

Optimal entanglement witnesses: a scalable data-driven approach

Irénée Frérot^{1,2,*} and Tommaso Roscilde^{3,†}

¹*ICFO-Institut de Ciències Fotoniques, The Barcelona Institute of Science and Technology,
Av. Carl Friedrich Gauss 3, 08860 Barcelona, Spain*

²*Max-Planck-Institut für Quantenoptik, D-85748 Garching, Germany*

³*Laboratoire de Physique, CNRS UMR 5672, Ecole Normale Supérieure de Lyon,
Université de Lyon, 46 Allée d'Italie, Lyon, F-69364, France*

Multipartite entanglement is a key resource allowing quantum devices to outperform their classical counterparts, and entanglement certification is fundamental to assess any quantum advantage. The only scalable certification scheme relies on entanglement witnessing, typically effective only for special entangled states. Here we focus on finite sets of measurements on quantum states (hereafter called quantum data); and we propose an approach which, given a particular spatial partitioning of the system of interest, can effectively ascertain whether or not the data set is compatible with a separable state. When compatibility is disproven, the approach produces the optimal entanglement witness for the quantum data at hand. Our approach is based on mapping separable states onto equilibrium classical field theories on a lattice; and on mapping the compatibility problem onto an inverse statistical problem, whose solution is reached in polynomial time whenever the classical field theory does not describe a glassy system. Our results pave the way for systematic entanglement certification in quantum devices, optimized with respect to the accessible observables.

Introduction. Preparing and processing strongly entangled many-body states, in both a controlled and scalable way, is the goal of all quantum simulators and computers. Indeed, as the efficient representation of generic entangled many-body states is impossible on classical machines, entanglement represents the key computational resource of these devices [1, 2]. As a consequence, developing generic and scalable methods to certify entanglement in multipartite systems stands as a grand challenge of quantum information science. Even more fundamentally, entanglement certification is a central task to probe the resilience of quantum correlations from the microscopic world to the macroscopic one [3].

Any practical method must circumvent the tomographic reconstruction of the full density matrix [4, 5] (which implies a number of measurements scaling exponentially with system size), and it should instead infer entanglement from the partial information contained in a given data set of measurement results (hereafter referred to as *quantum data*). When one adopts this data-driven strategy, the goal of entanglement certification is to establish whether or not the quantum data are compatible with a separable state [5–7]. Given an extended quantum system composed of N_{tot} degrees of freedom, grouped together into $N \leq N_{\text{tot}}$ clusters [see Fig. 1(a)], the state $\hat{\rho}$ of the system is separable [8] if it can be written in the form

$$\hat{\rho}_p := \int d\lambda p(\lambda) \hat{\rho}_{\text{prod}}(\lambda) \quad (1)$$

where $\hat{\rho}_{\text{prod}}(\lambda) = \otimes_{i=1}^N |\psi_i(\lambda_i)\rangle\langle\psi_i(\lambda_i)|$ is a product state of the partition, $|\psi_i(\lambda_i)\rangle$ being the state of the i -th clus-

ter, parametrized by parameters $\lambda = (\lambda_1, \dots, \lambda_i, \dots, \lambda_N)$, distributed according to $p(\lambda) \geq 0$. The distribution p fully specifies classical correlations across the partition. A multipartite entangled state $\hat{\rho}$, on the other hand, cannot be written in the above form. Given a set of observables \hat{A}_a ($a = 1, \dots, R$), multipartite entanglement is therefore *witnessed* by the quantum data set $\{\langle\hat{A}_a\rangle_{\hat{\rho}}\}_{a=1}^R$ [where $\langle\hat{A}_a\rangle_{\hat{\rho}} = \text{Tr}(\hat{A}_a\hat{\rho})$] if one proves that the latter cannot be reproduced by *any* separable state. This task is accomplished by proving that the quantum data violate an entanglement witness (EW) inequality, $\langle\hat{W}\rangle_{\hat{\rho}_p} = \sum_a W_a \langle\hat{A}_a\rangle_{\hat{\rho}_p} \geq B_{\text{sep}}$, valid for all separable states $\hat{\rho}_p$ [9]. Here W_a are suitable coefficients and B_{sep} is the so-called separable bound.

EW operators \hat{W} are generally defined based on the properties of special entangled states (e.g. squeezed states, total spin singlets, etc.) [9], and failure of a data set to violate a given EW inequality does not exclude the existence of a different violated inequality involving the same data, yet to be discovered. This may erroneously suggest that entanglement witnessing is limited by creativity and physical insight; and that the entanglement witnessing problem (“is a quantum data set compatible with a separable state?”) [5–7] is generically undecidable. The goal of our work is to show that this is *not* the case, and that the entanglement witnessing capability of a quantum data set can be *exhaustively* tested. Our key insight is that the problem of finding the distribution $p(\lambda)$, which defines the separable state reproducing at best the quantum data, is a statistical inference problem; and remarkably it has the structure of a convex optimization problem, whose solution can be attained in a time scaling polynomially with the partition size (under mild assumptions), and with the Hilbert space dimension of the subsystems composing the partition. When the optimal separable state fails to reproduce the quantum data, the distance between the quantum data set $\{\langle\hat{A}_a\rangle_{\hat{\rho}}\}$ and

*Electronic address: irenee.frerot@gmail.com

†Electronic address: tommaso.roskilde@ens-lyon.fr

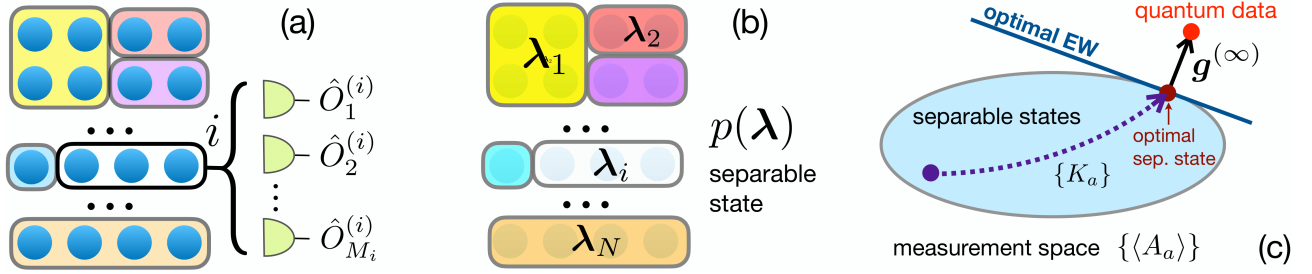


FIG. 1: (a) Partition of a quantum device into N clusters, each of which is subject to M_i measurements; (b) A separable state of the system is described as a probability distribution $p(\lambda)$ of local states defined by the $\{\lambda_i\}$ parameters; (c) Our algorithm builds a trajectory of separable states (parametrized by couplings $\{K_a\}$ defining $p(\lambda)$) which converges to the optimal state approximating at best some target quantum data. If the state fails to reproduce the quantum data exactly, the vector joining the optimal separable data and the quantum data reconstructs the optimal EW inequality.

the optimal separable set $\{\langle \hat{A}_a \rangle_{\hat{\rho}_p}\}$ allows one to reconstruct the optimal EW inequality violated by the quantum data. We benchmark our approach by establishing new EW inequalities satisfied by the low-temperature states of the Heisenberg antiferromagnetic chain and the quantum Ising chain; in the latter case, our new EW inequalities outperform all previously known EW criteria for multipartite entanglement. Our work parallels the recent mapping of the Bell-nonlocality detection problem onto an inverse statistical problem [10], and it offers an efficient scheme for entanglement detection in state-of-the-art quantum devices within a device-dependent scenario.

Quantum data set. For definiteness, we assume that on each subsystem $i = 1, \dots, N$, M_i local observables $\hat{O}_m^{(i)}$ can be measured ($m = 1, \dots, M_i$; e.g. the Pauli matrices $\hat{\sigma}_a^{(i)}$, $a \in \{x, y, z\}$ for individual qubits taken as subsystems). For convenience, we denote the local identity operator by $\hat{O}_0^{(i)} := \mathbb{1}$. In order to reveal entanglement, these local observables must be non-commuting ($[\hat{O}_m^{(i)}, \hat{O}_n^{(i)}] \neq 0$ for $1 \leq m < n \leq M_i$) [11]. From these local observables, we build p -body correlators of the form $\hat{O}_{\mathbf{m}} = \otimes_{i=1}^N \hat{O}_{m_i}^{(i)}$ where $m_i = 0$ for $N - p$ subsystems. Arbitrary observables can be built as linear combinations of correlators – such as e.g. powers of collective spin variables [12, 13] $\hat{J}_a = \sum_i \hat{\sigma}_a^{(i)}/2$ ($a = x, y, z$) for systems of qubits. Hence we shall assume that R observables of the form $\hat{A}_a = \sum_{\mathbf{m}} x_{\mathbf{m}}^{(a)} \hat{O}_{\mathbf{m}}$ can be measured, where the sum runs over all strings $\mathbf{m} = (m_1, \dots, m_N)$, and $x_{\mathbf{m}}^{(a)}$ are arbitrary real coefficients. The quantum data $\{\langle \hat{A}_a \rangle_{\hat{\rho}}\}_{a=1}^R$ form the basis for entanglement certification in our scheme. The problem of entanglement certification based on a data set has been discussed in the past, but the proposed methods either lack scalability [6], or are scalable only under some restrictive assumptions (short-range correlations, low-dimensional geometry) [7]. Our method aims at surpassing these limitations.

Mapping onto an inverse statistical problem. The key aspect behind our approach is the limited information content of separable states. The parameters λ specify-

ing the product state $\hat{\rho}_{\text{prod}}(\lambda)$ can indeed be chosen as $\sum_i (2d_i - 2) \sim \mathcal{O}(N)$ real parameters, where d_i is the dimension of the local Hilbert space of the i -th subsystem [14]. The average of the \hat{A}_a observable on a separable state reads

$$\langle \hat{A}_a \rangle_{\hat{\rho}_p} = \int d\lambda p(\lambda) \mathcal{A}_a(\lambda) =: \langle \mathcal{A}_a \rangle_p \quad (2)$$

where $\mathcal{A}_a(\lambda) = \sum_{\mathbf{m}} x_{\mathbf{m}}^{(a)} \prod_{i=1}^N o_{m_i}^{(i)}(\lambda_i)$ and $o_{m_i}^{(i)}(\lambda_i) = \langle \psi_i(\lambda_i) | \hat{O}_{m_i}^{(i)} | \psi_i(\lambda_i) \rangle$. Given a product state, the calculation of each term in the sum defining $\mathcal{A}_a(\lambda)$ is clearly an operation scaling as $\mathcal{O}(N)$. Once the quantum nature of the state has been absorbed in $\mathcal{A}_a(\lambda)$, the calculation of $\langle \hat{A}_a \rangle_{\hat{\rho}_p}$, Eq. (2), is a classical statistical average over the distribution p which, from a statistical physics viewpoint, can be regarded as the Boltzmann distribution $p(\lambda) =: \exp[-H(\lambda)]/\mathcal{Z}$ of a classical field theory on a lattice (normalized by the \mathcal{Z} factor), with a vector field λ_i defined on each of the N clusters [Fig. 1(b)]. The complexity of separable states is fundamentally inscribed in the effective Hamiltonian $H(\lambda)$, which is *a priori* arbitrary, namely it is specified by a number $\mathcal{O}(\exp(N))$ of parameters.

Once the classical statistical structure of the expectation values on separable states is exposed, the problem of reproducing the quantum data with a separable state takes the form of a statistical inference problem, whose solution is well known in statistical physics [15]. First of all, applying a maximum-entropy principle [16], the Hamiltonian can be parametrized without loss of generality with as many parameters as the elements of the quantum data set [17]:

$$H(\lambda) = - \sum_{a=1}^R K_a \mathcal{A}_a(\lambda). \quad (3)$$

The parameters $\mathbf{K} = \{K_a\}_{a=1}^R$ – the coupling constants of the classical field theory – are Lagrange multipliers whose optimization allows one to build the separable state $\hat{\rho}_p$ whose expectation values $\{\langle \hat{A}_a \rangle_{\hat{\rho}_p}\}$ best approximate the quantum data $\{\langle \hat{A}_a \rangle_{\hat{\rho}}\}$. The optimiza-

tion of \mathbf{K} can be efficiently achieved upon minimizing the cost function $L(\mathbf{K}) := \log \mathcal{Z}(\mathbf{K}) - \sum_a K_a \langle \hat{A}_a \rangle_{\hat{\rho}}$ [10, 15]. The a -th component of the gradient of L is $g_a := \frac{\partial L}{\partial K_a} = \langle \mathcal{A}_a \rangle_p - \langle \hat{A}_a \rangle_{\hat{\rho}}$, and its Hessian matrix is $\frac{\partial^2 L}{\partial K_a \partial K_b} = \langle \mathcal{A}_a \mathcal{A}_b \rangle_p - \langle \mathcal{A}_a \rangle_p \langle \mathcal{A}_b \rangle_p$, namely the covariance matrix of the $\mathcal{A}_a(\boldsymbol{\lambda})$ functions. Since the latter is a semi-definite positive matrix, L is a convex function. Therefore, a simple gradient-descent algorithm, which consists in iterating the update rule $K'_a = K_a - \epsilon[\langle \mathcal{A}_a \rangle_p - \langle \hat{A}_a \rangle_{\hat{\rho}}]$ with $\epsilon \ll 1$, or any improvement thereof, is guaranteed to reach the global optimum of the problem. In practice, this requires to repeatedly compute $\langle \mathcal{A}_a \rangle_p$ as in Eq. (2), a task efficiently accomplished *e.g.* by Markov-chain Monte Carlo sampling of $p(\boldsymbol{\lambda})$, whenever the Hamiltonian H does not describe a glassy system. The restriction to non-glassy systems is the only practical limitation of our approach [17]; and is ensured in the examples considered below by considering translationally invariant systems.

Construction of an optimal entanglement witness. As illustrated on Fig. 1(c), the algorithm converges to the distribution p which minimizes $|g|$ – the norm of the gradient of L . If the minimal distance $g^{(\min)}$ vanishes (within the error on the quantum data), i.e. if $\langle \hat{A}_a \rangle_{\hat{\rho}_p^{(\min)}} = \langle \hat{A}_a \rangle_{\hat{\rho}}$ for all $a = 1, \dots, R$, then entanglement cannot be assessed from the available data. But in the opposite case, the coupling constants K_a increase indefinitely along the optimization, and the coefficients of the gradient $g_a^{(\min)} = \langle \hat{A}_a \rangle_{\hat{\rho}_p^{(\min)}} - \langle \hat{A}_a \rangle_{\hat{\rho}}$ allow us to build a violated EW inequality. First, we define the normalized coefficients $W_a := -g_a^{(\min)} / |g^{(\min)}|$. The condition $|g^{(\min)}|^2 > 0$ is then rewritten as:

$$-\sum_a W_a \langle \hat{A}_a \rangle_{\hat{\rho}} < \min_{\hat{\rho}_p} \left\{ -\sum_a W_a \langle \hat{A}_a \rangle_{\hat{\rho}_p} \right\} =: B_{\text{sep}} \quad (4)$$

The linear combination $\hat{\mathcal{W}} := -\sum_{a=1}^R W_a \hat{A}_a$ is the data-driven EW operator. The separable bound B_{sep} , namely the minimal value of $\text{Tr}(\hat{\rho} \hat{\mathcal{W}})$ over separable states, is violated by the data set, ultimately proving that entanglement is present among the subsystems. The operator $\hat{\mathcal{W}}$ is optimal, in that any other normalized linear combination $\hat{\mathcal{W}}' = -\sum_a W'_a \hat{A}_a$ defines an EW inequality whose violation cannot exceed the violation of the inequality involving $\hat{\mathcal{W}}$. This property follows from the convexity of the set of separable states.

Complexity of the algorithm. If the quantum data contain correlation functions involving up to k points, the effective Hamiltonian H contains $\mathcal{O}(N^k)$ terms; therefore the computational cost of evaluating statistical averages of the kind of Eq. (2) with a precision of ϵ (using Monte Carlo sampling) scales as $\mathcal{O}(d^2 \epsilon^{-2} N^k)$, where $\mathcal{O}(d^2)$ is the cost of evaluating the local observables $o_{m_i}^{(i)}(\boldsymbol{\lambda}_i)$ when $d_i = d$. The polynomial scaling of the computational cost with the number N of parties and with the local Hilbert space dimension is the central asset of our approach.

Ensembles of qubits. Hereafter we shall specify our atten-

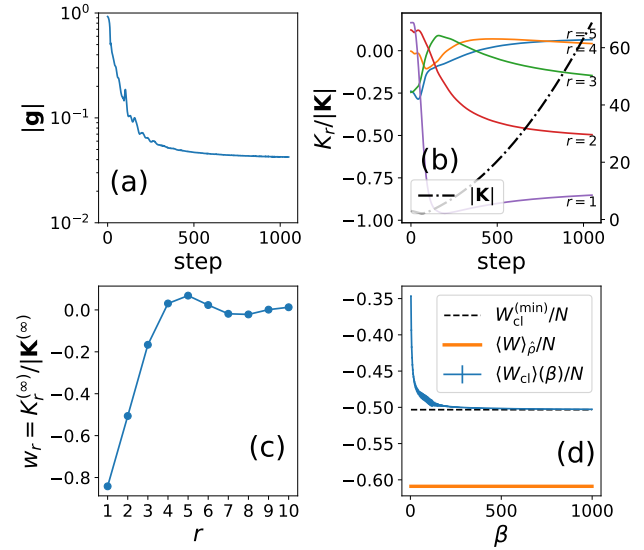


FIG. 2: Data-driven entanglement witness for the Heisenberg chain at $T/J = 1$. (a) Distance between the quantum data (all spin-spin correlators) and the optimized separable state (\mathbf{g} : gradient of the cost function), as a function of optimization steps in a Nesterov accelerated gradient descent ($\epsilon = 0.01$). Each step contains $10^5 - 10^7$ Monte Carlo steps to achieve a relative precision of 10% on the modulus of the gradient [17]. (b) Normalized coupling constants K_r in the classical Hamiltonian defining the separable state (solid lines, left axis), and overall amplitude $|K|$ (dashed-dotted line, right axis). (c) Normalized couplings K_r at the end of the algorithm; (d) The separable bound can be obtained via simulated annealing [18] by calculating $\langle \mathcal{W}_{\text{cl}} \rangle(\beta)$ against $\exp[-\beta \mathcal{W}_{\text{cl}}]$, ramping β from 0 to 1000. The minimum $\mathcal{W}_{\text{cl}}^{(\min)}$ is actually the lowest value recorded for \mathcal{W}_{cl} throughout the ramp.

tion to the case of systems of N qubits partitioned into subsystems consisting of single qubits; and quantum data will be assumed to consist of one- and two-point correlations, $\langle \hat{\sigma}_a^{(i)} \rangle_{\hat{\rho}}$ and $\langle \hat{\sigma}_a^{(i)} \hat{\sigma}_b^{(j)} \rangle_{\hat{\rho}}$ respectively, fully specifying all one- and two-qubit reduced density matrices. Product states are parametrized by the orientations $\{\boldsymbol{\lambda}_i\} = \{\mathbf{n}^{(i)}\}$ of each qubit on the Bloch sphere (where $\mathbf{n}^{(i)}$ are unit vectors), so that the effective Hamiltonian describes *classical Heisenberg spins* (namely, rotators), coupled via bilinear interactions and immersed in an external field:

$$H(\{\mathbf{n}^{(i)}\}) = -\sum_{i=1}^N \sum_{a=x,y,z} K_a^{(i)} n_a^{(i)} - \sum_{i < j} \sum_{a,b} K_{ab}^{(ij)} n_a^{(i)} n_b^{(j)}.$$

Heisenberg antiferromagnetic chain. The first example of entangled states that we study with our approach is the thermal equilibrium state of the $S = 1/2$ Heisenberg chain $\hat{H} = J \sum_{i=1}^N \hat{\mathbf{S}}^{(i)} \cdot \hat{\mathbf{S}}^{(i+1)}$, where $\hat{\mathbf{S}}^{(i)}$ are $S = 1/2$ spin operators, J is the exchange energy, and periodic boundary conditions (PBC) are assumed. Thermal equilibrium states $\hat{\rho} \propto \exp[-\hat{H}/k_B T]$ give $\langle \hat{\sigma}_a^{(i)} \rangle_{\hat{\rho}} = 0$ and $\langle \hat{\sigma}_a^{(i)} \hat{\sigma}_b^{(j)} \rangle_{\hat{\rho}} = \delta_{ab} C(|i - j|)$, due to rotational invariance

of the spin-spin couplings and translational invariance. These elementary symmetries of the quantum data are directly inherited by the classical Hamiltonian defining separable states aimed at reproducing them. The Hamiltonian takes the form of a classical long-range Heisenberg model $H(\{\mathbf{n}^{(i)}\}) = -\sum_{i<j} K_{|i-j|} \mathbf{n}^{(i)} \cdot \mathbf{n}^{(j)}$ with $K_r = K_{N-r}$. The most effective existing multipartite entanglement criterion for this quantum data is based on the collective spin, namely $\langle \hat{J}^2 \rangle = \sum_{ij} \langle \hat{\mathbf{S}}^{(i)} \cdot \hat{\mathbf{S}}^{(j)} \rangle < N/2$ [19, 20], which is verified for $t = T/J \lesssim 1.4$. This criterion is a permutationally invariant EW (PIEW), treating correlations at all distances on the same footing, and it cannot be optimal at sufficiently high temperatures, namely when the correlation length ξ becomes of the order of a few lattice spacings.

As a first validation of our approach, we search for the optimal EW based on two-body correlations $\langle \hat{\sigma}_a^{(i)} \hat{\sigma}_a^{(j)} \rangle$ by using as input quantum data the correlations (obtained via quantum Monte Carlo - QMC [17]) at $t = 1$ for $N = 64$ spins, at which $\xi = 0.72$. Because of their finite range we only used correlations up to a distance $r_{\max} = 10$. Fig. 2 illustrates the results of our approach. The saturation to a finite value of the distance between the quantum data and those of the optimized separable state (measured by the norm of the vector \mathbf{g} , see Fig. 2(a)) and the divergence of the couplings K_r (Fig. 2(b)) clearly indicate the success of entanglement witnessing. The optimal EW operator can be reconstructed in principle from the asymptotic value of the gradient vector $\mathbf{g}^{(\infty)}$ as $\hat{\mathcal{W}} = -\sum_{i=1}^N \sum_{a \in \{x,y,z\}} \sum_{r=1}^{r_{\max}} w_r \hat{\sigma}_a^{(i)} \hat{\sigma}_a^{(i+r)}$ with $w_r = -g_r^{(\infty)} / |\mathbf{g}^{(\infty)}|$. In practice, we found a more strongly violated EW inequality using the asymptotic couplings of the effective Hamiltonian, namely $w_r = K_r^{(\infty)} / |\mathbf{K}^{(\infty)}|$ - which display a clear spatial structure, shown in Fig. 2(c) (see [17] for the numerical values). The final step of the approach consists in determining the separable bound $B_{\text{sep}} = \min_{\hat{\rho}_p} \text{Tr}(\hat{\rho}_p \hat{\mathcal{W}})$. The latter can be obtained as the solution of a set of algebraic equations [21, 22]; here we rather obtain it by finding the ground-state energy of the classical Hamiltonian $\mathcal{W}_{\text{cl}} = -\sum_{i=1}^N \sum_{r=1}^{r_{\max}} w_r \mathbf{n}^{(i)} \cdot \mathbf{n}^{(i+r)}$ via temperature annealing [18] [Fig. 2(d)]. We observe that $B_{\text{sep}}/N = -0.5032$, while the quantum data reach $\langle \hat{\mathcal{W}} \rangle_{\hat{\rho}}/N = -0.6089$. In contrast, the best PIEW - properly normalized [17] - is violated by an amount of 0.04552. This result is *not* incremental, because the EW inequality we find is optimal among all those containing two-body correlators. Interestingly, for temperatures $t \gtrsim 1.4$ (at which the PIEW ceases to work) we found numerically impossible to prove that $\hat{\rho}(T)$ is entangled solely based on two-point correlators: this in turn shows that the maximal set of thermal states whose entanglement can be witnessed using two-point correlators is essentially captured by the PIEW. This will not be the case in our next example, in which our approach significantly extends the range of witnessed entangled thermal states.

Quantum Ising chain. Our final example is the quantum Ising model with Hamiltonian

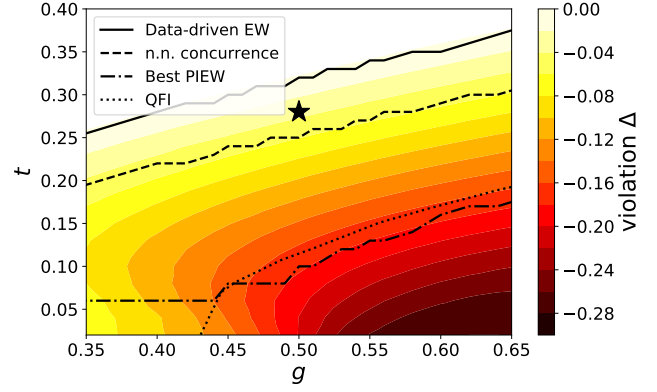


FIG. 3: Data-driven EW for the quantum Ising chain. Phase diagram around the QCP. The star corresponds to $t = 0.28$, $g = 0.5$, at which the quantum data used as input were calculated. The color represents the violation $\Delta = (\langle \hat{\mathcal{W}} \rangle_{\hat{\rho}} - B_{\text{sep}})/N$ of our data-driven EW. The various curves correspond to the temperature below which different entanglement criteria are satisfied (nearest-neighbour concurrence [23]; best PIEW [12]; and quantum Fisher information (QFI) of \hat{J}_z [24].

$\hat{H} = -J \sum_{i=1}^N (\hat{S}_z^{(i)} \hat{S}_z^{(i+1)} + g \hat{S}_x^{(i)})$, where J is the interaction strength and Jg the transverse field. In the ground state, the system displays a quantum critical point (QCP) at $g = g_c = 1/2$ between a ferromagnetic phase ($g < g_c$) and a paramagnetic phase ($g > g_c$) [25]. At finite temperature around the QCP, the system is known to exhibit robust entanglement [24, 26, 27]. Given the symmetries of the correlation functions ($\langle \hat{\sigma}_a^{(i)} \rangle_{\hat{\rho}} = 0$ for $a = y, z$; $\langle \hat{\sigma}_a^{(i)} \hat{\sigma}_b^{(j)} \rangle_{\hat{\rho}} \sim \delta_{ab}$), the classical Hamiltonian tailored to reproduce them is of the form: $H(\{\mathbf{n}^{(i)}\}) = -K_x \sum_{i=1}^N n_x^{(i)} - \sum_{a=x,y,z} \sum_{i<j} K_a^{|i-j|} n_a^{(i)} n_a^{(j)}$. As input quantum data, we consider the correlation functions of a chain of $N = 64$ spins with PBC at a temperature $t = T/J = 0.28$ for $g = 0.5$ - obtained as well via QMC. Given the finite correlation length, we only used correlators up to a distance $r_{\max} = 20$. Following the same procedure as described for the Heisenberg chain, we find an optimal EW operator which is spatially structured, of the form $\hat{\mathcal{W}} = -w_x \sum_{i=1}^N \hat{\sigma}_x^{(i)} - \sum_{a=x,y,z} \sum_{i<j} w_a^{|i-j|} \hat{\sigma}_a^{(i)} \hat{\sigma}_a^{(j)}$ (coefficients and separable bound in the Supplemental Material [17]). On Fig. 3, we show that this new EW criterion, optimal for the thermal state at $t = 0.28$, $g = 0.5$, allows one to prove entanglement for a larger set of thermal states than all the existing criteria in the literature (namely the nearest-neighbour concurrence [23], the PIEW [12], and the quantum Fisher information [24] - see [17] for further details).

Conclusions. We introduced a data-driven method to probe multipartite entanglement in many-body systems. This method relies on mapping separable states onto Boltzmann distributions for a classical field theory on a lattice. The classical degrees of freedom of this field the-

ory are dictated by the considered partitioning of the system. The structure of the corresponding classical Hamiltonian is dictated by the quantum data at hand; and its parameters are optimized in order to fit at best the quantum data. This method allows to exhaustively test the entanglement witnessing capability of a set of quantum data in a time scaling polynomially with the number of parties in the partition (if the size of quantum data is also polynomial); this is guaranteed whenever the classical field theory is not a model of a glass (namely when it does not feature disorder *and* frustration). This opens the way to the systematic certification of entanglement in intermediate-scale quantum devices.

Acknowledgments

We are very grateful to Antonio Acín for insightful discussions. IF acknowledges support from the

Government of Spain (FIS2020-TRANQI and Severo Ochoa CEX2019-000910-S), Fundació Cellex and Fundació Mir-Puig through an ICFO-MPQ Postdoctoral Fellowship, Generalitat de Catalunya (CERCA, AGAUR SGR 1381 and QuantumCAT), and MINECO-EU QuantERA MAQS (funded by The State Research Agency (AEI) PCI2019-111828-2 / 10.13039/501100011033). TR acknowledges support from ANR (EELS project) and QuantERA (MAQS project). Numerical computations have been performed at the Pôle Scientifique de Modélisation Numérique (PSMN).

-
- [1] Georgescu, I. M., Ashhab, S. & Nori, F. Quantum simulation. *Rev. Mod. Phys.* **86**, 153–185 (2014). URL <https://link.aps.org/doi/10.1103/RevModPhys.86.153>.
 - [2] Preskill, J. Quantum computing and the entanglement frontier (2012). 1203.5813.
 - [3] Fröwis, F., Sekatski, P., Dür, W., Gisin, N. & Sangouard, N. Macroscopic quantum states: Measures, fragility, and implementations. *Rev. Mod. Phys.* **90**, 025004 (2018). URL <https://link.aps.org/doi/10.1103/RevModPhys.90.025004>.
 - [4] Paris, M. & Rehacek, J. *Quantum State Estimation*. Lecture Notes in Physics (Springer Berlin Heidelberg, 2004). URL <https://www.springer.com/gp/book/9783540223290>.
 - [5] Doherty, A. C., Parrilo, P. A. & Spedalieri, F. M. Detecting multipartite entanglement. *Phys. Rev. A* **71**, 032333 (2005). URL <https://link.aps.org/doi/10.1103/PhysRevA.71.032333>.
 - [6] Gühne, O., Reimpell, M. & Werner, R. F. Lower bounds on entanglement measures from incomplete information. *Phys. Rev. A* **77**, 052317 (2008). URL <https://link.aps.org/doi/10.1103/PhysRevA.77.052317>.
 - [7] Navascués, M., Baccari, F. & Acín, A. Entanglement marginal problems (2020). 2006.09064.
 - [8] Werner, R. F. Quantum states with einstein-podolsky-rosen correlations admitting a hidden-variable model. *Phys. Rev. A* **40**, 4277–4281 (1989). URL <https://link.aps.org/doi/10.1103/PhysRevA.40.4277>.
 - [9] Gühne, O. & Tóth, G. Entanglement detection. *Physics Reports* **474**, 1–75 (2009). URL <http://www.sciencedirect.com/science/article/pii/S0370157309000623>.
 - [10] Frérot, I. & Roscilde, T. Detecting many-body bell non-locality by solving ising models (2020). URL <https://arxiv.org/abs/2004.07796>. 2004.07796.
 - [11] Otherwise one can always reproduce the correlations among these observables using separable states by choosing the local states $|\psi_i\rangle$ to be joint eigenstates of the observables, and reproducing with p the classical correlations among the observables.
 - [12] Tóth, G., Knapp, C., Gühne, O. & Briegel, H. J. Spin squeezing and entanglement. *Phys. Rev. A* **79**, 042334 (2009). URL <https://link.aps.org/doi/10.1103/PhysRevA.79.042334>.
 - [13] Vitagliano, G., Hyllus, P., Egusquiza, I. n. L. & Tóth, G. Spin squeezing inequalities for arbitrary spin. *Phys. Rev. Lett.* **107**, 240502 (2011). URL <https://link.aps.org/doi/10.1103/PhysRevLett.107.240502>.
 - [14] d_i complex amplitudes, up to normalization of the state and an arbitrary global phase.
 - [15] Nguyen, H. C., Zecchina, R. & Berg, J. Inverse statistical problems: from the inverse Ising problem to data science. *Advances in Physics* **66**, 197–261 (2017). URL <https://doi.org/10.1080/00018732.2017.1341604>.
 - [16] Jaynes, E. T. Information theory and statistical mechanics. *Phys. Rev.* **106**, 620–630 (1957). URL <https://link.aps.org/doi/10.1103/PhysRev.106.620>.
 - [17] See Supplemental Material (SM, which includes [28–35]) for a discussion of: 1) further details of the optimization algorithm and its relationship with existing complexity results; 2) the quantum data from Quantum Monte Carlo; and 3) the existing entanglement witnesses for the quantum Ising chain.
 - [18] Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. Optimization by simulated annealing. *Science* **220**, 671–680 (1983). URL <https://science.sciencemag.org/content/220/4598/671>.
 - [19] Tóth, G. Entanglement detection in optical lattices of bosonic atoms with collective measurements. *Phys. Rev. A* **69**, 052327 (2004). URL <https://link.aps.org/doi/10.1103/PhysRevA.69.052327>.
 - [20] Wieśniak, M., Vedral, V. & Brukner, Č. Magnetic susceptibility as a macroscopic entanglement witness. *New Journal of Physics* **7**, 258–258 (2005). URL <https://doi.org/10.1088%2F1367-2630%2F7%2F1%2F258>.
 - [21] Sperling, J. & Vogel, W. Multipartite entanglement witnesses. *Phys. Rev. Lett.* **111**, 110503 (2013). URL <https://link.aps.org/doi/10.1103/>

- PhysRevLett.111.110503.
- [22] Gerke, S., Vogel, W. & Sperling, J. Numerical construction of multipartite entanglement witnesses. *Phys. Rev. X* **8**, 031047 (2018). URL <https://link.aps.org/doi/10.1103/PhysRevX.8.031047>.
 - [23] Wootters, W. K. Entanglement of formation of an arbitrary state of two qubits. *Phys. Rev. Lett.* **80**, 2245–2248 (1998). URL <https://link.aps.org/doi/10.1103/PhysRevLett.80.2245>.
 - [24] Hauke, P., Heyl, M., Tagliacozzo, L. & Zoller, P. Measuring multipartite entanglement through dynamic susceptibilities. *Nat Phys* **12**, 778–782 (2016). URL <http://dx.doi.org/10.1038/nphys3700>.
 - [25] Sachdev, S. *Quantum Phase Transitions* (Cambridge University Press, 2011). URL <https://books.google.fr/books?id=F3IkpxwpqSgC>.
 - [26] Frérot, I. & Roscilde, T. Quantum critical metrology. *Phys. Rev. Lett.* **121**, 020402 (2018). URL <https://link.aps.org/doi/10.1103/PhysRevLett.121.020402>.
 - [27] Frérot, I. & Roscilde, T. Reconstructing the quantum critical fan of strongly correlated systems via quantum correlations. *arXiv:1805.03140 [cond-mat, physics:quant-ph]* (2018). URL <http://arxiv.org/abs/1805.03140>. ArXiv: 1805.03140.
 - [28] Boyd, S. & Vandenberghe, L. *Convex Optimization* (Cambridge University Press, 2004).
 - [29] Gurvits, L. Classical deterministic complexity of edmonds’ problem and quantum entanglement. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, STOC ’03, 10–19 (Association for Computing Machinery, New York, NY, USA, 2003). URL <https://doi.org/10.1145/780542.780545>.
 - [30] Barahona, F. On the computational complexity of ising spin glass models. *Journal of Physics A: Mathematical and General* **15**, 3241–3253 (1982). URL <https://doi.org/10.1088/2F0305-4470/2F15/2F10/2F028>.
 - [31] Creutz, M. Overrelaxation and monte carlo simulation. *Phys. Rev. D* **36**, 515–519 (1987). URL <https://link.aps.org/doi/10.1103/PhysRevD.36.515>.
 - [32] Syljuåsen, O. F. & Sandvik, A. W. Quantum monte carlo with directed loops. *Phys. Rev. E* **66**, 046701 (2002). URL <https://link.aps.org/doi/10.1103/PhysRevE.66.046701>.
 - [33] Boll, M. *et al.* Spin- and density-resolved microscopy of antiferromagnetic correlations in fermi-hubbard chains. *Science* **353**, 1257–1260 (2016). URL <https://science.sciencemag.org/content/353/6305/1257>.
 - [34] Browaeys, A. & Lahaye, T. Many-body physics with individually controlled rydberg atoms. *Nature Physics* **16**, 132–142 (2020). URL <https://doi.org/10.1038/s41567-019-0733-z>.
 - [35] Pezzè, L. & Smerzi, A. Quantum theory of phase estimation. In Tino, G. & Kasevich, M. (eds.) *Proceedings of the International School of Physics “Enrico Fermi”*, 691 (IOS Press, Amsterdam, 2014). URL <https://arxiv.org/abs/1411.5164>.

Supplemental Material

In this Supplemental Material, we provide: 1) further technical details on the variational algorithm described and implemented for the data presented in the main text; 2) on the generation of quantum data, used as input to

our algorithm, by quantum Monte Carlo; 3) on the relative versus absolute violation of the entanglement witnesses; 4) on the comparison with existing entanglement criteria. In the attached .csv files, the numerical coefficients of the entanglement witnesses discussed in the main text are given.

Appendix A: Details on the algorithm

1. Mapping of the separability problem onto an inverse statistical problem

General strategy. Following the notations of the main text, we assume that the quantum data consist of a collection of expectation values $\{\langle \hat{A} \rangle_\rho\}_{a=1}^R$ measured on the quantum device. Our constructive strategy to solve the separability problem is to try and reproduce these data with a separable state – the failure of this attempt marking the success of entanglement detection. As discussed in the main text, a separable state $\hat{\rho}_p$ is defined by an arbitrary probability distribution $p(\boldsymbol{\lambda})$ over local quantum states $|\psi_i(\boldsymbol{\lambda}_i)\rangle$. Our strategy is then to build an optimal $p_{\text{opt}}(\boldsymbol{\lambda})$, such that the corresponding separable state $\hat{\rho}_{p_{\text{opt}}}$ produces the best possible approximation to the available data attainable using separable states. For a given separable state $\hat{\rho}_p$, we have:

$$\langle \hat{A}_a \rangle_{\hat{\rho}_p} = \int d\boldsymbol{\lambda} p(\boldsymbol{\lambda}) \mathcal{A}_a(\boldsymbol{\lambda}) =: \langle \mathcal{A}_a \rangle_p \quad (\text{A1})$$

where $\mathcal{A}_a(\boldsymbol{\lambda}) = \sum_{\mathbf{m}} x_{\mathbf{m}}^{(a)} \prod_{i=1}^N o_{m_i}^{(i)}(\boldsymbol{\lambda}_i)$ and $o_{m_i}^{(i)}(\boldsymbol{\lambda}_i) = \langle \psi_i(\boldsymbol{\lambda}_i) | \hat{O}_{m_i}^{(i)} | \psi_i(\boldsymbol{\lambda}_i) \rangle$ (see the main text for the precise definition of the \hat{O} ’s operators).

Our approach is in essence a *variational approach*, in which we parametrize the distribution $p(\boldsymbol{\lambda})$ as a Boltzmann distribution $p(\boldsymbol{\lambda}) = \exp[-H(\boldsymbol{\lambda})]/\mathcal{Z}(\mathbf{K})$ associated with a classical Hamiltonian $H(\boldsymbol{\lambda}) = -\sum_{a=1}^R K_a \mathcal{A}_a(\boldsymbol{\lambda})$. Two crucial properties, on which we further elaborate in this section, make this choice of Ansatz especially suited to solve the separability problem. Firstly, the expressive power of this Ansatz is *complete*, which means that there is no loss of generality in looking for a separable state of this specific form: if a separable state of this form cannot reproduce the data, then no separable state whatsoever can do so. Secondly, the variational parameters K_a can be optimized by minimizing a *convex* cost function, whose gradient can be evaluated at a cost scaling polynomially with N (the number of local subsystems) and with d (the local Hilbert space dimension), under mild assumptions (specifically, the absence of glassiness of the classical model $H(\boldsymbol{\lambda})$).

Completeness of the Ansatz. If a distribution $p(\boldsymbol{\lambda})$ exists which reproduces the data set: $\langle \mathcal{A}_a \rangle_p = \langle \hat{A} \rangle_{\hat{\rho}}$ for all $a = 1, \dots, R$, it is generically not unique. One may therefore choose the distribution which, as a further specification, maximizes the Shannon entropy functional $S[p] = -\int d\boldsymbol{\lambda} p(\boldsymbol{\lambda}) \log p(\boldsymbol{\lambda})$. This amounts to removing

any other constraints on the distribution except that of reproducing the data set with its averages. Following the seminal work of Jaynes [16], maximizing $S[p]$ under the constraint of reproducing the data is achieved upon introducing Lagrange multipliers K_a , and minimizing the functional $F[p] = -S[p] - \sum_{a=1}^R K_a [\langle \mathcal{A}_a \rangle_p - \langle \hat{A}_a \rangle_\rho]$. Setting to zero the functional derivative with respect to $p(\boldsymbol{\lambda})$ yields as a solution the Boltzmann distribution $p(\boldsymbol{\lambda}) = \exp[\sum_a K_a \mathcal{A}_a(\boldsymbol{\lambda})] / \mathcal{Z}(\mathbf{K})$. The parameters K_a are hence exactly the tuning knobs that allow $p(\boldsymbol{\lambda})$ to satisfy the constraints to the best that a classical probability distribution can do.

To further understand this point, let us stress that throughout our work we assume (as it is reasonable to do) that the size of the quantum data set scales at most polynomially with system size, so that the number of constraints associated with the reproduction of the quantum data also scales polynomially. On the other hand a distribution $p(\boldsymbol{\lambda})$ is uniquely defined by an exponentially large number of constraints – as many as the possible values of the argument $\boldsymbol{\lambda}$. The exponentially many constraints, to be added in order to specify the distribution uniquely, cannot help it in any way in reproducing the quantum data. On the other hand, maximizing the entropy of the distribution precisely gets rid of the useless constraints beyond the ones imposed by the quantum data themselves. Once the least constrained distribution is achieved upon maximizing the entropy (subject to the constraint), varying the parameters K_a of the distribution exactly allows one to reproduce all the data sets which could potentially be produced by the most general distribution $p(\boldsymbol{\lambda})$. The Boltzmann distribution associated with the classical Hamiltonian $H(\boldsymbol{\lambda}) = -\sum_a K_a \mathcal{A}_a(\boldsymbol{\lambda})$ can therefore be viewed as an Ansatz whose expressive power of quantum data sets of is as high as one can possibly achieve with a classical distribution.

Optimizing the variational parameters. We then show that the parameters K_a can be optimized upon minimizing a *convex* cost function. Convexity is a crucial property of the whole procedure: if the optimization relied on a heuristic algorithm, then the failure to reproduce the quantum data could simply mean that the optimization has been trapped in some local minimum [28], and therefore the result would be inconclusive. As stated in the main text, a convex cost function for our problem is given by $L(\mathbf{K}) = \log \mathcal{Z}(\mathbf{K}) - \sum_a K_a \langle \hat{A}_a \rangle_\rho$. Another crucial aspect for the scalability of our algorithm is that the cost function $L(\mathbf{K})$ itself is never computed. Only its gradient $g_a = \partial L / \partial K_a = \langle \mathcal{A}_a \rangle_p - \langle \hat{A}_a \rangle_\rho$ is evaluated, and used to update the parameters K_a in a gradient-descent algorithm, or any improvement thereof (in this paper, we used the accelerated gradient-descent algorithm of Nesterov). Even though the cost function itself is never computed, its existence and its convexity are key to ensure the monotonous convergence of our algorithm towards the global optimum of the problem [28]. Specifically, together with the cost function, the norm of its gradient converges towards its minimal value; namely, the distri-

bution $p(\boldsymbol{\lambda})$ converges towards the best possible approximation to the data with a separable state. If a distribution $p(\boldsymbol{\lambda})$ exists which reproduces the data, then the gradient of the cost function vanishes, and it is impossible to detect entanglement from the available data. Notice that this is not a limitation of our approach, but on the contrary it represents a fundamental property of the data that our method exhibits. On the other hand, if the data lie outside of the convex region reachable by separable states, the cost function L is not bounded from below, and the gradient will stabilize to a finite value, leading to a runaway to infinity of the coupling constants K_a , and marking the success of entanglement detection – as further discussed in the main text.

Computational complexity. Finally, we would like to remark that the computational cost required to evaluate the gradient $g_a = \langle \mathcal{A}_a \rangle_p - \langle \hat{A}_a \rangle_\rho$ with a given relative precision of ϵ via Monte Carlo methods scales as $1/\epsilon^2$. One could imagine a fine-tuned situation in which the distance between the data under investigation and the separable set is exponentially small in the system size: $|\mathbf{g}| = \mathcal{O}[\exp(-N)]$, which would translate into a computational cost of our algorithm diverging exponentially with N . While such a situation cannot be excluded *a priori*, in any practical instance the quantum data come with a finite uncertainty – certainly not decreasing exponentially with the system size. Indeed, the best scaling of the relative uncertainty that one can expect is as $N^{-1/2}$, when considering collective observables which are the sums of $\mathcal{O}(N)$ nearly independent degrees of freedom (as it happens in systems with a finite correlation length), and the same benign scaling is shared by Monte Carlo estimates of the same quantities. On the other hand, exponentially decreasing precision would require exponentially large statistics, which is not a realistic assumption for any source of the quantum data set (be it experiments or numerical calculations). As a consequence, quantum data whose distance to the separable set scales exponentially with system size would inevitably be reproduced by our algorithm using a separable state *within their uncertainty*, and at a polynomial cost.

In the literature, the separability problem has been proved to be NP-hard in the bipartite case [29]. This implies that there exists instances requiring an exponential cost in the local Hilbert space dimension d . On the other hand, we are not aware of a similar complexity result in the multipartite case, namely for a fixed d ($d = 2$ in the qubit examples treated explicitly in this work), but increasing the number N of parties. For the multipartite separability problem with N qubits, we state in the main text that classical glassy models define the practical limitation to the scalability of our approach. We would like to emphasize that this assumption is rather conservative. Indeed, the classical models one has to sample in our approach involve continuous degrees of freedom (e.g. N classical rotators representing vectors on the Bloch sphere, defining the local quantum states, see Section A 2 below). While Ising spin glasses, which involve ± 1

variables, have been proved to be NP-hard [30], a similar result does not exist for frustrated disordered classical models involving rotators (to the best of our knowledge). This (classical) statistical-physics observation is consistent with the absence of formal proof of NP-hardness of the (quantum) multipartite separability problem.

Concerning the bipartite case ($N = 2$, increasing d), whose NP-hardness is proven [29], our algorithm has a cost which is polynomial in d , in apparent contradiction with the complexity result. First, we notice that the NP-hardness [29] concerns the situation where the full bipartite state ρ_{AB} (which is a $d^2 \times d^2$ Hermitian matrix of unit trace) is used as input. Our algorithm treats a more general situation, where 1- and 2-body correlations of the form $\langle \hat{A}_a \rangle$, $\langle \hat{B}_b \rangle$, $\langle \hat{A}_a \hat{B}_b \rangle$ are known [where \hat{A}_a ($a \in \{1, \dots, R_A\}$), and \hat{B}_b ($b \in \{1, \dots, R_B\}$) are local observables on A and B subsystems, respectively]. This knowledge is equivalent to the knowledge of ρ_{AB} if \hat{A}_a and \hat{B}_b form tomographically complete sets of observables (for instance, the $R_A = R_B = d^2 - 1$ generalized Gell-Mann matrices, which are the generators of $SU(d)$). In our approach, we parametrize separable states as Boltzmann distributions related to $H(\psi_A, \psi_B) = -\sum_a K_a \mathcal{A}_a(\psi_A) - \sum_b K_b \mathcal{B}_b(\psi_B) - \sum_{ab} K_{ab} \mathcal{A}(\psi_A) \mathcal{B}_b(\psi_B)$, where ψ_A and ψ_B represent local quantum states, parametrized by $2d-2$ classical variables, and where K_a, K_b, K_{ab} are $R_A + R_B + R_A R_B$ variational parameters. The NP-hardness result [29] implies that if one considers tomographically-complete sets of observables, then there exists instances of parameters K_a, K_b, K_{ab} for which sampling the corresponding Boltzmann distribution takes a time diverging exponentially with d . We cannot immediately identify to which hard statistical physics problem this situation would correspond – but certainly such hard instances must exist, as imposed by the result of Ref. [29]. In analogy with glassy problems, for these instances the energy landscape described by $H(\psi_A, \psi_B)$ should display a myriad of local minima separated by energy barriers whose height is proportional to d , making the sampling of the model via Monte-Carlo methods inefficient.

On a more constructive tone, we would like to remark that such complexity results only refer to worst-case instances. In the context of our approach, such worst-case instances could correspond to glassy models, and in the case of translationally invariant data considered in this paper, such glassiness is avoided by construction. Such instances are not expected to be generically encountered when analyzing data from present-day quantum simulators of non-disordered systems. Finally, we would like to emphasize that there is no risk of erroneously concluding that entanglement is present if such hard instances manifest themselves. We have already argued above that realistic quantum data cannot reveal entanglement in the case of exponentially small violations of witness inequalities. In the presence of glassiness, instead, one would be unable to run the simulation forward due to very large error bars in the Monte Carlo evaluation of the expectation values for separable states. As a consequence, one would

conclude that entanglement cannot be detected within the accuracy of the method.

2. Special case: partitioning the system into N qubits

In this work we introduce a variational algorithm to fit a given data set of expectation values by using separable states. In the case of qubits taken as individual subsystems, separable states are represented without loss of generality as Boltzmann distributions over classical Heisenberg spins $\mathbf{n}^{(i)}$ on the unit sphere (which represent pure states on the Bloch sphere for individual qubits). In the examples discussed in the main text, the data set contains one-qubit expectation values $\langle \hat{\sigma}_a^{(i)} \rangle_{\hat{\rho}}$ and two-qubit correlations $\langle \hat{\sigma}_a^{(i)} \hat{\sigma}_b^{(j)} \rangle_{\hat{\rho}}$. In the examples we considered (namely, the one-dimensional antiferromagnetic Heisenberg model, and the Ising model in a transverse field, both with periodic boundary conditions), correlations $\langle \hat{\sigma}_a^{(i)} \hat{\sigma}_b^{(j)} \rangle_{\hat{\rho}}$ vanish if $a \neq b$. Since we used translationally invariant chains (with periodic boundary conditions), the one-qubit data reduces to the average magnetization $\langle m_a \rangle_{\hat{\rho}} = \sum_{i=1}^N \langle \hat{\sigma}_a^{(i)} \rangle_{\hat{\rho}} / N$, and the two-qubit correlations depend only on the inter-qubit distance: $\langle C_a^{(r)} \rangle_{\hat{\rho}} = \langle \hat{\sigma}_a^{(i)} \hat{\sigma}_a^{(i+r)} \rangle_{\hat{\rho}}$. In the case of the Heisenberg model, which displays $SU(2)$ invariance, we have $m_a = 0$. In this case, we considered as quantum data $\langle C^{(r)} \rangle_{\hat{\rho}} = \langle C_x^{(r)} + C_y^{(r)} + C_z^{(r)} \rangle_{\hat{\rho}}$.

Correspondingly, the classical Hamiltonian aiming at reproducing the quantum data contains one- and two-body interactions terms (the latter truncated beyond a given distance r_{\max}). For the Heisenberg model, we get

$$H = - \sum_i \sum_{r=1}^{r_{\max}} K^{(r)} \mathbf{n}^{(i)} \cdot \mathbf{n}^{(i+r)} ; \quad (\text{A2})$$

while for the quantum Ising model, where $m_y = m_z = 0$, we have

$$H = -K_x \sum_{i=1}^N n_x^{(i)} - \sum_{a=x,y,z} \sum_i \sum_{r=1}^{r_{\max}} K_a^{(r)} n_a^{(i)} n_a^{(i+r)} . \quad (\text{A3})$$

The K 's coefficients are the variational parameters of our algorithm, which are optimized in an iterative manner. A simple gradient-descent algorithm consists in iterating the following update rule (for the Ising model):

$$K'_x = K_x - \epsilon [\langle m_x \rangle_p - \langle m_x \rangle_{\hat{\rho}}] \quad (\text{A4})$$

$$(K_a^{(r)})' = K_a^{(r)} - \epsilon [\langle C_a^{(r)} \rangle_p - \langle C_a^{(r)} \rangle_{\hat{\rho}}] \quad (\text{A5})$$

for $a \in \{x, y, z\}$, and $r \in \{1, 2, \dots, N/2\}$; and (for the Heisenberg model):

$$(K^{(r)})' = K^{(r)} - \epsilon [\langle C^{(r)} \rangle_p - \langle C^{(r)} \rangle_{\hat{\rho}}] . \quad (\text{A6})$$

In the above equations, $\langle \cdot \rangle_p$ is the expectation value on the Boltzmann distribution for the classical Hamiltonian

(whose couplings are the K 's coefficients), while $\langle \cdot \rangle_{\hat{\rho}}$ are the target quantum data. As discussed in the main text (see also [10]), ϵ is a small parameter, implementing a numerical gradient descent of the (convex) L function. In practice, we implemented the Nesterov's accelerated gradient-descent (NAG) algorithm, with $\epsilon = 0.01$.

Each step of the NAG algorithm requires to compute the Euclidean distance \mathbf{g} between the separable data and the quantum data, namely to compute $\langle m_x \rangle_p$ and $\langle C_a^{(r)} \rangle_p$ for the Ising model and $\langle C^{(r)} \rangle_p$ for the Heisenberg model. This was implemented using Markov-chain Monte Carlo. The number of Monte Carlo steps (defined below) implemented at each step of the NAG algorithm was chosen such that the relative error on \mathbf{g} be smaller than a given threshold η , which we chose as $\eta = 0.05$ for the Ising model, and $\eta = 0.1$ for the Heisenberg model. In other words, one step of the NAG algorithm is completed when:

$$\frac{2 \sum_{\alpha} |g_{\alpha}| \text{Err}(g_{\alpha})}{|\mathbf{g}|^2} < \eta, \quad (\text{A7})$$

where $\text{Err}(g_{\alpha})$ is the error on g_{α} , as estimated from the Monte Carlo algorithm. Each step of the Monte Carlo algorithm consisted of $2N$ iterations of single-spin Metropolis updates and of single-spin microcanonical overrelaxation updates [31]. The amplitude of the proposed Metropolis updates was adapted along the Monte Carlo simulation so that the move be accepted with frequency 0.5 ± 0.1 . Therefore, a single Monte Carlo step consists of $2N$ microcanonical updates, and of N accepted Metropolis updates (on average).

As the variational optimization of the K 's parameters progresses along the NAG algorithm, the norm of the gradient \mathbf{g} decreases, and therefore an increasing number of Monte Carlo steps is required at each step of the NAG algorithm in order to achieve the required relative precision of η . When the quantum data cannot be fitted by a separable state, \mathbf{g} stabilizes to a finite value. The number of steps of the NAG algorithm to achieve this convergence (and therefore the total number of Monte Carlo steps along the whole optimization) depends on the value of $|\mathbf{g}|$ as obtained at the end of the optimization. For the examples presented in the main text, about 10^3 steps of the NAG algorithm were necessary, each of them comprising $10^4 \div 10^7$ Monte Carlo steps.

Appendix B: Quantum data from Quantum Monte Carlo

Data-driven entanglement witnessing is fundamentally based on reliable quantum data on quantum many-body systems. Here we chose to use quantum Monte Carlo data for quantum spin chains at finite temperature, obtained using Stochastic Series Expansion [32], which provides numerically exact correlation functions for the model of interest (within the statistical error bar). Finite-temperature equilibrium calculations offer

the most reliable source of data for mixed states – which pose the real challenge in terms of entanglement detection, while for pure states any form of connected correlation is an entanglement witness. Beyond their significance in condensed matter physics and quantum statistical physics, the models we chose (quantum Heisenberg and quantum Ising chain) are also of direct relevance to several experiments in quantum simulation, see e.g. [33, 34] for recent examples.

Appendix C: Existing entanglement witnesses

In this section, we provide additional details on the existing entanglement witnesses against which the quantum data of the quantum Ising model were tested (Fig. 3 of the main text).

Concurrence. The concurrence [23] defines a necessary and sufficient condition for the separability of a two-qubits density matrix. We computed the concurrence between nearest-neighbours, after reconstructing the density matrix $\hat{\rho}_{12}$ from the knowledge of one- and two-qubits expectation values $\langle \hat{\sigma}_a^{(1)} \rangle_{\hat{\rho}}$, $\langle \hat{\sigma}_b^{(2)} \rangle_{\hat{\rho}}$ and $\langle \hat{\sigma}_a^{(1)} \hat{\sigma}_b^{(2)} \rangle_{\hat{\rho}}$ (with $a, b \in \{x, y, z\}$) [4]. The dashed line on Fig. 3 defines the temperature below which $\hat{\rho}_{12}$ is entangled. Since the concurrence criterion [23] is based on a subset of the full quantum data we considered (which contains all one- and two-qubits correlations functions, which is equivalent to all two-body reduced density matrices $\hat{\rho}_{ij}$, and not only $\hat{\rho}_{12}$), by construction our data-driven method must detect entanglement in a region of the phase diagram strictly larger than the one detected by the concurrence – a fact clearly visible on Fig. 3.

Permutationally-invariant entanglement witnesses. In Ref. [12], a complete family of 8 entanglement witnesses based on the two-qubits reduced density matrix averaged over all pairs, $\hat{\rho}_{\text{av},2} = 2 \sum_{i \neq j} \hat{\rho}_{ij} / [N(N-1)]$, was derived. Equivalently, $\hat{\rho}_{\text{av},2}$ is reconstructed from the knowledge of all one- and two-body correlations averaged over all permutations: $m_a := \sum_{i=1}^N \langle \hat{\sigma}_a^{(i)} \rangle_{\hat{\rho}}$ and $C_{ab} := \sum_{i \neq j} \langle \hat{\sigma}_a^{(i)} \hat{\sigma}_b^{(j)} \rangle_{\hat{\rho}}$. Since m_a and C_{ab} are coarse-grained features of the quantum data we have considered, if an EW inequality is violated by m_a and C_{ab} (namely if one of the 8 EW inequalities of ref. [12] is violated), then our data-driven algorithm must also reconstruct a violated entanglement witnesses – in general, a more strongly violated one. As illustrated on Fig. 3 for the quantum Ising model, for which we tested all 8 criteria for each parameters (t, g) (temperature and transverse field), this is clearly the case.

Quantum Fisher information. The quantum Fisher information (QFI) is another multipartite entanglement witness. Formally, the QFI quantifies the sensitivity of the state ρ to unitary transformations $\hat{\rho}(\phi) =$

$e^{-i\phi\hat{O}}\hat{\rho}e^{i\phi\hat{O}}$ with \hat{O} a quantum observable [35]. The QFI can be expressed as $\text{QFI}(\hat{O}, \hat{\rho}) = 2 \sum_{n \neq m} (p_n - p_m)^2 |\langle n | \hat{O} | m \rangle|^2 / (p_n + p_m)$, where $\hat{\rho}$ is diagonalized as $\hat{\rho} = \sum_n p_n |n\rangle\langle n|$. Here, we chose for \hat{O} the collective spin $J_z = \sum_{i=1}^N \hat{\sigma}_z^{(i)} / 2$, which is optimal to witness entanglement around the quantum critical point of the quantum Ising model [24, 26]. The inequality $\text{QFI}(\hat{J}_z, \hat{\rho}) \leq N$ is satisfied by all separable states, so that a QFI exceeding the system size is an entanglement witness [35]. In general, computing the QFI involves the knowledge of the full density matrix ρ , but the mapping of the quantum Ising chain onto a free-fermion model [25] makes this computation tractable [24]. Notice that computing the QFI requires knowledge beyond one- and two-body correlators, and therefore it goes beyond the data set we have considered. Hence there is no guarantee a priori that our method exceeds the EW capability of the QFI. Nevertheless, as illustrated on Fig. 3, the parameter region where entanglement is detected by the QFI is broadly included in the region where entanglement is detected via our data-driven algorithm.

Appendix D: Absolute versus relative violation of the entanglement witnesses

By construction, the optimal witness found by our approach is the one whose absolute violation is maximized. Namely, among all possible witness operators $\hat{W} = -\sum_a W_a \hat{A}_a$, properly normalized with the euclidian norm $\sum_a W_a^2 = 1$, our witness operator maximizes the *difference* $B_{\text{sep}} - \text{Tr}(\hat{\rho}\hat{W})$, where $B_{\text{sep}} = \min_{\hat{\rho}_{\text{sep}}} \text{Tr}(\hat{\rho}_{\text{sep}}\hat{W})$. As a consequence, it is the witness operator which is most robust to the uncertainty present on the quantum data, namely the one that requires the lowest amount of statistics producing the quantum data themselves.

On the other hand, it is also relevant to consider the *noise robustness* of a given entanglement witness, namely the robustness to a noisy, imperfect preparation of the quantum state $\hat{\rho}$ that should produce the quantum data. Noisy state preparation can be generically modeled as turning the target state into $(1 - \eta)\hat{\rho} + \eta\mathbb{1}/D$, where η parametrizes the strength of the noise, and D is the total Hilber space dimension. Assuming that all operators \hat{A}_a composing the witness are traceless (which is the case for tensor products of local Pauli matrices, as considered in this paper), this leads us to define the noise robustness as the maximal value of η such that $(1 - \eta_{\text{max}})\text{Tr}(\hat{\rho}\hat{W}) = B_{\text{sep}}$, namely $\eta_{\text{max}} = 1 - B_{\text{sep}}/\text{Tr}(\hat{\rho}\hat{W})$. There is no guarantee that the witnesses found by our approach are those whose noise robustness is maximal, and in fact it turns out not to be the case, as shown by the following example.

In the case of the Heisenberg chain, we have considered translationally-invariant entanglement witnesses of the

form $\hat{W} = -\sum_{a \in \{x,y,z\}} \sum_{i=1}^N \sum_{r=1}^{r_{\text{max}}} w_r \hat{\sigma}_a^{(i)} \hat{\sigma}_a^{(i+r)}$. Our convention has been to normalize them to $\sum_r w_r^2 = 1$. For a meaningful comparison with the PIEW $\langle [\sum_{i=1}^N \hat{\mathbf{S}}^{(i)}]^2 \rangle \geq N/2$, the latter should be properly normalized according to the same convention, namely: $\hat{W}_{\text{PIEW}} = (N - 1)^{-1/2} \sum_{a \in \{x,y,z\}} \sum_{i=1}^N \sum_{j \neq i} \hat{\sigma}_a^{(i)} \hat{\sigma}_a^{(j)}$, with a separable bound given by $-N/\sqrt{N-1}$. For the data considered in the main text (namely, a thermal state of the one-dimensional Heisenberg model at temperature $T/J = 1.00$ with $N = 64$ spins), we find a violation $-N/\sqrt{N-1} - \text{Tr}(\hat{\rho}\hat{W}_{\text{PIEW}}) = 0.04552$. In contrast, the optimal witness found by our data-driven approach exhibits a larger violation of 0.10570. On the other hand, the noise robustness of the PIEW is $\eta_{\text{max}} = 0.255$, while the noise robustness of the data-driven EW found by our approach is $\eta_{\text{max}} = 0.174$. This is qualitatively consistent with the observation that the PIEW is violated up a temperature of $T/J \approx 1.400$, which is higher than the temperature up to which the data-driven EW (optimal by construction at $T/J = 1.00$) is violated.

Appendix E: Detailed numerical values of the entanglement witnesses

The numerical coefficients of the entanglement witnesses reconstructed by our algorithm are given in this Section. For the Heisenberg model at temperature $T/J = 1$ (Fig. 2 of the main text), we discarded the correlations at distances beyond $r = 10$. The coefficients $K^{(r)}$ of the entanglement witness are given in Table I. The separable bound is $E/N = -0.503248446$ ($N = 64$). For

distance	$K^{(r)}$
1	-0.84192229
2	-0.50632705
3	-0.16643027
4	0.03072345
5	0.06820392
6	0.02322838
7	-0.01850183
8	-0.02146924
9	0.00126596
10	0.01267421

TABLE I: Coefficients of the data-driven entanglement witness presented on Fig. 2 of the main text (Heisenberg chain $T/J = 1.00$).

the quantum Ising model (Fig. 3 of the main text), the corresponding coefficients K_x , $K_x^{(r)}$, $K_y^{(r)}$ and $K_z^{(r)}$ are presented on Table II. The separable bound is $E/N = -0.465475151529285$ ($N = 64$).

distance	K_x	K_{xx}	K_{yy}	K_{zz}
0	0.078863109939705			
1		0.336212178177562	-0.53946940308646	0.701821871161535
2		0.006136971013906	-0.186782505835586	-0.213519552791139
3		-0.001090090530485	-0.00257250626173	-0.098573694386247
4		-0.016994651046548	0.023567559817685	0.057667124559048
5		0.003961220213939	-0.003352864510561	0.032995580923791
6		0.006728876469642	-0.006668943216078	-0.024548154286714
7		-0.001277958269768	0.003102691532182	-0.016751237825024
8		-0.000351393929815	0.001659840517445	0.012307823520514
9		0.002460403552907	-0.001929489352437	0.010912062898113
10		0.000818810042566	5.60569829502012E-05	-0.007413177478221
11		0.001440162056544	0.000798869250342	-0.00808541701429
12		0.001464201727649	-0.0003416423396	0.00459909928294
13		0.001744703197231	-0.000212938810296	0.006913583983173
14		0.001100010416802	0.000223069933552	-0.00293630357091
15		0.001061273315343	4.66867175896627E-06	-0.006226408282977
16		0.001462284817869	-0.000101546964095	0.001484817717649
17		0.001304160305372	3.77065191741915E-05	0.006024060541615
18		0.001179839161665	3.78654896794631E-05	-0.00071186404194
19		0.000350620499287	-2.25876364328412E-05	-0.006559629225319
20		0.004041162071268	-8.36015569473183E-06	0.003989414923715

TABLE II: Coefficients of the data-driven entanglement witness used on Fig. 3 of the main text (transverse-field Ising chain for $g = 0.5$ and $T/J = 0.28$).