

## Error reduction using machine learning on Ising worm simulation

---

**Jangho Kim<sup>a,\*</sup> and Wolfgang Unger<sup>b</sup>**

<sup>a</sup>*Institute for Advanced Simulation (IAS-4), Forschungszentrum Jülich,  
Wilhelm-Johnen-Straße, 52428 Jülich, Germany*

<sup>b</sup>*Fakultät für Physik, Universität Bielefeld,  
Universitätsstraße 25, D33619 Bielefeld, Germany*

*E-mail:* [j.kim@fz-juelich.de](mailto:j.kim@fz-juelich.de), [wunger@physik.uni-bielefeld.de](mailto:wunger@physik.uni-bielefeld.de)

We develop a method to improve on the statistical errors for higher moments using machine learning techniques. We present here results for the dual representation of the Ising model with an external field, derived via the high temperature expansion and simulated by the worm algorithm. We compare two ways of measuring the same set of observables, without and with machine learning: moments of the magnetization and the susceptibility can be improved by using the decision tree method to train the correlations between the higher moments and the second moment obtained from an integrated 2-point function. Those results are compared in small volumes to analytic predictions.

*The 39th International Symposium on Lattice Field Theory,  
8th-13th August, 2022,  
Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany*

---

\*Speaker

## 1. Ising dual representation

The dual representation of the Ising model is derived by introducing bond variables and integrating out the spin degrees of freedom [1–3]:

$$\begin{aligned} Z_{\text{Ising}} &= \sum_{\{s\}} e^{-\beta H(s)}, \quad H = -J \sum_{\langle i,j \rangle} s_i s_j + h \sum_i s_i \\ &= (2 \cosh(\beta h))^V \cosh(\beta J)^E \sum_{\substack{\{n_b, m_i\} \\ \partial\{n_b\} = \{m_i\}}} \tanh(\beta J)^{\sum_b n_b} \tanh(\beta h)^{\sum_i m_i}, \end{aligned} \quad (1)$$

where the first line is the standard spin representation of the Ising model and the second line is its dual representation derived by the high temperature character expansion in  $\beta = 1/T$ , and  $h$  is the external magnetic field. The dual variables of that representation are the monomers  $m_i \in \{0, 1\}$  defined on the lattice sites  $i$  and the dimers  $n_b \in \{0, 1\}$  defined on the bonds  $b$ , which are nearest neighbor pairs. Here,  $V$  is the volume,  $E$  is the number of bonds. Since spin summation only gives non-trivial contributions if after the expansion the spin at every site is raised to an even power, we obtain the constraint that the dimers form intersecting loops that are either closed or form strings that connect two monomers:  $\partial\{n_b\} = \{m_i\}$ . The dual representation is well suited to be simulated via Monte Carlo, in particular using the worm algorithm [1]. Then we can measure the number of monomers  $M = \sum_i m_i$  and its higher moments on each configuration obtained after a worm update, but also so-called improved estimators, such as the 2-point correlation functions during worm evolution [2]. It turns out that the averaged worm length is simply the connected susceptibility obtained via the integrated 2-point function

$$\langle G_2 \rangle = \frac{1}{V^2} \sum_{x,y} \langle G(x,y) \rangle = \langle \sigma^2 \rangle \quad (2)$$

with  $G(x,y) = G(x-y)$  due to translation symmetry.

The magnetization  $\langle \sigma \rangle$  and the susceptibility  $\chi$  can be written in terms of the total monomer number  $M$  as follows,

$$\langle \sigma^n \rangle = \frac{1}{(N\beta)^n} \frac{1}{Z} \frac{\partial^n Z}{\partial h^n} = \langle f_n \rangle \quad (3)$$

$$\langle \sigma \rangle = \tanh(\beta h) + \frac{\langle M \rangle}{\sinh(\beta h) \cosh(\beta h)} = \langle f_1 \rangle \quad (4)$$

$$\begin{aligned} \chi = \langle \sigma^2 \rangle - \langle \sigma \rangle^2 &= \frac{1}{N \cosh^2(\beta h)} - \frac{1}{N} \left( \frac{1}{\sinh^2(\beta h)} + \frac{1}{\cosh^2(\beta h)} \right) \langle M \rangle \\ &\quad + \frac{1}{(\sinh(\beta h) \cosh(\beta h))^2} \left( \langle M^2 \rangle - \langle M \rangle^2 \right) = \langle f_2 \rangle - \langle f_1 \rangle^2. \end{aligned} \quad (5)$$

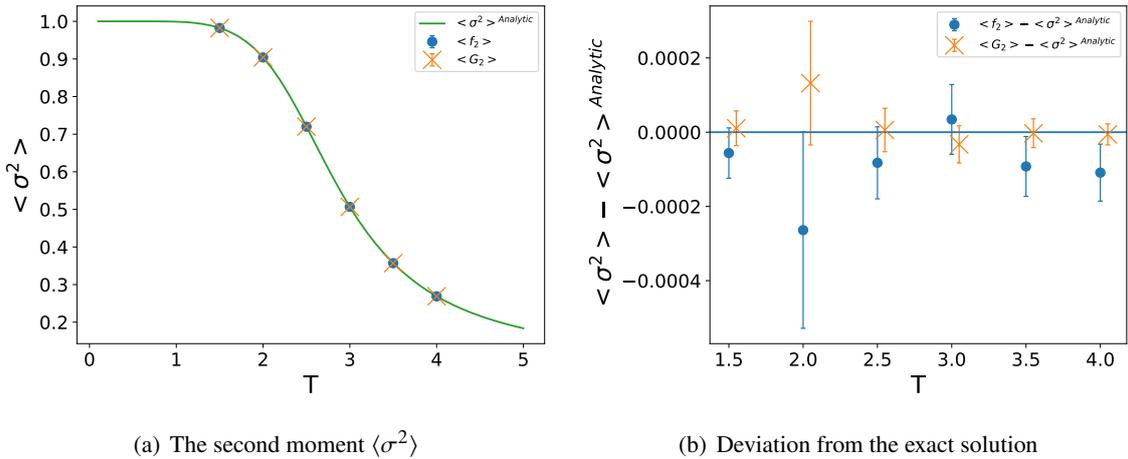
Here, we define  $f_n$  to distinguish the observables written in terms of  $M$  from the same observable written in terms of the improved estimator  $G_2$ . For example,  $f_2 = \sigma^2(M)$ . Note that they are not the same  $f_2 \neq G_2$  before ensemble averaging, as they have different distributions. We compare  $\langle f_2 \rangle$  and  $\langle G_2 \rangle$  with the exact solution in Fig. 1. The analytic solution is subtracted in Fig. 1(b) to see the statistical errors and the deviations from the analytic result. Since  $\langle G_2 \rangle$  as a worm estimator has better statistics and hence smaller error bars compared to  $\langle f_2 \rangle$  (see Fig 2(b)), also its mean value

is closer to the analytic result, see Fig. 2(a). This advantage of improved estimators concerning statistical errors is strong in particular at higher temperatures, including the vicinity of a critical point, but it weaker at low temperatures as the worm algorithm is less efficient here, since the average worm length is very large here.

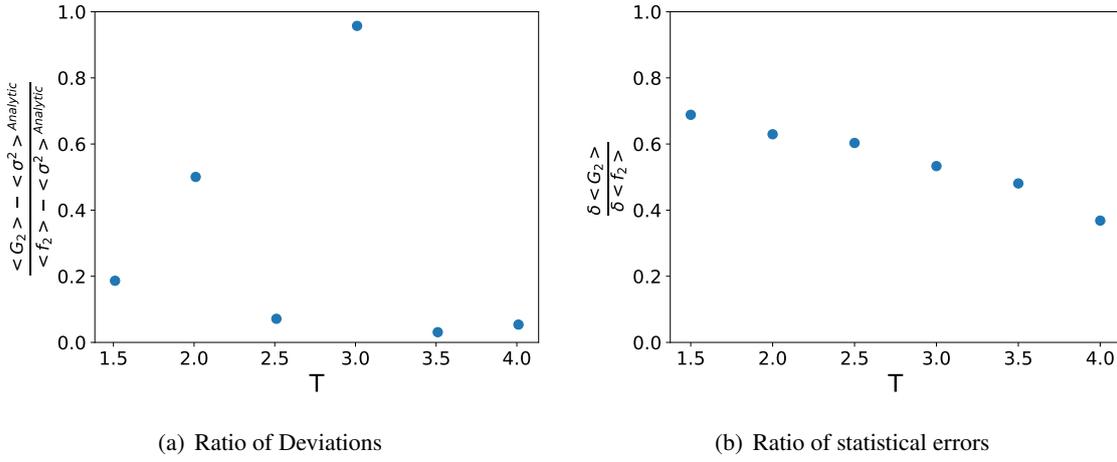
## 2. Machine learning strategy

The important observable to determine the critical temperature is the susceptibility  $\chi = \langle \sigma^2 \rangle - \langle \sigma \rangle^2$ . The pseudo-critical temperature can be determined from its peak via finite size scaling. Whereas  $\langle \sigma^2 \rangle$  can be determined by  $\langle G_2 \rangle$ , there is no improved estimator for  $\langle \sigma \rangle$ , and it has to be determined by  $\langle f_1 \rangle$ , which is less accurate. While  $\delta \langle G_2 \rangle$  is small, the dominant error of the susceptibility comes from  $\delta(\langle f_1 \rangle^2)$ . The goal of our machine learning strategy is to reduce the statistical error of the susceptibility by predicting a new observable  $\langle \tilde{G}_1 \rangle$ , which corresponds to  $\langle \sigma \rangle$ , with a reduced error. To obtain the general mapping of the distributions of the means  $\langle \sigma^2 \rangle$  and  $\langle \sigma \rangle$ , we consider as training data the correlation of bootstrap samples ( $n = 1000$ ) between  $\langle f_2 \rangle$  and  $\langle f_1 \rangle$ . With this, we train the machine this correlation on a  $4 \times 4$  lattice with the external field  $h = 0.2$ , as presented in Fig. 3(a). In Fig. 3(b), we present the machine learning prediction  $\langle \tilde{G}_1 \rangle$ . We obtain this mapping by applying the decision tree regression method from the scikit-learn library [4]. The decision tree regression method is used to select the closest data point of the input. The blue points are selected by the decision tree from the input of  $\langle G_2 \rangle$ . The red and blue crosses indicate the statistical error of  $\langle f_2 \rangle$ ,  $\langle f_1 \rangle$  and  $\langle G_2 \rangle$ ,  $\langle \tilde{G}_1 \rangle$ . Comparing with the analytic solution, the green star, the machine learning prediction  $\langle \tilde{G}_1 \rangle$  has smaller statistical error and deviation than  $\langle f_1 \rangle$ .

The distribution of the bootstrap samples is Gaussian. In Fig. 4(a), we have shown that both  $\langle f_2 \rangle$  and  $\langle G_2 \rangle$  are Gaussian distributions and  $\langle G_2 \rangle$  has smaller statistical error. Here, the purple line is the analytic result. After applying the decision tree regression, the machine learning prediction has also Gaussian distribution in Fig. 4(b).



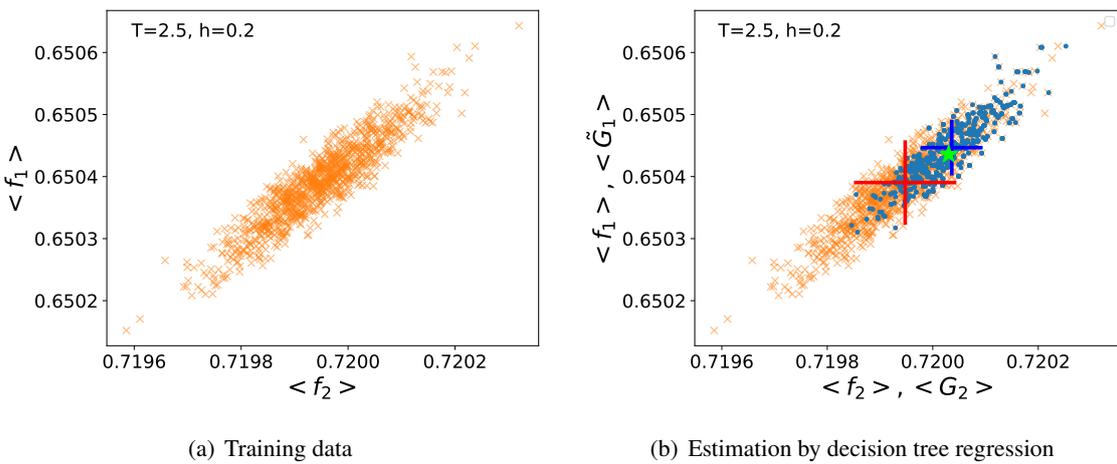
**Figure 1:** Left: Comparison of  $\langle f_2 \rangle$  and  $\langle G_2 \rangle$  with analytic solution on a  $4 \times 4$  lattice and the external field  $h = 0.2$ . Right: Subtracting the exact solution from  $\langle f_2 \rangle$  and  $\langle G_2 \rangle$ .



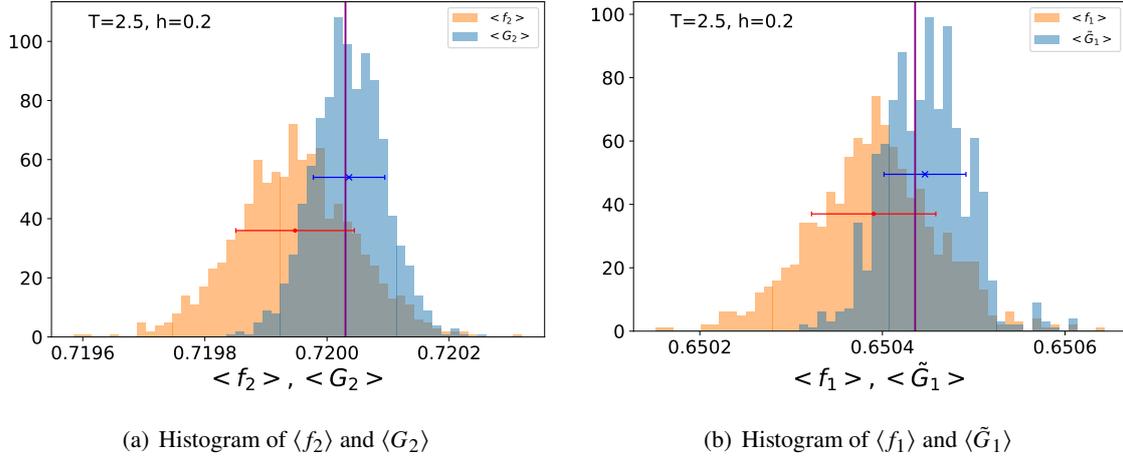
**Figure 2:** Left: Ratio of the deviations from analytic solution on a  $4 \times 4$  lattice for  $h = 0.2$ . Right: Ratio of the statistical errors of  $\langle f_2 \rangle$  and  $\langle G_2 \rangle$

The magnetization  $\langle \tilde{G}_1 \rangle$  obtained by machine learning for temperatures from  $T = 1.5$  to  $T = 4.0$  are presented in Fig. 5. In Fig. 6(a) and Fig. 6(b), we compare the deviation and statistical error of  $\langle \tilde{G}_1 \rangle$  and  $\langle f_1 \rangle$ . As a result, machine learning predictions are more accurate and closer to the true result. The statistical errors are reduced by about 40%.

The susceptibility can be obtained by subtracting  $\langle f_1 \rangle^2$  or  $\langle \tilde{G}_1 \rangle^2$  from  $\langle G_2 \rangle$ . In Fig. 7, we compare two ways of calculations. When the machine learning prediction  $\langle \tilde{G}_1 \rangle^2$  is subtracted, the results are closer to the analytic result. The reduction of deviations and statistical errors are presented in Fig. 8(a) and Fig. 8(b). The statistical errors are reduced by about 20-70% depending on the temperature.



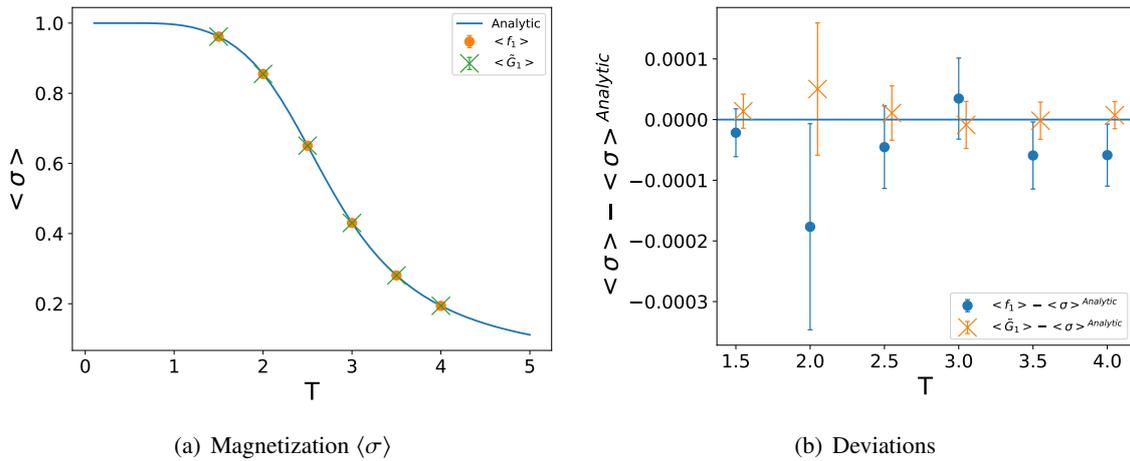
**Figure 3:** Left: Correlation between the bootstrap samples of  $\langle f_2 \rangle$  and  $\langle f_1 \rangle$  at  $T = 2.5$  and  $h = 0.2$ . Right: blue points are the decision tree regression prediction for  $\langle G_2 \rangle$  input. Green point is the analytic solution.



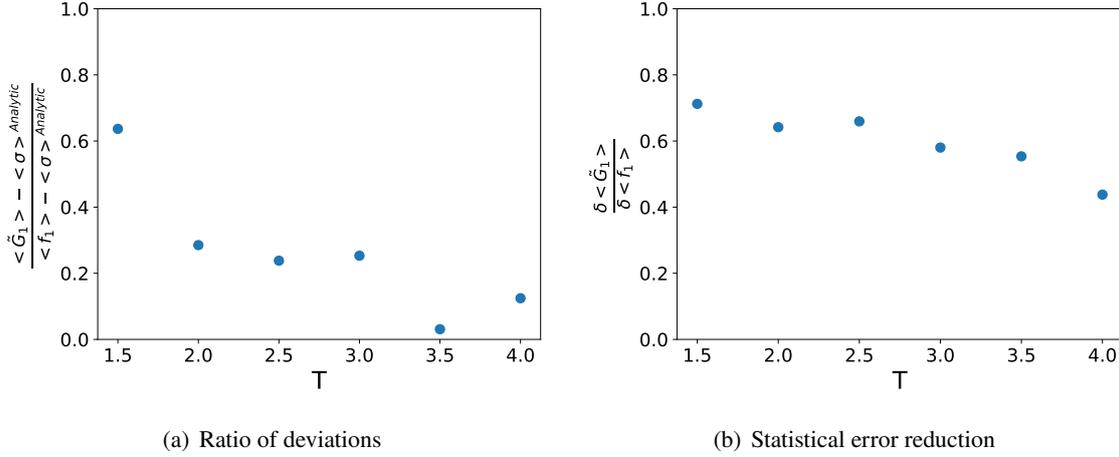
**Figure 4:** Left: histogram of  $\langle f_2 \rangle$  and  $\langle G_2 \rangle$  at  $T = 2.5$  and  $h = 0.2$ . Right: histogram of  $\langle f_1 \rangle$  and  $\langle \tilde{G}_1 \rangle$  at  $T = 2.5$  and  $h = 0.2$ .

### 3. Higher moments

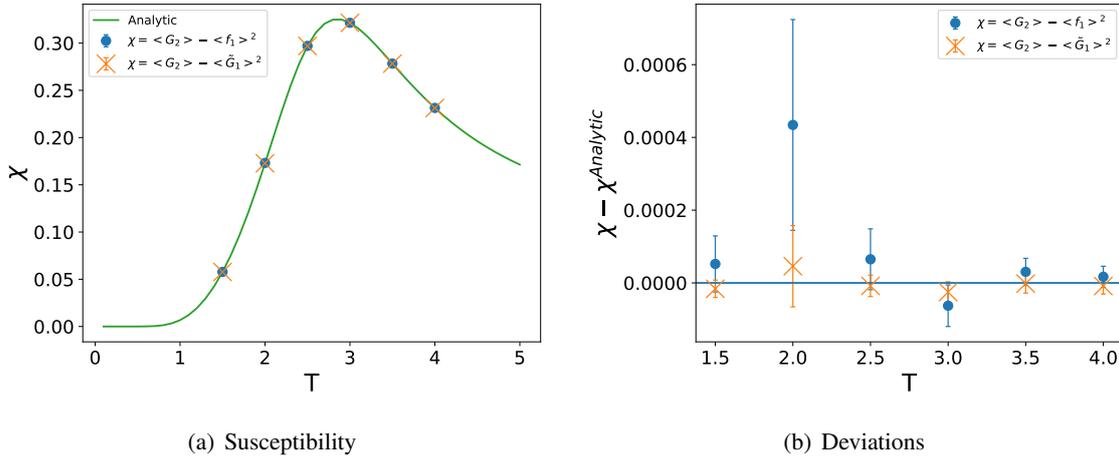
This method is applicable to higher moment, for instance,  $\langle \sigma^3 \rangle$  or  $\langle \sigma^4 \rangle$ . The effect of the statistical error reduction depends on how strong the correlation is, which can be quantified by the Pearson coefficient. In Fig. 9(b), we show the dependence of statistical error reduction with respect to Pearson coefficient of the correlation between  $\langle f_2 \rangle$  and  $\langle f_n \rangle$ , where  $n = 1, 3, 4$ . In the case of  $\langle f_1 \rangle$ , it is clear that the statistical error reduction is more effective at strong correlations. However, in the case of higher moments,  $n = 3, 4$ , the statistical errors are smaller but not as much as for the magnetization. The reason is that the accuracy of the input data affects the error reduction. In Fig. 2(b), the input data is less accurate at lower temperatures. Hence the error reduction is less



**Figure 5:** Left: Comparison of  $\langle f_1 \rangle$  and  $\langle \tilde{G}_1 \rangle$  with analytic solution on  $4 \times 4$  lattice with external field  $h = 0.2$ . Right: subtracting analytic solution from  $\langle f_1 \rangle$  and  $\langle \tilde{G}_1 \rangle$ .

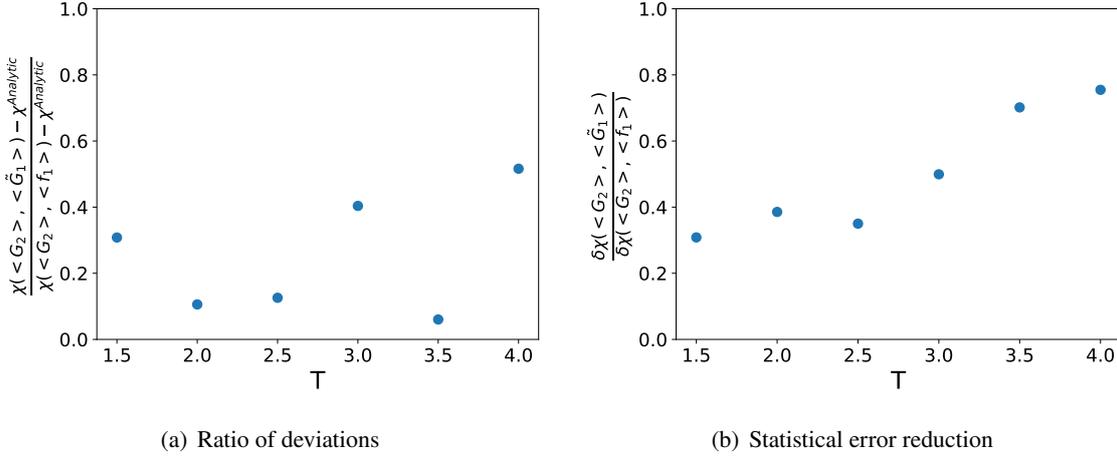


**Figure 6:** Left: Comparison of  $\langle f_1 \rangle$  and  $\langle \tilde{G}_1 \rangle$  with analytic solution on  $4 \times 4$  lattice. Right: Ratio of the statistical errors of  $\langle f_1 \rangle$  and  $\langle \tilde{G}_1 \rangle$

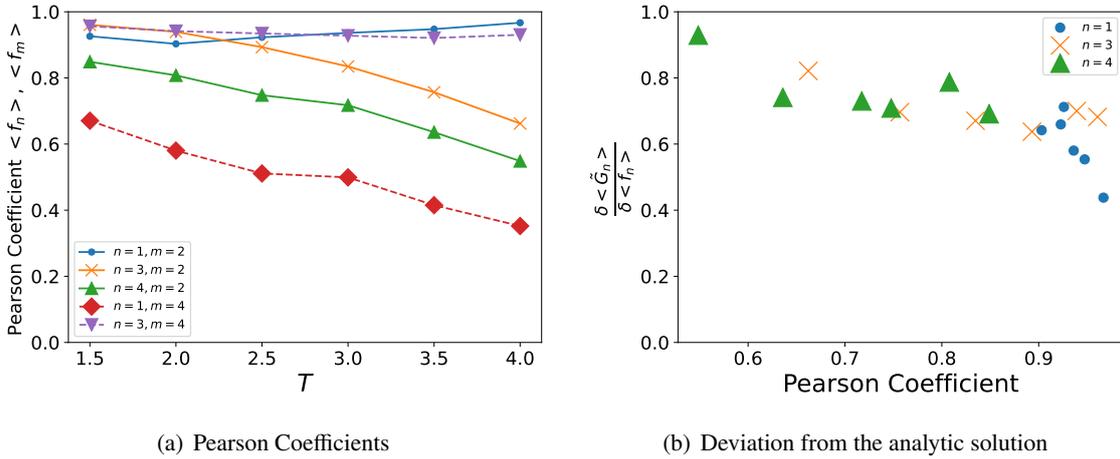


**Figure 7:** Comparison of  $\chi = \langle G_2 \rangle - \langle f_1 \rangle^2$  and  $\chi = \langle G_2 \rangle - \langle \tilde{G}_1 \rangle^2$  with analytic solution on  $4 \times 4$  lattice.

effective despite the Pearson coefficient is larger than 0.9. The two data points in Fig. 9(b) for  $n = 3$  with largest Pearson coefficient correspond to rather low temperatures  $T = 1.5, 2.0$ . Their Pearson coefficients are close to one but the error reduction is not large, about 30%. On the other hand, the data for  $n = 1$  with largest Pearson coefficient correspond to high temperature which has very accurate input data. Despite of accurate input data at high temperatures, the small error reduction for the other data is due to the weak correlation, as seen in Fig. 9(a) for  $n = 3, 4$  and  $m = 2$ . The Pearson coefficient for the correlation between  $\langle f_3 \rangle$  and  $\langle f_4 \rangle$  is large even at high temperatures. Hence, if we have the accurate input  $\langle G_4 \rangle$ ,  $\langle \tilde{G}_3 \rangle$  can be evaluated by our machine learning method precisely.  $\langle G_4 \rangle$  can be sampled by introducing a second worm: sampling the Ising model with such a two-worm algorithm [5], the four-point function  $G(x, y, z, w)$  and its integrated expectation value  $\langle G_4 \rangle$  can be directly measured as an improved estimator. Then the correlation between  $\langle f_3 \rangle$  and  $\langle f_4 \rangle$  can be learned from the input  $\langle G_4 \rangle$ , resulting in an error reduction for  $\langle \tilde{G}_3 \rangle$ .



**Figure 8:** Left: deviation from the analytic solution. Right: Statistical error reduction by machine learning method for susceptibility.



**Figure 9:** Pearson coefficient of the correlation between  $f_n$  and  $f_m$ , where  $m = 2, 4$ . Statistical error reduction with respect to Pearson coefficient.

#### 4. Conclusion

We developed an error reduction strategy using the decision tree method. We have tested this method for the Ising model in its dual representation and found that the error of the magnetization  $\langle \sigma \rangle$  is reduced about 40%. Moreover, for the susceptibility, the mean value of the machine learned prediction is closer to the analytic result. Applying this method to higher moments is less efficient because of the weak correlation of the worm estimator with the higher moments in terms of monomers. A two-worm algorithm with external field is required for its improvement, and is currently under investigation. We also plan to test this method for larger volumes and higher dimensions and determine observables for finite size scaling, such as the Binder cumulant, with reduced statistical error from the decision tree method. Finally, we aim is to apply this method to

strong coupling lattice QCD in the dual representation [6]. Here, we may benefit for improving on chiral and nuclear observables, to pinpoint the QCD phase diagram in the strong coupling regime.

## Acknowledgments

J.K. was supported in part by the NSFC and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through the funds provided to the Sino-German Collaborative Research Center TRR110 "Symmetries and the Emergence of Structure in QCD" (NSFC Grant No. 12070131001, DFG Project-ID 196253076 - TRR 110) W.U. acknowledges support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through the CRC-TR 211 'Strong-interaction matter under extreme conditions' – project number 315477589 – TRR 211.

## References

- [1] N. Prokof'ev and B. Svistunov, *Worm algorithms for classical statistical models*, *Physical Review Letters* **87** (sep, 2001).
- [2] U. Wolff, *Simulating the All-Order Strong Coupling Expansion I: Ising Model Demo*, *Nucl. Phys. B* **810** (2009) 491–502, [[0808.3934](#)].
- [3] C. Gabriel, *Dynamical properties of the Worm Algorithm*, *Thesis* (2002).
- [4] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, *Scikit-learn: Machine learning in Python*, *Journal of Machine Learning Research* **12** (2011) 2825–2830.
- [5] R. Brett, *Binder Cumulants for the Ising Model using Worm algorithms*, *Thesis* (2015).
- [6] G. Gagliardi, J. Kim, and W. Unger, *Dual Formulation and Phase Diagram of Lattice QCD in the Strong Coupling Regime*, *EPJ Web Conf.* **175** (2018) 07047, [[1710.07564](#)].