Science Advances

Supplementary Materials for

Easing the Monte Carlo sign problem

Dominik Hangleiter*, Ingo Roth, Daniel Nagaj, Jens Eisert

*Corresponding author. Email: dominik.hangleiter@fu-berlin.de

This PDF file includes:

Sections S1 to S4
Figs. S1 to S5
Appendices A and B
References

OVERVIEW

The plan for the technical part of this work is as follows: In Section S1 we sketch the idea of world-line QMC methods and explain how the sign problem arises there. In Section S2 we then discuss the relation between the average sign and non-stoquasticity. There, we construct examples showing that the two are in general unrelated (S2 A), but then continue to argue both analytically (S2 B) and numerically (S2 C) that the non-stoquasticity ν_1 defined in Eq. (1) of the main text is a meaningful and efficiently computable (S2 D) measure of the sign problem. In Section S3 we perform a proof-of-principle numerical study showing that easing is both feasible and meaningful for translationally invariant models with a sign problem. In Section S4 we then study the fundamental limitations of a systematic approach to the sign problem in proving the computational hardness of SignEasing when allowing for both orthogonal Clifford (Theorem 8) and general orthogonal transformations (Theorem 9). Details of the proofs are deferred to the appendices.

S1. THE SIGN PROBLEM OF QUANTUM MONTE CARLO

We begin the technical part of this work with an exposition of the basics of Quantum Monte Carlo methods. For the purpose of this work, we focus on the prominent *world-line Monte Carlo* method of calculating partition functions and thermal expectation values of a Hamiltonian H at inverse temperature β [47]. Here, both quantities are expressed as

$$Z_{\beta,H} \simeq \text{Tr}[T_m^m] = \sum_{\vec{\lambda} \in \Lambda_{m+1},\, \lambda_{m+1}=\lambda_1} a(\vec{\lambda}) \tag{1}$$

$$\langle O \rangle_{\beta,H} \simeq \frac{1}{Z_{\beta,H}} \text{Tr}[T_m^m O] = \frac{1}{Z_{\beta,H}} \sum_{\vec{\lambda} \in \Lambda_{m+1}} a(\vec{\lambda}) O(\lambda_m|\lambda_1), \tag{2}$$

for large enough $m \in \mathbb{N}$ *Monte Carlo steps* in terms of the amplitudes

$$a(\vec{\lambda}) = T_m(\lambda_1|\lambda_2) T_m(\lambda_2|\lambda_3) \cdots T_m(\lambda_m|\lambda_{m+1}), \tag{3}$$

on the configuration space $\Lambda_{m+1} = [\dim \mathcal{H}]^{\times(m+1)}$. Here, we have defined the transfer matrix $T_m(\lambda'|\lambda) = \langle \lambda'| \mathbb{1} - \beta H/m |\lambda \rangle$ and in general denote the entries of a matrix A as $A(\lambda_1|\lambda_2) = \langle \lambda_1|A|\lambda_2 \rangle$. The computation of the partition function involves a summation over all closed paths of length m (i.e., paths with periodic boundary conditions); the computation of general observables involves a summation over all open paths.

For non-negative path weights, both quantities may be rewritten as expectation values in a probability distribution $q(\vec{\lambda}) = a(\vec{\lambda})/\sum_{\vec{\lambda}} a(\vec{\lambda})$, which reduces to $q(\vec{\lambda}) = a(\vec{\lambda})/Z_{\beta,H}$ when computing the expectation value of diagonal observables. The sign problem is manifested in the fact that the off-diagonal entries of H may be positive potentially implying

that $a(\vec{\lambda}) < 0$. Therefore $q(\vec{\lambda})$ is in general a quasi-probability distribution.

To compute the quantities (1) and (2) via Monte Carlo sampling, one constructs a linear estimator as the expectation value $\langle f \rangle_p = \sum_{\vec{\lambda}} p(\vec{\lambda}) f(\vec{\lambda})$ of a random variable f distributed according to a probability distribution p. By Chebyshev's inequality the statistical error ϵ, i.e. the deviation from the mean, when averaging s samples of an i.i.d. random variable X is upper bounded by its variance

$$\epsilon \le \sqrt{\text{Var}(X)/(s(1-\delta))}, \tag{4}$$

with probability at least $1 - \delta$. Hence, to achieve any relative error $\tilde{\epsilon}$, the number of samples needs to grow with the variance of the random variable normalized by its expectation value. In fact, it can be easily shown that the variance-optimal estimator for the partition function $Z_{\beta,H}$ is given by the probability distribution $p(\vec{\lambda}) = |a(\vec{\lambda})|/\|a\|_{\ell_1}$ with $\|a\|_{\ell_1} = \sum_{\vec{\lambda}} |a(\vec{\lambda})|$ and the estimator $f(\vec{\lambda}) = \text{sign}(a(\vec{\lambda})) \cdot \|a\|_{\ell_1}$ [51]. The variance of this estimator is given by

$$\text{Var}_p(f) = \|a\|_{\ell_1}^2 (\|q\|_{\ell_1}^2 - 1) \tag{5}$$

and hence the relative error of the approximation by

$$\frac{\text{Var}_p(f)}{\langle f \rangle_p^2} = \|q\|_{\ell_1}^2 - 1 \equiv \langle \text{sign} \rangle_p^{-2} - 1, \tag{6}$$

where $\langle \text{sign} \rangle_p = 1/\|q\|_{\ell_1}$ is called the *average sign* of the quasi-probability distribution q. One may interpret the average sign as the ratio between the partition functions of the original system with Hamiltonian H acting on n qubits and a corresponding 'bosonic system' with Hamiltonian $H' = (H - 2H_\neg)$ as $\langle \text{sign} \rangle_p = \text{Tr}[e^{-\beta H}]/\text{Tr}[e^{-\beta H'}]$. Generically, such a quantity is expected to scale as $e^{-\beta n \Delta f}$, that is, inverse exponentially in the particle number n, the inverse temperature β, and the free energy density difference $\Delta f = f' - f \ge 0$ between 'bosonic' and original system [18].

In order to minimize the relative approximation error of a QMC algorithm, we therefore need to minimize the inverse average sign, or equivalently $\|q\|_{\ell_1}$, over the allowed set of basis choices which we denote by $\mathcal{U}$. To optimally ease the sign problem in terms of its sample (and hence computational) complexity one therefore needs to solve the following minimization problem

$$\min_{U \in \mathcal{U}} \|q\|_{\ell_1}^2 - 1 = \min_{U \in \mathcal{U}} \frac{\text{Tr}[|U T_m U^\dagger|^m]^2}{\text{Tr}[T_m^m]^2} - 1, \tag{7}$$

where as throughout this work $|\cdot|$ denotes taking the entry-wise absolute value and not the matrix absolute value.

S2. THE RELATION BETWEEN THE AVERAGE SIGN AND NON-STOQUASTICITY

The difficulty in dealing with the minimization problem (7) is manifold. First, determining the quantity $\|q\|_{\ell_1} = $

$\mathrm{Tr}[|T_m|^m]/\mathrm{Tr}[T_m^m]$ via QMC suffers from the very sign problem it quantifies: it can easily be checked that the relative variance of $\langle\mathrm{sign}\rangle_p$ is precisely given by $\langle\mathrm{sign}\rangle_p^{-2} - 1$. It thus inherits the complexity of computing the partition function $Z_{\beta,H}$ in the first place. Naïve optimization of the term $\mathrm{Tr}[|T_m|^m]/\mathrm{Tr}[T_m^m]$ even incurs the cost of diagonalizing the exponential-size matrices T_m and $|T_m|$. Second, the optimization problem is non-convex and highly non-linear in the unitary transformation $T \mapsto UTU^\dagger$ with $U \in \mathcal{U}$.

While it might be possible to minimize the unitarily dependent term $\mathrm{Tr}[|T_m|^m]$ and its gradient stochastically via QMC in some cases [55, 64], such approaches cannot yield certificates for the quality of the obtained basis as the average sign itself is not computed. Moreover, they are dependent on the distribution defined by $|T_m|$ being well-behaved (i.e., ergodic and satisfying detailed balance) for QMC algorithms.

It therefore seems infeasible to find a converging and efficient algorithm for minimizing the average sign for general Hamiltonians directly. Ideally, one could find a simple quantity measuring the non-stoquasticity of the Hamiltonian which can be connected to the inverse average sign in a meaningful way while at the same time admitting efficient evaluation.

A. Case studies

We now show that this hope is in vain in its most general formulation. Specifically, we provide an example of a Hamiltonian which has large positive entries but is nevertheless sign-problem free (has unit average sign) for specific choices of β and m, as well as an example of an Hamiltonian that is close to stoquastic but incurs an arbitrarily small average sign for certain choices of β and m in a specific QMC procedure.

Here, as throughout this work, whenever we consider systems of multiple qubits, for $A \in \mathbb{C}^{2\times2}$ we define

$$A_i = A \otimes \mathbb{1}_{\{i\}^c}, \tag{8}$$

to be the operator that acts as A on qubit i and trivially on its complement $\{i\}^c$.

Example 3 (Highly non-stoquastic but sign-problem free Hamiltonians). *Let us define a Hamiltonian term acting on two qubits with label i, j as*

$$h_{i,j} = -\frac{1}{2}(X_i X_j - Y_i Y_j) + X_i. \tag{9}$$

Then this Hamiltonian term is non-stoquastic with total weight $\nu_1(h_{i,j}) = 1$. What is more, the n-qubit Hamiltonian

$$H = \mathbb{1} + \sum_{i<j}^n h_{i,j} \tag{10}$$

is highly non-stoquastic with total weight $\nu_1(H) = n$. At the same time, the QMC algorithm for computing the partition function of H with parameters β, m, has average sign $\langle\mathrm{sign}\rangle_{\beta,m} = 1$.

Proof. We first determine the non-stoquasticity of H as

$$\nu_1(H) = \sum_i \nu_1(X_i) = n. \tag{11}$$

To see why the QMC algorithm has unit average sign, note that the transfer matrix $T_m = \mathbb{1} - \beta H/m$ has negative entries $T_m(\lambda|\lambda') < 0$ only if the parity of $\lambda \oplus \lambda'$ is odd since for these terms only a single X term contributes. Whenever $\lambda \oplus \lambda' = 0$, i.e., has even parity, we have $T_m(\lambda|\lambda') \geq 0$ since only $XX - YY$ terms or the diagonal contribute – both of which have non-negative matrix elements.

In the calculation of the partition function, the summation runs over closed paths only. But for any closed path $\lambda_1 \rightarrow \lambda_2 \rightarrow \cdots \rightarrow \lambda_m \rightarrow \lambda_1$, it is necessary that the total parity $(\lambda_1 \oplus \lambda_2) \oplus \ldots \oplus (\lambda_m \oplus \lambda_1)$ vanishes. In particular, this implies that every allowed path incurs an even number of odd-parity steps and therefore an even number of negative signs. Therefore, only non-negative paths contribute to the path integral and the average sign is attained at unity. $\qquad\square$

Example 4 (Barely non-stoquastic Hamiltonians with arbitrarily small average sign). *Let us define the 2-qubit Hamiltonian*

$$H_{a,b} = \frac{m}{\beta}\bigg(\mathbb{1}\otimes\mathbb{1} - \mathbb{1}\otimes X - \frac{1}{2}(X\otimes X + Y\otimes Y) \tag{12}$$

$$+ \frac{1}{2}[(a+b)X\otimes Z + (b-a)X\otimes\mathbb{1}]\bigg), \tag{13}$$

with $b \geq a > 0$ positive numbers and $m \in 2\mathbb{N} + 1$. The non-stoquasticity of $H_{a,b}$ is given by $\nu_1(H_{a,b}) = bm/(2\beta)$, the average sign of QMC with parameters β and m is dominated by $|\langle\mathrm{sign}\rangle_{\beta,m}| \leq C(b-a)/a$, where C is a constant. Thus, even for arbitrarily small non-stoquasticity we can make the sign problem unboundedly severe as we tune a to be close to b.

Proof. We derive the bound on the average sign. For the given Hamiltonian, the corresponding transfer matrix for a QMC algorithm for inverse temperature β with m steps is given by

$$T_m \equiv T_{a,b} = \begin{pmatrix} 0 & 1 & -b & 0 \\ 1 & 0 & 1 & a \\ -b & 1 & 0 & 1 \\ 0 & a & 1 & 0 \end{pmatrix}. \tag{14}$$

Recall that the average sign can be written as

$$\langle\mathrm{sign}\rangle_{\beta,m} = \frac{\mathrm{Tr}[T_m^m]}{\mathrm{Tr}[|T_m|^m]}. \tag{15}$$

We denote by $\overline{T}_m$ a matrix similar to T_m but where the positions of a and $-b$ are exchanged. Due to the symmetry of the problem we have that $\mathrm{Tr}[T_m^m] = \mathrm{Tr}[\overline{T}_m^m]$ and $\mathrm{Tr}[|T_m|^m] = $

$\mathrm{Tr}[|\overline{T}_m|^m]$. Hence,

$$\mathrm{Tr}[T_m^m] = \frac{1}{2}\left[\mathrm{Tr}[T_m^m] + \mathrm{Tr}[\overline{T}_m^m]\right] \tag{16}$$

$$= \frac{1}{2}\sum_{\vec{\lambda}\in\Lambda_m}\left[T_m(\lambda_1 \mid \lambda_2)\cdots T_m(\lambda_m \mid \lambda_1) \right. \tag{17}$$

$$\left. + \overline{T}_m(\lambda_1 \mid \lambda_2)\cdots \overline{T}_m(\lambda_m \mid \lambda_1)\right] \tag{18}$$

$$= \frac{1}{2}\sum_{\vec{\lambda}}\left[a^{f(\vec{\lambda})}(-b)^{g(\vec{\lambda})} + a^{g(\lambda)}(-b)^{f(\vec{\lambda})}\right], \tag{19}$$

where in the last line we have used the fact that every summand is a polynomial in the entries of $T_{a,b}$. The functions $g, f : \Lambda_m \to [m]$ describe the corresponding exponents. A little thought reveals that since all path are closed and m is odd $g(\vec{\lambda}) + f(\vec{\lambda})$ is larger than 1 and also odd for all $\vec{\lambda}$. We thus find that one of the two terms for each $\vec{\lambda}$ must be negative and

$$|\mathrm{Tr}\,T_m^m| \le \frac{1}{2}\sum_{\vec{\lambda}}|a^{f(\vec{\lambda})}b^{g(\vec{\lambda})} - b^{f(\lambda)}a^{g(\vec{\lambda})}| \tag{20}$$

$$\le \frac{1}{2}\sum_{\vec{\lambda}}(2^{g(\vec{\lambda})} - 1)a^{f(\vec{\lambda})+g(\vec{\lambda})-1}|b - a|. \tag{21}$$

Furthermore, we have

$$|\mathrm{Tr}\,|T_m|^m| = \frac{1}{2}\sum_{\vec{\lambda}}(a^{f(\vec{\lambda})}b^{g(\vec{\lambda})} + b^{f(\lambda)}a^{g(\vec{\lambda})}) \tag{22}$$

$$\ge \sum_{\vec{\lambda}}\left(a^{f(\vec{\lambda})+g(\vec{\lambda})}\right). \tag{23}$$

Combining these two bounds and using $g(\vec{\lambda}) \le m$, we conclude that

$$|\langle\mathrm{sign}\rangle| \le \left(2^{m-1} - \frac{1}{2}\right)\frac{|b-a|}{a}. \tag{24}$$

$\square$

The second example shows that in principle also Hamiltonians with arbitrarily small positive entries can cause a severe increase of the sampling complexity of specific Monte Carlo algorithms. Interestingly, in this situation the sign problem cannot be eased by a basis change: the average sign vanishes since the *unitarily invariant* term $|\mathrm{Tr}\,T_m^m|$ is tuned to be small. On the contrary, the sign problem can be completely avoided in this example by choosing the Monte-Carlo path length to be even instead of odd.

These simple examples illustrate the following important observation: The sign problem as measured by the average sign can in certain situations be avoided or intensified by fine-tuning the problem and parameters of the QMC procedure independently of the actual magnitude of the positive entries of the Hamiltonian. But such examples seem to rely on an intricate conspiracy of the structure of the Hamiltonian and the

chooen QMC procedure, e.g., the discretization. It is plausible to assume that the most pathological cases are unlikely to appear in practical applications, and can at least be rather easily overcome by slightly modifying the QMC algorithm.

B.   Measures of non-stoquasticity

In this work, our goal is to develop a more general methodology for the task of easing the sign problem that is independent of the details of the QMC algorithm and the combinatorial properties of potential paths that can be constructed from the entries of the transfer matrix. Very much in the spirit of the notion of stoquasticity we aim at a property of the Hamiltonian in a given basis to measure its deviation from having a good sampling complexity. Natural candidates for such a non-stoquasticity measure of a Hamiltonian are entry-wise norms of its positive entries. For any $p \ge 1$ we define the non-stoquasticitiy of H as

$$\nu_p(H) = D^{-1}\|H_\neg\|_{\ell_p}, \tag{25}$$

where $\|\cdot\|_{\ell_p}$ denotes the vector-ℓ_p norm. For every p, ν_p is efficiently computable for local Hamiltonians on bounded-degree graphs and therefore suitable for easing the sign problem of a large class of Hamiltonians by local basis choices. It is also obviously a measure of the non-stoquasticity in the sense that $\nu_p(H) = 0$ if and only if H is stoquastic. We note that we have chosen our definition such that the non-stoquasticity measure ν_p scales extensively in the number of non-stoquastic terms of a local Hamiltonian. This is because every non-stoquastic local Hamiltonian term creates on the order of 2^n positive matrix entries in a global n-qubit Hamiltonian matrix due to tensoring with identities on the complement of its support.

Our examples in the previous section show that it is notoriously difficult if not impossible to connect any notion of non-stoquasticity to the actual sample complexity incurred by a QMC procedure as measured by the inverse average sign. This is due to the dependence of the average sign on the combinatorics of Monte Carlo paths. However, those examples involved a large degree of fine-tuning. Therefore, one might hope to find a connection between non-stoquasticity and average sign for *generic* cases.

So let us look at the connection between optimizing a non-stoquasticity measure ν_p and optimizing the QMC sampling complexity as in (7). Our measure can be expressed in terms of the transfer matrix T_m as

$$\nu_p(H) = \frac{1}{D}\frac{m}{2\beta}\||T_m| - T_m\|_{\ell_p}, \tag{26}$$

where we assume that $\mathrm{diag}(\beta H/m) \le \mathbb{1}$.

Due to the unitary invariance of the trace, the optimization of the sampling complexity via (7) is equivalent to minimising

$$S(U) = \mathrm{Tr}[|UT_mU^\dagger|^m] - \mathrm{Tr}[T_m^m]. \tag{27}$$

Let us for the sake of clarity, suppress the explicit dependence on the unitary U and define $\hat{T}_m = UT_mU^\dagger$. If we define the

positive and negative entries of the transfer matrix respectively as $\Delta_\pm = \frac{1}{2}\left(|\hat{T}_m| \pm \hat{T}_m\right)$, we can write

$$S(U) = \text{Tr}[|\hat{T}_m|^m] - \text{Tr}[\hat{T}_m^m] \tag{28}$$

$$= 2 \sum_{\substack{\vec{s} \in \{\pm\}^m: \\ \vec{s} \text{ odd}}} \text{Tr}[\Delta_{s_1} \cdots \Delta_{s_m}]. \tag{29}$$

The summation in the last line is restricted to all $\vec{s} \in \{\pm\}^m$ with an odd number of negative signs. The resulting expression thus involves a summation over closed paths that contain an odd number of negative contributions such that $\Delta_{s_1}(\lambda_1|\lambda_2) \cdots \Delta_{s_m}(\lambda_m|\lambda_1) < 0$. In particular, every such path contains at least one step with a negative contribution.

The size of $S(U)$ thus depends both on the number of 'negative paths' and their individual weight. From Eq. (29) we find that the contribution to $S(U)$ of paths with exactly one negative step has the form

$$2m \sum_{\lambda_1, \lambda_2} \Delta_-(\lambda_1|\lambda_2)\Delta_+^{m-1}(\lambda_2|\lambda_1), \tag{30}$$

using the cyclicity of the trace. This expression (30) is a weighted sum over the negative entries of $\hat{T}_m$, where the weights are given by the contribution $\Delta_+^{m-1}(\lambda_2|\lambda_1)$ of all positive paths of length $m-1$.

For a transfer matrix in which the positive entries do not significantly differ and their distribution relative to the negative entries is unstructured, we have constant $\Delta_+^{m-1}(\lambda_2|\lambda_1) \approx \|\Delta_+^{m-1}\|_{\ell_\infty}$. Therefore, the linear term (30) scales approximately as

$$2m\|\Delta_-\|_{\ell_1}\|\Delta_+^{m-1}\|_{\ell_\infty} \propto D\,\nu_1(H). \tag{31}$$

For higher-order negative contributions, we expect that $S(U)$ or, correspondingly, the average sign scales as $\exp(c \cdot D\,\nu_1(H))$ for some $c > 0$. Our expectation is based on the following observation: in the calculation of the inverse average sign, all paths of length m with an odd number of negative steps contribute. Potentially, in each step every negative entry of T_m appears. Then the sum of all negative entries of T_m contributes. But the number of paths with $k \in 2\mathbb{N}_0 + 1$ negative steps scales as $\binom{m}{k}$ which leads to an exponential growth in $\|H_\neg\|_{\ell_1}$ and hence $D\,\nu_1(H)$. In the following section, we provide a brief numerical analysis confirming this expectation.

We further observe that if the positive entries of $\hat{T}_m$ are more structured, the weights appearing in Eq. (30) might deviate from a uniform distribution. In such a case, other ν_p-measures become meaningful as a measure of the inverse average sign since they saturate a corresponding Hölder bound.

C. Numerical analysis

In this subsection, we provide evidence that $\nu_1(H)$ is indeed a very much meaningful measure of the sample complexity and hence the inverse average sign by exactly calculating
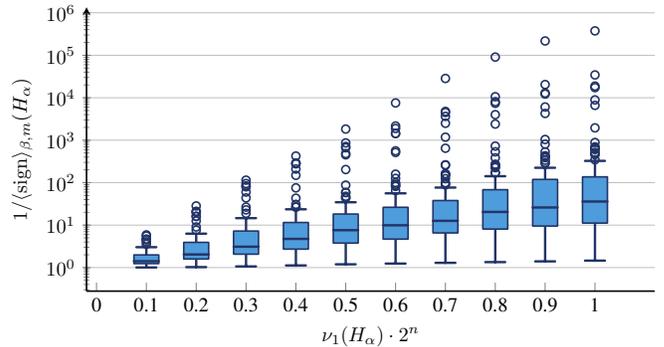

Figure S1. The figure shows the inverse average sign for 100 randomly chosen instances of 5-qubit Hamiltonians H_α for $\beta = 1$ and $m = 100$ Monte Carlo steps as a function of $d\nu_1(H_\alpha)$. We find a roughly exponential dependence of the inverse average sign with $\nu_1(H_\alpha)$ as $1/\langle\text{sign}\rangle_{\beta,m}(H_\alpha) \propto \exp(a \cdot d\nu_1(H_\alpha))$ for $a > 0$.

the inverse average sign as a function of $\nu_1(H)$. We do so by randomly drawing 2-local Hamiltonians on a line of n qubits of the form

$$H = \sum_{i=1}^{n} T_i(h), \tag{32}$$

where $h \in \mathbb{R}^{4 \times 4}$ is a nearest-neighbour interaction term and the translation operator T_i acts as $T_i(h) = \mathbb{1}_d^{\otimes(i-1)} \otimes h \otimes \mathbb{1}_d^{\otimes n-i-1}$. We choose each local term h in an i.i.d. fashion from the zero-centered Gaussian measure and projecting to Hermitian matrices. For each random instance H, we then consider the one-parameter Hamiltonian family

$$H_\alpha = \frac{H - H_\neg + \alpha H_\neg}{2^n \nu_1(H_\neg)}. \tag{33}$$

Note that $\nu_1(H_\alpha) = \alpha/2^n$. Fig. S1 shows that, generically, the average sign monotonously depends on the non-stoquasticity. Indeed, as expected for large m, the dependence is an exponential one.

D. Computing the non-stoquasticity

Above, we have argued that a key problem of the average sign lies in the fact that it is not efficiently computable whenever there is a sign problem. But how does the non-stoquasticity measure ν_1 fare in this respect? We now show that for arbitrary 2-local Hamiltonians the non-stoquasticity measure ν_1 can in fact be approximated up to an inverse polynomially small additive error in polynomial time. While this is not possible for arbitrary local Hamiltonians as a simple example shows, any ν_p-measure can be efficiently computed exactly in polynomial time for local Hamiltonians acting on bounded-degree graphs.

We write a real 2-local Hamiltonian with 1-local terms as

$$H_{2+1} = \sum_{i<j} \big(a_{i,j} X_i X_j + b_{i,j} Y_i Y_j + c_{i,j} Z_i Z_j \tag{34}$$
$$+ x_{i,j} X_i Z_j + x_{j,i} Z_i X_j \big) + \sum_i \big(\alpha_i X_i + \gamma_i Z_i \big),$$

parametrized by real coefficient vectors $a, b, c \in \mathbb{R}^{n(n-1)/2}$, $x \in \mathbb{R}^{n(n-1)}$ which are non-zero only on the edges $(i,j) \in E$ of the interaction Hamiltonian graph $G = (V, E)$ as well as vectors $\alpha, \gamma \in \mathbb{R}^n$. The Hamiltonian interaction graph is defined by a set V of sites or vertices and the edge set

$$E = \big\{ (i,j) \in V \times V :$$
$$\neg(a_{i,j} = b_{i,j} = c_{i,j} = x_{i,j} = x_{j,i} = 0) \big\}. \tag{35}$$

We call $\mathcal{N}(i) = \{j : (i,j) \in E\}$ the neighbourhood of site i on the graph G, $\deg(i) = |\mathcal{N}(i)|$ the degree of site i and $\deg(G) = \max_{i \in V} \deg(i)$ the degree of the overall graph. Notice that obtaining an expression for the non-stoquasticity is non-trivial since several local Hamiltonian terms may contribute to a particular entry of the global Hamiltonian matrix.

Lemma 5 (Non-stoquasticity of $(2+1)$-local Hamiltonians). *The non-stoquasticity measure ν_1 of real 2-local Hamiltonians with 1-local terms of the form H_{2+1} satisfies*

$$\nu_1(H_{2+1}) = \sum_{i<j} \nu_1(a_{i,j} X_i X_j + b_{i,j} Y_i Y_j)$$
$$+ \sum_i \nu_1 \bigg(\alpha_i X_i + \sum_{j \in \mathcal{N}_{XZ}(i)} x_{i,j} X_i Z_j \bigg), \tag{36}$$

and it holds that

$$\nu_1(a_{i,j} X_i X_j + b_{i,j} Y_i Y_j) = \tag{37}$$
$$\frac{1}{2} \sum_{i<j} \big(\max\{a_{i,j} + b_{i,j}, 0\} + \max\{a_{i,j} - b_{i,j}, 0\} \big),$$

$$\nu_1 \bigg(\alpha_i X_i + \sum_{j \in \mathcal{N}_{XZ}(i)} x_{i,j} X_i Z_j \bigg) = 2^{-\deg_{XZ}(i)}$$
$$\times \sum_{\lambda_{\mathcal{N}_{XZ}(i)}} \max \bigg\{ \alpha_i + \sum_{j \in \mathcal{N}_{XZ}(i)} (-1)^{\lambda_j} x_{i,j}, 0 \bigg\}. \tag{38}$$

Here, we have defined the XZ neighbourhood $\mathcal{N}_{XZ}(i) = \{j : x_{ij} \neq 0\}$ of site i as all vertices j connected to i by an $X_i Z_j$-edge. As useful shorthands, we also define the XZ degree $\deg_{XZ}(i) = |\mathcal{N}_{XZ}(i)|$ and the restriction of a spin configuration $\lambda \in \{0,1\}^n$ to an XZ neighbourhood as $\lambda_{\mathcal{N}_{XZ}(i)} = (\lambda_j)_{j \in \mathcal{N}_{XZ}(i)} \in \{0,1\}^{\deg_{XZ}(i)}$. We conceive of summation over an empty set (the case that $\mathcal{N}_{XZ}(i) = \{\}$) as resulting in 0 so that the corresponding term in Eq. (38) vanishes.

Notice that the non-stoquasticity of an XZ term does not depend on the sign of its weight, while for XX and X terms it crucially does.

Proof. We can determine the ℓ_1-norm of the off-diagonal part of the Hamiltonian H_{2+1} as

$$\|H_{2+1} - \mathrm{diag}(H_{2+1})\|_{\ell_1}$$
$$= \sum_{\lambda \in \{0,1\}^n} \bigg\{ \sum_{i<j} |a_{i,j} - (-1)^{\lambda_i + \lambda_j} b_{i,j}|$$
$$+ \sum_i \bigg| \alpha_i + \sum_{j \in \mathcal{N}_{XZ}(i)} (-1)^{\lambda_j} x_{i,j} \bigg| \bigg\}$$
$$= 2^{n-1} \sum_{i<j} \big(|a_{i,j} + b_{i,j}| + |a_{i,j} - b_{i,j}| \big)$$
$$+ \sum_i 2^{n - \deg_{XZ}(i)} \sum_{\lambda_{\mathcal{N}_{XZ}(i)}} \bigg| \alpha_i + \sum_{j \in \mathcal{N}_{XZ}(i)} (-1)^{\lambda_j} x_{i,j} \bigg|. \tag{39}$$

From Eq. (39) we can then directly calculate the non-stoquasticity ν_1 of H_2 by discarding all matrix entries with negative sign before taking the ℓ_1-norm and dividing by 2^n. $\square$

Now, clearly we can efficiently compute the term (37) for arbitrary graphs as the sum runs over at most $n(n-1)/2$ many terms. In the term (38), in contrast, the sum over spin configurations $\lambda_{\mathcal{N}_{XZ}(i)}$ in the XZ neighbourhood of site i runs over $2^{\deg_{XZ}(i)}$ many terms and hence this term is efficiently computable exactly in case the vertex degree $\deg_{XZ}(i)$ of any vertex i grows at most logarithmically with n. This shows that for bounded-degree graphs such as regular lattices the non-stoquasticity can be computed efficiently.

But what if the degree of the graph grows faster than logarithmically with n so that the sum runs over super-polynomially many non-trivial terms? The following Lemma shows that even in this case, that is, for 2-local Hamiltonians acting on arbitrary graphs, the non-stoquasticity can be efficiently approximated up to any inverse polynomially small additive error using Monte Carlo sampling.

Theorem 6. *The sum* (38) *can be efficiently approximated up to additive error ϵ via Monte Carlo sampling with failure probability δ from $16 \deg_{XZ}(i)(\max_j |x_{i,j}|)^2 \log(2/\delta)/\epsilon^2$ many iid. samples.*

Proof. For the proof we will use concentration of measure for Lipschitz functions. To this end we begin by noticing that the sum (38) can be rewritten as a uniform expectation value over $k_i = \deg_{XZ}(i)$ many Rademacher (± 1) random variables as

$$\sum_{\lambda_{\mathcal{N}_{XZ}(i)}} \max \bigg\{ \alpha_i + \sum_{j \in \mathcal{N}_{XZ}(i)} (-1)^{\lambda_j} x_{i,j}, 0 \bigg\}$$
$$= \mathbb{E}_{\sigma \in \{\pm 1\}^{k_i}} [f_{\alpha,x}^{(i)}(\sigma)], \tag{40}$$

where we have defined

$$f_{\alpha,x}^{(i)} : \mathbb{R}^{k_i} \to \mathbb{R}$$
$$s \mapsto \max \bigg\{ \alpha_i + \sum_{j \in \mathcal{N}_{XZ}(i)} s_j x_{i,j}, 0 \bigg\}. \tag{41}$$

It can easily be seen that the function $f_{\alpha,x}^{(i)}$ is Lipschitz with constant $\big(\max_j |x_{i,j}| \big) k_i^{1/2}$:

$$\big| f_{\alpha,x}^{(i)}(s) - f_{\alpha,x}^{(i)}(s') \big| = \bigg| \sum_{j=1}^{\deg_{XZ}(i)} x_{i,j}\big(s_i - s'_i\big) \bigg| \tag{42}$$

$$\leq \big(\max_j |x_{i,j}| \big) \|s - s'\|_{\ell_1} \tag{43}$$

$$\leq \big(\max_j |x_{i,j}| \big) k_i^{1/2} \|s - s'\|_{\ell_2}. \tag{44}$$

Here, we have used the fact that the ℓ_p norms on $\mathbb{R}^n$ satisfy Moreover, $f_{\alpha,x}^{(i)}$ is clearly *separately convex*, that is, for each $k = 1, 2, \ldots, k_i$ the function $s_j \mapsto f_{\alpha,x}^{(i)}(s_1, s_2, \ldots, s_{j-1}, s_j, s_{j+1}, s_{j+2}, \ldots, s_n)$ is convex for each fixed $(s_1, s_2, \ldots, s_{j-1}, s_{j+1}, s_{j+2}, \ldots, s_n) \in \mathbb{R}^{k_i-1}$.

We can then apply [65, Theorem 3.4] to obtain that the estimator

$$\hat{f}_{\alpha,x}^{(i)} = \frac{1}{m} \sum_{l=1}^{m} f_{\alpha,x}^{(i)}(\sigma^{(l)}), \tag{45}$$

for the m Rademacher vectors $\sigma^{(l)} \in \{\pm 1\}^{k_i}$ drawn iid. uniformly at random satisfies

$$\mathbb{P}_\sigma \left[\big| \hat{f}_{\alpha,x}^{(i)} - \mathbb{E}_\sigma[f_{\alpha,x}^{(i)}\sigma] \big| \geq \epsilon \right] \leq 2\mathrm{e}^{-\frac{m\epsilon^2}{16 k_i (\max_j |x_{i,j}|)^2}}. \tag{46}$$

This implies that with probability $1 - \delta$ the estimator satisfies

$$\big| \hat{f}_{\alpha,x}^{(i)} - \mathbb{E}_\sigma[f_{\alpha,x}^{(i)}\sigma] \big| \leq \epsilon \tag{47}$$

whenever the number m of independently drawn Rademacher vectors satisfies

$$m \geq \frac{16 \, k_i (\max_j |x_{i,j}|)^2}{\epsilon^2} \log(2/\delta). \tag{48}$$

$\square$

In total we thus obtain a polynomial worst-case complexity of computing the non-stoquasticity of a $(2+1)$-local Hamiltonian of the form (34) up to additive error ϵ with failure probability δ as given by

$$\frac{n(n-1)}{2} + \frac{16 \sum_i \deg_{XZ}(i)(\max_{i,j} x_{i,j})^2}{(\epsilon/n)^2} \log\left(\frac{2}{\delta}\right). \tag{49}$$

S3.   EASING THE SIGN PROBLEM: AN ALGORITHMIC APPROACH

To demonstrate the feasibility of SignEasing and to put our findings more closely into the context of practical condensed matter problems, we numerically optimize the non-stoquasticity of certain nearest-neighbour Hamiltonians in quasi one-dimensional ladder geometries. Such systems are effectively described by translation-invariant Hamiltonians on n d-dimensional quantum systems of the form (32) with

nearest-neighbour interaction term $h \in \mathbb{R}^{d^2 \times d^2}$. For the sake of simplicity, we specialize here to closed boundary conditions, identifying $n + 1 = 1$.

We then optimize the non-stoquasticity of H over on-site orthogonal basis choices of the type

$$H \mapsto O^{\otimes n} H (O^T)^{\otimes n}. \tag{50}$$

On-site transformations are particularly simple to handle as they preserve both locality and translation-invariance of the Hamiltonian. Due to the translation-invariance of the problem the global non-stoquasticity measure can be expressed locally in terms of the transformed term $O^{\otimes 2} h (O^T)^{\otimes 2}$ so that the problem has constant complexity in the system size. More precisely, for Hamiltonians of the form (32) we can express the non-stoquasticity measure $\nu_1(H) = n 2^{n-3} \tilde{\nu}_1(h)$ in terms of an effective local measure

$$\tilde{\nu}_1(h) = \sum_{\substack{ijk;lmn: \\ j \neq m, k = n}} \max \Big\{ 0, (h \otimes \mathbb{1} + \mathbb{1} \otimes h)_{ijk;lmn} \Big\}. \tag{51}$$

Optimizing $\nu_1(H)$ for the global Hamiltonian is therefore equivalent to the much smaller problem of minimizing $\tilde{\nu}_1(h)$. While the non-stoquasticity measure ν_1 can be efficiently evaluated, thus satisfying a necessary criterion for an efficient optimization algorithm, minimizing ν_1 may and in fact *will* still be a non-trivial task in general – an intuition we make rigorous below. This is because in optimizing the basis-dependent measure ν_1 over quasi-local basis choices one is faced with a highly non-convex optimization problem of a high-order polynomial over the sphere of orthogonal matrices. Among the best developed multi-purpose methods for optimization over group manifolds such as the orthogonal group are conjugate gradient descent methods [44]. Compared to simple gradient-descent algorithms, conjugate gradient algorithms are capable to better incorporate the underlying group structure to the effect that they achieve much faster runtimes and better convergence properties.

To practically minimize the non-stoquasticity $\tilde{\nu}_1$ over the orthogonal group $O(d)$ we have implemented a conjugate gradient descent algorithm following Ref. [44]. Our implementation is publicly available [45] and detailed in Appendix A.

We first benchmark the algorithm on Hamiltonians which we know to admit an on-site stoquastic basis. Specifically, we apply the algorithm to recover an on-site stoquastic basis of the random translation-invariant Hamiltonian

$$H = \sum_{i=1}^{n} T_i \left(O^{\otimes 2} (h - h_\neg)(O^T)^{\otimes 2} \right) \tag{52}$$

on n qudits where the two-local term $h \in \mathbb{R}^{d^2 \times d^2}$ is a Hamiltonian term with uniformly random spectrum expressed in a Haar-random basis and $O \in O(d)$ is a Haar-random on-site orthogonal matrix. In Fig. 1(a) we show the performance of the algorithm on randomly chosen instances of (52) for different values of the local dimension d. In all but very few instances our algorithm essentially recovers the stoquastic basis of the random Hamiltonian. By construction, this can only
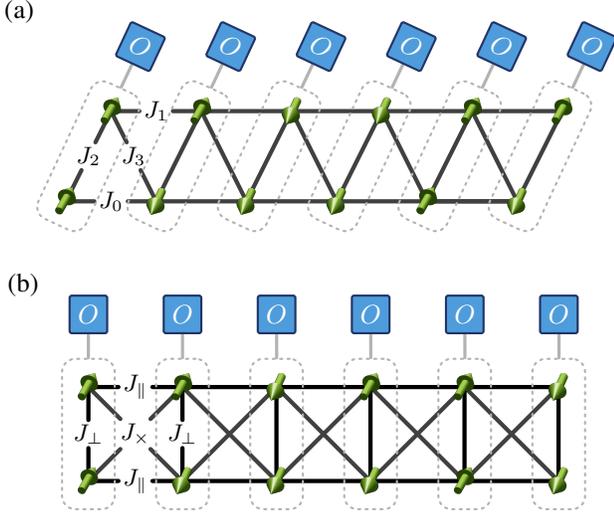
Figure S2. Quasi one-dimensional models with a sign problem. Figure *(a)* shows the lattice of the J_0-J_1-J_2-J_3-Heisenberg model on a triangular quasi one-dimensional lattice as given in Eq. (53). Figure *(b)* shows the lattice structure of the frustrated Heisenberg model (54) with couplings $J_\perp$, $J_\parallel$ and $J_\times$ on a square-lattice ladder with cross coupling. In our simulations, we group sites to dimers as indicated in the figures and then optimize the measure $\tilde{\nu}_1(h)$ of the effective 2-local terms h over on-site orthogonal transformations $O^{\otimes 2}$.

be due to the fact that the algorithm gets stuck in local minima, indicating a potential limitation of first-order optimization techniques as a tool for easing the sign problem of general Hamiltonians.

We then study frustrated anti-ferromagnetic Heisenberg Hamiltonians, i.e., Hamiltonians with positively weighted interaction terms $\vec{S}_i \cdot \vec{S}_j$, on different ladder geometries. Here, $\vec{S}_i = (X_i, Y_i, Z_i)^T$ is the spin operator at site i. The sign problem of frustrated ladder systems can in many cases actually be removed by going to a dimer basis [11, 13, 14]. However – and this is important for our approach – in those cases the sign problem is not removed by finding a *stoquastic local basis*, but rather by exploiting specific properties of the Monte Carlo simulation at hand, for example, that no negative paths occur in the simulation [11] or by exploiting specific properties of the Monte Carlo implementation at hand [13, 14]. Therefore, frustrated Heisenberg ladders constitute the ideal playground to explore the methodology of easing the sign problem by (quasi-)local basis choices.

The first model we study is the J_0-J_1-J_2-J_3-model [11] whose Hamiltonian is given by (see Fig. S2(a))

$$H_{\vec{J}} = \sum_{i=1}^{n} \left(J_0 \vec{S}_i^1 \vec{S}_{i+1}^1 + J_1 \vec{S}_i^2 \vec{S}_{i+1}^2 + J_2 \vec{S}_i^1 \vec{S}_i^2 + J_3 \vec{S}_{i+1}^1 \vec{S}_i^2 \right),$$

(53)

where $\vec{S}_i^1$ denotes the spin operator at site i on the lower rung and $\vec{S}_i^2$ on the upper rung of the ladder, respectively, and $J_i \geq 0$ for all i. Intriguingly, this Hamiltonian does

not have a sign problem in the singlet-triplet dimer basis even though the Hamiltonian is not stoquastic in that basis. However, there exists a *non-local* stoquastic basis for values of $J_3 \geq J_0 + J_1$ [11]. We show the results of optimizing the non-stoquasticity of $H_{\vec{J}}$ with $J_0 = J_1 = J$ over a translation-invariant dimer basis (see Fig. S2(a)) in Fig. 1(b). We initialize our simulations in a Haar random orthogonal on-site basis. Interestingly, we find an improvement of the non-stoquasticity under on-site orthogonal basis choices that does not seem to correlate with the region in which a non-local stoquastic basis was found in Ref. [11]. We view this as an indication that less local ansatz classes such as quasi-local circuits can further improve the non-stoquasticity for this model.

We now apply the algorithm to the anti-ferromagnetic Heisenberg ladder studied in Refs. [13, 14]. The Hamiltonian of this system is given by (see Fig. S2(b))

$$H_{J_\parallel, J_\perp, J_\times} = \sum_{i=1}^{n} \left(J_\parallel \left(\vec{S}_i^1 \vec{S}_{i+1}^1 + \vec{S}_i^2 \vec{S}_{i+1}^2 \right) + J_\perp \vec{S}_i^1 \vec{S}_i^2 \right.$$
$$\left. + J_\times \left(\vec{S}_i^1 \vec{S}_{i+1}^2 + \vec{S}_{i+1}^1 \vec{S}_i^2 \right) \right),$$

(54)

with interaction constants $J_\parallel, J_\perp, J_\times \geq 0$. For this geometry, the situation is somewhat more involved: Refs. [13, 14] find that a sign-problem free QMC procedure exists, albeit for a slightly different QMC procedure than we consider here, namely stochastic series expansion (SSE) Monte Carlo [36]. Similar to the world-line Monte Carlo method discussed here, SSE is based on an expansion of the exponential $\exp(-\beta H)$ albeit via a Taylor expansion as opposed to a product expansion. Specifically, for the partially frustrated case in which $J_\times \neq J_\parallel$ their solution of the sign problem exploits a specific sublattice structure of the Hamiltonian in combination with the SSE approach. We optimize the non-stoquasticity of dimer basis choices as shown in Fig. S2(b) when starting from a random initial point that is close to the identity. Our results, shown in Fig. 1(c), qualitatively reflect the findings of Wessel *et al.* [14] for SSE in terms of stoquasticity in that the non-stoquasticity can be significantly reduced for the fully frustrated case $J_\parallel = J_\times$, while it can be merely slightly improved for the partially frustrated case.

At the same time, the algorithm does not recover a fully stoquastic basis for the frustrated ladder model $H_{J_\parallel, J_\perp, J_\times}$ as might be expected. There may be several reasons for this: either the nearly sign-problem free QMC procedure found in Refs. [13, 14] is in fact specific to SSE in that no stoquastic dimer basis and hence no sign-problem free world-line Monte Carlo method exists in the orbit of orthogonal dimer bases, or the conjugate gradient algorithm gets stuck in local minima. In any case, the performance of our algorithm for both frustrated ladders demonstrates that the optimization landscape is generically an extremely rugged one, reflecting the computational hardness of the optimization problem in general.

We now turn to showing the improvement of the average sign concomitant with the improvement in non-stoquasticity in Fig. S3. We first observe that Figs. S3(a) and (b) are compatible with an exponential dependence of the inverse aver-
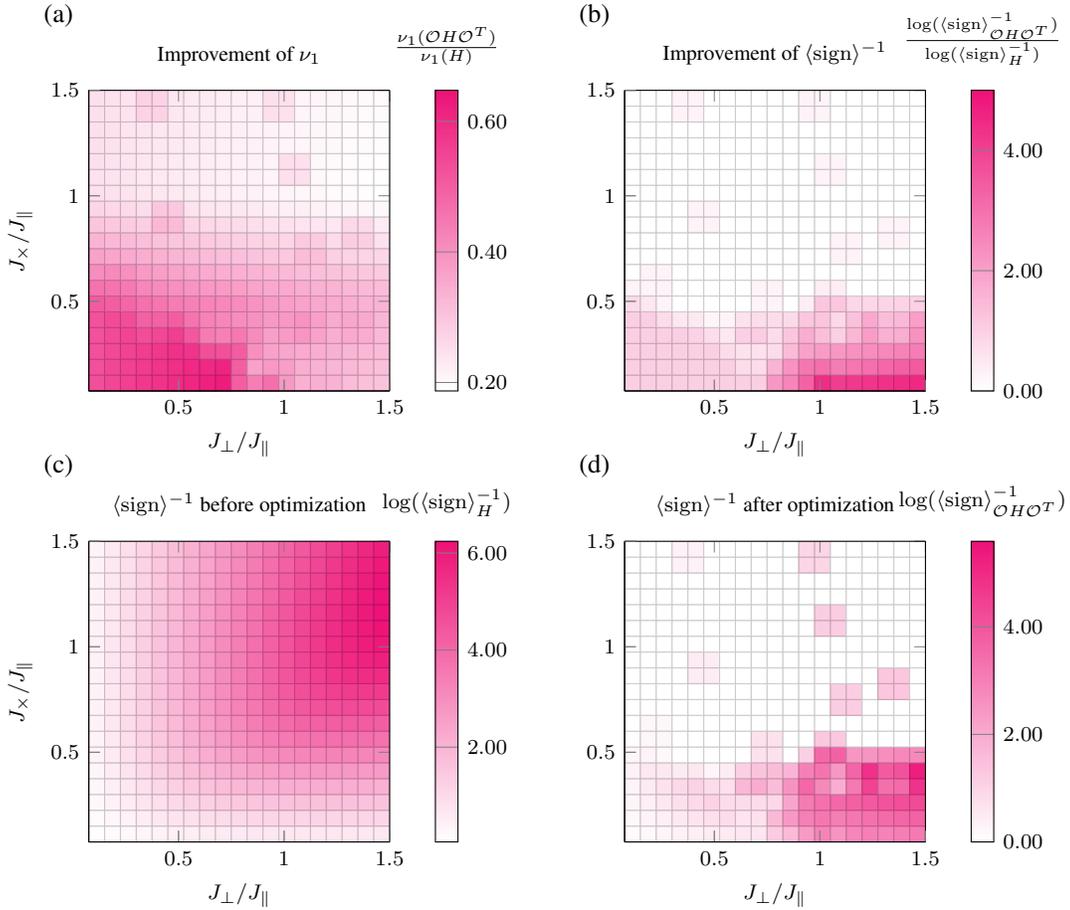
Figure S3. Improvement of the average sign concomitant with an improvement of the non-stoquasticity of the frustrated Heisenberg ladder (54). Figure *(a)* shows the optimized non-stoquasticity ν_1 in terms of its relative improvement compared to the computational basis. *(b)* We expect the inverse average sign to scale exponentially in the non-stoquasticity. Therefore, we plot the ratio of the logarithm of the inverse average sign before optimization to that after optimization. We compute the average sign via exact diagonalization for a ladder of 2×4-sites, $m = 100$ Monte Carlo steps and inverse temperature $\beta = 1$. We also plot the logarithm of the inverse average sign *(c)* before and *(d)* after optimization of a local orthogonal basis.

age sign on the non-stoquasticity $\langle \text{sign} \rangle^{-1} \propto \exp(c\nu_1(H))$ as conjectured above: in the regions in which a significant improvement of the non-stoquasticity could be achieved by local basis choices, the inverse average sign could also be strongly improved. Importantly, while the Hamiltonian could not be made entirely stoquastic, the improvement in the inverse average sign reaches an extent to which nearly no sign problem remains in those regions. This shows that also moderate improvements in non-stoquasticity can yield tremendous improvements of the average sign. At the same time, a severe sign problem remains – and actually becomes worse – in a small region of the parameter space (around $J_\perp/J_\parallel \gtrsim 3/4$ and $J_\times/J_\parallel \lesssim 1/2$) even though the non-stoquasticity could be reduced to some extent in that region. This may reflect open questions about the relation between average sign and non-stoquasticity that arose in our earlier discussion in Section S2: while in generic cases the two notions of severeness of the sign problem are expected to be closely related, there is no general simple correspondence between them.

Our findings demonstrate both the feasibility of minimizing the non-stoquasticity in order to ease the sign problem by optimizing over suitably chosen ansatz classes of unitary/orthogonal transformations and potential obstacles to a universal solution of the sign problem. In particular, for translation-invariant problems – while it may well be computationally infeasible – the complexity of the optimization problem only scales with the locality of the Hamiltonian, the local dimension and the depth of the circuit. We expect, however, that there exists no algorithm with polynomial runtime in all of these parameters that always solves the optimisation problem.

Our findings also indicate that more general ansatz classes yield the potential to further improve non-stoquasticity. Different classes of orthogonal transformations can be straightforwardly incorporated in our algorithmic approach. A detailed study of different ansatz classes and their potential for easing the sign problem is, however, beyond the scope of this work. It is the subject of ongoing and future efforts to study the optimal basis choice in terms of the non-stoquasticity for

both deeper circuits and further models as well as the connection between the average sign and the non-stoquasticity.

S4. EASING THE SIGN PROBLEM: COMPUTATIONAL COMPLEXITY

Let us now focus on a more fundamental question, namely, how far an approach that optimizes the non-stoquasticity can carry in principle. We have explored the potential of easing using state-of-the-art optimization algorithms; let us now turn to its boundaries, a glimpse of which we have already witnessed in the shape of a rugged optimization landscape. The method of choice for this task is the theory of computational complexity.

We analyze the computational complexity of easing the sign problem under particularly simple basis choices, namely, real on-site Clifford and orthogonal transformations. In both cases, we show that easing the sign problem is an NP-complete task even in cases in which deciding whether the sign problem can be cured is efficiently solvable [21], namely for XYZ Hamiltonians as given by

$$H_{\mathrm{XYZ}} = \sum_{i<j} \left(a_{i,j} X_i X_j + b_{i,j} Y_i Y_j + c_{i,j} Z_i Z_j \right). \quad (55)$$

Like Refs. [18–21], we allow for arbitrary interaction graphs.

A central ingredient in proving Theorem 2 is an expression for the non-stoquasticity measure ν_1 of strictly 2-local Hamiltonians of the form

$$H_2 = \sum_{i<j} \left(a_{i,j} X_i X_j + c_{i,j} Z_i Z_j + x_{i,j} X_i Z_j + x_{j,i} Z_i X_j \right). \quad (56)$$

It is sufficient to restrict to Hamiltonians of the form (56) because the orbit of XYZ Hamiltonians under on-site orthogonal (Clifford) transformations does not reach YY terms.

It is a direct consequence of Lemma 5 that

$$\nu_1(H_2) = \sum_{i<j} \nu_1(a_{i,j} X_i X_j) + \sum_i \nu_1 \left(\sum_{j \in \mathcal{N}_{XZ}(i)} x_{i,j} X_i Z_j \right), \quad (57)$$

where the XZ neighbourhood $\mathcal{N}_{XZ}(i)$ of a vertex i and related notions are defined in Sec. S2 D. More specifically, following Eqs. (37) and (38) we find that

$$\nu_1(a_{i,j} X_i X_j) = \sum_{i<j} \max\{a_{i,j}, 0\}, \quad (58)$$

$$\nu_1 \left(\sum_{j \in \mathcal{N}_{XZ}(i)} x_{i,j} X_i Z_j \right) = 2^{-\deg_{XZ}(i)}$$
$$\times \sum_{\lambda_{\mathcal{N}_{XZ}(i)}} \max \left\{ \sum_{j \in \mathcal{N}_{XZ}(i)} (-1)^{\lambda_j} x_{i,j}, 0 \right\}. \quad (59)$$

Since for the proof of hardness we need analytical expressions of the non-stoquasticity, we cannot resort to the sampling algorithm to evaluate the non-stoquasticity of XZ terms

as proposed in Sec. S2 D. We analytically bound the contribution of a vertex with non-trivial XZ neighbourhood with the following lemma.

Lemma 7 (XZ non-stoquasticity). *The following bound is true for $k \in \mathbb{N}$*

$$\sum_{\lambda \in \{0,1\}^k} \max \left\{ \sum_{j=1}^k (-1)^{\lambda_j} x_j, 0 \right\} \geq \max_j |x_j| \cdot 2^{k-1}. \quad (60)$$

Proof. Let us assume wlog. that $x_1 \geq x_2 \geq \ldots \geq x_k \geq 0$, all terms being positive and non-increasingly ordered. This does not restrict generality as all possible combinations of signs appear in the sum (60). We prove the claim by induction. For $k = 1$, the statement is true by immediate inspection. For the induction step, we use the following inequality for $a, b \in \mathbb{R}$

$$\max\{a + b, 0\} + \max\{a - b, 0\} \geq 2 \max\{a, 0\}, \quad (61)$$

which can be easily checked by checking the three cases $a \geq |b|$, $a \leq -|b|$ and $-|b| < a < |b|$. We then calculate

$$\sum_{\lambda \in \{0,1\}^k} \max \left\{ \sum_{j=1}^k (-1)^{\lambda_j} x_j, 0 \right\} \quad (62)$$

$$= \sum_{\lambda_1,\ldots,\lambda_{k-1} \in \{0,1\}} \max \left\{ x_k + \sum_{j=1}^{k-1} (-1)^{\lambda_j} x_j, 0 \right\}$$

$$+ \sum_{\lambda_1,\ldots,\lambda_{k-1} \in \{0,1\}} \max \left\{ -x_k + \sum_{j=1}^{k-1} (-1)^{\lambda_j} x_j, 0 \right\}. \quad (63)$$

$$\geq 2 \sum_{\lambda' \in \{0,1\}^{k-1}} \max \left\{ \sum_{j=1}^{k-1} (-1)^{\lambda'_j} x_j, 0 \right\} \quad (64)$$

$$\overset{\mathrm{I.H.}}{\geq} 2 \cdot 2^{k-2} |x_1| = 2^{k-1} |x_1|, \quad (65)$$

where we have used (61) in the second to last and the induction hypothesis in the last step. This proves the claim. $\square$

In the proof of Theorem 2 we will use that Lemma 5 implies that every term $a_{i,j} X_i X_j$ contributes an additional cost $\max\{a_{i,j}, 0\}$ to the non-stoquasticity of H_2. Moreover, since $\max_{j \in [k]} |x_j| \geq \sum_{j=1}^k |x_j|/k$, Lemmas 5 and 7 imply that we can view every term $x_{i,j} X_i Z_j$ of H_2 as contributing at least a cost $|x_{i,j}|/(2 \deg(G))$ to the non-stoquasticity of H_2, where G is the interaction graph of H_2.

We are now ready to show that with respect to the non-stoquasticity measure ν_1 easing the sign problem for 2-local XYZ Hamiltonians with on-site Cliffords is NP-complete on arbitrary graphs. We restate Theorem 2i here.

Theorem 8 (SignEasing under orthogonal Clifford transformations). *SignEasing is* NP-*complete for 2-local Hamiltonians on an arbitrary graph, in particular for XYZ Hamiltonians, under on-site orthogonal Clifford transformations, that is, the real group generated by $\{X, Z, W\}$ with W the Hadamard matrix.*

Proof. Clearly the problem is in NP, since one can simply receive a (polynomial-size) description of the transformation in the Yes-case, and then calculate the measure of non-stoquasticity efficiently for XYZ Hamiltonians, verifying the solution.

To prove NP-hardness, we encode the MAXCUT problem in the SignEasing problem. A MAXCUT instance can be phrased in terms of asking whether an anti-ferromagnetic (AF) Ising model on a graph $G = (V, E)$ with $e = |E|$ edges on $v = |V|$ spins

$$H = \sum_{(i,j) \in E} Z_i Z_j, \qquad (66)$$

has ground-state energy $\lambda_{\min}(H)$ below A or above B with constants $B - A \geq 1/\mathsf{poly}(v)$. This is because in the Ising model one gets energy -1 for a $(0,1)$ or $(1,0)$ -edge and $+1$ for a $(0,0)$ or $(1,1)$ edge.

Let us now encode the MAXCUT problem phrased in terms of the AF Ising model problem into SignEasing for the XYZ Hamiltonian. We will design a Hamiltonian H', and ask if on-site orthogonal Clifford transformations can decrease its measure of non-stoquasticity ν_1 below A, or whether it remains above B for any Clifford basis choice.

For each AF edge between spins i, j in the AF Ising model, the new Hamiltonian H' will have an edge

$$h_{i,j} = X_i X_j. \qquad (67)$$

On top of that, for every edge $(i,j) \in E$ we add one ancilla qubit $\xi_{i,j}$ as shown in Figure 3, and interactions

$$h_{i,j}^{(\xi)} = C \left(Z_i Z_j - Z_i Z_{\xi_{i,j}} - Z_{\xi_{i,j}} Z_j \right), \qquad (68)$$

where $C = 4 \deg(G)$. Note that the additional terms are diagonal and hence stoquastic. The total new Hamiltonian then reads

$$H' = \sum_{(i,j) \in E} \left[X_i X_j + C \left(Z_i Z_j - Z_i Z_{\xi_{i,j}} - Z_{\xi_{i,j}} Z_j \right) \right], \qquad (69)$$

and acts on $n = v + e$ qubits. We construct H' so that an attempt to decrease the non-stoquasticity ν_1 by swapping Z and X operators via Hadamard transformations will fail, and so the best one can do is to choose a sign in front of each local X operator. Of course, this then becomes the original, hard, MAXCUT problem in disguise. Let us prove this.

Let $\mathcal{N}'(i) = \{j : (i,j) \in E'\} = \mathcal{N}(i) \cup \{\xi_{i,j} : (i,j) \in E\}$ be the neighbourhood of site i on the augmented graph $G' = (V', E')$ on which H' lives. We start the proof by observing that the degree $\deg'(i) = |\mathcal{N}'(i)|$ of a site i on the augmented graph satisifies $\deg'(i) = 2 \deg(i)$, and hence $\deg'(G) = 2 \deg(G)$.

a. Orthogonal Clifford transformations First, let us note that any element of the orthogonal single-qubit Clifford group can be written as

$$C = \pm W^w X^x Z^z, \qquad (70)$$

where we denote the Hadamard matrix with W and $w, x, z \in \{0, 1\}$. Since the global sign is irrelevant, a real n-qubit Clifford of the form $C = C_1 \otimes \cdots \otimes C_n$ is parametrized by binary vectors $\vec{w}, \vec{x}, \vec{z} \in \{0, 1\}^n$.

How does H' transform under real single-qubit Clifford transformations? By definition $CZC^\dagger \in \{\pm Z, \pm X\}$ and likewise for X. Therefore, the transformed Hamiltonian will be of the form (56). Throughout the proof, we will use that every term $a_{i,j} X_i X_j$ contributes at least $\max\{a_{i,j}, 0\}$ to the non-stoquasticity, while every term $x_{i,j} X_i Z_j$ contributes at least $|x_{i,j}|/(2 \deg(G')) = |x_{i,j}|/(4 \deg(G))$ as shown by Lemmas 5 and 7 above.

We now show that MAXCUT can be embedded into SignEasing under on-site orthogonal Clifford transformations. To do so, we need to show two things: first, that in the yes-case that $\lambda_{\min}(H) \leq A$, the non-stoquasticity of H' can be brought below A using on-site orthogonal Clifford transformations. Second, we show that in the no-case that $\lambda_{\min}(H) \geq B$, the non-stoquasticity of H' cannot be brought below B using on-site orthogonal Clifford transformations.

b. Yes-case: (Diagonal) transformations that map X to $\pm X$ ($\vec{w} = 0$). These transformations only change the sign in front of X_i, keeping its form. At the same time they only change the signs of the $Z_i Z_j$ terms, keeping them diagonal and hence stoquastic. The transformed $X_i X_j$ terms (67) will be stoquastic if and only if exactly one of the transformations of the X at sites i, j is a Z-flip.

We can view the coefficient z_i as a spin s_i in the original AF Ising model: for $z_i = 1$, $X_i \to -X_i$, corresponding to a spin $s_i = 1$ in the original AF Ising model, while for $z_i = 0$, $X_i \to X_i$, which we view it as the Ising spin $s_i = 0$.
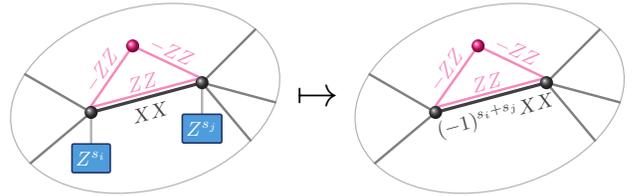
Each such Clifford transformation thus corresponds to a particular state of the original AF Ising model as given by a spin configuration $\vec{s} \in \{0, 1\}^v$. Whenever the transformations on neighbouring sites result in a stoquastic interaction $-X_i X_j$ in the transformed XYZ Hamiltonian, we have a $(0, 1)$ or $(1, 0)$ anti-ferromagnetic Ising edge with cost 0. On the other hand, each non-stoquastic $X_i X_j$ term in the XYZ Hamiltonian has cost 1, while the corresponding edge in the Ising model is $(0, 0)$ or $(1, 1)$ also with cost 1.

What is the amount of sign easing we can hope to achieve? We have argued above that only diagonal transformations which map $X_i \mapsto \pm X_i$ potentially ease the sign problem since we designed the interactions so that a Hadamard transformation always incurs a larger cost than keeping an $X_i X_j$ term non-stoquastic. For those transformations, we have a one-to-one correspondence with the ground state of the original AF Ising model. Hence, the original AF Ising model ground state energy $\lambda_{\min}(H)$ is also the optimal number of

non-stoquastic terms $X_i X_j$ which one can achieve via on-site orthogonal Clifford transformations, each adding an additional cost 1 to the non-stoquasticity measure ν_1.

In the yes-case we can therefore achieve non-stoquasticity

$$\nu_1(yes) \le A, \tag{71}$$

by choosing $\vec{x}, \vec{w} = 0$ and $(z_1, \ldots, z_v)^T = \vec{s}_0$, the ground state of H.

We now show that in the no-case, the non-stoquasticity measure will be at least

$$\nu_1(no) \ge B. \tag{72}$$

c. No-case: (Hadamard) transformations that map X to $\pm Z$ ($\vec{w} \ne 0$). We have designed the additional Hamiltonian term (68) so that such transformations result in large non-stoquasticity. Specifically, we show that for any choice of $\vec{z}$, choosing $\vec{x} = \vec{w} = 0$ achieves the optimal non-stoquasticity in the orbit of orthogonal Clifford transformations.

It is sufficient to show that any Clifford transformation on an edge (i, j) (and its ancilla qubit $\xi_{i,j}$) that is non-stoquastic for a given choice of $\vec{z}$ can only increase the non-stoquasticity.

To begin with, note that choosing $x_i = 1$ results in $Z_i \mapsto -Z_i$, $X_i \mapsto X_i$, and choosing $w_i = 1$ maps $X_i \mapsto Z_i$ and $Z_i \mapsto X_i$. We obtain the following transformation rules of Pauli Z_i and X_i, given choices of x_i and w_i:

x_i	w_i	Z_i	X_i
0	0	Z_i	X_i
0	1	X_i	Z_i
1	0	$-Z_i$	X_i
1	1	$-X_i$	Z_i

First, suppose that for an edge (i, j), a Hadamard transformation is performed on qubit i, but not j so that we have $w_i = 1$, $w_j = 0$. Then for some choice of x_i, x_j the transformed edge is given by
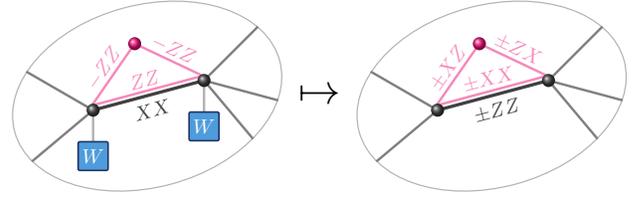
$$W_i(X_i X_j \pm C Z_i Z_j)W_i = Z_i X_j \pm C X_i Z_j, \tag{73}$$
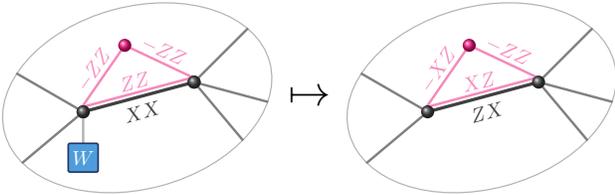
and has non-stoquasticity at least $(C + 1)/(2 \deg G')$.

Now, suppose that a Hadamard transformation is performed on both qubit i and j but not its ancilla qubit. Then the transformed term is given by

$$W_i W_j X_i^{x_i} X_j^{x_j} X_{\xi_{i,j}}^{x_{\xi_{i,j}}} (h_{i,j} + h_{i,j}^{(\xi)}) X_i^{x_i} X_j^{x_j} X_{\xi_{i,j}}^{x_{\xi_{i,j}}} W_i W_j$$
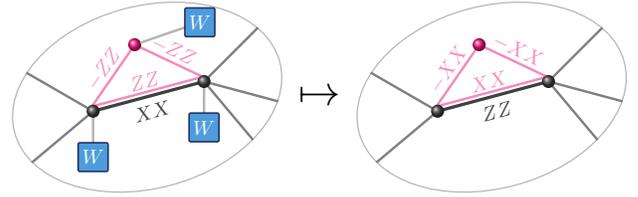$$= \pm Z_i Z_j + C \left(\pm X_i X_j \pm X_i Z_{\xi_{i,j}} \pm Z_{\xi_{i,j}} X_j \right), \tag{74}$$

with non-stoquasticity cost at least

$$\nu_1 \left(C(\pm X_i Z_{\xi_{i,j}} \pm Z_{\xi_{i,j}} X_j) \right) = 2C/(2 \deg G'). \tag{75}$$

Could the edge be possibly cured by performing a Hadamard transformation on the ancilla qubit as well? In this case, we get

$$W_i W_j W_{\xi_{i,j}} (h_{i,j} + h_{i,j}^{(\xi)}) W_i W_j W_{\xi_{i,j}}$$
$$= Z_i Z_j + C \left(X_i X_j - X_i X_{\xi_{i,j}} - X_{\xi_{i,j}} X_j \right), \tag{76}$$

with non-stoquasticity

$$\nu_1(W_i W_j W_{\xi_{i,j}} (h_{i,j} + h_{i,j}^{(\xi)}) W_i W_j W_{\xi_{i,j}}) = C. \tag{77}$$

Because of the frustrated arrangement of the signs of the ZZ terms, no local sign flip of those terms (achieved by choices of $x_i, x_j, x_{\xi_{i,j}} \ne 0$) can cure the sign problem of an ancillary triangle, leaving it lower bounded by (77).

On the other hand, the original cost incurred from local sign flips via Z-transformations, is given by

$$\nu_1(X_i X_j) = 1, \tag{78}$$

which is always smaller than the cost incurred if additional X or W transformations are applied since we chose C such that $C/(2 \deg G') = 1$. Therefore, in the no-case of the original AF Ising model the non-stoquasticity of H' cannot be brought below

$$\nu_1(no) \ge B, \tag{79}$$

with the optimal choice achieved for $\vec{x}, \vec{w} = 0$ and $(z_1, \ldots, z_v)^T = \vec{s}_0$.

$\square$

Theorem 9 (SignEasing under orthogonal transformations). *SignEasing is NP-complete for 2-local Hamiltonians on an arbitrary graph, in particular, for XYZ Hamiltonians under on-site orthogonal transformations.*

Proof. We proceed analogously to the proof for orthogonal Clifford transformations, showing that in the yes-case, there exists a product orthogonal transformation $O = O_1 \cdots O_n$ such that $\nu_1(OH'O^T) \leq A$, while in the no-case there exists no such transformation P with $\nu_1(PH'P^T) \leq B$.

The yes-case is clear: In this case, the energy of $\vec{s}$ is below A. Then by our construction, the sign problem can be eased below A with the orthogonal transformation

$$O = \prod_{i \in V} Z_i^{s_i}. \tag{80}$$

We now need to show that in the no-case, any orthogonal transformation incurs non-stoquasticity above B. We first remark that the orthogonal group $O(2)$ decomposes into two sectors with determinant ± 1, respectively. Therefore, any 2×2 orthogonal matrix can be written as

$$O_a(\theta) = \begin{pmatrix} \cos\theta & -a\sin\theta \\ \sin\theta & a\cos\theta \end{pmatrix}, \tag{81}$$

which for $a = \det(O_a(\theta)) = -1$ is a reflection and for $a = \det(O_a(\theta)) = +1$ a rotation by an angle θ. Note that the following composition laws hold

$$O_{-1}(\theta)O_1(\phi) = O_{-1}(\theta - \phi), \tag{82}$$
$$O_1(\theta)O_{-1}(\phi) = O_{-1}(\theta + \phi), \tag{83}$$
$$O_1(\theta)O_1(\phi) = O_1(\theta + \phi), \tag{84}$$
$$O_{-1}(\theta)O_{-1}(\phi) = O_1(\theta - \phi). \tag{85}$$

Now observe three facts: First, any reflection by an angle θ can be written as a product of a reflection across the X-axis and a rotation as

$$\begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} Z = R(\theta)Z, \tag{86}$$

where $R(\theta)$ is the rotation by an angle θ. Second, for any Hermitian matrix H and any angle θ, it holds that

$$O(\theta)HO(\theta)^T = O(\theta + \pi)HO(\theta + \pi)^T, \tag{87}$$

so that it suffices to restrict to angles $\theta \in [-\pi/4, 3\pi/4]$ in an interval of length π. Third, a rotation by an angle $\pi/2$ can be decomposed into two reflections as

$$R(\pi/2) = XZ. \tag{88}$$

Taken together, these facts imply that an arbitrary single-qubit orthogonal transformation is given by

$$O(\theta, z, p) = R\left(\theta + (\pi/2)^p\right) \cdot Z^z = R(\theta) \cdot X^p Z^{p+z}, \tag{89}$$

where the rotation angle is given by $\theta \in [-\pi/4, \pi/4]$, $z \in \{0,1\}$ fixes whether or not a Z-flip is applied, and $p \in \{0,1\}$ mods out a rotation by an angle $\pi/2$. Now define $O(\vec{\theta}, \vec{z}, \vec{p}) = \prod_i O_i(\theta_i, z_i, p_i)$.

We now need to show that, in the no-case, for any choice of $\vec{\theta} \in [-\pi/4, \pi/4]^n$, $\vec{z}, \vec{p} \in \{0,1\}^n$ it holds that
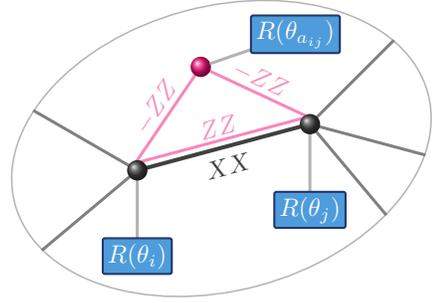
$$\nu_1\left(O(\vec{\theta}, \vec{z}, \vec{p})H'O(\vec{\theta}, \vec{z}, \vec{p})^T\right) \geq b. \tag{90}$$

To complete the proof, we use that the action of $R(\theta)$ on Pauli-X and Z matrices is given by

$$R(\theta)ZR(\theta)^T = \cos(2\theta)Z + \sin(2\theta)X, \tag{91}$$
$$R(\theta)XR(\theta)^T = \cos(2\theta)X - \sin(2\theta)Z. \tag{92}$$

In Appendix B, we show that given any choice of $\vec{z}$ and $\vec{p}$, no choice of $\vec{\theta}$ can decrease the non-stoquasticity of an edge with non-zero non-stoquasticity below 1. That is, we analyze transformations of the following form:

We do this by using the standard form (89) of the local transformations in terms of X, Z and restricted rotation matrices with angles in $[-\pi/4, \pi/4]$. We split the proof into three parts, analyzing three different (continous) regions for rotation angle choices. The only difference in the construction when compared to the Clifford-case is that here we choose $C = (2 \deg G')^2$. This completes the proof. $\qquad \square$

Notice that since MAXCUT is NP-hard already on subgraphs of the double-layered square lattice [46], which has constant degree, there are hard instances of the SignEasing problem on low-dimensional lattices with constant coupling strength.

We remark that one can easily extend the proofs of Theorems 8 and 9 for different non-stoquasticity measures ν_p with $1 < p < \infty$. To see this, note that the decision problem for $\nu_p(H)$ is equivalent to the problem for $\nu_p(H)^p$.

For terms $x_{i,j}X_iZ_j$ we can then use the trivial bound

$$\sum_{\lambda \in \{0,1\}^k} \max\left\{\sum_{j=1}^{k}(-1)^{\lambda_j}x_{i,j}, 0\right\}^p \geq 2^{p(k-\deg(i))}\sum_j |x_j|^p \tag{93}$$

instead of Lemma 7. Thus, every term $x_{i,j}X_iZ_j$ contributes at least $2^{-p\deg G'}|x_{i,j}|^p$ to ν_p^p, while a term $a_{i,j}X_iX_j$ contributes

$$\nu_p(a_{i,j}X_iX_j)^p = (\max\{a_{i,j}, 0\})^p, \tag{94}$$

to the total non-stoquasticity of H'.

For general ℓ_p-non-stoquasticity measures ν_p one therefore need merely choose $C = 2^{\deg(G')} = 4^{\deg(G)}$ to prove Theorem 8 and $C = 2^{2\deg(G')} = 16^{\deg(G)}$ for Theorem 9.

Appendix A: Conjugate gradient descent for sign easing

In this appendix, we provide details on the numerical implementation [45] of the conjugate gradient descent algorithm for minimizing the non-stoquasticity ν_1 over orthogonal circuits. In this work, we focus on the ansatz class of translation-invariant on-site orthogonal transformations, but other classes such as constant-depth circuits can be implemented analogously.

1. Translation-invariant formulation of the non-stoquasticity measure

We begin by deriving a simple expression for the non-stoquasticity measure for translation-invariant, one-dimensional, nearest-neighbour Hamiltonians. Our formulation is based on the observation that by translation-invariance, the measure $\nu_1(H)$ only depends on the local Hamiltonian term h and is therefore independent of the system size. Since Hamiltonian terms with overlapping support can contribute to the same matrix element in the global Hamiltonian matrix it is thus sufficient to optimize certain sums of matrix elements of h rather than make h itself stoquastic – a condition required only by the stronger notion of term-wise sxtoquasticity.

Recall that we consider real nearest-neighbour Hamiltonians with closed boundary conditions on a line of n local systems with dimension d

$$H = \sum_{i=1}^{n} T_i(h), \tag{A1}$$

where $h \in \mathbb{R}^{d^2 \times d^2}$ is the local term and we use the notation $T_i(h) = \mathbb{1}^{\otimes(i-1)} \otimes h \otimes \mathbb{1}^{\otimes(n-i-1)}$. Closed boundary conditions identify $n + 1 = 1$.

Specifically, we can calculate

$$\nu_1(H) = n2^{n-3} \sum_{\substack{i,j,k;l,m,n: \\ k \neq l, m = n}} \max\left\{0, (h \otimes \mathbb{1} + \mathbb{1} \otimes h)_{i,k,m;j,l,n}\right\} \equiv n2^{n-3}\tilde{\nu}_1(h). \tag{A2}$$

This reduces the problem of minimizing $\nu_1(H)$ over on-site orthogonal bases to minimizing $\tilde{\nu}_1(h)$ over such bases. To derive Eq. (A2) we re-express the global non-stoquasticity measure as follows:

$$\nu_1(H) = \sum_{\substack{(i_1,\ldots,i_n) \neq \\ (j_1,\ldots,j_n)}} \max\left\{0, \langle i_1,\ldots,i_n| \sum_{l=1}^{n} T_l(h)|j_1,\ldots,j_n\rangle\right\} \tag{A3}$$

$$= \sum_{p=1}^{n} \sum_{\substack{i_1,\ldots,i_n, \\ j_p, j_{p+1}: \\ i_{p+1} \neq j_{p+1}}} \max\left\{0, \langle i_1,\ldots,i_n|(T_p(h) + T_{p+1}(h))|i_1,\ldots,i_{p-1},j_p,j_{p+1},i_{p+2},\ldots,i_n\rangle\right\} \tag{A4}$$

$$= 2^{n-3} \sum_{p=1}^{n} \sum_{\substack{i_p,\ldots,i_{p+2}, \\ j_p, j_{p+1}: \\ i_{p+1} \neq j_{p+1}}} \max\left\{0, \langle i_p,\ldots,i_{p+2}|(h \otimes \mathbb{1} + \mathbb{1} \otimes h)|j_p,j_{p+1},i_{p+2}\rangle\right\} \tag{A5}$$

$$= n2^{n-3} \sum_{\substack{i_1,\ldots,i_3, \\ j_1, j_2: \\ i_2 \neq j_2}} \max\left\{0, \langle i_1,i_2,i_3|(h \otimes \mathbb{1} + \mathbb{1} \otimes h)|j_1,j_2,i_3\rangle\right\}. \tag{A6}$$

In the first step, we have used that the condition $(i_1,\ldots,i_n) \neq (j_1,\ldots,j_n)$ implies that at least one of the indices differs. Summing over p, we can let this index be the $(p+1)^{\text{st}}$ one. To avoid double-counting, we then divide the sum over all strings which differ on at most two nearest neighbours into a sum over all strings which potentially differ on two nearest neighbours *left* of the $(p+1)^{\text{st}}$ index. Patches which differ on more than two nearest neighbours vanish for nearest-neighbour Hamiltonians. In the second step, we use that all terms with support left of the pth qubit vanish, since the basis strings differ on the $(p+1)^{\text{st}}$ qubit. In the last step, we use the translation-invariance again to account for the sum over p, incurring a factor n.

2. Gradient of the objective function

As an ansatz class we choose on-site orthogonal transformations in $O(d)$, where d is the dimension of a local constituent of the system. More precisely, we consider transformations of the type (see Eq. (50))

$$H = \sum_{i=1}^{n} T_i(h) \mapsto O^{\otimes n} H (O^T)^{\otimes n}, \tag{A7}$$

which locally amounts to

$$h \mapsto h(O) := (O \otimes O) h (O^T \otimes O^T). \tag{A8}$$

The key ingredient for the conjugate gradient descent algorithm is the derivative of the objective function $\tilde{\nu}_1(h(O))$ with respect to the orthogonal matrix O. The gradient is given by

$$\frac{\partial}{\partial O} \tilde{\nu}_1(h(O)) = \sum_{i,j} \frac{\partial \tilde{\nu}_1}{\partial h(i|j)} \cdot \frac{\partial h(i|j)}{\partial O}. \tag{A9}$$

We can further expand the terms of the global gradient (A9): in particular, we can express the gradient of the effective local terms as a conjugation $h(O) = \mathcal{C}(O) h \mathcal{C}^\dagger(O)$ of the local term h by the orthorgonal circuit $\mathcal{C}(O) = O \otimes O$. We will now derive expressions for the measure and the gradients that the algorithm has to evaluate.

a. The objective function gradient We first determine the gradient of the objective function. Since we will also make use of different measures ν_p as defined in Eq. (26), we write the objective function for different $p \geq 1$ as

$$\tilde{\nu}_p^p(h) = \sum_{\substack{i,j,k;l,m,n:\\ k \neq l, m=n}} \max\left\{ (h \otimes \mathbb{1} + \mathbb{1} \otimes h)_{i,k,m;j,l,n}, 0 \right\}^p. \tag{A10}$$

Then, the gradient of the objective function is given by

$$\left. \frac{\partial \tilde{\nu}_p^p}{\partial h(x|y)} \right|_{h(0)} = p \sum_{\substack{i,j,k;l,m,n:\\ k \neq l, m=n}} (|x\rangle\langle y| \otimes \mathbb{1} + \mathbb{1} \otimes |x\rangle\langle y|)_{i,k,m;j,l,n} \max\left\{ (h \otimes \mathbb{1} + \mathbb{1} \otimes h)_{i,k,m;j,l,n}, 0 \right\}^{p-1} \tag{A11}$$

b. Gradient of the transformed Hamiltonian We can expand the gradient of the transformed Hamiltonian term by the orthogonal matrix as

$$\frac{\partial h(i|j)}{\partial O} = \sum_{m,n} \frac{\partial \operatorname{adj}_h(\mathcal{C})(i|j)}{\partial \mathcal{C}(m|n)} \frac{\partial \mathcal{C}(m|n)}{\partial O,} \tag{A12}$$

where by $\operatorname{adj}_h(\mathcal{C})$ we denote the adjunction map $h \mapsto \mathcal{C} h \mathcal{C}^T$. The derivative of the adjoint action of $\mathcal{C}$ on h is given by

$$\frac{\partial \operatorname{adj}_h(\mathcal{C})(i|j)}{\partial \mathcal{C}(m|n)} = \sum_{k,l} \frac{\partial}{\partial \mathcal{C}(m|n)} \mathcal{C}(i|k) h(k|l) \mathcal{C}(j|l) \tag{A13}$$

$$= \sum_{k,l} [\delta_{m,i} \delta_{n,k} h(k|l) \mathcal{C}(j|l) + \mathcal{C}(i|k) h(k|l) \delta_{m,j} \delta_{n,l}] \tag{A14}$$

$$= \langle m|i\rangle\langle j|\mathcal{C} h^T|n\rangle + \langle m|j\rangle\langle i|\mathcal{C} h|n\rangle \tag{A15}$$

From this expression, we can directly read off its matrix form

$$\frac{\partial \operatorname{ad}_h(\mathcal{C})(i|j)}{\partial \mathcal{C}} = |i\rangle\langle j|\mathcal{C} h^T + |j\rangle\langle i|\mathcal{C} h. \tag{A16}$$

It remains to compute the gradient of the circuit with respect to the orthogonal matrix. We obtain with $\langle m| = \langle m_1| \otimes \langle m_2|$

$$\frac{\partial \mathcal{C}(m|n)}{\partial O(k|l)} = \sum_{i=1}^{2} \delta_{m_i,k} \delta_{n_i,l} O(m_1|n_1) O(m_2|n_2) = \sum_{i=1}^{2} \langle m_i|k\rangle\langle l|n_i\rangle O(m_1|n_1) O(m_2|n_2), \tag{A17}$$

which expressed in matrix form is then given by

$$\frac{\partial \mathcal{C}_{l_e}(m|n)}{\partial O} = |n_1\rangle\langle m_1|\langle m_2|O|n_2\rangle + |n_2\rangle\langle m_2|\langle m_1|O|n_1\rangle. \tag{A18}$$

3. Algorithmic procedure

We start our conjugate gradient algorithm either at the identity matrix (with or without a small perturbation) or a Haar-randomly chosen orthogonal matrix as indicated at the respective places in the main text.

Since the minimization of the $\tilde{\nu}_1$-measure is numerically not well behaved, we improve the performance of the algorithm in several ways. First, we observe that the measure $\tilde{\nu}_2$ given by the Frobenius norm of the non-stoquastic part of the Hamiltonian is numerically much better behaved. This is due to the ℓ_2-norm being continuously differentiable while the ℓ_1-norm is only subdifferentible. In particular, at its minima the gradient of the ℓ_1-norm is discontinuous and never vanishes. For this reason, rather than optimizing the ℓ_1-norm of the non-stoquastic part, we optimize a smooth approximation thereof as given by [66]

$$\nu_{1,\alpha}(H) := \sum_{i \neq j} f_\alpha\left(H_{i,j}\right), \tag{A19}$$

with $f_\alpha(x) = x + \alpha^{-1} \log(1 + \exp(-\alpha x))$.

To achieve the best possible performance, we then carry out a hybrid approach: First, we pre-optimize by minimizing $\tilde{\nu}_2$ using our conjugate gradient descent algorithm. Second, starting at the minimizer obtained in the Frobenius norm optimization, we minimize the smooth non-stoquasticity measure $\tilde{\nu}_{1,\alpha}$. We choose the value $\alpha = 50$, $\alpha = 100$ and $\alpha = 40$, for the random stoquastic Hamiltonians, the J_0-J_1-J_2-J_3-model, and the frustrated ladder model, respectively. We then compare the result to a direct minimization of the non-stoquasticity $\tilde{\nu}_{1,\alpha}$ starting from the original initial point and choose whichever of the minimizations performed best. The exact details of our optimization algorithms can be found at Ref. [45] together with code to reproduce the figures shown here. Our code framework can be easily adapted for optimization of other Hamiltonian models and more general circuit architectures.

Appendix B: Orthogonal transformations of the penalty terms (proof of Theorem 9)

Proof of Theorem 9 (continued). As in the proof for orthogonal Clifford transformations, we will show that for any given choice of Z-transformations one cannot further decrease the non-stoquasticity by exploiting the additional freedom offered by the full orthogonal group. Above, we have argued that applying an arbitrary orthogonal transformation at a single site can be reduced to applying $R(\theta) X^p Z^{p+z}$ with $\theta \in [-\pi/4, \pi/4]$ and $z, p \in \{0, 1\}$. We will now show that for any choice of $\vec{z}$ and $\vec{p}$, a rotation by angles $\vec{\theta} \in [-\pi/4, \pi/4]^n$ cannot decrease the non-stoquasticity any further.

Analogously to the proof for Clifford-transformations, we discuss all possible transformations by dividing them into different cases. In each case the non-stoquasticity of an uncured edge (i, j) and its ancilla qubit $\xi_{i,j}$ cannot be eased below its previous value of 1. The additional difficulty we encouter here is that the orthogonal group is continuous as opposed to the orthogonal Clifford group, which is a discrete and rather 'small' group.

Given a choice of $\vec{z}$, consider an edge (i, j) with a non-trivial contribution to ν_1 and its corresponding ancilla qubit $\xi_{i,j}$. We begin, supposing that $p_i = p_j = p_{\xi_{i,j}} = 0$ so that the X-flips act trivially on all three qubits. We now analyze the effect of the remaining rotations $R(\theta)$ on each of the qubits.
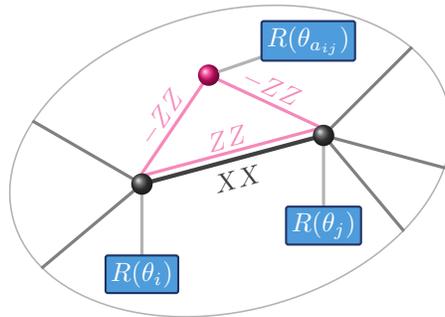
Specifically, we apply rotations with angles $\theta_i/2$, $\theta_j/2$, $\theta_{\xi_{i,j}}/2$ with $\theta_i, \theta_j, \theta_{\xi_{i,j}} \in [-\pi/2, \pi/2]$ to the three qubits. Note that we consider rotations by half-angles $\theta \to \theta/2$ while at the same time doubling the interval $[-\pi/4, \pi/4] \to [-\pi/2, \pi/2]$ to ease
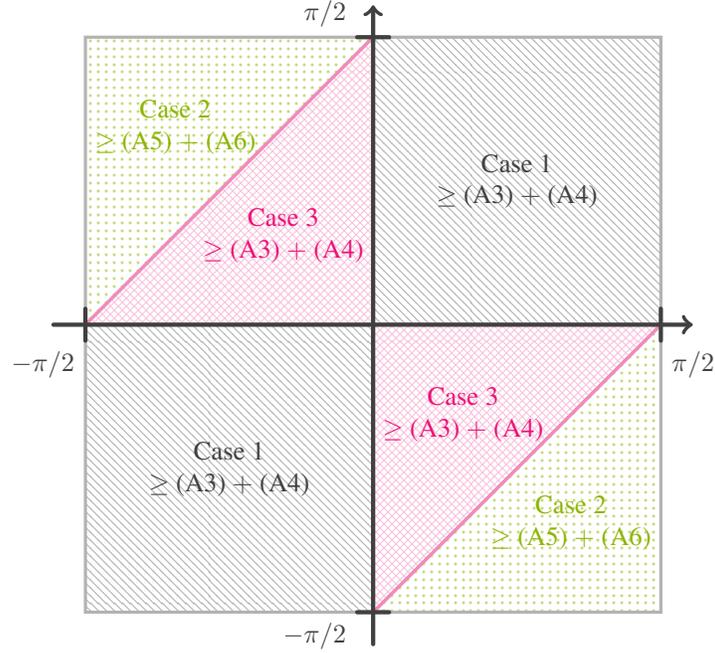
Figure S4. The derivation of a lower bound on the non-stoquasticity of an edge (i,j) and its ancilla qubit $\xi_{i,j}$ on which rotations $R_i(\theta_i/2)R_j(\theta_j/2)R_{\xi_{i,j}}(\theta_{\xi_{i,j}}/2)$ with $\theta_i,\theta_j,\theta_{\xi_{i,j}} \in [\pi/2,\pi/2]$ are applied can be divided into three distinct cases. In each case we use different terms of the expression (B3)-(B6) to lower bound the non-stoquasticity.

notation later in the proof. The effect of rotations on each vertex of an edge (i,j) is given by

$$
\begin{aligned}
X_i X_j + C Z_i Z_j \mapsto \\
&[C\cos(\theta_i)\cos(\theta_j) - \sin(\theta_i)\sin(\theta_j)]Z_i Z_j \\
&+ [C\cos(\theta_i)\sin(\theta_j) - \sin(\theta_i)\cos(\theta_j)]Z_i X_j \\
&+ [C\sin(\theta_i)\cos(\theta_j) - \cos(\theta_i)\sin(\theta_j)]X_i Z_j \\
&+ [C\sin(\theta_i)\sin(\theta_j) + \cos(\theta_i)\cos(\theta_j)]X_i X_j,
\end{aligned}
\tag{B1}
$$

and likewise for the edges $(i,\xi_{i,j})$ and $(j,\xi_{i,j})$. The non-stoquasticity of the transformed Hamiltonian terms corresponding to the three qubits is then given by

$$
\nu_1\left(R_i(\theta_i/2)R_j(\theta_j/2)R_{\xi_{i,j}}(\theta_{\xi_{i,j}}/2)\left(X_i X_j + C(Z_i Z_j - Z_i Z_{\xi_{i,j}} - Z_j Z_{\xi_{i,j}})\right)R_i(\theta_i/2)^T R_j(\theta_j/2)^T R_{\xi_{i,j}}(\theta_{\xi_{i,j}}/2)^T\right)
\tag{B2}
$$

$$
\geq \max\left\{C\sin(\theta_i)\sin(\theta_j) + \cos(\theta_i)\cos(\theta_j), 0\right\}
\tag{B3}
$$

$$
+ (2\deg G')^{-1} \cdot \left(\left|C\sin(\theta_i)\cos(\theta_j) - \cos(\theta_i)\sin(\theta_j)\right| + \left|C\sin(\theta_j)\cos(\theta_i) - \cos(\theta_j)\sin(\theta_i)\right|\right)
\tag{B4}
$$

$$
+ \left(\max\left\{-C\sin(\theta_i)\sin(\theta_{\xi_{i,j}}), 0\right\} + \max\left\{-C\sin(\theta_j)\sin(\theta_{\xi_{i,j}}), 0\right\}\right)
\tag{B5}
$$

$$
+ (2\deg G')^{-1} \cdot C\left(\left|\cos(\theta_i)\sin(\theta_{\xi_{i,j}})\right| + \left|\sin(\theta_i)\cos(\theta_{\xi_{i,j}})\right| + \left|\cos(\theta_j)\sin(\theta_{\xi_{i,j}})\right| + \left|\sin(\theta_j)\cos(\theta_{\xi_{i,j}})\right|\right).
\tag{B6}
$$

Note that the terms (B3) and (B5) stem from the XX interactions with a positive sign and therefore depend on the signs of $\sin$ terms. Conversely, the terms (B4) and (B6) stem from the XZ interactions and therefore involve absolute values.

We divide the allowed rotations into different sectors corresponding to the different combinations of the signs of $\sin(\theta_i)$ and $\sin(\theta_j)$ as shown in Fig. S4. Which of the terms in (B3) and (B5) are non-trivial precisely depends on these combinations. We divide the cases as follows: first, the sectors in which $\mathrm{sign}(\theta_i) = \mathrm{sign}(\theta_j)$. Second, the sectors in which $\mathrm{sign}(\theta_i) = -\mathrm{sign}(\theta_j)$ and $|\theta_i| + |\theta_j| \leq \pi/2$. Third, the sectors in which $\mathrm{sign}(\theta_i) = -\mathrm{sign}(\theta_j)$ and $|\theta_i| + |\theta_j| \geq \pi/2$. Taken together, the three cases cover the entire range of allowed angles. Moreover, in all cases we allow arbitrary choices of $\theta_{\xi_{i,j}} \in [-\pi/2, \pi/2]$. We now proceed to lower-bound the non-stoquasticity of the Hamiltonian terms acting on the three qubits, where in each case we will use different terms of Eqs. (B3)-(B6).
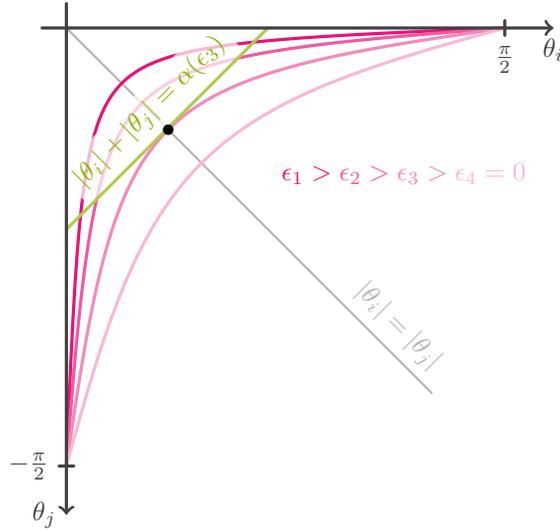
Figure S5. Illustration of the lower bound in case 3: Every θ_i and θ_j such that $|\theta_i| + |\theta_j| \leq \pi/2$ defines a contour line $\epsilon(\theta_i, \theta_j) = \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4 = 0$ (shades of pink). Given θ_i, θ_j and defining $\bar{\epsilon} = \epsilon(\theta_i, \theta_j)$, Lemma 10 implies the lower bound $|\theta_i| + |\theta_j| \geq \alpha(\bar{\epsilon})$ as defined in Eq. (B22). We then obtain (B3) + (B4) $\geq \bar{\epsilon} + (C + 1)\sin(\alpha(\bar{\epsilon})) \geq 1$.

We first discuss the case in which $\text{sign}(\theta_i) = \text{sign}(\theta_j)$. In this case, it suffices to consider the terms (B3) and (B4). Observe that in this case both $C\sin(\theta_i)\sin(\theta_j) \geq 0$ and $\cos(\theta_i)\cos(\theta_j) \geq 0$. Moreover, to our choice of C and noting that $(|C\sin(\theta_i)\cos(\theta_j) - \cos(\theta_i)\sin(\theta_j)| + |C\sin(\theta_j)\cos(\theta_i) - \cos(\theta_j)\sin(\theta_i)|) \geq C|\sin(\theta_i - \theta_j)|$, we obtain

$$(B3) + (B4) \geq \cos(\theta_i - \theta_j) + C/(2\deg G') \cdot |\sin(\theta_i - \theta_j)| \geq 1. \tag{B7}$$

Second, we discuss the case in which $\text{sign}(\theta_i) = -\text{sign}(\theta_j)$ and $\pi/2 < |\theta_i| + |\theta_j| \leq \pi$. In this case, we consider the terms (B5) and (B6):

$$(B5) + (B6) \geq C/(2\deg G') \cdot \Big(|\cos(\theta_{\xi_{i,j}})|(\sin(|\theta_i|) + \sin(|\theta_j|)) \tag{B8}$$

$$+ |\sin(\theta_{\xi_{i,j}})| \left(2\deg(G') \min\{\sin(|\theta_i|), \sin(|\theta_j|)\} + |\cos(\theta_i)| + |\cos(\theta_j)| \right) \Big) \tag{B9}$$

$$\geq 1. \tag{B10}$$

Here, we have used that

$$\sin(|\theta_i|) + \sin(|\theta_j|) = 2\sin\left(\frac{|x| + |y|}{2}\right)\cos\left(\frac{|x| - |y|}{2}\right) \geq 1, \tag{B11}$$

for $\pi/2 \leq |\theta_i| + |\theta_j| \leq \pi$ so that $|x| - |y| \leq \pi/2$ and the definition of $C = (2\deg G')^2$.

Finally, we have the remaining case $\text{sign}(\theta_i) = -\text{sign}(\theta_j)$ and $|\theta_i| + |\theta_j| \leq \pi/2$. In this case, it is again sufficient to consider terms (B3) and (B4). This is the hardest case since the sin terms in (B3) increase much faster than the cos terms decrease due to the factor of $C > 1$. Therefore, we cannot find a bound in terms of a sum-of-angles rule as in the previous cases. Instead, to lower-bound the terms terms (B3) and (B4) in this case, we proceed as follows: We reduce the problem of minimizing the sum (B3) + (B4) ≥ 1 to a one-dimensional problem by noting two facts: first, the term (B4) depends only on the sum $|\theta_i| + |\theta_j|$. Moreover, it increases monotonously in this sum. Second, for every choice of θ_i, θ_j, the value of (B3) takes on its minimal value $\bar{\epsilon}$ at $|\theta_i| = |\theta_j| =: \alpha(\bar{\epsilon})/2$. Hence, the sum (B3) + (B4) is lower bounded by the sum of $\bar{\epsilon}$ and (B4) evaluated at $\alpha(\bar{\epsilon})$. The proof is concluded by a lower bound on the latter term. We now elaborate those steps one-by-one.

Let us begin by expressing (B4) as a function of $|\theta_i| + |\theta_j|$

$$|C\sin(\theta_i)\cos(\theta_j) - \cos(\theta_i)\sin(\theta_j)| + |C\sin(\theta_j)\cos(\theta_i) - \cos(\theta_j)\sin(\theta_i)| \tag{B12}$$

$$= C\sin(|\theta_i|)\cos(\theta_j) + \cos(\theta_i)\sin(|\theta_j|) + C\sin(|\theta_j|)\cos(\theta_i) + \cos(\theta_j)\sin(|\theta_i|) \tag{B13}$$

$$= (C + 1)\sin(|\theta_i| + |\theta_j|), \tag{B14}$$

where we have used the fact that $\text{sign}(\sin(\theta_i)) = -\text{sign}(\sin(\theta_j))$ and that the cosines are non-negative. For $|\theta_i| + |\theta_j| \leq \pi/2$ this is a monotonously increasing function in $|\theta_i| + |\theta_j|$. We now define the value of the term (B3) to be

$$\epsilon(\theta_i, \theta_j) := \max\{-C \sin|\theta_i| \sin|\theta_j| + \cos|\theta_i| \cos|\theta_j|, 0\} \geq 0. \tag{B15}$$

For every choice of $\alpha = \alpha(\theta_i, \theta_j) := |\theta_i| + |\theta_j|$, the minimal value of

$$(B3) + (B4) = \epsilon(\theta_i, \theta_j) + (C+1)/(2 \deg G') \cdot \sin(\alpha(\theta_i, \theta_j)), \tag{B16}$$

is therefore attained at the minimal value of $\epsilon(\theta_i, \theta_j)$ subject to the constraint $|\theta_i| + |\theta_j| = \alpha \leq \pi/2$. Covnersely, the value is attained at the minimal value of $\alpha(\theta_i, \theta_j)$ subject to the constraint (B15). This reduces the problem to a one dimensional problem, which we exploit explicitly in the following lemma. The intuition behind this lemma is shown in Fig. S5.

Lemma 10. *For any fixed value $\pi/2 \geq |\theta_i| + |\theta_j| = \alpha \geq 0$, the minimal value $\bar{\epsilon}(\alpha)$ of $\epsilon(\theta_i, \theta_j)$ is achieved at $|\theta_i| = |\theta_j| = \alpha/2$. Moreover, for every θ_i, θ_j such that $\epsilon(\theta_i, \theta_j) \geq \bar{\epsilon}(\alpha)$ it holds that $|\theta_i| + |\theta_j| \geq \alpha$.*

Proof. Let $|\theta_i| = (\alpha - \delta)/2$, $|\theta_j| = (\alpha + \delta)/2$ for $0 \leq \delta \leq \pi/2$. Then

$$\epsilon(\theta_i, \theta_j) = -C \sin\left(\frac{\alpha - \delta}{2}\right) \sin\left(\frac{\alpha + \delta}{2}\right) + \cos\left(\frac{\alpha + \delta}{2}\right) \cos\left(\frac{\alpha - \delta}{2}\right) \tag{B17}$$

$$= \frac{C}{2}(\cos\alpha - \cos\delta) + \frac{1}{2}(\cos\alpha + \cos\delta) \tag{B18}$$

$$= \frac{1}{2}((C+1)\cos\alpha - (C-1)\cos\delta), \tag{B19}$$

which is minimal at $\delta = 0$.

The second part of the lemma can be be seen by contraposition: Assume $|\theta_i| + |\theta_j| < \alpha(\bar{\epsilon})$. Then

$$\epsilon(\theta_i, \theta_j) = \frac{1}{2}((C+1)\cos(|\theta_i| + |\theta_j|) - (C-1)\cos(|\theta_i| - |\theta_j|)) \tag{B20}$$

$$\leq \frac{1}{2}(C+1)\cos(|\theta_i| + |\theta_j|) < \frac{1}{2}(C+1)\cos\alpha \equiv \bar{\epsilon}(\alpha), \tag{B21}$$

where we have used the assumption and the monotonicity of the cosine in the interval $[0, \pi/2]$ in the last inequality and its non-negativity in the interval $[-\pi/2, \pi/2]$ in the second to last inequality. □

Now, for every choice of θ_i, θ_j, the minimal value of $\bar{\epsilon}(\alpha)$ of $\epsilon(\theta_i, \theta_j)$ corresponding to $\alpha = |\theta_i| + |\theta_j|$ is attained at $|\theta_i| = |\theta_j| = \alpha/2$. Correspondingly, we can re-express α in terms of $\bar{\epsilon}$ as

$$\bar{\epsilon} \equiv \bar{\epsilon}(\alpha) = -C \sin^2(\alpha/2) + \cos^2(\alpha/2) \quad \Leftrightarrow \quad \alpha(\bar{\epsilon}) = 2 \arctan\sqrt{\frac{1 - \bar{\epsilon}}{C + \bar{\epsilon}}}. \tag{B22}$$

The second part of Lemma 10 states that for all θ_i, θ_j such that $\epsilon(\theta_i, \theta_j) \geq \bar{\epsilon}(\alpha) \geq 0$ we have $|\theta_i| + |\theta_j| \geq \alpha(\bar{\epsilon})$ and consequently $\sin(|\theta_i| + |\theta_j|) \geq \sin(\alpha(\bar{\epsilon}))$. Given θ_i, θ_j and defining $\bar{\epsilon} := \epsilon(\theta_i, \theta_j)$ this implies the lower bound

$$(B3) + (B4) \geq \bar{\epsilon} + (C+1)/(2 \deg G') \cdot \sin(\alpha(\bar{\epsilon})), \tag{B23}$$

where we have used the equivalence (B22).

It remains to lower-bound $\sin(\alpha(\bar{\epsilon}))$. Define $x = \sqrt{(1 - \bar{\epsilon})/(C + \bar{\epsilon})}$. We can then rewrite

$$\sin(\alpha(\bar{\epsilon})) = \sin(2 \arctan x) = 2 \sin(\arctan x) \cos(\arctan x) = 2\frac{x}{1 + x^2} \geq x, \tag{B24}$$

for $x \leq 1$, where we have used that $\sin(\arctan(x)) = x \cos(\arctan x) = x/\sqrt{1 + x^2}$. We can also bound

$$x = \sqrt{\frac{1 - \bar{\epsilon}}{C + \bar{\epsilon}}} \geq \sqrt{\frac{1}{C}}\sqrt{\frac{1 - \bar{\epsilon}}{1 + \bar{\epsilon}/C}} \geq \sqrt{\frac{1}{C}}\sqrt{\frac{1 - \bar{\epsilon}}{1 + \bar{\epsilon}}} \geq \frac{1 - \bar{\epsilon}}{\sqrt{C}}, \tag{B25}$$

where the last inequality can be seen by squaring both sides and using $0 \leq \bar{\epsilon} < 1$. Combining everything we obtain

$$(B3) + (B4) \geq \bar{\epsilon} + \frac{\sqrt{C}}{2 \deg G'} \cdot (1 - \bar{\epsilon}) = \bar{\epsilon}\left(1 - \frac{\sqrt{C}}{2 \deg G'}\right) + \frac{\sqrt{C}}{2 \deg G'} = 1. \tag{B26}$$

due to our choice of $C = (2 \deg G')^2$.

To conclude the proof, we discuss the effect of applying X-flips to each of the sites. Applying $X_i X_j$ or $X_{\xi_{i,j}}$ merely alters the signs of the terms in (B5). But since we did not constrain the sign of $\theta_{\xi_{i,j}}$ in the proof, everything remains unchanged. Suppose an X-flip is applied to either qubit i or j, or either both qubit i and $\xi_{i,j}$ or either both qubit j and $\xi_{i,j}$. Assuming wlog. that qubit i is X-flipped, we achieve the same lower bounds as before by identifying $\theta_i \mapsto -\theta_i$.

$\square$

REFERENCES AND NOTES

1. J. E. Hirsch, R. L. Sugar, D. J. Scalapino, R. Blankenbecler, Monte Carlo simulations of one-dimensional fermion systems. *Phys. Rev. B* **26**, 5033–5055 (1982).

2. M. Troyer, F. Alet, S. Trebst, S. Wessel, Non-local updates for quantum Monte Carlo simulations. *AIP Conf. Proc.* **690**, 156–169 (2003).

3. L. Pollet, Recent developments in quantum Monte-Carlo simulations with applications for cold gases. *Rep. Prog. Phys.* **75**, 094501 (2012).

4. S. Trotzky, L. Pollet, F. Gerbier, U. Schnorrberger, I. Bloch, N. V. Prokof'ev, B. Svistunov, M. Troyer, Suppression of the critical temperature for superfluidity near the Mott transition. *Nat. Phys.* **6**, 998–1004 (2010).

5. N. Hatano, M. Suzuki, Representation basis in quantum Monte Carlo calculations and the negative-sign problem. *Phys. Lett. A* **163**, 246–249 (1992).

6. M. B. Hastings, How quantum are non-negative wavefunctions? *J. Math. Phys.* **57**, 015210 (2015).

7. C. Wu, J.-p. Hu, S.-c. Zhang, Exact SO(5) symmetry in the spin-3/2 fermionic system. *Phys. Rev. Lett.* **91**, 186402 (2003).

8. K. Okunishi, K. Harada, Symmetry-protected topological order and negative-sign problem for **SO**(*N*) bilinear-biquadratic chains. *Phys. Rev. B* **89**, 134422 (2014).

9. Z.-X. Li, Y.-F. Jiang, H. Yao, Solving the fermion sign problem in quantum Monte Carlo simulations by Majorana representation. *Phys. Rev. B* **91**, 241117 (2015).

10. Z.-X. Li, Y.-F. Jiang, H. Yao, Majorana-time-reversal symmetries: A fundamental principle for sign-problem-free quantum Monte Carlo simulations. *Phys. Rev. Lett.* **117**, 267002 (2016).

11. T. Nakamura, Vanishing of the negative-sign problem of quantum Monte Carlo simulations in one-dimensional frustrated spin systems. *Phys. Rev. B* **57**, R3197–R3200 (1998).

12. F. Alet, K. Damle, S. Pujari, Sign-problem-free Monte Carlo simulation of certain frustrated quantum magnets. *Phys. Rev. Lett.* **117**, 197203 (2016).

13. A. Honecker, S. Wessel, R. Kerkdyk, T. Pruschke, F. Mila, B. Normand, Thermodynamic properties of highly frustrated quantum spin ladders: Influence of many-particle bound states. *Phys. Rev. B* **93**, 054408 (2016).

14. S. Wessel, B. Normand, F. Mila, A. Honecker, Efficient quantum Monte Carlo simulations of highly frustrated magnets: The frustrated spin–1/2 ladder. *SciPost Phys.* **3**, 005 (2017).

15. J. I. Cirac, D. Perez-Garcia, N. Schuch, F. Verstraete, Matrix product unitaries: Structure, symmetries, and topological invariants. *J. Stat. Mech.* **2017**, 083105 (2017).

16. M. B. Şahinoğlu, S. K. Shukla, F. Bi, X. Chen, Matrix product representation of locality preserving unitaries. *Phys. Rev. B* **98**, 245122 (2018).

17. W. Dobrautz, H. Luo, A. Alavi, Compact numerical solutions to the two-dimensional repulsive Hubbard model obtained via nonunitary similarity transformations. *Phys. Rev. B* **99**, 075119 (2019).

18. M. Troyer, U.-J. Wiese, Computational complexity and fundamental limitations to fermionic quantum Monte Carlo simulations. *Phys. Rev. Lett.* **94**, 170201 (2005).

19. M. Marvian, D. A. Lidar, I. Hen, On the computational complexity of curing non-stoquastic Hamiltonians. *Nat. Commun.* **10**, 1571 (2019).

20. J. Klassen, M. Marvian, S. Piddock, M. Ioannou, I. Hen, B. Terhal, Hardness and ease of curing the sign problem for two-local qubit Hamiltonians. arXiv:1906.08800.3000 [quant-ph] (20 June 2019).

21. J. Klassen, B. M. Terhal, Two-local qubit Hamiltonians: When are they stoquastic? *Quantum* **3**, 139 (2019).

22. H. Shinaoka, Y. Nomura, S. Biermann, M. Troyer, P. Werner, Negative sign problem in continuous-time quantum Monte Carlo: Optimal choice of single-particle basis for impurity problems. *Phys. Rev. B* **92**, 195126 (2015).

23. J. R. McClean, A. Aspuru-Guzik, Clock quantum Monte Carlo: An imaginary-time method for real-time quantum dynamics. *Phys. Rev. A* **91**, 012311 (2015).

24. R. E. Thomas, Q. Sun, A. Alavi, G. H. Booth, Stochastic multiconfigurational self-consistent field theory. *J. Chem. Theory Comput.* **11**, 5316–5325 (2015).

25. W. Bietenholz, A. Pochinsky, U.-J. Wiese, Meron-cluster simulation of the θ vacuum in the 2D O(3) model. *Phys. Rev. Lett.* **75**, 4524–4527 (1995).

26. S. Chandrasekharan, U.-J. Wiese, Meron-cluster solution of fermion sign problems. *Phys. Rev. Lett.* **83**, 3116–3119 (1999).

27. P. Henelius, A. W. Sandvik, Sign problem in Monte Carlo simulations of frustrated quantum spin systems. *Phys. Rev. B* **62**, 1102–1113 (2000).

28. M. Nyfeler, F.-J. Jiang, F. Kämpfer, U.-J. Wiese, Nested cluster algorithm for frustrated quantum antiferromagnets. *Phys. Rev. Lett.* **100**, 247206 (2008).

29. E. Huffman, S. Chandrasekharan, Solution to sign problems in models of interacting fermions and quantum spins. *Phys. Rev. E* **94**, 043311 (2016).

30. C. T. Hann, E. Huffman, S. Chandrasekharan, Solution to the sign problem in a frustrated quantum impurity model. *Ann. Phys.* **376**, 63–75 (2017).

31. I. Hen, Resolution of the sign problem for a frustrated triplet of spins. *Phys. Rev. E* **99**, 033306 (2019).

32. S. Foucart, H. Rauhut, A mathematical introduction to compressive sensing, in *Applied and Numerical Harmonic Analysis* (Springer, 2013).

33. H.-J. Mikeska, A. K. Kolezhuk, One-dimensional magnetism, in *Quantum Magnetism*, U. Schollwöck, J. Richter, D. J. J. Farnell, R. F. Bishop, Eds. (Lecture Notes in Physics, Springer Berlin Heidelberg, 2004), pp. 1–83.

34. E. Dagotto, T. M. Rice, Surprises on the way from one- to two-dimensional quantum magnets: The ladder materials. *Science* **271**, 618–623 (1996).

35. M. Takano, Spin ladder compounds. *Phys. C* **263**, 468–474 (1996).

36. A. W. Sandvik, Computational studies of quantum spin systems. *AIP Conf. Proc.* **1297**, 135–338 (2010).

37. T. Meng, T. Neupert, M. Greiter, R. Thomale, Coupled-wire construction of chiral spin liquids. *Phys. Rev. B* **91**, 241106 (2015).

38. P.-H. Huang, J.-H. Chen, A. E. Feiguin, C. Chamon, C. Mudry, Coupled spin-$\frac{1}{2}$ ladders as microscopic models for non-Abelian chiral spin liquids. *Phys. Rev. B* **95**, 144413 (2017).

39. A. Nietner, C. Krumnow, E. J. Bergholtz, J. Eisert, Composite symmetry protected topological order and effective models. *Phys. Rev. B* **96**, 235138 (2017).

40. Y. Yoshida, H. Ito, M. Maesato, Y. Shimizu, H. Hayama, T. Hiramatsu, Y. Nakamura, H. Kishida, T. Koretsune, C. Hotta, G. Saito, Spin-disordered quantum phases in a quasi-one-dimensional triangular lattice. *Nat. Phys.* **11**, 679–683 (2015).

41. K. T. Lai, M. Valldor, Coexistence of spin ordering on ladders and spin dimer formation in a new-structure-type compound $Sr_2Co_3S_2O_3$. *Sci. Rep.* **7**, 43767 (2017).

42. V. I. Iglovikov, E. Khatami, R. T. Scalettar, Geometry dependence of the sign problem in quantum Monte Carlo simulations. *Phys. Rev. B* **92**, 045110 (2015).

43. J. Carrasquilla, Z. Hao, R. G. Melko, A two-dimensional spin liquid in quantum kagome ice. *Nat. Commun.* **6**, 7421 (2015).

44. T. Abrudan, J. Eriksson, V. Koivunen, Conjugate gradient algorithm for optimization under unitary matrix constraint. *Signal Process.* **89**, 1704–1714 (2009).

45. D. Hangleiter, I. Roth, Gitlab repository (2019); https://gitlab.com/ingo.roth/signease.

46. F. Barahona, On the computational complexity of Ising spin glass models. *J. Phys. A* **15**, 3241–3253 (1982).

47. D. P. Landau, K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge Univ. Press, 2000).

48. G. H. Booth, A. J. W. Thom, A. Alavi, Fermion Monte Carlo without fixed nodes: A game of life, death, and annihilation in Slater determinant space. *J. Chem. Phys.* **131**, 054106 (2009).

49. C. M. Dawson, H. L. Haselgrove, A. P. Hines, D. Mortimer, M. A. Nielsen, T. J. Osborne, Quantum computing and polynomial equations over the finite field Z_2. *Quant. Inf. Comp.* **5**, 102–112 (2005).

50. S. P. Jordan, D. Gosset, P. J. Love, Quantum-Merlin-Arthur–complete problems for stoquastic Hamiltonians and Markov matrices. *Phys. Rev. A* **81**, 032331 (2010).

51. H. Pashayan, J. J. Wallman, S. D. Bartlett, Estimating outcome probabilities of quantum circuits using quasiprobabilities. *Phys. Rev. Lett.* **115**, 070501 (2015).

52. K. N. Anagnostopoulos, J. Nishimura, New approach to the complex-action problem and its application to a nonperturbative study of superstring theory. *Phys. Rev. D* **66**, 106008 (2002).

53. Z. Ringel, D. L. Kovrizhin, Quantized gravitational responses, the sign problem, and quantum complexity. *Sci. Adv.* **3**, e1701758 (2017).

54. M. A. Levin, X.-G. Wen, String-net condensation: A physical mechanism for topological phases. *Phys. Rev. B* **71**, 045110 (2005).

55. R. Levy, B. K. Clark, Mitigating the sign problem through basis rotations. arXiv:1907.02076 [cond-mat.str-el] (3 July 2019).

56. G. Torlai, J. Carrasquilla, M. T. Fishman, R. G. Melko, M. P. A. Fisher, Wavefunction positivization via automatic differentiation. arXiv:1906.04654 [quant-ph] (11 June 2019).

57. A. J. Kim, P. Werner, R. Valentí, Alleviating the sign problem in quantum Monte Carlo simulations of spin-orbit-coupled multi-orbital Hubbard Models. *Phys. Rev. B.* **101**, 045108 (2019).

58. Y. Shoukry, P. Nuzzo, A. L. Sangiovanni-Vincentelli, S. A. Seshia, G. J. Pappas, P. Tabuada, SMC: Satisfiability modulo convex optimization, in *Proceedings of the 20th International Conference on Hybrid Systems: Computation and Control, HSCC '17* (ACM, 2017), pp. 19–28.

59. S. Bravyi, D. P. DiVincenzo, R. I. Oliveira, B. M. Terhal, The complexity of stoquastic local Hamiltonian problems. *Quant. Inf. Comp.* **8**, 0361–0385 (2008).

60. S. Bravyi, B. Terhal, Complexity of stoquastic frustration-free Hamiltonians. *SIAM J. Comput.* **39**, 1462 (2009).

61. T. Cubitt, A. Montanaro, Complexity classification of local Hamiltonian problems. *SIAM J. Comput.* **45**, 268–316 (2016).

62. T. S. Cubitt, A. Montanaro, S. Piddock, Universal quantum Hamiltonians. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 9497–9502 (2018).

63. ISO 16269-4:2010, *Statistical Interpretation of Data—Part 4: Detection and Treatment of Outliers* (International Organization for Standardization, 2010).

64. J. S. Spencer, N. S. Blunt, W. M. C. Foulkes, The sign problem and population dynamics in the full configuration interaction quantum Monte Carlo method. *J. Chem. Phys.* **136**, 054110 (2012).

65. M. J. Wainwright, *High-Dimensional Statistics: A Non-Asymptotic Viewpoint* (Cambridge Univ. Press, ed. 1, 2019).

66. M. Schmidt, G. Fung, R. Rosales, Fast optimization methods for L1 regularization: A comparative study and two new approaches*, in *Machine Learning: ECML 2007*, J. N. Kok, J. Koronacki, R. L. de Mantaras, S. Matwin, D. Mladenič, A. Skowron (Lecture Notes in Computer Science, Springer, 2007), pp. 286–297.