

Observation of Entangled States of a Fully Controlled 20-Qubit System

Nicolai Friis,^{1,‡} Oliver Marty,^{2,‡} Christine Maier,^{3,4} Cornelius Hempel,^{3,4,†} Milan Holzäpfel,² Petar Jurcevic,^{3,4} Martin B. Plenio,² Marcus Huber,¹ Christian Roos,³ Rainer Blatt,^{3,4} and Ben Lanyon^{3,*}

¹*Institute for Quantum Optics and Quantum Information-IQOQI Vienna, Austrian Academy of Sciences, Boltzmannngasse 3, 1090 Vienna, Austria*

²*Institut für Theoretische Physik and IQST, Albert-Einstein-Allee 11, Universität Ulm, 89069 Ulm, Germany*

³*Institute for Quantum Optics and Quantum Information, Austrian Academy of Sciences, Technikerstraße 21a, 6020 Innsbruck, Austria*

⁴*Institut für Experimentalphysik, Universität Innsbruck, Technikerstraße 25, 6020 Innsbruck, Austria*



(Received 11 December 2017; revised manuscript received 13 February 2018; published 10 April 2018)

We generate and characterize entangled states of a register of 20 individually controlled qubits, where each qubit is encoded into the electronic state of a trapped atomic ion. Entanglement is generated amongst the qubits during the out-of-equilibrium dynamics of an Ising-type Hamiltonian, engineered via laser fields. Since the qubit-qubit interactions decay with distance, entanglement is generated at early times predominantly between neighboring groups of qubits. We characterize entanglement between these groups by designing and applying witnesses for genuine multipartite entanglement. Our results show that, during the dynamical evolution, all neighboring qubit pairs, triplets, most quadruplets, and some quintuplets simultaneously develop genuine multipartite entanglement. Witnessing genuine multipartite entanglement in larger groups of qubits in our system remains an open challenge.

DOI: [10.1103/PhysRevX.8.021012](https://doi.org/10.1103/PhysRevX.8.021012)

Subject Areas: Atomic and Molecular Physics,
Quantum Physics, Quantum Information

I. INTRODUCTION

The ability to generate quantum entanglement [1,2] between large numbers of spatially separated and individually controllable quantum systems—such as qubits—is of fundamental importance to a broad range of current research endeavors, including studies of nonlocality [3], quantum computing [4], quantum simulation [5], quantum communication [6,7], and quantum metrology [8–11]. For example, in order for quantum computers and simulators to go beyond the capabilities of conventional computers, large amounts of entanglement (or other quantum correlations) must be generated between their components [12].

As such, there is an ongoing effort to generate and characterize entangled states of increasing numbers of qubits, in systems which permit preparation of arbitrary

initial states, the control of interactions between constituent particles, and readout of individual sites. In such systems, the largest number of qubits entangled to date is 14, achieved in a trapped-ion system [13], followed by 10 entangled superconducting qubits [14], and 10 entangled photonic qubits [15].

Since every qubit added to an experimental system doubles the Hilbert space dimension in which the collective quantum state is described, the task of characterization of an unknown state in the laboratory can soon become a significant challenge. Indeed, all generated entangled states of more than 6 qubits to date have been of a highly symmetric form, such as Greenberger-Horne-Zeilinger (GHZ) or W states, for which efficient characterization techniques exist [16]. How to generate and detect more complex multiqubit entangled states remains an open challenge.

In this paper, we report on the deterministic generation of complex entangled states of 20 trapped-ion qubits and their partial characterization via custom-built witnesses for genuine multipartite entanglement (GME). Our states are complex in the sense that they are generated during quench dynamics of an engineered many-body Hamiltonian and their exact description requires specifying a number of parameters that grows exponentially in the number of qubits involved. Each qubit in our system can be, and is in this work, individually manipulated and read

*ben.lanyon@uibk.ac.at

†Present address: ARC Centre for Engineered Quantum Systems, School of Physics, The University of Sydney, 2006 New South Wales, Australia.

‡These authors contributed equally to this work.

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

out, as required for universal quantum computation and quantum simulation.

This paper is structured as follows. Section II presents the experimental system and explains how the 20-qubit quantum states are generated and measured in the laboratory. Section III presents results of basic properties measured for the generated 20-qubit states, using established methods. Section IV introduces and applies analytically derived GME witnesses to reveal genuine tripartite entanglement in all groups of 3 neighboring qubits. To go beyond tripartite correlations, we then turn to more computationally demanding witnesses in Sec. V, which enable GME to be detected in groups of up to 5 neighboring qubits. Finally, we discuss our results and possible future directions in Sec. VI.

II. EXPERIMENTAL SETUP

Our register of $N = 20$ qubits is realized using a 1D string of $^{40}\text{Ca}^+$ ions confined in a linear Paul trap, with axial (radial) center-of-mass vibrational frequency of 220 kHz (2.712 MHz) [17]. A qubit is encoded into two long-lived states of the outer valence electron in each ion. That is, the computational basis states of the qubits are chosen as $|0\rangle = |S_{J=1/2, m_j=1/2}\rangle$, $|1\rangle = |D_{J=5/2, m_j=5/2}\rangle$, which are connected by an electric quadrupole transition at 729 nm [18].

Under the influence of laser-induced forces that off-resonantly drive all 40 transverse normal vibrational modes of the ion string, the interactions between the qubits are well described by an “XY” model in a dominant transverse field [19–21], with Hamiltonian

$$H_{XY} = \hbar \sum_{i < j} J_{ij} (\sigma_i^+ \sigma_j^- + \sigma_i^- \sigma_j^+) + \hbar B \sum_j \sigma_j^z. \quad (1)$$

Here J_{ij} is an $N \times N$ qubit-qubit coupling matrix, σ_i^+ (σ_i^-) is the qubit raising (lowering) operator for qubit i , B is the transverse field strength ($B \gg \max\{|J_{ij}|\}$), and $\sigma_j^z \equiv Z_j$ is the Pauli Z matrix for qubit j with eigenvectors satisfying $Z|0\rangle = -|0\rangle$ and $Z|1\rangle = |1\rangle$. Interactions reduce approximately with a power law $J_{ij} \propto 1/|i-j|^\alpha$ with qubit separation number $|i-j|$, where in this work $\alpha \approx 1.1$.

The ground state of H_{XY} has all qubits in the state $|0\rangle$. The excited states are split into m uncoupled and non-degenerate manifolds. Each manifold contains an integer number of qubit excitations (qubits in the state $|1\rangle$), with the m th manifold containing states with m qubit excitations. In previous work, we have shown that an initial state consisting of a single localized qubit excitation coherently disperses in the system, distributing quantum correlations as it propagates [22]. Here, we study the entanglement generated during the time evolution of the initial 20-qubit Néel-ordered product state $|\psi(t=0)\rangle = |1, 0, 1, \dots\rangle$ under H_{XY} . That is, we study the state in the laboratory that is

ideally described by $|\psi(t)\rangle = \exp(-iH_{XY}t)|\psi(0)\rangle$. The initial Néel state $|\psi(0)\rangle$ contains localized qubit excitations at every other site, which should ideally coherently disperse in the subsequent dynamics, entangling groups of neighboring qubits, as we have previously shown for neighboring pairs with a string of up to 14 qubits [19]. While Ref. [19] presented scalable tomography techniques [23–25], here we study multipartite entanglement dynamics in the system.

The initial state is prepared as follows. Standard Doppler cooling, optical pumping, and resolved-sideband cooling prepare the initial qubit state $|0, 0, 0, \dots, 0\rangle$ and all 40 transverse vibrational string modes into the motional ground state [19,20]. Next, a combination of qubit-resonant laser beams that illuminate all ions simultaneously and off-resonant single-ion-focused laser beams flip every second qubit, preparing the state $|\psi(0)\rangle$. The interactions (laser-induced forces) simulating H_{XY} are then turned on.

After the desired evolution time t , the interactions are turned off and the state [ideally $|\psi(t)\rangle$] is measured via qubit-state-dependent resonance fluorescence, using a single-ion-resolved electron multiplying charge coupled device (EMCCD) camera. Specifically, detecting a fluorescing (nonfluorescing) ion corresponds to the measurement outcome $|0\rangle$ ($|1\rangle$). Such a measurement corresponds to projecting each of the 20 qubits into the eigenstates of the Pauli Z operator, for which there are 2^{20} possible outcomes each corresponding to a 20-qubit projective measurement outcome. After repeated state preparations and measurements in the “Z basis,” any single-qubit expectation value ($\langle Z_i \rangle$), 2-qubit correlator ($\langle Z_i Z_j \rangle$), or indeed any other n -qubit “Z-type” correlator can be estimated between any qubits (up to $n = 20$). Performing single-qubit operations, with a single-ion-focused laser, before the aforementioned measurement process enables projective measurement of any qubit in any desired basis, and therefore the construction of any multiqubit correlation function. That is, in this work full local control over the individual qubits is available and necessary for state preparation and analysis. A conceptual schematic of our experimental protocol is presented in Fig. 1.

One approach to studying the entanglement properties of an N -qubit system is to perform full quantum state tomography to estimate the N -qubit density matrix and then develop and apply entanglement measures to that matrix. While this is technically possible for our 20-qubit system (i.e., the required measurements can each be performed in principle), it is practically not feasible as, e.g., billions of measurement bases are required. In general, the number of measurement bases required for full state tomography grows exponentially in N as 3^N . Several of us have recently shown that matrix product state (MPS) tomography can provide a pure-state estimate of states generated in quantum systems with finite-range interactions, using a number of measurements (and all other resources) that scales efficiently (polynomially) with the system size [19]. However,

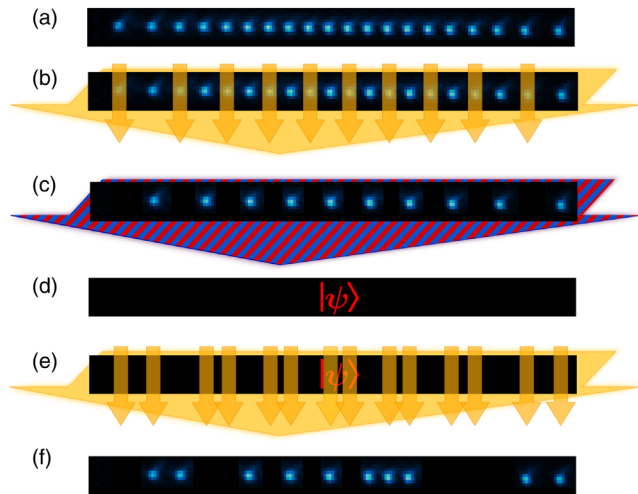


FIG. 1. Conceptual schematic of the experiment. The experimental sequence proceeds as follows. (a) Standard Doppler cooling, optical pumping, and resolved-sideband cooling prepare the initial qubit state $|0, 0, 0, \dots, 0\rangle$ and all 40 transverse vibrational string modes close (< 1 phonon per mode) to the motional ground state [19,20]. An image of the 20 ions in our trap during state-dependent fluorescence detection of the state $|0, 0, 0, \dots, 0\rangle$ is shown. The string length is $108 \mu\text{m}$. (b) A combination of single-ion-focused and global laser beams prepares the initial state $|\psi(t=0)\rangle = |1, 0, 1, \dots\rangle$. (c) A bichromatic light field is applied to $|\psi(t=0)\rangle$, subsequently inducing qubit-qubit interactions (not shown) as described in the text [19–21]. (d) After any desired evolution time, the interactions are turned off, leaving a nonclassical state of qubits (well approximated by) $|\psi(t)\rangle$. (e) A combination of single-ion-focused and global laser beams is used to rotate the basis of individual qubits, determining the desired measurement basis in the next step. An example laser pattern is shown. (f) The standard state-dependent resonant fluorescence technique is used to determine the state of each qubit (see Methods in Ref. [19] for more details). An example outcome, imaged on an EMCCD camera, is shown schematically. The ions are then cooled again and initialized, ready for the sequence to be repeated.

MPS tomography failed to produce a useful pure-state description in our present 20-qubit system, probably due to errors in preparation of the initial state that lead to mixed states and the long-range nature of the interactions present in our 20-qubit Hamiltonian.

A more favorable approach to detecting and characterizing entanglement in N -qubit systems is to develop entanglement witnesses that are not a function of every element of the density matrix and can be directly measured in the laboratory with a practical number of measurements.

III. INITIAL RESULTS: MAGNETIZATION AND ENTANGLEMENT

As the first experimental step, we prepare the time-evolved state of our system [ideally described by $|\psi(t)\rangle$] and measure each qubit in the Z basis. The dynamical

evolution of $\langle Z_i \rangle$ (proportional to the probability of finding a qubit excitation at site i) shows how the multiple excitations in the initial state disperse, and then partially refocus at later times, in close agreement with predictions from the exact model [see Figs. 2(a) and 2(b)]. One sees in the data and theory [Figs. 2(a) and 2(b)] that the qubit excitations at the ends of the string disperse more slowly than those in the center of the string (the green color representing “delocalized” excitations appears at later times for qubits farther from the center of the string). The qubit-qubit interactions (J_{ij}) are not homogeneous across the string, leading to slower evolution for qubits farther away from the center. J_{ij} for spin pairs near the string center is approximately 25% larger than for spin pairs located at either end of the string. The physical origin of this effect is a combination of the Gaussian intensity profile of the laser fields involved in generating the interactions (the laser beam is centered on the middle ions and therefore weaker on the outer ions) and the intrinsic interaction inhomogeneity across the string set by the ion-string vibrational mode frequencies, for our chosen trapping parameters. Both effects are taken into account in the theoretical model presented in Fig. 2(b).

As the final experimental step, we perform the set of measurements that would be sufficient to reconstruct (via full quantum state tomography) the density matrices of all neighboring $k = 3$ qubits (qubit triplets), during the simulator dynamics. This set consists of $3^k = 27$ measurement bases (with 1000 measurements performed per basis), corresponding to all possible combinations of choosing

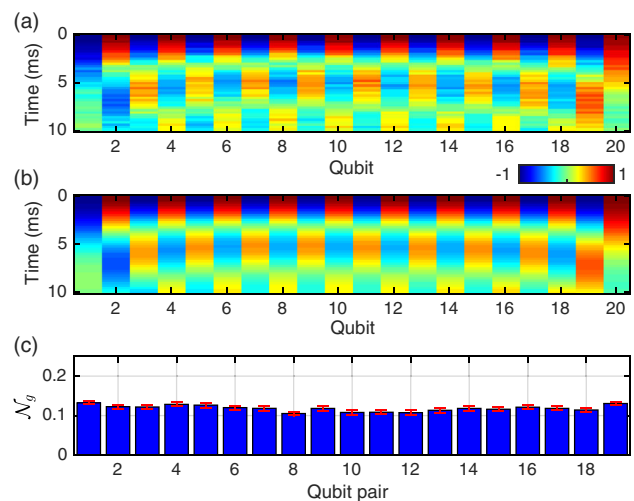


FIG. 2. Quantum dynamics of the 20-qubit register. (a) Experimental data. Single-qubit magnetization $\langle Z_i \rangle$. (b) Theoretical model. Single-qubit magnetization $\langle Z_i \rangle$ for the exact time-evolved state $|\psi(t)\rangle$. (c) Experimental data. Entanglement in all 19 neighboring qubit pairs at evolution time $t = 2$ ms, quantified by the genuine multipartite negativity [26,27]. Values for entanglement are calculated from experimentally reconstructed 2-qubit density matrices. Errors are 1 s.d. derived from Monte Carlo simulation of finite measurement number.

three Pauli operators. We carry out a simple scheme (choice of measured Pauli operators) that allows measurements on all 18 neighboring qubit triplets (out of the 20-qubit string) to be performed in parallel, requiring a total of only twenty-seven 20-qubit measurement bases. From this data set, we could reconstruct the density matrices of all single qubits, neighboring qubit pairs, and neighboring qubit triplets. Generalizing this approach to arbitrary k , all $N - k + 1$ groups of neighboring k -qubit density matrices in an N -qubit string can be fully characterized by measuring in 3^k bases (independent of the number of qubits N). For fixed k , this measurement approach is clearly efficient (constant overhead) in the system size N . We nonetheless stop at $k = 3$, as the number of measurement bases for $k = 4$ is already quite demanding and, as we show, $k = 3$ is already sufficient to observe genuine multipartite entanglement in groups of up to five qubits.

We reconstruct the density matrices of all neighboring qubit pairs from the experimental data, via the standard maximum likelihood method, which finds the most likely physical density matrix to have produced the data. For each of the reconstructed 2-qubit states we evaluate the genuine multipartite negativity \mathcal{N}_g , an established measure for GME [26,27]. A positive value of \mathcal{N}_g for a given k -qubit state implies the existence of genuine k -partite entanglement in this state, since \mathcal{N}_g vanishes for all biseparable states. For two qubits, \mathcal{N}_g is directly related to the logarithmic negativity [28]. More details on \mathcal{N}_g are given in Sec. V. From the results one sees that all neighboring qubit pairs become entangled during the time evolution of the system, as is shown for $t = 2$ ms in Fig. 2(c). Error bars, on properties calculated using the tomographically reconstructed density matrices, are derived from the finite number of measurements (1000 for each global basis) used to estimate expectation values and calculated using the standard Monte Carlo method [29].

Naturally, one may wonder if entanglement extends beyond qubit pairs, for instance, in the form of bipartite entanglement between distant qubits or in terms of genuine multipartite entanglement between groups of more than two (adjacent) qubits. In fact, one may even be tempted to ask, is multipartite entanglement not implied if every neighboring qubit pair is entangled? The answer to this question is simply no. One can indeed have states that feature entanglement in every 2-qubit reduction, yet still feature only bipartite entanglement [30]. Nonetheless, in theory, it is often possible to detect GME purely from inspection of the reduced density matrices of overlapping groups of qubits [16]. This is the basis for our first approach to detecting GME, presented in the next section, where we derive GME witnesses purely from neighboring two-body observables.

In general, the task of determining if and how an N -qubit quantum system is entangled is highly nontrivial. For an arbitrary *known* mixed state (e.g., reconstructed via full

state tomography), the problem is at least NP hard and even the best known relaxations are semidefinite programs (SDPs) that are not feasible beyond 5 qubits with our computers [26,31]. However, if the density matrix is close to a given pure target state $|\psi_T\rangle$, then targeted witnesses can be constructed to detect its entanglement without resorting to full tomography [16]. There, the canonical ansatz would be to estimate the fidelity to the ideal target state $\text{Tr}(\rho|\psi_T\rangle\langle\psi_T|)$ and check whether it is above the maximum possible fidelity of a biseparable state. That is, a corresponding witness would be $W = \beta\mathbb{1} - |\psi_T\rangle\langle\psi_T|$, where $\beta := \min_{A|\bar{A}} \|\text{Tr}_{\bar{A}}(|\psi_T\rangle\langle\psi_T|)\|_\infty$; see, e.g., Ref. [16], Sec. 3.6. While this witness could in principle be successful in detecting GME if the experimental state is indeed very close to the intended pure state, it suffers from poor noise resistance [32] and the task of determining the state fidelity for arbitrary pure states still requires a number of measurement settings that scales exponentially in qubit number [33].

For example, if the state is “well conditioned” [33], i.e., if only few Pauli-expectation values are of a significant size and all others vanish, one could estimate the fidelity via randomized measurements [33] with effort that scales efficiently in qubit number. Although our states are not well conditioned, in Ref. [19] we implemented this randomized measurement strategy to obtain a fidelity estimate for a 14-qubit version of the states presented here, at one time step. That experiment required preparing 5×10^5 sequential copies of the state and involved over 5 hours of data taking (with periodic recalibration of experimental parameters). Numerical simulations of the randomized measurement technique applied to the 20-qubit states considered in this work show that more than 3 times the number of copies, and therefore impractical measurement time, would be required to yield accurate fidelity estimations. As such, we aim to develop novel, and more time-efficient, approaches to characterizing entanglement in our 20-qubit system. As a first step in this direction, we focus on the dynamics of subsystem entanglement percolating through the system and are able to make statements about the entanglement using measurements completed in a few tens of minutes in our system.

IV. GME WITNESSES BASED ON 2-QUBIT OBSERVABLES

In this section, we construct analytical witnesses for GME based on 2-qubit observables and use them to detect GME in groups of up to three neighboring qubits ($k = 3$) within the 20-qubit ($N = 20$) register. Recall that a (multipartite) pure state $|\psi\rangle$ is called biseparable if there exists a bipartition $A|B$ such that $|\psi\rangle = |\phi\rangle_A|\chi\rangle_B$ for some $|\phi\rangle_A$ and $|\chi\rangle_B$, and is called genuinely multipartite entangled otherwise. Mixed states are GME if their density operators cannot be written as convex combinations of biseparable pure states. For more details, see Appendix A 2.

Following the observation in the previous section of strong entanglement between neighboring qubits, the first type of GME witnesses we consider is based on average fidelities of the 2-qubit density matrices with Bell states. As such, only expectation values of pairs of Pauli operators, on k -qubit subsets of choice, are required. Linear combinations of these expectation values are then evaluated and compared to their respective thresholds for biseparable k -qubit states. Surpassing a k -qubit biseparability threshold then detects genuine k -partite entanglement.

We now present a short technical summary of our method, and refer the reader to Appendix A for more details. The main quantity of interest for detecting k -qubit GME in this section is the k -qubit symmetric average Bell fidelity $\bar{\mathcal{F}}_{\text{Bell}}^{(k)}$, which we define as

$$\bar{\mathcal{F}}_{\text{Bell}}^{(k)} := \frac{1}{4b_k} \left(b_k + \sum_{\substack{i,j=1 \\ i < j}}^k (|\langle \tilde{X}_i \tilde{X}_j \rangle| + |\langle \tilde{Y}_i \tilde{Y}_j \rangle| + |\langle \tilde{Z}_i \tilde{Z}_j \rangle|) \right), \quad (2)$$

where $b_k = \binom{k}{2} = \frac{1}{2} \{(k!)/[(k-2)!]\}$, and the subscripts i and j denote operators acting nontrivially only on the i th and j th qubits; i.e., $O_i \equiv \mathbb{1}_1 \otimes \dots \otimes \mathbb{1}_{i-1} \otimes O_i \otimes \mathbb{1}_{i+1} \otimes \dots \otimes \mathbb{1}_N$. The operator triple $\tilde{X}_i = U_i X U_i^\dagger$, $\tilde{Y}_i = U_i Y U_i^\dagger$, and $\tilde{Z}_i = U_i Z U_i^\dagger$ is chosen unitarily equivalent to the usual triple of Pauli operators X , Y , and Z , although the unitary $U_i \in SU(2)$ may be chosen differently for each qubit (for each i). This ensures that $\bar{\mathcal{F}}_{\text{Bell}}^{(k)}$ can be written as a linear combination (the absolute values can be replaced by appropriate sign changes) of pairs of Pauli operators. As we show in detail in Appendix A, any quantum state of k qubits for which

$$\bar{\mathcal{F}}_{\text{Bell}}^{(k)} > \begin{cases} \frac{1}{12}(3 + \sqrt{15}) & \text{for } k = 3 \\ \frac{1}{4}(1 + \sqrt{3}) - \frac{1}{2k}(\sqrt{3} - 1) & \text{for } k \geq 4 \end{cases} \quad (3)$$

is genuinely k -partite entangled for any choice of U_1, \dots, U_k . For example, for $k = 3$ and $k = 4$ one can detect GME for $\bar{\mathcal{F}}_{\text{Bell}}^{(3)} > \frac{1}{12}(3 + \sqrt{15}) \approx 0.573$ and for $\bar{\mathcal{F}}_{\text{Bell}}^{(4)} > \frac{1}{8}(3 + \sqrt{3}) \approx 0.592$, respectively. Meanwhile, the threshold for $k = 2$ qubits, $\bar{\mathcal{F}}_{\text{Bell}}^{(2)} > 0.5$, is a well-known result.

If the underlying N -partite quantum state is known (the U_i are known), one could directly measure all of the $3b_k$ 2-qubit correlators $\langle \tilde{O}_i \tilde{O}_j \rangle$ appearing in Eq. (2) for optimally chosen $\{U_i\}$. However, when the optimal local measurements are unknown, one strategy is to measure the $6k$ k -qubit basis settings corresponding to the set

$$\left\{ O_{XY}^{(i)}, O_{YX}^{(i)}, O_{XZ}^{(i)}, O_{ZX}^{(i)}, O_{YZ}^{(i)}, O_{ZY}^{(i)} \right\}_{i=1, \dots, k}, \quad (4)$$

where $O_{AB}^{(i)} = A_i \prod_{i \neq j} B_j$, and perform the optimization when evaluating the corresponding results. From the 2^k outcomes of each of these simultaneous measurements of k qubits one can obtain the expectation values of all pairwise combinations of Pauli operators.

For our purposes, we exploit the fact that the results from the twenty-seven 20-qubit measurement bases already taken in the laboratory are also sufficient to calculate all the expectation values appearing in the witnesses $\bar{\mathcal{F}}_{\text{Bell}}^{(k)}$ for $k = 2$ and $k = 3$. That is, they contain as a subset, all the 2-qubit observables required to calculate $\bar{\mathcal{F}}_{\text{Bell}}^{(2)}$ and $\bar{\mathcal{F}}_{\text{Bell}}^{(3)}$, without knowledge of the states. The results, for increasing system interaction times and for the optimization of the U_i limited to the X - Y plane for each qubit, are shown in Fig. 3. First, the witness for bipartite entanglement ($\bar{\mathcal{F}}_{\text{Bell}}^{(2)} > 0.5$) reaffirms that all qubits are (bipartite) entangled with their direct neighbors throughout the interaction time (from time 0.5 to 3.5 ms). Second, genuine tripartite entanglement between neighboring qubit triples builds up more slowly and is initially detected at time 1.5 ms. At time 2 ms, most triples of neighboring qubits are genuinely tripartite entangled, before the GME gradually disappears again at later times.

The experimental uncertainties (error bars) in Fig. 3 originate from a finite number of measurements used to estimate expectation values. Specifically, the error bars show an estimate of 1 standard deviation of the mean. When estimating the standard deviation of the mean, one must consider possible correlations between 2-qubit expectation values if they are estimated from outcomes of the same 20-qubit measurement basis. We estimate the relevant variance as described in Ref. [19] (therein see Supplemental Material Secs. IV.A.4 and IV.A.5, pp. 9–13).

The small deviations between the theoretical and measured dynamics of $\bar{\mathcal{F}}_{\text{Bell}}^{(2)}$ and $\bar{\mathcal{F}}_{\text{Bell}}^{(3)}$ in Fig. 3 are due to experimental imperfections, which we discuss in the next section.

A number of interesting observations, based on analyses beyond those presented in Fig. 3, are now made. First, up to the evolution time presented in Fig. 3, entanglement between any 2-qubits spaced farther apart than direct neighbors was never detected—in agreement with the ideal theoretical model. For instance, on these timescales, qubit 1 does not directly become entangled with qubit 3 alone, but qubits 1, 2, and 3 do become genuinely tripartite entangled with each other. In fact, the absence of next-nearest-neighbor pairwise entanglement was necessary to detect 3-qubit GME. Specifically, we found that a GME witness based only on the entanglement between direct neighbors (see Appendix A 2 b) is not able to verify genuine tripartite entanglement, and it was only possible to do so once the (separable) correlations between non-neighboring qubits (e.g., qubits 1 and 3) are also taken into account.

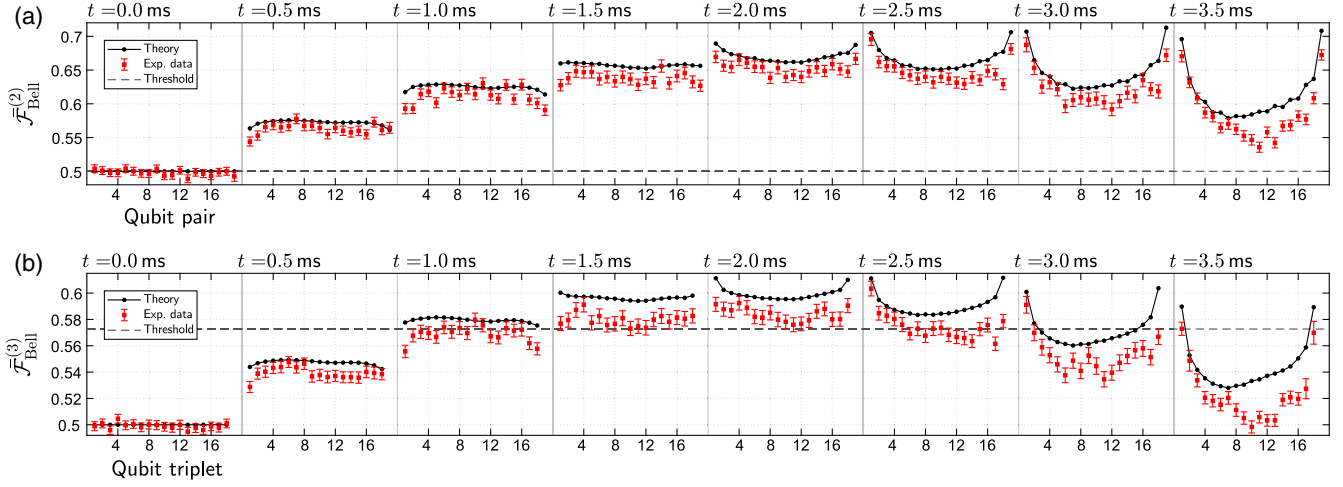


FIG. 3. Entanglement witnesses based on symmetric average Bell fidelities. The experimental results (red) and theoretical predictions based on ideal time-evolved state $|\psi(t)\rangle$ (blue) for the entanglement witnesses $\bar{\mathcal{F}}_{\text{Bell}}^{(2)}$ and $\bar{\mathcal{F}}_{\text{Bell}}^{(3)}$ are shown in (a) and (b), respectively. The horizontally arranged panels show the results at different time steps, 0.0 ms, 0.5 ms, 1.0 ms, etc., with intervals of 0.5 ms during the time evolution of the 20-qubit chain, starting with 0 ms (the initial state). Within each panel, each dot represents a pair (a) or triple (b) of neighboring qubits. The horizontal dashed lines indicate the detection thresholds for bipartite [(a) $\bar{\mathcal{F}}_{\text{Bell}}^{(2)} > 0.5$] and genuine tripartite entanglement [(b) $\bar{\mathcal{F}}_{\text{Bell}}^{(3)} > \frac{1}{12}(3 + \sqrt{15}) \approx 0.573$], respectively. The error bars for each qubit pair or triple represent 1 standard deviation of the mean in each direction. It can be seen in (a) that all qubits immediately become entangled with their direct neighbors and remain entangled throughout, whereas genuine tripartite entanglement is detected in time step 1.0 ms for the first time. At time step 2.0 ms, the witness $\bar{\mathcal{F}}_{\text{Bell}}^{(3)}$ indicates that all neighboring qubit triples are genuinely tripartite entangled simultaneously (although the witness is less than 1 standard deviation above the threshold for two of these triples).

Second, although there are states for which the witness of Eqs. (2) and (3) could be used to detect GME between more than 3 parties [34], it is not sensitive enough to do so for the states presented here in our setup (neither for the theoretical predictions nor for the experimental data).

To address the question of whether genuine multipartite quantum correlations occur in groups of more than 3 qubits, in our setup, we hence turn to more computationally demanding procedures, which we present in the next section.

The observed and predicted entanglement peak amplitude and dynamics for qubits near the center and those near the ends (Fig. 3) are markedly different. We attribute those differences to the interaction inhomogeneity across the qubit string and boundary effects.

V. GME WITNESSES BASED ON NUMERICAL SEARCH

In this section, we present and apply a method that employs a numerical search to find k -qubit witnesses for GME. This search is computationally intensive: an optimization is performed that takes computational resources that increase exponentially with k . Finding a GME witness operator for mixed 5-qubit states is already at the practical limit of our available computers and algorithms. Nonetheless, we find witnesses that succeed in detecting GME in groups of up to 5 qubits in our 20-qubit experimental system. In the following, we give a brief

overview of the new witnesses and defer to Appendix B for a more detailed discussion of the technical aspects.

We make use of the genuine multipartite negativity (\mathcal{N}_g), an established measure for GME [26,27]. A positive value of \mathcal{N}_g for a given k -qubit state implies the existence of genuine k -partite entanglement in this state, since \mathcal{N}_g vanishes for all biseparable states. The \mathcal{N}_g can be calculated given knowledge of the density matrix [26,27]. However, we have not performed a tomographically complete set of measurements for more than $k = 3$ qubits (we do not have the density matrices for the state in the lab, for $k > 3$). Our approach, to detect GME in any given group i , of k qubits, is to find a k -qubit witness operator $Q_i^{(k)}$ whose expectation value provides a lower bound on the k -qubit \mathcal{N}_g , and which can be written as a function of the set of measurements that were carried out in our experiment. We now provide more details on this approach.

We perform a search to find a k -qubit operator $Q_i^{(k)}$, subject to two important constraints. First, we search for an operator which both maximizes the following inequality,

$$-\text{Tr}(Q_i^{(k)} \rho_i^k) \equiv \mathcal{S}_i^{(k)} \leq \mathcal{N}_g(\rho_i^k), \quad (5)$$

for a specific k -qubit state of interest ρ_i^k , and satisfies the inequality for all possible k -qubit states. We call $Q_i^{(k)}$ a quantitative entanglement witness (QEW) because it provides a lower bound on \mathcal{N}_g . It is straightforward to

constrain the search in this way and also to verify that any given $Q_i^{(k)}$ is a QEW. When searching for the optimal witness, we use a theoretical model for the time-evolved k -qubit state for ρ_i^k . Second, we include the additional constraint that the $Q_i^{(k)}$ can be written as a linear function of the k -qubit measurement operators (projectors) that were done, involving qubit group i . Specifically, we restrict $Q_i^{(k)}$ to the form

$$Q_i^{(k)} = \sum_{\vec{s}, \vec{\alpha}} c_{i; \vec{s}, \vec{\alpha}}^{(k)} P_{\vec{s}, \vec{\alpha}}^{(k)}, \quad (6)$$

where the projectors $P_{\vec{s}, \vec{\alpha}}^{(k)}$ correspond to the marginal distributions of the twenty-seven 20-qubit projective measurement settings carried out in the lab (Sec. III), and $c_{i; \vec{s}, \vec{\alpha}}^{(k)}$ denote some coefficients. Here, \vec{s} and $\vec{\alpha}$ label, respectively, the qubit outcome and the local basis of the measurements; see also Appendix B 2.

The search for the optimal witnesses operator $Q_i^{(k)}$ is carried out using a semidefinite program. The run time of the SDP is polynomial in the dimension of the Hilbert space [35,36], but the dimension of our Hilbert space naturally

increases exponentially with the number of qubits k . This makes the optimization demanding already for medium numbers of qubits: Our available computational resources are not sufficient to determine optimal witnesses for states of more than 5 qubits.

The $Q_i^{(k)}$ which satisfies Eqs. (5) and (6), and maximizes the left-hand side of Eq. (5), determines an optimal witness tailored to the target state (from a theoretical model) and the available measurements. Once this optimal $Q_i^{(k)}$ is found we can calculate its expectation value from the outcomes of the measurement done on the state in the laboratory. A witness expectation value ($S_i^{(k)}$) larger than zero then detects k -qubit GME ($\mathcal{N}_g > 0$), for the i th group of k qubits.

The experimental results for $k = 3$ presented in Fig. 4(a) show that all neighboring qubit triplets soon develop GME during the dynamics, to within many standard deviations, reaching a maximum at $t = 2$ ms. Furthermore, for the times 2, 2.5, and 3 ms, Figs. 4(b) and 4(c) show that GME is detected in the majority of all neighboring groups of 4 and 5 qubits, to within at least 1 standard deviation of experimental uncertainty.

Figure 4 compares the witness results obtained from the data with those derived from two theoretical models. The

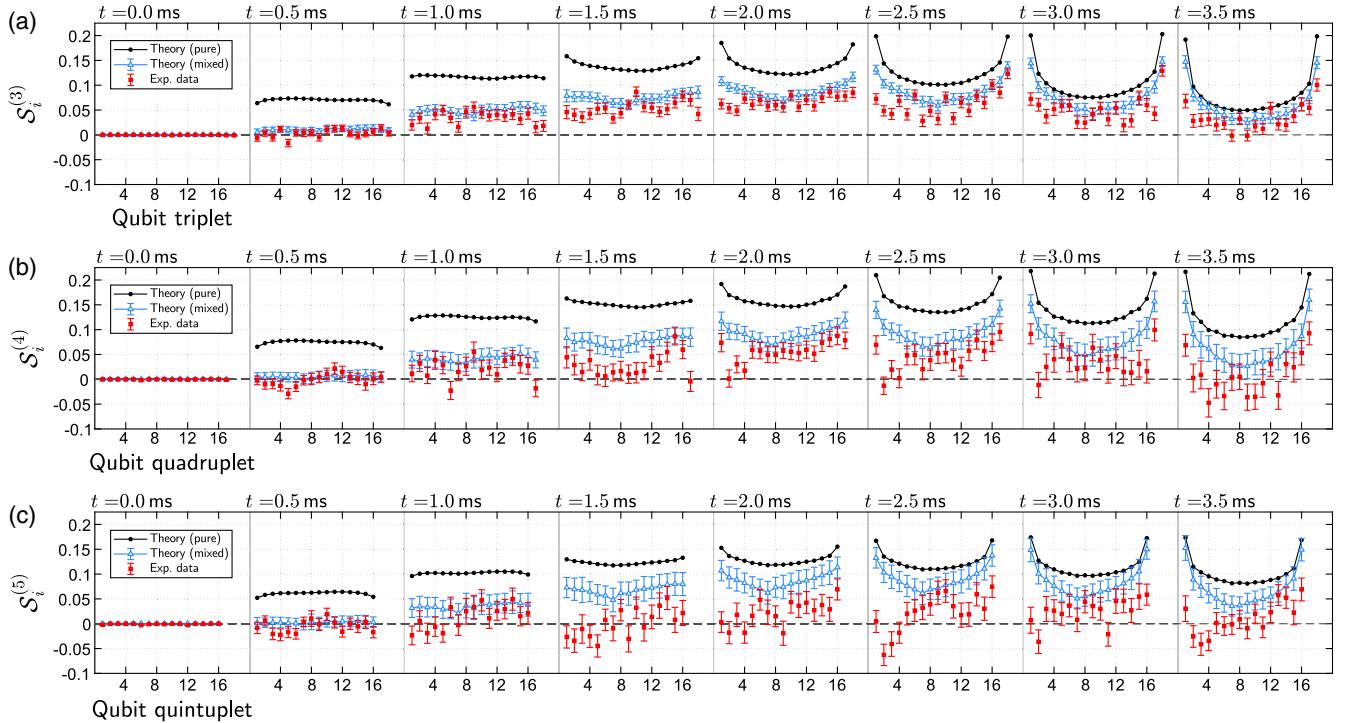


FIG. 4. Entanglement witnesses derived by a numerical search. The entanglement witnesses $S_i^{(3)}$, $S_i^{(4)}$, and $S_i^{(5)}$ (see Sec. V) are shown in (a)–(c), respectively. The horizontally arranged subpanels show the results at different system evolution time steps with intervals of 0.5 ms during the time evolution of the 20-qubit chain, starting with 0 ms (the initial state). Within each panel, each dot represents results for a given triplet [(a) $k = 3$], quadruplet [(b) $k = 4$], or quintuplet [(c) $k = 5$] of neighboring qubits. GME is detected if the witness is positive. The different theory plots are for the pure model (black circles without error bars) and mixed model (blue triangles with error bars), as described in the text. The red squares with error bars show experimental data. All error bars show 1 standard deviation of the mean and originate from a finite number of numerically simulated measurements per measurement basis (1000).

first “pure” model employs the perfect pure 20-qubit time-evolved state $[|\psi(t)\rangle]$ and uses exact knowledge of k -qubit density matrices to optimize and apply the witnesses. The witness expectation value for the pure model yields $\mathcal{S}_i^{(k)} = \mathcal{N}_g(\rho_i^k)$. Although the pure model succeeds in qualitatively describing the multipartite entanglement dynamics, the witness expectation values from the data are generally offset to lower values. A more sophisticated “mixed” model is able to explain part of this offset, which includes known imperfections in preparing the initial Néel-ordered state. Specifically, out of 1000 attempts to generate the Néel state, we observe the correct output state 829 times. In the remaining 171 cases, 146 correspond to single qubit flip errors and the rest to errors with two or more qubit flips. We attribute these errors to uncontrolled fluctuations in laser intensity and frequency, and model them as leading to the preparation of a statistical mixture of those different logical initial states, with corresponding weights. The witnesses used for the data were obtained by a search involving the mixed-model reduced states for the targets. We attribute the remaining small differences between data and theory, in Fig. 4, to additional mixing processes that occur in the laser-induced qubit-qubit interactions.

The mixed theory predictions in Fig. 4 include error bars due to the use of a finite number of numerically simulated measurements per measurement basis (1000). Error bars indicate 1 standard deviation of the mean, which is estimated as in Sec. IV, and show that the fluctuations in the data are largely consistent with those expected from such statistical noise. We conclude that, in order to witness 4- and 5-partite GME with greater statistical significance in future work, we could benefit from taking more measurements. However, it will be challenging to ensure that the experimental configuration remains stable over the longer time required to take such additional measurements.

The sizes of the error bars on both data and mixed theory points, in Fig. 4, increase with increasing k . This can be understood as follows: there are more measurement outcomes available in the data for the $k = 3$ witness calculations than for larger k . Amongst the twenty-seven 20-qubit measurement bases, 3-qubit measurements are repeated (duplicated) more often in the measurement pattern than 4-qubit or 5-qubit measurements, leading to better statistics.

VI. DISCUSSION AND CONCLUSION

We have experimentally generated and detected the presence of entanglement in a register of 20 qubits. In particular, we detected the dynamical evolution of genuine multipartite entanglement in the system following a quench, and developed new characterization techniques to do so. While we cannot say that 20-qubit GME was generated, we can say that every qubit simultaneously

became genuine multipartite entangled with a least two of its neighbors and, in most cases, three and four of its neighbors.

Our experimental apparatus represents the largest joint system of individually controllable subsystems to date where the presence of entanglement has been demonstrated. Each qubit can be individually controlled and qubit-qubit interactions can be turned on and off as desired (and tuned to have various forms). As such, our system has the capability to perform universal quantum simulation and quantum computation.

Confirming GME beyond groups of 5 qubits, even for the ideal states, is currently beyond our available classical computational resources and algorithms. A possible approach to overcome that problem is to exploit symmetries in the system and initial state, to reduce the size of the search space for witnesses. Another is to tune the experimental system Hamiltonian into regimes where more symmetries are apparent or approximated, e.g., infinite or nearest-neighbor-only qubit-qubit interaction ranges.

Finally, witnesses based on average Bell-state fidelities are straightforward to use and measure in the lab. As with all such witnesses, they detect entanglement without the need to carry out state tomography and can be evaluated with only a few measurements. This can be important for the detection of weakly entangled states, where estimates based on state tomography are known to overestimate entanglement [37]. Our witnesses based on brute-force numerical searching have the advantage of placing the least constraints on the form of the state in the lab and the measurements that should be taken: this can be important in the case of unknown local rotations of qubits during the dynamics. As such, our witnesses should find application beyond the present trapped-ion setting.

ACKNOWLEDGMENTS

The work in Innsbruck was supported by the Austrian Science Fund (FWF) under Grant No. P25354-N20, by the European Commission via the integrated project SIQS, and by the Army Research Laboratory under Cooperative Agreement No. W911NF-15-2-0060. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. B. L. acknowledges funding from the Austrian Science Fund (FWF) through the START Project No. Y 849-N20. N. F. and M. Huber acknowledge funding from the Austrian Science Fund (FWF) through the START Project No. Y879-N27 and the joint Czech-Austrian project MultiQUEST (I 3053-N27). The work in Ulm was supported by an Alexander von Humboldt Professorship, the ERC Synergy grant BioQ, the EU projects QUCHIP, and the U.S. Army Research Office Grant No. W91-1NF-14-1-0133, as well as by computational resources by the state of Baden-Württemberg through

bwHPC and the German Research Foundation (DFG) through Grant No. INST 40/467-1 FUGG.

APPENDIX A: GME WITNESSES BASED ON BIPARTITE CORRELATORS

In this section, we introduce a method for the detection of genuine multipartite entanglement in N -qubit systems. This method is based on 2-qubit observables and does not require full state tomography. In particular, our detection criteria can be phrased as biseparability thresholds for average Bell fidelities, i.e., expectation values of linear combinations of pairs of Pauli operators. At the heart of this method lies the *anticommutativity theorem* (ACT) from Refs. [38,39], which we use to provide bounds on the average Bell fidelities. Although our approach is not able to detect all types of GME in multipartite systems, its advantage lies in providing linear entanglement witnesses that can be practically evaluated with only a few measurements. More specifically, our approach does not require obtaining a good estimate of the N -partite correlation tensor [40,41] with 3^N components, but instead only needs at most $6N$ measurements of strings of N local Pauli operators to test for N -partite GME. As we discuss, the linearity of the witness also makes it amenable to a simple treatment of the potentially correlated statistical errors arising from deriving expectation values of bipartite observables from simultaneous measurements of N qubits.

Following the brief description of these results in Sec. IV of the main text, we now present more detailed derivations of the quantities and bounds that we consider. In Appendix A 1, we briefly define and motivate the basic quantities of interest, before we construct our GME witnesses in Appendix A 2.

1. Framework

In this section, we explain the basic quantities and notions of interest, i.e., the anticommutativity theorem of Ref. [39] and the average Bell fidelities to establish a basis for the more detailed discussion of GME that is to follow in Appendix A 2.

a. Anticommutativity theorem

Let us consider a set $\{A_n\}_{n=1,2,\dots,k}$ of self-adjoint, normalized, anticommuting operators on a Hilbert space \mathcal{H} with $\dim(\mathcal{H}) = d$, i.e.,

$$\text{Tr}\{A_m, A_n\}_+ = \text{Tr}(A_m A_n + A_n A_m) = 2d\delta_{mn}, \quad (\text{A1})$$

for all $m, n = 1, \dots, k$. The ACT [38,39] then states that for all states $\rho \in L(\mathcal{H})$,

$$\sum_{n=1}^k \langle A_n \rangle_\rho^2 \leq \max_n \langle A_n^2 \rangle_\rho. \quad (\text{A2})$$

A simple example for the applicability of this theorem is the set of single-qubit Pauli operators $\{X, Y, Z\}$. Since all of these operators anticommute and square to the identity, the ACT then simply requires that

$$\langle X \rangle_\rho^2 + \langle Y \rangle_\rho^2 + \langle Z \rangle_\rho^2 \leq \langle \mathbb{1} \rangle_\rho = 1. \quad (\text{A3})$$

In other words, for single-qubit Pauli operators, the ACT is equivalent to demanding that Bloch vectors are (sub) normalized, i.e., positivity of the density operator ρ . A less trivial example of the ACT arises for 2 qubits. Consider the set of operators

$$\{X_1 X_2, Y_1 Y_2, Z_1 Z_2, X_2 X_3, Y_2 Y_3, Z_2 Z_3\}, \quad (\text{A4})$$

where the shorthand notation for N -qubit operators is

$$O_i \equiv \mathbb{1}_1 \otimes \dots \otimes \mathbb{1}_{i-1} \otimes O_i \otimes \mathbb{1}_{i+1} \otimes \dots \otimes \mathbb{1}_N, \quad (\text{A5})$$

and $O \in \{X, Y, Z\}$. We can sort the six operators in the set displayed in Eq. (A4) into three pairs of anticommuting operators; e.g.,

$$\{X_1 X_2, Y_2 Y_3\}_+ = 0, \quad (\text{A6a})$$

$$\{Y_1 Y_2, Z_2 Z_3\}_+ = 0, \quad (\text{A6b})$$

$$\{Z_1 Z_2, X_2 X_3\}_+ = 0. \quad (\text{A6c})$$

Since the spectra of all six operators are $\{\pm 1\}$ (with twofold degeneracy), we further have $\langle O_i O_{i+1} \rangle^2 \leq 1$, and the ACT theorem hence tells us that)

$$\langle X_1 X_2 \rangle_\rho^2 + \langle Y_2 Y_3 \rangle_\rho^2 \leq 1, \quad (\text{A7a})$$

$$\langle Y_1 Y_2 \rangle_\rho^2 + \langle Z_2 Z_3 \rangle_\rho^2 \leq 1, \quad (\text{A7b})$$

$$\langle Z_1 Z_2 \rangle_\rho^2 + \langle X_2 X_3 \rangle_\rho^2 \leq 1. \quad (\text{A7c})$$

To see where these bounds can be of use, let us next examine fidelities with 2-qubit Bell states.

b. Average Bell-state fidelities

We now want to consider ways of quantifying how close a given 2-qubit state is to a maximally entangled Bell state. To this end, note that the density operators for the four Bell states can be written in a generalized Bloch decomposition as

$$\rho_{\psi^-} = |\psi^-\rangle\langle\psi^-| = \frac{1}{4}(\mathbb{1}_{1,2} - X_1 X_2 - Y_1 Y_2 - Z_1 Z_2), \quad (\text{A8a})$$

$$\rho_{\psi^+} = |\psi^+\rangle\langle\psi^+| = \frac{1}{4}(\mathbb{1}_{1,2} + X_1 X_2 + Y_1 Y_2 - Z_1 Z_2), \quad (\text{A8b})$$

$$\rho_{\phi^-} = |\phi^-\rangle\langle\phi^-| = \frac{1}{4}(\mathbb{1}_{1,2} - X_1X_2 + Y_1Y_2 + Z_1Z_2), \quad (\text{A8c})$$

$$\rho_{\phi^+} = |\phi^+\rangle\langle\phi^+| = \frac{1}{4}(\mathbb{1}_{1,2} + X_1X_2 - Y_1Y_2 + Z_1Z_2). \quad (\text{A8d})$$

For any 2-qubit density operator ρ , we can then compute the fidelity with any of the Bell states. For this purpose we use the Uhlmann fidelity \mathcal{F} , given by

$$\mathcal{F}(\rho, \sigma) = \left(\text{Tr} \sqrt{\sqrt{\sigma} \rho \sqrt{\sigma}} \right)^2, \quad (\text{A9})$$

which reduces to

$$\mathcal{F}(\rho, |\psi\rangle\langle\psi|) = \langle\psi|\rho|\psi\rangle = \text{Tr}(\rho|\psi\rangle\langle\psi|) \quad (\text{A10})$$

if one of the arguments is a pure state. For example, for the Bell state $|\psi^-\rangle$ one can use Eq. (A8a) and the fact that all Pauli operators are traceless to find

$$\mathcal{F}(\rho, \rho_{\psi^-}) = \frac{1}{4}(1 - \langle X_1X_2 \rangle_\rho - \langle Y_1Y_2 \rangle_\rho - \langle Z_1Z_2 \rangle_\rho). \quad (\text{A11})$$

Since the only difference to the fidelities with any of the other Bell states are the relative signs between the different expectation values, we can immediately note that the fidelity of ρ with any of the four Bell states is bounded according to

$$\mathcal{F}(\rho, \rho_{\text{Bell}}) \leq \frac{1}{4}(1 + |\langle X_1X_2 \rangle| + |\langle Y_1Y_2 \rangle| + |\langle Z_1Z_2 \rangle|), \quad (\text{A12})$$

where we have dropped the subscript for the state ρ on the expectation values for brevity.

c. Nearest-neighbor average Bell fidelity

When the system consists of more than 2 qubits, we can evaluate the fidelity with 2-qubit Bell states for any two of the constituent qubits. For simplicity, let us first consider the nearest neighbors for now and examine the case of three qubits. The average fidelity with arbitrary nearest-neighbor Bell states is then

$$\frac{1}{2}[\mathcal{F}(\rho, \rho_{\text{Bell},12}) + \mathcal{F}(\rho, \rho_{\text{Bell},23})] \leq \bar{\mathcal{F}}_{\text{NN Bell}}, \quad (\text{A13})$$

where we define the quantity $\bar{\mathcal{F}}_{\text{NN Bell}}$ as the upper bound

$$\bar{\mathcal{F}}_{\text{NN Bell}} := \frac{1}{8}(2 + |\langle X_1X_2 \rangle| + |\langle Y_1Y_2 \rangle| + |\langle Z_1Z_2 \rangle| + |\langle X_2X_3 \rangle| + |\langle Y_2Y_3 \rangle| + |\langle Z_2Z_3 \rangle|), \quad (\text{A14})$$

but we refer to $\bar{\mathcal{F}}_{\text{NN Bell}}$ as the average nearest-neighbor Bell fidelity from now on for simplicity. Next, we make use of the relation between the 1-norm $\|\vec{a}\|_1 = \sum_{i=1}^n |a_i|$ and the 2-norm $\|\vec{a}\|_2 = (\sum_{i=1}^n |a_i|^2)^{1/2}$ in an n -dimensional vector space, i.e., the fact that

$$\|\vec{a}\|_1 = \sum_{i=1}^n |a_i| \times 1 = |(\vec{a}, \vec{1})| \leq \|\vec{a}\|_2 \left(\sum_{i=1}^n 1^2 \right)^{1/2} = \sqrt{n} \|\vec{a}\|_2, \quad (\text{A15})$$

where we have taken $\vec{1} = (1, 1, \dots, 1)^T$ to be a vector whose components (with respect to whichever basis is chosen for \vec{a}) are all equal to 1, and we have used the Cauchy-Schwarz inequality $|(\vec{a}, \vec{b})| \leq \|\vec{a}\|_2 \|\vec{b}\|_2$. Combining this with the ACT theorem from Eq. (A2), we find, e.g.,

$$|\langle X_1X_2 \rangle| + |\langle Y_2Y_3 \rangle| \leq \sqrt{2}(\langle X_1X_2 \rangle^2 + \langle Y_2Y_3 \rangle^2) \leq \sqrt{2}. \quad (\text{A16})$$

Applying the same procedure to the other pairs of expectation values of anticommuting operators in Eq. (A14), we arrive at the bound

$$\bar{\mathcal{F}}_{\text{NN Bell}} \leq \frac{1}{8}(2 + 3\sqrt{2}). \quad (\text{A17})$$

The average nearest-neighbor Bell state fidelity can of course be generalized to N qubits; i.e., the upper bound on the average nearest-neighbor Bell fidelity is

$$\frac{1}{N-1} \sum_{i=1}^{N-1} \mathcal{F}(\rho, \rho_{\text{Bell},i(i+1)}) \leq \bar{\mathcal{F}}_{\text{NN Bell}}^{(N)}, \quad (\text{A18})$$

where we have defined

$$\begin{aligned} \bar{\mathcal{F}}_{\text{NN Bell}}^{(N)} &:= \frac{1}{4(N-1)} \left((N-1) + \sum_{i=1}^{N-1} \sum_{O=X,Y,Z} |\langle O_i O_{i+1} \rangle| \right) \\ &= \frac{1}{4(N-1)} [(N-1) + |\langle X_1X_2 \rangle| + |\langle Y_1Y_2 \rangle| \\ &\quad + |\langle Z_1Z_2 \rangle| + |\langle X_2X_3 \rangle| + |\langle Y_2Y_3 \rangle| \\ &\quad + |\langle Z_2Z_3 \rangle| + \dots + |\langle X_{N-1}X_N \rangle| \\ &\quad + |\langle Y_{N-1}Y_N \rangle| + |\langle Z_{N-1}Z_N \rangle|]. \end{aligned} \quad (\text{A19})$$

The expression on the right-hand side contains $N-1$ triples of expectation values. If N is odd, then $N-1$ is even, and each expectation value of an operator $O_i O_{i+1}$ can be paired with another expectation value of an operator $O'_{i+1} O'_{i+2}$ that anticommutes with it; i.e., $O, O' \in \{X, Y, Z\}$ and $\{O_i O_{i+1}, O'_{i+1} O'_{i+2}\}_+ = 0$. The bound of Eq. (A16) can hence be used $\lceil [3(N-1)]/2 \rceil$ times, and we arrive at

$$\begin{aligned}\bar{\mathcal{F}}_{\text{NN Bell}}^{(N\text{odd})} &\leq \frac{1}{4(N-1)} \left((N-1) + \frac{3(N-1)}{2} \sqrt{2} \right) \\ &= \frac{1}{8} (2 + 3\sqrt{2}).\end{aligned}\quad (\text{A20})$$

However, when N is even, one triple of expectation values (without loss of generality for $i = N-1$) remains unpaired and can only be bounded by

$$|\langle X_{N-1} X_N \rangle| + |\langle Y_{N-1} Y_N \rangle| + |\langle Z_{N-1} Z_N \rangle| \leq 3. \quad (\text{A21})$$

Thus we arrive at the following upper bound on the nearest-neighbor average Bell fidelity for arbitrary N -qubit states; i.e.,

$$\bar{\mathcal{F}}_{\text{NN Bell}}^{(N)} \leq \begin{cases} \frac{1}{8} (2 + 3\sqrt{2}) & (N \text{ odd}) \\ \frac{1}{4(N-1)} \left[(N-1) + \frac{3(N-2)\sqrt{2}}{2} + 3 \right] & (N \text{ even}). \end{cases} \quad (\text{A22})$$

d. Symmetric average Bell fidelity

Instead of restricting the analysis to nearest neighbors as in Eq. (A14), one can of course also average over all pairings of 2 qubits, obtaining a quantity that is symmetric with respect to the exchange of any 2 qubits. Noting that there are $b_N = \binom{N}{2} = \frac{1}{2} \{ (N!) / [(N-2)!] \}$ different such pairings, we have the upper bound

$$\frac{1}{b_N} \sum_{\substack{i,j=1 \\ i < j}}^N \mathcal{F}(\rho, \rho_{\text{Bell},ij}) \leq \bar{\mathcal{F}}_{\text{Bell}}^{(N)}, \quad (\text{A23})$$

with the definition

$$\bar{\mathcal{F}}_{\text{Bell}}^{(N)} := \frac{1}{4b_N} \left(b_N + \sum_{i,j=1}^N \sum_{O=X,Y,Z} |\langle O_i O_j \rangle| \right). \quad (\text{A24})$$

For, instance, for 3 qubits we have $b_3 = 3$ and the symmetric average Bell fidelity reads

$$\begin{aligned}\bar{\mathcal{F}}_{\text{Bell}}^{(3)} &= \frac{1}{12} (3 + |\langle X_1 X_2 \rangle| + |\langle X_2 X_3 \rangle| + |\langle X_1 X_3 \rangle| \\ &\quad + |\langle Y_1 Y_3 \rangle| + |\langle Y_1 Y_2 \rangle| + |\langle Y_2 Y_3 \rangle| \\ &\quad + |\langle Z_2 Z_3 \rangle| + |\langle Z_1 Z_3 \rangle| + |\langle Z_1 Z_2 \rangle|).\end{aligned}\quad (\text{A25})$$

Here, we have arranged the expectation values such that it becomes immediately obvious that the triples of operators corresponding to expectation values listed directly below or above each other mutually anticommute. We can then apply

the bound of Eq. (A15) and the ACT of Eq. (A2), e.g., as illustrated for the terms

$$\begin{aligned}|\langle X_1 X_2 \rangle| + |\langle Y_1 Y_3 \rangle| + |\langle Z_2 Z_3 \rangle| \\ \leq \sqrt{3} (|\langle X_1 X_2 \rangle|^2 + |\langle Y_1 Y_3 \rangle|^2 + |\langle Z_2 Z_3 \rangle|^2) \leq \sqrt{3}.\end{aligned}\quad (\text{A26})$$

We thus arrive at the bound

$$\bar{\mathcal{F}}_{\text{Bell}}^{(3)} \leq \frac{1}{12} (3 + 3\sqrt{3}) = \frac{1}{4} (1 + \sqrt{3}). \quad (\text{A27})$$

In fact, the same bound applies for arbitrary numbers of qubits, since all $3b_N$ expectation values can be collected in groups of 3 mutually anticommuting operators. To see this, we use an inductive proof. Assume that we have found b_N groups of three anticommuting operators for $N \geq 3$ qubits and we wish to add another qubit. This means that we have to additionally consider the operators

$$\begin{array}{lll} X_1 X_{N+1}, & X_2 X_{N+1}, & X_3 X_{N+1}, \dots, X_N X_{N+1}, \\ Y_1 Y_{N+1}, & Y_1 Y_{N+1}, & Y_2 Y_{N+1}, \dots, Y_{N-1} Y_{N+1}, \\ Z_{N-1} Z_{N+1}, & Z_N Z_{N+1}, & Z_1 Z_{N+1}, \dots, Z_{N-2} Z_{N+1}. \end{array}$$

All columns contain three mutually anticommuting operators for $N \geq 3$. If the original $3b_N$ operators can be arranged in mutually anticommuting triples, then also the new set of $3b_{N+1}$ operators can be grouped in this way, which concludes the inductive step. We have already demonstrated that this statement is true for $N = 3$ and have hence shown that for any $N \geq 3$ we have the bound

$$\bar{\mathcal{F}}_{\text{Bell}}^{(N \geq 3)} \leq \bar{\mathcal{F}}_{\text{Bell}}^{(N) \text{ max}} := \frac{1}{4} (1 + \sqrt{3}). \quad (\text{A28})$$

Having established these general bounds that apply for arbitrary quantum states, we next examine how these bounds can be improved upon when the states in question are biseparable. This will allow us to formulate criteria for the detection of genuine multipartite entanglement.

2. GME witnesses

In this section, we establish upper bounds for the nearest-neighbor and symmetric average Bell fidelities for biseparable states. These new upper bounds are below the respective bounds of Eqs. (A22) and (A28) and hence leave room for GME states in between. That is, any states for which the combinations of expectation values discussed above provide values beyond these biseparability bounds are GME. As we shall see, the biseparability bounds for nearest-neighbor Bell fidelities are not directly useful for detecting GME, but these bounds serve as a simple example for discussing the method of construction, which will be

helpful for identifying GME witnesses based on symmetric average Bell fidelities.

a. Outline of the technique

In the following we consider bipartitions $A|B$ of the set $\kappa = \{1, 2, \dots, N\}$ of all N qubits; that is, we split κ into two sets,

$$A = \{a_1, a_2, \dots, a_k | a_i \in \kappa, a_i \neq a_j \ \forall i \neq j\}, \quad (\text{A29a})$$

$$B = \{b_1, b_2, \dots, b_{N-k} | b_i \in \kappa, b_i \neq b_j \ \forall i \neq j\}, \quad (\text{A29b})$$

such that $A \cup B = \kappa$ and $A \cap B = \emptyset$. For N qubits, one has $2^{N-1} - 1$ different bipartitions.

Before we continue, let us briefly recall the definitions of biseparability and genuine multipartite entanglement. In general, a pure, N -partite state $|\psi\rangle \in \mathcal{H}_{1,2,\dots,N} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \dots \mathcal{H}_N$ is called k -separable if it can be written as a tensor product with respect to some partition of $\mathcal{H}_{1,2,\dots,N}$ into $k \leq N$ subsystems. As a special case of this definition, a pure state is called biseparable if it can be written as a tensor product with respect to some bipartition, i.e., if there exists a bipartition $A|B$ such that $|\psi\rangle = |\phi\rangle_A |\chi\rangle_B$. Conversely, a pure state $|\psi\rangle \in \mathcal{H}_{1,2,\dots,N}$ that is not biseparable is called genuinely N -partite entangled. A mixed state with density operator ρ is considered to be genuinely multipartite entangled if it cannot be written as a convex combination of biseparable states, that is, if it cannot be written as

$$\rho_{\text{bisep}} = \sum_i p_i |\psi_{\text{bisep}}^{(i)}\rangle \langle \psi_{\text{bisep}}^{(i)}|, \quad (\text{A30})$$

where $\sum_i p_i = 1$ with $0 \leq p_i \leq 1$ and $|\psi_{\text{bisep}}^{(i)}\rangle$ are biseparable pure states. Note that the $|\psi_{\text{bisep}}^{(i)}\rangle$ for different i can be separable with respect to different bipartitions.

Now, consider a bipartition $A|B$ and an operator $O_i O_j$ such that $i \in A$ and $j \in B$. If the system is in a pure state $|\psi\rangle$ that is separable with respect to this bipartition, i.e., if $|\psi\rangle_{AB} = |\phi\rangle_A |\chi\rangle_B$, then we have

$$\langle O_i O_j \rangle_\psi = \langle O_i \rangle_\phi \langle O_j \rangle_\chi. \quad (\text{A31})$$

When we have a triple of operators $X_i X_j$, $Y_i Y_j$, and $Z_i Z_j$ for such a separable state across $A|B$, we have

$$\begin{aligned} & |\langle X_i X_j \rangle| + |\langle Y_i Y_j \rangle| + |\langle Z_i Z_j \rangle| \\ &= |\langle X_i \rangle| |\langle X_j \rangle| + |\langle Y_i \rangle| |\langle Y_j \rangle| + |\langle Z_i \rangle| |\langle Z_j \rangle| \\ &\leq \prod_{n=i,j} \sqrt{|\langle X_n \rangle|^2 + |\langle Y_n \rangle|^2 + |\langle Z_n \rangle|^2} \leq 1, \end{aligned} \quad (\text{A32})$$

where we have used the Cauchy-Schwarz inequality in the second-to-last step and the subnormalization of the Bloch vector in the last step. The inequality Eq. (A32) can be used to bound the Bell fidelities for pure biseparable states for different bipartitions.

As an example, consider again the nearest-neighbor average Bell fidelity for three qubits from Eq. (A14). For the bipartition $1|23$, we can apply (A32) to the first three expectation values in Eq. (A14), while the remaining three can each be bounded by 1. A similar argument can be made for the bipartition $12|3$ by exchanging the roles of the two triples of expectation values, such that

$$\bar{\mathcal{F}}_{\text{NN Bell}}^{1|23,12|3} \leq \frac{1}{8}(2 + 1 + 3) = \frac{3}{4}, \quad (\text{A33})$$

where the superscripts indicate that the inequality is satisfied for states that are biseparable with respect to (at least one of) the listed bipartitions. When we examine the bipartition $2|13$, the situation is slightly different, since Eq. (A32) can be used for all expectation values, and we have

$$\bar{\mathcal{F}}_{\text{NN Bell}}^{2|13} \leq \frac{1}{8}(2 + 1 + 1) = \frac{1}{4}. \quad (\text{A34})$$

Any pure 3-qubit state that is separable with respect to one or more of these bipartitions (any pure, biseparable state of three qubits) must hence satisfy $\bar{\mathcal{F}}_{\text{NN Bell}} \leq \frac{3}{4}$. Moreover, since any mixed state is considered to be biseparable when it can be written as a convex combination of biseparable pure states (not necessarily with respect to the same bipartition), all mixed, biseparable states must also respect this bound. Conversely, the first 3 qubits of any state ρ for which

$$\begin{aligned} & \frac{1}{8}(2 + |\langle X_1 X_2 \rangle| + |\langle Y_1 Y_2 \rangle| + |\langle Z_1 Z_2 \rangle| \\ &+ |\langle X_2 X_3 \rangle| + |\langle Y_2 Y_3 \rangle| + |\langle Z_2 Z_3 \rangle|) > \frac{3}{4} \end{aligned} \quad (\text{A35})$$

are genuinely 3-partite entangled.

b. Nearest-neighbor average Bell fidelity as GME witness

In principle, the nearest-neighbor Bell fidelity could hence provide a detection criterion for GME that can be generalized to N qubits. However, at this point a remark on the detection power of this quantity is in order, since even some paradigmatic cases of genuinely tripartite entangled states for 3 qubits cannot be detected with this bound. That is, for the 3-qubit GHZ and W states $|\psi_{\text{GHZ}}^{(3)}\rangle$ and $|\psi_W^{(3)}\rangle$ (or the local unitarily equivalent 2-excitation Dicke state $|\psi_{D,2}^{(3)}\rangle$), given by

$$|\psi_{\text{GHZ}}^{(3)}\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle), \quad (\text{A36a})$$

$$|\psi_W^{(3)}\rangle = \frac{1}{\sqrt{3}}(|100\rangle + |010\rangle + |001\rangle), \quad (\text{A36b})$$

$$|\psi_{D,2}^{(3)}\rangle = \frac{1}{\sqrt{3}}(|110\rangle + |101\rangle + |011\rangle), \quad (\text{A36c})$$

one finds nearest-neighbor average Bell fidelities of

$$\bar{\mathcal{F}}_{\text{NN Bell}}^{(3)}(|\psi_{\text{GHZ}}^{(3)}\rangle) = \frac{1}{2}, \quad (\text{A37a})$$

$$\bar{\mathcal{F}}_{\text{NN Bell}}^{(3)}(|\psi_W^{(3)}\rangle) = \frac{2}{3}, \quad (\text{A37b})$$

$$\bar{\mathcal{F}}_{\text{NN Bell}}^{(3)}(|\psi_{D,2}^{(3)}\rangle) = \frac{2}{3}, \quad (\text{A37c})$$

whereas the corresponding bound for detecting GME is $\frac{3}{4}$.

One can hence try to improve the method or find an alternative. One way to improve the bound is by way of taking into account the purity of the biseparable states. That is, if we consider again the worst-case bipartition $1|23$ for 3 qubits under the assumption that the state is separable with respect to this cut, i.e., that $|\psi\rangle_{123} = |\phi\rangle_1 |\chi\rangle_{23}$, we have

$$\begin{aligned} \bar{\mathcal{F}}_{\text{NN Bell}}^{1|23} &\leq \frac{1}{8} \left(2 + \sqrt{|\langle X_2 \rangle|^2 + |\langle Y_2 \rangle|^2 + |\langle Z_2 \rangle|^2} \right. \\ &\quad \left. + |\langle X_2 X_3 \rangle| + |\langle Y_2 Y_3 \rangle| + |\langle Z_2 Z_3 \rangle| \right) \\ &= \frac{1}{8} \left(2 + |\vec{b}| + \sum_{n=1}^3 |t_{nn}| \right), \end{aligned} \quad (\text{A38})$$

where we have used the Bloch vector \vec{b} of the second qubit and the correlation tensor $t = (t_{mn})$ of qubits 2 and 3. In other words, the state $|\chi\rangle_{23}$ can be written in a generalized Bloch decomposition as

$$\rho_\chi = |\chi\rangle\langle\chi| = \frac{1}{4} \left(\mathbb{1} + \vec{b} \cdot \vec{\sigma} \otimes \mathbb{1} + \mathbb{1} \otimes \vec{c} \cdot \vec{\sigma} + \sum_{i,j=1}^3 t_{ij} \sigma_i \otimes \sigma_j \right), \quad (\text{A39})$$

where $\vec{\sigma} = (\sigma_n)$ is the vector of Pauli operators ($\sigma_1 = X, \sigma_2 = Y, \sigma_3 = Z$). Now, since $|\chi\rangle_{23}$ is a pure state, we have $\text{Tr}(\rho_\chi^2) = 1$, which translates to

$$\frac{1}{4} \left(1 + |\vec{b}|^2 + |\vec{c}|^2 + \sum_{m,n=1}^3 |t_{mn}|^2 \right) = 1, \quad (\text{A40})$$

and we can hence derive the bound

$$|\vec{b}|^2 + \sum_{n=1}^3 |t_{nn}|^2 \leq 3. \quad (\text{A41})$$

Interpreting $|\vec{b}|$ and $|t_{nn}|$ ($n = 1, 2, 3$) as coordinates in \mathbb{R}^4 , we find that Eq. (A41) defines a four-dimensional sphere of radius $\sqrt{3}$. The sum of the coordinates is then maximal when all coordinates take the same value $\sqrt{3}/4$. Inserting into Eq. (A38), we then get the bound

$$\bar{\mathcal{F}}_{\text{NN Bell}}^{1|23} \leq \frac{1}{8} \left(2 + 4\sqrt{\frac{3}{4}} \right) = \frac{1}{4} (1 + \sqrt{3}) \approx 0.683013. \quad (\text{A42})$$

Since the bipartition $12|3$ is equivalent and for $2|13$ we have the lower value $\bar{\mathcal{F}}_{\text{NN Bell}}^{2|13} \leq \frac{1}{2}$, the ‘‘improved’’ biseparability bound for the nearest-neighbor average Bell fidelity for three qubits is $\frac{1}{4}(1 + \sqrt{3}) \approx 0.683013$. This value is still above the fidelity $\frac{2}{3}$ obtained for pure GME states of 3 qubits in Eq. (A44). In addition, we have also conducted a numerical search which did not reveal any pure 3-qubit states with nearest-neighbor average Bell fidelities beyond $\frac{2}{3}$. At the same time, one can find biseparable states that give values for $\bar{\mathcal{F}}_{\text{NN Bell}}^{(3)}$ that are very close to $\frac{2}{3}$. For instance, for the state $\rho_{\text{bisep}} = |0\rangle\langle 0|_1 \otimes \tilde{\rho}_{23}$, where $|0\rangle_1$ is an eigenstate of Z , and the (nearly pure) state $\tilde{\rho}$ has Bloch vectors $\vec{b} = \vec{c} = (0, 0, 0.447)^T$ and a diagonal correlation matrix $t = \text{diag}\{0.894, -0.894, 1\}$, we find $\bar{\mathcal{F}}_{\text{NN Bell}}^{(3)} = 0.654375$. We therefore conclude that, in their present form, GME witnesses based on the nearest-neighbor average Bell fidelity are practically irrelevant for 3 qubits and there is no reason to expect an improvement for more than 3 qubits.

We therefore now continue with an analysis of a different quantity, the symmetric average Bell fidelity.

c. Symmetric average Bell fidelity as GME witness

In this section, we discuss the usefulness of the symmetric average Bell fidelity as a witness for GME. To this end, we again need to identify the bipartitions providing the worst (largest) upper bound for $\bar{\mathcal{F}}_{\text{Bell}}^{(N)}$ under the assumption of separability with respect to the respective bipartition. Since the combination of expectation values that we consider now is symmetric under the exchange of any 2 qubits, this task is rather straightforward.

First, we consider the case of 3 qubits separately, where all three possible bipartitions (i.e., $1|23$, $2|13$, and $12|3$) are equivalent. If the system state is pure and separable with respect to any of these bipartitions, two of the triples of expectation values in Eq. (A25) are ‘‘cut’’ by the bipartition and can be bounded by 1, while the remaining triple consists of three commuting observables, whose

expectation values are jointly bounded by 3. For any labeling of the 3 qubits, we hence have

$$\bar{\mathcal{F}}_{\text{Bell}}^{1|23} \leq \frac{1}{12}(3 + 1 + 1 + 3) = \frac{2}{3}. \quad (\text{A43})$$

As we discussed in Appendix A 2 b, this bound has to be compared with values achievable with pure GME states. For the 3-qubit GHZ-, W-, and 2-excitation Dicke states, we find symmetric average Bell fidelities

$$\bar{\mathcal{F}}_{\text{Bell}}^{(3)}(|\psi_{\text{GHZ}}^{(3)}\rangle) = \frac{1}{2}, \quad (\text{A44a})$$

$$\bar{\mathcal{F}}_{\text{Bell}}^{(3)}(|\psi_W^{(3)}\rangle) = \frac{2}{3}, \quad (\text{A44b})$$

$$\bar{\mathcal{F}}_{\text{Bell}}^{(3)}(|\psi_{D,2}^{(3)}\rangle) = \frac{2}{3}, \quad (\text{A44c})$$

which happen to coincide with the corresponding nearest-neighbor average Bell fidelities of Eq. (A44). We must hence try to improve the bound using a similar trick as before in Appendix A 2 b. Again assuming a biseparable pure state for the bipartition 1|23, we can write

$$\bar{\mathcal{F}}_{\text{Bell}}^{1|23} \leq \frac{1}{12} \left(3 + |\vec{b}| + |\vec{c}| + \sum_{n=1}^3 |t_{nn}| \right), \quad (\text{A45})$$

and in analogy to Eq. (A41) we can bound each of the moduli $|\vec{b}|$, $|\vec{c}|$, and $|t_{nn}|$ (for $n = 1, 2, 3$) by $\sqrt{3/5}$, which gives the bound

$$\bar{\mathcal{F}}_{\text{Bell}}^{(3)\text{bisep}} \leq \frac{1}{12} \left(3 + 5\sqrt{\frac{3}{5}} \right) = \frac{1}{12}(3 + \sqrt{15}) \approx 0.572749. \quad (\text{A46})$$

Using numerical optimization, we can also provide a pure biseparable state that comes very close to this bound. That is, for the state $\rho_{\text{bisep}} = |0\rangle\langle 0|_1 \otimes \tilde{\rho}_{23}$, where $|0\rangle_1$ is an eigenstate of Z , and the pure state $\tilde{\rho}$ has Bloch vectors $\vec{b} = \vec{c} = [0, 0, (1/\sqrt{2})]^T$ and a diagonal correlation matrix $t = \text{diag}\{(1/\sqrt{2}), -(1/\sqrt{2}), 1\}$, we find $\bar{\mathcal{F}}_{\text{Bell}}^{(3)} = 0.569036$. As before, the pure state biseparability bound extends to mixed states via convexity. Thus, any 3-qubit state for which the combination of (moduli of) expectation values on the right-hand side of Eq. (A25) exceeds $\frac{1}{12}(3 + \sqrt{15})$ must be genuinely tripartite entangled.

Second, let us turn to the case of 4 qubits, where we are interested in bounding the quantity

$$\begin{aligned} \bar{\mathcal{F}}_{\text{Bell}}^{(4)} = & \frac{1}{24} (6 + |\langle X_1 X_2 \rangle| + |\langle X_1 X_3 \rangle| + |\langle X_1 X_4 \rangle| \\ & + |\langle X_2 X_3 \rangle| + |\langle X_2 X_4 \rangle| + |\langle X_3 X_4 \rangle| \\ & + |\langle Y_1 Y_2 \rangle| + |\langle Y_1 Y_3 \rangle| + |\langle Y_1 Y_4 \rangle| \\ & + |\langle Y_2 Y_3 \rangle| + |\langle Y_2 Y_4 \rangle| + |\langle Y_3 Y_4 \rangle| \\ & + |\langle Z_1 Z_2 \rangle| + |\langle Z_1 Z_3 \rangle| + |\langle Z_1 Z_4 \rangle| \\ & + |\langle Z_2 Z_3 \rangle| + |\langle Z_2 Z_4 \rangle| + |\langle Z_3 Z_4 \rangle|). \end{aligned} \quad (\text{A47})$$

For any pure state that is separable with respect to a bipartition into 1 versus 3 qubits, we find three triples of expectation values that are cut (each bounded by 1), while three triples pertaining to the same subsystem can be combined into mutually anticommuting triples, each bounded by $\sqrt{3}$, obtaining

$$\bar{\mathcal{F}}_{\text{Bell}}^{1|234} \leq \frac{1}{24} (6 + 3 + 3\sqrt{3}) = \frac{1}{8} (3 + \sqrt{3}) \approx 0.591506. \quad (\text{A48})$$

Instead, we can also use the bound arising from the purity of the reduced state of qubits 234 of the biseparable pure state. Since the local dimension for these three qubits is $2^3 = 8$ and we have 12 terms appearing [the Bloch vectors of the i th qubit $|\vec{a}_i|$ ($i = 2, 3, 4$) and the correlations tensor elements $|t_{nn}^{23}|$, $|t_{nn}^{24}|$, and $|t_{nn}^{34}|$ for $n = 1, 2, 3$], we find the bound

$$\bar{\mathcal{F}}_{\text{Bell}}^{1|234} \leq \frac{1}{24} \left(6 + 12\sqrt{\frac{8-1}{12}} \right) = \frac{1}{8} (6 + \sqrt{84}) \approx 0.631881. \quad (\text{A49})$$

However, this upper bound is larger than that arising just from using the ACT, and Eq. (A49) is therefore of no further consequence.

The only other possible type of bipartition of 4 qubits is into two sets of 2 qubits. In this case, four triples are cut by the bipartition, but in each set one triple of unpaired operators remains (jointly bounded by 3), such that we have

$$\bar{\mathcal{F}}_{\text{Bell}}^{12|34} \leq \frac{1}{24} (6 + 4 + 3 + 3) = \frac{2}{3}. \quad (\text{A50})$$

As we have argued before, a biseparability bound for 3 qubits that is larger or equal to $\frac{2}{3}$ is not very useful since even pure GME states (e.g., the 4-qubit Dicke state with two excitations) achieve only this value. We hence again turn to using the purity of the subsystems for a biseparable state $|\psi\rangle_{1243} = |\phi\rangle_{12}|\chi\rangle_{34}$. In this case, the symmetric average Bell fidelity can be bounded by

$$\begin{aligned} \bar{\mathcal{F}}_{\text{Bell}}^{12|34} &\leq \frac{1}{24} \left(6 + |\vec{a}_1| |\vec{a}_3| + |\vec{a}_1| |\vec{a}_4| + |\vec{a}_2| |\vec{a}_3| + |\vec{a}_2| |\vec{a}_4| \right. \\ &\quad \left. + \sum_{n=1,2,3} |t_{nn}^{12}| + \sum_{n=1,2,3} |t_{nn}^{34}| \right). \end{aligned} \quad (\text{A51})$$

Here, we encounter a different optimization problem than before, since we no longer seek to maximize the sum of absolute values, but some quantities (e.g., $|\vec{a}_1|$ and $|\vec{a}_3|$) are coupled. However, due to the symmetric form (with respect to the exchange of qubits 12 with 34) of the expression, we may write

$$\bar{\mathcal{F}}_{\text{Bell}}^{12|34} \leq \frac{1}{24} \left(6 + 2 \left(|\vec{a}_1|^2 + |\vec{a}_2|^2 + \sum_{n=1,2,3} |t_{nn}^{12}| \right) \right). \quad (\text{A52})$$

We hence seek to maximize $f(a, t) = 3t + 2a^2$ under the constraints $2a^2 + 3t^2 = 3$ and $a^2 \leq 1$, which is achieved for $a = 1$ and $t = (1/\sqrt{3})$, and hence

$$\bar{\mathcal{F}}_{\text{Bell}}^{12|34} \leq \frac{1}{24} \left(6 + 2 \left[2 + \frac{3}{\sqrt{3}} \right] \right) = \frac{1}{12} (5 + \sqrt{3}) \approx 0.561004. \quad (\text{A53})$$

Since this value is smaller than that for the bipartition $1|234$ in Eq. (A48), we can identify the bound of Eq. (A48) with the biseparability bound for the symmetric average Bell fidelity for 4 qubits; i.e.,

$$\bar{\mathcal{F}}_{\text{Bell}}^{(4)\text{bisep}} \leq \frac{1}{8} (3 + \sqrt{3}) \approx 0.591506. \quad (\text{A54})$$

For more than 4 qubits, we can derive more general expressions using the method based on the ACT, while the exponentially increasing subsystem dimension makes bounds based on the subsystem purity unfeasible. Consider a system of $N \geq 4$ qubits that is in a separable pure state with respect to a bipartition into a single qubit versus the remaining $N-1$ qubits. One may then identify $N-1$ triples of expectation values that factorize and can be bounded by one, while the remaining $b_{N-1} = b_N - (N-1)$ triples form anticommuting sets of three which are each bounded by $\sqrt{3}$. We thus have

$$\begin{aligned} \bar{\mathcal{F}}_{\text{Bell}}^{1|23\dots N} &\leq \frac{1}{4} \left(1 + \frac{1}{b_N} \left\{ N-1 + [b_N - (N-1)]\sqrt{3} \right\} \right) \\ &= \frac{1}{4} (1 + \sqrt{3}) - \frac{1}{2N} (\sqrt{3} - 1), \end{aligned} \quad (\text{A55})$$

where we have made use of the fact that $[(N-1)/b_N] = \{[(N-1)2(N-2)!/(N!)]\} = (2/N)$. Note that, as required, this bound reduces to the result obtained in Eq. (A48) for $N = 4$. Intuitively, it is now clear that other bipartitions will provide smaller upper bounds, since more operators are

affected by the factorization. The exception being the case $N = 4$, where we have already seen that the separation into two sets of two provides a larger upper bound since each side then features unpaired expectation values.

To confirm this, let us briefly consider bipartitions into 2 and $N-2$ qubits for $N \geq 5$. In such a case, $2(N-2)$ expectation values factorize for the respective pure, separable states, and one triple of operators pertaining to the two isolated qubits cannot be paired with anticommuting partners, whereas $b_{N-2} = b_N - (N-2) = b_N - 2(N-1) + 1$ triples of operators can be matched up in this way. Thus, we have

$$\begin{aligned} \bar{\mathcal{F}}_{\text{Bell}}^{12|34\dots N} &\leq \frac{1}{4} \left(1 + \frac{1}{b_N} \left[2(N-2) + 3 + (b_N - 2N + 3)\sqrt{3} \right] \right) \\ &= \frac{1}{4} (1 + \sqrt{3}) \left(1 + \frac{1}{b_N} \right) - \frac{1}{N} (\sqrt{3} - 1). \end{aligned} \quad (\text{A56})$$

This expression provides a smaller upper bound when

$$\frac{1}{2N} (\sqrt{3} - 1) > \frac{1}{4b_N} (1 + \sqrt{3}) \quad (\text{A57a})$$

$$\Rightarrow 1 > \frac{1}{2(N-1)} (1 + \sqrt{3})^2, \quad (\text{A57b})$$

which is the case for $N \geq 5$, as expected. We have also confirmed that this intuition holds for bipartitions into k versus $N-k$ qubits for $3 \leq k \leq N-3$. We can hence formulate the biseparability bound based on the symmetric average Bell fidelity for arbitrary numbers of qubits in the following way. For any biseparable state of N qubits, the symmetric average Bell fidelity satisfies

$$\begin{aligned} \bar{\mathcal{F}}_{\text{Bell}}^{(N)} &\leq \bar{\mathcal{F}}_{\text{Bell}}^{(N)\text{bisep}} \\ &:= \begin{cases} \frac{1}{12} (3 + \sqrt{15}) & \text{for } N = 3 \\ \frac{1}{4} (1 + \sqrt{3}) - \frac{1}{2N} (\sqrt{3} - 1) & \text{for } N \geq 4. \end{cases} \end{aligned} \quad (\text{A58})$$

Conversely, any state that violates the inequality Eq. (A58) is genuinely N -partite entangled. Before we discuss the practical usefulness of these witnesses, let us briefly analyze possible improvements in Appendix A 2 d.

d. Optimizing GME witnesses based on bipartite fidelities

To keep the notation simple during the derivations, we have thus far used only expectation values of pairs of the same Pauli operators, i.e., of the form $|\langle O_i O_j \rangle|$. In practice, this corresponds to measuring the real part of certain off-diagonal elements of the density operator. To see this, consider a 2-qubit state ρ and note that

$$\text{Tr}[\rho(X_1X_2 + Y_1Y_2)] = 4\text{Re}(\langle 01|\rho|10\rangle). \quad (\text{A59})$$

Of course, the off-diagonal element $\langle 01|\rho|10\rangle$ need not be real for a given 2-qubit state. Here, one may note that the derivations of all bounds that we have considered so far are invariant under local unitary transformations. That is, we can replace the triple of operators $\{X_i, Y_i, Z_i\}$ for the i th qubit with the rotated operators $\tilde{O}_i = U_i O_i U_i^\dagger$ for any unitary U_i . This is the case because such a rotation maps a triple of anticommuting operators to another triple of anticommuting operators, and the length of the Bloch vectors also is left invariant. For instance, one could perform a rotation in the equatorial plane of the Bloch sphere, and map

$$X_i \mapsto \tilde{X}_i = \cos(\theta_i)X_i - \sin(\theta_i)Y_i, \quad (\text{A60})$$

$$Y_i \mapsto \tilde{Y}_i = \sin(\theta_i)X_i + \cos(\theta_i)Y_i. \quad (\text{A61})$$

In the example of Eq. (A59) this means we can pick $\theta_1 = 0$ and $\theta_2 = -(\pi/2)$ to obtain

$$\begin{aligned} \text{Tr}[\rho(\tilde{X}_1\tilde{X}_2 + \tilde{Y}_1\tilde{Y}_2)] &= \text{Tr}[\rho(X_1Y_2 - Y_1X_2)] \\ &= 4\text{Im}(\langle 01|\rho|10\rangle). \end{aligned} \quad (\text{A62})$$

In particular, there exist rotation angles θ_1 and θ_2 such that

$$\text{Tr}[\rho(\tilde{X}_1\tilde{X}_2 + \tilde{Y}_1\tilde{Y}_2)] = 4|\langle 01|\rho|10\rangle|. \quad (\text{A63})$$

In general, one hence has the freedom of N independent transformations $U_i \in U(2)$ to optimize the GME witnesses presented so far. In an experimental setting, this optimization can be done *a priori* if the quantum state ρ that one expects to produce (approximately) in the experiment is known. However, if the underlying state is unknown, one may also measure all combinations of 2-qubit Pauli operators for all pairs of qubits within the set of N qubits [amounting to $9b_N = \frac{9}{2}N(N-1)$ 2-qubit measurements] and perform the optimization on the experimental data. For instance, the results for $\bar{\mathcal{F}}_{\text{Bell}}^{(3)}$ presented in Fig. 3 of the main text have been obtained by such a postprocessing of available measurement data, and the corresponding optimization has been restricted to rotations in the X - Y planes, as shown in Eq. (A61).

In addition to *a posteriori* optimization, one may perform some of these 2-qubit measurements on different pairs simultaneously if the individual outcomes for each qubit are recorded. For instance, in a register of $N \geq 3$ qubits one may obtain the expectation values of X_1X_2 and X_2Y_3 from measuring the first and second qubit in the eigenbasis of X and the third in the eigenbasis of Y and recording all three outcomes in each run. For measurements of this kind, such as have been performed in our

experiment, data used to estimate different 2-qubit expectation values may be correlated, which has to be taken into account in the calculation of the estimate for the variance of the GME witness. This is explained in more detail in Ref. [19] (therein see Supplemental Material Secs. IV. A. 4 and IV. A. 5, pp. 9–13).

e. Usefulness of GME witnesses based on bipartite fidelities

A crucial question when employing these witnesses is of course whether or not states exist that can be detected by them. To analyze this problem, we compare the upper bound $\bar{\mathcal{F}}_{\text{Bell}}^{(N)\text{bisep}}$ for biseparable states with the upper bound $\bar{\mathcal{F}}_{\text{Bell}}^{(N)\text{max}}$ for arbitrary states from Eq. (A28) by calculating their distance as a function of the number of qubits. We find the expression

$$\bar{\mathcal{F}}_{\text{Bell}}^{(N)\text{max}} - \bar{\mathcal{F}}_{\text{Bell}}^{(N)\text{bisep}} = \begin{cases} \frac{1}{12}\sqrt{3}(3 - \sqrt{5}) & (N = 3) \\ \frac{1}{2N}(\sqrt{3} - 1) & (N \geq 4), \end{cases} \quad (\text{A64})$$

where the numerical values for 3 and 4 qubits are $\frac{1}{12}\sqrt{3}(3 - \sqrt{5}) \approx 0.110264$ and $\frac{1}{8}(\sqrt{3} - 1) \approx 0.0915064$, respectively. The gap between the bounds is hence largest for $N = 3$, and shrinks with increasing number of qubits. It is hence expected that there is some finite N for which no GME states exist that are detected by our witnesses, and at this point, we cannot say for which N this occurs.

For 3 qubits, we have already found examples of genuinely tripartite entangled states that can be detected, i.e., the 3-qubit W state $|\psi_W^{(3)}\rangle$ and the 2-excitation Dicke state $|\psi_{D,2}^{(3)}\rangle$, which provide symmetric average Bell fidelities of $\frac{2}{3}$. The experimental results discussed in the main text (see Fig. 3) further show that $\bar{\mathcal{F}}_{\text{Bell}}^{(3)}$ is also a useful witness for mixed states produced in realistic situations. Beyond 3 qubits, the witnesses $\bar{\mathcal{F}}_{\text{Bell}}^{(N \geq 4)}$ (optimized only over rotations in the X - Y plane; see Appendix A 2 d) have not been able to detect GME in our experimental setting. However, we know that 4-qubit states exist, e.g., the 4-qubit 2-excitation Dicke state $|\psi_{D,2}^{(4)}\rangle$ which could be detected in this way, since $\bar{\mathcal{F}}_{\text{Bell}}^{(4)}(|\psi_{D,2}^{(4)}\rangle) = \frac{2}{3}$. Unfortunately, the 2-excitation Dicke state for 5 qubits only provides a value of $\bar{\mathcal{F}}_{\text{Bell}}^{(5)}(|\psi_{D,2}^{(5)}\rangle) = 0.6$, whereas $\bar{\mathcal{F}}_{\text{Bell}}^{(N)\text{bisep}} = \frac{1}{20}(7 + 3\sqrt{3}) \approx 0.61$. Tentative searches for other 5-qubit states for which $\bar{\mathcal{F}}_{\text{Bell}}^{(5)}$ exceeds the biseparability bound have been unsuccessful thus far. The question of whether GME states exist that can be detected with our method for $N \geq 5$ hence remains open.

APPENDIX B: CONSTRUCTION OF WITNESSES BASED ON NUMERICAL SEARCH

1. Genuine multipartite negativity

We now discuss the genuine negativity (GMN) of Ref. [27] that we use to quantify GME in the experiment. We present its definition as a convex-roof construction and the alternative way of writing it in terms of a semidefinite program, which turns it into a numerically computable measure of entanglement for an arbitrary mixed state.

We start by introducing the notation and presenting preliminary definitions. Note that a bipartition of $\{1, \dots, N\}$ can be specified by a subset $A \subset \{1, \dots, N\}$ and its complement $\bar{A} = \{1, \dots, N\} \setminus A$. With this, we define the partial transposition on A for an operator $X_A \otimes X_{\bar{A}}$, where X_A and $X_{\bar{A}}$ act on the Hilbert spaces associated to A and \bar{A} , respectively, as $(X_A \otimes X_{\bar{A}})^{T_A} = X_A^T \otimes X_{\bar{A}}$. The definition then extends to any operator on the N -particle Hilbert space by linearity. Next, the negativity of a bipartite quantum state ρ with respect to the bipartition $A|\bar{A}$ is given by the sum of the negative eigenvalues of the partially transposed density matrix; i.e., $\mathcal{N}_{A|\bar{A}}(\rho) = \sum_{\lambda_i \leq 0} |\lambda_i(\rho^{T_A})| = \mathcal{N}_{A|\bar{A}}(\rho) = [(\|\rho^{T_A}\|_1 - 1)/2]$, where $\lambda_i(X)$ denotes the i th eigenvalue of an operator X . With this, the GMN of an N -particle state ρ is given by

$$\mathcal{N}_g(\rho) = \min_{\{p_i, \rho_i\}} \sum_i p_i \min_{A|\bar{A}} \mathcal{N}_{A|\bar{A}}(\rho_i), \quad (\text{B1})$$

where the inner minimization is over the possible bipartitions $A|\bar{A}$ of $\{1, \dots, N\}$ and the outer minimization is over decompositions $\rho = \sum_i p_i \rho_i$, where $\{p_i\}_i$ is a probability distribution and ρ_i are density matrices. For pure states, this reduces to $\mathcal{N}_g(|\psi\rangle\langle\psi|) = \min_{A|\bar{A}} \mathcal{N}_{A|\bar{A}}(|\psi\rangle\langle\psi|)$.

For mixed states the GMN can still efficiently be computed using numerical tools from the field of semidefinite programming. This follows from the fact that the GMN is given by the optimal value of the following optimization:

$$\begin{aligned} \mathcal{N}_g(\rho) &= \max_{Q, P_A, R_A} [-\text{Tr}(Q\rho)] \\ Q &= K_A + Q_A^{T_A} \quad \forall A|\bar{A}, \\ 0 &\leq K_A \quad \text{and} \quad 0 \leq R_A \leq 1, \end{aligned} \quad (\text{B2})$$

where Q , K_A , and R_A are operators acting on the Hilbert space. The GMN is zero for all (bi)separable states and, therefore, a nonzero value provides a way to certify GME. More precisely, the GMN is nonzero for any state that cannot be written as a positive partial transpose (PPT) mixture. Recall that a multipartite state ρ is called a PPT mixture if it admits a mixed state decomposition $\rho = \sum_A p_A \rho_A$, where $\{p_A\}_A$ is a probability distribution and ρ_A has a positive partial transposition (we say, ρ_A is PPT) with respect to the bipartition $A|\bar{A}$. That is, formally,

$\rho_A^{T_A} \geq 0$. As noted earlier, a state is called biseparable if it can be written as a convex combination of states that are separable with respect to one bipartition $A|\bar{A}$. Since any separable state is PPT, every biseparable state can be written as a PPT mixture. Consequently, a state with nonzero GMN is GME.

Moreover, the GMN quantifies the entanglement in the sense that a state ρ is more entangled than σ if $\mathcal{N}_g(\rho) \geq \mathcal{N}_g(\sigma)$. The underlying mathematical property is that the GMN is nonincreasing under so-called full local operations and classical communication operations. In particular, from this property it follows that no GME state can be generated from a non-GME state with local operations only.

A further beneficial property of writing the GMN as in Eq. (B2) is that this yields an entanglement witness, that is, an observable that provides ideally a sharp lower bound to the GMN as in Eq. (5) in the main text and may be accessed experimentally. For our purposes note that whether a witness can be measured depends on the measurements that are available in our experiment. We therefore discuss those measurements next. The procedure of obtaining the entanglement witnesses that are accessible for us is then described subsequently in Appendix B 3 in more detail.

2. Accessible measurements on k neighboring sites

Here, we discuss the operators that can be measured locally on k consecutive sites with the data available in our experiment. Our starting point is hence the set of all possible observables whose measurement outcomes can be obtained from the 27 measurement settings that we mentioned in Sec. III. As we noted there, these measurement settings can be used to obtain estimates of the expectation values of all possible products of the identity and the three Pauli operators on all groups of three neighboring qubits in the chain of 20 qubits. On such triples of neighboring qubits this thus allows us to estimate the expectation value of any operator, in particular, any possible entanglement witness. However, the situation is different for more than three neighboring sites. Regarding this case, recall that in the 27 settings we consider, the 20 qubits are measured simultaneously such that the accessible information turns out to be more than just knowing the three-body reductions of neighboring qubits. Also as a consequence of how we choose these 27 settings, each of them has the property that the local bases in which sites i and $i + 3$ are measured are identical. To illustrate what this implies for the accessible operators, let us consider the example of four sites. In terms of projective qubit measurements, a basis of operators that we can measure with the 27 settings is given by $\mathcal{B}_4 = \{P_{\vec{s}, \vec{\alpha}}^{(4)}\}_{\vec{s}, \vec{\alpha}}$, where $P_{\vec{s}, \vec{\alpha}}^{(4)}$ denotes the projector onto the state

$$|\vec{s}, \vec{\alpha}\rangle = |s_1, s_2, s_3, s_4\rangle_{\alpha_1, \alpha_2, \alpha_3, \alpha_1}, \quad (\text{B3})$$

where $\vec{s} \equiv (s_1, s_2, s_3, s_4) \in \{\uparrow, \downarrow\}^{\times 4}$ and $\vec{\alpha} \equiv (\alpha_1, \alpha_2, \alpha_3) \in \{1, 2, 3\}^{\times 3}$. In words, these operators comprise all projective qubit measurements where the first and the fourth qubit are measured in the same direction. Notably, with the Pauli operators X, Y , and Z , this set of operators spans the same subspace as the operators $\{\sigma_{\alpha_1} \otimes \sigma_{\alpha_2} \otimes \sigma_{\alpha_3} \otimes \sigma_{\alpha_1}\}_{\vec{\alpha}}$, where $\vec{\alpha} \in \{0, 1, 2, 3\}^{\times 3}$ and $\sigma_0 = \mathbb{1}$, $\sigma_1 = X$, $\sigma_2 = Y$, $\sigma_3 = Z$. For the construction of the witness we will, however, use the projectors as our basis set. Accordingly, generalizing this to arbitrary number of sites, we denote the projectors that form a basis of the operators that can be measured on k consecutive sites with the 27 settings by \mathcal{B}_k . As a further remark, let us mention that the number of elements in \mathcal{B}_k grows exponentially with k as $|\mathcal{B}_k| = 27 \times 2^k$.

3. Accessible quantitative entanglement witnesses

Next, we turn to the construction of an entanglement witness that is fully accessible from the available information provided by the 27 measurement settings. The main step is to solve the optimization given in Eq. (B6) below, which provides a quantitative witness. The witness can then be evaluated using the experimental data.

We distinguish between a simple entanglement witness, which is an operator Q that has a positive expectation value $\text{Tr}(Q\rho_{\text{sep}}) \geq 0$ for any separable state ρ_{sep} and for which there exists at least one entangled state ρ with $\text{Tr}(Q\rho) \leq 0$, and a *quantitative* entanglement witness, i.e., an operator that fulfills the property of being an entanglement witness and additionally provides a lower bound to the GMN via (minus) its expectation value $\mathcal{S} = -\text{Tr}(Q\rho)$ of the form

$$\mathcal{S} \leq \mathcal{N}_g(\rho) \quad (\text{B4})$$

for any ρ , as noted in the main text.

Now, considering k neighboring sites in the chain of 20 qubits, the available information from the 27 settings is determined by the projectors in the set \mathcal{B}_k as described above. Then, an entanglement witness \hat{W} is accessible from this information if it can be written as a linear combination of operators from \mathcal{B}_k , i.e., if

$$Q_i^{(k)} = \sum_{\vec{s}, \vec{\alpha}} c_{i;\vec{s}, \vec{\alpha}}^{(k)} P_{\vec{s}, \vec{\alpha}}^{(k)}, \quad (\text{B5})$$

with coefficients $c_{i;\vec{s}, \vec{\alpha}}^{(k)} \in \mathbb{R}$, since, in this case, it is fully determined by the set \mathcal{B}_k .

Here, for a given state $\rho_i^{(k)}$ (as defined in the main text), we can optimize over the coefficients $c_{i;\vec{s}, \vec{\alpha}}^{(k)}$ in order to find a quantitative witness that provides the best lower bound to the GMN of ρ . As the computation of the GMN itself, see Eq. (B2), this optimization can be expressed as a semidefinite program. That is, with the definition $p_{i;\vec{s}, \vec{\alpha}}^{(k)} = \text{Tr}(P_{\vec{s}, \vec{\alpha}}^{(k)} \rho_i^{(k)})$, the solution of

$$\begin{aligned} \mathcal{S}_i^{(k)}(\{P_{\vec{s}, \vec{\alpha}}^{(k)}, P_{i;\vec{s}, \vec{\alpha}}^{(k)}\}) &= \max_{c_{i;\vec{s}, \vec{\alpha}}^{(k)}, K_A, R_A} \left(-\sum_{\vec{s}, \vec{\alpha}} c_{i;\vec{s}, \vec{\alpha}}^{(k)} P_{\vec{s}, \vec{\alpha}}^{(k)} \right) \\ \sum_{\vec{s}, \vec{\alpha}} c_{i;\vec{s}, \vec{\alpha}}^{(k)} P_{\vec{s}, \vec{\alpha}}^{(k)} &\geq K_A + R_A^T A \quad \forall A | \bar{A}, \\ 0 &\leq K_A \quad \text{and} \quad 0 \leq R_A \leq \mathbb{1}, \end{aligned} \quad (\text{B6})$$

is the best lower bound of the form $\mathcal{S}_i^{(k)}(\{P_{\vec{s}, \vec{\alpha}}^{(k)}, P_{i;\vec{s}, \vec{\alpha}}^{(k)}\}) = -\text{Tr}(Q_i^{(k)} \rho_i^{(k)}) \leq \mathcal{N}_g(\rho_i^{(k)})$, with $Q_i^{(k)}$ a quantitative witness as in Eq. (B5) and with the coefficients for which the maximum in Eq. (B6) is achieved. Note that for brevity we denote both the optimization parameters in Eq. (B6) as well as, in the following, the optimal coefficients by $c_{i;\vec{s}, \vec{\alpha}}^{(k)}$. As a further remark, we note that $\mathcal{S}_i^{(k)}(\{P_{\vec{s}, \vec{\alpha}}^{(k)}, P_{i;\vec{s}, \vec{\alpha}}^{(k)}\})$ and the optimal witness Q depend only on the (accessible) probabilities $p_{\vec{s}, \vec{\alpha}}$ (and the corresponding projectors), such that these quantities are sufficient to determine the bound.

In order to obtain the bounds in Fig. 4 we determine an entanglement witness W that can be decomposed as a sum of projectors as in Eq. (B5) and, hence, it can readily be evaluated using the frequencies measured in the experiment. The basic steps to determine a quantitative witness are then (as also described in the main text) as follows. We first perform a numerical simulation to obtain the time-evolved state $\rho_{\text{sim}}(t)$ and determine its reduced density matrices on neighboring sites. Then, for every reduction $\rho_{\text{sim}}^{\mathcal{X}}(t) = \text{Tr}_{\setminus \mathcal{X}}[\rho_{\text{sim}}(t)]$, where $\mathcal{X} \subset \{1, \dots, N\}$ denotes the sites corresponding to the Hilbert space on which $\rho_{\text{sim}}^{\mathcal{X}}(t)$ acts on, we solve the optimization of Eq. (B6) (plus some practical amendments; see Appendix B 4) with $p_{i;\vec{s}, \vec{\alpha}}^{(k)} = \text{Tr}[P_{\vec{s}, \vec{\alpha}}^{(k)} \rho_{\text{sim}}^{\mathcal{X}}(t)]$ as input to obtain the witness.

In the numerical simulation we include mixing due to an imperfect initial state, as described in the main text. To this end, we use an initial state that is diagonal in the Z bases as $\rho_{\text{sim}}(0) = \sum_{\vec{s}} f_{\vec{s}} |\vec{s}\rangle \langle \vec{s}|$, where $|\vec{s}\rangle \equiv |\vec{s}, (3, \dots, 3)\rangle$ is a state of the product basis in the Z direction [see Eq. (B3) for the notation we use] and $f_{\vec{s}}$ is the frequency of the qubit configuration \vec{s} that we observe using the experimental data obtained at $t = 0$ ms. Since, besides the ideal initial state with alternating qubits, only 42 other configurations with nonzero frequency $f_{\vec{s}}$ occur, the participating configurations are readily evolved separately and mixed in order to obtain a density matrix $\rho_{\text{sim}}(t)$ at later times.

The optimization of Eq. (B6) used to determine the QEW $Q_i^{(k)}$ depends on the probabilities $p_{i;\vec{s}, \vec{\alpha}}^{(k)}$ but not on the underlying quantum state $\rho_i^{(k)}$. One may be tempted to insert experimentally measured estimates for the probabilities $p_{i;\vec{s}, \vec{\alpha}}^{(k)}$ to determine a witness which is optimal for the (unknown) quantum state of the experiment. However, there is, in general, no quantum state $\rho_i^{(k)}$ which satisfies

$\text{Tr}(P_{\vec{s},\vec{\alpha}}^{(k)}\rho_i^{(k)}) = \tilde{p}_{i;\vec{s},\vec{\alpha}}^{(k)}$ where the $\tilde{p}_{i;\vec{s},\vec{\alpha}}^{(k)}$ are experimentally estimated probabilities because the $\tilde{p}_{i;\vec{s},\vec{\alpha}}^{(k)}$ are affected by statistical noise from a finite number of measurements. As a consequence, the optimization may fail to be feasible. Therefore, we use the probabilities $p_{i;\vec{s},\vec{\alpha}}^{(k)}$ from the reduced density matrices of the numerically simulated state $\rho_{\text{sim}}(t)$ in the optimization of Eq. (B6) in order to determine the QEW $Q_i^{(k)}$. After $Q_i^{(k)}$ has been determined, its expectation value, which provides the lower bounds in Fig. 4, is determined from the experimental data.

Instead of using the numerically simulated state $\rho_{\text{sim}}(t)$ to determine the witness, we furthermore test states obtained from quantum state tomography. For this purpose we split the data into two statistically independent sets of samples and use one set for tomography and the other to evaluate the witness. We pursue this strategy using global pure-state MPS reconstruction and find no significant advantage of the approach over using the numerically simulated state $\rho_{\text{sim}}(t)$ as a starting point.

After performing the optimization of Eq. (B6) to determine $Q_i^{(k)}$, we perform an additional optimization which makes the QEW more robust to statistical noise and experimental imperfection. The lower bounds presented in Fig. 4 of the main text are based on this improved witness, and the remaining steps are discussed in the next section.

4. Additional conditions for practical witnesses

To make the witnesses less sensitive to experimental imperfections, we modify the coefficients $c_{i;\vec{s},\vec{\alpha}}^{(k)}$ as discussed in the following. The major step is to reduce the number of nonzero coefficients. This can be achieved by employing an interactive l_1 -relaxation method with weight updating that is intended to find sparse solutions [42]. We incorporate this into our considerations by solving the following optimization in the n th iteration step:

$$\begin{aligned} & \min_{c_{i;\vec{s},\vec{\alpha}}^{(k),n}} \sum_{i;\vec{s},\vec{\alpha}} \frac{|c_{i;\vec{s},\vec{\alpha}}^{(k),n}|}{|c_{i;\vec{s},\vec{\alpha}}^{(k),n-1}| + \epsilon} \\ & \text{Tr}[(Q_i^{(k),n} - Q_i^{(k),0})\rho_i^{(k)}] \leq \epsilon, \\ & Q_i^{(k),n} \geq K_A + R_A^{T_A} \quad \forall A|\bar{A}, \end{aligned} \quad (\text{B7})$$

where $Q_i^{(k),n} = \sum_{\vec{s},\vec{\alpha}} c_{i;\vec{s},\vec{\alpha}}^{(k),n} P_{\vec{s},\vec{\alpha}}^{(k)}$, and initially the coefficients are obtained from Eq. (B6) with $c_{\vec{s},\vec{\alpha}}^{(k),0} = c_{i;\vec{s},\vec{\alpha}}^{(k)}$ with the coefficients that are found from solving Eq. (B6). We allow the bound to diminish by at most ϵ from the optimal one in order to find a sparse solution. In the results we present, we choose $\epsilon = 5 \times 10^{-3}$ and use three iteration steps. Except for the tolerance ϵ , this step does not deteriorate the bound. We then further add two simple constraints to Eqs. (B6) and

(B7). First, we choose the coefficients as $c_{i;\vec{s},\vec{\alpha}}^{(k),n} \in [-1, 1]$. Second, we add the semidefinite condition $K_A \leq \mathbb{1}$ to Eq. (B6). Note that this condition has previously been employed in an earlier version of the GMN; see Refs. [26,27]. However, there the quantitative witnesses are exact functions of the operators K_A and R_A , whereas here they are only related via an inequality. Solving Eqs. (B6) and (B7) with the constraints introduced in this section results in the bounds $S_i^{(k)}$ presented in the main text. We observe that the additional modifications significantly help to improve the lower bounds.

-
- [1] M. B. Plenio and S. Virmani, *An Introduction to Entanglement Measures*, *Quantum Inf. Comput.* **7**, 1 (2007).
 - [2] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, *Quantum Entanglement*, *Rev. Mod. Phys.* **81**, 865 (2009).
 - [3] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, *Bell Nonlocality*, *Rev. Mod. Phys.* **86**, 419 (2014).
 - [4] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, U.K., 2000).
 - [5] S. Lloyd, *Universal Quantum Simulators*, *Science* **273**, 1073 (1996).
 - [6] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, *Quantum Cryptography*, *Rev. Mod. Phys.* **74**, 145 (2002).
 - [7] H. J. Kimble, *The Quantum Internet*, *Nature (London)* **453**, 1023 (2008).
 - [8] D. J. Wineland, J. J. Bollinger, W. M. Itano, F. L. Moore, and D. J. Heinzen, *Spin Squeezing and Reduced Quantum Noise in Spectroscopy*, *Phys. Rev. A* **46**, R6797 (1992).
 - [9] S. F. Huelga, C. Macchiavello, T. Pellizzari, A. K. Ekert, M. B. Plenio, and J. I. Cirac, *On the Improvement of Frequency Standards with Quantum Entanglement*, *Phys. Rev. Lett.* **79**, 3865 (1997).
 - [10] V. Giovannetti, S. Lloyd, and L. Maccone, *Quantum Metrology*, *Phys. Rev. Lett.* **96**, 010401 (2006).
 - [11] N. Friis, D. Orsucci, M. Skotiniotis, P. Sekatski, V. Dunjko, H. J. Briegel, and W. Dür, *Flexible Resources for Quantum Metrology*, *New J. Phys.* **19**, 063044 (2017).
 - [12] G. Vidal, *Efficient Classical Simulation of Slightly Entangled Quantum Computations*, *Phys. Rev. Lett.* **91**, 147902 (2003).
 - [13] T. Monz, P. Schindler, J. T. Barreiro, M. Chwalla, D. Nigg, W. A. Coish, M. Harlander, W. Hänsel, M. Hennrich, and R. Blatt, *14-Qubit Entanglement: Creation and Coherence*, *Phys. Rev. Lett.* **106**, 130506 (2011).
 - [14] C. Song, K. Xu, W. Liu, C. P. Yang, S.-B. Zheng, H. Deng, Q. Xie, K. Huang, Q. Guo, L. Zhang, P. Zhang, D. Xu, D. Zheng, X. Zhu, H. Wang, Y. A. Chen, C. Y. Lu, S. Han, and J. W. Pan, *10-Qubit Entanglement and Parallel Logic Operations with a Superconducting Circuit*, *Phys. Rev. Lett.* **119**, 180511 (2017).
 - [15] X.-L. Wang, L.-K. Chen, W. Li, H.-L. Huang, C. Liu, C. Chen, Y.-H. Luo, Z.-E. Su, D. Wu, Z.-D. Li, H. Lu, Y. Hu, X. Jiang, C.-Z. Peng, L. Li, N.-L. Liu, Y.-A. Chen, C.-Y. Lu,

- and J.-W. Pan, *Experimental Ten-Photon Entanglement*, *Phys. Rev. Lett.* **117**, 210502 (2016).
- [16] O. Gühne and G. Tóth, *Entanglement Detection*, *Phys. Rep.* **474**, 1 (2009).
- [17] C. Hempel, Ph.D. thesis, University of Innsbruck, 2014.
- [18] P. Schindler, D. Nigg, T. Monz, J. T. Barreiro, E. Martinez, S. X. Wang, S. Quint, M. F. Brandl, V. Nebendahl, C. F. Roos, M. Chwalla, M. Hennrich, and R. Blatt, *A Quantum Information Processor with Trapped Ions*, *New J. Phys.* **15**, 123012 (2013).
- [19] B. P. Lanyon, C. Maier, M. Holzäpfel, T. Baumgratz, C. Hempel, P. Jurcevic, I. Dhand, A. S. Buyskikh, A. J. Daley, M. Cramer, M. B. Plenio, R. Blatt, and C. F. Roos, *Efficient Tomography of a Quantum Many-Body System*, *Nat. Phys.* **13**, 1158 (2017).
- [20] Petar Jurcevic, Ph.D. thesis, University of Innsbruck, 2017.
- [21] K. Kim, M.-S. Chang, R. Islam, S. Korenblit, L.-M. Duan, and C. Monroe, *Entanglement and Tunable Spin-Spin Couplings between Trapped Ions Using Multiple Transverse Modes*, *Phys. Rev. Lett.* **103**, 120502 (2009).
- [22] P. Jurcevic, B. P. Lanyon, P. Hauke, C. Hempel, P. Zoller, R. Blatt, and C. F. Roos, *Quasiparticle Engineering and Entanglement Propagation in a Quantum Many-Body System*, *Nature (London)* **511**, 202 (2014).
- [23] M. Cramer and M. B. Plenio, *Reconstructing Quantum States Efficiently*, [arXiv:1002.3780](https://arxiv.org/abs/1002.3780).
- [24] S. T. Flammia, D. Gross, S. D. Bartlett, and R. Somma, *Heralded Polynomial-Time Quantum State Tomography*, [arXiv:1002.3839](https://arxiv.org/abs/1002.3839).
- [25] M. Cramer, M. B. Plenio, S. T. Flammia, R. Somma, D. Gross, S. D. Bartlett, O. Landon-Cardinal, D. Poulin, and Y.-K. Liu, *Efficient Quantum State Tomography*, *Nat. Commun.* **1**, 149 (2010).
- [26] B. Jungnitsch, T. Moroder, and O. Gühne, *Taming Multiparticle Entanglement*, *Phys. Rev. Lett.* **106**, 190502 (2011).
- [27] M. Hofmann, T. Moroder, and O. Gühne, *Analytical Characterization of the Genuine Multiparticle Negativity*, *J. Phys. A* **47**, 155301 (2014).
- [28] M. B. Plenio, *Logarithmic Negativity: A Full Entanglement Monotone That Is Not Convex*, *Phys. Rev. Lett.* **95**, 090503 (2005).
- [29] B. Efron and R. Tibshirani, *Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy*, *Stat. Sci.* **1**, 54 (1986).
- [30] Take, for example, the k -qubit state with density operator $\rho = \frac{1}{2} \left(\bigotimes_{i=1}^{k/2} |\phi_{2i-1,2i}^+\rangle \langle \phi_{2i-1,2i}^+| + \bigotimes_{i=1}^{k/2} |\phi_{2i,2i+1}^+\rangle \langle \phi_{2i,2i+1}^+| \right)$, where k is assumed to be even and $|\phi_{i,j}^+\rangle = (1/\sqrt{2})(|0\rangle_i |0\rangle_j + |1\rangle_i |1\rangle_j)$, with the second index being understood as modulo k . The state ρ is a convex combination of $(k/2)$ -separable states, yet the reduced state $\rho_{i,i+1} = \frac{1}{2}(|\phi^+\rangle \langle \phi^+| + \frac{1}{4}\mathbb{1})$ of every pair of neighboring qubits is entangled.
- [31] C. Lancien, O. Gühne, R. Sengupta, and M. Huber, *Relaxations of Separability in Multipartite Systems: Semidefinite Programs, Witnesses and Volumes*, *J. Phys. A* **48**, 505302 (2015).
- [32] For qubits $\beta \leq \frac{1}{2}$ and thus the best possible noise resistance is 50% white noise. While this may seem strong from a bipartite intuition, many multipartite states in fact have asymptotically perfect white noise resistance, i.e., approaching 100% polynomially in system dimension, which would inevitably be missed by such simple witness constructions.
- [33] S. T. Flammia and Y.-K. Liu, *Direct Fidelity Estimation from Few Pauli Measurements*, *Phys. Rev. Lett.* **106**, 230501 (2011).
- [34] For instance, pure 4-qubit Dicke states with two excitations would yield $\bar{\mathcal{F}}_{\text{Bell}}^{(4)} = \frac{2}{3} > 0.592$. Beyond 4 qubits, we cannot say with certainty whether states exist for which our witness could certify GME in principle. For a discussion, see Appendix A 2 c.
- [35] S. Boyd and L. Vandenberghe, *Convex Optimization*, 7th ed. (Cambridge University Press, Cambridge, U.K., 2004).
- [36] R. H. Tütüncü, K. C. Toh, and M. J. Todd, *Solving Semidefinite-Quadratic-Linear Programs Using SDPT3*, *Math. Program., Ser. B* **95**, 189 (2003).
- [37] C. Schwemmer, L. Knips, D. Richart, H. Weinfurter, T. Moroder, M. Kleinmann, and O. Gühne, *Systematic Errors in Current Quantum State Tomography Tools*, *Phys. Rev. Lett.* **114**, 080403 (2015).
- [38] G. Tóth and O. Gühne, *Entanglement Detection in the Stabilizer Formalism*, *Phys. Rev. A* **72**, 022340 (2005).
- [39] A. Asadian, P. Erker, M. Huber, and C. Klöckl, *Heisenberg-Weyl Observables: Bloch Vectors in Phase Space*, *Phys. Rev. A* **94**, 010301(R) (2016).
- [40] J.-Y. Wu, H. Kampermann, D. Bruß, C. Klöckl, and M. Huber, *Determining Lower Bounds on a Measure of Multipartite Entanglement from Few Local Observables*, *Phys. Rev. A* **86**, 022319 (2012).
- [41] C. Klöckl and M. Huber, *Characterizing Multipartite Entanglement without Shared Reference Frames*, *Phys. Rev. A* **91**, 042339 (2015).
- [42] E. J. Candès, M. B. Wakin, and S. P. Boyd, *Enhancing Sparsity by Reweighted ℓ_1 Minimization*, *J. Fourier Anal. Appl.* **14**, 877 (2008).