All-mode renormalization for tensor network with stochastic noise

Erika Arai
Department of Physics, Nara Women's University, Nara 630-8506, Japan

Hiroshi Ohki
Department of Physics, Nara Women's University, Nara 630-8506, Japan
and RIKEN BNL Research Center, Brookhaven National Laboratory, Upton New York 11973, USA

Shinji Takeda
Institute for Theoretical Physics, Kanazawa University, Kanazawa 920-1192, Japan

Masaaki Tomii
Physics Department, University of Connecticut, Storrs, Connecticut 06269-3046, USA

(Received 23 November 2022; accepted 31 May 2023; published 20 June 2023)

In usual (nonstochastic) tensor network calculations, the truncated singular value decomposition is often used for approximating a tensor, and it causes systematic errors. By introducing stochastic noise in the approximation, however, one can avoid such systematic errors at the expense of statistical errors, which can be straightforwardly controlled. Therefore in principle, exact results can be obtained even at finite bond dimension up to the statistical errors. A previous study of the unbiased method implemented in tensor renormalization group algorithm, however, showed that the statistical errors for physical quantity are not negligible, and furthermore the computational cost is linearly proportional to a system volume. In this paper, we introduce a new way of stochastic noise such that the statistical error is suppressed, and moreover, in order to reduce the computational cost we propose common noise method whose cost is proportional to the logarithm of volume. We find that the method provides better accuracy for the free energy compared with the truncated singular value decomposition when applying to tensor renormalization group for Ising model on square lattice. Although the common noise method introduces systematic error originated from a correlation of noises, we show that the error can be described by a simple functional form in terms of the number of noises, thus the error can be straightforwardly controlled in an actual analysis. We also apply the method to the graph independent local truncation algorithm and show that the accuracy is further improved.

DOI: 10.1103/PhysRevD.107.114515

I. INTRODUCTION

A numerical renormalization group (RG) technique based on the tensor networks becomes a popular and powerful tool to study the quantum lattice field theory and the many-body systems in condensed matter physics. A typical and simple algorithm for a tensor contraction is the tensor renormalization group (TRG) approach [1], where the tensor contraction is performed using the low-rank approximation via the singular value decomposition (SVD). The lattice topology is unchanged after a tensor reconstruction step, thus this coarse graining procedure can be repeated and a large volume simulation can be easily realized. Another striking feature of the TRG is that there is no sign problem and hence TRG-related methods can be applied to systems such as a complex action with θ term, the real-time dynamics that are not easily accessible by the Monte Carlo methods [2–14].[1]

A key ingredient of the TRG is using the truncated SVD when decomposing a tensor, and this plays an important role to realize an efficient and sustainable coarse-graining algorithm. The truncation of the lower modes, however, causes a systematic error, and in general it is difficult to predict a scaling property of the truncation error especially

[1]For recent lattice studies of tensor network with applied quantum computing and for a system with higher dimensions, see reviews [15,16] and references therein.

when repeating several coarse-graining steps. Such a truncation technique is commonly used in the improved coarse-graining algorithms such as tensor network renormalization (TNR) [17], loop-TNR [18], Gilt [19] and also in the variety of efficient cost reduction algorithms [20–34]. In fact, for high precision calculations a careful treatment of the systematic error estimate is needed, so that it is important to pursue a possibility of alternative algorithms that can improve the error evaluation method as well as the numerical accuracy itself. One of possibilities to remove the systematic error is to use the Monte Carlo tensor network (MCTN) [35,36], in which a Monte Carlo sampling of the singular modes is employed and there is only a statistical error. For different Monte Carlo approaches in variational methods in a tensor network representation see [37–46].

In this work, following the strategy of the MCTN we propose a new stochastic method for tensor decomposition and apply it to the TRG. A basic idea of using the stochastic method is as follows. In the usual truncated SVD, the lower modes are just discarded, and such a low rank approximation is known to be the best one from a local point of view but not necessarily for global quantity like a partition function. The contribution of the lower modes, however, can be incorporated into decomposed tensors by combining them with stochastic noises. By using the tensors which are compact but contain all singular modes (all-mode renormalization), the coarse graining can be done as in the case of the TRG. Note that in the stochastic method the truncation error can be replaced with an statistical error due to the noise, and a stochastic determination of the partition function and any related physical quantities is possible. We examine two types of spatial distribution of noise: position-dependent and -independent ways (for the latter case we refer it as the common noise method). As for the position-dependent noise method it is applicable to the system that has no translation invariance thus the computational complexity scales as its volume. In this case there is no systematic error and it provides an unbiased result, which will be directly confirmed by a numerical calculation. As for the common noise method, the order of the computational cost remains the same as the TRG, while as will be discussed there is a noise cross contamination effect, which turns out to be the only source of the residual systematic error. Nevertheless we find a significant improvement of the accuracy compared to the TRG. Since this residual systematic error has a simple scaling property due to a nature of the random noise, we provide a systematic error evaluation method, which is independent of model dynamics and tensor network algorithms. Thus our new stochastic noise method combining with the deterministic algorithm can actually improve the error estimation as well as the numerical accuracy. Moreover, our method is so simple that can easily be applied to a complicated system and combined with improved tensor network algorithms.

Our method shares the same idea as the MCTN, but there are some practically important differences. The MCTN uses a subset of the singular modes for a tensor decomposition that are randomly chosen with an appropriate probability distribution (see [35] for details), while in our case all the singular modes are manifestly included thanks to the random noise vectors. Another important difference is that we propose the common noise method that is not considered in [35,36], where the position-dependent method in our language is only proposed and examined. As mentioned before, the common noise method has the desired properties: it provides better accuracy, the residual systematic error is under control, and the computational cost is the same order as the TRG. Therefore we consider that the common noise method is practically useful for future applications.

The paper is organized as follows. In Sec. II we review the TRG algorithm and introduce a tensor decomposition method using noise vectors. In Sec. III after introducing an ensemble method, we define the position-dependent and the common noise method. We test the noise methods for the 2D Ising model in Sec. IV. We discuss a possible application to other TRG algorithms and show some numerical results when applying to the Gilt in Sec V. Section VI is devoted to the conclusion. In the Appendix, the results for the specific heat are shown as an example of related thermodynamical quantities. Our preliminary result has been published in [47].

II. TENSOR DECOMPOSITION WITH RANDOM NOISE VECTOR

We first review the TRG method for the 2D Ising model on a square lattice. The partition function Z for the model is given as a trace over tensor indices in a tensor network representation,

$$Z \equiv \sum_{\{\sigma\}} e^{-\beta H[\sigma]} = \sum_{ijk\cdots} T_{ijkl} T_{mnip} \cdots = \mathrm{Tr}[\otimes T], \quad (1)$$

where T_{ijkl} is the initial tensor. We express the tensor in a matrix representation, $T_{ijkl} = M_{ij;kl} \equiv M_{ab}$, and $T_{ijkl} = \tilde{M}_{jk;li} \equiv \tilde{M}_{cd}$ with different combination of the indices. Using the SVD, we represent a matrix M with a rank R as follows:

$$M_{ab} = \sum_{s=1}^{R} \sqrt{\Lambda_s}\, u_{as} \sqrt{\Lambda_s} v_{sb}, \quad (2)$$

where Λ_s are singular values $(\Lambda_1 \geq \Lambda_2 \geq \Lambda_3 \geq \cdots)$. By truncating the lower modes and keeping the largest $D_{\mathrm{svd}} (\leq R)$ mode, the fourth-order tensor is approximated by a product of two third-order tensors

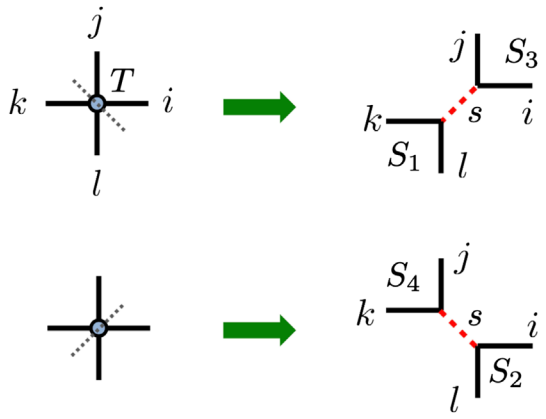

FIG. 1. Decomposition of a tensor T_{ijkl} to two third-order tensors $S_{3as}S_{1sb}$ (upper panel) and $S_{4cs}S_{2sd}$ (lower panel).

$$T_{ijkl} = M_{ab} \simeq \sum_{s=1}^{D_{\rm svd}} \sqrt{\Lambda_s}\, u_{as} \sqrt{\Lambda_s} v_{sb} = \sum_{s=1}^{D_{\rm svd}} S_{3as} S_{1sb}, \quad (3)$$

where $S_{3as} = \sqrt{\Lambda_s} u_{as}$ and $S_{1sb} = \sqrt{\Lambda_s} v_{sb}$ are the third-order tensors. Similarly $S_{2,4}$ are also defined to approximate the other type of a tensor decomposition, $\tilde{M}_{cd} \simeq \sum_{s=1}^{D_{\rm svd}} S_{4cs} S_{2sd}$. See Fig. 1 for pictorial expression of the decompositions. We then obtain a coarse grained tensor by contracting all old indices (i, j, k, l) of the four third-order tensors of $S_{1,2,3,4}$ (see Fig. 2)

$$T'_{i'j'k'l'} = \sum_{ijkl} S_{1i';ji} S_{2j';kj} S_{3lk;k'} S_{4il;l'} \equiv {\rm Tr}[S_{1i'} S_{2j'} S_{3k'} S_{4l'}], \quad (4)$$

where Tr stands for contracting all old indices. By repeating the coarse-graining n-times, we obtain a renormalized tensor $T^{(n)}_{i'j'k'l'}$ as

$$T^{(n)}_{i'j'k'l'} = {\rm Tr}\left[S^{(n-1)}_{1i'} S^{(n-1)}_{2j'} S^{(n-1)}_{3k'} S^{(n-1)}_{4l'}\right], \quad (n > 0) \quad (5)$$

where $T^{(0)} = T$ and $S^{(0)}_{1,2,3,4} = S_{1,2,3,4}$. The bond dimension of the renormalized tensor is $D_{\rm svd}$ and the dominant computational cost for contractions scales as $\mathcal{O}(D^6_{\rm svd})$.

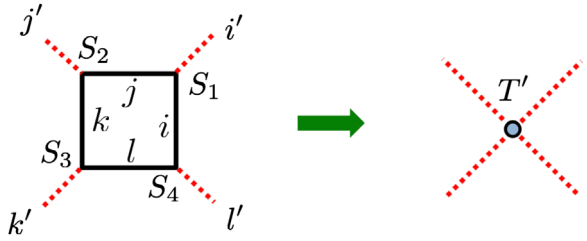

FIG. 2. Contraction to create a coarse grained tensor $T'_{i'j'k'l'}$ from the third-order tensors $S_{1,2,3,4}$.

Next let us explain our noise method. In the tensor decomposition step in the TRG, the truncated SVD is used, and it causes systematic errors. In order to reduce the systematic uncertainty with a limited number of bond dimensions, we introduce D-dimensional noise vectors $\eta_r = (\eta_{1r}, \eta_{2r}, \ldots, \eta_{Dr})^T$ for $r = 1, \ldots, N_r$, where each component η_{ir} is an element of Z_N. We then define a $D \times N_r$ matrix $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_{N_r})$, which satisfies the completeness condition

$$\frac{1}{N_r} \boldsymbol{\eta} \cdot \boldsymbol{\eta}^\dagger = \frac{1}{N_r} \sum_{r=1}^{N_r} \eta_r \cdot \eta_r^\dagger = \mathbf{1}_{D\times D} + \mathcal{O}(1/\sqrt{N_r}), \quad (6)$$

where we note that for Z_N noise the diagonal part is exactly unity and the statistical fluctuation appears only in the off-diagonal part. We use the random noise vector[2] in the decomposition of a matrix using SVD. Substituting $\mathbf{1}_{D\times D} \simeq \frac{1}{N_r} \boldsymbol{\eta} \cdot \boldsymbol{\eta}^\dagger$ into the lower mode subspace in Eq. (2) with $D = R - D_{\rm svd}$, the matrix M is given by

$$\begin{aligned} M_{ab} = & \sum_{s=1}^{D_{\rm svd}} \left(\sqrt{\Lambda_s} u_{as} \sqrt{\Lambda_s} v_{sb} \right) \\ & + \sum_{s,t=D_{\rm svd}+1}^{R} \sqrt{\Lambda_s}\, u_{as} \left(\frac{1}{N_r} \boldsymbol{\eta} \cdot \boldsymbol{\eta}^\dagger \right)_{s-D_{\rm svd}, t-D_{\rm svd}} \sqrt{\Lambda_t} v_{tb} \\ & + \mathcal{O}\left(\sqrt{\frac{\Lambda_{D_{\rm svd}+1} \Lambda_{D_{\rm svd}+2}}{N_r}} \right), \end{aligned} \quad (7)$$

where the first term represents the contributions from the largest $D_{\rm svd}$ singular value modes and the second term contains the residual contributions that are discarded in the original TRG. Since the diagonal element of $\frac{1}{N_r} \boldsymbol{\eta} \cdot \boldsymbol{\eta}^\dagger$ is unity for Z_N noise, we obtain the following expression:

$$M_{ab} = \sum_{s=1}^{R} (\sqrt{\Lambda_s} u_{as} \sqrt{\Lambda_s} v_{sb}) + \mathcal{O}\left(\sqrt{\frac{\Lambda_{D_{\rm svd}+1} \Lambda_{D_{\rm svd}+2}}{N_r}} \right). \quad (8)$$

Thus we can see that all the singular modes are explicitly included in the decomposition as shown in the first term and the statistical fluctuation represented in the second term is suppressed by smaller singular values $\Lambda_{D_{\rm svd}+1}$, $\Lambda_{D_{\rm svd}+2} \ll \Lambda_1$. This decomposition suggests to modify the third-order tensor $S_{1,2,3,4}$ to include the noise vector parts. For this purpose, we define the following modified third-order tensors as a function of the noise vectors,

[2]Although $\boldsymbol{\eta}$ is a matrix, it is convention to call it noise vector.

$$\bar{S}_{3as}(\boldsymbol{\eta}) \equiv \begin{cases} \sqrt{\Lambda_s} u_{as} & (1 \le s \le D_{\rm svd}) \\ \sum_{i=D_{\rm svd}+1}^{R} \sqrt{\frac{\Lambda_i}{N_r}} u_{ai} \eta_{i-D_{\rm svd},s-D_{\rm svd}} & (D_{\rm svd}+1 \le s \le D_{\rm svd}+N_r) \end{cases},$$
$$\bar{S}_{1sb}(\boldsymbol{\eta}) \equiv \begin{cases} \sqrt{\Lambda_s} v_{sb} & (1 \le s \le D_{\rm svd}) \\ \sum_{i=D_{\rm svd}+1}^{R} \sqrt{\frac{\Lambda_i}{N_r}} \eta^*_{i-D_{\rm svd},s-D_{\rm svd}} v_{ib} & (D_{\rm svd}+1 \le s \le D_{\rm svd}+N_r) \end{cases}, \tag{9}$$

and $\bar{S}_{2,4}(\boldsymbol{\eta})$ are also defined similarly. Using the modified tensors $\bar{S}_{1,2,3,4}$ and Eq. (6) we immediately see that the original matrix M is recovered $M_{ab} = \lim_{N_r\to\infty} \sum_{s=1}^{D_{\rm svd}+N_r} \bar{S}_{3as}\bar{S}_{1sb}$ for any value of $D_{\rm svd}$. We also note that in practice Z_2 noise is useful to avoid a complex valued tensor and we will exclusively use the real valued noise in our calculation given in Sec. IV.

Since the modified third-order tensors contain all the singular modes, there is only a statistical error due to the noise vectors instead of the truncation error, which ensures that the original matrix should be reproduced up to the statistical error. Note that a similar decomposition has already been employed in the low-mode approximation for the Dirac propagators in lattice QCD [48], where the Dirac propagator is dominated by a low-mode part of the Dirac eigenvectors, and a high-mode part is stochastically evaluated using the random noise vectors. The key property of the low-mode approximation of the inverse matrix is reflected in our method.

III. NOISE ENSEMBLE METHOD

First we simply use the modified third-order tensor $\bar{S}$ instead of the original one in the TRG, however, we do not observe any improvement on the accuracy of the physical quantities, because the low-rank approximation in the noise method is not as good as the truncated SVD in the sense of the Frobenius norm due to the theorem by Eckhart, Young, and Mirsky. Thus simply increasing the noise dimension N_r may not give a better accuracy compared to the normal TRG when the same number of the bond dimension is used. Therefore, instead of increasing N_r, we adopt an ensemble approach. Namely, we generate an ensemble of random noise vectors $\boldsymbol{\eta}^{[\ell]}$ $(\ell = 1, \ldots, N)$ with N being the total number of statistics. Our strategy is that a matrix decomposition is approximately obtained from an ensemble average of the modified third-order tensors with keeping a smaller value of N_r.

$$M_{ab} = \frac{1}{N}\sum_{\ell=1}^{N}\sum_{s=1}^{D_{\rm svd}+N_r} \bar{S}_{3as}(\boldsymbol{\eta}^{[\ell]})\bar{S}_{1sb}(\boldsymbol{\eta}^{[\ell]}) + \mathcal{O}\left(\sqrt{\frac{\Lambda_{D_{\rm svd}+1}\Lambda_{D_{\rm svd}+2}}{N_r N}}\right). \tag{10}$$

Thus, an original matrix can be reproduced in the limit of infinite statistics ($N \to \infty$) while keeping N_r finite. Our new stochastic method can be implemented in the TRG, and in the next subsections, we discuss how to obtain a renormalized tensor with two kinds of spatial noise distributions: position-dependent and -independent ways.

A. Coarse graining with position-dependent noise

The first step is to decompose the original tensor as in Eq. (10) and obtain the third-order tensors $\bar{S}_{1,2,3,4}$. Since the decomposition is separately done for each lattice site i, the noise vector is position dependent $\boldsymbol{\eta}_i^{[\ell]}$. Using $\bar{S}$ we define a renormalized tensor $T^{(1)[\ell]}_{i'j'k'l'}(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3, \boldsymbol{\eta}_4)$ as

$$T^{(1)[\ell]}_{i'j'k'l'}(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3, \boldsymbol{\eta}_4) = {\rm Tr}\left[\bar{S}_{1i'}(\boldsymbol{\eta}_1^{[\ell]})\bar{S}_{2j'}(\boldsymbol{\eta}_2^{[\ell]})\bar{S}_{3k'}(\boldsymbol{\eta}_3^{[\ell]})\bar{S}_{4l'}(\boldsymbol{\eta}_4^{[\ell]})\right], \tag{11}$$

where the tensor indices i', j', k', and l' run from 1 to $D_{\rm svd}+N_r$, and hence the bond dimension $D_{\rm cut}$ for a renormalized tensor $T^{(1)}$ is given as $D_{\rm cut} = D_{\rm svd} + N_r$. By iterating this RG process n times, we obtain a sequence of renormalized tensors for each sample ℓ,

$$T^{(0)} \to T^{(1)[\ell]} \to T^{(2)[\ell]} \cdots \to T^{(n)[\ell]}, \quad (\ell = 1, \ldots, N). \tag{12}$$

Since the sets of random noise vectors are generated for each RG step and for each site independently, the noise vectors satisfy the following property in orthogonality,

$$\frac{1}{N}\sum_{\ell=1}^{N}\frac{1}{N_r}\boldsymbol{\eta}_i^{[\ell]} \cdot \boldsymbol{\eta}_j^{[\ell]\dagger} = \delta_{i,j}\mathbf{1}_{D\times D} + \mathcal{O}(1/\sqrt{N_r N}), \tag{13}$$

and this orthogonal relation can be generalized to arbitrary order tensor products. For example, the tensor product of two independent noise vectors of $\boldsymbol{\eta}_i^{[\ell]}$ and $\boldsymbol{\eta}_j^{[\ell]}$ for different sites $i \neq j$ should also have an orthogonality

$$\frac{1}{N}\sum_{\ell=1}^{N}\frac{1}{N_r}(\boldsymbol{\eta}_i^{[\ell]} \cdot \boldsymbol{\eta}_i^{[\ell]\dagger})_{ab} \otimes \frac{1}{N_r}(\boldsymbol{\eta}_j^{[\ell]} \cdot \boldsymbol{\eta}_j^{[\ell]\dagger})_{cd} = (\mathbf{1}_{D\times D})_{ab} \otimes (\mathbf{1}_{D\times D})_{cd} + \mathcal{O}(1/\sqrt{N_r N}) \quad (\text{for } i \neq j). \tag{14}$$

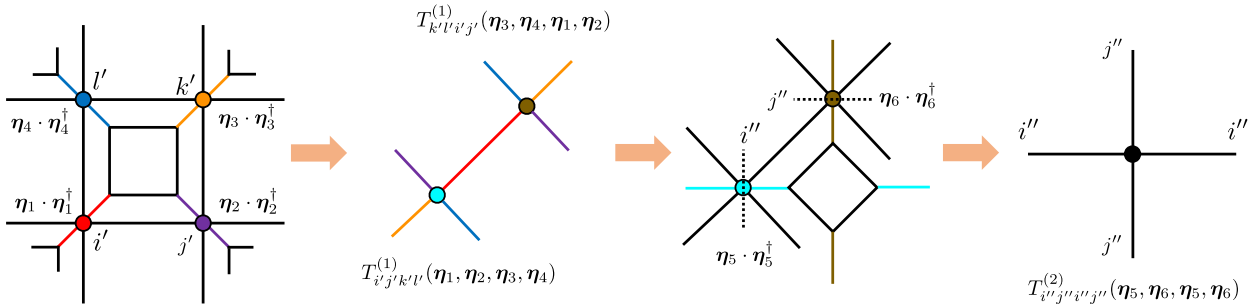

FIG. 3. Coarse-graining of the tensor network using the position-dependent noise for 2×2 lattice with the periodic boundary condition for all directions. In the initial lattice four independent noise vectors $\boldsymbol{\eta}_{1,2,3,4}$ are distributed to the four sites, and the initial tensors are decomposed. The contraction for each plaquette produces a renormalized tensor $T^{(1)}$ on 1×2 lattice. And then two different noise vectors $\boldsymbol{\eta}_{5,6}$ are distributed and the tensors are decomposed. Finally after contracting for one plaquette, a trace of a renormalized tensor $T^{(2)}$ is obtained.

Thanks to these properties, the truncation error in the tensor decomposition is completely replaced with the statistical error of the random noise. The exact partition function Z_V for a finite volume $V=2^n$ can be obtained from the renormalized tensor $T^{(n)[\ell]}$ in the limit of infinite statistics

$$Z_V=\lim_{N\to\infty}\frac{1}{N}\sum_{\ell=1}^{N}Z(T^{(n)[\ell]}), \tag{15}$$

where $Z(T^{(n)[\ell]})$ is a sample of the partition function,

$$Z(T^{(n)[\ell]})=\mathrm{Tr}[T^{(n)[\ell]}]=\sum_{i,j=1}^{D_{\mathrm{cut}}}T^{(n)[\ell]}_{ijij}. \tag{16}$$

See Fig. 3 for multiple coarse-graining procedures in this method.

At each coarse-graining step, we obtain an ensemble of tensor configurations so that this method is similar to a Monte Carlo method for configuration generations, while in our case the tensor variables are stochastically generated in analogous to a random walk and hence there is no sign problem and the individual configurations are completely uncorrelated. Thus individual configurations can be efficiently generated by using massive parallel computers. Furthermore since the contribution of all modes of SVD are maintained in the coarse-graining step, it is mathematically guaranteed that an exact result should be obtained within the statistical error. On the other hand, the lattice homogeneity is completely broken due to the position-dependent noise, thus we need to separately calculate a renormalized tensor for each plaquette, where we also have to take into account the boundary conditions in a finite volume. Therefore, the system volume V has to be fixed beforehand, which restricts on the maximum number of the coarse-graining processes. If we repeat the processes until obtaining a single tensor, the total computational cost will be as expensive as $\mathcal{O}(NVD_{\mathrm{cut}}^6)$, so that it spoils one of advantages in the original TRG methods: a logarithmic computational cost on the volume. All the above properties are already pointed out in the previous studies of the MCTN [35,36]. The difference between the MCTN and our position-dependent noise method lies in the probability distribution of which modes are taken in. Thanks to the random noise vectors, the renormalized tensors are compact but contain all singular modes (all-mode renormalization). We consider that our case is rather simple to implement and can easily control the statistical error. In the next section we shall numerically study the position-dependent method in details.

B. Coarse graining with common noise

In order to preserve the lattice homogeneity we consider a common (position-independent) noise. For this purpose we use a common set of noise vectors to obtain $\bar{S}_{1,3}$ or $\bar{S}_{2,4}$,

$$T_{ijkl}=M_{ab}\simeq\frac{1}{N}\sum_{\ell=1}^{N}\sum_{s=1}^{D_{\mathrm{svd}}+N_r}\bar{S}_{3as}(\boldsymbol{\eta}_1{}^{[\ell]})\bar{S}_{1sb}(\boldsymbol{\eta}_1{}^{[\ell]}), \tag{17}$$

$$T_{ijkl}=\tilde{M}_{cd}\simeq\frac{1}{N}\sum_{\ell=1}^{N}\sum_{s=1}^{D_{\mathrm{svd}}+N_r}\bar{S}_{4cs}(\boldsymbol{\eta}_2{}^{[\ell]})\bar{S}_{2sd}(\boldsymbol{\eta}_2{}^{[\ell]}), \tag{18}$$

where two sets of noise vectors $\boldsymbol{\eta}_1^{[\ell]}$ and $\boldsymbol{\eta}_2^{[\ell]}$ are independent with each other in general (see Fig. 4). The renormalized tensors are then obtained as

$$T^{(1)[\ell]}_{i'j'k'l'}(\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\boldsymbol{\eta}_1,\boldsymbol{\eta}_2)=\mathrm{Tr}\left[\bar{S}_{1i'}(\boldsymbol{\eta}_1^{[\ell]})\bar{S}_{2j'}(\boldsymbol{\eta}_2^{[\ell]})\bar{S}_{3k'}(\boldsymbol{\eta}_1^{[\ell]})\bar{S}_{4l'}(\boldsymbol{\eta}_2^{[\ell]})\right]. \tag{19}$$

Since in this method the same renormalized tensor is obtained at every site after a coarse graining, we only need to calculate a tensor contraction at a single plaquette for each RG step. Thus the order of the computational cost per sample remains the same as that of the original TRG. We refer this method as common noise method on normal lattice.

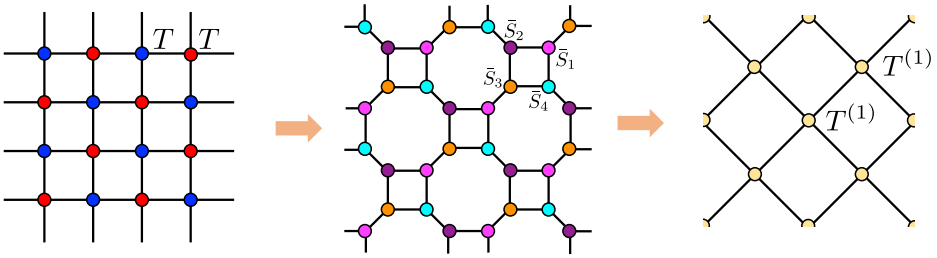

FIG. 4. A coarse-graining of the tensor network with common noises. In left panel, an initial tensor T is uniformly put on each site, while even and odd sites are distinguished by the different sets of the noise vectors represented by two different colors on sites, that is, the common noise $\boldsymbol{\eta}_1^{[\ell]}$ ($\boldsymbol{\eta}_2^{[\ell]}$) is used on all red (blue) sites. In middle panel, the third-order tensors $\bar{S}_{1,2,3,4}$ are obtained by a tensor decomposition. Right panel shows that after coarse graining a renormalized tensor $T^{(1)}$ is commonly obtained for each site, namely we obtain a uniform tensor network consisting of $T^{(1)}$.

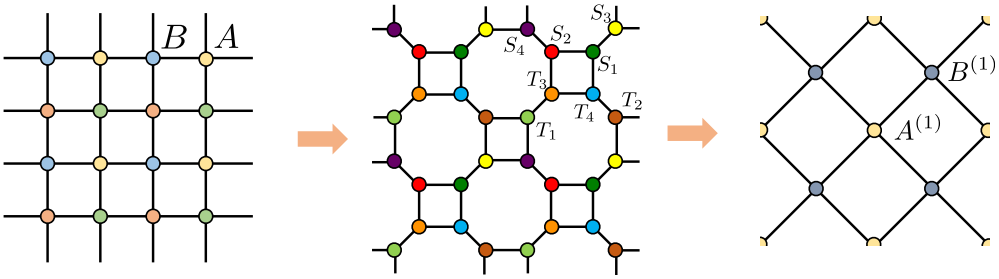

FIG. 5. Coarse graining of the tensor network with common noises for checkerboard lattice.

We note, however, that there exists an unwanted additional systematic error that arises from a contact term of the same noise vectors. Let $\boldsymbol{\eta}$ be a noise vector, which satisfies the relation

$$\frac{1}{N_r}(\boldsymbol{\eta}\cdot\boldsymbol{\eta}^\dagger)_{ab} = \frac{1}{N_r}\sum_r^{N_r}\eta_{ar}\eta_{br}^* = \delta_{ab} + \mathcal{O}\left(1/\sqrt{N_r}\right). \quad (20)$$

Since we have to commonly use $\boldsymbol{\eta}$ for tensor decompositions at two different sites, e.g., see the argument of $\bar{S}_1$ and $\bar{S}_3$ in Eq. (19), we encounter the following tensor product in the tensor contraction for a single plaquette as in Eq. (19),

$$\begin{aligned}
&\frac{1}{N_r^2}(\boldsymbol{\eta}\cdot\boldsymbol{\eta}^\dagger)_{ab}\otimes(\boldsymbol{\eta}\cdot\boldsymbol{\eta}^\dagger)_{cd}\\
&=\frac{1}{N_r^2}\sum_{r_1=1}^{N_r}\eta_{ar_1}\eta_{br_1}^*\sum_{r_2=1}^{N_r}\eta_{cr_2}\eta_{dr_2}^*,\\
&=\frac{N_r-1}{N_r^2}\sum_{r_1=1}^{N_r}\eta_{ar_1}\eta_{br_1}^*\left(\delta_{cd}+\mathcal{O}\left(1/\sqrt{N_r}\right)\right)\\
&\quad+\frac{1}{N_r^2}\sum_{r_1=1}^{N_r}\eta_{ar_1}\eta_{br_1}^*\eta_{cr_1}\eta_{dr_1}^*+\mathcal{O}\left(1/\sqrt{N_r}\right),\\
&=\delta_{ab}\delta_{cd}+\frac{1}{N_r}\delta_{ad}\delta_{cb}(1-\delta_{ab}\delta_{cd})+\mathcal{O}\left(1/\sqrt{N_r}\right), \quad (21)
\end{aligned}$$

where the second term is the contact term due to a multiple use of noise vectors, which causes a noise cross contamination effect and violates the orthogonal relation.[3] In terms of the tensor networks this contamination would generate an unphysical network that links between two distant tensors. Nevertheless, from Eq. (21), we find that this residual systematic error is proportional to $1/N_r$, thus it is reduced for large N_r. Moreover, thanks to its simple functional form in terms of N_r, the systematic error in the free energy can be controlled in a straightforward way. (See Fig. 13 for $1/N_r$ scaling in actual simulation result of the common noise method in the next section.)

To reduce the noise correlation, equivalently, systematic error of the common noise method we consider a checkerboard lattice, where four different sets of the noise vectors are distributed at four sites on a plaquette, respectively (see Fig. 5). Suppose that in the initial tensor network two different tensors A and B are on even and odd sites, respectively. We decompose these two tensors with four different noise vectors distributed at four sites so that we obtain eight different third-order tensors $T_{1,2,3,4}$ and $S_{1,2,3,4}$. After contracting these tensors we obtain two different renormalized tensors $A^{(1)}$ and $B^{(2)}$ defined as

$$A^{(1)} = \mathrm{Tr}[T_1T_2S_3S_4], \qquad B^{(1)} = \mathrm{Tr}[S_1S_2T_3T_4]. \quad (22)$$

[3]We implicitly assume Z_N random noises with $N \geq 3$ in Eq. (21). In the case of Z_2 noise, there may exist additional contact terms of order $1/N_r$, since two vectors of η_r and η_r^* can not be distinguished.

As shown in Fig. 5, after the coarse-graining the same checkerboard lattice structure appears, where the two different tensors $A^{(1)}$ and $B^{(1)}$ reside on even and odd sites, respectively. Thus the iterative coarse-graining process is possible. The computational cost in the checkerboard lattice case is doubled compared to the normal lattice case, since both the tensor decompositions and contractions should be performed for even and odd sites separately. We note that although the noise correlation of the checkerboard lattice is reduced compared with the normal lattice, there still exists the contamination error since two distant noise vectors come into contact after multi coarse-graining processes. After all, the existence of a contact term is inevitable to preserve the lattice homogeneity, and we end up with taking the residual systematic error into account in the common noise methods. We will discuss a detailed study of the systematic error and its scaling property in the next section.

IV. NUMERICAL TEST IN 2D ISING MODEL

We implement the noise methods in the TRG for Ising model on square lattice and study the efficiency of the position-dependent and the common noise methods in comparison with the original TRG as well as the analytic results on finite volumes [49] and the Onsager's exact result in infinite volume limit. We employ the Z_2 noise, i.e., η_{ir} takes ± 1 randomly and calculate the free energy density with volume $V = 2^n$. In what follows, we refer to n as the number of the coarse-graining steps, N the total number of statistics, and $D_{\rm cut} = D_{\rm svd} + N_r$ the total number of the bond dimensions, where $D_{\rm svd}$ and N_r are the bond dimensions for the singular modes included exactly and the dimension of the noise vectors, respectively.

A. Position-dependent noise

We first examine the position-dependent noise method. As shown in Eq. (15), there is no systematic error even for a finite $D_{\rm cut}$ in this method, so an exact partition function Z_V on finite volume $V = 2^n$ is obtained as an ensemble average of the renormalized tensor $T^{(n)}$ at nth coarse-graining step. The free energy density f_V is calculated as

$$f_V = -\frac{T}{V}\log Z_V, \tag{23}$$

where Z_V is the mean value of the partition function given in Eq. (15). To estimate a statistical error we utilize the jackknife method, by which the statistical fluctuation of the logarithmic function is easily estimated. We note, however, that in the stochastic method a negative Z sample $[Z(T^{(n)[\ell]}) < 0]$ can be occasionally generated, and some of the jackknife samples also become negative due to a limited number of statistics. Since we have to exclude negative jackknife samples from the error estimate, the statistical error of f_V is not estimated correctly. We will mention it when we present such a result. On the other hand, to quote the central value for f_V we take the ensemble average of both positive and negative Z samples. An implication of negative Z contributions is discussed in the following analysis.

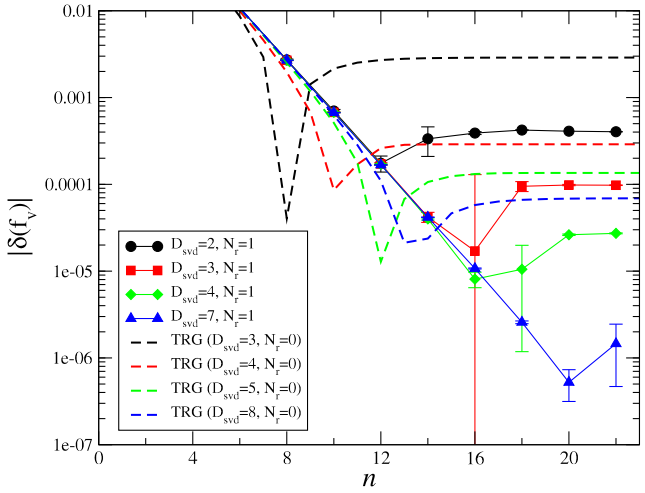

FIG. 6. The absolute value of the relative error $|\delta(f_V)|$ at $T = T_c$ as a function of the coarse-graining step. The data points represent the results for the position-dependent noise method for $N = 500$. The dashed curves show the results obtained from the TRG method with the bond dimension $D_{\rm cut} = D_{\rm svd} + N_r$, e.g., the black dashed line represents the TRG results with $D_{\rm cut} = 3$. Note that there are negative jackknife samples for $n \geq 16$ that are excluded from the jackknife error estimate.

Figure 6 shows the absolute value of relative error of the free energy $|\delta(f)|$, where $\delta(f) = (f - f^{\rm exact})/f^{\rm exact}$ is a relative error from the Onsager's exact result $f^{\rm exact}$ in the infinite volume. In the figure, we show the results for several values of $D_{\rm svd}$ with fixed $N_r = 1$ at the critical temperature $T = T_c$. We use $N = 500$ statistics for each parameter, where we exclude some negative jackknife samples from the error analysis. As shown in the figure, we find that our results with $\mathcal{O}(100)$ statistics basically give a much more accurate result than the TRG for each coarse-graining step. We also observe that the results tend to be flat as increasing the coarse-graining step n. This flattening behavior apparently contradicts what explained in the previous section, that is, the results of the position-dependent noise method should not have systematic errors so as to be consistent with the exact result. A possible reason for the behavior is a lack of statistics and the error is not correctly estimated. In fact, when computing a partition function itself using a statistical method, one needs extremely high statistics especially for large volume case since the partition function has a broad distribution whose width grows exponentially with respect to a system volume (see, e.g., [35,50]). Such a distribution of the partition function can be understood as follows. Since

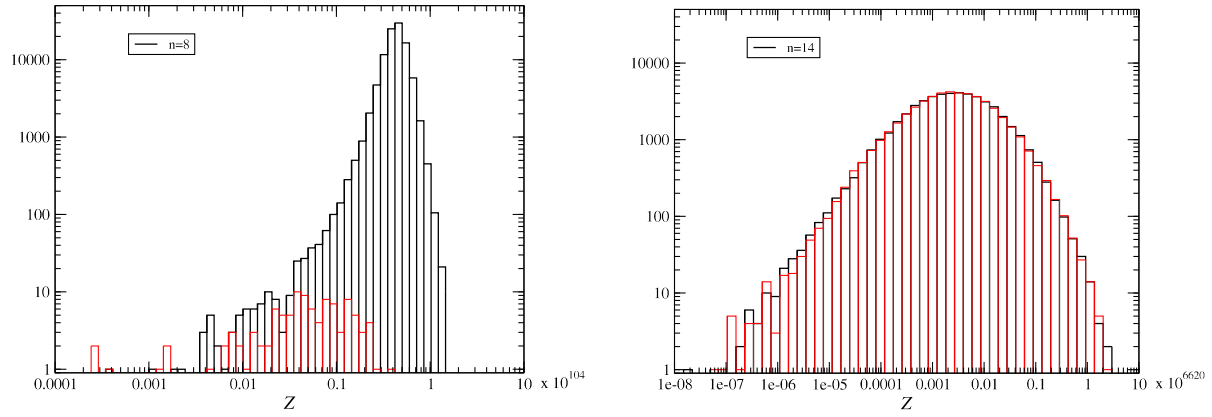

FIG. 7. Histogram of Z for $V = 2^8$ (left) and $V = 2^{14}$ (right) for 100000 samples with $D_{\mathrm{svd}} = 2$ and $N_r = 1$ at $T = T_c$, where the numbers of positive and negative Z samples are indicated as black and red bars, respectively.

we approximate the exact partition function Z_V^{ex} as tensor products

$$Z_V^{\mathrm{ex}} = \mathrm{Tr}\left[\prod_{i=1}^{2^n} \otimes T^{(0)}\right] \simeq \frac{1}{N}\sum_{\ell=1}^{N} \mathrm{Tr}\left[\prod_{i=1}^{2^{n-1}} \otimes T^{(1)[\ell]}\right] \simeq \cdots \simeq \frac{1}{N}\sum_{\ell=1}^{N} \mathrm{Tr}[T^{(n)[\ell]}], \tag{24}$$

where each $T^{(n)[\ell]}$ behaves like a random variable due to noise vectors, the partition function given as a product of the random tensors is expected to have a log-normal distribution. To see if the situation holds true, in Fig. 7, we plot the histograms of $Z(T^{(n)[\ell]})$ whose distribution is well described by a log-normal distribution, and central value is proportional to the exponential of the volume V, thus the distribution of Z is confirmed to follow the log-normal one. In addition, we also find that a probability of negative samples increase as volume (n) increases, and the total amount of the negative Z contributions will become almost the same as that of the positive Z contributions. The result at $n = 14$ actually is the case. In this case it is notoriously difficult to numerically calculate any thermodynamical quantity based on the stochastic approach. To directly see this, we show the sample size dependence of $\delta(f_V)$ and $|\delta_V(f_V)|$ in Fig. 8, where $\delta_V(f_V) = (f_V - f_V^{\mathrm{analytic}})/f_V^{\mathrm{analytic}}$ is a relative error from the analytic result on finite volume (f_V^{analytic}) [49]. In the case of $n \leq 12$, the results with sufficient statistics ($N \gtrsim 100$) are consistent with f_V^{analytic} within the statistical error and become more precise with more statistics. These trends are clearly observed on the right panel, where $|\delta_V(f_V)|$ with $n \leq 12$ is consistent with zero and its central value and statistical error decrease with increasing N. We thus numerically confirm that the position-dependent noise method provides an unbiased result for $n \leq 12$ as it should be. However, in the case of $n = 14$, the mean value f_V becomes unstable and it is difficult to obtain the $1/\sqrt{N}$ scaling of the error. In order to alleviate this numerical difficulty in large n region D_{cut} should be increased.

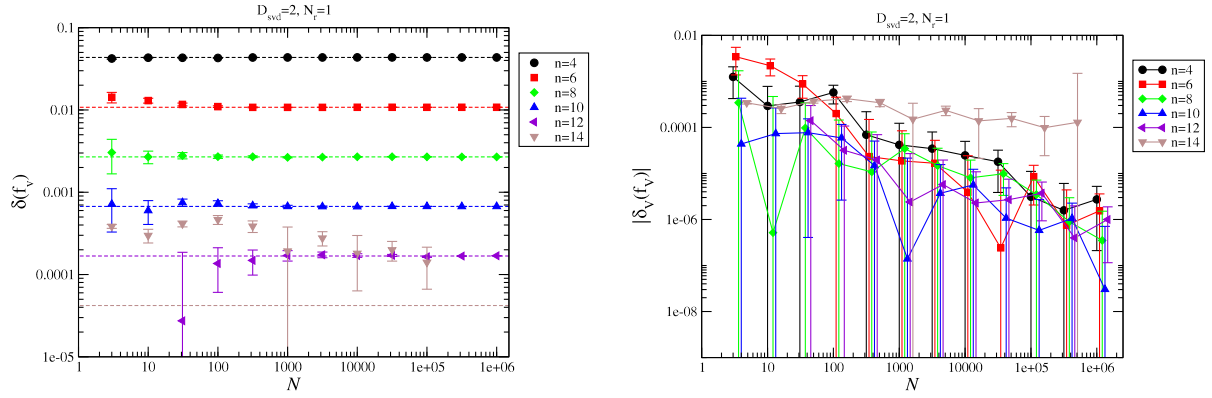

FIG. 8. Sample size dependence for $\delta(f_V)$ (left panel) and $|\delta_V(f_V)|$ (right panel) at $T = T_c$ with $D_{\mathrm{svd}} = 2, N_r = 1$ for various volumes of $V = 2^n$ with $n = 4, 6, 8, 10, 12$, and 14. The horizontal dashed lines show the analytic results on finite volumes [49]. In the right panel, the values are slightly shifted along the x axis for clarity. Note that there are negative jackknife samples in $N = 10^3$ and $N = 10^{5.5}$ statistics at $n = 14$ that are excluded from the jackknife error estimate.

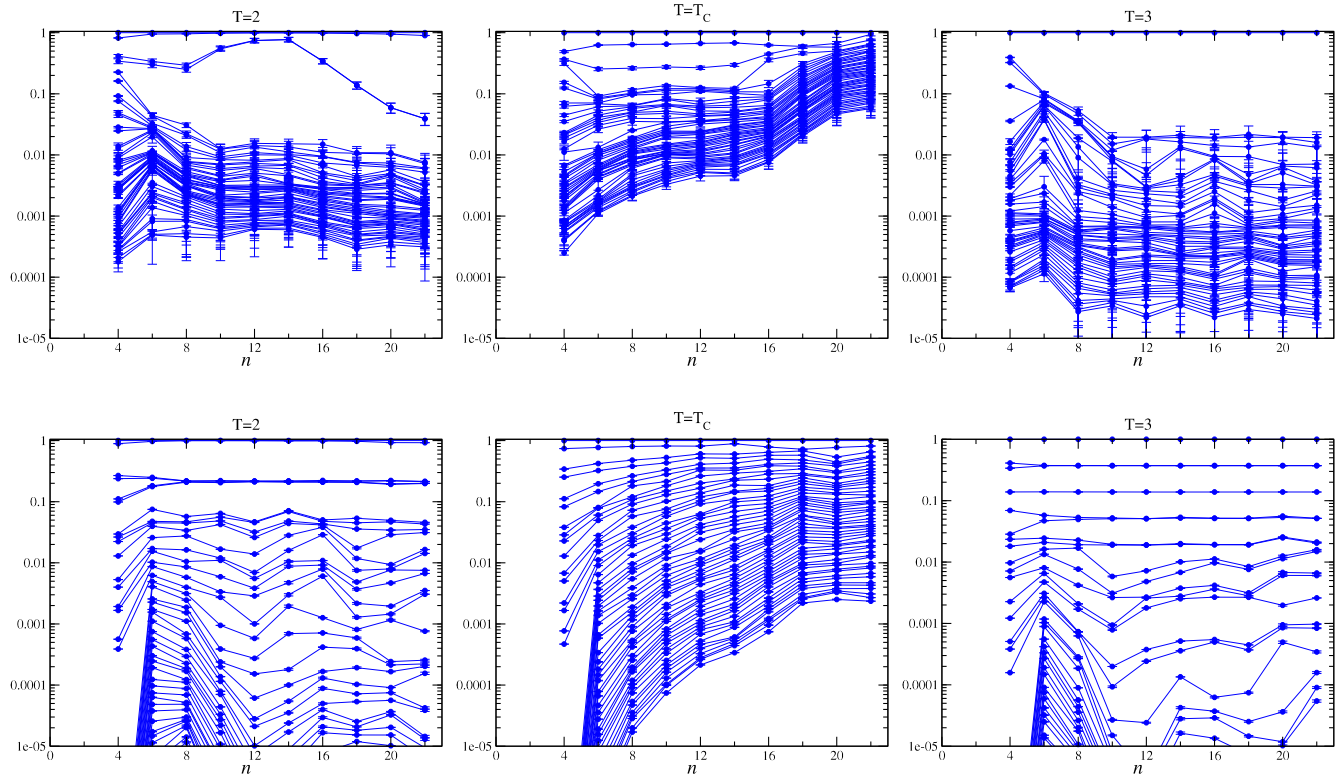

FIG. 9. The upper panels show RG flow of the singular value spectra for the ensemble averaged tensors. In the lower panel we show RG flow of the ensemble average of the singular values obtained from each sample of the renormalized tensor. The 50 largest singular values normalized by the largest one are plotted as a function of the RG steps for three different temperatures of $T = 2, T_c$ and 3 with $D_{\mathrm{svd}} = 7$, $N_r = 1$. We use 500 statistics for each.

Next we investigate the RG flow of the renormalized tensors. Since the position-dependent noise method can provide exact results even with the finite bond dimension, it is interesting to see if the renormalized tensor follows a physically correct RG flow in large volume limit. In the upper panels of Fig. 9 we show the RG flow of the singular value spectra of the ensemble averaged renormalized tensor

$$\frac{1}{N}\sum_{\ell=1}^{N} T^{(n)[\ell]}. \tag{25}$$

We use three different temperatures below and above T_c. At $T > T_c$, the renormalized tensor becomes to be dominated by only one singular mode, and the other singular values are strongly suppressed. On the other hand at $T < T_c$, we find that two degenerate singular values dominate the renormalized tensors. At $T = T_c$, all the singular values are densely distributed. These behaviors are consistent with the expectations from the real space RG transformation of the Ising model. A similar observation has already been made in other improved tensor network algorithms that can remove short distance correlations and take into account the environment effects of the tensor networks [17,19]. It should be noted here that we use exactly the same coarse-graining process as the original TRG. Nevertheless our method seems to reproduce a physically correct RG flow of the renormalized tensors despite its simple and easy algorithm. One of the reasons for this is that in our method all the singular modes are included in the renormalized tensors hence the long-range correlation are manifestly taken into account. As seen in Eq. (25), it is also important to take the ensemble average for the renormalized tensor first then one should see its singular values. To make this point clear, we investigate the RG flow of the singular values from an individual sample of renormalized tensor, and then it turns out to obey the same spectral flow as in the TRG as shown in lower panels of Fig. 9. Therefore we see that the ensemble averaging procedure reduces the statistical noise and makes the renormalized tensor to be close to the correct RG flow.

B. Common noise method

In this subsection we basically show results of the common noise method on normal lattice. We shall explicitly mention whether the normal lattice or the checkerboard lattice when necessary.

Even for the common noise method, it is natural to use Eq. (23) to calculate the free energy density. The average of the partition function Z_V, however, has the systematic errors due to the noise cross contamination effect as

discussed in Sec. III B. Furthermore, as explained in Sec. IV A, the partition function itself has an exponentially broad distribution which suffers from the long tail effect and the error may not be correctly estimated. Therefore it is not necessary to adhere f_V in Eq. (23) and one may pursue a different way to compute the free energy density. One of possibility, we propose here, is to calculate an ensemble average of the log of the partition function,

$$\bar{f}_V = -\frac{T}{V}\langle \log Z^{(n)} \rangle, \quad \langle \log Z^{(n)} \rangle = \frac{1}{N}\sum_{\ell=1}^{N} \log Z(T^{(n)[\ell]}), \tag{26}$$

where the statistical error is estimated from the variance. Needless to say, $\bar{f}_V$ also contains the systematic error and $\bar{f}_V \neq f_V$. However, it turns out that $\bar{f}_V$ has a much better control of the systematic error than f_V as we see later [see paragraph including Eqs. (27)–(32)]. Again, it should be noted here that there may exist some negative samples of $Z(T^{(n)[\ell]}) < 0$ in the common noise method as well. Since individual samples are required to calculate $\bar{f}_V$ in Eq. (26), we have to exclude negative samples from the measurement, and hence there exists an additional systematic uncertainty in the common noise method. As shown later the appearance probability of negative samples depends on both the physical system and the noise parameters. In order to clarify the presence of such an additional systematic uncertainty and study its impact we consider the following distinction when we present the results. If we do not observe any negative Z samples in the ensemble, then we indicate the result with a filled symbol. On the other hand, if there are any negative Z samples in the ensemble, we exclude these samples from the analysis and the corresponding result is indicated by an empty symbol. In this case, the mean value has the additional systematic uncertainty.

First we study the volume dependence of the relative deviation for $\bar{f}_V$ at the critical temperature $T = T_c$ for several parameters of noise vectors as shown in Fig. 10, where the total number of bond dimensions $D_{\text{cut}} = D_{\text{svd}} + N_r$ is fixed to 50. We use 50 statistics for each parameter. In comparison, the results obtained from the original TRG ($D_{\text{svd}} = 50$, $N_r = 0$) are also shown. We observe a plateau (a fixed-point tensor) at $n \gtrsim 28$ for all the parameters as in the case of the original TRG. On the contrary to the position-dependent noise method, this plateau indicates that there exists a systematic error that can not be reduced by increasing the statistics. As discussed in Eq. (21), this is the noise cross contamination effect due to a multiple use of the same noise vectors. The detailed analysis of this residual systematic error $\propto 1/N_r$ will be discussed near the end of this subsection. As for the negative samples, since this is caused by a large fluctuation due to the random noise, the appearance probability of

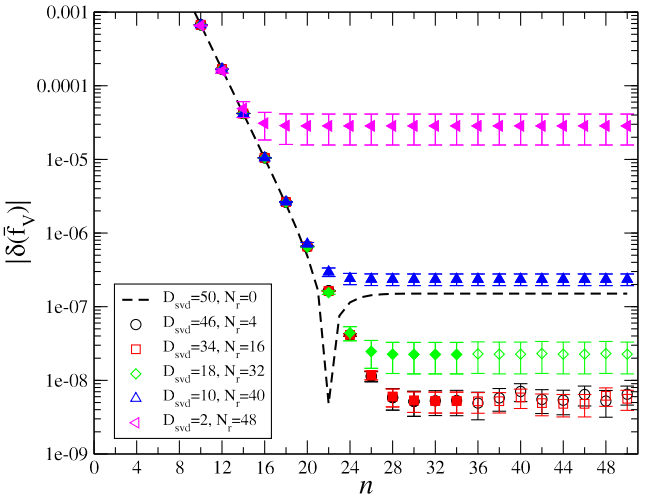

FIG. 10. The coarse-graining step dependence of relative deviation between the common noise method and the exact result for the free energy density $|\delta(\bar{f})| = |(\bar{f} - f^{\text{exact}})/f^{\text{exact}}|$ at the critical temperature $T = T_c$. The noise dimension is varied while keeping the total bond dimension $D_{\text{cut}} = D_{\text{svd}} + N_r = 50$ fixed. The number of statistics is $N = 50$ for all cases. The TRG result ($D_{\text{svd}} = 50$, $N_r = 0$) is also shown for comparison. The data with filled symbols indicate that we do not observe any negative sample. The data with empty symbols indicate that there are some negative samples in the ensemble that are excluded from the analysis.

negative samples can be suppressed by increasing N_r as shown in Table I. In practice, we find that with fixed $N_r = \mathcal{O}(1)$ a negative sample [$Z(T^{(n)[\ell]}) < 0$] seems to be generated after reaching a plateau. In order to avoid such the additional systematic uncertainty D_{cut} should be increased in accordance with increasing n. We also find that in contrast to the position-dependent noise method, the number of the negative samples does not drastically change as increasing n.

As for the accuracy of the common noise method, we obtain a significantly improved accuracy for smaller values of $N_r = 4$ and 16. On the other hand, the results with larger values of N_r (with smaller D_{svd}) have worse accuracy. This is because the tensor approximation via the SVD is not good for such smaller D_{svd} as shown in Eq. (8), which should also lead to a significant increase of the noise contamination error as decreasing D_{svd}. Thus in order to minimize the error one should consider an optimization for the parameters of D_{svd} and N_r. By changing these two

TABLE I. Number of the negative samples at $D_{\text{cut}} = D_{\text{svd}} + N_r = 50$ for total $N = 50$ samples.

n	1–30	32	34	36	38	40	42	44	46	48	50
$D_{\text{svd}} = 46$, $N_r = 4$	0	1	2	2	3	6	3	3	4	6	5
$D_{\text{svd}} = 34$, $N_r = 16$	0	0	0	3	3	4	2	2	2	4	2
$D_{\text{svd}} = 18$, $N_r = 32$	0	0	0	1	1	1	2	1	1	2	1

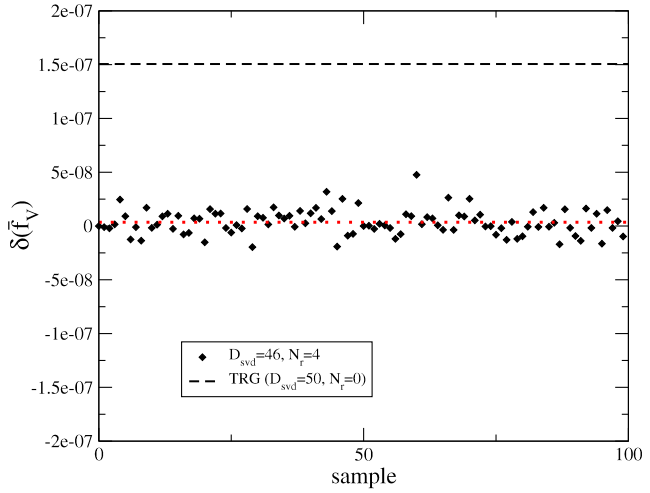

FIG. 11. Sample dependence of the relative errors for the free energy density $\delta(\bar{f}_V)$ at the step $n = 30$ and $T = T_c$. The red dotted and black dashed lines represent the mean value of $N = 100$ samples and the absolute value of TRG result, respectively.

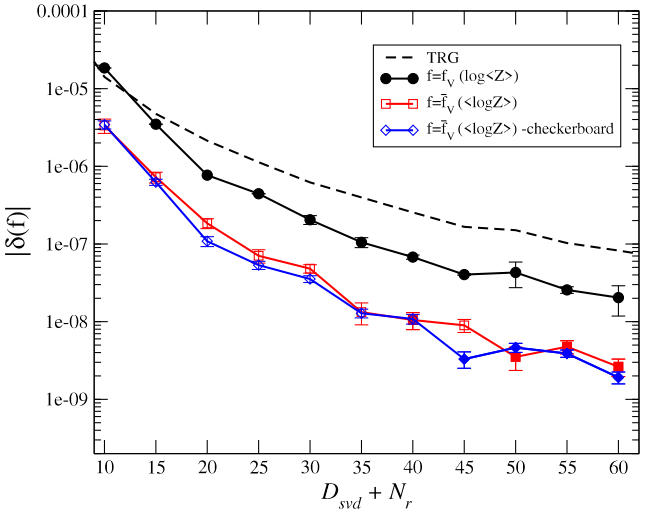

FIG. 12. $D_{\rm cut}$ dependence of the relative errors of the free energy density on $V = 2^n$ with $n = 30$ and $N_r = 4$ at $T = T_c$ for the common noise method with normal lattice (black circle and red square) and checkerboard lattice (blue diamond). We use $N = 100$ samples for each parameter. For comparison the TRG results with the bond dimension $D_{\rm cut} = D_{\rm svd} + N_r$ ($N_r = 0$) are also plotted.

parameters, we could minimize the noise contamination error. We find that typically a few number of $N_r \ll D_{\rm svd}$ is sufficient to obtain better accuracy than the original TRG.

Figure 11 shows a scatter plot for 100 samples of the results with $D_{\rm svd} = 46$ and $N_r = 4$. We find that all the samples have much better accuracy than the original TRG and the fluctuation is well controlled. In fact the plot indicates that the systematic error of TRG with $D = 50$, of $\mathcal{O}(\Lambda_{51})$, is larger than the statistical fluctuation of our method with $D_{\rm svd} = 46$ and $N_r = 4$, of $\mathcal{O}(\Lambda_{47}/\sqrt{4})$, which is estimated from Eq. (10). Since our noise method utilizes all SVD modes in the tensor decomposition and the noises are spatially correlated, such a small fluctuation is not surprising, as long as the additional systematic error due to the noise correlation is under control. The mean value is stable against the changes of the statistics compared to the position-dependent noise method, that is, only a few statistics is sufficient to obtain a reliable estimate.

Figure 12 shows the $D_{\rm cut}$ dependence of the relative error of the free energy at $T = T_c$ with fixed $N_r = 4$ at $n = 30$. Similar to the original TRG, we observe a monotonically decreasing of $|\delta(\bar{f})|$ as increasing the bond dimension $D_{\rm cut}$. We compare two methods of normal and checkerboard lattices in the figure. We observe a slightly better accuracy for the checkerboard lattice as expected. We also compare two evaluations of the free energy, f_V in Eq. (23) and $\bar{f}_V$ in Eq. (26). As shown in the figure, a significant improvement in $\bar{f}_V$ rather than f_V is observed, which means that the systematic error of $\bar{f}_V$ is much smaller than that of f_V.[4]

[4]Note that the statistical error for f_V may not be correctly estimated for such small statistics due to the exponentially broad distribution.

We shall briefly discuss a possible reason for the error reduction mechanism in $\bar{f}_V$. Let us consider a spin-0 toy model instead of the original Ising model, namely a renormalized tensor A has a bond dimension $D_{\rm cut} = 1$. For simplicity we assume that A consists of two terms as

$$A = A_{\rm ex}(1 + \delta_A), \tag{27}$$

where the first term $A_{\rm ex}$ is an exact part, and the second term δ_A represents a statistical fluctuation due to random noises with $\langle \delta_A \rangle = 0$ where $\langle \cdots \rangle$ denotes the ensemble average as given in (26). Assuming $|\delta_A| \ll 1$, the partition function $Z^{(n)}$ is then given as

$$Z^{(n)} \equiv A^V = A_{\rm ex}^V \left(1 + V\delta_A + \frac{V(V-1)}{2}(\delta_A)^2\right) + \mathcal{O}((\delta_A)^3), \tag{28}$$

and its mean value is

$$\langle Z^{(n)} \rangle = A_{\rm ex}^V \left(1 + \frac{V(V-1)}{2}\langle \delta_A^2 \rangle\right) + \mathcal{O}((\delta_A)^3). \tag{29}$$

In the case of f_V it is given as the log of the mean value

$$f_V = -\frac{T}{V}\log\left(\langle Z^{(n)} \rangle\right) \simeq f_V^{\rm ex} - T\frac{(V-1)}{2}\langle \delta_A^2 \rangle, \tag{30}$$

where the exact free energy is given by $f_V^{\rm ex} = -T \log A_{\rm ex}$. On the other hand, in the case of $\bar{f}_V$ it is given as

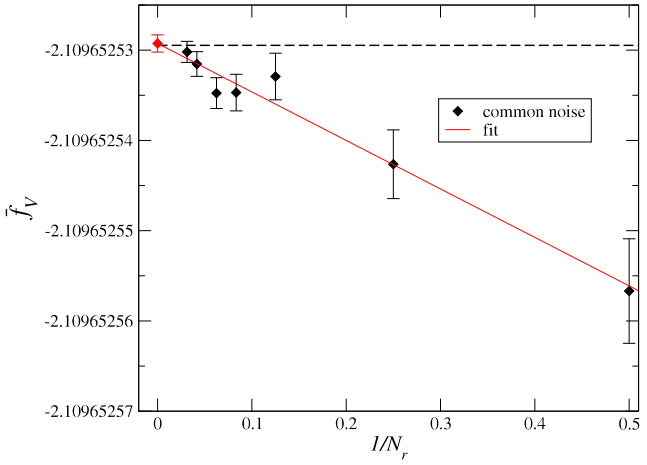

FIG. 13. N_r dependence of the free energy density $\bar{f}_V$ with fixed $D_{\mathrm{svd}} = 20$ at the coarse-graining step $n = 20$, $T = T_c$, and $N = 5000$ statistics for each parameters. The dashed line represents the analytic result on $V = 2^{20}$.

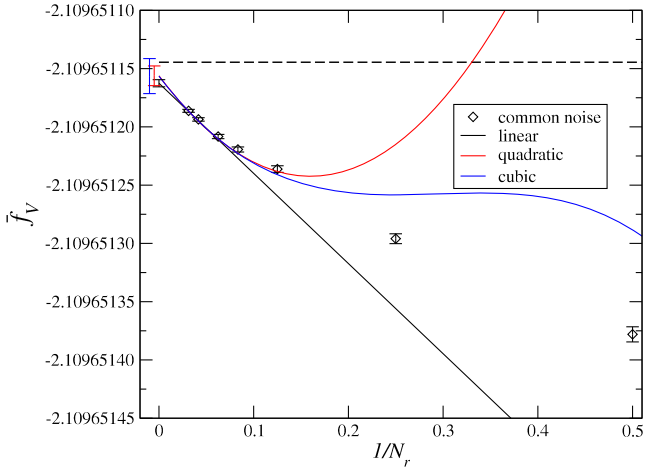

FIG. 14. N_r dependence of the free energy density $\bar{f}_V$ with fixed $D_{\mathrm{svd}} = 20$ at the coarse-graining step $n = 40$, $T = T_c$, and $N = 5000$ statistics for each parameters. The dashed line represents the analytic result on $V = 2^{40}$.

$$\bar{f}_V = -\frac{T}{V}\langle \log Z^{(n)} \rangle \simeq f_V^{\mathrm{ex}} + T\frac{\langle \delta_A^2 \rangle}{2}. \tag{31}$$

Equations (30) and (31) tell us that f_V and $\bar{f}_V$ are not simply equal to the exact one due to the noise cross contamination effect at leading order $\langle \delta_A^2 \rangle$, however, according to the argument in Eq. (21) the systematic error scales with $\langle \delta_A^2 \rangle \propto 1/N_r$, therefore such a systematic error can be removed for sufficiently large N_r,

$$\lim_{N_r\to\infty} f_V = \lim_{N_r\to\infty} \bar{f}_V = f_V^{\mathrm{ex}}. \tag{32}$$

For finite N_r, there is the systematic error $\langle \delta_A^2 \rangle$ for both methods, but the coefficient for $\bar{f}_V$ is much smaller than that of f_V, which is proportional to volume[5] [see Eq. (30)]. In fact, the error reduction in $\bar{f}_V$ is observed in all the parameter regions for both the normal and checkerboard lattices. As for the additional systematic uncertainty in $\bar{f}_V$ due to negative samples, it can be avoided by taking sufficiently large N_r, so Eq. (32) should hold.

Next we study the scaling property of the systematic error. As shown in the previous section, the residual systematic error in the common noise method should be given by the noise cross contamination effect in Eq. (21), and it is important to study the N_r dependence of the results. Figure 13 shows the free energy density at the coarse-graining step $n = 20$ as a function of $1/N_r$ with fixed $D_{\mathrm{svd}} = 20$. N_r is varied in the range 2–32. As shown in the figure, since we observe a clear linear dependence of $1/N_r$, we can carry out a fit analysis to obtain a result in the $N_r \to \infty$ limit. We consider a linear fit function of $\bar{f}_V = d_0 + d_1/N_r$ and use all the data for a fit. The fit result is given as $d_0 = -2.1096525293(10)$ with $\chi^2/\mathrm{d.o.f.} = 0.7$, which is consistent with the analytic value at $V = 2^{20}$. For larger volumes we show the results at the step $n = 40$ with fixed $D_{\mathrm{svd}} = 20$ as a function of $1/N_r$ in Fig. 14. In contract to the results with $n = 20$, we see a curvature so that higher order corrections of $1/N_r$ are required to fit these data. We consider higher order polynomial fit functions

$$\bar{f}_V = d_0 + d_1\frac{1}{N_r} + d_2\frac{1}{N_r^2} + d_3\frac{1}{N_r^3}. \tag{33}$$

The fit results are also shown in the figure and Table II. Our results are well fitted by polynomial functions in $1/N_r$. We obtain $\bar{f}_{2^{40}} = -2.1096511637(28)$ for cubic functions in $N_r = \infty$ limit, which is consistent with the analytic result at $V = 2^{40}$. Thus we obtain much better results than the original TRG, and the noise cross contamination effect is found to be well controlled by the $1/N_r$ corrections.[6]

We note here that since this contamination is the only systematic error, the $1/N_r$ scaling is a universal feature of the common noise method, which should not depend on the details of the systems and the coarse-graining processes.

[5]In the case of the position-dependent noise method, the noise cross contamination effect $\langle \delta_A^2 \rangle$ should be replaced with $\langle \delta_{A^i} \rangle \langle \delta_{A^j} \rangle$ $(i \neq j)$, where δ_{A^i} is a noise fluctuation on site i. A broad distribution with volume dependence discussed in the previous section is also understood from this analysis.

[6]We note that in the case of $n = 40$ we observe negative samples in the entire region of N_r, and hence the results have the additional systematic uncertainty. Thus there is a possibility that the curvature in $n = 40$ is caused by the systematic effect due to the negative Z samples. To correctly study a $1/N_r$ scaling we would need more data for a much larger N_r region where there is no negative Z sample.

TABLE II. Fit results for $\bar{f}_V$ at $n = 40$ using a polynomial fit form in Eq. (33). The value of χ^2/d.o.f. and the fit ranges are also tabulated.

	d_0	$d_1 \times 10^{-7}$	$d_2 \times 10^{-6}$	$d_3 \times 10^{-6}$	χ^2/d.o.f.	Fit range
Linear	−2.1096511625(31)	−7.74(72)	⋯	⋯	6×10^{-3}	$N_r \geq 16$
Quadratic	−2.1096511562(85)	−1.08(34)	3.4(3.1)	⋯	0.3	$N_r \geq 12$
Cubic	−2.109651156(15)	−1.08(7.6)	4(11)	−4(50)	0.3	$N_r \geq 8$

Therefore this scaling property is quite different from the original TRG. As for the MCTN, since the common noise method has not been examined in [35,36], it is not obvious if one could see a clear scaling property of the systematic error. It may be interesting to study the scaling property in the common noise method for the MCTN as well.

Finally we shall discuss the sample size dependence of the performance in comparison with the TRG. In order to contextualize the numerical cost of our method, which depends on the number of statistics N, we estimate a computational time τ as [28]

$$\tau = \begin{cases} D_{\mathrm{cut}}^6 & \text{for TRG} \\ N D_{\mathrm{cut}}^6 & \text{for common noise method} \end{cases}. \tag{34}$$

The left panel of Fig. 15 shows the relative error as a function of D_{cut} for various N. The mean values for different N are consistent with each other. A substantial reduction of the systematic error is found to be statistically significant even for a small N. In the right panel of Fig. 15, we show the same results as a function of τ. For smaller values of $\tau < 10^{10}$ the numerical cost becomes comparable with the TRG with increasing N. On the other hand, for larger τ the common noise method has better performance even for large values of N. We note that the actual elapsed time can be even more reduced by utilizing parallel computing, since individual samples can be independently generated.

We note that other thermodynamical quantities are also calculated based on the stochastic approach. As an example, we show the results for the specific heat that can be obtained by numerical differences of $\bar{f}_V$ in Appendix. It should be also noted that the common noise method can be straightforwardly implemented in other tensor networks without any substantial increase in computational cost and any iterative process. Thus it is interesting to apply this method to other models. In the next section, we test this method with an advanced algorithm to explore a possibility of further improvement.

V. APPLICATION TO OTHER COARSE-GRAINING ALGORITHMS

As shown in the previous section, our methods provide a very simple and useful way to reduce and control the systematic errors. This is a distinguished feature of our methods in sharp contrast to other improved tensor network algorithms that are designed to exhibit a correct RG flow. In addition, our methods can be easily applied to other tensor networks as long as the truncated SVD is used in a matrix decomposition. In this regard let us demonstrate how our methods can work in other tensor network algorithms. Here as an example, we consider Gilt-TNR [19]. It is known that the Gilt procedure also provides a simple algorithm to systematically truncate the bond dimensions of tensor networks. Furthermore Gilt-TNR, which is a combination

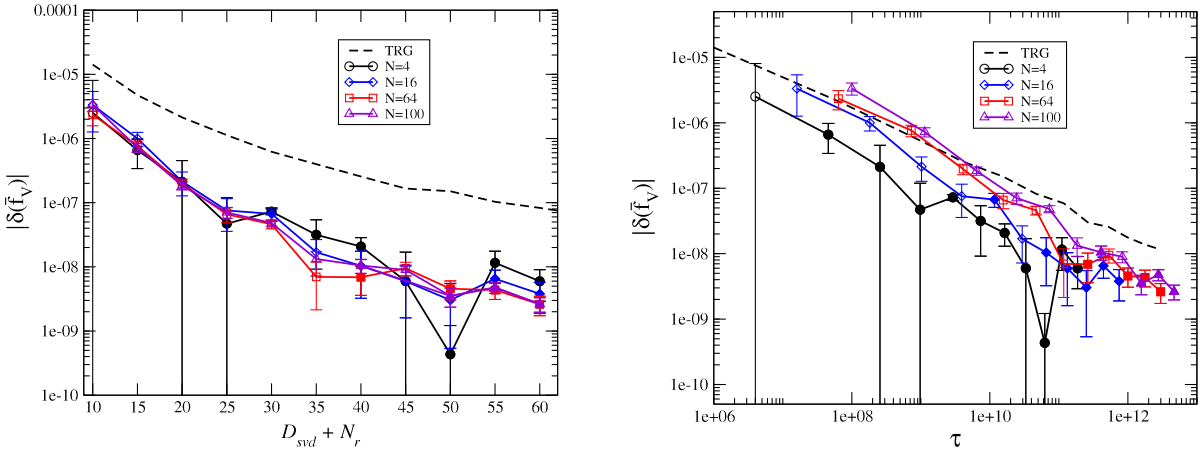

FIG. 15. D_{cut} (left panel) and τ (right panel) dependence of the relative errors $|\delta(\bar{f}_V)|$ for various statistics on $V = 2^{30}$ with fixed $N_r = 4$ at $T = T_c$.

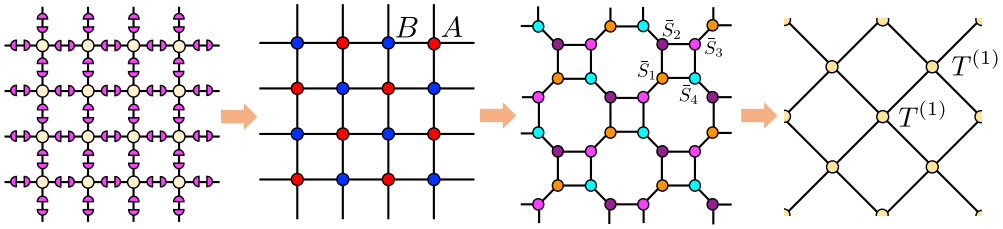

FIG. 16. Coarse graining of Gilt-TNR in combining with the common noise method. The original tensor T resides on every site, and four rank-reduced matrices obtained via Gilt are inserted into four links on a plaquette. These matrices are decomposed by full SVDs and absorbed into neighboring tensors T, by which two rank-reduced tensors of A and B are obtained. These A and B are decomposed by the common noise methods. In the initial lattice four different noise vectors $\boldsymbol{\eta}_{1,2,3,4}$ are distributed to the four sites. The tensor contraction for one plaquette is carried out, and a renormalized tensor $T^{(1)}$ is obtained.

of Gilt and TRG, also exhibits a physically correct RG flow of the renormalized tensors with better precisions. From the practical point of view, it is interesting to see whether there still is room for numerical improvement by combining with our noise methods.

We explain an implementation of the common noise method to Gilt-TNR as follows. In the case of 2D Ising model on square lattice, a tensor network on a plaquette is approximated by inserting four matrices $R_{1,2,3,4}$ to four links of the plaquette via Gilt (see Fig. 16). Then the original tensor T on each site is modified to A or B by absorbing the neighboring matrices $R_{1,2,3,4}$ via SVDs, by which a part of short distance correlations is removed and the bond dimensions D_{cut} can be reduced to D'_{cut}, which depends on the threshold parameter ϵ in Gilt as well as the dynamics. Here we simply apply the common noise method when decomposing two matrices A and B. Since the tensor network consisting of A and B has a checkerboard structure, we can employ the common noise method

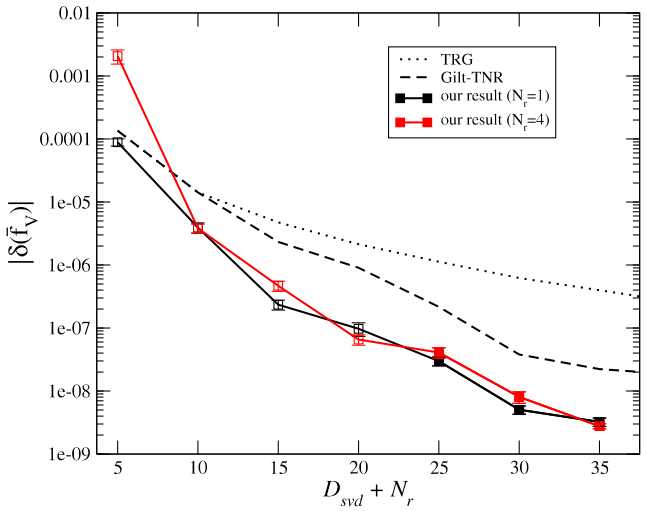

FIG. 17. Relative deviation of the free energy at $T = T_c$ with fixed values of $N_r = 1$ (black), $N_r = 4$ (red), and the threshold parameter $\epsilon = 8 \times 10^{-7}$. The number of statistics is 100 for each result. In comparison, the results for TRG (dotted) and Gilt-TNR (dashed) are also shown with same parameters.

on checkerboard lattice as explained in the previous section. Moreover, after coarse graining the tensor network still retains a homogeneity on a checkerboard lattice as seen in Fig. 16. For the numerical test we use the original source code for 2D Ising model on square lattice [51] with small modifications to suit our purposes.

We present a benchmark result in Fig. 17 which shows the relative errors $|\delta(\bar{f}_V)|$ as a function of D_{cut} at $T = T_c$ with $N = 100$ on $V = 2^{51}$ lattice. For a comparison we also show the results for Gilt-TNR as well as the original TRG. We commonly choose the Gilt threshold parameter $\epsilon = 8 \times 10^{-7}$ as given in [19]. As shown in the figure, our noise methods systematically improve the accuracy. Our results with even smaller bond dimensions can achieve a relative error of $\mathcal{O}(10^{-9})$, that is a typical precision limit of Gilt-TNR for this range of parameters. We note that the computational cost per sample is the same order of Gilt-TNR, while the reduced number of the bond dimension via Gilt depends not only on the threshold parameter ϵ but also on samples due to the random noises. Thus our method is shown to be effective even for an improved algorithm without any changes of the original tensor networks.

VI. CONCLUSION

Following the idea of the MCTN, we have proposed a new stochastic method by utilizing random noise vectors combining with the singular value decomposition, where the rank-reduced tensor manifestly contains all the singular modes thanks to the random noise vectors. In the method we generate tensor ensembles, and the partition functions and any related physical quantities are statistically calculated. We have tested two types of the noise distribution for 2D Ising model.

In the case of the position-dependent noise method since there is no systematic error, our result is exact in the sense that no matter how accurate result one could obtain by increasing the statistics while keeping the bond dimension finite. We note that the computational cost is as expensive as $\mathcal{O}(ND_{\mathrm{cut}}^6 V)$, which should be comparable to the cost of TRG-related methods for a system without translation invariance. We also confirm that the RG flow of the

singular value spectra are consistent with the expectation from the real space RG transformation despite its simple and easy algorithm. On the other hand, in the case of the common noise method, the computational cost scales as $\mathcal{O}(ND_{\mathrm{cut}}^6 \log V)$ and we obtain a better accuracy than the original TRG with a limited number of statistics. While there exists a residual systematic error due to a multiple use of the noise vectors, this error is found to be under control by a model independent $1/N_r$ scaling, where N_r is the number of the noise dimensions. Thus our stochastic method actually improves the error evaluation method as well as the numerical accuracy. This model independent property is in sharp contrast to other tensor network algorithms using the truncated SVD. It should also be emphasized that our method is very simple and does not require any iterative process, so it is easily implemented even to other improved tensor networks. As a nontrivial example, we have applied our method to the Gilt-TNR, where an even more error reduction has been obtained. An interesting future direction will be an application to more complicated system with higher dimensionality.

ACKNOWLEDGMENTS

H. O. is supported in part by JSPS KAKENHI Grants No. 21K03554 and No. 22H00138. S. T. is supported in part by JSPS KAKENHI Grants No. 20H00148, No. 21K03531, and No. 22H01222. M. T. is supported by U.S. DOE Grants No. DE-SC0010339 and No. DE-SC0021147. This work was supported by MEXT KAKENHI Grant-in-Aid for Transformative Research Areas A "Extreme Universe" No. 22H05251.

APPENDIX: SPECIFIC HEAT

As an example of thermodynamical quantities, we study the specific heat based on the stochastic approach for the common noise method. Figure 18 shows THE temperature dependence of the free energy $\bar{f}_V$ and the absolute value of the relative error $|\delta(\bar{f}_V)|$ around the critical temperature $T = T_c$. From this results, we calculate the specific heat C_V from numerical differentials of $\bar{f}_V$ with a finite interval ΔT of temperature T.

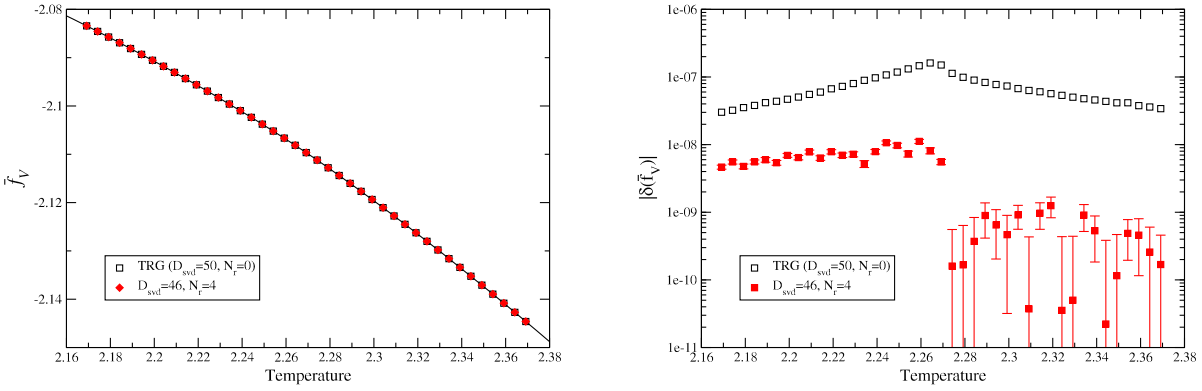

FIG. 18. Temperature dependence of $\bar{f}_V$ (left panel) and $|\delta(\bar{f}_V)|$ (right panel) on $V = 2^{30}$ with $N = 100$ statistics is shown in comparison with the results for the TRG (square). The solid curve in the left panel represents the Onsager's exact results.

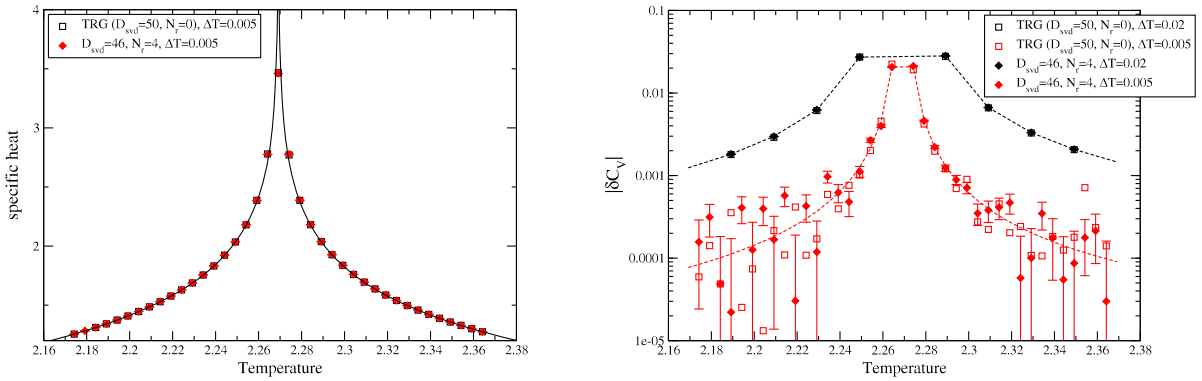

FIG. 19. Temperature dependence of C_V (left panel) and $|\delta C_V|$ (right panel) on $V = 2^{30}$ with $N = 100$ statistics is shown in comparison with the results for the TRG (square). The solid curve in the left panel represents the Onsager's exact results. The dashed curves in the right panel represent the analytic results on the finite volume obtained from numerical differences of the free energy f_V^{analytic} with a finite interval ΔT.

In Fig. 19 we show T dependence of C_V and the absolute value of the relative error $|\delta C_V|$ around the critical temperature $T = T_c$, where $\delta C_V = (C_V - C^{\text{exact}})/C^{\text{exact}}$ is a relative error from the Onsager's exact result for the specific heat C^{exact}. We show the results for $|\delta C_V|$ obtained with two values of ΔT, 0.02 and 0.005, which indicate the significance of the finite-ΔT effect on the results from both TRG and the stochastic methods. Decreasing ΔT can in principle reduce this effect but in stead increases the significance of other errors, the roundoff effects on both results and the statistical error on the result from the stochastic method, as seen for $\Delta T = 0.005$. The enhancement of the statistical error on numerical differential is known. In fact, it has been observed in [27] that when using a stochastic method, the precision of the free energy is better while the specific heat obtained by the numerical difference is comparable with that of the TRG. The increase of $|\delta C_V|$ near T_c should be because of the finite-volume effect. Since these dominant finite-volume and finite-ΔT effects are the same for both TRG and the stochastic method, the accuracy of these methods are comparable.

[1] M. Levin and C. P. Nave, Phys. Rev. Lett. **99**, 120601 (2007).
[2] Y. Shimizu and Y. Kuramashi, Phys. Rev. D **90**, 014508 (2014).
[3] S. Hong and D.-H. Kim, J. Phys. Soc. Jpn. **91**, 084003 (2022).
[4] A. Bazavov, S. Catterall, R. G. Jha, and J. Unmuth-Yockey, Phys. Rev. D **99**, 114507 (2019).
[5] D. Kadoh, Y. Kuramashi, Y. Nakamura, R. Sakai, S. Takeda, and Y. Yoshimura, J. High Energy Phys. 03 (2018) 141.
[6] D. Kadoh, Y. Kuramashi, Y. Nakamura, R. Sakai, S. Takeda, and Y. Yoshimura, J. High Energy Phys. 05 (2019) 184.
[7] Y. Yoshimura, Y. Kuramashi, Y. Nakamura, S. Takeda, and R. Sakai, Phys. Rev. D **97**, 054511 (2018).
[8] H. Kawauchi and S. Takeda, Phys. Rev. D **93**, 114503 (2016).
[9] Y. Kuramashi and Y. Yoshimura, J. High Energy Phys. 04 (2020) 089.
[10] S. S. Jahromi, R. Orús, M. Kargarian, and A. Langari, Phys. Rev. B **97**, 115161 (2018).
[11] R. Sakai, S. Takeda, and Y. Yoshimura, Prog. Theor. Exp. Phys. **2017**, 063B07 (2017).
[12] Y. Shimizu and Y. Kuramashi, Phys. Rev. D **90**, 074503 (2014).
[13] Y. Shimizu and Y. Kuramashi, Phys. Rev. D **97**, 034502 (2018).
[14] S. Takeda and Y. Yoshimura, Prog. Theor. Exp. Phys. **2015**, 043B01 (2015).
[15] M. C. Bañuls *et al.*, Eur. Phys. J. D **74**, 165 (2020).
[16] D. Kadoh, Proc. Sci. LATTICE2021 (**2022**) 633.
[17] G. Evenbly and G. Vidal, Phys. Rev. Lett. **115**, 180405 (2015).
[18] S. Yang, Z.-C. Gu, and X.-G. Wen, Phys. Rev. Lett. **118**, 110504 (2017).
[19] M. Hauru, C. Delcamp, and S. Mizera, Phys. Rev. B **97**, 045111 (2018).
[20] Z. Y. Xie, H. C. Jiang, Q. N. Chen, Z. Y. Weng, and T. Xiang, Phys. Rev. Lett. **103**, 160601 (2009).
[21] Z.-Y. Xie, J. Chen, M.-P. Qin, J. W. Zhu, L.-P. Yang, and T. Xiang, Phys. Rev. B **86**, 045139 (2012).
[22] H.-H. Zhao, Z.-Y. Xie, T. Xiang, and M. Imada, Phys. Rev. B **93**, 125115 (2016).
[23] S. Wang, Z.-Y. Xie, J. Chen, B. Normand, and T. Xiang, Chin. Phys. Lett. **31**, 070503 (2014).
[24] K. Harada, Phys. Rev. B **97**, 045124 (2018).
[25] S. Morita, R. Igarashi, H.-H. Zhao, and N. Kawashima, Phys. Rev. E **97**, 033310 (2018).
[26] G. Evenbly, Phys. Rev. B **95**, 045117 (2017).
[27] Y. Nakamura, H. Oba, and S. Takeda, Phys. Rev. B **99**, 155101 (2019).
[28] D. Adachi, T. Okubo, and S. Todo, Phys. Rev. B **102**, 054432 (2020).
[29] D. Kadoh and K. Nakayama, arXiv:1912.02414.
[30] W. Lan and G. Evenbly, Phys. Rev. B **100**, 235118 (2019).
[31] H. Oba, Prog. Theor. Exp. Phys. **2020**, 013B02 (2020).
[32] S. Morita and N. Kawashima, Phys. Rev. B **103**, 045131 (2021).
[33] D. Adachi, T. Okubo, and S. Todo, Phys. Rev. B **105**, L060402 (2020).
[34] M. C. Bañuls, K. Cichy, Y.-J. Kao, C. J. D. Lin, Y.-P. Lin, and D. T. L. Tan, Phys. Rev. D **100**, 094504 (2019).
[35] A. J. Ferris, arXiv:1507.00767.
[36] W. Huggins, C. D. Freeman, M. Stoudenmire, N. M. Tubman, and K. B. Whaley, arXiv:1710.03757.
[37] O. Sikora, H.-W. Chang, C.-P. Chou, F. Pollmann, and Y.-J. Kao, Phys. Rev. B **91**, 165113 (2015).
[38] A. W. Sandvik and G. Vidal, Phys. Rev. Lett. **99**, 220602 (2007).
[39] L. Wang, I. Pižorn, and F. Verstraete, Phys. Rev. B **83**, 134421 (2011).
[40] A. J. Ferris and G. Vidal, Phys. Rev. B **85**, 165146 (2012).
[41] A. J. Ferris and G. Vidal, Phys. Rev. B **85**, 165147 (2012).
[42] S. Iblisdir, New J. Phys. **16**, 103022 (2014).
[43] L. R. Schwarz, A. Alavi, and G. H. Booth, Phys. Rev. Lett. **118**, 176403 (2017).
[44] H.-H. Zhao, K. Ido, S. Morita, and M. Imada, Phys. Rev. B **96**, 085103 (2017).
[45] T. Vieijra, J. Haegeman, F. Verstraete, and L. Vanderstraeten, Phys. Rev. B **104**, 235141 (2021).
[46] M. Qin, Phys. Rev. B **102**, 125143 (2020).
[47] H. Ohki, M. Tomii, and E. Arai, Proc. Sci. LATTICE2021 (**2022**) 051.

[48] J. Foley, K. Jimmy Juge, A. O'Cais, M. Peardon, S. M. Ryan, and J.-I. Skullerud, Comput. Phys. Commun. **172**, 145 (2005).
[49] B. Kaufman, Phys. Rev. **76**, 1232 (1949).
[50] J. C. Plefka and S. Samuel, Phys. Rev. D **56**, 44 (1997).
[51] M. Hauru, C. Delcamp, and S. Mizera, Phys. Rev. B **97**, 045111 (2018).