Constrained sampling method for analytic continuation

Anders W. Sandvik
Department of Physics, Boston University, 590 Commonwealth Avenue, Boston, Massachusetts 02215, USA
and Institute of Physics, Chinese Academy of Sciences, P.O. Box 603, Beijing 100190, China
(Received 20 February 2015; revised manuscript received 19 October 2016; published 27 December 2016)

A method for analytic continuation of imaginary-time correlation functions (here obtained in quantum Monte Carlo simulations) to real-frequency spectral functions is proposed. Stochastically sampling a spectrum parametrized by a large number of δ functions, treated as a statistical-mechanics problem, it avoids distortions caused by (as demonstrated here) configurational entropy in previous sampling methods. The key development is the suppression of entropy by constraining the spectral weight to within identifiable optimal bounds and imposing a set number of peaks. As a test case, the dynamic structure factor of the $S = 1/2$ Heisenberg chain is computed. Very good agreement is found with Bethe ansatz results in the ground state (including a sharp edge) and with exact diagonalization of small systems at elevated temperatures.

DOI: 10.1103/PhysRevE.94.063308

I. INTRODUCTION

Obtaining real-frequency dynamic response functions from imaginary-time correlations remains one of the outstanding challenges for quantum Monte Carlo (QMC) and related simulation methods (e.g., lattice QCD). The general form of the problem is to invert the relationship

$$G(\tau) = \int d\omega\, A(w) K(\tau,\omega), \tag{1}$$

where a QMC estimate $\tilde{G}(\tau)$ of the correlation function $G(\tau)$ is available, $A(\omega)$ is the spectral function sought, and the kernel $K(\tau,\omega)$ depends on the type of spectral function. Similar to an inverse Laplace transform, there is no general closed form for $A(\omega)$. Only broad features of $A(\omega)$ can be resolved in numerical analytic continuation, because information on fine structure is only present at a level of precision of $G(\tau)$ that is not attainable in practice. Nevertheless, one can extract important dynamical features and the key question is how to do that with the maximum fidelity, given $\tilde{G}(\tau)$ and its statistical errors. Significant progress will be presented here.

The maximum entropy (ME) method [1] was adapted to the particulars of QMC some time ago [2]. Overcoming problems of previous approaches [3,4], it quickly became a standard tool [5]. The ME method has an appealing (though often over-stated) footing in probability theory, but in many cases the entropic prior regularizes the spectrum too heavily, leading to excessive broadening and distortions. To avoid this, an alternative line of methods has been developed [6–10] (and applied to diverse systems [11–14]) that does not impose the entropic prior, instead using stochastic sampling of $A(\omega)$ with the likelihood function

$$P(A) \propto \exp(-\chi^2/2\Theta), \tag{2}$$

where χ^2 is the standard measure of the goodness of the fit of $G(\tau)$ obtained from $A(\omega)$ according to Eq. (1) to the QMC-computed $\tilde{G}(\tau)$ with its full covariance matrix [5,8] for a set of imaginary-time points $\{\tau_i\}$. The spectrum is typically parametrized as a sum of a large number of δ functions, though other forms have also been proposed [10]. The sampling temperature Θ in Eq. (2) acts as a regularization parameter, and the main remaining issue has been how to choose its value properly.

An important insight was gained by Beach [7], showing that a mean-field treatment of the sampling approach gives the ME method, with Θ corresponding to the entropic weight. Subsequently, Syljuåsen argued for fixing $\Theta = 1$ [8] (as had also been done by White in earlier work [15]). A recent variant of the method by Fuchs *et al.* uses Bayesian inference to determine Θ [9].

Here a previously overlooked problem with the sampling approach is pointed out and a solution is offered that improves the performance of the method to the point that a sharp edge of the spectrum can be resolved without having to resort to imposing such a feature by hand (e.g., by fitting to a functional form containing a sharp edge [16]). The key insight is that, when parametrizing $A(\omega)$ with N δ functions and treating these as the configuration space of a statistical mechanics problem with χ^2 corresponding to the energy, the configurational entropy at a fixed value of Θ (not to be confused with the information entropy of the ME methods, though the two are ultimately related) increases when N is increasing, thereby forcing $A(\omega)$ away from a good fit. This happens primarily because χ^2 does not have the normal extensive property of an energy function. Spectral weight is therefore forced out by entropic pressure beyond the bounds of the true spectrum, leading also to severe distortions of other parts of the spectrum. Ways to counteract this entropic catastrophe will be presented.

II. MODEL AND METHOD

The method will here be demonstrated for the dynamic spin structure factor of the $S = 1/2$ Heisenberg spin chain, with the Hamiltonian

$$H = \sum_{i=1}^{L} \mathbf{S}_i \cdot \mathbf{S}_{i+1}. \tag{3}$$

The stochastic series expansion QMC algorithm [17] is used to compute the correlation function

$$G_q(\tau) = \left\langle S^z_{-q}(\tau) S^z_q(0) \right\rangle, \tag{4}$$

where S^z_q is the Fourier transform of the spins. With the kernel $K(\tau,\omega) = \pi^{-1} e^{-\tau\omega}$ in Eq. (1) and $\omega \in (-\infty,\infty)$, $A(\omega)$ is

the dynamic structure factor $S(q,\omega)$. At inverse temperature $\beta = 1/T$ it satisfies $S(q,-\omega) = e^{-\beta\omega}S(q,\omega)$. In the method to be discussed, it is more practical to define $A_q(\omega) = S(q,\omega)(1+e^{-\beta\omega})$, so that

$$K(\tau,\omega) = \frac{1}{\pi}\frac{e^{-\tau\omega}+e^{-(\beta-\tau)\omega}}{1+e^{-\beta\omega}} \tag{5}$$

and integrating over $\omega \in (0,\infty)$ in Eq. (1).

Here $G_q(\tau)$ is computed for a set $\tau \in \{\tau_1,\ldots,\tau_M\}$ with $\tau_j = (j-1)\Delta_\tau$ and because of symmetry properties, only the range $0 \leqslant \tau \leqslant \beta/2$ has to be considered. For large τ the statistical errors may become too large and the number of points M is therefore adjusted in this work so that the relative error never exceeds 10%.

With $A_q(\omega)$ parametrized as

$$A_q(\omega) = \sum_{n=1}^{N} a_n \delta(\omega-\omega_n), \quad \omega_n = (n-1/2)\Delta_\omega, \tag{6}$$

the weights $\{a_n\}$ will first be importance sampled using Eq. (2) with $\Theta = 1$ and later with a modified form. Different types of updates are carried out to transfer weight between two or more δ functions, with the normalization $G_q(0)$ conserved to achieve a high acceptance rate [6,8]. Conservation of higher moments can also be incorporated [6] but will not be done here. Single-weight updates account for the (small) normalization fluctuations.

The $T = 0$ results for $S(q,\omega)$ are available from Bethe ansatz (BA) calculations including two- and four-spinon processes, which accounts for almost all spectral weight [18]. Comparisons will be made with these results for a system with 500 spins [19] as well as with exact diagonalization results for an $L = 16$ chain at $T > 0$ [20].

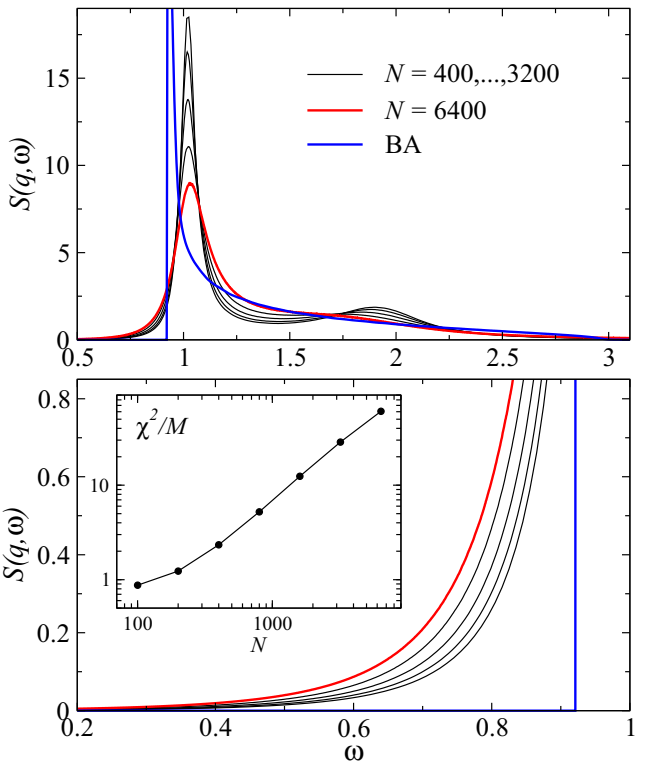

FIG. 1. Dynamic structure factor at $q = 0.8\pi$ obtained by unconstrained sampling for $\omega \in [0,4]$ and different N of the form 100×2^n (peak decreasing with increasing N), compared with a BA result [18,19]. The bottom panel shows details of the low-frequency part. The inset shows the goodness of the fit versus N.

III. UNCONSTRAINED SAMPLING

To illustrate the entropic problem with the sampling method in the $\Theta = 1$ formulation [8], results for $L = 500$ and $q = 0.8\pi$ are shown in Fig. 1. The QMC calculations were carried out at inverse temperature $\beta = 500$, which for all practical purposes gives $T = 0$ results for $G_q(\tau)$ at the momentum considered. The time spacing was $\Delta_\tau = 1/4$ and the number of data points $M = 33$. The relative statistical error of $G_q(\tau)$ was $\approx 10^{-5}$ at $\tau_1 = 0$ and ≈ 0.1 at τ_M. Figure 1 shows results obtained with several different numbers of δ functions in the spectrum. Comparing with the BA result, a striking feature is how the low-energy weight in the region below the actual spectral edge increases with increasing N (and the weight similarly increases also above the upper bound at $\omega \approx 3$), while the peak is suppressed. The main peak is too far to the right and there is a second, spurious peak at higher ω that is more prominent for small N. Overall, the results look similar to those of Ref. [8], where only a fixed $N = 1000$ was used.

From a statistical-mechanics point of view, it is clear that the sampling method suffers an entropic catastrophe for large N, with growing weight outside the bounds of the actual spectrum and therefore a rapidly increasing χ^2. Results indicating a similar problem with the Bayesian selection of Θ can be seen in Fig. 7 of Ref. [9]. To counteract the entropy, several modifications of the sampling method will be introduced next.

IV. CONSTRAINED SAMPLING AT $T = 0$

If the spectral bounds are known one can prevent the entropy-driven leakage of weight and, presumably, the associated distortions of the spectrum within the bounds. Normally the bounds are not known, however, but, as will be shown below, they can be approximately determined using the data. Before discussing how this is done, another important feature reducing the configurational entropy will be incorporated.

With the spectrum parametrized as in (6), no particular shape is imposed and when N becomes sufficiently large any spectrum can be reproduced in principle. In practice, however, one can only hope to resolve some prominent features of the spectrum. In particular, it is difficult to resolve a large number of closely spaced peaks. In many cases one has some prior information, e.g., one may know that the spectrum should have one or two peaks. In other cases, recognizing the generic limitations of analytic continuation, one may want to use a spectrum with the smallest number of peaks consistent with the QMC data. It is easy to impose a fixed number of peaks in sampling a δ-function sum (6), by starting with a spectrum with the desired number of peaks and only proposing updates that do not create or destroy peaks. Here a one-peak spectrum $A_q(w)$ will be considered [which implies a single peak also in $S(q,\omega)$, unless T is very high and a small peak at low ω can appear], but the procedures can be very easily generalized to any number of peaks.

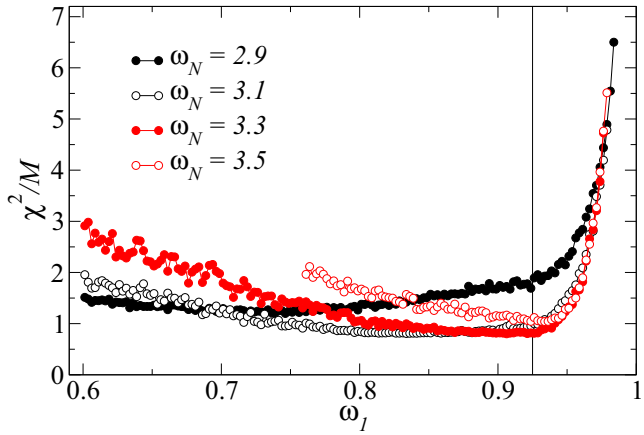

FIG. 2. Goodness of fit versus the lower bound of the spectrum for an $L = 500$ chain at $q = 0.8\pi$, for several choices of the upper bound ω_N and $\Delta_\omega = 0.0025$. The vertical line shows the location of the edge of the BA spectrum.

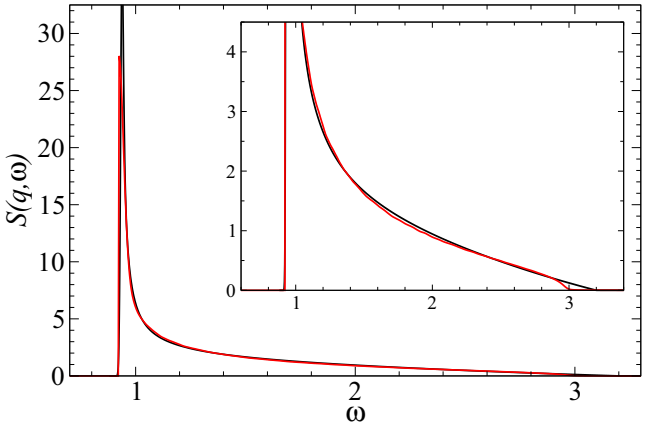

FIG. 3. The $T \to 0$ dynamic structure factor at $q = 0.8\pi$ for $L = 500$, obtained after two adjustments of the spectral bounds (black curve). The BA result [18,19] is shown with the red curve.

The bounds of the spectrum can be approximately determined by following the goodness of the fit as a function of the frequencies ω_1 and ω_N in Eq. (6). Fixing one of the bounds, ω_N say, a minimum in χ^2 versus ω_1 has to exist for large N, because the entropic effect is reduced as ω_1 is increased (provided of course that the true spectrum has vanishing or very small low-frequency weight), thereby reducing χ^2 until ω_1 starts to extend into the region of significant weight, whence χ^2 must increase. Figure 2 shows results of such scans for the normalized goodness of fit χ^2/M (with M used instead of the unknown number of degrees of freedom $N_{\rm DOF}$ [5]). The minimum χ^2/M is indeed for ω_1 close to the lower spectral edge and there is a sharp increase when ω_1 is pushed beyond the edge. The upper edge can be roughly determined to within 5%–10% of the location of the sharp decay in weight at $\omega \approx 3.0$ in the BA spectrum. The χ^2 minimum becomes more prominent for large N (hence making it easier to determine the bounds), in accord with the entropic scenario.

When determining the spectral bounds it is safe to allow χ^2 to deviate by a statistically insignificant amount proportional to $M^{1/2}$ from the best value $\chi^2_{\rm min}$ [given that the width of the χ^2 distribution is $(2N_{\rm DOF})^{1/2}$ and $M \sim N_{\rm DOF}$], going toward higher ω_1, where χ^2 grows very rapidly, and also toward higher ω_N, where the spectrum is less sensitive to the exact location of the bound. For the lower bound in the case of a spectrum with a sharp edge, as is the case here, one should not push ω_1 beyond the point where the peak of the spectrum is at the lower bound. One may also determine ω_1 by separately analyzing the large-τ behavior, though that is not always an easy task unless the lower edge is a well isolated δ function.

A faster way to identify the spectral bounds is to begin with a high upper edge (beyond what is expected for the true spectrum) and identify the best lower bound under that condition. With the lower bound fixed at its optimum, the upper bound can be optimized next. Iterating this procedure once or twice typically leads to excellent bounds very close to those obtained in a two-dimensional search. The results of such a procedure for a small spacing $\Delta_\omega = 0.001$ is shown in Fig. 3. The agreement with the BA calculation (which for $q = 0.8\pi$ misses about 2% of the known total spectral weight) is remarkably good. The peak location is off by only 1%, the lower bound slightly below it deviates by less than 0.5% from the true edge, and the nontrivial profile is essentially reproduced.

V. CONSTRAINED SAMPLING AT $T > 0$

In addition to the entropy-driven leakage of spectral weight outside the correct bounds, there is another entropic effect in the sampling of the single-peak spectrum at high (physical) temperature. In such a spectrum the volume of the accessible configuration space as a function of the peak height a_m (located at the mth δ function) is given by

$$V(a_m) = \frac{(a_m - a_0)^{m-1}}{(m-1)!} \frac{a_m^{N-m}}{(N-m)!}, \tag{7}$$

where a_0 is a floor imposed on the spectrum at the low-frequency bound $a_1 \geqslant a_0$, which again is regarded as an adjustable parameter to be optimized by monitoring $\chi^2(a_0)$. The floor at the high-frequency bound does not appear explicitly, being at 0 since the spectrum always decays to 0 when $\omega \to \infty$, unlike at $w \to 0$. Sampling a spectrum (6) without any data, i.e., with $\chi^2 = 0$ in Eq. (2), the fact that the configurational entropy $\ln(V)$ increases rapidly with a_m will drive the peak to infinite height (since no normalization is imposed). Sampling with χ^2 will of course counteract this effect, but still the entropy will unduly favor a sharp peak when N is large. This is not a serious issue in the $T = 0$ case discussed above (unless N is much larger than in Fig. 3), because this spectrum has a very sharp peak. However, at high T the peak entropy will cause problems, unless this version of the entropic catastrophe is counteracted by dividing the probability (2) by $V(a_m)$.

In order to obtain continuity as a function of T, considering that no entropic counterweighting was required above at $T = 0$, the following probability is used:

$$P(A) \propto \exp\{-\chi^2/2 - \lambda \ln[V(a_m)]\}, \tag{8}$$

where λ is also to be optimized using $\chi^2(\lambda)$. In practice, it was found that $\lambda = 1$ gives good solutions when the floor

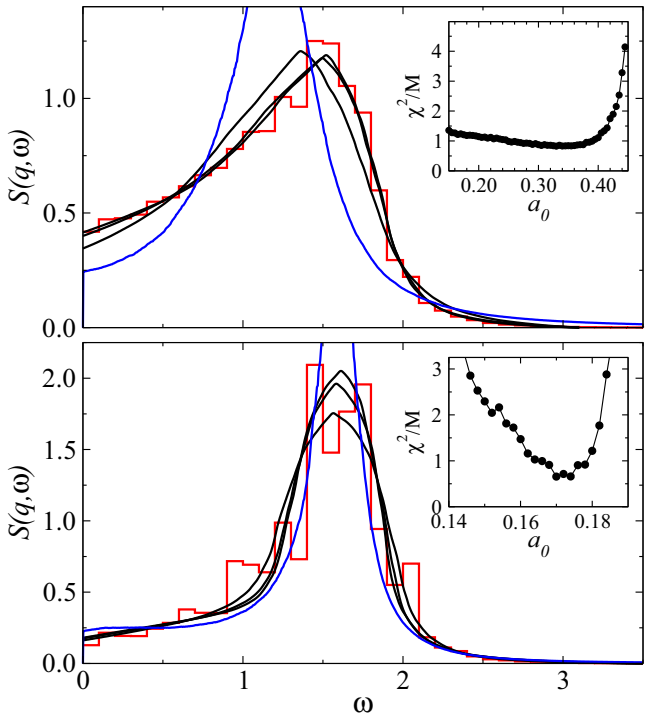

FIG. 4. Dynamic $q = \pi/2$ structure factor for $L = 16$ chains at $T = 1$ (top) and 0.5 (bottom). The histogram (red) represents exact diagonalization results. The black curves were obtained with $\lambda = 1$ in Eq. (8) and three different values of the floor a_0: at the minimum $\chi^2(a_0)$ (curves with the lowest a_0) and for higher values where χ^2/M is approximately its minimum value plus $M^{-1/2}$ and $2M^{-1/2}$. The insets show $\chi^2(a_0)/M$. The curves with higher peaks (blue) are from unconstrained sampling. The upper spectral bounds were also chosen according to a χ^2 criterion, as discussed in the text.

$a_0 > 0$, while optimizing $\lambda \in [0,1]$ is better when $a_0 = 0$. Optimizing λ after identifying the spectral bounds in the $T = 0$ case discussed above gave $\lambda \approx 0$ and no significant change in the spectrum from Fig. 3. The optimal λ varies monotonically as T is increased.

The form (8) and the optimization procedures can be easily generalized to more than one peak. An even better form of the probability with entropy suppression may possibly be obtained by using $V(a_m)$ at fixed normalization, which, however, is a much more complicated function that has not yet been evaluated in closed form.

Figure 4 shows results at $T = 1$ and $1/2$ for an $L = 16$ chain, obtained using $\lambda = 1$ and scanning over a grid of a_0 values. Exact diagonalization results for the spectrum are represented by histograms [20] and one of course cannot expect to resolve the fine structures in such a spectrum by analytic continuation of QMC results. With the single-peak property imposed, however, one can observe very good agreement with the broad features, including very reasonable values for the low-energy limit, when choosing a_0 such that χ^2 is close to its minimum value. In practice, it is better to go slightly beyond the floor value minimizing χ^2. When a_0 is taken past the minimizing value χ^2 is seen growing rapidly and the spectrum does not change much initially in this region, though it changes noticeably at high T for smaller a_0. Since the best value $\chi^2_{\rm min}$ can fluctuate of the order $M^{1/2}$, it is statistically sound to choose a_0 where $\chi^2 \approx \chi^2_{\rm min} + M^{1/2}$, where the solution typically has stabilized before χ^2 increases sharply. The solution is again not very sensitive to the upper bound as long as ω_N is reasonably close to the value to optimizing χ^2. One can again determine the two parameters in an iterative fashion, adjusting a_0 first with a high ω_N, then adjusting ω_N to where $\chi^2 \approx \chi^2_{\rm min} + M^{1/2}$ (above the point where χ^2 is minimized), and repeating this once or twice.

Results of this optimized constrained sampling scheme are seen in Fig. 4 to be much better than those of unconstrained sampling, which leads to excessively sharp peaks. One can also counteract the peak sharpness in the unconstrained case, e.g., by imposing a ceiling on the weights a_i in the sampling. However, results of such a procedure are still not as good as with the constrained sampling, where the form (7) provides a more natural mechanism for suppressing the entropy and the shape of the spectrum comes out remarkably well.

VI. DISCUSSION

The main result of this work is the identification of configurational entropy as a detriment to stochastic analytic continuation with the sampling temperature $\Theta = 1$ [8]. A remarkable improvement in fidelity can be achieved with respect to other methods by suppressing the entropy in various ways. An important aspect of these procedures is that the average spectrum no longer depends on the number of δ functions N used to parametrize it, once N is sufficiently large for discretization effects on the scale of the main spectral features to become unimportant.

A bottleneck of the method is that sampling has to be carried out for many values of the parameters to be optimized: ω_1, ω_N, a_0, and λ. However, in practice good results can be obtained with simple scans over a single parameter as follows. For fixed ω_N, if $\omega_1 = \Delta_\omega/2$ is found to be optimal, then a_0 is adjusted with $\lambda = 1$. If the optimum is at $a_0 = 0$, then λ is optimized. If $\omega_1 > 0$ is optimal one should subsequently also optimize λ. Very good results for long Heisenberg chains were obtained in this way for the full range of temperatures, where comparisons can be made with the results of time-dependent density-matrix renormalization calculations [21].

ACKNOWLEDGMENTS

I would like to thank K. Beach, J.-S. Caux, Z. Y. Meng, B. Normand, O. Syljuåsen, and T. Xiang for valuable discussions and J.-S. Caux also for providing his BA data. This research was supported by the NSF under Grant No. DMR-1410126 and by the Simons Foundation.

[1] S. F. Gull and J. Skilling, Proc. IEEE **131**, 646 (1984).
[2] R. N. Silver, D. S. Sivia, and J. E. Gubernatis, Phys. Rev. B **41**, 2380 (1990); J. E. Gubernatis, M. Jarrell, R. N. Silver, and D. S. Sivia, *ibid.* **44**, 6011 (1991).

[3] H.-B. Schüttler and D. J. Scalapino, Phys. Rev. Lett. **55**, 1204 (1985); Phys. Rev. B **34**, 4744 (1986).
[4] S. R. White, D. J. Scalapino, R. L. Sugar, and N. E. Bickers, Phys. Rev. Lett. **63**, 1523 (1989).
[5] M. Jarrell and J. E. Gubernatis, Phys. Rep. **269**, 133 (1996).
[6] A. W. Sandvik, Phys. Rev. B **57**, 10287 (1998).
[7] K. S. D. Beach, arXiv:cond-mat/0403055.
[8] O. F. Syljuåsen, Phys. Rev. B **78**, 174429 (2008).
[9] S. Fuchs, T. Pruschke, and M. Jarrell, Phys. Rev. E **81**, 056701 (2010).
[10] Q.-S. Wu, Y.-L. Wang, Z. Fang, and X. Dai, Chin. Phys. Lett. **30**, 090201 (2013).
[11] D. R. Reichman and E. Rabani, J. Chem. Phys. **131**, 054502 (2009).
[12] D. N. Aristov, C. Brünger, F. F. Assaad, M. N. Kiselev, A. Weichselbaum, S. Capponi, and F. Alet, Phys. Rev. B **82**, 174410 (2010).
[13] H. Feldner, Z. Y. Meng, T. C. Lang, F. F. Assaad, S. Wessel, and A. Honecker, Phys. Rev. Lett. **106**, 226401 (2011).
[14] F. Goth, D. J. Luitz, and F. F. Assaad, Phys. Rev. B **88**, 075110 (2013).
[15] S. R. White, in *Computer Simulation Studies in Condensed Matter Physics III*, edited by D. P. Landau, K. K. Mon, and H.-B. Schuttler (Springer, Berlin, 1991).
[16] A. W. Sandvik and R. R. P. Singh, Phys. Rev. Lett. **86**, 528 (2001).
[17] A. W. Sandvik, AIP Conf. Proc. **1297**, 135 (2010); J. Phys. A **25**, 3667 (1992); A. W. Sandvik, R. R. P. Singh, and D. K. Campbell, Phys. Rev. B **56**, 14510 (1997).
[18] J.-S. Caux and J. M. Maillet, Phys. Rev. Lett. **95**, 077201 (2005); J.-S. Caux, R. Hagemans, and J. M. Maillet, J. Stat. Mech. (2005) P09003; R. G. Pereira, J. Sirker, J.-S. Caux, R. Hagemans, J. M. Maillet, S. R. White, and I. Affleck, Phys. Rev. Lett. **96**, 257202 (2006).
[19] Data provided by J.-S. Caux (private communication).
[20] O. A. Starykh, A. W. Sandvik, and R. R. P. Singh, Phys. Rev. B **55**, 14953 (1997).
[21] T. Barthel, U. Schollwöck, and S. R. White, Phys. Rev. B **79**, 245101 (2009).