

Predicting the optimal noise strength for solving optimization problems with analog Ising machines

Leen Mys¹,* Guy Verschaffelt¹, and Guy Van der Sande¹

Applied Physics Research Group, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium



(Received 11 September 2025; revised 21 January 2026; accepted 17 February 2026; published 20 March 2026)

Analog Ising machines are dedicated hardware solvers designed to solve NP hard optimization problems. However, the global optimum is often not found as the system gets stuck in local minima. While several strategies exist to increase the chance of escaping local minima, these methods often require extensive parameter tuning. In this work, we investigate the injection of large noise as a standalone scheme and in combination with annealing to improve the success rate and the time-to-target (TTT) of analog Ising machines for MaxCut problems. We demonstrate that optimizing the noise improves the TTT by several orders of magnitude and renders both approaches competitive with state-of-the-art schemes, such as chaotic amplitude control. Moreover, we are able to predict an optimal noise value based on the problem connectivity, statistics of the coupling matrix and the coupling strength, thereby eliminating the need for costly parameter optimization.

DOI: [10.1103/czsz-z8ng](https://doi.org/10.1103/czsz-z8ng)

I. INTRODUCTION

As the demand for energy-efficient and high-speed computational power continues to grow, there is increasing interest in alternative computing systems to complement or replace the traditional von Neumann system for specific highly demanding tasks. Combinatorial optimization problems, in particular, would greatly benefit from faster and more efficient computing paradigms. A nontraditional approach to solve optimization problems in general is the so-called Ising machine (IM), inspired by the NP-complete Ising problem. In an Ising problem the goal is to find the spin configuration of a magnetic material that minimizes its energy. An IM is a piece of hardware where the Ising problem is recreated using artificial spins. It relies on the natural tendency of the Ising spins to evolve toward their ground state. So, it naturally minimizes the Ising energy function

$$H_{\text{Ising}} = -\frac{1}{2} \sum_{ij} J_{ij} \sigma_i \sigma_j - \sum_i h_i \sigma_i, \quad (1)$$

where J_{ij} is the coupling interaction between two spins, N is the number of spins in the problem, h_i is the external

field, and $\sigma_i \in \{-1, 1\}$ is the binary spin. The optimization problem to be solved is encoded in the coupling interaction $\mathbf{J}$ and the local field $\mathbf{h}$ while the ground-state configuration is the solution of the problem. It has been shown that all combinatorial optimization problems can be mapped polynomially onto an Ising problem. There are already a range of problems for which it is known how to efficiently map them on an Ising problem [1]. For example, IMs can solve a wide range of real-life problems [2] ranging from job scheduling [3] to protein folding [4], traffic-flow optimization [5], and finance-related problems [6,7].

Here, we concentrate on analog IMs without local fields ($h_i = 0$). To facilitate optimization of an Ising problem, the spin amplitudes are allowed to take a real value. We then refer to analog spin amplitudes indicated by x_i in analog IMs. This analog spin system can be mapped back to the original binary spin system by taking the sign of the spin amplitudes ($\sigma_i = \text{sign}(x_i)$). When using analog spins, it must be verified that the ground state of the analog spin problem, to which the spins evolve naturally, results in the optimal energy of the original binary problem. To ensure this, nonlinear terms are often employed to impose a binary structure on the analog spin during the dynamic evolution [8]. The choice of a nonlinear term can also affect the efficiency of the search for the ground state [9].

There exists a multitude of physical implementations of IMs using electronic [10] or photonic [11] platforms. Possible optical analog Ising implementations use (degenerate) optical parametric oscillators (DOPOs) [2,12–15], opto-electric oscillators (OEOs) [16,17], or optical Kerr resonators [18]. Different physical implementations induce

*Contact author: leen.mys@vub.be

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

different nonlinear terms, thereby influencing the efficiency of the IM [9]. A major challenge in analog IMs is that the spins can become trapped in a local minimum, preventing convergence to the desired global optimum [8]. Several factors or techniques can limit the chance of being trapped, including annealing [7,19], tweaking nonlinearities [9], stochastic perturbations through noise injections [20,21], simulated bifurcation [22,23], or using amplitude control schemes [24].

Although both annealing and noise injection are commonly used in analog IMs [9,25], it remains unclear under which conditions one approach is more beneficial than the other. Furthermore, recent studies have shown that injecting high levels of noise can be beneficial in improving the performance of IMs in general [26–30], and analog IMs in particular [20,21]. However, in most of these studies, noise is treated primarily as a tunable optimization parameter, without a deeper investigation into why or how it affects the time-to-solution and success rate. The presence of an optimal noise level has been reported in Ref. [21], but was not further explored and was only demonstrated to be optimal for the success rate. We aim to predict the optimal noise level from basic system characteristics, such as the distribution of the elements of the coupling matrix or the coupling strength, rather than relying on exhaustive parameter scans or Bayesian optimization methods. Although effective, these methods are often computationally expensive or require prior knowledge of the ground-state energy, which may not be available in practice.

In this work, we will investigate the role of noise injection in analog IMs, both as a standalone scheme and in combination with annealing. We benchmark the performances of the models including noise against the considered state-of-the-art chaotic amplitude control (CAC) IM using MaxCut problems to better understand the effect of noise in both situations. Additionally, we explore how the optimal noise regime depends on problem characteristics, particularly the average number of connections per spin and the standard deviation of the elements of the coupling matrix. We find a relationship between the noise and some basic system parameters, allowing us to predict the optimal noise strength.

In Sec. II, we will introduce the different schemes, nonlinearities, and benchmarks that have been used in our investigation. We then use one benchmark as an example to illustrate the influence of noise and investigate the relationship between different system parameters in Sec. III. Section IV will dive deeper in the performance differences between the different models. After that, we will investigate the relationship between noise and the system parameters in Sec. V. We analyze small problems ($N \lesssim 100$) and large problems ($N > 100$) separately. Afterwards, we will verify the found relationship between noise and the system parameters and conclude the paper with some discussion and future work in Sec. VI.

II. MODELS AND BENCHMARKS

An analog IM is a (physical) system of interconnected analog spins. The analog spins have the tendency to evolve naturally towards the configuration of minimal energy. This evolution can be modeled by the differential equation

$$\frac{dx_i}{dt} = G \left[x_i, \alpha, \beta \sum_{j=1}^N J_{ij} x_j, \gamma F(t) \right]. \quad (2)$$

Here, G is the nonlinear transfer function, α is the linear gain, β is the coupling strength, $x_i \in \mathbb{R}$ is the analog spin amplitude, t is the renormalized time, and $\gamma F(t)$ is a Gaussian white-noise term with $\langle F(t) \rangle = 0$ and $\langle F(t)F(t') \rangle = 2\delta(t - t')$. As is described in Ref. [9], the transfer function depends on the specific implementation and hardware of the IM. As this paper is only based on simulations and theory, we use transfer functions, which minimize the number of parameters to simplify the analysis. Therefore, we will use the so-called polynomial and clipped transfer function. For these transfer functions, the parameter α can be eliminated by (re)defining x_i , t , and β , as described in Ref. [25]. The differential equation for the polynomial nonlinearity is

$$\frac{dx_i}{dt} = -x_i - x_i^3 + \beta \sum_{j=1}^N J_{ij} x_j + \gamma F(t). \quad (3)$$

This is the more basic nonlinearity for several reasons: it is the normal form of the pitchfork bifurcation of the spins. The polynomial transfer function also describes the traditional optical setup of a coherent IM using DOPO [2,12–15]. Furthermore, it is the Taylor expansion of the other transfer function [9]. In this study, it is used to provide the necessary insight into the working of noise. The performance of the IM with the clipped nonlinearity is substantially better than the performance of the polynomial nonlinearity. Its performance is comparable with the other nonlinearities investigated in Ref. [9]. The differential equation for the clipped nonlinearity is

$$\frac{dx_i}{dt} = \begin{cases} -x_i - x_i^3 + \beta \sum_{j=1}^N J_{ij} x_j + \gamma F(t), & |x_i| \leq x_{\text{cutoff}} \\ 0, & |x_i| > x_{\text{cutoff}} \end{cases} \quad (4)$$

We have investigated several values for x_{cutoff} and its effect on the performance of the IM. The exact value does not seem to be pivotal, as long as x_{cutoff} is smaller than the average spin amplitude (see Appendix A). The $-x_i^3$ term is sometimes omitted in the literature for the clipped nonlinearity, but adding or omitting this term seems not to influence the performance. Here we choose to keep this term and take $x_{\text{cutoff}} = 1$.

We simulate the natural evolution of the spin system towards the ground state. To this end, Eqs. (3) and (A1) are

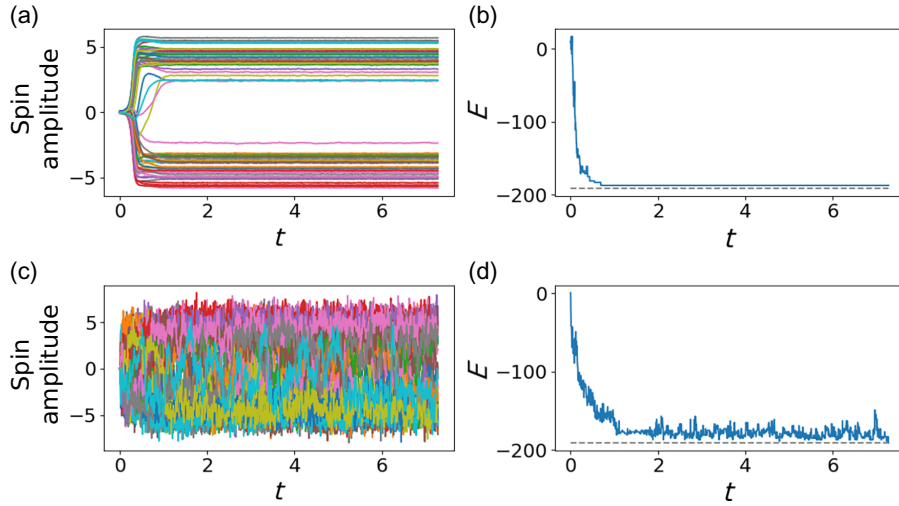


FIG. 1. (a),(c) The spin amplitudes and (b),(d) energy evolution of g05_60.3 for $\beta = 3$ and $\gamma = 0.1$ (a),(b) or $\gamma = 9.4$ (c),(d). Both trials are run for $t = 75$ and $dt = 0.005$. The ground state (grey dotted line; $E = -191$) is reached at $t = 7.285$ for $\gamma = 9.4$ and never for low noise.

solved numerically using the Euler method. The evolution is shown in Fig. 1 for problem g05_60.3 of the BiqMac library [31]. In Fig. 1(a) the time evolution of all the 60 spin amplitudes is demonstrated over time for $\beta = 3$. Starting from their almost zero position, the spin amplitudes will evolve naturally to a clear “up” or “down” position. Notice that there is a substantial difference between the amplitudes of the spins, often referred to as amplitude inhomogeneity. Amplitude inhomogeneity increases the chance of the spins getting stuck in nonoptimal spin configurations [8]. The corresponding evolution of the energy of the system is shown in Fig. 1(b). The energy level quickly drops to a level close to the ground-state energy, here shown by the grey dotted line. In this example, the system never reaches the actual ground-state energy, but is trapped in a local minimum just above it. In Figs. 1(c) and 1(d), the spin amplitudes and energy evolution are shown for a system with more noise. While $\gamma = 0.1$ in Figs. 1(a) and 1(b), it is increased to $\gamma = 9.4$ for (c) and (d). Figures 1(c) and 1(d) demonstrate the effect of large noise, which means that $\gamma \gtrsim \beta$. Now there is no clear distinction between up and down spins in the spin amplitudes’ evolution. While initially the energy also drops quickly, it keeps fluctuating. It finally reaches the ground state at $t = 7.285$. This is a first indication that injecting large noise can help to reach the ground state.

We evaluate the performance of an IM by simulating multiple independent runs. In each run, the parameters are kept fixed and the initial conditions are randomly generated. We use two metrics to quantify performance: the success rate (SR) and either the time-to-solution (TTS) or the time-to-target (TTT). The success rate is defined as the fraction of runs in which the ground-state energy or a pre-defined target energy is reached. When the ground-state

energy is known and success corresponds to finding this optimal solution, we use the TTS. When the ground-state energy is unknown or when success is defined by reaching a target energy, we use the TTT. Both the TTS and TTT are computed as

$$\text{TTS or TTT} = \frac{\sum_i t_i}{N_{\text{success}}} \frac{\log 0.01}{\log(1 - \text{SR})}, \quad (5)$$

Here N_{success} is the number of successful runs. t_i is defined as the simulation time required to reach the ground state or target energy in the i th successful run. One iteration consists of updating all the spin amplitudes once using the spin amplitudes of the previous iteration. For the first iteration, the spin amplitudes are randomly initialized using a uniform distribution with a boundary $|x_i| < 0.1$. For practical reasons, our simulations have a runtime limit. For this, we distinguish between small problems ($N \lesssim 100$) and large problems ($N > 100$). For small problems, the runtime limit consists of 15 000 numerical iterations. Large problems need more time to reach the ground state. Therefore, the runtime limit is 20 000 iterations for large problems. More information about the numerical simulations can be found in Sec. VII.

To optimize the performance of an IM, several schemes are proposed in the literature. Most schemes focus on avoiding becoming trapped in local minima, as this problem drastically increases the TTS and lowers the success rate [9,32]. Here, we will focus on the following three schemes: The first scheme relies on the injection of noise into the system. This perturbs the spin amplitudes, helping the system escape local minima. This occurs because the noise induces spin flips even when doing so is not energetically favorable. This mechanism enables the system to

explore a broader region of the phase space. This scheme is referred to as the noise-only scheme.

In the second scheme, β is annealed linearly, i.e., $\beta = \beta_0 + v t$ where v is the annealing speed. The working of this scheme was partly explored in [25]. Here, it was shown that for some problems annealing always leads to the desired ground state as long as the annealing speed is kept low enough. These problems were therefore labeled “Ising easy” in [25]. There are, however, also problems for which annealing leads to very low success rates, certainly so if the noise strength γ is low. These problems are called “Ising hard” in [25]. Noise is crucial to effectively solve Ising hard problems. Notice that if a problem is Ising easy or hard, depends on the chosen nonlinearity and parameters. For example, it is possible that a problem is Ising hard for the polynomial nonlinearity and Ising easy for the clipped nonlinearity.

The third and last scheme we include in our analysis is chaotic amplitude control (CAC), which was introduced by Leleu *et al.* in Ref. [24]. CAC is one of the most highly performant analog IM schemes. We will use it as a baseline for the other two schemes. The parameters of this scheme will be tuned using Bayesian optimization [33] for $\beta = 1$. Afterwards, we will scan over β using the optimized parameters for $\beta = 1$. This scan will be similar to those we will perform for the other two schemes. As this scheme is inherently chaotic, it is not beneficial to add noise, according to Ref. [34]. Therefore, we will not investigate the influence of noise for this scheme. We will only use it as a baseline when comparing performances.

The influence of noise in both the noise-only and annealing schemes is investigated in Sec. III. In Sec. IV, we compare the performance of the noise-only scheme and the annealing scheme, using CAC as a baseline. Finally, in Sec. V, we focus on the noise-only scheme to derive relationships between the optimal noise strength and basic system parameters. These results are then extrapolated to the annealing scheme in Sec. VC. All this will be done using simulations of MaxCut problems. In MaxCut problems, a graph structure needs to be partitioned in two parts. The optimal solution corresponds to the solution with the maximum cut number $\frac{1}{4} \left(\sum_{ij} J_{ij} - \sum_{ij}^N J_{ij} \sigma_i \sigma_j \right)$. This is related to minimizing the Ising energy. Therefore, this type of problem can be easily implemented on an IM. Maximizing the cut number and minimizing the energy is used interchangeably in the remainder of the paper. We use instances from the BiqMac library and the SuiteSparse Matrix Collection [35]. Furthermore, extra problems are generated using rudy [36]. Rudy is a random graph generator that was also used to generate the MaxCut instances from the BiqMac library and the Gset problems from the SuiteSparse Matrix Collection [31,37]. The coupling weights in the MaxCut problems studied here are discrete and uniformly distributed.

III. INFLUENCE OF NOISE IN THE ANNEALING AND THE NOISE-ONLY SCHEME

Here, we study the effect of Gaussian noise on both the polynomial and clipped nonlinearities for the annealing and the noise-only scheme. Our objective is to maximize the success rate and minimize the TTS. When these two goals conflict, the TTS is prioritized, as we consider it to be the most relevant metric for assessing the efficiency of IMs.

To illustrate the impact of the noise strength, we begin by examining the MaxCut instance g05_60.3 of the Biq Mac library. This problem was selected because it clearly exhibits a relationship between noise and coupling strength, is not Ising easy for the polynomial nonlinearity [25], and is representative for the other problems studied. In Fig. 2, we present parameter sweeps over β , γ , and v as heat maps. The figure covers the polynomial and the clipped nonlinearity and the annealing and the noise-only scheme. The TTS and the success rate for a scan over β and γ using the noise-only scheme and polynomial nonlinearity is shown in Fig. 2(a). The lowest TTSs are found for large γ and large β , which we define here as $\beta > 1$. For example, $\beta = 3$ and $\gamma \approx 9.4$ results in a low TTS and a success rate of 0.99 for the polynomial nonlinearity. This is also demonstrated in Fig. 1(b). Additionally, we see that for each β , there is an optimal noise that maximizes the success rate and minimizes the TTS. When going away from the optimum, the success rate goes to 0. There appears to be a linear relationship between β and the optimal noise for that coupling strength. When following this linear relationship from low β to high β , the success rate and TTS are further optimized, which agrees with our first observation that the lowest TTSs are found for high γ and β . This linear relationship will be investigated in more detail later in Sec. V.

In Fig. 2(b), the heatmap shows a scan over β and γ for the TTS and success rate for the clipped nonlinearity. Again there is an optimal noise per β and the TTS decreases for high β and γ . However, the linear relationship between β and γ does not exist anymore. Here, it seems to be replaced by a square-root behavior, which we will verify later in Sec. V. Due to this different relationship, the noise values that lead to low TTS are not as high as for the polynomial nonlinearity.

Similar results are obtained for annealing in Fig. 2(c) for the polynomial nonlinearity and in Fig. 2(d) for the clipped nonlinearity. Since the coupling strength β is annealed during each run, it cannot be used as a scan parameter. Instead, we scan over γ and the annealing speed $v = (\beta - \beta_0)/t$. We choose the range for the annealing speed such that a parameter region is explored comparable to the region for the noise-only scheme in Fig. 2. The initial coupling strength is set to $\beta_0 = 0$. For the highest annealing speed ($v = 0.1$), β reaches a value of 7.5 within the simulation

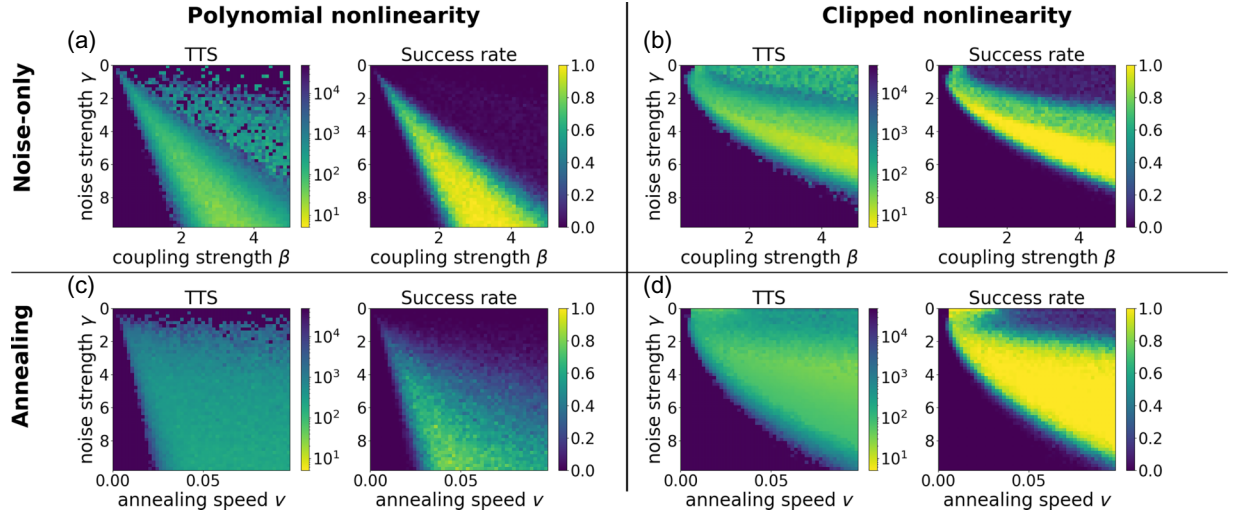


FIG. 2. The success rate and the time-to-solution as a function of γ and β for the noise-only scheme (a),(b) and γ and v for annealing (c),(d) as heat map. The results for the polynomial nonlinearity can be seen at the left (a),(c) and the results for the clipped nonlinearity at the right (b),(d).

time limit. As in the noise-only scheme, we observe that β relates linearly to γ for the polynomial nonlinearity. For the clipped nonlinearity, there is a square-root relationship between β and γ . The key difference between the noise-only scheme and the annealing scheme is that this relationship now describes the optimal noise strength as a function of the annealing speed v , rather than as a function of the coupling strength β .

The existence of an optimal noise level is to be expected. With insufficient noise, spin flips are rare. A lack of spin flips prevents the system from effective exploration of the phase space and causes the system to remain trapped in local minima. Conversely, excessive noise drives the system into a Boltzmann sampling regime [38], which reduces the likelihood of reaching the ground state. Therefore, an intermediate, optimal noise level exists, depending on β . A similar effect is seen when annealing. If the velocity is too low in comparison to the noise, the time limit of our simulations will prevent β from reaching high enough values to escape the region of Boltzmann sampling. An annealing speed that is too large would push the system out of the region where the coupling strength is ideal for the amount of noise. The spin amplitudes will not have the time to reach low energy states before the noise becomes too small to significantly affect them.

In this example shown in Fig. 2, the TTS is significantly lower for the noise-only scheme than for annealing with noise. While the MaxCut problem g05_60.3 is used to showcase the significance of noise in the TTS and success rate, similar behavior is observed in all the investigated MaxCut problems (see Appendix B for other examples). Although not all problems achieve success rates close to 1, almost all exhibit an optimal noise level for each β .

IV. COMPARISON OF DIFFERENT SCHEMES AND NONLINEARITIES

Different IM implementations are modeled by different nonlinearities. Additionally, several schemes are proposed in the literature to improve the efficiency of an IM. In this section, we compare the performance of different schemes and nonlinearities using optimized parameters. Specifically, we compare the annealing scheme and the noise-only scheme to the baseline scheme (i.e., CAC). This comparison is performed for both the polynomial nonlinearity and the clipped nonlinearity. For clarity, we refer to a combination of a scheme and a nonlinearity as a model. From this point forward, we focus on TTS. Even when multiple models achieve near-unity success rates, their TTS can differ substantially.

We compare the TTS for the polynomial nonlinearity and the clipped nonlinearity for all the g05 Biq Mac problems and G1-40 problems of the GSet. For this, we use the best TTS per model of each problem obtained from a parameter scan similar to the one shown in Fig. 2. The ratio between the best TTS for each model and the best TTS across all models ($\text{TTS}_{\text{model}}/\text{TTS}_{\text{best}}$) is shown for each g05 problem in Fig. 3 on a logarithmic scale. We compare the relative TTS for five models: noise-only scheme (red) and annealing scheme (blue) with the polynomial nonlinearity and the clipped nonlinearity, and CAC (green). To provide a sense of absolute performance, the best TTS value for each problem is displayed above the bars in Fig. 3. We can count the number of times each model yields the best TTS for the 30 g05 BiqMac problems. The noise-only scheme always outperforms annealing for the polynomial as well as the clipped nonlinearity. The noise-only scheme with the clipped nonlinearity delivers

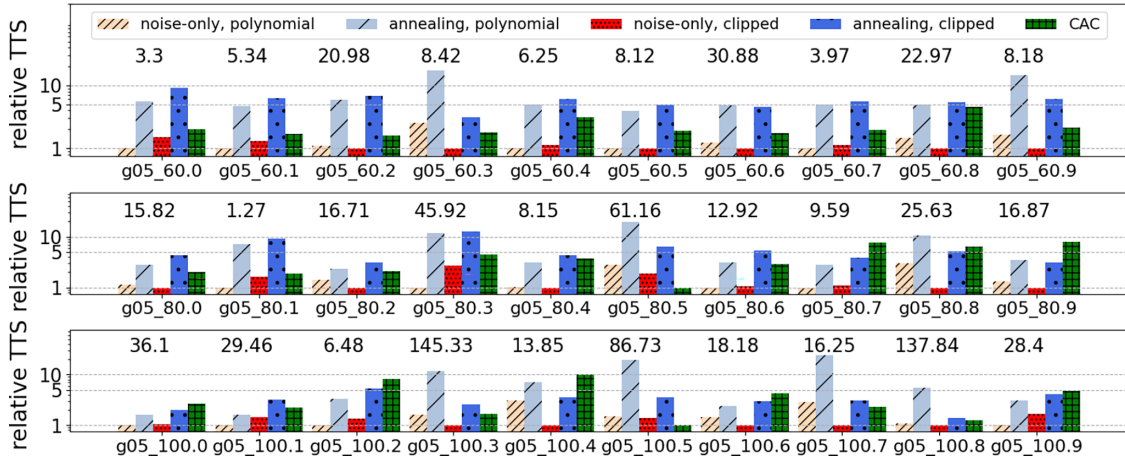


FIG. 3. Time-to-solution of the noise-only scheme and the annealing scheme for the polynomial and clipped nonlinearity and CAC relative to the best TTS of these models. The TTS of the best model is mentioned above the bars to give an idea of the timescale. All the problems are from the BiqMac library.

for 16 out of the 30 benchmark problems the best TTS. So this model performs the best for the g05 BiqMac problems. The second-best performing model is the noise-only scheme with the polynomial nonlinearity. This model delivers the best TTS for 12 out of the 30 benchmark problems. The two last best values for the TTS are obtained by CAC.

The best TTS is provided 16 times by the clipped nonlinearity, 12 times by the polynomial nonlinearity, and 2 times by CAC. Thus, CAC rarely achieves a better TTS for small problems ($N \lesssim 100$) such as the BiqMac problems.

The relative best TTT of the noise-only scheme and the annealing scheme for the polynomial and clipped nonlinearity and of CAC for the Gset are shown in Fig. 4. The different models are displayed with the same colors and

patterns as in Fig. 3. Due to the difficulty level of these problems and the absence of a known ground state, a run is considered successful if it reaches 99.5% of the highest cut value mentioned in the literature [39]. There are five problems omitted (G32, G33, G34, G39, G40), as neither CAC, nor any of the other models reaches 99.5% of the best cut value in the given simulation time. In Fig. 4, we see that the best TTS is often given by CAC: it outperformed the other models in 28 of the 35 problems. However, the other models and specifically annealing with the clipped nonlinearity performs very similar to CAC: only in 8 of 35 cases is the TTT of the second-best model 5 times larger than that of CAC. When comparing the noise-only scheme with the annealing scheme, we see that annealing outperforms the noise-only scheme in 32 of the 35 problems.

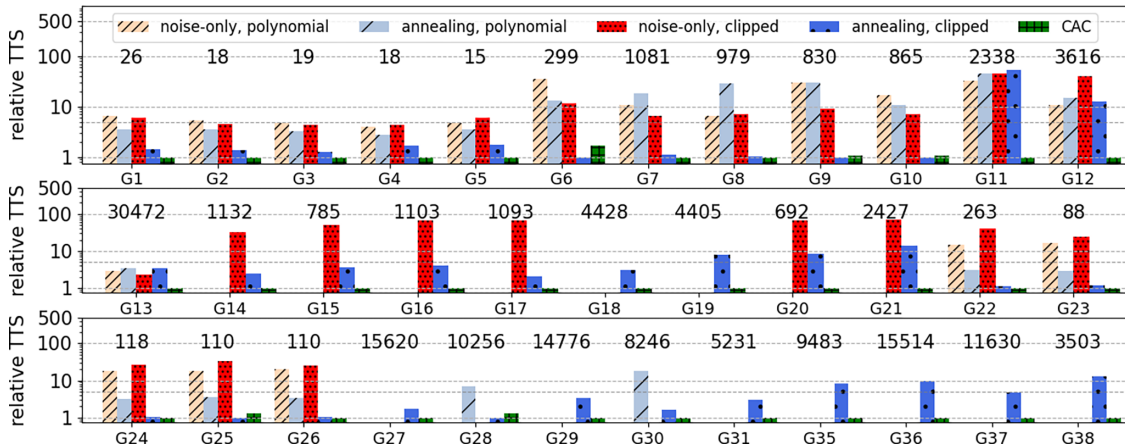


FIG. 4. TTT of the polynomial and clipped nonlinearity for the noise-only scheme and the annealing scheme and of CAC relative to the best TTT of these models. The TTT of the best model is mentioned above the bars. When no bar is visible, 99.5% of the best-known cut value is never reached by that model. All the problems are from the Gset.

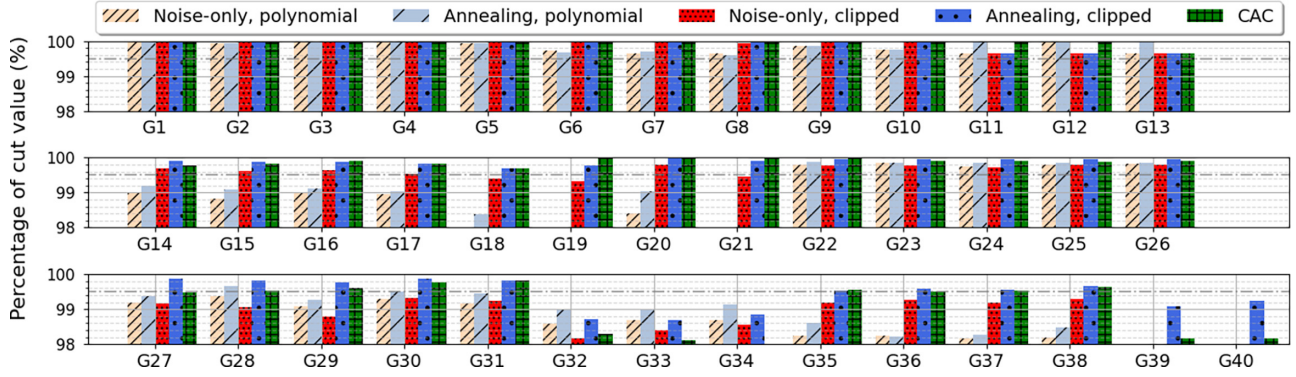


FIG. 5. The highest percentage of the best-known cut value that is reached with annealing, the noise-only scheme and CAC for the Gset. When the bar reaches 100%, the highest found value and the best-known value are equal.

Specifically annealing with the clipped nonlinearity gives the best results, which are often in the same order as CAC.

As the TTT is based on 99.5% of the best-known cut value, it is also interesting to examine the cut value. For each problem, we determine the highest cut value found within the time limit per run of $t = 75$ during the parameter scan or Bayesian optimization. In Fig. 5, it is displayed which percentage of the best-known cut value is found for each model. Notice that the parameters yielding the highest cut value need not be identical to the parameters producing the lowest TTT. A bar reaching 100% indicates that the model achieved the best-known solution. CAC often reaches the best-known solution and reaches in general higher cut values than the other models. Annealing with the clipped nonlinearity performs almost equally good as CAC. Annealing mostly outperforms the noise-only scheme: annealing finds a lower energy than the noise-only scheme in 37 out of the 40 problems. For 33 problems, this energy was found by the clipped nonlinearity.

Comparing the results for the smaller BiqMac problems with those for the Gset reveals some contrasting trends. For the BiqMac problems, the noise-only scheme always results in better TTS, but annealing reaches higher cut values and lower TTTs for the Gset problems. This indicates that annealing with large noise is a better scheme for larger and more difficult problem sets, such as Gset, whereas the noise-only scheme performs better for simple tasks. The comparison with CAC shows that injecting large noise seems to speed up the process of finding the target. For small problems or when it is not pivotal to find the lowest possible energy value, injecting noise is faster than CAC. For larger problems, annealing with the clipped nonlinearity obtains similar efficiencies as CAC. So, large noise significantly improves the performance of previously slower nonlinearities, making them competitive with fast and promising approaches such as CAC.

V. RELATION BETWEEN THE COUPLING STRENGTH/THE ANNEALING SPEED AND THE NOISE STRENGTH

In the previous section, we have compared several nonlinearities and schemes, using the minimum TTS (or TTT) obtained from parameter scans. We demonstrated that with optimal noise, these nonlinearities are competitive with CAC. Now we investigate which parameter values lead to the best TTS. We are particularly interested in the relation between the optimal noise strength γ_{opt} and the coupling strength β . From Fig. 2, we noticed a roughly linear relationship between the noise and coupling strength for the polynomial nonlinearity and a square-root dependence for the clipped nonlinearity. From here on, we will focus on the noise-only scheme, as the relationship between γ_{opt} and β looks clearer for the noise-only scheme than for annealing in all the observed problems, as can be seen in Fig. 2 and in Figs. 14–16 in Appendix B. This is because the area of low TTS is smaller and more delineated in the noise-only scheme. Additionally, the coupling strength β and the annealing speed v are related. Therefore, it will be possible to derive the relationship between the noise strength γ_{opt} and the annealing speed v from the relationship between the noise strength γ_{opt} and the coupling strength β . It is therefore sufficient to investigate the relationship between γ_{opt} and β . We will look at small ($N \lesssim 100$) and large problems ($N > 100$) separately, as they have a different definition of success. For large problems, the best-known solution is never reached or only a few times, so it is not possible to analyze it. Therefore, the success rate is based on a target energy, which is a bit higher than the actual ground-state energy. For the small problems, this is no issue. This difference in the definition in success rate will influence the relation between the coupling strength and the noise strength.

A. Small problems

First, we will consider small problems ($N \lesssim 100$), starting with our illustrative problem g05_60.3 from the BiqMac library. For this problem, the relationship between β and γ_{opt} is shown in Fig. 6(a) for the polynomial nonlinearity and in Fig. 6(b) for the clipped nonlinearity. We determined γ_{opt} for each β using parameter scans similar to Fig. 2. To minimize the influence of randomness on the relationship, we have repeated the whole parameter scan 10 times. The optimal noise value is determined for both the TTS (blue) and the success rate (red). For the TTS, there is always one minimal TTS, but for the success rate, multiple values of the noise strength yield the best success rate. This can also be seen in Fig. 2, where the success rate maximizes to one for multiple parameter values. We always choose the lowest noise strength for which the best success rate is reached as the optimal noise value. The standard deviation of the γ_{opt} values is displayed as a color spread in Fig. 6, to give an idea of the differences between the repeated parameter scans. We not only show the best noise value per coupling strength, but also the corresponding TTS and success rate. As the noise strength is only scanned in the interval $[2, 10]$, the linear relationship in Fig. 6 flattens when the maximum noise is reached.

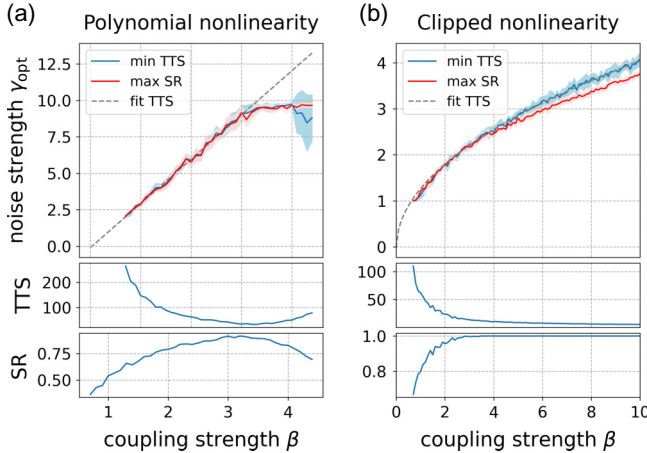


FIG. 6. The optimal noise strength as a function of the coupling strength β for the g05_60.3 problem. The noise strength that leads to the maximal success rate is in red and to the minimal TTS in blue. This is shown for the polynomial nonlinearity (a) or clipped nonlinearity (b). The minimal TTS and maximal success rate are added as two small figures under the main figure. The standard deviation (shown as light-colored areas) of the γ_{opt} values is determined by repeating the whole parameter scan 10 times. For each parameter scan, the noise value leading to the maximal success rate and the minimal TTS are determined. The relationship is fitted with a line through the origin for the polynomial nonlinearity and with a square root for the clipped nonlinearity. Due to the scan limits, this relationship can fail for too high or low coupling strength values.

The best TTS is fitted with a straight line through the origin ($\gamma_{\text{opt}} = A\beta$) for the polynomial nonlinearity and with a square root ($\gamma_{\text{opt}} = B\sqrt{\beta}$) for the clipped nonlinearity. When β increases, we see that the TTS decreases and success rate increases for both nonlinearities. So the best TTS values consistently occur at high noise and coupling strengths. For $\beta \lesssim 2$, the TTS changes several orders of magnitude when changing β . After $\beta \approx 2$, the decline in TTS slows down. While further increases in β and thus also in γ_{opt} are still beneficial, the variations in the TTS will not exceed an order of magnitude. It is thus possible to find better values for the TTS in Fig. 3 by expanding the range to higher β and γ_{opt} . However, the improvement will be less than an order of magnitude. Therefore, we will typically set $\beta \approx 2$ when we are searching for the best TTS without extensive parameter scans.

We are interested in the slope A of the green-dotted fit in Fig. 6 and we investigate if this slope is determined by some easy determinable problem characteristics, such as N or the number of connections. We assume that for the noise to make an impact on the spin amplitude, it has to have a magnitude similar to the size of the rest of the polynomial nonlinear transfer function. As both are on average zero, we look at the standard deviation.

$$\text{std} \left(-x_i - x_i^3 + \beta \sum_j J_{ij} x_j \right) \sim \text{std}(\gamma_{\text{opt}} F(t)). \quad (6)$$

We assume that $|-x_i - x_i^3| \ll |\beta \sum_j J_{ij} x_j|$, which corresponds to high β and/or a high number of connections per spin (resulting in a large sum). Furthermore, we enforce amplitude homogeneity ($\forall i \in \{0, 1, \dots, n-1\} : x = |x_i|$). To simplify the notation, we contract the sign of $J_{ij} x_j$ in s_j . This results in

$$\beta \text{std} \left(x \sum_{j, J_{ij} \neq 0} |J_{ij}| s_j \right) \sim \text{std}(\gamma_{\text{opt}} F(t)). \quad (7)$$

We notice that $|J_{ij}|$ and s_j are independent. Thus

$$\beta x \text{std} \left(\sum_{j, J_{ij} \neq 0} s_j \right) \text{std}(|J_{ij}|) \sim \text{std}(\gamma_{\text{opt}} F(t)). \quad (8)$$

When simulating problems, we observe that the spins have approximately a 50% chance to be “up” and 50% to be “down.” So s_j is approximately binomial distributed (see Appendix C), and we can approximate the standard deviation as $\text{std}(\sum_{j, J_{ij} \neq 0} s_j) = \sqrt{C}$, where C is the average number of connections per spin. Furthermore, $|J_{ij}|$ follows a discrete uniform distribution $\mathcal{U}\{0, J_{\text{max}}\}$, where J_{max} is the maximum value of J_{ij} . Thus $\text{std}(|J_{ij}|) =$

$\sqrt{((J_{\max} + 1)^2 - 1)/12}$. This leads us to the formula

$$\gamma_{\text{opt}} \sim \beta \sqrt{C \frac{(J_{\max} + 1)^2 - 1}{12}} x. \quad (9)$$

When each spin is connected to more than half the other spins ($C > N/2$) and the connections are purely anti-ferromagnetic, there are for sure connected spins, which will have the same direction. More specific, each spin is connected with at least $C - N/2$ spins, which have the same direction. This means that $C - N/2$ spins have a negative impact on the sum $\sum_j s_j$ and another $C - N/2$ connected spins are needed to neutralize them, making the sum 0 again. This impacts the derived formula (9), as now $\text{std}(\sum_{j, J_{ij} \neq 0} s_j) = \sqrt{N - C}$ and thus

$$\gamma_{\text{opt}} \sim \beta \sqrt{(N - C) \frac{(J_{\max} + 1)^2 - 1}{12}} x.$$

We have derived that the slope $A = \gamma_{\text{opt}}/\beta$ depends on the number of connections per spin and the standard deviation of the coupling matrix elements. However, it should be noted that the assumptions, such as amplitude homogeneity, only hold approximately in practice. Therefore, it is necessary to verify this relationship with numerical simulations. To this end, we analyze several MaxCut problems with varying connectivity and coupling matrix properties. For each problem, there is a linear relationship between the optimal noise and coupling strength when using the polynomial nonlinearity, as illustrated in Fig. 6. We determine the slope $A = \gamma_{\text{opt}}/\beta$ by performing a linear fit to this relationship. In Fig. 7, we show the resulting slope A as a function of the number of connections per spin C and the maximum value of the coupling element $J_{\max}$. We simulate several problems for each combination of C , N , and type of coupling matrix. We call problems similar when they have the same value for C , N , $J_{\max}$, and the same distribution of the coupling matrix elements. Each data point represents the mean slope of a group of similar problems. The error bar indicates the standard deviation of the slope. In Fig. 7, all the problems have 100 or less spins. We divided the problems in four groups: the BiqMac problems with the uniform distribution $\mathcal{U}\{-J_{\max}, J_{\max}\}$ as distribution of the coupling matrix elements (blue), the BiqMac problems with $\mathcal{U}\{0, J_{\max}\}$ as distribution (red), the self-generated problems with $\mathcal{U}\{-J_{\max}, J_{\max}\}$ as distribution (green), and the self-generated problems with $N = 100$, $C > N/2$, and $\mathcal{U}\{0, 1\}$ as the distribution of the coupling matrix elements (yellow). Notice that for the last group of problems, the slope A depends of $N - C$ rather than C , reflecting a slightly different scaling behavior.

The slope A is not always determined in the same way. For all the problems, the relationship between β and γ_{opt} is fitted on the basis of the minimum TTS or the maximum success rate. Mostly, the fit parameters for these two

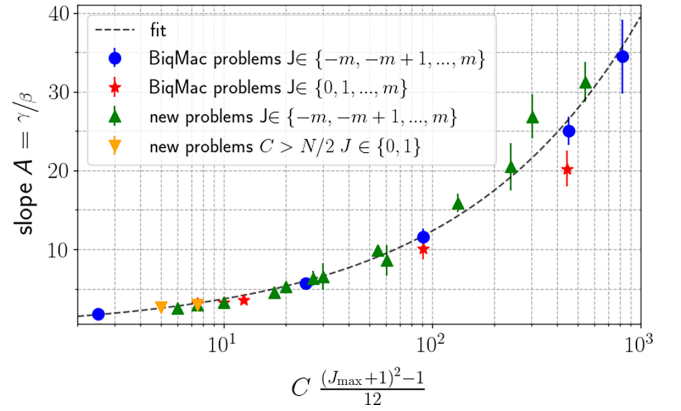


FIG. 7. A is the slope of the linear fit relating the optimal noise strength γ_{opt} to the coupling strength β . Here A is shown as a function of the average number of connections per spin C and maximum coupling element $J_{\max}$. The four groups of problems are as follows: BiqMac problems with mixed interactions (blue dots), antiferromagnetic BiqMac problems (red stars), self-generated problems with mixed interactions (green triangles), and self-generated problems with predominantly antiferromagnetic connections (yellow triangles).

were very similar. In extreme cases, such as very easy or very hard problems, the TTS and success rate lead to different relationships. For easy problems, we used the minimum TTS to fit and for difficult problems, we base the fit on the maximum success rate. In Fig. 7, we also show a fit of the slope A as a function of C and $J_{\max}$ using Eq. (9), so $A = \gamma_{\text{opt}}/\beta \sim c_1 \sqrt{C((J_{\max} + 1)^2 - 1)/12} + c_2$ where c_1, c_2 are fit parameters. While the exact coupling matrix is required to accurately deduce the optimal noise level, we see that the number of connections per spin and the distribution of J_{ij} predicts the ratio $\gamma_{\text{opt}}/\beta$ fairly well. In Fig. 7, $c_1 = 1.26 \pm 0.04$ and $c_2 = 0.29 \pm 0.12$ with a R^2 of 0.96. Hence, this scaling law can be used to predict the best noise value for a chosen β in random problems without local fields.

B. Large problems

We now investigate MaxCut problems between 200 and 2000 spins with densities 0.5, 0.1, 0.06, and 0.01. We generated them using the method `rnd_graph` of `rudy`. We estimate the ground-state energy relying on a Bayes optimization of a CAC IM [24]. These problems, together with the random problems in Gset, namely G1-G10 and G22-G31, are used to qualify the influence of the number of connections on the slope $A = \gamma_{\text{opt}}/\beta$ in Fig. 8(a). The similar Gset problems are grouped together and have an error bar to show the variation between the problems. Here $J_{\max}$ is always kept 1, as these larger problems are already more difficult to solve and varying the distribution of J_{ij} would further increase the difficulty level. The slopes A are fitted using $A = \gamma_{\text{opt}}/\beta = c_1 0.25 \sqrt{C} + c_2$, leading

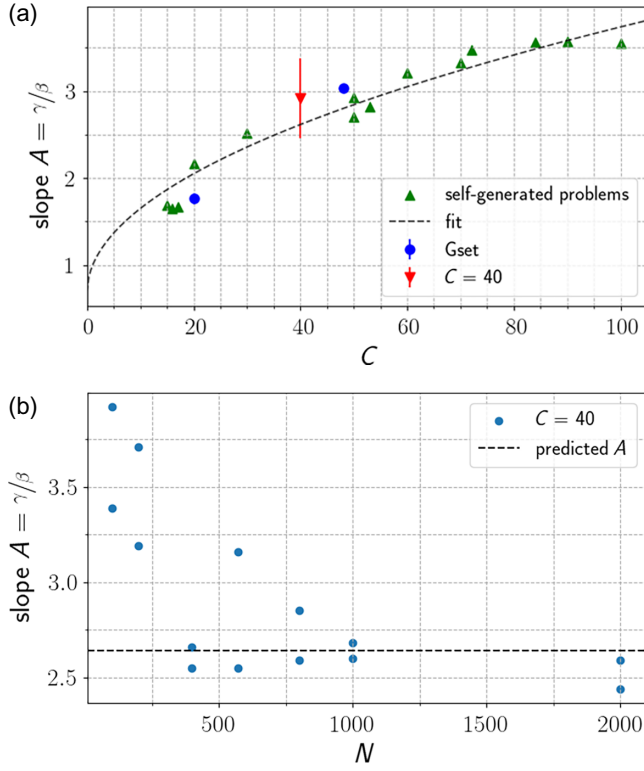


FIG. 8. (a) A is the slope of the linear fit relating the optimal noise strength γ_{opt} to the coupling strength β . Here A is shown as a function of the average number of connections per spin C . The Gset problems with the same C , N and distribution of J are averaged and shown with an error bar. The average slope of self-generated $C = 40$ problems of subfigure (b) is also included (red triangle). All the other self-generated problems (green triangles) have different values for C or N , so these data points are not averaged. The data are fitted using Eq. (9), with $J_{\text{max}} = 1$. (b) The slope A as a function of the number of spins N . The predicted slope using the fit in subfigure (a) is shown as a grey dotted line.

to $c_1 = 1.22 \pm 0.17$ and $c_2 = 0.2 \pm 0.3$ with a R -squared value of 0.76.

As most problems are newly generated, we cannot rely on an established best-known cut value. Moreover, since we consider large problem instances and we first must determine the best cut value, the simulations are more time consuming. Due to computation time constraints, each data point in Fig. 8(a) corresponds to a single problem instance. In contrast, the data points in Fig. 7 represent averages over at least three similar problems. These two issues likely explain why there is more variation in Fig. 8(a) and the fit quality is lower compared to Fig. 7. Including more problems and determining the highest cut value more accurately would likely make the relationship more apparent and improve the R -squared value.

We also want to further verify that only the average number of connections per spin C and the distribution of J_{ij} influences the relationship between the noise strength

γ_{opt} and the coupling strength β . We will investigate the influence of N on this relationship for a certain connectivity. In Fig. 8(b), the slope A of self-generated problems with $C = 40$ and $J_{\text{max}} = 1$ but different problem sizes are shown as a function of N . As a reference, we also displayed their predicted slope A calculated using the fit in Fig. 8(a) as a black dotted line. For large problems, the size of the problem does not seem to influence the slope A . For smaller problems, there is more variation in A and it is less clear if there is no link between N and A . This variation on A is quite large in comparison to the Gset problems, as can be seen in Fig. 8(a). There are several explanations for this. First of all, smaller problems are more prone to statistic variations. Additionally, the variation can be due to how well the ground-state energy is known. Since the Gset problems are often used, the best-known cut value probably approaches the optimum cut value quite well. The highest cut value of the self-generated problems was only estimated by us using CAC. We often encountered higher cut values during the simulations with all models.

C. Predicting the best parameter set

Both the small and the large problems follow the theoretical relationship (9). Here, we investigate whether the ideal noise value can be predicted using the fits from Fig. 7 and Fig. 8(a). We first focus on small problems, where the success rate is based on the ground-state energy. To illustrate this, we first look at two specific problems: g05_60.3 and w09_100.3 from the BiqMac library. We use the noise-only scheme with the polynomial nonlinearity. We keep β fixed at 2, and then predict the optimal noise strength using the fit from Fig. 7. Figure 9 shows the TTS as a function of the noise strength for $\beta = 2$. The predicted optimal noise for each problem is shown as a blue dot. While the predicted optimal noise does not always yield the best TTS or success rate, the values of both metrics at this noise level are fairly close to their optimum for $\beta = 2$.

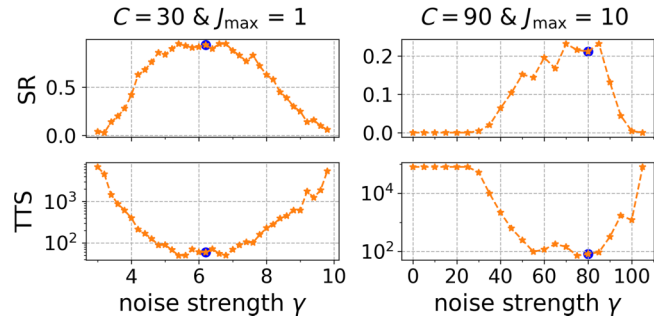


FIG. 9. The TTS and success rate as a function of the noise strength γ . We demonstrate this for the problems g05_60.3 ($C = 30$ & $J_{\text{max}} = 1$) and w09_100.3 ($C = 90$ & $J_{\text{max}} = 10$) for $\beta = 2$. The predicted optimal noise strengths for $\beta = 2$ using the fit in Fig. 7 are shown as blue dots.

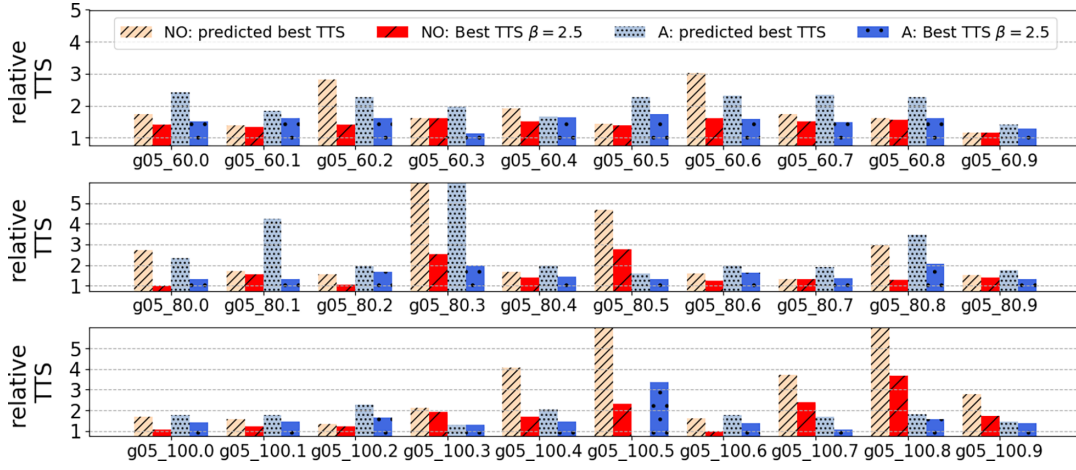


FIG. 10. The TTS using the predicted noise value for $\beta = 2.5$ and the best TTS for $\beta = 2.5$ relative to the best TTS using the polynomial nonlinearity for BiqMac problems. The two left red striped bars are for the noise-only scheme (abbreviated as NO in the legend) and the two right blue dotted bars represent annealing (abbreviated as A in the legend). For both annealing and noise, the TTSs are relative to the best TTS for that problem and nonlinearity. As we want the predicted TTS to be as good as possible, we want our relative TTS to be close to 1.

We now turn from illustrative examples to a systematic evaluation over all g05 BiqMac problems. We assess the accuracy of the predicted noise strength for the noise-only scheme and the annealing scheme. For the noise-only scheme, we choose $\beta = 2.5$ as this is already large enough to improve the TTS drastically and is not too big to lead to numerical instabilities in the simulations. Nonetheless, choosing β a bit higher or lower would provide similar results. The corresponding annealing speed v is obtained from the relation between the annealing speed and the coupling strength: $v = \beta/t$. Since the time required to reach the ground state varies between runs, we use an estimated average time t_A . Here, we set $t_A = 50$ for all problems, although any value between 25–75 is adequate.

To examine the influence of our choice for the coupling strength β or the annealing speed v , we compare three TTS values. First, we consider the TTS obtained using the predicted noise strength at $\beta = 2.5$ (or at $v = 0.05$

for the annealing scheme), denoted $\text{TTS}_{\text{predict}}$. Second, we consider the optimal TTS over all noise values for fixed $\beta = 2.5$ (or for fixed $v = 0.05$ in the annealing case), denoted $\text{TTS}_{\text{fixed}}$. Finally, we consider the optimal TTS over all noise and coupling strengths (or annealing speeds), denoted TTS_{best} . We use the TTS_{best} as the baseline and compare the ratios of the other TTS values to it. In Fig. 10, we display $\text{TTS}_{\text{predict}}/\text{TTS}_{\text{best}}$ and $\text{TTS}_{\text{fixed}}/\text{TTS}_{\text{best}}$ for each g05 BiqMac problem for the noise-only scheme and the annealing scheme. Using the predicted noise value results in TTS values, which are about twice TTS_{best} . The difference with $\text{TTS}_{\text{fixed}}$ is even smaller, indicating that our prediction of the best noise for a certain β value works fine. Nonetheless, there are some instances where the TTS_{best} is significantly lower than the $\text{TTS}_{\text{predict}}$. For example, the ground state is never reached for problem g05_100.5 during annealing using the predicted noise value. Due to the overall low success rate for this problem, the TTS does

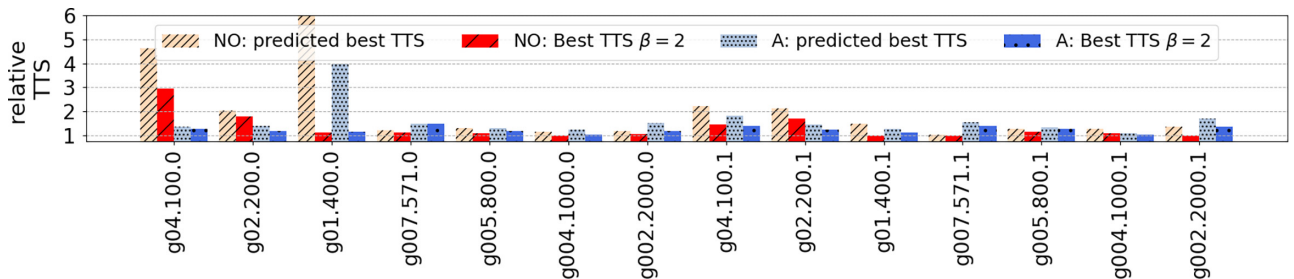


FIG. 11. The TTT using the predicted noise value for $\beta = 2$ and the best TTT for $\beta = 2$ relative to the best TTT using the polynomial nonlinearity for the $C = 40$ problems. The two left red striped bars are for the noise-only scheme and the two right blue dotted bars represent annealing. For both annealing and noise, the TTT is relative to the best TTT for that problem and nonlinearity. As we want the predicted TTT to be as good as possible, we want our relative TTT to be close to 1.

not change smoothly anymore, making it harder to predict. The same occurs for the problem g05_80.3. For problem g05_100.8, the TTS_{predict} is only a factor of 2 higher than TTS_{fixed} , but the TTS can be much improved by choosing a higher β value.

The same comparison is done in Fig. 11 for the $C = 40$ problem from Fig. 8(b) to verify the fit for the large problems. Here we take β fixed at a value of 2 and use the fit in Fig. 8(a) to predict γ_{opt} . (Remark that we choose a different β value, as the predicted noise values for higher β values fell outside of the scanned noise range.) For annealing, the corresponding annealing speed is then $v = 0.04$. Similar to the BiqMac problems, we find that using the fitted relationship allows us to identify a decent parameter set, which gives TTT_{predict} values that are very close to the best TTT (TTT_{best}). For annealing, the TTT_{predict} is never larger than twice the TTT_{best} . This also indicates that our choice of β was sufficiently good. In general, a larger coupling strength leads to a better TTT, but the choice of β is often constrained by some numerical or physical limitations. Choosing a large β , that is experimentally or numerically feasible, almost always produces good results. Thus, despite wide variation on the slope A in Fig. 8(b), predicting the best parameter set using the slope of the fit together with a large β yields consistently good outcomes.

VI. DISCUSSION AND CONCLUSION

We have investigated the significance of injecting noise in an IM when it evolves to the ground state. It is evident that choosing appropriate noise and coupling strengths can significantly improve the TTS and success rate. By comparing the annealing scheme and the noise-only scheme to CAC, we observe that adding large noise could make these models competitive. While the noise-only scheme performs faster for smaller problems, annealing with large noise proves more effective for larger problems.

The noise strength can be selected based on the connectivity and the distribution of the coupling matrix elements. This eliminates the need to scan over noise values. For estimating the noise strength, we use the fits of the theoretical correlation (9). We verified the predicted noise strength for the Biq Mac MaxCut problems in Fig. 10 and the self-generated $C = 40$ problems in Fig. 11. We observe that it is possible to predict a good noise value using the fitting parameters in Fig. 7 and in Fig. 8(a). Furthermore, just choosing β high enough ($\beta > 2$) almost always leads to good TTS. The specific coupling matrix still influences the optimal noise strength. However, a rough estimation using the connectivity and the coupling matrix distribution generally yields a TTS (or TTT), which differs from the best TTS (or TTT) by no more than a factor of 2.

Our analysis has focused on the relationship between the noise strength γ_{opt} and coupling strength β , ignoring

the annealing cases. Nonetheless, annealing is faster and has a higher success rate for more difficult problems. We choose to focus on the relationship between γ_{opt} and β as the relation between the annealing speed v and the noise strength γ_{opt} can be derived from it as $v = \beta/t$. As the time to reach the ground state varies across runs, we used an estimated average time t_A . We found that the exact value is not pivotal, as long as it roughly approximates the time needed to reach the ground state. In fact, there is always a wider spread in the good parameter values for annealing, as can already be seen in Fig. 2. This phenomenon is even more pronounced for larger problems, such as the Gset, where annealing results in lower TTT than the noise-only scheme.

Our results are based on numerical simulations, but this study also provides a guideline to improve experimental setups. A key requirement for testing our predictions experimentally is the ability to control the noise strength. There are several ways to inject a controlled amount of noise in physical systems. The exact implementation depends on the experimental platform and hardware components. Consequently, it is not possible to prescribe a single, universal approach. Instead, we will discuss a few possibilities for several experimental setups. In many setups, an FPGA is used to efficiently compute the coupling terms. Noise can be generated internally on the FPGA, using a pseudorandom number generator. Alternatively, the noise can be generated externally and injected into the system. In Ref. [38], the authors use a broadband noise signal generated by an electrical analog noise source and amplified by an electronic amplifier. In other setups, the noise strength can be tuned by controlling intrinsic noise sources [20,29], for example, by adjusting the camera exposure time.

In this study, the focus is on the polynomial nonlinearity and the clipped nonlinearity, even when it is already shown that other nonlinearities, in particular the sigmoid nonlinearity, work better. We restricted our study to nonlinearities that could be explored using a single parameter, but we suspect that injecting noise is also beneficial for other nonlinearities. Especially, we expect the behavior of the sigmoid nonlinearity to be similar to that of the clipped nonlinearity.

Furthermore, it would be interesting to explore coupling matrices whose elements are drawn from distributions other than discrete uniform ones. We conjecture that using the standard deviation of J_{ij} will also work for different distributions of the coupling matrix elements, but testing this conjecture will be done in future work. We also remark here that due to computational limits, the role of J_{max} was only checked for small problems. Nevertheless, it is highly likely that J_{max} influences the optimal noise strength similarly for larger problems. So the fitted relationships can be used to predict the optimal noise value for different types of problems and different problem sizes, eliminating the need for scans of the noise strength. This makes it

possible to solve alternative problems efficiently without prior extensive parameter scans.

VII. METHODS

All results are obtained from numerical simulations of Eqs. (3) and (A1). The differential equations are solved using a first-order time discretization, also known as the Euler method. The three schemes are implemented in c++, to allow for a consistent comparison. To keep the simulations computationally feasible, we set a maximum of 15 000 iterations per run for the small problems and 20 000 iterations for the large problems. We take a time step of $\Delta t = 0.005$. We verified that this time step is sufficiently small such that it does not influence the results. For all problems, we computed the transient success rate by repeating each run multiple times: 100 runs for the g05 problems, 200–300 runs for the self-generated problems, and 500 runs for the Gset. Larger problems typically have lower success rates and therefore require more repetitions to obtain reliable statistics. Noise is injected at every time step as explained in Ref. [40].

To quantify the processing speed of an IM, we use the time-to-solution. The TTS estimates the average time required to find the ground-state energy with 99% confidence. Several definitions of TTS exist in the literature. In the following, we introduce two commonly used definitions. A widely used definition is [34,41]

$$\text{TTS} = \min_{t_R} t_R \frac{\log(0.01)}{\log(1 - \text{SR})}, \quad (10)$$

where t_R is the run time of a single attempt and SR is the success rate. The success rate is defined as the probability of reaching a predefined target energy in a single run of duration t_R . In the case of the TTS, this target corresponds to the ground-state energy. So, the success rate is defined as $\text{SR} = N_{\text{success}}/N$, where N is the total number of independent runs and N_{success} is the number of runs in which the target is reached. The run time t_R is optimized to obtain the minimum TTS.

Another commonly used definition is employed in Refs. [9,24]. In this approach, each simulation is stopped once the ground state is reached. The TTS is then defined as

$$\text{TTS} = \frac{\sum_i t_i}{N_{\text{success}}} \frac{\log(0.01)}{\log(1 - \text{SR})}, \quad (11)$$

where t_i is the simulation time required to reach the ground state in the i th successful run. To facilitate comparison of results from different implementations, we use a re-normalized time variable t . This removes the explicit dependence on the system bandwidth τ , which varies across physical implementations. The renormalized time is defined as $t = t'/\tau$, where t' denotes the physical time.

Since the time is renormalized, the TTS is dimensionless. The physical time-to-solution of a specific Ising machine setup can be obtained by multiplying the dimensionless TTS from this paper by the characteristic bandwidth of the physical IM implementation.

In this work, we use the second definition. It avoids an additional optimization over the run time and is therefore less computationally expensive, which is key for large numerical studies. To verify that this choice does not affect our conclusions, we compared both definitions by examining the behavior of the TTS in parameter scans for several benchmark problems. In particular, we performed scans over the coupling strength and the noise strength, such as those shown in Fig. 2. Both definitions show the same qualitative behavior across these scans, and the resulting TTS values are of the same order of magnitude.

Although these definitions will result in slightly different TTS, the conclusions will not change. We compared for several problems the TTS for both definitions and the TTS behaved similarly. In general, the TTS using Eq. (10) is the double of the TTS using Eq. (11). The relative TTS were the same independent on the choice of definition of TTS. As we are only comparing TTS in this study and a TTS also differs slightly between simulations, the small difference between these definitions is negligible and unimportant in this study.

When the target energy corresponds to a fixed percentage of the best-known cut value rather than the true best cut value, we use the TTT. The TTT is computed in the same way as the TTS but is used to clearly distinguish approximate optimization from a ground-state search. In this work, we use TTS for problems where the optimal solution is known, such as the small MaxCut problems from the BiqMac library. We use TTT for larger problems, such as the Gset instances, where the optimal solution is unknown and performance is evaluated relative to a target energy. The best TTS or TTT obtained with CAC is determined using Bayesian optimization with the BayesOpt library [33], using 40 initial samples and 120 iterations. Table I lists the parameters that are optimized and their boundaries. More information about the parameters can be found in Ref. [34]. We use the same symbols and equations for CAC as in Ref. [34]. For the Gset problems, the coupling strength β is included as an optimization parameter. For the BiqMac problems, β is not optimized using Bayesian optimization, but an additional sweep over β is performed for 50 values between 0 and 5. The best TTS/TTT for the noise-only scheme and the annealing scheme are determined by a parameter scan of 50×50 (BiqMac) or 25×25 (Gset and self-generated problems) over noise and coupling strength.

We consider MaxCut problems from three different sources: the BiqMac library [31], the Gset problems from the SuiteSparse library [35], and additional self-generated problems created using the random generator rudy. For the BiqMac problems, the exact ground-state energies are

TABLE I. The parameters of CAC that are optimized using Bayesian optimization, together with the preset ranges.

Symbol	Meaning	Min	Max
α	Target amplitude baseline	0.5	5
δ	Sensitivity to energy variations	1	20
ρ	Amplitude and gain variation	0	10
ξ_0	Rate of change of error variables for $t = 0$	0	0.01
Γ	Rate of increase of ξ	0	0.01
n_x	Number of iterations for nonlinear terms	1	20
n_e	Number of iterations for error terms	1	20
Δt	Max number of iterations before ξ resets	10	1000
β	Coupling strength	0.1	5

known and provided with the dataset. These values are used to compute the success rate and the time-to-solution. For the Gset problems, the exact ground-state energies are unknown. In this case, we use the best-known solutions reported in Ref. [39] as target energies.

For the self-generated problems, the ground-state energies are estimated using CAC. We apply Bayesian optimization with the BayesOpt library to minimize the energy, and we use the lowest energy found as the target value. Exact solvers are not used in this work because they are computationally prohibitive for large problem sizes. For consistency, we determine the ground-state energies in the same way for small and large problems. So we do not use exact solvers for small problems, even though exact solvers would give the exact solution with certainty. We claim that CAC almost always gives the ground-state energy of small problems for the following two reasons. For small benchmark problems with known solutions, CAC consistently reaches the exact ground state. In addition, for the small self-generated problems, the solutions found by CAC were never improved by the other schemes

considered in this study. This supports the appropriateness of using CAC as a ground-state estimator for small self-generated instances.

ACKNOWLEDGMENTS

This research was funded by the Research Foundation Flanders (FWO) under Grants No. G0A6L25N and No. G006020N. Additional funding was provided by the EOS project “Photonic Ising Machines.” This project (EOS No. 40007536) has received funding from the FWO and F.R.S.-FNRS under the Excellence of Science (EOS) programme. The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation Flanders (FWO) and the Flemish Government—department WEWIS.

L.M. performed the simulations and wrote the paper. G.V. and G.V.d.S. supervised the project. All authors discussed the results and reviewed the paper.

DATA AVAILABILITY

Some of the data that support the findings of this article are openly available [31,36]. All raw data corresponding to the findings in the paper are not publicly available upon publication because it is not technically feasible and/or the cost of preparing, depositing, and hosting the data would be prohibitive within the terms of this research project. The data are available from the authors upon reasonable request.

APPENDIX A: CLIPPED NONLINEARITY

We compared the TTS for several clipped nonlinearities: we investigate the influence of x_{cutoff} by taking $x_{\text{cutoff}} \in \{0.5, 1, 2\}$. Additionally, we examined the influence of $-x_i^3$ of omitting the cubic term $x_{\text{cutoff}} = 1$ in the dynamical

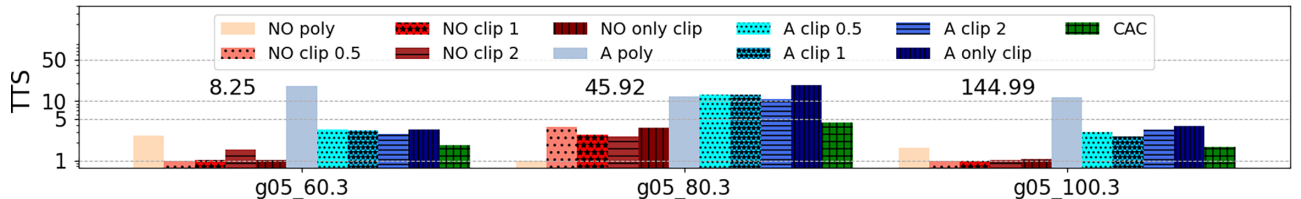


FIG. 12. The TTS of the noise-only scheme (red, stripes) and annealing (blue) scheme for the polynomial nonlinearity and four different clipped nonlinearities ($x_{\text{cutoff}} \in \{0.5, 1, 2\}$ and $x_{\text{cutoff}} = 1$ when $-x_i^3$ is omitted) relative to the best TTS. The relative TTS for CAC is shown as a reference. In the legend, the several nonlinearities are abbreviated: NO stands for the noise-only scheme, A for the annealing scheme, poly for the polynomial nonlinearity, clip c for the clipped nonlinearity with $x_{\text{cutoff}} = c$, and only clip for omitting $-x_i^3$.

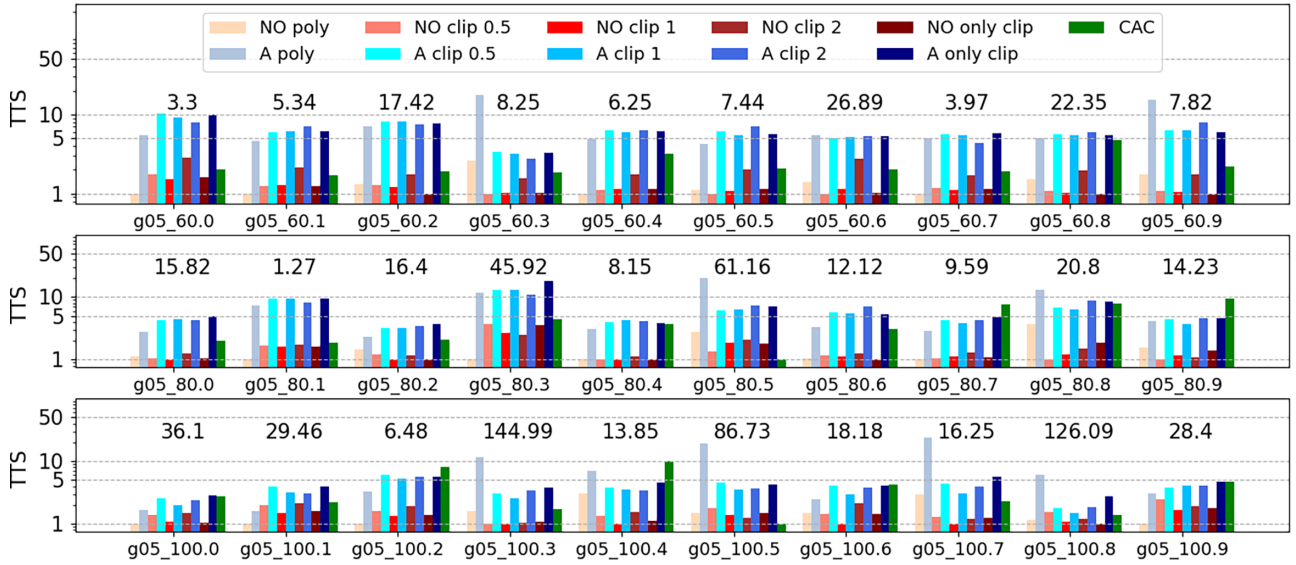


FIG. 13. The TTS of the noise-only scheme (red) and annealing scheme (blue) of the polynomial nonlinearity, several clipped nonlinearities ($x_{\text{cutoff}} \in \{0.5, 1, 2\}$ with $-x_i^3$ included and $x_{\text{cutoff}} = 1$ without $-x_i^3$) and CAC relative to the best TTS of these models. The TTS of the best model is mentioned above to give an idea of the timescale. In the legend, the several nonlinearities are abbreviated: NO stands for the noise-only scheme, A for the annealing scheme, poly for the polynomial nonlinearity, clip c for the clipped nonlinearity with $x_{\text{cutoff}} = c$, and only clip for omitting $-x_i^3$.

equation

$$\frac{dx_i}{dt} = \begin{cases} -x_i - x_i^3 + \beta \sum_{j=1}^N J_{ij} x_j + \gamma F(t), & x_i \leq x_{\text{cutoff}} \\ 0, & x_i > x_{\text{cutoff}}. \end{cases} \quad (\text{A1})$$

The TTS of these clipped nonlinearities is shown in Fig. 12 for three exemplary BiqMac problems of 60 (g05_60.3), 80 (g05_80.3), and 100 (g05_100.3) spins. As a reference, the results for the polynomial nonlinearity and CAC (green) are also included. The complete comparison of the TTS for all g05 BiqMac problems can be found in Fig. 13. Like Fig. 12, it shows the TTS for 11 different models, and both figures have the same color code for the models. While in Fig. 12, the models are grouped per scheme, in Fig. 13, the models are grouped per nonlinearity.

Both figures deliver the same message: Fig. 13 provides an overview of all considered g05 problem instances, whereas Fig. 12 is included for better readability. Although the clip values and the different clipped nonlinearities influence the TTS, the differences in TTS are modest and fall mostly within the expected variability due to the limited number of repetitions. Additionally, using excessively high clip values is generally counterproductive; the clip value 2 almost always leads to a higher TTS. Therefore, we only looked at $x_{\text{cutoff}} = 1$ with the term $-x_i^3$ included in the differential equation for all the problems. Furthermore, the parameter set that leads to the best TTS differs depending on x_{cutoff} . Larger cutoff values require stronger noise for a given coupling strength. Since varying x_{cutoff}

does not substantially affect the TTS, we do not further investigate its influence on the relationship between noise strength and coupling strength. Instead, it suffices to analyze this relationship and the influence of C for a single choice of x_{cutoff} .

APPENDIX B: INFLUENCE OF NOISE FOR SEVERAL PROBLEMS

In Fig. 2 of the paper, we studied the influence of noise strength and coupling strength β on the TTS and success rate. Here we give the results for nine other problems of the BiqMac library with different problem size ($N \in \{60, 80, 100\}$) in Figs. 14–16. The figures are build up similar to Fig. 2. The TTS and success rate are shown for the noise-only scheme and the annealing scheme for both the polynomial and clipped nonlinearity. For these nine problems, we also see a linear relationship between γ and β or v for the polynomial nonlinearity and a square-root dependence between γ and β or v for the clipped nonlinearity. Furthermore, we see a larger spread on the good parameter sets for the annealing scheme than for the noise-only scheme.

APPENDIX C: THE BINOMIAL DISTRIBUTION

During the derivation of the relationship between β and γ_{opt} , we reasoned that $\sum_j s_j$ follows approximately a binomial distribution. We conclude this from the fact that the spin signs have equal chances to be 1 and -1 . In a binomial distribution, s_j have to be independent events, but

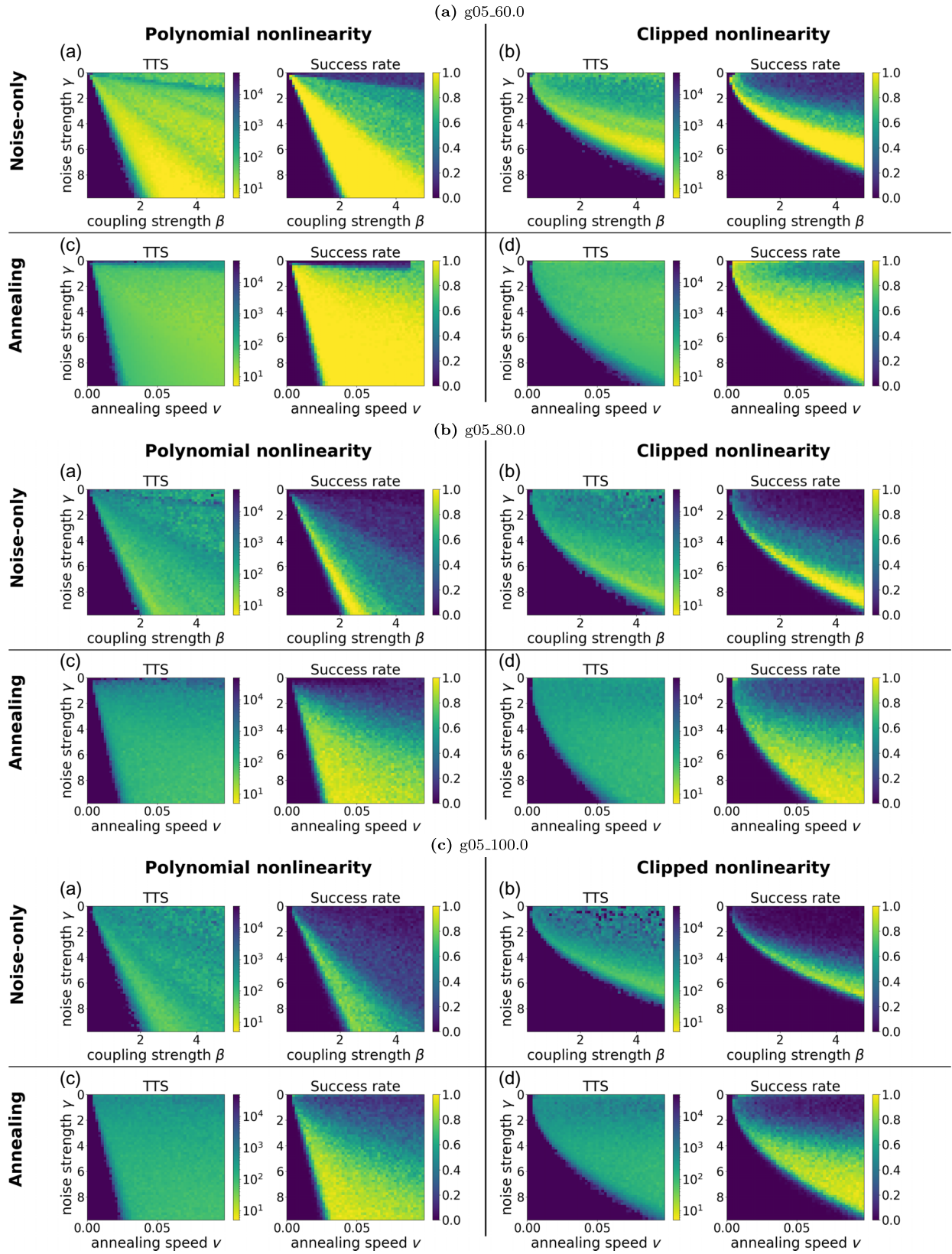


FIG. 14. The success rate and the time-to-solution in a function of γ and β for the noise-only scheme (a),(b) and γ and v for annealing (c),(d) as heat map for three problems. From top to bottom, these problems are g05_60.0, g05_80.0, and g05_100.0 from the BiqMac library. The results for the polynomial nonlinearity can be seen at the left (a),(c) and the results for the clipped nonlinearity at the right (b),(d).

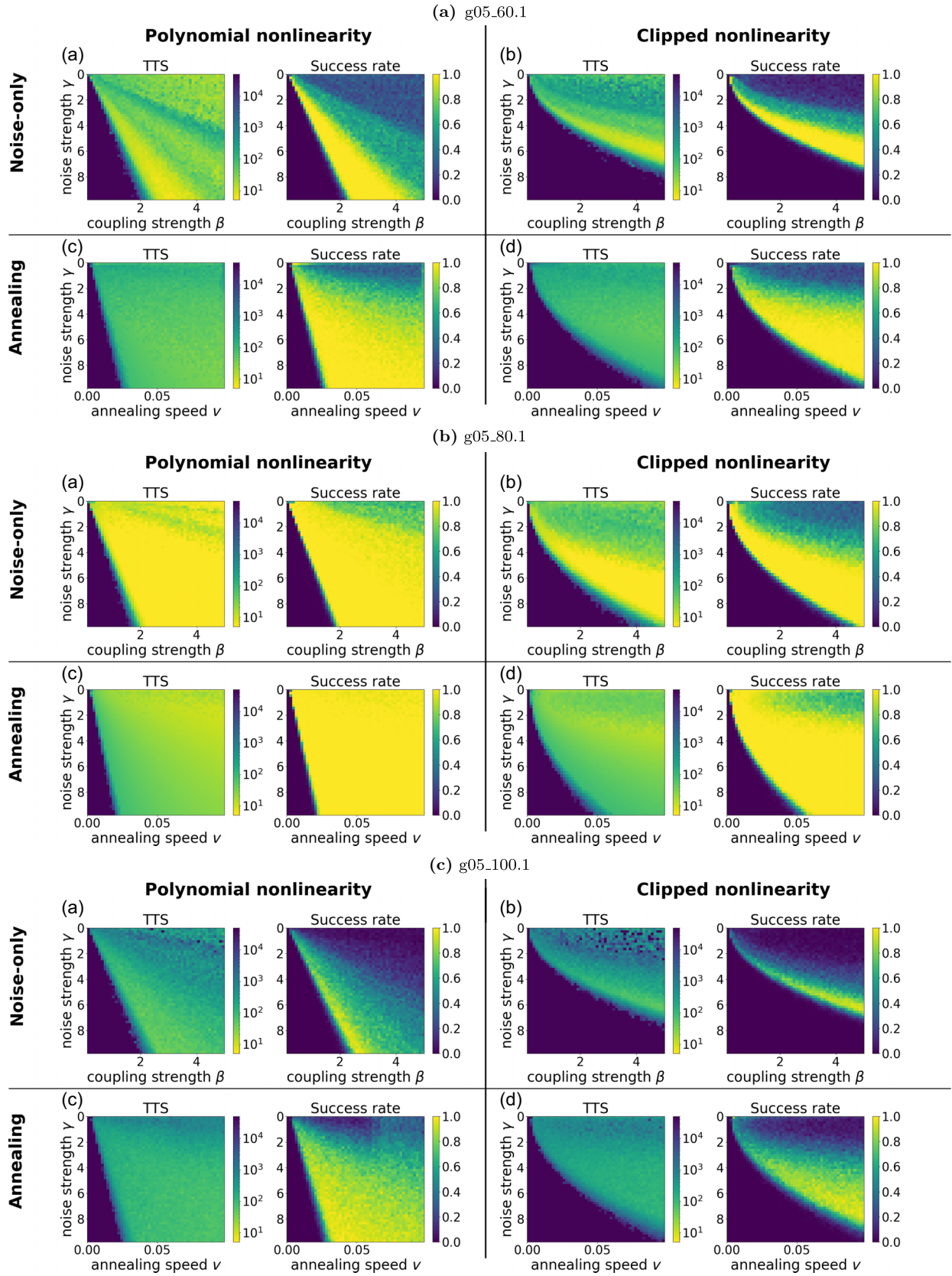


FIG. 15. The success rate and the time-to-solution in function of γ and β for the noise-only scheme (a),(b) and γ and ν for annealing (c),(d) as for three problems. From top to bottom, these problems are g05_60.1, g05_80.0-1, and g05_100.0-1 from the BiqMac library. The results for the polynomial nonlinearity can be seen at the left (a),(c) and the results for the clipped nonlinearity at the right (b),(d).

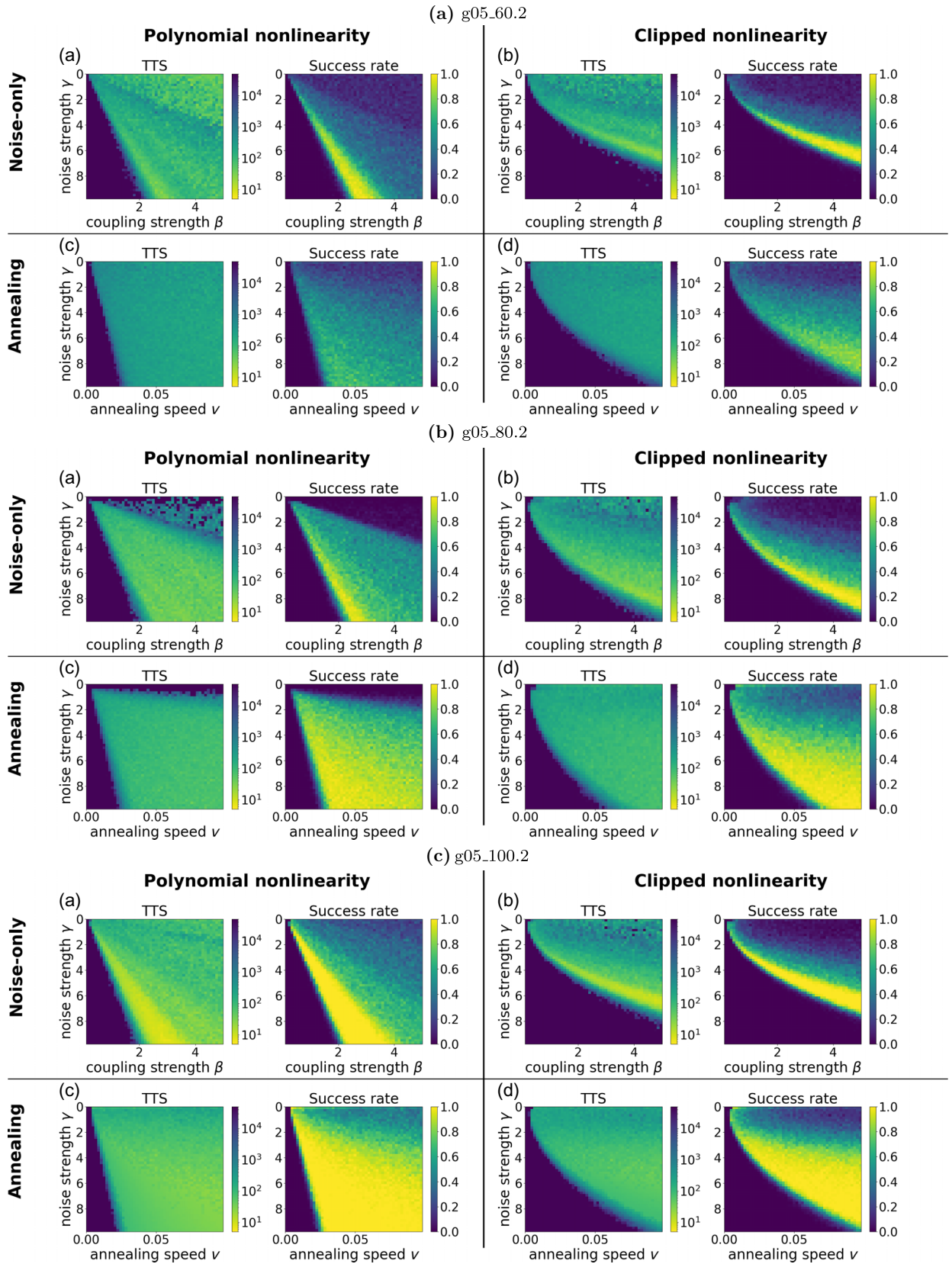


FIG. 16. The success rate and the time-to-solution in function of γ and β for the noise-only scheme (a),(b) and γ and ν for annealing (c),(d) as heat map for three problems. From top to bottom, these problems are g05_60.2, g05_80.2, and g05_100.2 from the BiqMac library. The results for the polynomial nonlinearity can be seen at the left (a),(c) and the results for the clipped nonlinearity at the right (b), (d).

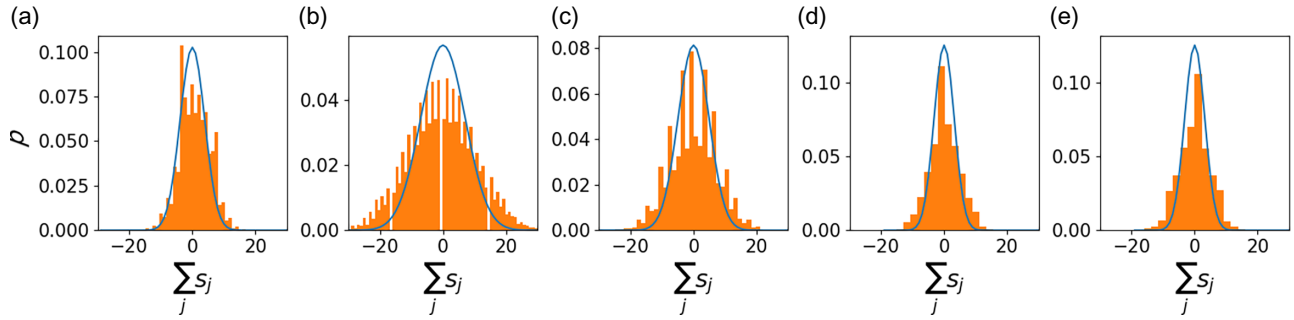


FIG. 17. The distribution of $\sum_j s_j$ for five different problems: (a) BiqMac problem g05_60.3, (b) BiqMac problem pm1d_100.0, (c) Gset problem G1, (d) Gset problem G22, (e) Gset problem G27. The blue line gives the expected binomial distribution for these respective problems.

here they are connected via the coupling matrix. However, the added noise will make them more independent and we are looking over all local minima, so then we will be able to approximate them as independent events. To verify this claim, we check this for five different MaxCut problems using Boltzmann sampling for $\sum_j s_j$. The resulting distribution and the expected binomial distribution are shown in Fig. 17. Clearly, these distributions match very well.

-
- [1] A. Lucas, Ising formulations of many NP problems, *Front. Phys.* **2**, 74887 (2014).
 - [2] H. Takesue, K. Inaba, T. Honjo, Y. Yamada, T. Ikuta, Y. Yonezu, T. Inagaki, T. Umeki, and R. Kasahara, Finding independent sets in large-scale graphs with a coherent Ising machine, *Sci. Adv.* **11**, eads7223 (2025).
 - [3] E. G. Rieffel, D. Venturelli, B. O’Gorman, M. B. Do, E. M. Prystay, and V. N. Smelyanskiy, A case study in programming a quantum annealer for hard operational planning problems, *Quantum Inf. Process.* **14**, 1 (2015).
 - [4] A. Perdomo, C. Truncik, I. Tubert-Brohman, G. Rose, and A. Aspuru-Guzik, Construction of model Hamiltonians for adiabatic quantum computation and its application to finding low-energy conformations of lattice protein models, *Phys. Rev. A* **78**, 012320 (2008).
 - [5] F. Neukart, G. Compostella, C. Seidel, D. von Dollen, S. Yarkoni, and B. Parney, Traffic flow optimization using a quantum annealer, *Front. ICT* **4**, 29 (2017).
 - [6] R. Orús, S. Mugel, and E. Lizaso, Quantum computing for finance: Overview and prospects, *Rev. Phys.* **4**, 100028 (2019).
 - [7] M. W. Johnson *et al.*, Quantum annealing with manufactured spins, *Nature* **473**, 194 (2011).
 - [8] T. Leleu, Y. Yamamoto, S. Utsunomiya, and K. Aihara, Combinatorial optimization using dynamical phase transitions in driven-dissipative systems, *Phys. Rev. E* **95**, 022118 (2017).
 - [9] F. Böhm, T. V. Vaerenbergh, G. Verschaffelt, and G. Van der Sande, Order-of-magnitude differences in computational performance of analog Ising machines induced by the choice of nonlinearity, *Commun. Phys.* **4**, 1 (2021).
 - [10] T. Zhang, Q. Tao, B. Liu, A. Grimaldi, E. Raimondo, M. Jiménez, M. J. Avedillo, J. Nuñez, B. Linares-Barranco, T. Serrano-Gotarredona, G. Finocchio, and J. Han, A review of Ising machines implemented in conventional and emerging technologies, *IEEE Trans. Nanotechnol.* **23**, 704 (2024).
 - [11] N. Mohseni, P. L. McMahon, and T. Byrnes, Ising machines as hardware solvers of combinatorial optimization problems, *Nat. Rev. Phys.* **4**, 363 (2022).
 - [12] P. L. McMahon, A. Marandi, Y. Haribara, R. Hamerly, C. Langrock, S. Tamate, T. Inagaki, H. Takesue, S. Utsunomiya, K. Aihara, R. L. Byer, M. M. Fejer, H. Mabuchi, and Y. Yamamoto, A fully programmable 100-spin coherent Ising machine with all-to-all connections, *Science* **354**, 614 (2016).
 - [13] T. Inagaki, K. Inaba, R. Hamerly, K. Inoue, Y. Yamamoto, and H. Takesue, Large-scale Ising spin network based on degenerate optical parametric oscillators, *Nat. Photonics* **10**, 415 (2016).
 - [14] T. Inagaki, Y. Haribara, K. Igarashi, T. Sonobe, S. Tamate, T. Honjo, A. Marandi, P. L. McMahon, T. Umeki, K. Enbutsu, O. Tadanaga, H. Takenouchi, K. Aihara, K.-i. Kawarabayashi, K. Inoue, S. Utsunomiya, and H. Takesue, A coherent Ising machine for 2000-node optimization problems, *Science* **354**, 603 (2016).
 - [15] Y. Haribara, S. Utsunomiya, and Y. Yamamoto, Computational principle and performance evaluation of coherent Ising machine based on degenerate optical parametric oscillator network, *Entropy* **18**, 151 (2016).
 - [16] F. Böhm, G. Verschaffelt, and G. Van der Sande, A poor man’s coherent Ising machine based on opto-electronic feedback systems for solving optimization problems, *Nat. Commun.* **10**, 3538 (2019).
 - [17] Q. Cen, H. Ding, T. Hao, S. Guan, Z. Qin, J. Lyu, W. Li, N. Zhu, K. Xu, Y. Dai, and M. Li, Large-scale coherent Ising machine based on optoelectronic parametric oscillator, *Light: Sci. Appl.* **11**, 333 (2022).
 - [18] L. Quinn, Y. Xu, J. Fatome, G.-L. Oppo, S. G. Murdoch, M. Erkintalo, and S. Coen, Coherent Ising machine based on polarization symmetry breaking in a driven Kerr resonator, *Nat. Commun.* **17**, 2100 (2026).
 - [19] A. B. Finnila, M. A. Gomez, C. Sebenik, C. Stenson, and J. D. Doll, Quantum annealing: A new method for minimizing multidimensional functions, *Chem. Phys. Lett.* **219**, 343 (1994).

- [20] D. Pierangeli, G. Marcucci, D. Brunner, and C. Conti, Noise-enhanced spatial-photonic Ising machine, *Nano photonics* **9**, 4109 (2020).
- [21] R. Shi, F. Böhm, T. V. Vaerenbergh, and P. Bienstman, Enhancing the performance of coherent Ising machines in the large-noise regime with a fifth-order nonlinearity, *Opt. Express* **32**, 21681 (2024).
- [22] T. Kashimata, M. Yamasaki, R. Hidaka, and K. Tatsumura, Efficient and scalable architecture for multiple-chip implementation of simulated bifurcation machines, *IEEE Access* **12**, 36606 (2024).
- [23] H. Goto, K. Endo, M. Suzuki, Y. Sakai, T. Kanao, Y. Hamakawa, R. Hidaka, M. Yamasaki, and K. Tatsumura, High-performance combinatorial optimization based on classical mechanics, *Sci. Adv.* **7**, eabe7953 (2021).
- [24] T. Leleu, Y. Yamamoto, P. L. McMahon, and K. Aihara, Destabilization of local minima in analog spin systems by correction of amplitude heterogeneity, *Phys. Rev. Lett.* **122**, 040607 (2019).
- [25] J. Lamers, G. Verschaffelt, and G. Van der Sande, Using continuation methods to analyse the difficulty of problems solved by Ising machines, *Commun. Phys.* **7**, 378 (2024).
- [26] Z. Liao, K. Ma, M. S. Sarker, S. Tang, H. Yamahara, M. Seki, and H. Tabata, Quantum analog annealing of gain-dissipative Ising machine driven by colored Gaussian noise, *Adv. Theory Simul.* **5**, 2100497 (2022).
- [27] Z. Liao, K. Ma, M. S. Sarker, H. Yamahara, M. Seki, and H. Tabata, Overdamped Ising machine with stochastic resonance phenomena in large noise condition, *Nonlinear Dyn.* **112**, 8967 (2024).
- [28] C. Roques-Carmes, Y. Shen, C. Zanoci, M. Prabhu, F. Atieh, L. Jing, T. Dubček, C. Mao, M. R. Johnson, V. Čeperić, J. D. Joannopoulos, D. Englund, and M. Soljačić, Heuristic recurrent algorithms for photonic Ising machines, *Nat. Commun.* **11**, 249 (2020).
- [29] K. Lee, S. Chowdhury, and K. Y. Camsari, Noise-augmented chaotic Ising machines for combinatorial optimization and sampling, *Commun. Phys.* **8**, 35 (2025).
- [30] W. Lee, H. Kim, H. Jung, Y. Choi, J. Jeon, and C. Kim, Correlation free large-scale probabilistic computing using a true-random chaotic oscillator p-bit, *Sci. Rep.* **15**, 8018 (2025).
- [31] A. Wiegele, BiqMac library—BINARY quadratic and MAX cut library, <https://biqmac.aau.at/biqmaclib.html> (2007).
- [32] D. Dobrynin, A. Renaudineau, M. Hizzani, D. Strukov, M. Mohseni, and J. P. Strachan, Energy landscapes of combinatorial optimization in Ising machines, *Phys. Rev. E* **110**, 045308 (2024).
- [33] R. Martinez-Cantin, BayesOpt: A Bayesian optimization library for nonlinear optimization, experimental design and bandits, *J. Mach. Learn. Res.* **15**, 3915 (2014).
- [34] T. Leleu, F. Khoyratee, T. Levi, R. Hamerly, T. Kohno, and K. Aihara, Scaling advantage of chaotic amplitude control for high-performance combinatorial optimization, *Commun. Phys.* **4**, 1 (2021).
- [35] T. A. Davis and Y. Hu, The University of Florida sparse matrix collection, *ACM Trans. Math. Softw.* **38**, 1 (2011).
- [36] Y. Ye, Gset, <https://web.stanford.edu/yyye/yyye/Gset/>, accessed: 2025-04-22.
- [37] C. Helmberg and F. Rendl, A Spectral Bundle Method for Semidefinite Programming, *SIAM J. Optim.* **10**, 673 (2000).
- [38] F. Böhm, D. Alonso-Urquijo, G. Verschaffelt, and G. Van der Sande, Noise-injected analog Ising machines enable ultrafast statistical sampling and machine learning, *Nat. Commun.* **13**, 5847 (2022).
- [39] Y. Matsuda, Benchmarking the MAX-CUT problem on the simulated bifurcation machine, <https://medium.com/toshiba-sbm/benchmarking-the-max-cut-problem-on-the-simulated-bifurcation-machine-e26e1127c0b0> (2019).
- [40] M. San Miguel and R. Toral, in *Instabilities and Nonequilibrium Structures VI*, edited by E. Tirapegui, J. Martínez, and R. Tiemann (Springer Netherlands, Dordrecht, 2000), p. 35.
- [41] T. Sevenants, G. Van der Sande, and G. Verschaffelt, Requirements on bit resolution in optical Ising machine implementations with analog spin variables, *Commun. Phys.* **8**, 1 (2025).