Finite-temperature Lanczos for anisotropic spin systems using triple-hybrid high-performance computing

Jürgen Schnack[1, *]

[1] *Fakultät für Physik, Universität Bielefeld, Postfach 100131, D-33501 Bielefeld, Germany*

Lanczos concepts are often realized on supercomputers. In view of modern architectures of high-performance computing triple-hybrid schemes employing MPI, CPU-openMP as well as GPU-openMP should be utilized. The present article sketches how such a scheme could be utilized in order to meet the demands of the finite-temperature Lanczos method for anisotropic spin systems.

arXiv:2607.25585v1 [cond-mat.str-el] 28 Jul 2026

I. INTRODUCTION

Thermodynamic equilibrium expectation values can be accurately approximated using trace estimators together with Krylov space representations. These methods are heavily employed for spin models as well as Hubbard models but are not restricted to these. The trace is estimated by a simple evaluation of an expectation value with respect to a random vector [1–4], i.e.,

$$\mathrm{tr}\left(\underset{\sim}{A}\right) \approx \langle r | \underset{\sim}{A} | r \rangle \frac{\dim(\mathcal{H})}{\langle r | r \rangle} . \tag{1}$$

Here, $\underset{\sim}{A}$ is the operator of interest, and $|r\rangle$ is a vector randomly drawn from a (potentially high-dimensional) Hilbert space. The complex components r_ν of the vector $|r\rangle$ with respect to a chosen orthonormal basis $\{|\nu\rangle\}$,

$$|r\rangle = \sum_\nu r_\nu |\nu\rangle , \tag{2}$$

are supposed to follow a Gaussian distribution with zero mean, but the method works robustly with other choices of random numbers [5].

In equilibrium statistical quantum physics such traces concern operators as $\exp\{-\beta\underset{\sim}{H}\}$ and $\underset{\sim}{Q}\exp\{-\beta\underset{\sim}{H}\}$ yielding the partition function and thermal expectation values of observables $\underset{\sim}{Q}$, respectively. For correlated electron systems magnetic observables are often evaluated this way, see e.g. [3, 6–16] for a small selection of references, but the method is also used elsewhere, see e.g. [17].

In this paper, we focus on the finite-temperature Lanczos method (FTLM) [3, 7, 18], a variant of the above idea, and apply it to anisotropic spin systems. The Hamiltonian of such systems might contain anisotropic exchange, dipolar interactions [19, 20] or single-ion anisotropies [21] together with isotropic terms such as Heisenberg exchange interactions. It was noted earlier that the typically very good convergence of FTLM for Heisenberg systems worsens drastically under anisotropy in particular at low temperatures [8, 10, 22, 23]. We follow the numerically costly workarounds proposed in [8, 10] and present an implementation for high-performance computing. With two examples we will demonstrate the applicability of FTLM also for very anisotropic spin systems.

The paper is organized as follows. In Section II we introduce model and methods. Results are discussed in Sec. III. The article closes with a discussion in Section IV.

II. MODEL AND METHODS

A. Spin model

The Hamiltonian that is employed in the following consists of Heisenberg exchange interactions, single-ion anisotropies, the Zeeman term as well as dipolar interactions

$$\begin{aligned}\underset{\sim}{H} =& 2\sum_{i<j} J_{ij}\,\underset{\sim}{\vec{s}}_i \cdot \underset{\sim}{\vec{s}}_j + \sum_i d_i \left(\underset{\sim}{\vec{s}}_i \cdot \vec{e}_i\right)^2 + \mu_B \vec{B} \cdot \sum_i g_i \underset{\sim}{\vec{s}}_i \\ &+ \frac{\mu_0 \mu_B^2}{4\pi} \sum_{i<j} \frac{g^2}{r_{ij}^3} \left(\underset{\sim}{\vec{s}}_i \cdot \underset{\sim}{\vec{s}}_j - 3 \cdot \underset{\sim}{\vec{s}}_i \cdot \vec{e}_{ij} \otimes \vec{e}_{ij} \cdot \underset{\sim}{\vec{s}}_j\right) .\end{aligned} \tag{3}$$

Here, $\underset{\sim}{\vec{s}}_j$ denotes the spin vector operator at site j, and a tilde is used to mark operators in general. Single-ion anisotropy is incorporated in its simplest form with anisotropy axes along local directions $\vec{e}_i$.

If the dimension of the related Hilbert space is sufficiently small the respective matrix representation can be diagonalized numerically, if possible using symmetries, see e.g. [24–26]. However, since the dimension, $d = \prod_i (2s_i + 1)$, grows exponentially with the number of spins, exact diagonalization is rather limited. This is even more true for anisotropic spin systems since they typically lack symmetries. Then, approximate methods such as FTLM are needed.

B. Finite-temperature Lanczos method

The finite-temperature Lanczos method (FTLM) [3, 7, 18] approximates an equilibrium observable as

$$O^{\mathrm{FTLM}}(T, \vec{B}) \approx \frac{\sum_{r=1}^{R} \langle r | \underset{\sim}{Q} e^{-\beta \underset{\sim}{H}} | r \rangle}{\sum_{r=1}^{R} \langle r | e^{-\beta \underset{\sim}{H}} | r \rangle} . \tag{4}$$

* jschnack@uni-bielefeld.de

Here, the set of R random vectors $|\,r\,\rangle$ is the same in numerator and denominator. If the Hamiltonian $\underset{\sim}{H}$ possesses symmetries, these can be employed by decomposing the full Hilbert space into mutually orthogonal subspaces according to the irreducible representations of the employed symmetry [10, 27] yielding for the partition function

$$Z^{\mathrm{FTLM}}(T,\vec{B}) \approx \sum_{\gamma=1}^{\Gamma} \frac{\dim(\mathcal{H}(\gamma))}{R} \times \sum_{r=1}^{R}\sum_{n=1}^{N_L} e^{-\beta\epsilon_n^{(r)}} |\langle\, n(r)\,|\,r\,\rangle|^2 \ . \quad (5)$$

$\mathcal{H}(\gamma)$ is the subspace belonging to the irreducible representation γ, N_L is the dimension of the generated Krylov space, and $|\,n(r)\,\rangle$ is the n-th eigenvector of $\underset{\sim}{H}$ in this Krylov space with seed $|\,r\,\rangle$ and energy eigenvalue $\epsilon_n^{(r)}$. N_L is typically chosen such, that the ground state energy in the respective subspace is converged to numerical accuracy, which often already happens for N_L of the order of 100 [5]. $\beta = 1/(k_B T)$ denotes the inverse temperature.

The arguably most often employed symmetry is related to the conservation of the z-component of the total spin $\underset{\sim}{S^z}$ that is for instance present in Heisenberg and XXZ models. Investigations of anisotropic systems where the $\underset{\sim}{S^z}$-symmetry does not hold revealed that this symmetry is responsible for the astonishing accuracy and fast convergence of magnetic observables in FTLM. Both, accuracy as well as fast convergence worsen massively if anisotropy breaking $\underset{\sim}{S^z}$-symmetry is present – the stronger the worse. It was argued that both problems can be cured by making use of the always present time-reversal symmetry by utilizing pairs of mutually time-reversed random vectors for the FTLM approximation [10]. We will follow this suggestion in the present paper.

A second concern is related to the asymmetry in (4), which was first discussed in [8]. At the level of the trace we observe that

$$\mathrm{Tr}\left[\underset{\sim}{Q} e^{-\beta \underset{\sim}{H}}\right] = \mathrm{Tr}\left[e^{-\frac{1}{2}\beta\underset{\sim}{H}}\underset{\sim}{Q}e^{-\frac{1}{2}\beta\underset{\sim}{H}}\right] , \quad (6)$$

but this is in general not true for the expectation value

$$\langle\, r\,|\,\underset{\sim}{Q}e^{-\beta\underset{\sim}{H}}\,|\,r\,\rangle \neq \langle\, r\,|\,e^{-\frac{1}{2}\beta\underset{\sim}{H}}\underset{\sim}{Q}e^{-\frac{1}{2}\beta\underset{\sim}{H}}\,|\,r\,\rangle \ . \quad (7)$$

The authors of [8] demonstrated a massively lowered accuracy for small temperatures if the asymmetric form (4) is employed, but depending on the anisotropy this problem is not restricted to low temperatures. One also notices, that if $\underset{\sim}{Q}$ commutes with $\underset{\sim}{H}$ the expressions in (7) are the same which is at the heart of the good convergence for $\underset{\sim}{S^z}$-symmetric problems.

In the remainder of this paper we will thus employ the symmetric version

$$O^{\mathrm{sFTLM}}(T,B) \approx \frac{\sum_{r=1}^{R}\langle\, r\,|\,e^{-\frac{1}{2}\beta\underset{\sim}{H}}\underset{\sim}{Q}e^{-\frac{1}{2}\beta\underset{\sim}{H}}\,|\,r\,\rangle}{\sum_{r=1}^{R}\langle\, r\,|\,e^{-\frac{1}{2}\beta\underset{\sim}{H}}e^{-\frac{1}{2}\beta\underset{\sim}{H}}\,|\,r\,\rangle} \quad (8)$$

and term it symmetric FTLM (sFTLM = low-temperature FTLM in [8]). Although looking rather innocent, this improvement comes at the cost of another factor N_L in computing time because the expectation value

$$\langle\, r\,|\,e^{-\frac{1}{2}\beta\underset{\sim}{H}}\underset{\sim}{Q}e^{-\frac{1}{2}\beta\underset{\sim}{H}}\,|\,r\,\rangle \quad (9)$$
$$\approx \sum_{m=1}^{N_L}\sum_{n=1}^{N_L} e^{-\frac{1}{2}\beta(\epsilon_m^{(r)}+\epsilon_n^{(r)})} \times\langle\, r\,|\,m(r)\,\rangle\langle\, m(r)\,|\,\underset{\sim}{Q}\,|\,n(r)\,\rangle\langle\, n(r)\,|\,r\,\rangle \ ,$$

now contains a double sum, compare to (5). Since $N_L \sim 100$ this dramatically increases the computing time [15]. In addition, all eigenvectors $|\,n(r)\,\rangle$ need to be stored.

C. Triple-hybrid high-performance computing

Lanczos concepts [28] are often realized on supercomputers, see [14, 29–37] for some of the largest more recent calculations. The paradigm that high-performance computations should be programmed using a MPI/openMP hybrid scheme as employed for instance in the publicly available program *spinpack* [38, 39] moves towards triple-hybrid MPI/CPU-openMP/GPU-openMP with recent updates of supercomputers such as the Phase-2 at the Leibniz supercomputing center (LRZ) of the Bavarian Academy of Sciences in Germany. Now, a program should be parallelized across MPI ranks, CPU threads, and GPU engines where GPU is the graphics processing unit (*vulgo* graphics card). In the following, a working approach is presented for the sparse matrix vector multiplications that are at the heart of every Lanczos procedure.

The approach presented here distributes the Hamiltonian matrix and the Lanczos vectors across the ranks of a MPI parallelization and takes care of the sparsity at the level of the resulting small sub-matrices. In a practical realization a *rank* would correspond to a GPU, i.e., for a system such as Phase-2@LRZ with nodes of 4 GPUs one would address four times as many ranks as *nodes*. For technical clarity, a node denotes a pc-like part of a supercomputer, a rank is one of the parallel MPI processes of which there are *size* many.

In a first step, the Lanczos vectors used for the buildup of the Krylov space will be split into as many (almost) equal sections as there are ranks, i.e., $ranks = 0, 1, \ldots, size-1$. How many Lanczos vectors are stored depends on whether one wants to perform reorthogonalization against earlier vectors. The minimal number is two predecessor vectors to construct the new Lanczos vector. In a similar manner the Hamiltonian matrix is split into as many slices, see Fig. 1. The idea is to rotate the sections of one of the vectors among the MPI ranks and leave the matrix stripes stationary where they are created. In the scheme presented in Fig. 1 the source

vector v_{i-1} will be rotated. Thus the stripes of the Hamiltonian presented in the figure belong to H. It is as well possible to rotate the target vector (v_i), then the stripes would belong to H^T.

Figure 1 shows the first MPI-step of the matrix vector multiplication that is in parallel executed on each rank. The initially blank vector v_i is filled section-wise on each rank.

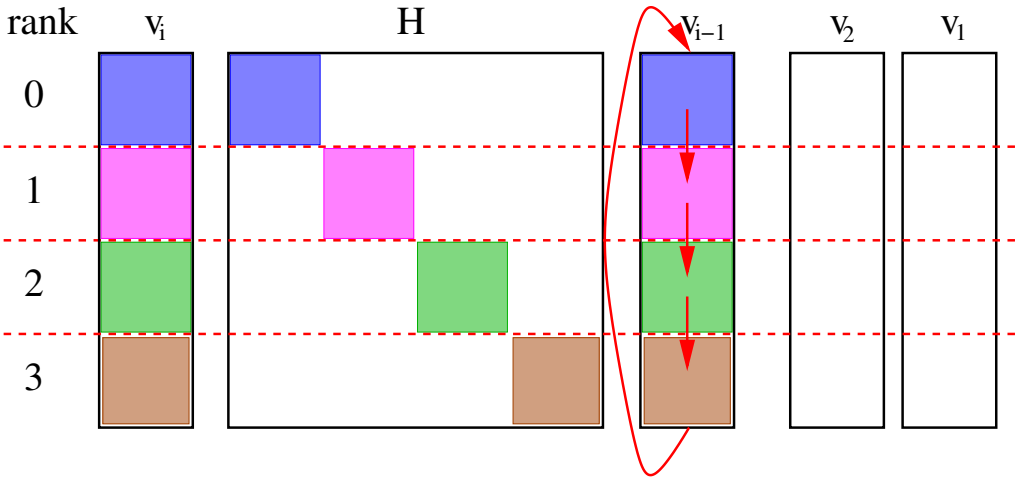

Figure 1. Sketch of the MPI parallelization of a matrix-vector multiplication. Each rank hosts a section of all vectors and a stripe of the matrix. The first step fills the initially blank target vector v_i with the contributions of the respective sections of v_{i-1} on the same rank. The needed parts of the matrix are given by colored blocks.

For the next MPI-step of the matrix-vector multiplication the sections of the source vector v_{i-1} are rotated using *mpi_sendrecv* which yields the configuration shown in Fig. 2. Now, the contributions of the respective sections of v_{i-1} on the same rank are added by using the correct part of the matrix strip given by the colored boxes. This process is continued until the vector v_{i-1} has completed a full cycle and is again of the structure shown in Fig. 1.

In a Lanczos procedure v_{i-1} and v_{i-2} have to be projected out of v_i, and for complete reorthogonalization all previous v_j. The advantage of the organization in stripes is that this can be done in parallel on every rank without any communication with other ranks (since we use an orthonormal basis). Only for the normalization of v_i we need an *mpi_allreduce* command to add up the contributions of every rank.

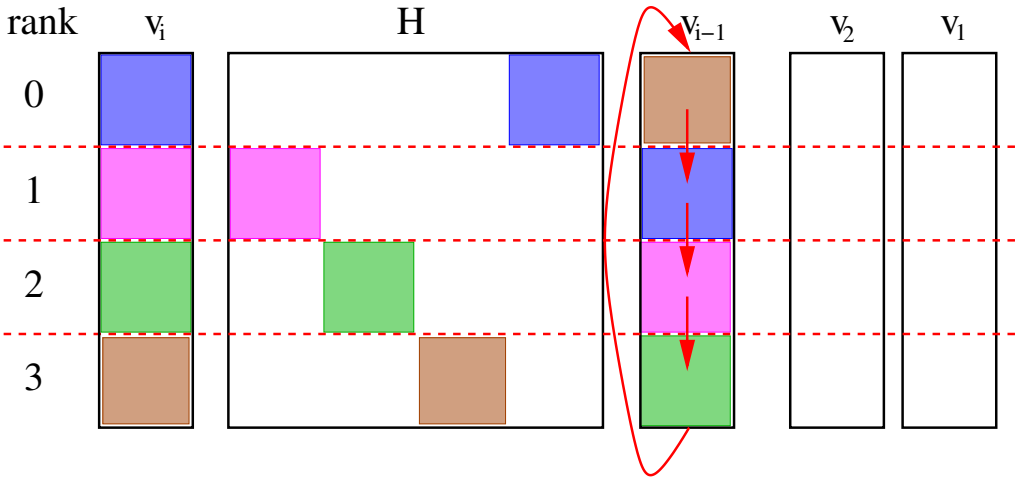

Figure 2. Second MPI-step of the matrix-vector multiplication; the contributions of the respective sections of v_{i-1} on the same rank are added by using the correct part of the matrix strip given by the colored boxes.

The multiplication of the submatrices with the respective sections of the vector is done on the GPU. For this purpose, the matrix stripes should be set up and then (permanently) stored on the GPU using *omp target teams* pragmas so that only the processed vector sections need to be uploaded to the GPU and the resulting sections downloaded, respectively. To this end, the needed space is allocated on the GPU, the sparse matrix (the full stripe

of H on each rank) is uploaded only once at the beginning, and the vectors on each rank are only updated using *omp target update to/from* pragmas.

For an easy manipulation, every stripe of H is subdivided into *size* blocks (colored boxed in Figs. 1 and 2) by transforming the two-dimensional stripe into a sequence of blocks, i.e., by reorganizing the two-dimensional array into three dimensions. Each of these matrix blocks is then stored as sparse matrix. We employ the COO (coordinate format) storage mode, i.e., store each sparse matrix as three one-dimensional arrays containing indices m and n of the non-zero matrix elements H_{mn}. Hermiticity is not employed by storing only matrix elements with $m \leq n$ since this would slow down the openMP loops for the matrix-vector multiplication on the GPU.

On the GPU the matrix vector multiplication is done by employing openMP for the loop about the entries of the three vectors. This would lead to *race conditions* due to the multiple appearance of the same index of the target vector in COO format, i.e., it would lead to wrong assignments of entries of v_i which would worsen the stability problems inherent in Lanczos procedures [40]. The problem could be cured by using the compressed sparse row (CSR) format, where a parallelization over row pointers would prevent the race condition, however, we used an alternative approach of grouping the indices into disjoint sets as shown in the example in Table I. This also simplifies the parallel creation of the matrix in COO format.

Table I. Example on how the sparse matrices are stored using the auxiliary index iCore to structure the respective vectors into blocks of disjoint indices m.

glob. index	iCore	loc. index	m	n	H_{mn}
1	1	1	1	5	30.1
2		2	1	7	−8.9
3		3	2	3	1.8
4		4	2	7	−2.5
5		5	2	9	3.4
6	2	1	3	2	1.8
7		2	3	5	−15.6

In Table I, the global index enumerates the non-zero matrix elements as well as the entries of the index vectors which in a simple openMP parallelization about the global index would lead to race conditions for identical m. This is cured with the help of an auxiliary index $iCore$ that groups indices into disjoint sets so that identical m are treated one after the other since the GPU-openMP parallelization runs about the index $iCore$.

III. RESULTS

A. HPC

The summary of the above is that the outlined programming scheme works. The program was compiled using the Intel® oneAPI® suite, and it runs on flexible numbers of ranks as tested on the Phase-2 supercomputer at LRZ employing nodes with two Intel® Xeon® Platinum 8480+ CPUs (Sapphire Rapids) and four Intel® Data Center GPUs Max 1550 (Ponte Vecchio).

A few characteristics can be deduced from the test runs performed at the LRZ and on a local AMD machine. The total linear dimension of the Hilbert space is $(2s+1)^N$ for N equal spins, and the number of non-zero matrix elements per row is of order $o(N^2)$ since the dipolar interaction connects all spins pairwise. However, the number of non-zero elements for each block of a stripe of the matrix in Figs. 1 and 2 can be very different which leads to an imbalance. For large projects with several nodes (up to 64 tested) MPI transfer times seem to dominate the run time. The time to load vectors to and from the GPU is small in comparison [41]. In the scheme employed so far additional idle times occur since each node has got many CPU cores which wait while the GPU is dealing with the matrix. The same is true if the CPU cores work and the GPU waits. Our experience collected so far suggests that one should fill the nodes with as much matrix and vector data as allowed by the available RAM. This reduces all communication times.

Realistic future improvements could be achieved by processing several random vectors at the same time, for instance the two time-reversed vectors, in order to attenuate random memory access on the GPU due to unstructured indices n in the sparse storage format. This would better the ratio of run time to load time of the GPU. In addition, one could alternate between vectors, i.e., process one group on the ranks while the other is MPI-transferred between ranks and vice versa. A more intricate improvement would be the parallel use of CPUs and GPUs.

B. FTLM for anisotropic spin rings

The scientific goal of the outlined numerical procedure is to model anisotropic spin systems by means of the approximate finite-temperature Lanczos procedure in cases where they are too large for exact diagonalization of the Hamiltonian and also too large for an FTLM approximation on a local workstation even with GPU [42]. The example spin system discussed in the following possesses a rather large single-ion anisotropy as e.g. in the case of dysprosium, manganese, or cobalt, see [43] for an example. In addition, anisotropic exchange might occur such as dipolar interactions. In order to estimate the accuracy and convergence of the above scheme we restrict our investigation to a fictitious ring molecule with $N = 6$ spins of spin quantum number $s = 5/2$ that can irrespective of the terms in the Hamiltonian be diagonalized numerically exactly. The Hamilton operator as given by Eq. (3) contains a nearest-neighbor Heisenberg exchange $J = 0.02$ K, strong non-collinear single-ion anisotropy $d_i = -20$ K as well as all-to-all dipolar interactions. The

single-ion anisotropy is dominant in the chosen example; the parameters are motivated by dysprosium ring molecules as e.g. in [43], but the specific values do not really matter for the statements made in the following.

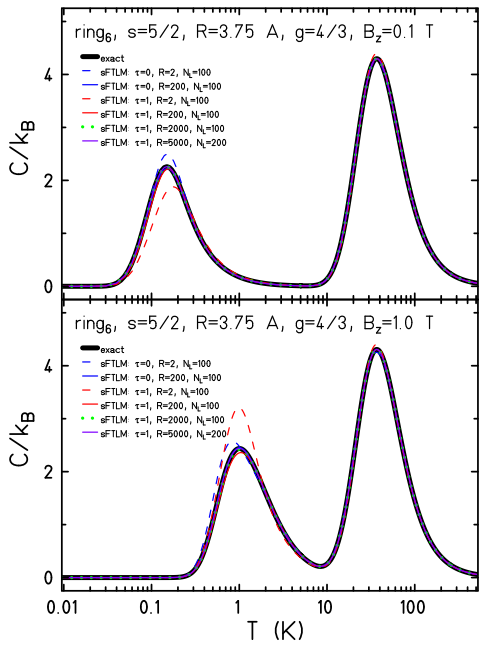

Figure 3. Heat capacity of a fictitious spin ring of $N = 6$ spins with $s = 5/2$ with nearest-neighbor Heisenberg exchange, strong single-ion easy-axis anisotropy as well as all-to-all dipolar interaction for $B = 0.1$ T (top) and $B = 1$ T (bottom). $\tau = 1$ or 0 denotes the use or not-use of time reversed random vectors; N_L provides the number of Lanczos steps for a single random vector, and R gives the number of employed random vectors according to (8) and (9).

Figure 3 shows as a first result the heat capacity for two values of the magnetic field applied along the z-axis which is the symmetry axis of the ring. The various curves origin from using time reversed random vectors ($\tau = 1$) or not ($\tau = 0$) together with N_L Lanczos steps for each random vector and averages of the partition function over R random vectors according to (8) and (9). One notices that the heat capacity is astonishingly robust and well approximated by already rather small numbers of random vectors and Lanczos steps of the order of 100. This even holds for the low-temperature peak of the heat capacity. Only for single-figured numbers of random vectors the accuracy is visibly lower, but only around the low-temperature peak. We conjecture that the reason is given by the fact, that the approximate energy eigenvalues converge much faster than the eigenstates as is known from perturbation theory. Therefore, the heat capacity, which depends only on the energy eigenvalues, appears rather accurate and robust with smaller effort. This changes drastically for the magnetization.

The magnetization, displayed in Fig. 4, turns out to be much more delicate, a behavior that can be traced

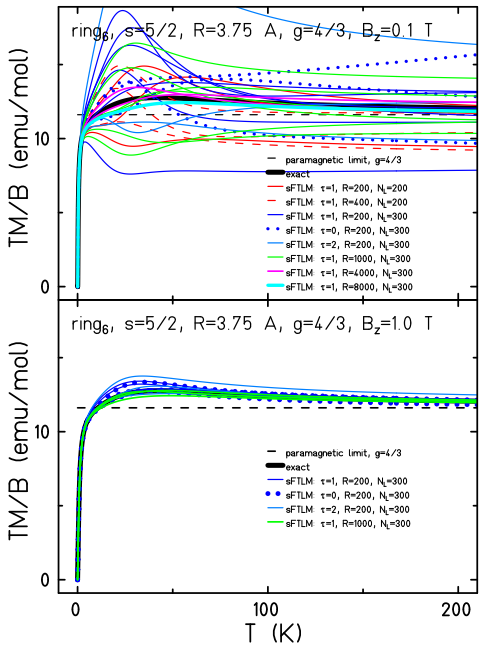

Figure 4. Susceptibility, MT/B, of a fictitious spin ring of $N = 6$ spins with $s = 5/2$ with nearest-neighbor Heisenberg exchange, strong single-ion easy-axis anisotropy as well as all-to-all dipolar interaction for $B = 0.1$ T (top) and $B = 1$ T (bottom). $\tau = 1$ or 0 denotes the use or not-use of time reversed random vectors; N_L provides the number of Lanczos steps for a single random vector, and R gives the number of employed random vectors according to (8) and (9).

back to the fact that now not only the approximate energy values $\epsilon_m^{(r)}$ matter, but also the matrix elements of the magnetization, compare (9). In addition, the typical presentation of susceptibility times temperature magnifies high-temperature inaccuracies. Such inaccuracies are more prominent for smaller values of the magnetic field. In such cases, the Zeeman term might be of the same order as FTLM accuracy fluctuations in energy and weights. On top of this, fluctuations of the magnetization are enlarged by dividing by a small field B.

This all changes for the larger field of $B = 1$ T (Fig. 4, top) which converges much more quickly for smaller numbers of both random vectors and Lanczos steps. In addition, the explicit use of time-reversal symmetry is no longer needed.

The various scenarios shown in Fig. 4 can be summarized as follows. Small numbers of random vectors up to at least $R = 200$ lead to wrong results, in particular for the case of small fields such as $B = 0.1$ T (top panel). The symmetric cases with $R = 200$ and $\tau = 1$ (blue curves) are somewhat better, but still wrong. Fortunately, even without the correct reference untrustworthy curves can be identified since they, e.g., do not approach the paramagnetic limit. The finer details, for instance where the curves bend, converge rather slowly in accordance with the observations in Ref. [8]. It came

as a surprise that even for $R = 8,000$ (cyan curve) the approximation does not converge to the exact result for the small field of $B = 0.1$ T and lower temperatures (top panel).

C. Mn_9 – a double pyramid

As an encouraging outlook we would like to present an example where the authors stated in their recent article that a numerical treatment of this spin system appears impossible to them, see [44].

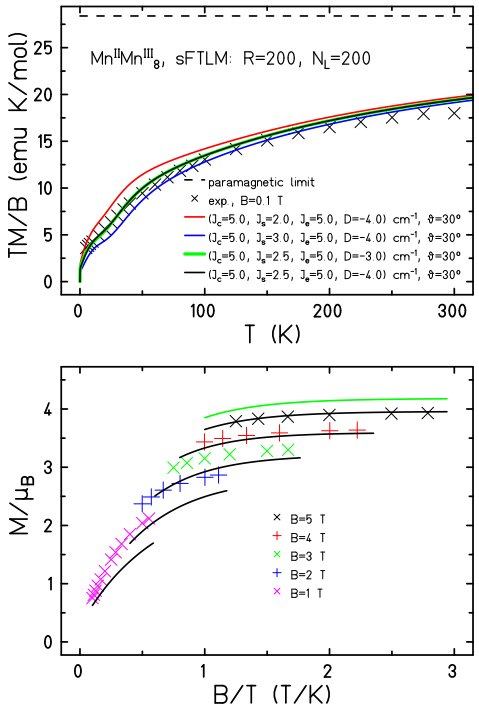

Figure 5. Susceptibility, MT/B, of Mn_9. N_L provides the number of Lanczos steps for a single random vector, and R gives the number of employed random vectors according to (8) and (9). Data taken from [44].

We model Mn_9 $=Mn^{II}Mn^{III}_8$ as a centered cube with three exchange interactions: J_c between the central Mn^{II} of spin $s = 5/2$ and the eight Mn^{III} of spin $s = 2$ at the vertices of the cube; J_s between nearest neighbors on the top and bottom squares, and J_e along the edges between top and bottom Mn^{III}. In addition, Mn^{III} possesses an easy axis of strength D that is tilted by 30° with respect to the symmetry axis through top and bottom square and directed along $\phi = \pm 10°$ of the x-axis. For details see [44]. We use powder averages about the vertices of a dodecahedron.

Since this is a paper on numerical methods we did not aim at a perfect fit, but rather want to demonstrate that with the proposed numerical scheme large anisotropic spin systems can be modelled. Therefore, we picked a few reasonable parameter set guided by our collected experience with Mn-based hourglas molecules [45–47]. Our best results, green and black curves in Fig. 5, show that Mn_9 with Hilbert space dimension of 2,343,750 can be modelled with resonable effort. We used 4 nodes on SuperMuc Phase 2 with 16 GPUs in total. For each field value and 10 directions the calculation took about 100 minutes. The convergence of this system is better than for the ring system discussed before since the single-ion anisotropy is much smaller compared to exchange and field. Therefore, the low-energy density of states is denser which enhances the convergence of FTLM.

If one would like to improve the agreement of theory and data, then more refined and maybe less symmetric parameterizations, compare[44], would have to be investigated.

IV. DISCUSSION

In view of the necessary large number of random vectors of the order of a few thousand as well as $N_L \approx 200 \ldots 300$ Lanczos steps for each of them, the use of massively parallel schemes is indispensable. A Hamiltonian matrix is applied $R \cdot N_L$ times, i.e., a hundred thousand times – this task is indeed very well suited for a GPU [36, 42]

There are a few possible future improvements that one could and should investigate. One option might be an improvement of the storage mode of the Hamiltonian matrix, compare Table I, and connected to this a more cache-efficient reading of the source vector on the GPU, e.g., by manipulating a group of vectors at the same time.

Another possible improvement would be a simultaneous use of GPUs and CPUs which are both idle while the other works. However, an obstacle is always given by the fear to develop too specialized code that does not run elsewhere or on potentiall different near-future architectures.

A third improvement concerns the rate of convergence of the Lanczos procedure when approximating a density of states that exhibits pronounced bunching. Here a representation by Chebyshev polynomials as proposed in [4] might be the better approach, however, without smoothing kernel as we demonstrated in [48].

ACKNOWLEDGMENT

I would like to thank my group at Bielefeld University for many fruitful discussions and for the patience with which they did stand my enthusiastic but long musings on matrices, MPI, openMP, and FORTRAN. I would like to thank Hans-Hermann Frese for a good introduction to MPI.

I would also like to thank Martin Ohlerich of the Leibniz supercomputing center (LRZ) of the Bavarian

Academy of Sciences for very many insightful discussions on the various computational problems and strategies.

Thanks go also to Jörg Schulenburg (Magdeburg University), Hans De Raedt and Fengping Jin (FZ Jülich), Momme Allalen (LRZ) as well as Christopher Danken and Tobias Klöffel (Intel) for discussions on HPC of matrix vector multiplications.

Computing time at the LRZ trough projects pn39ga & pn76ta is gratefully acknowledged.

[1] J. Skilling, Maximum entropy and Bayesian methods (Kluwer, Dordrecht, 1988) Chap. The eigenvalues of mega-dimensional matrices, pp. 455–466.
[2] M. Hutchinson, A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines, Comm. Statist. Simulation Comput. **18**, 1059 (1989).
[3] J. Jaklič and P. Prelovšek, Lanczos method for the calculation of finite-temperature quantities in correlated systems, Phys. Rev. B **49**, 5065 (1994).
[4] A. Weiße, G. Wellein, A. Alvermann, and H. Fehske, The kernel polynomial method, Rev. Mod. Phys. **78**, 275 (2006).
[5] J. Schnack, J. Richter, and R. Steinigeweg, Accuracy of the finite-temperature Lanczos method compared to simple typicality-based estimates, Phys. Rev. Research **2**, 013186 (2020).
[6] H. De Raedt and P. de Vries, Simulation of two and three-dimensional disordered systems: Lifshitz tails and localization properties, Z. Phys. B Condens. Matter **77**, 243 (1989).
[7] J. Jaklič and P. Prelovšek, Finite-temperature properties of doped antiferromagnets, Adv. Phys. **49**, 1 (2000).
[8] M. Aichhorn, M. Daghofer, H. G. Evertz, and W. von der Linden, Low-temperature Lanczos method for strongly correlated systems, Phys. Rev. B **67**, 161103(R) (2003).
[9] I. Zerec, B. Schmidt, and P. Thalmeier, Kondo lattice model studied with the finite temperature Lanczos method, Phys. Rev. B **73**, 245108 (2006).
[10] O. Hanebaum and J. Schnack, Advanced finite-temperature Lanczos method for anisotropic spin systems, Eur. Phys. J. B **87**, 194 (2014).
[11] R. Steinigeweg, J. Herbrych, F. Pollmann, and W. Brenig, Typicality approach to the optical conductivity in thermal and many-body localized phases, Phys. Rev. B **94**, 180401(R) (2016).
[12] B. Schmidt and P. Thalmeier, Frustrated two dimensional quantum magnets, Phys. Rep. **703**, 1 (2017).
[13] S. Okamoto, G. Alvarez, E. Dagotto, and T. Tohyama, Accuracy of the microcanonical Lanczos method to compute real-frequency dynamical spectral functions of quantum models at finite temperatures, Phys. Rev. E **97**, 043308 (2018).
[14] J. Schnack, J. Schulenburg, and J. Richter, Magnetism of the $N = 42$ kagome lattice antiferromagnet, Phys. Rev. B **98**, 094423 (2018).
[15] I. Rousochatzakis, S. Kourtis, J. Knolle, R. Moessner, and N. B. Perkins, Quantum spin liquid at finite temperature: Proximate dynamics and persistent typicality, Phys. Rev. B **100**, 045117 (2019).
[16] F. Jin, D. Willsch, M. Willsch, H. Lagemann, K. Michielsen, and H. De Raedt, Random state technology, J. Phys. Soc. Jpn. **90**, 012001 (2021).
[17] U. Manthe and F. Huarte-Larranaga, Partition functions for reaction rate calculations: statistical sampling and MCTDH propagation, Chem. Phys. Lett. **349**, 321 (2001).
[18] E. Pavarini, E. Koch, R. Scalettar, and R. M. Martin, eds., The physics of correlated insulators, metals, and superconductors (Verlag des Forschungszentrum Jülich, 2017) Chap. The Finite Temperature Lanczos Method and its Applications by P. Prelovšek, ISBN 978-3-95806-224-5, http://hdl.handle.net/2128/15283.
[19] S. M. Winter, K. Riedl, D. Kaib, R. Coldea, and R. Valentí, Probing $\alpha - \mathrm{RuCl_3}$ beyond magnetic order: Effects of temperature and magnetic field, Phys. Rev. Lett. **120**, 077203 (2018).
[20] Z. Ma, D. Ni, D. A. S. Kaib, K. MacFarquharson, J. S. Pearce, R. J. Cava, R. Valentí, R. Coldea, and A. I. Coldea, Anisotropic magnetic interactions in a candidate Kitaev spin liquid close to a metal-insulator transition, Communications Physics **7**, 390 (2024).
[21] D. Schray, D. Westerbeck, J. Braun, Y. Lan, S. Gomez-Coca, W. Wernsdorfer, E. Ruiz, C. E. Anson, J. Schnack, and A. K. Powell, Fe-Gd ferromagnetic cyclic coordination cluster $[\mathrm{Fe_4^{III}Gd_4^{III}(teaH)_8(N_3)_8(H_2O)}]$ with magnetic anisotropy – theory and experiment, Inorg. Chem. **62**, 6642 (2023).
[22] T. Munehisa, An improved finite temperature Lanczos method and its application to the spin-1/2 Heisenberg model on the kagome lattice, World Journal of Condensed Matter Physics **4**, 134 (2014).
[23] K. Morita and T. Tohyama, Finite-temperature properties of the kitaev-heisenberg models on kagome and triangular lattices studied by improved finite-temperature Lanczos methods, Phys. Rev. Research **2**, 013205 (2020).
[24] O. Waldmann, Symmetry and energy spectrum of high-nuclearity spin clusters, Phys. Rev. B **61**, 6138 (2000).
[25] R. Schnalle and J. Schnack, Calculating the energy spectra of magnetic molecules: application of real- and spin-space symmetries, Int. Rev. Phys. Chem. **29**, 403 (2010).
[26] S. Ghassemi Tabrizi and T. D. Kühne, Simultaneous spin and point-group adaptation in exact diagonalization of spin clusters, Magnetism **5**, 8 (2025).
[27] J. Schnack and O. Wendland, Properties of highly frustrated magnetic molecules studied by the finite-temperature Lanczos method, Eur. Phys. J. B **78**, 535 (2010).
[28] C. Lanczos, An iteration method for the solution of the eigenvalue problem of linear differential and integral operators, J. Res. Nat. Bur. Stand. **45**, 255 (1950).
[29] A. Wietek and A. M. Läuchli, Sublattice coding algorithm and distributed memory parallelization for large-scale exact diagonalizations of quantum many-body systems, Phys. Rev. E **98**, 033309 (2018).
[30] A. M. Läuchli, J. Sudan, and R. Moessner, $s = \frac{1}{2}$ kagome Heisenberg antiferromagnet revisited, Phys. Rev. B **100**, 155142 (2019).
[31] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C.

Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J. R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Y. Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. Rieffel, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis, Quantum supremacy using a programmable superconducting processor, Nature **574**, 505 (2019).

[32] H. Nakano, N. Todoroki, and T. Sakai, Haldane gaps of large-s Heisenberg antiferromagnetic chains and asymptotic behavior, J. Phys. Soc. Jpn. **88**, 114702 (2019).

[33] H. De Raedt, F. Jin, D. Willsch, M. Willsch, N. Yoshioka, N. Ito, S. Yuan, and K. Michielsen, Massively parallel quantum computer simulator, eleven years later, Comp. Phys. Commun. **237**, 47 (2019).

[34] J. Schnack, J. Schulenburg, A. Honecker, and J. Richter, Magnon crystallization in the kagome lattice antiferromagnet, Phys. Rev. Lett. **125**, 117207 (2020).

[35] H. Nakano, H. Tadano, N. Todoroki, and T. Sakai, The Haldane gap of the $s = 1$ Heisenberg antiferromagnetic chain, J. Phys. Soc. Jpn. **91**, 074701 (2022).

[36] D. Willsch, M. Willsch, F. Jin, K. Michielsen, and H. De Raedt, GPU-accelerated simulations of quantum annealing and the quantum approximate optimization algorithm, Comp. Phys. Commun. **278**, 108411 (2022).

[37] H. Nakano and T. Sakai, Large-scale numerical-diagonalization study of the Shastry–Sutherland model, JPS Conf. Proc. **38**, 011166 (2023).

[38] J. Schulenburg, *spinpack 2.58*, Magdeburg University (2019).

[39] J. Richter and J. Schulenburg, The spin-1/2 $J_1 - J_2$ Heisenberg antiferromagnet on the square lattice: Exact diagonalization for $N = 40$ spins, Eur. Phys. J. B **73**, 117 (2010).

[40] J. Eckseler, M. Pieper, and J. Schnack, Escaping the Krylov space during the finite-precision Lanczos algorithm, Phys. Rev. E **112**, 025306 (2025).

[41] Thanks to help by Tobias Klöffel (Intel).

[42] S. G. Tabrizi and T. D. Kühne, GPU-accelerated finite-temperature Lanczos method for Heisenberg spin systems with non-Abelian permutation symmetries, arXiv:2607.17429 (2026).

[43] S. Biswas, S. Das, J. Acharya, V. Kumar, J. van Leusen, P. Kögerler, J. M. Herrera, E. Colacio, and V. Chandrasekhar, Homometallic Dy^{III} complexes of varying nuclearity from 2 to 21: Synthesis, structure, and magnetism, Chem. Eur. J. **23**, 5154 (2017).

[44] K. H. Baka, M. M. Kallio, V. Vieru, A. M. Mowson, E. Koumousi, C. Raptopoulou, V. Psycharis, W. Wernsdorfer, E. Moreno-Pineda, G. Christou, L. Chibotaru, and T. C. Stamatatos, Exchange-transferred anisotropy: An isotropic mn^{ii} ion governs single-molecule magnetism in a $\{Mn_9\}$ double-decker metallacrown, Inorg. Chem. Front. (2026).

[45] V. Hoeke, K. Gieb, P. Müller, L. Ungur, L. F. Chibotaru, M. Heidemeier, E. Krickemeyer, A. Stammler, H. Bögge, C. Schröder, J. Schnack, and T. Glaser, Hysteresis in the ground and excited spin state up to 10 T of a $[Mn_6^{III}Mn^{III}]^{3+}$ triplesalen single-molecule magnet, Chem. Sci. **3**, 2868 (2012).

[46] V. Hoeke, M. Heidemeier, E. Krickemeyer, A. Stammler, H. Bögge, J. Schnack, and T. Glaser, Structural influences on the exchange coupling and zero-field splitting in the single-molecule magnet $[Mn_6^{III}Mn^{III}]^{3+}$, Dalton Trans. **41**, 12942 (2012).

[47] T. Glaser, V. Hoeke, K. Gieb, J. Schnack, C. Schröder, and P. Müller, Quantum tunneling of the magnetization in $[Mn_6^{III}M]^{3+}$ (M= Cr^{III}, Mn^{III}) SMMs: Impact of molecular and crystal symmetry, Coord. Chem. Rev. **289-290**, 261 (2015).

[48] H. Schlüter, F. Gayk, H.-J. Schmidt, A. Honecker, and J. Schnack, Accuracy of the typicality approach using Chebyshev polynomials, Z. Naturforsch. A **76**, 823 (2021).