ is the CETS), and the subscript 5 refers to the ancilla register in which the phase of the W-generated QPE is stored.

To increase the probability of measuring the eigenphase 0 in the QPE using the Szegedy operator W (and to reduce the effects of the systematics that will be discussed in the following), it is possible to exploit two facts: The first one is that CETS states with close temperatures have a good overlap, which differs from 1 by a quantity of the order of $(\Delta\beta)^2$ [14]. The second fact is that, in the infinite temperature limit ($\beta = 0$), the CETS is formally equivalent to the maximally entangled state in the computational basis [14], which can be easily prepared with a combination of Hadamard and C-NOT gates. To obtain, with high probability, the CETS at the desired inverse temperature β, we can thus resort to Quantum Simulated Annealing (QSA) [32] by creating a sequence of CETS starting from the analytically known one at $\beta = 0$, and then lowering the temperature in steps of $\Delta\beta = \beta/n_a$, where n_a is the length of the annealing sequence:

$$|\alpha_0^0\rangle \xrightarrow{\Pi_1} |\alpha_0^1\rangle \xrightarrow{\Pi_2} \cdots \xrightarrow{\Pi_{n_a}} |\alpha_0^{n_a}\rangle. \tag{12}$$

In this equation $|\alpha_0^j\rangle$ is the CETS at $\beta_j = j\beta/n_a$ (with $j \in \{0, \ldots, n_a\}$) and $\Pi_j \equiv \Pi(\beta_j)$.

Therefore, the Q^{2}MA can be summarized as follows:

(1) Start at $j = 0$ with the maximally entangled state in the computational basis and initialize all ancilla registers to zero.
(2) Compute the Szegedy operator W corresponding to $\beta_j = j\beta/n_a$ and use it to perform a QPE on the ancilla register 5.
(3) Perform a classical measurement on $|w\rangle_5$; if the result is 0, then proceed to the next step (the state has been correctly projected into the CETS at β_j); otherwise, reset all quantum registers and restart from step 1.
(4) Iterate steps 2 and 3 with $j \to j+1$ until $j+1 = n_a$.

At the end of the algorithm, we obtain the final state

$$|0\rangle_5|0\rangle_4|0\rangle_3 \sum_i \sqrt{e^{-\beta E_i}/Z(\beta)}\,|\varphi_i\rangle_2|\tilde{\varphi}_i\rangle_1, \tag{13}$$

which is equivalent to Eq. (9) as far as the probability of selecting a state with energy E_i is concerned, i.e., once all the ancilla registers 3–5 have been traced out. In practice, to measure observables on the CETS, one performs a QPE using the auxiliary register $|\Delta E\rangle_3$, followed by a measure on the same register, in order to extract with the correct probability energy eigenstates on which to perform the measure. Finally, all quantum registers are reset and the algorithm is restarted, preparing the CETS for another measurement.

We now come back to the choice of the kick operator K. To the best of our knowledge, this point has never been fully addressed in the literature, and an operator satisfying all the "optimal" requirements is assumed to exist and to have been selected in the discussion of the algorithm. However, the selection of K is far from trivial since such an operator has to generate an ergodic selection probability in the basis of the Hamiltonian eigenstates, which is obviously unknown for nontrivial problems.

If K does not generate an ergodic Markov chain, the eigenvalue 1 of the stochastic matrix associated with the Markov chain (and thus of the Szegedy operator W) can have degeneracy larger than 1. This implies that the projection on the phase zero of the Szegedy-generated QPE does not ensure the selection of the CETS. An operative way to eliminate this problem, or at least to reduce its consequences, is to use different kick operators in the different annealing steps $j = 1, \ldots, n_a$, thus projecting at step j on the eigenspace corresponding to the eigenvalue 1 of the Szegedy operator W_j built using the kick operator K_j. Since the CETS always corresponds to an eigenstate with eigenvalue 1 of the Szegedy operator for any K_j, in this way we expect to correct the ergodicity problem of the single-kick naive implementation of the Q^2MA.

Let us stress that, in principle, one should require all the kick operators, and the corresponding Szegedy operators, to be considered at the same time for each annealing step, in order to guarantee that CETS is the only possible outcome in the case of acceptance. However, the implementation in which different single kick operators are used in different annealing steps is expected to work well in practice, at least as long as there is a large overlap between CETS corresponding to different annealing steps, i.e., as long as the annealing procedure is slow enough. In particular, it is reasonable to guess that the convergence of the annealing process scales as $1/n_a$, provided that n_a is larger than the minimum number of Szegedy operators required to univocally identify the CETS. Since the aim of the algorithm is to generate a (large) sample of measures from which to extract averages and standard errors, the operators K_j could also be generated stochastically, as long as the CETS is most likely identified (i.e., with probability 1).

C. Sources of systematical errors

Several sources of systematical errors exist both in the QMS and in the Q^2MA simulation schemes. In the following, we try to understand their impact on the simulation results in a controlled setting by using a simple toy model to separately investigate the different contributions.

To avoid overcomplicating our analysis, and since here the focus is on the quantum algorithms themselves, we consider a basic spin model, for which no digitization error is present, unlike more complex systems, like QCD and systems with continuous gauge symmetries.

It should be clear that a common source of systematical error, in both QMS and Q^2MA, is also the digitization error of the energies used in the QPE performed by using the system Hamiltonian. In the QMS this step is required by the oracle of Eq. (5), while in the Q^2MA this step is hidden in the construction of the generalized Szegedy operator [14]. Note that problems related to the accuracy of the energy representation adopted are present virtually in any importance sampling Monte Carlo computation (also classical ones), although this issue is usually overlooked in most practical computations [33,34]. Since our aim is to test the effectiveness of quantum algorithms in computing thermal averages, the sensitivity to the energy digitization is a relevant property to be investigated. For the model we study (as for all two-level systems), it is, however, possible to find an encoding in which energies have no digitization error. This is obviously impossible for generic systems with incommensurate energy levels, but it allows us to focus on other sources of systematical error which are instead intrinsic to the algorithms we are studying (i.e., not simply inherited by the energy QPE). In the following, when not specified, we will always assume that such an exact encoding of the energy is being used. We will investigate what happens when such a constraint is relaxed in a later section.

The systematical error that is specific to the QMS, and that is avoided by the Q^2MA, is the one related to the thermalization time of the algorithm: Hamiltonian eigenstates are sampled with the Boltzmann statistics only asymptotically for large times, i.e., after many iterations. The leading correction to this asymptotic behavior scales as $\exp[-(1-\lambda_2)t]$, where t is the number of Monte Carlo steps (the so-called Monte Carlo time) and λ_2 is the second-largest eigenvalue of the Markov chain transition matrix, the largest eigenvalue being 1 by construction. In a classical Monte Carlo sampling, nothing but the computational power prevents us from using very long Markov chains, and the algorithm is thus stochastically exact since typical statistical errors scale as $1/\sqrt{t}$. This is also the case for the QMS only if one is interested in the thermal averages of observables compatible with the Hamiltonian. If instead the aim is to compute $\langle A \rangle$, with an operator A which does not commute with the Hamiltonian, the Markov chain breaks down: By measuring A on the energy eigenstate $|\varphi_k\rangle$ extracted by the QMS, the state is projected to an eigenvector of A, which is generically a linear combination of different energy eigenstates. Therefore, after such a measure, the QMS needs to be reinitialized. The number of updates performed between subsequent measures of A thus puts an upper bound on the attainable accuracy.

In the Q^2MA, QPE is used not only with the time evolution generated by the Hamiltonian but also with the unitary Szegedy operator. In this case, the typical systematical error is the digitization error on the phase of this evolution, related to the number of qubits used for the register $|w\rangle_5$ in Eq. (10). We noted before that, for simple enough systems, it is possible to find an exact encoding of the energies, thus removing the systematic associated to the Hamiltonian-related QPE. An analogous procedure is generically not possible for the QPE with the Szegedy operator W as time evolution since W and its eigenvalues depend on the (inverse) temperature value β.

Analogously as for the QMS, which becomes exact in the large-time limit, the Q^2MA is also exact in the adiabatic

limit of infinite annealing steps, provided that the single kick K generates an ergodic chain. Indeed, if we denote by ϵ the magnitude of the error introduced on the state by the inaccuracy of the Szegedy QPE, the global error at each step of the annealing procedure is of the order of $\epsilon(\Delta\beta)^2 = \epsilon\beta^2/n_a^2$, and the final error is thus $O(\epsilon\beta^2/n_a)$ [14]. Note that if the Szegedy QPEs were performed exactly (i.e., $\epsilon = 0$), the algorithm would be exact for any number of annealing steps n_a; however, the success probability of step 3 in Sec. II B would be extremely small if n_a is not large enough.

When different kick operators K_i are used to effectively restore ergodicity (see the discussion at the end of Sec. II B), this is *a priori* no longer true: The eigenspace corresponding to the eigenvalue 1 of each Szegedy operator W_i has dimensionality larger than 1, so the global error at each step of the annealing procedure can be, at least in principle, of the order of $\epsilon\Delta\beta = \epsilon\beta/n_a$. That would make the final error independent of the number of annealing steps; i.e., the quantum annealing would just increase the success probability of the CETS generation, with no effect on the systematical error, so that the only way of removing systematics from the final estimate would be to reduce ϵ, i.e., to reduce the inaccuracy of the Szegedy QPE.

Contrary to such expectations, as we will show in the following, the systematic errors of the Q^{2}MA algorithm appear, instead, to depend on the number of annealing steps, as if a single ergodic kick operator was used. A possible interpretation is that, once the system collapses onto the correct CETS during the annealing procedure (actually one surely starts from the correct CETS at $\beta = 0$), it is highly probable to stay on CETS due to the good overlap of CETS states at different annealing steps, thus making single-step errors of $O(\Delta\beta^2)$ again.

III. MODEL AND METRICS

The first part of this section is dedicated to a description of the particular quantum system used as a test bed for our analysis. Then, we introduce some quantities that will be used to quantify the systematical errors of the explored quantum algorithms.

A. Frustrated triangle

To compare the QMS and Q^{2}MA algorithms described in the previous sections, we consider a system of three quantum spin-1/2 variables with Hamiltonian [26]

$$H = J(\sigma_x \otimes \sigma_x \otimes \mathbb{1} + \sigma_x \otimes \mathbb{1} \otimes \sigma_x + \mathbb{1} \otimes \sigma_x \otimes \sigma_x), \tag{14}$$

where σ_j stands for the usual Pauli matrices, $\mathbb{1}$ is the 2×2 identity operator, and the coupling J is positive (i.e., antiferromagnetic), to make the system frustrated.

It is easy to find a basis of the Hilbert space which makes the problem trivial: On the basis in which all σ_x operators are diagonal, it is immediate to see that two distinct degenerate energy levels exist: the fundamental one with energy $E_0 = -J$ and degeneracy 6 (corresponding to the case of two spins aligned) and the excited one with energy $E_1 = 3J$ and degeneracy 2 (corresponding to three spins aligned). However, to mimic a realistic situation, we work in the standard computational basis, where all σ_z operators are diagonal. Using this basis to evaluate the thermodynamical quantities by means of the Trotter-Suzuki decomposition, it is possible to verify that here the standard path-integral importance sampling Monte Carlo fails, due to a sign problem (see Ref. [26]). This problem is obviously absent in the quantum computational approach.

As discussed in Secs. II A and II B, both QMS and Q^{2}MA require the application of unitary operators to sample the state space: In the QMS, we need the set of operators denoted by $\mathcal{C}$ in Sec. II A to evolve the Markov chain; for the Q^{2}MA, we need the kick operators K_i to build the Szegedy operators W_j (see Sec. II B). For both algorithms, the adopted unitary operators are Hadamard gates (Had) acting on a single qubit of the register of the system state, i.e., $\mathcal{C} \equiv \{K_0, K_1, K_2\}$, where $K_0 = \mathbb{1} \otimes \mathbb{1} \otimes \mathrm{Had}$, $K_1 = \mathbb{1} \otimes \mathrm{Had} \otimes \mathbb{1}$, and $K_2 = \mathrm{Had} \otimes \mathbb{1} \otimes \mathbb{1}$.

B. Quantifying systematical errors

Systematical errors induce biases in the thermal averages computed by means of QMS or Q^{2}MA. In the simple test system adopted here, these biases can be identified by comparing the numerically estimated values with the analytically known exact ones. Since systematic errors affect, in a different way, the various observables, to quantify them we decided to use the following three metrics:

(1) The bias in the expectation value of the Hamiltonian (i.e., in the internal energy):

$$d_{\mathrm{Ene}} \equiv |\overline{E} - \langle E \rangle_{\mathrm{exact}}|, \tag{15}$$

where $\bar{E}$ is the mean energy estimated by using the quantum algorithm, while $\langle E \rangle_{\mathrm{exact}}$ is the exact expectation value of the energy at (inverse) temperature β, given by

$$\langle E \rangle_{\mathrm{exact}} = 3J \frac{e^{-3\beta J} - e^{\beta J}}{e^{-3\beta J} + 3e^{\beta J}}. \tag{16}$$

(2) The bias in the expectation value of an observable A which does not commute with H. For this purpose, we choose $A \equiv \sigma_x \otimes \sigma_x \otimes (\mathbb{1} + \sigma_y)$ and define

$$d_{\mathrm{Aop}} \equiv |\bar{A} - \langle A \rangle_{\mathrm{exact}}|, \tag{17}$$

where $\bar{A}$ is the value estimated by using the quantum algorithm and $\langle A \rangle_{\mathrm{exact}} = \langle E \rangle_{\mathrm{exact}}/3$ is the analytically known result.

(3) The distance in the space of the density matrices

$$d_{\mathrm{TrD}} \equiv \frac{1}{2}\|\bar{\rho} - \rho_{\mathrm{exact}}\|_1, \tag{18}$$

where for a matrix M the norm $\|M\|_1 = \mathrm{Tr}\sqrt{M^\dagger M}$ is used.

The figure of merit in Eq. (18) is particularly significant since it directly quantifies the bias in the probability distribution and not only in some specific average value, which could be small. However, its definition requires some comments. While the computation of $\bar{E}$ (and analogously of $\bar{A}$) can be carried out (at least from a theoretical point of view) on a quantum simulator by using $\bar{E} = \frac{1}{N}\sum_{i=1}^{N} E_i$, where E_i is the energy observed in the ith draw, this is not the case for $\bar{\rho}$ since it is not possible to measure at once the density matrix ρ_i of the ith draw. However, since we are testing the algorithms using a quantum simulator and not a real quantum machine, for the purpose of investigating the systematical errors, we can actually pick up the full state of the algorithm, extract ρ_i, and compute $\bar{\rho} \equiv \frac{1}{N}\sum_i \rho_i$.

IV. RESULTS

The particular features of the explored model permit us to disentangle effects related to the inexact energy representation from the QPE of other systematics. For this reason, in the first part of this section we work within an exact energy representation scheme, switching to the inexact one in the second part.

The practical implementations of the explored algorithms are carried out on a hybrid quantum-classical algorithm emulator [35] developed by one of the authors (G.C.).

A. Exact energy representation

The frustrated triangle can be described with an exact encoding of its degrees of freedom using three qubits only. Moreover, as mentioned in Sec. III A, the system has two energy levels, which can be represented exactly with one qubit for the energy registers, thus removing any systematical error related to the QPE with the Hamiltonian. Energy differences (needed for the Q^2MA) can thus be exactly represented by using two qubits.

As previously discussed, when this exact energy encoding is used, QMS and Q^2MA have different sources of systematical errors, which will be investigated below.

1. Quantum Metropolis Sampling

As discussed in Sec. II C, the QMS is not stochastically exact when used to compute the thermal average of an observable which does not commute with the Hamiltonian (as for the operator A defined in Sec. III B) since the measurement of A breaks the Markov chain evolution. The only parameter which controls the size of the systematical error introduced by this breaking is the number of updates performed between consecutive measures of A.

Once a measurement of A has been performed, two different strategies are possible: One can either restart the chain from the beginning or make the state (which is now an eigenstate of A) collapse back to an energy eigenstate (possibly different from the original one) by performing a QPE on the energy register followed by a measurement [26]. In the first case, the number of updates between different measurements is called thermalization time, while in the second case it is more natural to call it rethermalization time. As in Ref. [26], we follow the second strategy, which seems to be slightly more efficient.

The density matrix after the ith rethermalization step is evaluated, before breaking the state with the A-measurement, by reading the state from the system register $|\psi_{k(i)}\rangle$ and building the projector $\rho_i \equiv |\psi_{k(i)}\rangle\langle\psi_{k(i)}|$. Of course, any single instance of ρ_i will be far from the exact density matrix ρ_{exact}, irrespective of the number of rethermalization steps. However, in the large sample limit, the (nonvanishing) discrepancy between $\bar{\rho}$ and ρ_{exact} is only due to the systematic introduced by rethermalization, and it is expected to vanish in the limit of an infinite number of rethermalization steps.

Results obtained for the three accuracy metrics d_{Ene}, d_{Aop}, and d_{TrD} introduced in the previous section are shown in Fig. 1 (respectively, top, middle, and bottom panels), for three values of the inverse temperature $\beta J = 0.25$, 0.5, and 1.0. The statistics accumulated for the different points is not homogeneous since runs are stopped when d_{TrD} reaches a fixed accuracy. For this reason, runs performed at larger values of rethermalization steps, for which the bias is smaller, are significantly longer than the ones performed at smaller values. The stopping criterium adopted, together with the fact that d_{TrD} is the most observable independent among the adopted metrics, also explains why data reported in the top and central panels, respectively, for d_{Ene} and d_{Aop}, have significantly larger relative errors compared to those in the bottom panel, for d_{TrD}. It is also clear that, for all the metrics considered, the systematic bias approaches zero exponentially in the number of rethermalization steps, as expected on theoretical grounds.

Note that the size of the systematic errors observed in Fig. 1 (at a fixed number of rethermalization steps) decreases with decreasing β. This is the same behavior expected in standard Monte Carlo sampling, which can be easily explained: In the high-temperature limit almost any proposed update is accepted, and as a consequence, the exponential autocorrelation time of the Markov chain decreases by decreasing β. Since the size of the systematics depends on the ratio between the rethermalization time and the autocorrelation time, systematics decrease when increasing the temperature at fixed rethermalization steps.

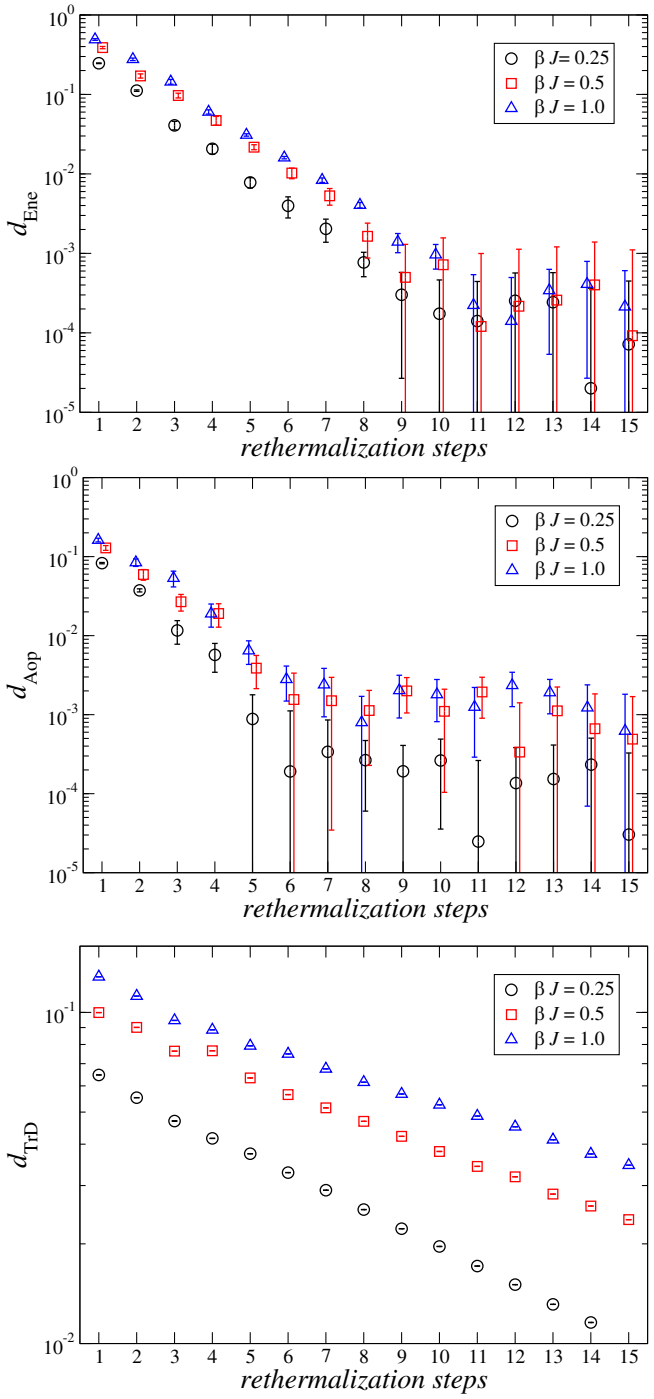

FIG. 1. QMS: behavior of d_{Ene} (top panel), d_{Aop} (middle panel), and d_{TrD} (lower panel), as a function of the number of rethermalization steps, for $\beta J = 0.25$, 0.5, and 1.0.

2. *Quantum-Quantum Metropolis Algorithm*

As explained in Sec. II C, when using an exact energy representation, the systematic errors of the Q^{2}MA algorithm are only due to the inexactness of the Szegedy QPE and to the finite number of annealing steps. However, we recall that it is generally not possible to make the Szegedy QPE exact, as the W operator depends on β. Results presented in this section have been obtained by using the range [0.5, 1] for the Szegedy QPE (we are only interested in the eigenvalue 1) and fixing the number of qubits in the $|w\rangle_5$ register to 3, studying the dependence of systematic errors on the number of annealing steps. Consistent results are obtained when using four qubits for the register $|w\rangle_5$.

Contrary to what happens in the QMS case, it is not possible to measure A and E during the same run, and a new CETS reconstruction is required after each measurement. Since here we are only interested in systematics related to the incorrect determination of CETS, we evaluated the density matrix after each run of the algorithm and then exploited the use of an emulator (rather than a real machine) to determine the exact average values of A and E corresponding to the given density matrix. Therefore, statistical errors shown in the following analysis are only due to fluctuations in the CETS determination from run to run and are in fact small and well below the symbol size in most cases.

In Fig. 2, we present the results for d_{Ene} (top panel), d_{Aop} (middle panel), and d_{TrD} (bottom panel) for the same three values of temperature explored in the previous subsection. In order to achieve a target precision on d_{TrD}, the runs performed with higher numbers of annealing steps required more Q^{2}MA iterations, as the systematic error decreases with increasing steps. It can be observed that error bars are visible and not homogeneous only in the middle panel, where the systematic error also shows a nonmonotonic behavior as a function of the number of annealing steps n_a: This is partially due to a change of sign in the bias of A as a function of n_a, which occurs accidentally for this particular choice of parameters.

A clear difference of Fig. 2, with respect to Fig. 1, is that the scale is logarithmic on both axes. The reason is that, in this case, systematics appear to decrease polynomially (instead of exponentially) with the number of annealing steps. In particular, it is interesting to notice that, for large enough n_a, i.e., when the annealing step is small enough, data are very compatible with a $1/n_a$ behavior. This result agrees with the prediction reported in Ref. [14] for a single ergodic kick operator. However, in the light of the discussion reported in Sec. II, the result is far from trivial, and it can be interpreted heuristically as evidence that randomly alternating different kick operators in the different annealing steps effectively reproduces, in the large annealing step limit, the ideal behavior predicted for a single ergodic kick operator.

We close this section by noting that the scaling of the systematic errors at fixed annealing steps is the opposite of the naively expected one. As discussed in the previous section, systematic errors typically decrease when decreasing β in a standard Markov chain Monte Carlo; in Fig. 2 we instead see the systematics decreasing by increasing the value of β. While a detailed analysis of this phenomenon goes beyond the scope of the present study, it is natural to

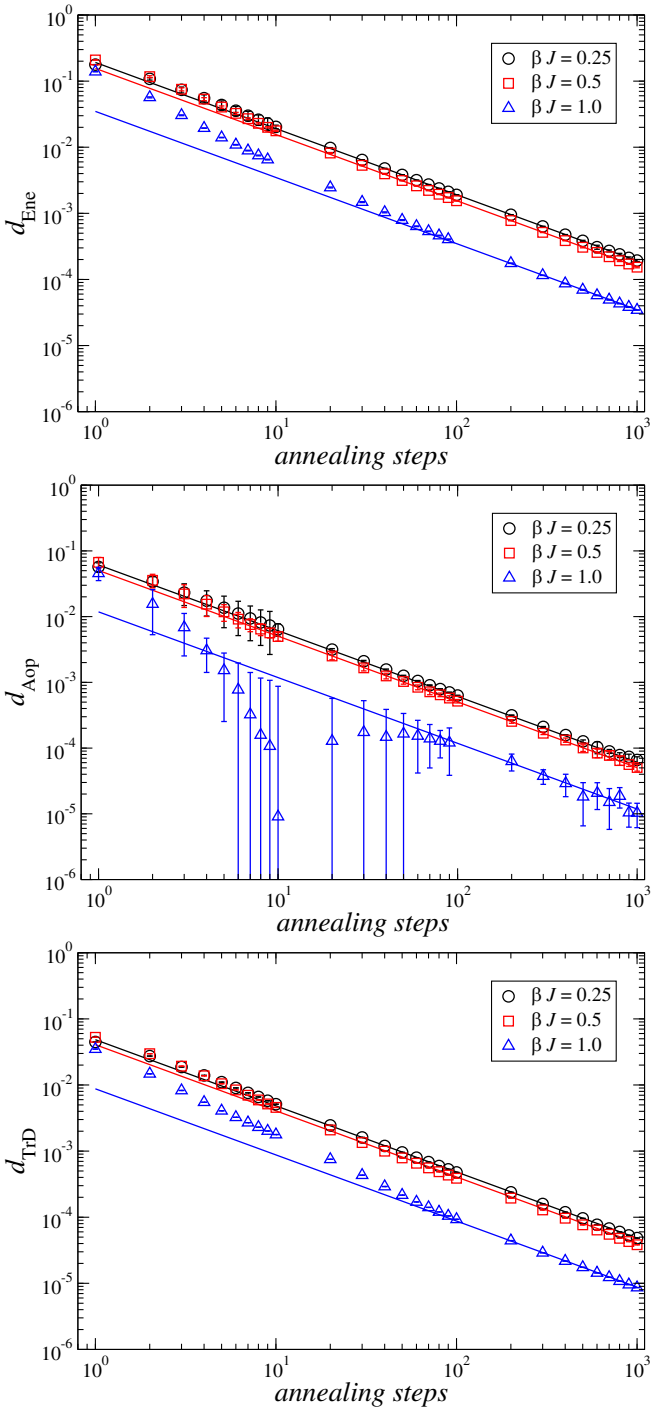

FIG. 2. Q^{2}MA: behavior of d_{Ene} (top panel), d_{Aop} (middle panel), and d_{TrD} (bottom panel), as a function of the annealing steps n_a, for $\beta J = 0.25$, 0.5 and 1.0, when using three qubits in the register $|w\rangle_5$; see Eq. (10). Solid lines represent fits to the $1/n_a$ behavior, well reproducing data at large enough n_a.

interpret it as a consequence of the fact that for large β values the spectral gap of the Szegedy operator increases with eigenvalue 1, which becomes more clearly separated from the rest of the spectrum. As a consequence, the annealing becomes more efficient since there is an increasing probability for the state to be projected on the correct CETS $|\alpha_0^i\rangle$ after each subsequent annealing step.

B. Inexact energy representation

In the previous analysis, working with an exact energy representation was instrumental to isolate some algorithmic-specific sources of systematical errors. This was possible because the considered quantum system only has two different exactly known energy levels, but it is clearly unrealistic in any case of direct physical interest. In such general cases, one has to use an inexact energy representation with n_e qubits in the energy register(s), thus introducing further systematics.

Denoting by E_0 and E_1 the exact energy eigenstates of the frustrated triangle system (see Sec. III A), we consider for the QPE an interval larger than $[E_0, E_1]$ by an amount of about 2δ. Two natural prescriptions exist to place the 2^{n_e} grid points of the QPE: the "fixed extrema grid" and the "refined grid" cases.

In the fixed extrema case the interval is fixed to be $[E_0 - \delta, E_1 + \delta]$, and the 2^{n_e} points are uniformly distributed in this interval. As n_e is increased, the grid becomes finer, but the position of all the grid points changes with n_e. On the other hand, in the refined grid case we use the interval

$$[E_0 - \delta, E_1 + \delta + (E_1 - E_0 + 2\delta)(1 - 2^{1-n_e})], \qquad (19)$$

which is chosen in such a way that, by increasing n_e, all points of the coarser grid are also present in the finer grid (with the exception of the largest value).

Figure 3 shows the behavior of $\langle E\rangle$ as a function of the number of rethermalization steps for different values of n_e in the fixed extrema scheme. It is clear that, for a large enough number of rethermalization steps, a plateau emerges in $\langle E\rangle$, signaling the presence of a systematical difference between $\langle E\rangle$ and $\langle E\rangle_{\mathrm{exact}}$. This systematic is expected to vanish for $n_e \to \infty$; however, the approach to the large n_e limit is obviously nonmonotonic, at least in the range of n_e values explored. This could make the diagnostic of the convergence nontrivial in realistic cases, in which the exact expectation value is not known and the number of values of n_e is limited by the hardware capabilities.

For this reason we also investigated the refined grid method, which could be expected to have a smoother approach to the large n_e limit. However, this is not the case, as can be seen from Fig. 4, in which a comparison of the two approaches is performed using ten rethermalization steps (note that this number of rethermalization steps is well within the plateau of Fig. 3). Both the approaches show a nonmonotonic scaling for an intermediate range of n_e values and converge to the correct asymptotic result for $n_e > 8$. It is, however, reasonable to guess the nonmonotonic behavior to continue also for larger values of n_e, where it is hidden by the statistical accuracy of our data.

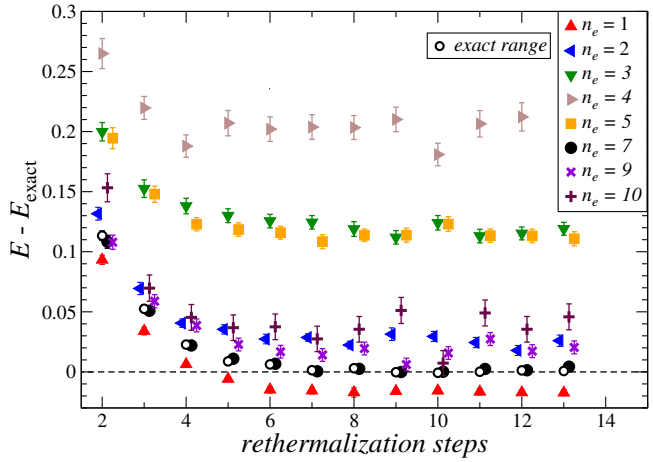

FIG. 3. Energy discrepancy for the QMS algorithm with an inexact QPE grid with fixed extrema $[-1.1, 3.1]$ (in units of J) at $\beta J = 0.25$.

The reason for the nonmonotonic behavior is related to the fact that, for a given value of n_e, an eigenvalue can, by chance, be well within one of the QPE grid intervals or close to the boundary between two consecutive grid intervals. Which of the two cases happens depends on n_e and changes by increasing n_e, as can be seen in Fig. 5. Obviously, the oscillations induced by this effect get smaller and smaller as n_e is increased; ultimately, convergence is reached with an arbitrary accuracy, but the approach to the asymptotic value presents oscillations. This effect is particularly evident in the system studied in this paper since the spectrum is very simple, consisting of just two points. In more complex systems, with a less simple spectrum, it seems reasonable to assume this discretization effect is less significant, with some form of self-averaging happening.

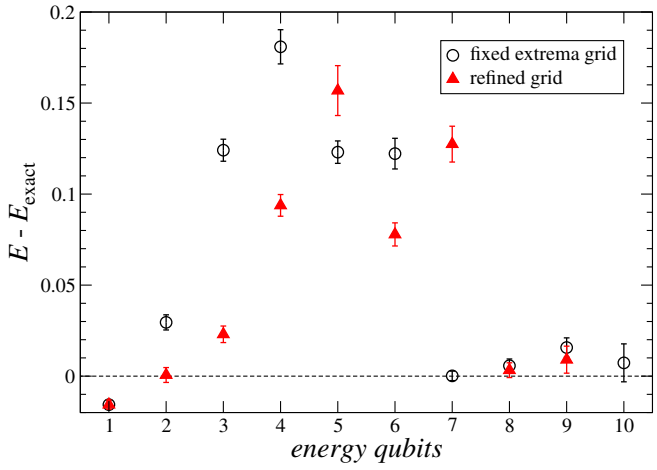

FIG. 4. Effects of the two grid prescriptions for the QMS algorithm on energy discrepancies with inexact QPE grids at $\beta J = 0.25$ and with ten rethermalization steps.

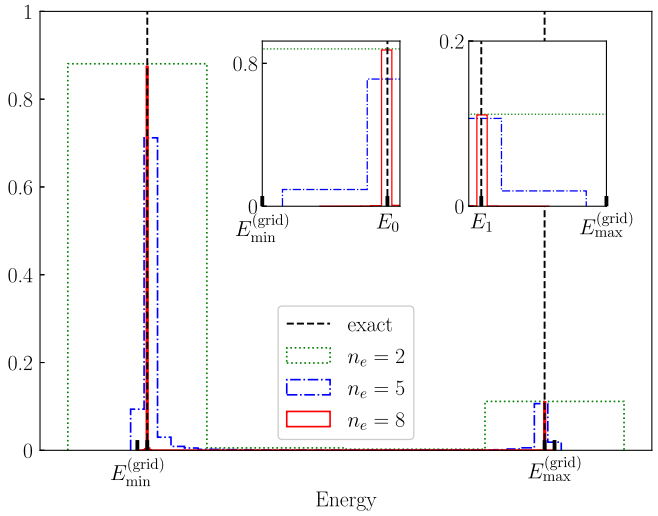

FIG. 5. Comparison between the exact energy distribution and the one computed from the QMS using an energy grid in the QPE with fixed extrema and with an inexact range $[E_{min}^{(grid)}, E_{max}^{(grid)}] = [-1.1, 3.1]$ (in units of J), for three values of the number of qubits of the energy register n_e. In all cases the figure refers to measures separated by ten rethermalization steps. Insets provide a zoom of the regions close to the exact energy levels E_0 and E_1.

We tried to repeat a similar analysis to investigate the effect of an inexact energy representation also in the Q^2MA case; however, the results turned out to be much less clear, a fact that is probably due to several aspects. First of all, the way in which energies enter the construction of the Szegedy operator is much more involved than the way in which they enter the Metropolis filter in the QMS algorithm, so "error propagation" is nontrivial for the limited number of qubits that can be used in the simulator. Another important point is the fact that in the Q^2MA other sources of systematics are also present, like the discretization error in the Szegedy-related QPE, and different systematics can interact in a nontrivial way with each other. For this reason we were not able to identify a reasonable trend in our data for the case of the inexact energy representation in the Q^2MA.

V. DISCUSSION AND CONCLUSIONS

This study is a step along a research line dedicated to the exploration of quantum algorithms for the computation of quantum thermal averages, in view of future applications to complex and interesting physical systems, like the fundamental theory of strong interactions, when technological developments will allow for reliable and scalable quantum machines.

In Ref. [26], some of the present authors already explored the QMS algorithm and applied it to a frustrated three-spin system. In this case, we have considered the same physical system to develop a practical implementation of the Q^2MA, proposed in Ref. [14]. This algorithm, in

principle, represents a conceptual improvement over the QMS, as it enjoys a further quantum advantage. Such an advantage stems from the fact that, while the QMS performs a Markov chain among the quantum states of the system which is classical in spirit and which faces the difficulties of the no-cloning theorem, the Q^2MA acts like a quantum searching algorithm, where the searched state is the so-called CETS, which is a pure state in a doubled Hilbert space, whose amplitudes encode the thermal distribution of the target system. Provided that the CETS is found, the algorithm is exact, in principle; however, an annealing procedure, in which one finds iteratively different CETS states corresponding to a descending sequence of temperatures, is used to improve the success probability of the searching algorithm.

A first result of our investigation is that the practical implementation of the Q^2MA algorithm might be far less trivial. The searching algorithm is based on the construction of a Szegedy operator, which is assumed to have a single eigenstate with eigenvalue $\lambda = 1$, corresponding to the CETS, which is actually the eigenstate the algorithm looks for. On the other hand, the construction of the Szegedy operator is based on the definition of a kick operator K, which is representative of a Markov chain and should be ergodic to guarantee the nondegeneracy of the $\lambda = 1$ eigenstate: If this is not the case, the algorithm is not guaranteed to find the correct CETS, leading to possible systematics.

As we have discussed in Sec. II, finding an ergodic kick operator is a nontrivial assumption. However, a possible conceptual modification of the algorithm is to make use of different kick operators, which are randomly alternated during the annealing procedure: Even if the single kick operators are nonergodic, the fact that a CETS corresponding to close-enough temperatures has a good overlap is expected to strongly enhance the probability that the correct CETS is found along the annealing sequence, at least if the annealing step is small enough.

As we have shown, this conceptual modification works well in practice, so one is able to reproduce quantum thermal averages through annealing with different kick operators; on the contrary, using a single nonergodic operator does not work. However, a consequence of this modification is that now the annealing step, or in other words the number of annealing steps n_a, is not only relevant to the success probability (i.e., of finding an eigenstate of the Szegedy operator with eigenvalue 1) but also to the fact that the selected state is actually the CETS. In other words, there is a systematic effect in the algorithm, related to the fact that one might find the wrong state, and this systematic is a function of n_a. In particular, we have shown that the error scales as $1/n_a$, at least for large enough n_a.

Returning to the comparison with the QMS, the final situation seems different from initial expectations. In the QMS, the main algorithm-specific systematics is related to the number of rethermalization steps performed along the Markov chain: In this case, the systematic error is exponentially suppressed in such a number. The outcome of our study is that the main algorithm-specific systematics of Q^2MA scales as the inverse of the number of annealing steps, which is a less favorable polynomial scaling, compared to QMS.

Finally, we have explored the systematic effects related to the inexact representation of the system energy spectrum obtained through the QPE, which is a problem affecting a large class of quantum algorithms and could be avoided for the particular explored system, due to its simplicity. In this case, the systematic error is expected to be suppressed as the number of qubits used to represent the energy register is increased, i.e., as the grid of possible outcomes of the QPE is made finer and finer. We have shown that this is indeed the case when the number of qubits is large enough, with a nontrivial intermediate regime due to the interplay between the energy level spacing, the grid spacing, and the overall range explorable by QPE. As a final comment, one should consider that this intermediate regime is likely to be less relevant for real and more complex many-body systems, for which the distribution of energy levels is more chaotic.

Future developments along the same research line should consider and compare different approaches, like those based on variational quantum algorithms or on quantum simulators [36], as well as less trivial models, including gauge degrees of freedom.

ACKNOWLEDGMENTS

This study was carried out within the National Centre on HPC, Big Data and Quantum Computing—SPOKE 10 (Quantum Computing) and received funding from the European Union Next-GenerationEU—National Recovery and Resilience Plan (NRRP)—MISSION 4 COMPONENT 2, INVESTMENT N. 1.4—CUP N. I53C22000690001. This work reflects only the authors' views and opinions, not those of the European Union nor the European Commission. It is a pleasure to acknowledge inspiring discussions with Raffaele Tripiccione, Leonardo Cosmai, and Fabio Schifano during the early stages of this work. We thank Man-Hong Yung and Alán Aspuru-Guzik for correspondence. The work of C. B. is supported by the Spanish Research Agency (Agencia Estatal de Investigación) through the grant IFT Centro de Excelencia Severo Ochoa CEX2020-001007-S and, partially, by Grant No. PID2021-127526NB-I00, both funded by MCIN/AEI/10.13039/501100011033. S. T. acknowledges support from projects PRIN 2017 Taming complexity via Quantum Strategies: a Hybrid Integrated Photonic approach (QUSHIP) Id. 2017SRN-BRK and PRO3 Quantum Pathfinder. The research of K. Z. has been supported by the University of Pisa under the "PRA—Progetti di Ricerca di Ateneo" (Institutional Research Grants)—Project No. PRA 2020-2021 92 "Quantum Computing, Technologies and Applications." Numerical simulations have been performed at the IT

Center of the University of Pisa and on the `Marconi100` machines at CINECA, based on the agreement between INFN and CINECA, under projects INF22_npqcd and INF23_npqcd.

APPENDIX: RESOURCE REQUIREMENTS

Both the QMS and the Q^{2}MA algorithms require resources which are not yet available (at the time of writing) on real quantum hardware, even for the application to the three-spin frustrated triangle system. In terms of the number q of qubits, the QMS algorithm requires a register with q_{sys} qubits to represent pure states of the system, a register with a variable number r_{ene} of qubits to represent energy estimates, and a 1-qubit register for the Metropolis test, for a total of

$$q^{(\mathrm{QMS})} = q_{\mathrm{sys}} + r_{ene} + 1_{acc}. \tag{A1}$$

The Q^{2}MA algorithm requires a register with $2q_{\mathrm{sys}}$ qubits to represent the two copies of the system, a register with a variable number of qubits $r_{\Delta E}$ to represent energy differences, a register with r_{Sz} qubits to perform QPE for the Szegedy operator, and a 1-qubit register for the annealing test, for a total of

$$q^{(\mathrm{Q^2MA})} = 2q_{\mathrm{sys}} + r_{\Delta E} + r_{Sz} + 1_{ann}. \tag{A2}$$

In terms of algorithmic depth, information about the gate count estimates for 1-qubit (n_1) and 2-qubit (n_2) gates for different code sections is reported in Table I for a run of the QMS algorithm, and in Table II for a run of the Q^{2}MA algorithm. For both algorithms, the reported numbers refer to the case of the frustrated triangle in the exact energy representation ($q_{\mathrm{sys}} = 3$, $r_{ene} = 1$, $r_{\Delta E} = 2$, and $r_{Sz} = 3$), which results in a total usage of $q^{(\mathrm{QMS})} = 5$ qubits in the QMS case and $q^{(\mathrm{Q^2MA})} = 11$ in the Q^{2}MA one.

The exact number of 1-qubit and 2-qubit gates depends on the specific decomposition of multiqubit gates, which might be dependent on the hardware adopted and other circuit optimization strategies. For example, to implement the oracles, we used multiple-control Toffoli (MCT_{q_c}) gates with $q_c \geq 3$ qubits. These gates are represented efficiently as a single operation in the emulator used [37], but, in terms of the actual representation with 1-qubit and 2-qubit gates, on some hardware one might need to use $n_1^{(\mathrm{MCT}_{q_c})} = 18q_c - 27$ gates acting on one qubit and $n_2^{(\mathrm{MCT}_{q_c})} = 12q_c - 17$ gates acting on two qubits (for an improved representation see, for example, Ref. [38]).

TABLE I. Gate counts for the QMS algorithm with 20 rethermalization steps for the frustrated triangle at $\beta = 0.25$. The Metropolis acceptance rate is about 90%. The (possibly overlapping) named sections are the following: "Metro step" refers to a single Metropolis step, "revert" to the revert procedure once a move is rejected, "sample" to the extraction of a single sample, and "measure" to the preparation for measurement. The relative number of executions for each section with respect to a single measurement is reported in the "frequency" column.

Section	Frequency	n_1	n_2
Metro step	20	∼240	∼200
Revert	2	∼500	∼450
Sample	1	∼5200	∼4500
Measure	1	∼120	∼110

TABLE II. Gate counts for the Q^{2}MA algorithm with 100 iterations and 100 annealing steps for the frustrated triangle at $\beta = 0.25$. The (possibly overlapping) named sections are the following: "Ann. step" refers to a single annealing step, "annealing" to the whole annealing sequence before measurement, and "measure" to the preparation for measurement. The relative number of executions for each section with respect to a single measurement is reported in the "frequency" column.

Section	Frequency	n_1	n_2
Annealing step	100	1.6×10^3	1.8×10^3
Annealing	1	1.6×10^5	1.8×10^5
Measure	1	1.2×10^2	1.2×10^2

[1] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, Nat. Phys. **14**, 595 (2018).
[2] F. Arute *et al.*, Nature (London) **574**, 505 (2019).
[3] Q. Zhu *et al.*, Sci. Bull. **67**, 240 (2022).
[4] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Rev. Mod. Phys. **92**, 015003 (2020).
[5] K. Bharti *et al.*, Rev. Mod. Phys. **94**, 015004 (2022).
[6] S. L. Shapiro and S. A. Teukolsky, *Black Holes, White Dwarfs, and Neutron Stars: The Physics of Compact Objects* (Wiley-Interscience, New York, 1983).
[7] F. Rajagopal and Krishna Wilczek, *The Condensed Matter Physics of QCD*, edited by M. Shifman, At the Frontier of Particle Physics Vol. 3 (World Scientific, Singapore, 2001).
[8] D. Poulin and P. Wocjan, Phys. Rev. Lett. **103**, 220502 (2009).

[9] E. Bilgin and S. Boixo, Phys. Rev. Lett. **105**, 170405 (2010).
[10] A. Riera, C. Gogolin, and J. Eisert, Phys. Rev. Lett. **108**, 080402 (2012).
[11] J. Wu and T. H. Hsieh, Phys. Rev. Lett. **123**, 220502 (2019).
[12] D. Zhu, S. Johri, N. M. Linke, K. A. Landsman, C. Huerta Alderete, N. H. Nguyen, A. Y. Matsuura, T. H. Hsieh, and C. Monroe, Proc. Natl. Acad. Sci. U.S.A. **117**, 25402 (2020).
[13] K. Temme, T. J. Osborne, K. G. Vollbrecht, D. Poulin, and F. Verstraete, Nature (London) **471**, 87 (2011).
[14] M.-H. Yung and A. Aspuru-Guzik, Proc. Natl. Acad. Sci. U.S.A. **109**, 754 (2012).
[15] A. N. Chowdhury and R. D. Somma, Quantum Inf. Comput. **17**, 0041 (2017).
[16] J. E. Moussa, arXiv:1903.01451.
[17] M. Motta, M. Sun, A. T. K. Tan, M. J. O'Rourke, E. Ye, A. J. Minnich, F. G. S. L. Brandão, and G. K.-L. Chan, Nat. Phys. **16**, 205 (2020).
[18] S.-N. Sun, M. Motta, R. N. Tazhigulov, A. T. K. Tan, G. K.-L. Chan, and A. J. Minnich, PRX Quantum **2**, 010317 (2021).
[19] S. Lu, M. C. Bañuls, and J. I. Cirac, PRX Quantum **2**, 020321 (2021).
[20] A. Yamamoto, Phys. Rev. D **105**, 094501 (2022).
[21] J. Selisko, M. Amsler, T. Hammerschmidt, R. Drautz, and T. Eckl, arXiv:2208.07621.
[22] Z. Davoudi, N. Mueller, and C. Powers, Phys. Rev. Lett. **131**, 081901 (2023).
[23] C. Ball and T. D. Cohen, Nucl. Phys. **A1038**, 122708 (2023).
[24] C. Powers, L. Bassman Oftelie, D. Camps, and W. de Jong, Sci. Rep. **13**, 1986 (2023).
[25] M. Fromm, O. Philipsen, M. Spannowsky, and C. Winterowd, arXiv:2306.06057.
[26] G. Clemente *et al.* (QuBiPF Collaboration), Phys. Rev. D **101**, 074510 (2020).
[27] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, England, 2010), 10th Anniversary Edition.
[28] G. Benenti, G. Casati, D. Rossini, and G. Strini, *Quantum Computation and Quantum Information* (World Scientific, Singapore, 2018).
[29] A. Y. Kitaev, arXiv:quant-ph/9511026.
[30] R. Cleve, A. Ekert, C. Macchiavello, and M. Mosca, Proc. R. Soc. A **454**, 1969 (1998).
[31] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[32] R. D. Somma, S. Boixo, H. Barnum, and E. Knill, Phys. Rev. Lett. **101**, 130504 (2008).
[33] G. O. Roberts, J. S. Rosenthal, and P. O. Schwartz, J. Appl. Probab. **35**, 1 (1998).
[34] L. Breyer, G. O. Roberts, and J. S. Rosenthal, Stat. Probab. Lett. **53**, 123 (2001).
[35] G. Clemente, SUQA: Simulator for Universal Quantum Algorithms, https://github.com/QC-PISA/suqa.
[36] M. C. Bañuls *et al.*, Eur. Phys. J. D **74**, 165 (2020).
[37] G. Clemente, A. Crippa, and K. Jansen, Phys. Rev. D **106**, 114511 (2022).
[38] V. V. Shende and I. L. Markov, Quantum Inf. Comput. **9**, 461 (2009).