Programmable digital quantum simulation of 2D Fermi-Hubbard dynamics using 72 superconducting qubits

Phasecraft[†]

(Dated: November 3, 2025)

Simulating the time-dynamics of quantum many-body systems was the original use of quantum computers proposed by Feynman [1], motivated by the critical role of quantum interactions between electrons in the properties of materials and molecules. Accurately simulating such systems remains one of the most promising applications of general-purpose digital quantum computers, in which all the parameters of the model can be programmed and any desired physical quantity output. However, performing such simulations on today's quantum computers at a scale beyond the reach of classical methods requires advances in the efficiency of simulation algorithms and error mitigation techniques. Here we demonstrate programmable digital quantum simulation of the dynamics of the 2D Fermi-Hubbard model – one of the best-known simplified models of electrons in crystalline solids – at a scale beyond exact classical simulation. We implement simulations of this model on lattice sizes up to 6×6 using 72 qubits on Google's Willow quantum processor, across a range of physical parameters, including on-site electron-electron interaction strength and magnetic flux, and study phenomena including formation of magnetic polarons [2], i.e. charge carriers surrounded by local magnetic polarisation, dynamical symmetry breaking in stripe-ordered states [3], attraction of charge carriers on an entangled state known as a valence bond solid [4], and the approach to equilibrium through thermalisation [5]. We validate our results against exact calculations in parameter regimes where these are feasible, and compare them to approximate classical simulations performed using tensor network and operator propagation methods. Our results demonstrate that programmable digital quantum simulation of many-body interacting electron models is now competitive on state-of-the-art quantum hardware.

Accurate digital simulations of the properties of interacting fermions, modelling electrons in crystalline solids or in molecules, have the potential to accelerate the discovery of novel materials and chemicals on which modern technology such as batteries [6–9] and photovoltaics [10] depends. Critically, as in other areas where computer simulation has revolutionised the way we model and design physical artefacts, accurate digital quantum simulation promises the ability to rapidly model and calculate properties of a wide range of candidate materials or molecules, significantly faster and more flexibly than synthesising samples in the laboratory.

Simulating the time-dynamics of quantum systems evolving under a many-body Hamiltonian – time-dynamics simulation for short – is one of the most extensively studied families of quantum algorithms [11], with a history dating back to the advent of the field [1], and many important algorithmic developments since [12, 13]. It has been proven that time-dynamics simulation can be performed efficiently on a quantum computer, and that classical computers cannot perform this task efficiently [14] – barring the unlikely scenario that they are somehow able to simulate all quantum computers efficiently.

This is also borne out in practice. Whilst classical algorithms such as density functional theory, quantum Monte-Carlo and tensor network methods are often very effective at computing equilibrium properties of quantum many-body systems, simulating the time-dynamics of such systems is more computationally challenging on classical computers. Exact state-vector simulations – wherein the full quantum state is represented and operated upon in memory – are limited to small sizes. Meanwhile, approximate simulation methods such as tensor networks and operator-propagation methods can achieve larger scales, but their accuracy diminishes as entanglement and correlations build up in the model. Quantum computers hold the promise of significantly outperforming classical methods, while still offering the versatility of programmable digital simulation.

Modelling the electronic structure of periodic crystalline materials is one of the more promising near-term applications of digital quantum simulation [15–17]. Due to symmetries and local structure, these systems can be compressed into comparatively small circuits [17]. But time-dynamics simulation of real crystalline materials systems is still beyond the reach of current quantum computers.

However, the simulation of simplified theoretical models of electronic lattice systems, at scales beyond what is possible for exact state-vector simulation and which are challenging for leading approximate classical methods, are now within reach of state-of-the-art quantum hardware. Amongst the simplest and best-studied of these is the Fermi-Hubbard model, a single-band approximation of transition metals with $3d$ orbitals exhibiting local repulsion. Despite its apparent simplicity, this model has a complex phase diagram, non-trivial dynamics and rich physics phenomenology [18].

The Fermi-Hubbard Hamiltonian

$$H = -J \sum_{\langle i,j \rangle, \sigma} (e^{i\phi_{i,j}} c^\dagger_{i,\sigma} c_{j,\sigma} + h.c.) + U \sum_i n_{i,\uparrow} n_{i,\downarrow} \quad (1)$$

represents electrons (fermions) with spin $\sigma = \uparrow$ or $\downarrow$, hopping between neighbouring sites i and j on a 2D square lattice with hopping strength $-J$, and interacting locally with strength U when two electrons meet at the same site. Here, $c^\dagger_{i,\sigma}$ and $c_{i,\sigma}$ are fermionic creation and annihilation

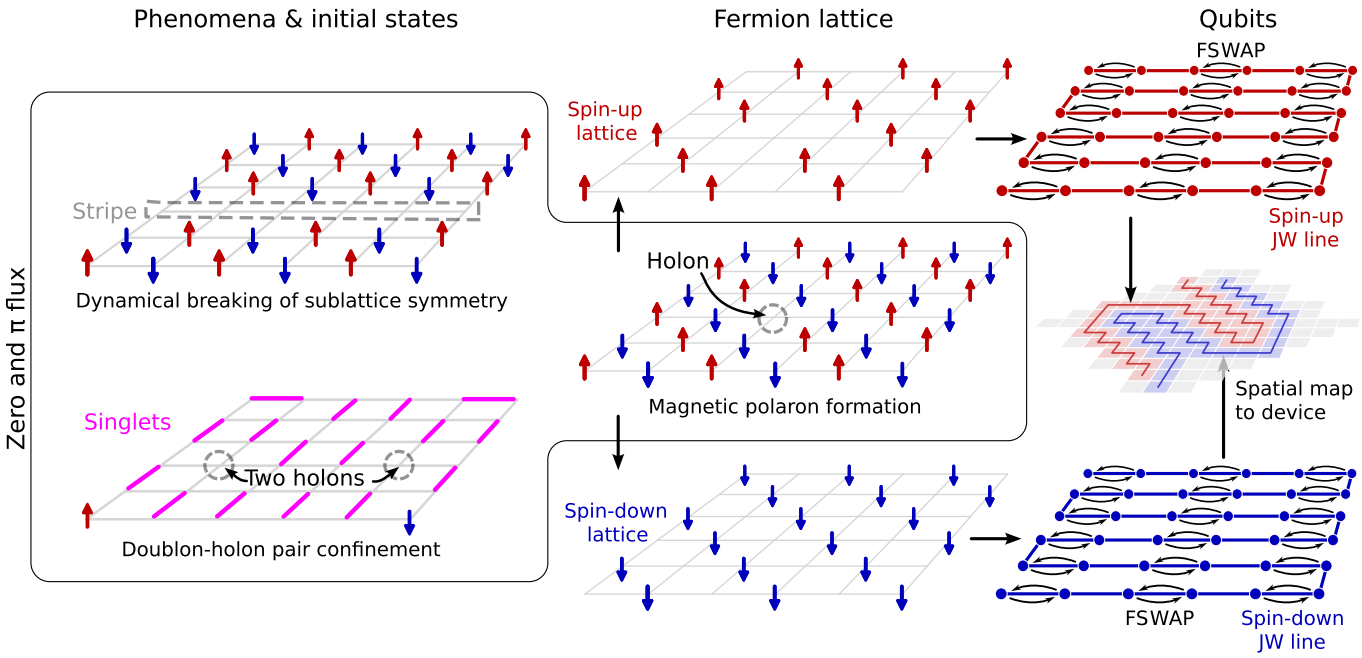


FIG. 1: **Selection of states considered and their representation on hardware. Left,** Initial states and physical phenomena considered in this work. Starting from a product state or a product of singlets, we evolve up to time 1.2 in units of inverse hopping and measure in the computational basis. We then compute expectation values of observables, and analyse signals based directly on the measured bitstrings. We perform approximate classical simulations to benchmark the results. **Right,** An illustration of the representation of the Fermi-Hubbard model on hardware. Each spin sector of the model is separately mapped to a Jordan-Wigner representation in which qubits are ordered along a line. Time dynamics is simulated via fermionic swap networks along this line. The two representations are then arranged on the hardware in an interlocking pattern.

operators for fermions with spin σ at site i, and the local density operator $n_{i,\sigma} = c^\dagger_{i,\sigma} c_{i,\sigma}$ counts the number of such fermions. The Peierls phases $\phi_{i,j}$ model an external magnetic flux through the system, and we choose them in a translationally invariant pattern (see Appendices A and D). As only the ratio U/J matters, we use the standard convention of setting $J = 1$. Throughout this work, we use natural units where Planck's constant $\hbar = 1$.

In this work, we realise a fully programmable, digital quantum simulation of the time dynamics of the 2D Fermi-Hubbard model on a square lattice, using Google's Willow superconducting quantum processor [19], at a scale that is beyond direct classical state-vector simulation, and challenging for tensor network and other leading classical approximate methods. We use this digital simulator to explore several aspects of Fermi-Hubbard dynamics on a 2D square lattice. Specifically, we consider Fermi-Hubbard lattice sizes up to 6×6 sites, requiring a total of up to 72 qubits. Using different initial states (Fig. 1), we study the formation of magnetic polarons emerging from a single holon (a quasiparticle indicating absence of electrons on a site) on a chequerboard pattern of up and down spins, known as a Néel state, the dynamical breaking of bipartite sublattice symmetry on a hole-doped state; the confinement of charge carriers in a valence bond solid state, a state where neighbouring spins combine into spin-zero objects (singlets); and the thermalisation of the system as a function of interaction strength. We explore these dynamical phenomena over a range of interaction strengths, and for systems with zero and π magnetic flux per plaquette. (See below and Fig. 2 for a discussion of the physics simulation results.)

We are able to resolve a signal from total quantum circuit depths of up to 294, with 4372 two-qubit gates in total, allowing time-dynamics simulation out to time 1.2. We illustrate the layout of the experiment on the device in Fig. 1.

Such deep circuits are made possible by two complementary technical developments. First, continued improvements in the performance and scale of superconducting quantum chips [19], combined with their high clock speed. Second, efficient design of simulation algorithms, circuit implementations, and error mitigation techniques tailored to the target application and hardware. Specifically, we achieve significant constant-factor efficiency savings over standard implementations of fermionic quantum simulations by tailoring the fermion representation to the qubit layout on the device, and using gate implementations of the Fermi-Hubbard Hamiltonian terms that are also optimised for this. (See Implementation and Appendix D for details.)

We also apply training by fermionic linear optics

(TFLO) [20], a training-based error mitigation method which exploits the fact that interacting fermion models such as Fermi-Hubbard become classically exactly solvable when $U = 0$, whilst the quantum circuit structure – and hence the hardware noise – changes minimally as the parameter U is varied, only requiring modifying the rotation angles of certain single-qubit gates. This error mitigation partially corrects systematic components of the effects of noise on few-body observables. For quantities that cannot be reduced to few-body observables, we also implement a novel TFLO-based maximum entropy shot-reweighting method to correct for noise directly on the bitstrings ("shots") sampled from the hardware. We combine these with a suite of other standard methods, including noise averaging and suppression, post-selection on fermion-number conservation, symmetry averaging, Gaussian process regression (GPR) on time series data, and more. (See Implementation for an overview, and Appendix E for extensive discussion and performance comparison of different noise mitigation methods.)

A primary motivation for simulating the Fermi-Hubbard model on quantum hardware is that, despite the importance of the model, classical methods still struggle with this task. Although significant progress has been made in recent decades in approximate simulation of equilibrium properties, such as ground states and phase diagrams – with modern supercomputers enabling good numerical results on lattices up to 60×60 [21, 22] – non-equilibrium and dynamical properties of the Fermi-Hubbard model remain extremely challenging to simulate classically.

Exact state-vector simulations are limited by available memory, with the largest to date able to simulating gate operations on at most 45 qubits [23, 24]. On the Fermi-Hubbard model, this corresponds to systems slightly larger than 4×5. Approximate classical computations can be performed at larger lattice sizes. For example, tensor-network-based simulations of Fermi-Hubbard dynamics, which approximate the quantum state of the system via classically efficient representations, have been demonstrated to approximate the dynamics of 6×5 systems for single-site observables out to time 1 [25]. Meanwhile, Majorana propagation methods can compute expectation values of time-evolved few-body observables under complex quantum circuits [26]. However, as these methods involve approximations, just as for quantum computation their accuracy must be validated separately. Comparisons between the leading approximate classical methods and the quantum computation are described below. (A more extensive discussion can be found in the appendices.)

To the best of our knowledge, all previous digital quantum simulations of the dynamics of the spinful Fermi-Hubbard model on quantum computers are either 1D instances [27–30] (which are tractable by tensor network methods) or relatively small 2D instances [31] which are amenable to exact state-vector simulation.

Analogue quantum simulation is an alternative quantum simulation method, where the many-body quantum Hamiltonian of interest is engineered directly into the interactions between, e.g. cold atoms trapped in an optical lattice [32], or between the ions in an ion trap [33]. This has been performed successfully at scale for multiple decades now. These approaches can successfully implement the Fermi-Hubbard model on thousands of sites [34, 35], well beyond the number of qubits currently available on state-of-the-art digital quantum computers, and recent implementations have been able to simulate dynamical properties as well as equilibrium quantities [36–41]. However, pure analogue simulation approaches have drawbacks compared to digital simulation. In particular, the fact that (by definition) arbitrary quantum circuits cannot be implemented in purely analogue quantum simulators limits somewhat which states the simulation can be initialised in, and also restricts which observables can be accessed at the end of the simulation. Hybrid analogue-digital simulations afford more flexibility and allow the same range of initial states to be prepared as digital quantum circuits, as demonstrated e.g. in work by Google [42]. However, unlike some analogue simulations which have physical fermions in the simulator itself, hybrid analogue-digital simulations on Google's superconducting hardware are currently restricted to using the underlying bosonic degrees of freedom.

Digital quantum simulation allows circuits that prepare different initial states, and circuits that extract different output observables, to be flexibly programmed in and run. This allows a rich variety of regimes and physics to be simulated, as the range of results discussed in the following section attests to.

PHYSICS SIMULATION RESULTS

The programmability of digital quantum simulations on general-purpose quantum computing hardware, combined with the high gate speed of superconducting qubits, enables iterating over a wide range of physically interesting scenarios with minimal modifications to the circuits.

The doped Fermi-Hubbard model at intermediate-coupling hosts a variety of complex quantum phases and phenomena that may be observed in real materials. Dynamical effects can shed light on the mechanisms underlying these phenomena [43]. We are therefore interested in the non-equilibrium dynamics of the Fermi-Hubbard model from physically relevant initial states that can be prepared with low-depth quantum circuits. In the limit of very large interaction strength and close to half filling, the presence of doubly-occupied sites (doublons) is energetically penalised. At exactly half filling, and $U \gg |J|$, the Hamiltonian (Eq. (1)) is well approximated by the antiferromagnetic Heisenberg Hamiltonian [18]

$$H_{\text{Heis}} = \frac{4J^2}{U} \sum_{\langle i,j \rangle} \left(\mathbf{S}_i \cdot \mathbf{S}_j - \frac{1}{4} \right), \tag{2}$$

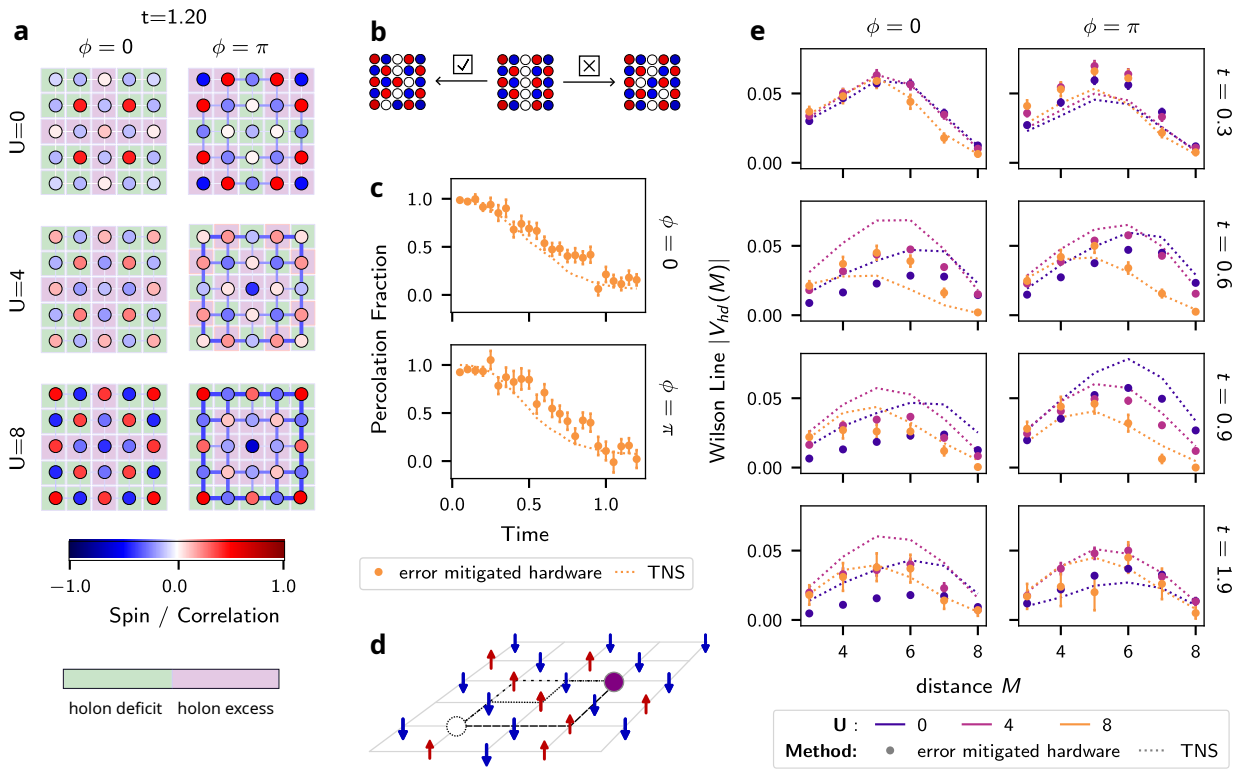


FIG. 2: **Fermi-Hubbard dynamics phenomenology explored at the 5×5 scale. a,** Holon density (squares in the background), local spin (circles in the foreground) and nearest neighbour spin correlations (links). We observe that a localisation of the excess of holon density (purple squares) correlates with the strongest development of nearest-neighbour correlations as the interaction and flux increase. Note that for nonzero flux and interaction, a local spin moment is trapped in the middle, where a lack of neighbour correlation is observed. **b,** Illustration of paths used to compute the percolation fraction. **c,** Error mitigated probability as a function of time of a stripe of holons to connect two opposite ends, computed by: postselecting on shots with less than four doublons and with negative spin-spin correlation across sites. While mean-field approaches [3] assume that the main dynamics of a stripe is to oscillate around its initial position, we observe that the stripe quickly breaks apart, even for the largest interactions considered ($U = 8$). The largest tensor network simulation that we considered ($\chi = 2048$) agrees with this result. **d,** Estimating a Wilson line between a holon (open circle) and a doublon (filled circle): We find a shot with a holon and a doublon at Manhattan distance M, and compute $\sum_{m,n\in\gamma} S^z_n S^z_m$ for consecutive sites along a path γ. We sum over all possible paths of the same length. Here we show three paths for $M = 4$. **e,** Time-snapshots of the Wilson line between a holon and a doublon as a function of their separation, generated by the spin background. We observe that the potential decreases as a function of time and increases as a function of distance up to a separation ~ 6, with the experimental results showing the strongest suppression at zero interaction and flux, as expected for the non-interacting system. The classical simulations seem to converge to a different value. The decrease of the Wilson line for larger distances appears to be a finite-size effect, as going to even larger system sizes, we see a movement of the peak to larger separations (see Appendix B 3).

where $\mathbf{S}_j$ is the spin operator at site j, so low-energy states favour anti-correlation of neighbouring spins. For this reason, we choose the Néel state as one of the initial states, and also an entangled initial state consisting of a singlet covering of the lattice. With these motivations in mind, our experiments are performed at small hole-doping on the $L_x \times L_y$ lattice, where we perturb the initial state by removing particles. We consider interactions $U = 0, 4, 8$ and several initial states with different hole-dopings, ranging from 1 to $\min(L_x, L_y)$. Each initial state has a given number of fermions N_e and total spin $S^z = 0, 1$, depending on the parity of $L_x \times L_y - N_e$. The initial state configuration corresponds to either a Néel background with holons placed at various locations, or a singlet covering with two holons. All these states have an energy density corresponding to a temperature $T \sim J$ [44] (see Appendix A). Finally, for some of these configurations, we study the evolution of the Fermi-Hubbard model with both zero and π flux per plaquette. While the simulation of the Fermi-Hubbard

model with and without magnetic flux differs at the circuit level just by some single-qubit gates, the physics of these two models differs dramatically. At half filling and without interactions, the model with zero flux is a typical metal at low energies, while the π flux system is a Dirac semimetal [45].

Away from half filling, the interplay between spin and charge degrees of freedom in the dynamics of the system is not well understood as the interaction strength increases. We study the formation and dependence on flux of a magnetic polaron [2], a quasiparticle – previously observed in analogue simulations [46–48] – believed to be intimately linked to the mechanism of high-temperature superconductivity in cuprates [49], by evolving a system with initially one holon in the middle of an otherwise Néel-ordered state. As shown in Fig. 2a, correlations between nearest neighbours build up over time in the π flux system, compared with the zero flux case. Around the original position of the holon, the magnetic order is suppressed, and the local spin moment is trapped. A complete discussion of this appears in Appendix B 1.

As the electron density is reduced from half filling, several phases are expected to appear as a function of (hole) doping at low energy in the Fermi-Hubbard model [50]. In particular, for hole-doping $n_h \sim 1/8$, stripe order is expected to occur [51, 52]. We introduce a stripe of holons across the Néel ordered background and investigate the stability of this holon stripe over time. In the high interaction strength regime, the mean-field description of the stripe evolution hinges on the existence of bipartite lattice symmetry [3]. We study how this symmetry breaks dynamically due to the proliferation of doublon-holon pairs. Starting from a stripe of holons, we study the sampled configurations as a function of time, and count the number of shots with a path connecting opposite ends (see Fig. 2 b). To counter the effect of proliferation, we postselect on the experimental shots that have three or fewer doublons (Fig. 2c). There, we also postselect on the shots that have negative spin-spin correlation to counter the effect of the melting of the antiferromagnetic order. We observe that, even within this subset, the bipartite symmetry is broken. This is a consequence of virtual holon-doublon pairs that appear, evolve and disappear in the system. This phenomenon seems insensitive to the flux (zero or π) considered. Several other measures regarding this effect are shown in Appendix B 2.

We also consider an initial entangled state, corresponding to an almost columnar valence-bond solid where the electrons between some edges in the lattice form singlets. This is motivated by the expectation that quantum fluctuations can prevent the development of magnetic long-range order, so $\langle S_i^z \rangle = 0$ [53–55]. These states on the square lattice spontaneously break translational invariance. In our setting, we choose the arrangement of singlets in a way that minimises the circuit cost whilst retaining a columnar order in the bulk. The configuration of singlets is shown in Fig. 1, bottom left, and its circuit construction is discussed in Appendix D 3 b. Trivially, in these states, the spin and charge propagation differ as the charge evolves to the homogeneous state, but the local spin expectation vanishes. We study the evolution of the potential between holons and doublon pairs at a (Manhattan) distance M mediated by the spinons (fermion quasiparticles carrying spin but no charge). This potential appears naturally in a description of the Fermi-Hubbard model in terms of holons, doublons and spinons (see Appendix B 3) in terms of extended objects known as Wilson lines [56] (see Fig. 2 d). As we observe for a 5×5 system in Fig. 2e, the expectation of the potential is roughly independent of the interactions at early times, for both the zero- and π flux systems, separating for longer times into increasing potentials at longer distances, with a smaller potential appearing in the non-interacting regime in the zero flux case. For π flux, we do not see a significant decrease of the potential at longer times, even for zero interactions. The decrease in the magnitude of the potential for the biggest distances can be attributed to finite-size effects, as – going to larger system sizes – the peak of the curve moves to larger distances. Note that these predictions differ from the ones obtained through classical simulation. A full discussion of this appears in Appendix B 3.

As is clear from the experimental results, the system is far from equilibrated. We study the approach to equilibration by considering the wavefunction's spread in the computational basis, using the inverse participation ratio (IPR) on a subsystem. For a fully delocalised state, corresponding to an equal superposition over all possible basis states, the IPR becomes $1/D$ where D is the Hilbert space dimension. On the other hand, for a state with support on a single basis state the IPR takes the value 1. Results on the different subsystems' IPRs that we consider, in terms of charge, spin and occupation, appear in Appendix B 4.

Finally, we explore the evolution of the mean distance between two holons placed in a Néel ordered state. As with the other input states, the dynamics induce holon-doublon pair proliferation. We find that, while the average distance of doublons rapidly converges to the value expected by homogeneous random placement, the distance between holons shows a nontrivial dependence on the interactions. This is discussed in detail in Appendix B 5.

VALIDATION OF THE SIMULATION RESULTS

There is an inherent challenge in validating the output of quantum simulations in regimes that cannot be computed exactly classically. A striking benefit of simulating fermionic systems, such as the Fermi-Hubbard model, is that they naturally afford access to a parameter regime that can be verified classically; but at the same time, changing only trivial parameters in the quantum circuit gives access to computations in a regime that is classically hard to access. Specifically, for $U = 0$ (but

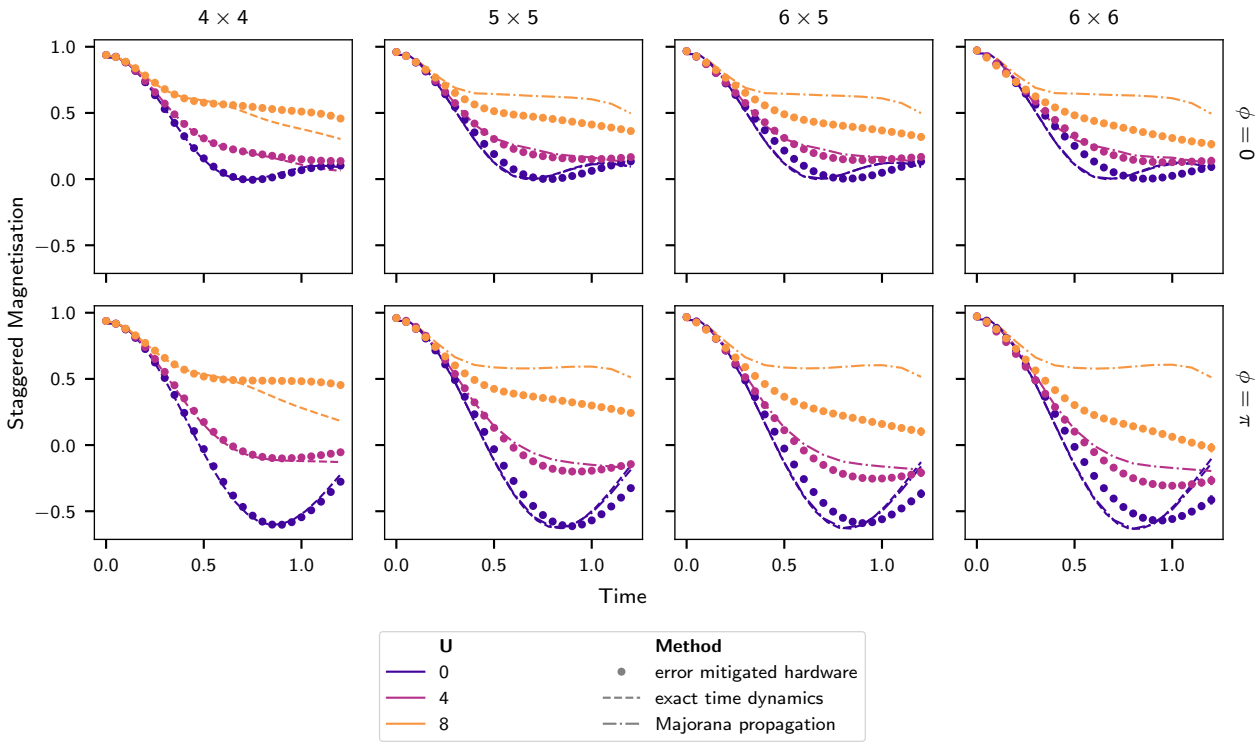


FIG. 3: **Dynamics of the staggered magnetisation at increasing scale.** The staggered magnetisation is $M_s = \frac{1}{L_x L_y}\sum_i (-1)^i S_i^z$ where S_j^z is the spin in the z direction at site j, and L_x, L_y are the system dimensions, ranging from 4×4 to 6×6, with and without a magnetic flux ϕ passing through the plane of the model. Different interaction strengths U are indicated by colour. Error-mitigated hardware data is given by points, with solid curves corresponding to smoothed fits. For the 4×4 case, exact state vector simulations are possible at all values of U, indicated by the dashed lines in that panel. For larger lattice sizes, exact classical simulations computed by FLO are shown with dashed lines for the non-interacting case ($U = 0$), and Majorana propagation simulations are shown with dash-dotted lines. At $U = 0$ the Majorana propagation simulations are expected to be accurate simulations of the ideal quantum circuit output, so differences to the exact time dynamics are indicative of Trotter error. Similar plots illustrating performance with scaling can be found in Appendix L.

not for $U \neq 0$), the exact dynamics on many initial states can be computed efficiently on a classical computer using Fermionic Linear Optics (FLO) [57] for all system sizes. Indeed, for $U = 0$ this method allows us to efficiently compute both the true, continuous-time evolution under the Fermi-Hubbard Hamiltonian, and to exactly simulate the quantum circuit (which only approximates the true Hamiltonian dynamics). Meanwhile, the difference between the time-dynamics simulation circuit with $U = 0$ and the circuit with $U \neq 0$ is just a change to the rotation angles of a set of single-qubit gates. This minor change is not expected to have a large impact on the performance of these gates. It is therefore reasonable to expect that the variation of the interaction strength U in the Fermi-Hubbard simulation does not impact the performance of the underlying digital simulation circuit in a substantial way. Thus, we can build evidence of the correct performance of the quantum computation in the $U = 0$ regime, whilst being able to access the harder-to-verify $U \neq 0$ regime via a minor change to the circuit parameters. (It should be noted, however, that the TFLO error mitigation used to attenuate the systematic components of device noise has been trained on $U = 0$ data, so caution is required in interpreting the experimental fit to the $U = 0$ curves as experimental validation. See Implementation.)

We find that, up to systems of size 6×6, the dynamics of a number of observables can be accurately retrieved by the quantum simulator, validated against FLO. This is illustrated in Fig. 3, which gives an example for one particular observable quantity – the staggered magnetisation. For 4×4 we can also compare against the ground-truth computed by exact state-vector simulation for all values of U. We see that for $U = 0$ and 4, the quantum simulation results match well with the exact calculations. For $U = 8$, however, we already see a deviation from the exact results at late times. As discussed, for larger lattice sizes where state-vector simulation is prohibitive, we can only compare with exact calculations for $U = 0$. Changing U does induce a substantial change in the dynamics, which qualitatively matches what is expected on physi-

cal grounds (see physics results discussion, above). For $U > 0$ we can still compare to Majorana propagation calculations, which are expected to be accurate for the staggered magnetisation as this quantity depends only on 1-body observables. We see that the quantum simulation is in good agreement with Majorana propagation for $U = 0$ and 4, but once again deviates significantly for $U = 8$. The quantum circuit depth is kept identical across all time points and for all U values at a given system size, so the underlying hardware noise is not expected to differ significantly at each fixed size across all the data points shown. But it is likely that the TFLO noise-mitigation method becomes less effective at values of U far from 0, possibly by a change in the impact of circuit errors on the observables, which may account for this effect at $U = 8$. However we note that as an approximate method the correctness of the Majorana propagation calculations cannot be guaranteed away from $U = 0$.

In addition, we use cross-entropy benchmarking for statistical validation of the correctness of the quantum hardware. Our results (Appendix I) indicate that error-mitigated hardware samples achieve a linear XEB fidelity of at least 1% with the ideal state at late evolution times, for a range of initial states and system sizes up to the largest circuits with 72 qubits.

BENCHMARKING AGAINST APPROXIMATE CLASSICAL SIMULATION METHODS

Quantum circuits on 72 qubits are currently well beyond the largest exact state-vector simulations. Nevertheless, approximate classical methods exist that are able to approximately simulate the quantum dynamics of significantly larger systems. In particular, state-of-the-art tensor network methods can accurately approximate the time-evolved state of many-body quantum systems on hundreds of qubits [58, 59], as long as the entanglement does not build up too rapidly as the state evolves. Amongst tensor network methods, those based on matrix product states under the time-dependent variational principle (TDVP) have been found previously to perform particularly well for approximately simulating the exact time evolution of many-body systems [60, 61], and in parallel work targetting the Fermi-Hubbard model specifically [62] we also found that TDVP performed the most reliably. We therefore take TDVP [63] here as the main tensor-network comparator. (See Appendix F and [62] for more extensive discussion and comparison with other tensor network methods.)

Complementing these tensor network approaches, operator propagation methods such as Pauli propagation [64] and Majorana propagation [26] are capable of approximating expectation values of time-evolved observables even when the state is highly entangled, as long as the time-evolved observable can be expressed as a sum of a small number of terms. Majorana propagation performs particularly well in simulating layers of fermionic Trotter dynamics [26]. Therefore, we have also implemented Majorana propagation computations explicitly targeting the Trotterised quantum circuits, which approximate the continuous dynamics of the system. (See Appendix G for more detailed discussion of Majorana propagation computations.)

To benchmark our digital quantum simulations against leading classical methods, we carried out extensive classical simulations using TDVP and (where applicable) Majorana propagation to mirror every configuration of the Fermi-Hubbard model and all data taken on the quantum simulator. These classical computations took over 100 CPU years on the Google Cloud Platform and generated over 25TB of data.

Quantities that are derived from expectation values of small numbers of few-body observables are expected to be reproduced accurately by both Majorana propagation and by tensor network methods. This includes many of the quantities discussed in the physics results above. This is borne out in the results, where we see (e.g. in Figs. 2 and 3) that the classical simulations and error-mitigated quantum simulations are in qualitative – and in many cases quantitative – agreement with one another and with the exact dynamics at $U = 0$. Indeed, because the quantum simulation is subject to hardware noise and errors, whereas the classical approximate computations are not, we see that the classical methods often achieve better agreement with the exact $U = 0$ calculations than the quantum simulation for such quantities.

If significant quantum many-body correlations and entanglement build up in the system over the course of the dynamics, quantities and observables that are sensitive to these become challenging for tensor network and operator propagation methods to reproduce accurately. On selected 4-site (8-qubit) through to 7-site (14-qubit) observables, we find that the tensor network simulations disagree with the exact calculations for $U = 0$, whereas the error-mitigated quantum simulations are in good agreement with the exact results throughout the time evolution. Meanwhile, these observables are too large and spread too quickly for Majorana propagation methods to be practically feasible for the larger observables. However, if we apply our novel training-based error-mitigation method to bit-strings sampled from the tensor-network simulation, we find that this brings the tensor network results into good agreement with the exact $U = 0$ calculations. For $U > 0$, where exact classical calculations are not possible, the tensor network simulation results (both with and without applying error-mitigation to the classical data) do not match the results obtained from the digital quantum simulation. Appendix H contains a more extensive discussion of these and further results comparing quantum simulation and classical methods on many-body quantities.

CONCLUSIONS

In this work, we have presented the first successful digital quantum simulation of the seminal spinful 2D Fermi-Hubbard model on lattice sizes up to 6×6. This is enabled by: highly efficient algorithmic implementation methods optimised for the specific hardware connectivity and target problem, that minimise the quantum circuit depth and gate count; and novel fermionic training-based noise mitigation methods, which leverage the fact that interacting electron models reduce to non-interacting fermion models for particular parameter choices, where exact classical numerical computations are possible. We also use exact calculations (for quantities that are not used to train the error mitigation) to validate that the digital quantum simulator is performing sufficiently accurately to produce physically meaningful results.

The wide range of Hubbard dynamics phenomenology we investigate is also accessible via modern, heuristic classical computational approaches, including tensor network and operator propagation methods. We have benchmarked the quantum simulation against these classical methods by running medium- and large-scale numerical computations of the quantities studied in this work. In some cases, the classical methods do not match the results obtained on the quantum hardware. We do, however, observe that the quality of the quantum simulation gradually degrades with system size and quantum gate count, as expected on pre-fault-tolerant quantum hardware.

These results demonstrate that meaningful, programmable digital quantum simulations of many-body interacting electron models can now be performed on state-of-the-art quantum hardware using state-of-the-art algorithmic techniques, heralding the era of general-purpose digital quantum physics simulation. Further algorithmic and implementation improvements are feasible. In particular, exploiting continuously tunable gate parameters would give a significant additional reduction in circuit depth, and further improvements can be made in particular to the fermionic training-based noise mitigation methods that may enhance their performance. As quantum hardware and algorithmic implementations continue to improve, we expect more complex electronic structure models related to periodic crystalline materials to come within reach.

IMPLEMENTATION

We give here a high-level overview of the implementation of the experiment. This includes: the algorithm and circuits employed for the dynamical simulation of the model on hardware; the error mitigation techniques employed; the classical simulation techniques used to compare against; and the statistical benchmarking techniques used to indicate high-quality performance of the device. A brief discussion of the hardware used is given in Appendix C.

Algorithm and Circuits

The time evolution of the system is approximated by three layers of second-order Trotterised time dynamics (approximately equivalent to 6 layers of first-order Trotterised dynamics). The number of Trotter layers is kept constant for all simulation times, with different times accessed by increasing the Trotter step size. The maximum simulation time attained is $t = 1.2$, with a maximum time step per Trotter layer of 0.4. The comparison of simulated quantum dynamics against ground truth suggests that the Trotter error remains small for $U = 0$ at all times simulated here. This is corroborated by full state-vector simulations of the dynamics of small systems, where we see that the Trotter error of up to weight 4 observables remains low for $U = 0$ and $U = 4$, but becomes significantly more pronounced for $U = 8$ at intermediate- to late-times. Simulations of 4×4 dynamics suggest that for $U = 0$ Trotter error is negligible at all times, for $U = 4$ Trotter error begins to play an important role at approximately $t = 0.9$, and for $U = 8$ Trotter error begins to play an important role at about $t = 0.5$. We discuss this in more detail in Appendix J.

Each spin sector of the fermionic system is represented by a separate Jordan-Wigner (JW) [65] transform arranged in a zig-zag pattern following the square connectivity of the device. The arrangement of the JW transforms is such that spin up and down modes associated with the same site are as close as possible given the constraints of the hardware. The smallest simulation requires 34 qubits, 1220 two-qubit gates and a circuit depth of 198, to simulate the 4×4 system. While the largest simulation requires 74 qubits, 4372 two-qubit gates and a circuit depth of 294 to simulate the 6×6 system – the maximum we could reasonably fit on the hardware devices given layout constraints. We employ a fermionic swap (fswap) network [66, 67], which dynamically rearranges the mode ordering of the JW transform in order to bring vertically adjacent modes next to one another. This amortizes the cost of executing vertical hopping terms across these modes, which would otherwise be much more costly to execute.

Since (2-qubit) gates are a major source of error, and the device has a limited coherence time, improving the circuit's gate count and depth are both critical to reducing noise in the computation. Here we achieve this by adopting a range of circuit optimisation techniques, some specifically tailored to fermionic simulation, including:

- employing the term ordering symmetry of the second order Trotter layer to minimize fermionic swaps;
- merging terms within and between second order Trotter layers;
- identifying opportunities to replace qubit swaps with fermionic swaps, which require fewer 2-qubit gates to implement on the hardware;
- merging fermionic hopping terms with fermionic swaps;
- making a judicious choice of JW layout to ensure minimal cost of spin-spin interactions.

Whilst the impact of each of the above is incremental, the combined effect of the whole set of optimisations leads to a significant reduction in the gate count. A full account of the circuit constructions and costs are given in Appendix D.

Error mitigation

Hardware noise and errors remain a major obstacle to the extraction of scientifically meaningful results from any current digital quantum simulation. In order to achieve meaningful results, it is necessary to combine multiple algorithmic design techniques (described above, and in more detail in Appendix D) with suitable error mitigation methods.

We employ a number of error mitigation techniques to improve our signal dramatically. The most impactful of these are (i) post-selection by occupation sector, (ii) raining with Fermionic Linear Optics (TFLO), (iii) Maximum Entropy Shot-Reweighting (MESR) and (iv) Gaussian process regression (GPR).

Post-selection by occupation sectors exploits the fact that the dynamics of the system conserves particle numbers in both the spin-up and spin-down sectors. Whenever density measurements are performed in all modes, we can therefore verify whether the spin-up and spin-down counts have been preserved. whenever that is not the case, we can know with certainty that there has been a significant error in the computation, and discard that run.

TFLO is a training-based noise mitigation method that exploits the fact that the Fermi-Hubbard model reduces to a free-fermion model for $U = 0$, hence can be classically efficiently simulated using FLO. The quantum circuit structure at the gate-level remains very similar for $U \neq 0$, and so we can expect the underlying hardware noise profile to be similar. The deviation of the noisy

signal from the ground truth at $U = 0$ is used to generate a linear fit on a set of observables corresponding to those being targeted for correction. The linear fit is then employed to apply linear noise correction for observables measured at non-zero values of U. We discuss this method in more detail in Appendix E 3. We note here that a critical motivating factor in deploying a fixed number of Trotter layers for all time steps, despite fewer layers in principle being required at low times, is to achieve consistency in training data for TFLO. An added benefit of this is that the contribution of device noise to the signal is largely decorrelated from the time parameter of the dynamics, as the 2-qubit circuit depth and gate count remain fixed across all data points.

For a number of signals, we are interested in retrieving statistics over patterns found in shots. We introduce and employ in this work a Maximum-Entropy Shot-Reweighting (MESR) technique which mirrors the power of TFLO in a shot-based setting. This technique learns a set of Lagrange parameters which skew the distribution of shots, produced by $U = 0$ dynamics, towards a distribution for which expectation values of local observables more closely match the FLO ground-truth and TFLO mitigated observables. This trained skewing process is then applied to the shots produced by $U \neq 0$ dynamics, before computing the output quantity of interest. This is discussed in more detail in Appendix E 10.

Unless otherwise stated, "error-mitigated" in this paper means mitigated by TFLO or MESR.

Notably, all of the above methods are bespoke to interacting fermion models, such as the Fermi-Hubbard models we are simulating here, targeting key specific properties of the system. This makes them particularly effective in mitigating errors for the simulations of interest.

These experiments target dynamical properties of the system, and aim to produce observables that vary smoothly with time. We can leverage this assumption of continuity for error mitigation via Gaussian process regression (GPR). GPR is a standard method for generating smooth curves from data points under an assumption of continuity and a maximum rate of change. We discuss this method in more detail in Appendix E 7.

Additional error mitigation methods we employ include Pauli- and readout-twirling, dynamical decoupling, and symmetrising over the symmetries of the Hamiltonian, circuit and initial states. All are discussed in more detail in Appendix E.

The set of techniques we have described here, and the details of their implementation, were settled upon after an exhaustive iterative process of comparisons against alternative implementations and techniques. We do not include the full account of these here. However, we note that we did compare our methods to another prominent error mitigation technique called probabilistic error amplification (PEA) with zero noise extrapolation (ZNE), and found TFLO to perform more favourably in our setting. We include a discussion of this in Appendix E 11.

One concern regarding employing TFLO and MESR as error mitigation techniques is that we risk confounding our ability to verify the performance of our simulations when comparing against $U = 0$ dynamics. This is because the signal is trained on the behaviour of $U = 0$, and although the aim is to use this to learn the generic impact of hardware noise on our signals, we may instead simply learn the specific impact that the combination of hardware noise and algorithmic errors has on shifting the $U = 0$ signal. Although we can not guarantee that this effect plays no role in the performance of our $U = 0$ signal, there are a number of factors that allow us, with care, to still validate the simulation performance on $U = 0$ results.

First, it is important to dispel the impression that TFLO is simply taking the curves and directly shifting them to align with $U = 0$. The linear fit training data is taken over all times, and the varying signals we observe require significantly more parameters to specify than are present in the linear fit. Second, in the case of a 4×4 system (see Fig. 64 in Appendix L), where the ground truth is available via exact state-vector simulations, we see that TFLO performs well for all values of U, even when the behaviour of the signal is dramatically different from that at $U = 0$. This suggests that the dominant effect of TFLO is not to retrieve a $U = 0$ signal, but rather to undo the effect of hardware noise. Finally, TFLO is trained on few-body observables. Where we consider many-body observables with support on a large number of qubits, for example in the case of the optimised observables shown in Fig. 50 and other many-body quantities not derivable directly from expectation values of observables (discussed in Appendix H), we use MESR with FLO or TFLO training data. This performs well, despite these many-body quantities being in principle independent of the 1- and 2-body observables being trained on.

Classical simulation

Some of the most performant of the tensor network methods targeting dynamical simulation include time-evolving matrix product states via the time-evolving block decimation [68] or the time-dependent variational principle (TDVP) [69] algorithms, time-dependent density matrix renormalisation group (DMRG) for fermionic matrix product states (fMPS) [70], and recent circuit-model DMRG simulators [25, 71] and projected entangled pair states (PEPS) algorithms [59, 72, 73].

In parallel work [62], we have tested multiple tensor network techniques: direct tensor contraction; time-evolving matrix product states via the time-dependent variational principle (TDVP) method; and fermionic matrix product states. These tests indicate that TDVP performs most reliably, and so for our main comparisons we include TVDP simulations of the dynamics of the system. Importantly, TDVP simulates continuous time dynamics, not circuit dynamics directly. It does this by iteratively evolving the Schrödinger equation for short time steps,

and at each step projecting the state onto the subset of MPS states of a given bond dimension. We perform simulations up to a bond dimension of 2048 with time steps of $\Delta t = 0.01$, and also perform bond-dimension extrapolation analysis in some cases. See Appendix F for a more detailed description of these simulations.

Majorana propagation (MP) is an approximate simulation method wherein the target observables are evolved under the circuit dynamics in the Heisenberg picture, and tracked as an operator expansion in Majorana monomials (products of individual majorana fermion operators). Under this evolution, the operator expansion can balloon into an exponential number of monomials, which must be truncated in order to remain efficient. The truncation criteria employed here are: number of Majoranas in the monomial, fixed at 10; and coefficient magnitude, fixed at 3×10^{-5}. We have found Majorana propagation to be a particularly performant method for simulating multiple layers of fermionic Trotter dynamics. We employ this method throughout to approximate the exact circuit dynamics. See Appendix G for a more detailed description of our simulations.

Validation and benchmarking against exact and approximate classical methods

In Appendix H we look at the structure of many-body correlations in local measurement outcomes for time-evolved states, which are expected to be challenging for tensor networks to track throughout the time-evolution. The total variation distance between the probability distribution over local measurement outcomes on the quantum hardware, and the probability distribution over outcomes of the same measurements predicted by classical methods, is the most fundamental of such quantities. It is not practical to compute this for very large subsets of qubits, because each k-qubit subset requires 2^k probabilities to be computed, and there are $\binom{n}{k}$ subsets. However, the total variation distance for all 5-site subsets (10 qubits each) can be calculated with reasonable computational effort, allowing these to be compared. This is only practically feasible because the high clock-speed of superconducting circuit quantum hardware allows large numbers of samples to be collected.

As any observable lower-bounds the total variation distance between the quantum state produced on the hardware and the classical approximation thereof, via differences in its expectation value measured on the quantum simulator and computed via approximate classical methods, we can also access divergences between the classical predictions and the quantum simulation results by optimising the difference over many-body observables. As we are computing simpler quantities – expectation values of observables for a fixed subset of sites, rather than total variation distance for all subsets of a certain size – this is feasible up to somewhat larger 7-site (14-qubit) observables in our experiments. In optimising over observables, however, it is important to avoid a spurious and statistically insignificant difference through optimising over a larger – or even comparable – number of parameters than the number of bit-strings sampled on the quantum simulation. As we are able to take large numbers of samples over many time-points in the quantum simulation of the Fermi-Hubbard dynamics, thanks to the high clock-speed of superconducting quantum circuit hardware, the 16,384 parameters being optimised over in the largest 14-qubit computational-bases observables we consider, remains significantly smaller than the total number of samples in the datasets it is being applied to, which is on the order of 150,000 in each data set even after post-selection (see error mitigation, above). Appendix H contains a more detailed discussion of how we perform the optimisation over observables and of the results.

In Appendix I we adapt the method of linear XEB benchmarking [74–78] to the setting of FLO circuits, and compare both unmitigated and MESR-mitigated hardware samples against a noiseless, untrotterised FLO simulator, to provide efficient statistical evidence of the global overlap of the ideal state with the output of the $U = 0$ digital simulation of the Fermi-Hubbard dynamics. In contrast with applications of linear XEB to random quantum circuits, our verification protocol utilizes samples generated in physically meaningful FLO circuits, allowing for efficient usage of hardware samples and end-to-end benchmarking of the digital simulation against the ideal target physical state.

ACKNOWLEDGEMENTS

The hardware data presented in this work was taken remotely on 105-qubit and 72-qubit processors, with access provided via Google's Quantum Engine. Calibration and support were provided by the Quantum Hardware Residency Program. We thank the Google Quantum AI team for providing the quantum systems and support that enabled these results. The views expressed in this work are solely those of the authors and do not reflect the policy of Google or the Google Quantum AI team. We thank Eliott Rosenberg for helpful comments and the entire Google team for collaborating on this work. Thanks to Dmitry Abanin for discussions on Fermi-Hubbard physics, and Zlatko Minev for advice on PEA. We thank Mohammed Alghadeer, Jaehong Choi, Tyler Cochran, Norhan Eassa, Gaurav Gyawali, Shashwat Kumar, and Ricky Oliver for hardware calibrations that enabled the data collection throughout this project. We also thank Andrew Childs for discussions on time-dynamic simulations. This work received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 817581).

† Phasecraft

Faisal Alam[1], Jan Lukas Bosse[1], Ieva Čepaitė[1], Adrian Chapman[2], Laura Clinton[1], Marcos Crichigno[2], Elizabeth Crosson[2], Toby Cubitt[1,3], Charles Derby[1], Oliver Dowinton[1], Paul K. Faehrmann[1,4], Steve Flammia[2,5], Brian Flynn[1], Filippo Maria Gambetta[1], Raúl García-Patrón[1,6], Max Hunter-Gordon[1], Glenn Jones[1], Abhishek Khedkar[1], Joel Klassen[1], Michael Kreshchuk[2], Edward Harry McMullan[1], Lana Mineh[1], Ashley Montanaro[1,7], Caterina Mora[1], John J. L. Morton[1,8], Dhrumil Patel[2,5], Pete Rolph[1], Raul A. Santos[1], James R. Seddon[1], Evan Sheridan[1], Wilfrid Somogyi[1], Marika Svensson[1], Niam Vaishnav[1], Sabrina Yue Wang[1], Gethin Wright[1]

[1] Phasecraft Ltd, London, UK
[2] Phasecraft Inc, Washington DC, USA
[3] Department of Computer Science, University College London, UK
[4] Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, 14195 Berlin, Germany
[5] Department of Computer Science, Virginia Tech, USA
[6] School of Informatics, QSL, University of Edinburgh, UK
[7] University of Bristol, UK
[8] Department of Electrical and Electronic Engineering, UCL, London, UK

Author contributions– JK, RAS, TC conceived the project and provided overall leadership and oversight. JK, MHG, CD, JLB, EHM, RAS, NV designed the quantum circuits. MHG, EHM, JLB implemented and executed the circuits on hardware. MHG, EHM, JLB, LC, JRS, JK, AM, TC, SYW, EC developed and applied the error mitigation techniques. BF, SYW, JK, GW, RAS, FMG, AK, PR, LM performed the tensor network simulations. SYW, JLB, BF, AM performed the FLO simulations. JLB implemented the Majorana propagation. LC, JK performed the Trotter Error analysis. EC, TC, SYW, MHG, MS, BF, NV performed the XEB analysis. RAS, SYW, FMG, MK, MHG, JK, TC, EC performed the physics analysis. MHG, BF, EHM, JLB, PR, GW developed the underpinning methods and infrastructure used for data analysis. All authors contributed to the writing and revision of the manuscript and the Supplementary Information.

TECHNICAL APPENDICES

Appendix A: Fermi-Hubbard physics at zero and π flux

In this work, we study the time evolution of several states under the Fermi-Hubbard Hamiltonian

$$H = -J \sum_{\langle i,j\rangle \in \mathcal{L},\sigma} (e^{i\phi_{ij}} c^\dagger_{i,\sigma} c_{j,\sigma} + \text{h.c.}) + U \sum_{i\in\mathcal{L}} n_{i,\uparrow} n_{i,\downarrow}. \tag{A1}$$

Here, the first term represents the hopping (with amplitude $-J$) of electrons on a lattice $\mathcal{L}$ between nearest neighbour sites, while the second represents the (repulsive) interaction between electrons occupying the same position. The fermionic operators $c^\dagger_{i,\sigma}$ ($c_{i,\sigma}$) create (destroy) an electron at lattice site $i = (i_x, i_y)$ of spin $\sigma = \{\uparrow, \downarrow\}$ and satisfy the usual anticommutation relations $\{c_{i,\sigma}, c^\dagger_{j,\sigma'}\} = \delta_{ij}\delta_{\sigma\sigma'}$. The operator $n_{j,\sigma} = c^\dagger_{i,\sigma} c_{i,\sigma}$ is the local electron density at site i of spin σ. The Peierls phases ϕ_{ij} encode a magnetic flux ϕ in a direction perpendicular to the sample by enforcing that the sum of Peierls phases around each individual square face (plaquette) of the lattice add up to ϕ. In our experimental setup, we consider rectangular lattices of $L_x \times L_y$ sites with open boundary conditions.

This system has several conserved quantities, which we employ with varying degrees to improve the experimental signal. The most useful conserved quantity is total number and spin conservation: The transformation $c_{j,\sigma} \to e^{i\alpha_\sigma} c_{j,\sigma}$ leaves the Hamiltonian invariant. The conserved quantities associated with this $U(1) \times U(1)$ symmetry are total number $N = N_\uparrow + N_\downarrow$ and total spin $S^z = N_\uparrow - N_\downarrow$, where $N_\sigma = \sum_i n_{i,\sigma}$.

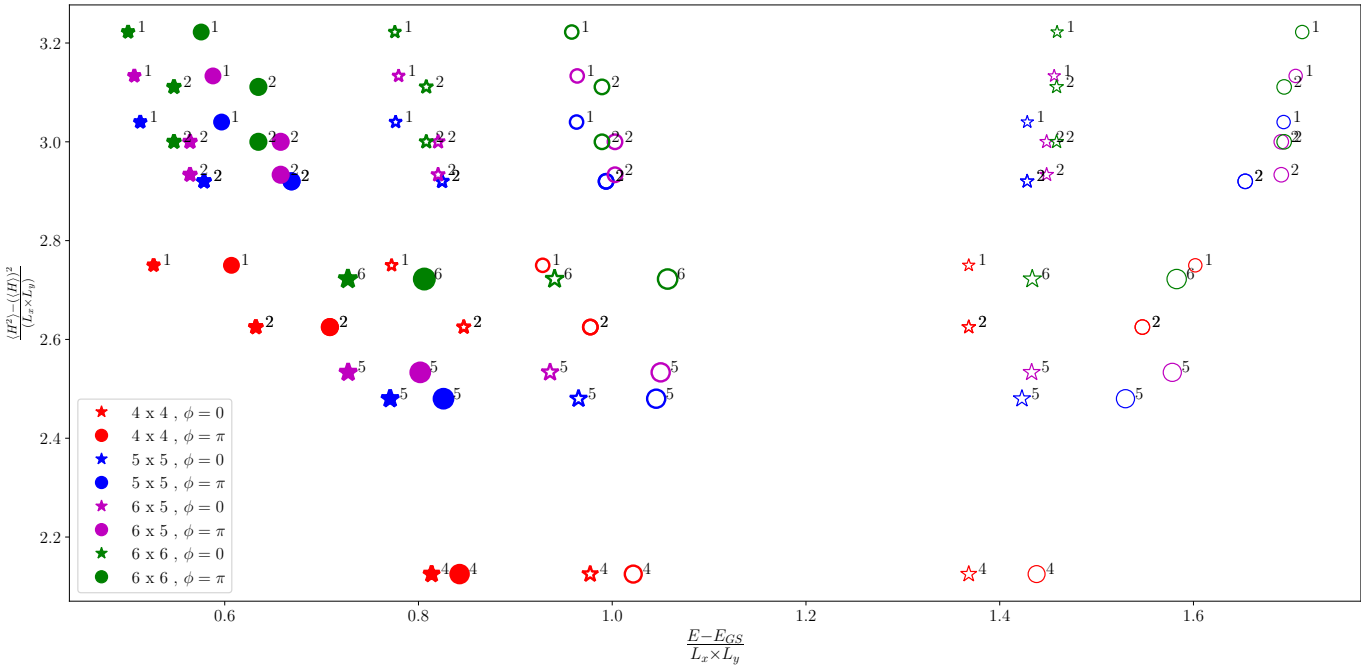


FIG. 4: Landscape of the systems studied in this work. Here we plot the energy of an initial state measured with respect to the ground state energy (obtained with DMRG) in the same number of particle, flux sector and its energy variance, both normalised by the system's size. Different colours represent different sizes. The numbers close to each marker indicate the number of holes in the system. $U = 0$ ($U = 4$) systems are shown with thin (thick) lines. Solid markers represent $U = 8$. Circles correspond to systems without magnetic flux, while stars represent systems with flux.

We are interested in understanding the quenched dynamics of physically relevant low-energy states that can be prepared with minimal circuit cost. In the limit of very large interactions and near half filling, the presence of doubly occupied sites (doublons) is energetically penalised. At exactly half filling, and $U \gg |J|$, the low-energy manifold of the Hamiltonian (Eq. (A1)) is well approximated (in the sense that corrections are of order $(|J|/U)^2$) by the antiferromagnetic Heisenberg Hamiltonian

$$H_{\text{Heis}} = \frac{4J^2}{U} \sum_{\langle i,j\rangle} \left(\mathbf{S}_i \cdot \mathbf{S}_j - \frac{1}{4} \right) \tag{A2}$$

where $\mathbf{S}_j$ is the spin operator at site j, so low-energy states favour anti-correlation of neighboring spins. For this reason, we choose the Néel state as one of the initial states. With these motivations in mind, our experiments are performed at different levels of small hole-doping, where we perturb the initial state by removing particles. We consider interactions $U = 0, 4, 8$ in units of the hopping strength J, which we set to one. We consider several initial states with different hole-dopings, ranging from one to $\min(L_x, L_y)$. Each initial state has a given number of electrons N_e and total spin $S^z = 0, 1$, depending on the parity of $L_x \times L_y - N_e$. The initial state configuration corresponds to either a Néel background with the some holons placed at various locations, or a singlet covering with two holons. A singlet across two sites i and j is defined as

$$|\text{singlet}\rangle = \frac{c^\dagger_{i,\uparrow} c^\dagger_{j,\downarrow} - c^\dagger_{i,\downarrow} c^\dagger_{j,\uparrow}}{\sqrt{2}} |\Omega\rangle \tag{A3}$$

where $|\Omega\rangle$ is the vacuum state, and a singlet covering is simply the tensor product over all singlet pairs, $\otimes_{s=1}^{N_{\text{singlets}}} |\text{singlet}_s\rangle$. In Fig. 4, we show the variance and energy (with respect to the ground state of the corresponding N_e, S^z sector) per site of each of the states considered. Using the energy density of these states, we can assign a corresponding temperature, understood as the temperature of a Gibbs state with the same energy, using the equation of state of the model [44]. All these state have an energy density of the order of the hopping J. We observe here a manifestation of Lieb's π flux phase theorem [79], namely that, for all configurations of fixed system size, interaction, and filling, the one that attains the lowest energy is the π flux system. For the states with two holons, the singlet states have either strictly smaller variance (for sizes 6×5 and 6×6) or equal energy variance (sizes 4×4 and 5×5) compared with the corresponding Néel ordered state.

While the simulation of the Fermi-Hubbard model with and without magnetic flux differs at the circuit level just in some single-qubit gates, the physics of these two models is completely different at low energies. As an illustration, we show the band dispersion of the Fermi-Hubbard model at zero and π flux in Fig. 5 (left and right).

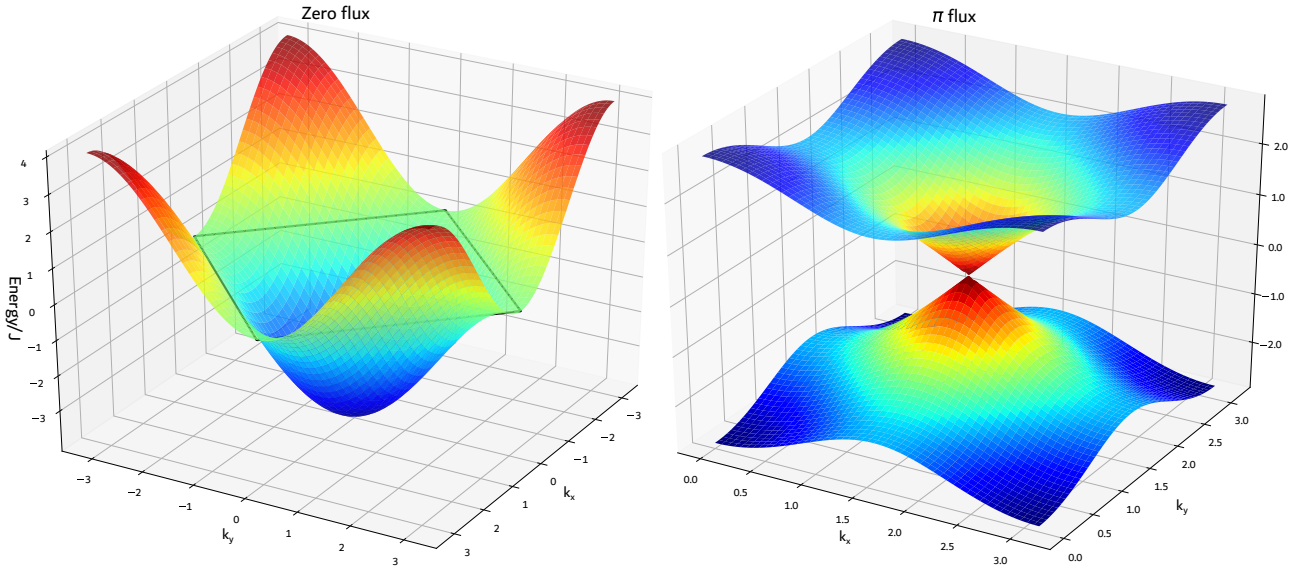


FIG. 5: Single particle band in the zero flux (left) and π flux regime (right, only one Dirac point shown) at $U = 0$ for the Fermi-Hubbard model. The dispersion relation for the zero flux case is given by $E(k)_{\phi=0} = -2\cos(k_x) - 2\cos(k_y)$, while for the π flux case it is $E(k)_{\phi=\pi} = \pm 2\sqrt{\cos^2(k_x) + \cos^2(k_y)}$, for wavenumbers k_x and k_y. At half filling, the system with no flux has a square Fermi surface, while the π flux system has only four points. This state is also known as a Dirac semimetal.

At zero interaction, the ground state of the Fermi-Hubbard model with and without flux corresponds to two copies of spinless electrons filling the single-particle bands. For zero flux, the system has a cosine band with a minimum around zero momentum. At half filling, the Fermi surface is a square and the system (square contour in Fig. 5 left) is very unstable to interactions as Umklapp processes are allowed by momentum conservation. At π flux the unit cell is enlarged and the system becomes a Dirac semimetal, violating Luttinger's theorem [80]. At half filling, the Fermi surface becomes 4 points, where the dispersion relation around each has a conical singularity (see Fig. 5 right) and at low energy is described by two-dimensional Dirac fermions.

Appendix B: Experimental Results

In this section, we collate the experimental results obtained. Unless otherwise specified, all results were obtained using error mitigation as described in Appendix E.

1. Dynamics of a single holon

At exactly half filling, the ground state has antiferromagnetic correlations. A holon moving around flips the spins in its path, creating aligned neighbours that break the Néel order pattern. The energy cost of this makes the movement of the holon different from a freely propagating one. In this section, we study the movement of charge and spin on an initial state where a single holon is placed in the center on an otherwise Néel background.

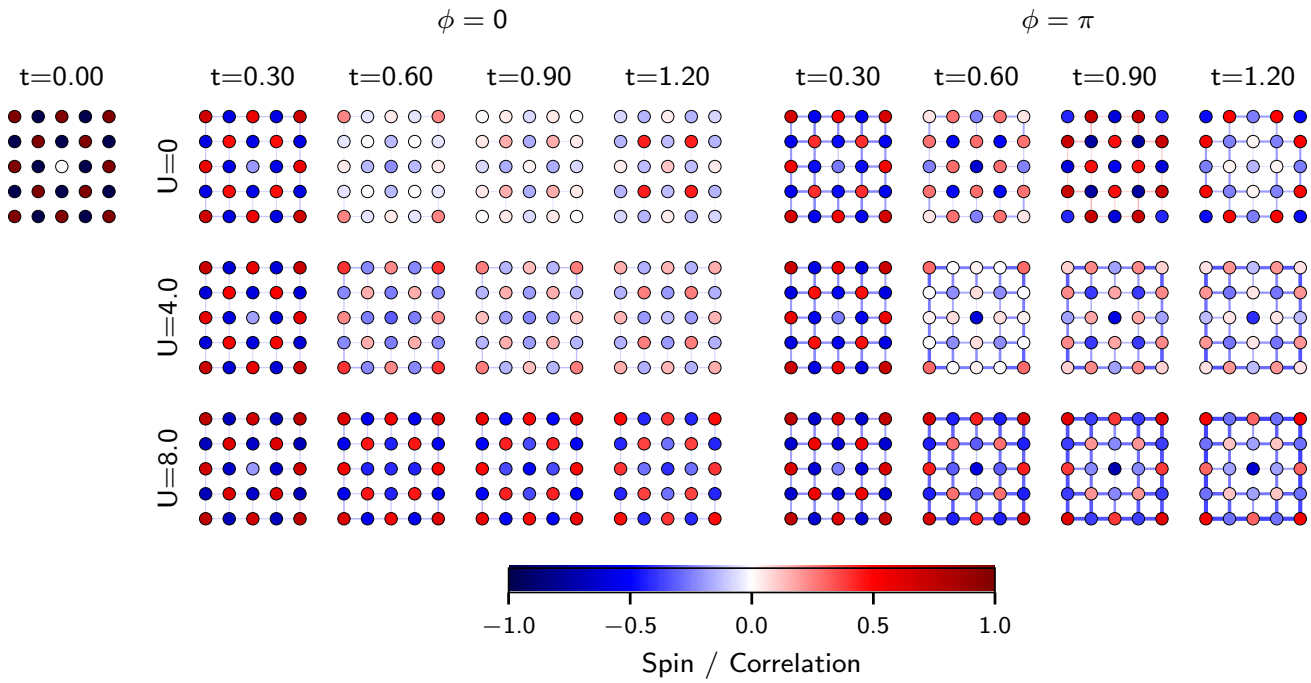


FIG. 6: Spin dynamics for $\phi = 0$ (left insert) and $\phi = \pi$ (right insert) for the 5×5 lattice with a central hole. Showing the evolution of spins B1 at each site (circles) and nearest neighbour spin correlators B2 (links) in a 5×5 system. Obtained from TFLO mitigated experimental results.

First, we show the dynamics of the error-mitigated local spin

$$\langle S_i^z(t)\rangle = \langle n_{i\uparrow}(t) - n_{i\downarrow}(t)\rangle, \tag{B1}$$

and nearest neighbour spin correlation

$$C_{i,e}(t) := \langle S_i^z(t) S_{i+e}^z(t)\rangle, \tag{B2}$$

where e is a primitive lattice vector. We observe that for zero flux, the spin dynamics slows down as the interaction increases, and the original Néel order gradually melts towards the uniform local spin zero state at a rate that decreases as the interaction increases. For non-zero flux, the reduction of the magnetic correlations slows down with increasing interactions, as in the case of zero flux, but some striking differences appear. At zero interaction, the dynamics is fast and we see a revival of the spin density with time, which is expected from the exact $U = 0$ evolution. As the interaction increases, we see two phenomena occurring simultaneously, the melting of the Néel order and the increase of antiferromagnetic correlations between neighbours. For $U = 0$, note that the Néel order flips around $t = 0.9$, with the roles of spin up and spin down interchanged with respect to the original configuration. Antiferromagnetic order between sites that have the same sign of the spin density appears around the edge. At the position where the

holon was initially located, we see a suppression of the magnetic correlations between neighbours. We can investigate how this signal correlates with the presence of a holon at the same place and time where the magnetic distortion is observed. A straightforward measure of the local holon density $\langle h_i(t)\rangle = \langle (1-n_{i\uparrow})(1-n_{i\downarrow})\rangle$ is not sufficient, as the number of holon-doublon pairs increases in time (see e.g. Fig. 9 left most panel). We measure instead the deviation from the average holon density

$$\langle h_i(t)\rangle_{\text{dev}} := \langle (1-n_{i\downarrow}(t))(1-n_{i\uparrow}(t))\rangle - \frac{1}{L_x L_y}\sum_j \langle (1-n_{j\downarrow}(t))(1-n_{j\uparrow}(t))\rangle, \tag{B3}$$

The resulting signal is shown in Fig. 7. At early times (up to $t \sim 0.3$), the evolution is essentially ballistic and independent of interactions and flux, which is a result that also appears in Fig. 9 (left and center left). For both zero and nonzero flux, the effect of interactions in the holon density deviation $\langle h_i(t)\rangle_{\text{dev}}$ is not substantial. The effect of flux, on the other hand, is more profound. For π flux, we observe that the approach to the average uniform value is faster than for the zero flux case. Precisely in the region where there is an excess of holon density, we see a perturbation of the nearest neighbour correlation. This signals the appearance of a magnetic polaron, which has been observed in the dynamics of the Fermi-Hubbard model in analogue experiments [46].

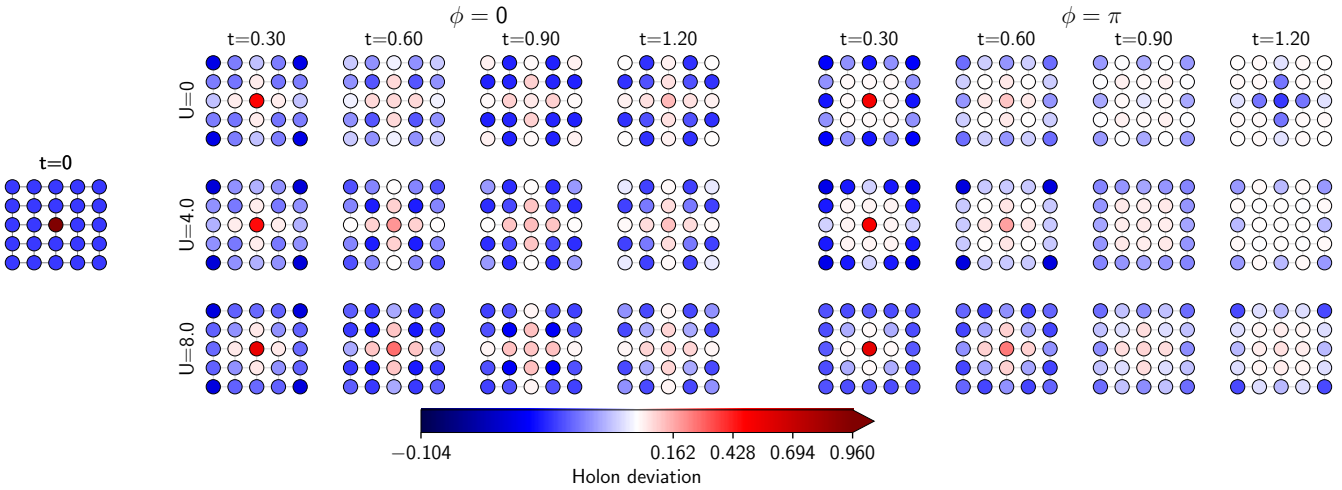


FIG. 7: Deviation from the average holon density in Eq. (B3) as a function of time for $\phi = 0$ (left insert) and $\phi = \pi$ (right insert) for the 5×5 lattice with a central hole. Obtained from TFLO-mitigated experimental results (for more details see Appendix E). Note the nonlinear colour scale, chosen for better contrast.

Finally, we look at some global observables for the evolution of this initial state. An observable that we will use to assess global properties of all the initial states presented is the average doublon density, n_d, defined as

$$n_d = \frac{1}{L_x L_y}\left\langle \sum_j n_{j\uparrow} n_{j\downarrow} \right\rangle, \tag{B4}$$

which is shown in Fig. 9 (left). Note that the number of holons minus the number of doublons is conserved, as they are created in pairs. Increasing the (repulsive) interaction, the total doublon number decreases as generating doublons is energetically more costly, which translates into a faster time oscillation of the components in the evolution of the initial state with more doublons, decreasing their contribution. To quantify the delocalisation of the original holon, we compute the root-mean-squared holon observable, defined as

$$\text{holon}_{\text{RMS}} := \sum_i |i|^2 \langle (1-n_{i\uparrow})(1-n_{i\downarrow})\rangle, \tag{B5}$$

where the (Euclidean) distances are measured with respect to the initial position of the holon at $t = 0$. This signal measures the spread of the holon's size in time. As mentioned before, we see (Fig. 9 center left) that the dynamics of the holon is universal (i.e U independent) and ballistic for small times, as expected from the Heisenberg equation of motion $\dot{n}_d = i\langle [H, n_d]\rangle = i\langle [H_0, n_d]\rangle$ where the derivative at $t = 0$ only depends on the quadratic part of the Hamiltonian H_0 as witnessed by the doublon number and RMS holon observables.

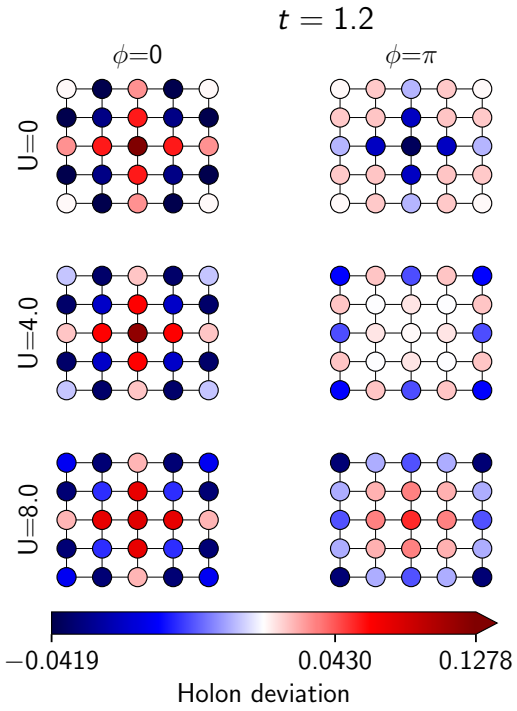


FIG. 8: Deviation from the average holon density in Eq. (B3) for $\phi = 0$ (left column) and $\phi = \pi$ (right column) for the 5×5 lattice with a central hole at the last time point, isolated for better contrast. These results are obtained from TFLO-mitigated experimental results (for more details, see Appendix E).

The staggered magnetisation, shown in Fig. 9 (center right) is defined as

$$M_s = \frac{1}{L_x L_y} \sum_i (-1)^i S_i^z, \tag{B6}$$

and together with the staggered spin-spin correlation Fig. 9 (right)

$$(S^z S^z)_{\text{stag}} := \frac{1}{(L_x L_y)^2} \sum_{i,j} (-1)^{|i-j|} \langle S_i^z S_j^z \rangle, \tag{B7}$$

quantifies the decay of the initial Néel order. Looking at these global observables, we see that the effect of the flux is to promote a faster decay in general and a revival of the magnetic order in the non-interacting case.

We assess the quality of the quantum simulation using several methods. In Fig. 9 we include exact dynamics and exact circuit simulation for $U = 0$ to gain some understanding of the Trotter error of the simulation (a thorough analysis of Trotter errors appears in Appendix J). For all the parameters considered (flux and interaction) we also perform tensor network simulations based on matrix product states, evolving the system using the time dependent variational principle (TDVP). We scale the bond dimension χ in these and all results shown below from 256 to 2048 (for more details see Appendix F). Another classical simulation technique that we explore to benchmark our results is Majorana propagation [26] (more details in Appendix G). We observe that the mitigated experimental results in general agree qualitatively with the classical benchmarks.

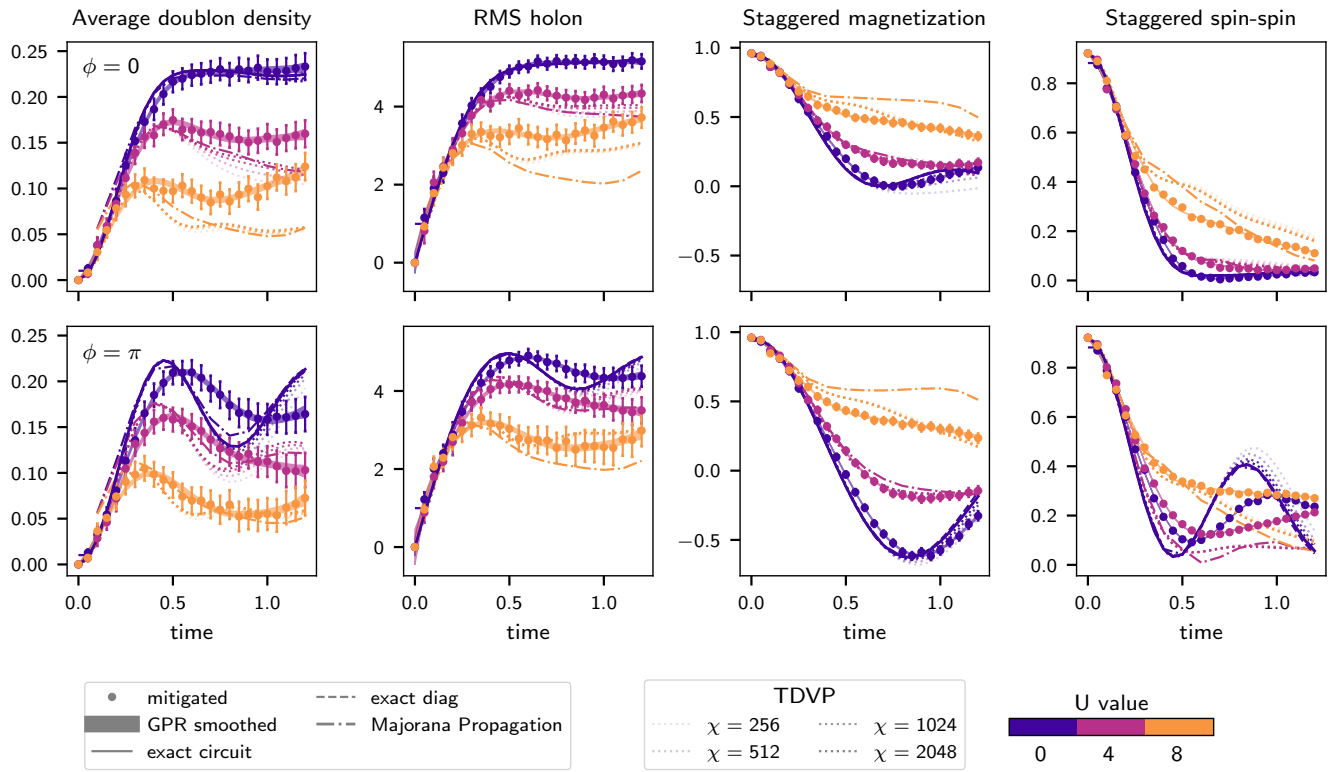


FIG. 9: Average doublon density B4, RMS holon B5, staggered magnetisation B6 and staggered spin-spin B7 as a function of time for $\phi = 0$ (top row) and $\phi = \pi$ (bottom row) for the 5×5 lattice with a central hole. Using 62 qubits and 3172 two-qubit gates. Here we also plot the results obtained using tensor network simulations at different bond dimensions, and exact circuit and TDS evolution when possible ($U = 0$).

2. Evolution of holon stripe

As the electron density is reduced from half filling, several orders are expected to appear as a function of (hole) doping at low energy in the Fermi-Hubbard model [50]. In particular, for hole-doping $n_h \sim 1/8$, [51, 52] stripe order is expected to appear. At high interaction, disturbing the Néel background is energetically costly and the main mode that a stripe can evolve is by fluctuating around its position. The mean-field approach to this dynamics has been studied in [3], where the concept of bipartite symmetry of the spin background plays a central role. In this approach, the magnetic order is studied once the holes have been removed, using the concept of squeezed space [81]. At finite interactions, different mechanisms can disrupt the stripe dynamics. One is the proliferation of doublon-holon pairs, that change the number of available holons and make the percolation of a stripe more probable. Another effect is the fluctuation of the magnetic order, that obscures the bipartite lattice symmetry order considered in [3]. In this section, we study the fluctuations away from the mean-field dynamics of the stripe.

We analyse the dynamics of a stripe by evolving an initial state consisting of a Néel ordered background with a stripe of holons in the middle running from top to bottom as shown schematically in Fig. 10 (leftmost panel) for a 5×5 system. We do this in two different regimes, evolving with the Fermi Hubbard Hamiltonian with zero and π flux per plaquette. Note that the Néel pattern that we choose in the initial state is such that in the squeezed space, where we remove the holons, the system is truly bipartite.

In Fig. 10 we show the (experimental, error mitigated) average spin density Eq. (B1) and the nearest-neighbour spin correlation Eq. (B2) The effect of interactions is radically different for zero and π flux. At zero flux, the initial configuration melts more slowly as the interaction increases, with the spin density tending to the uniform zero magnetisation value. For the Dirac semimetal regime $\phi = \pi$, the situation is somewhat different. For zero interaction, we see a fast melting and revival of the antiferromagnetic order, with the holon stripe reappearing. For increasing interaction, a slowdown of the dynamics is apparent, accompanied by an increase in the neighbour spin anti-correlation as time evolves. To complement this qualitative description, we study global properties that quantify these effects. Together with the error-mitigated experimental signals, we show the corresponding results obtained using classical

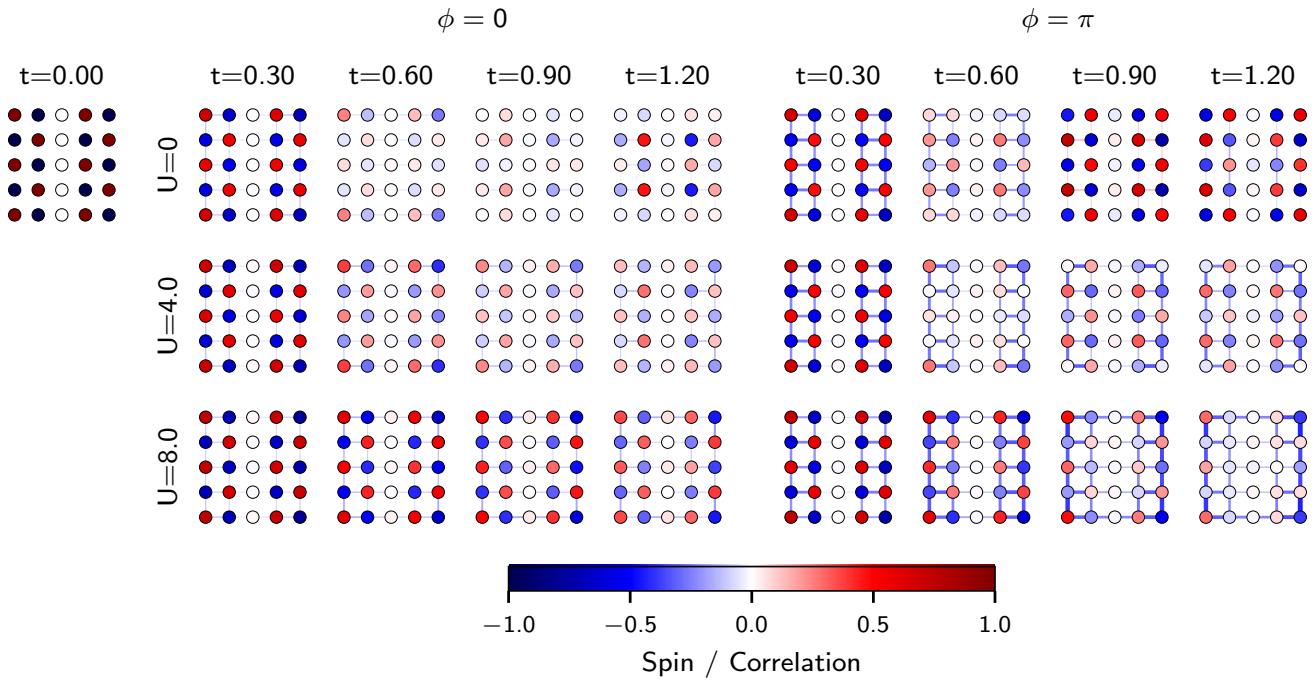


FIG. 10: Spin dynamics for $\phi = 0$ (left insert) and $\phi = \pi$ (right insert) for the 5×5 lattice with a central holon stripe. This displays the evolution of spins defined in Eq. (B1) at each site (circular nodes) and nearest neighbour spin correlators defined in Eq. (B2) (whereby the magnitude of the negative correlation is denoted in the links' thickness) in a 5×5 system. Obtained from TFLO mitigated experimental results. Observe the enhancement of spin correlations in the Dirac semimetal regime ($\phi = \pi$).

tensor networks simulations with the TVDP algorithm (see Appendix F for more details) and exact simulation for zero interaction. In Fig. 11 (left), we plot the evolution of the average density of doublons in the system for $\phi = 0, \pi$. The proliferation of doublons (and holons) depends strongly on the interaction, as expected. The asymptotic value of the doublon density approaches its maximum value

$$n_d(U = 0, t \to \infty) = \frac{1}{4}\left(\frac{L_x - 1}{L_x}\right)^2, \tag{B8}$$

at $U = 0$ and zero flux, where L_x is the side length of the system.

Note that this trend also appears for non-zero flux, but the experimental data becomes less reliable in this scenario as the error bars are amplified.

In Fig. 11 (center left), we show the dynamics of the root-mean-square (RMS) holon density. This observable is defined as

$$\text{stripe}_{\text{RMS}} = \sum_i (i_x - i_{x_0})^2 \langle n_i^h \rangle, \tag{B9}$$

where we sum over the distance perpendicular to the initial stripe located at x_0. It measures the spread of the initial holon stripe in time. Here we see that interactions reduce the spread, in line with the general expectation that interactions reduce the effective velocity of propagation of mobile quasiparticles. Including π flux, the trend remains. The role of magnetisation is captured by the stripe staggered magnetisation $M_{\text{stag}}^{\text{stripe}}$ (Fig. 11, center right) and the stripe staggered spin-spin $(S^z S^z)_{\text{stag}}$ (Fig. 11, right) observables, defined respectively as

$$M_{\text{stag}}^{\text{stripe}} := \frac{1}{L_x L_y} \sum_i (-1)^{i_x + i_y} \text{sign}(i_x - i_{x_0}) \langle S_z^i \rangle \tag{B10}$$

$$(S^z S^z)_{\text{stag}} := \frac{1}{(L_x L_y)^2} \sum_{i,j} (-1)^{|i-j|} \langle S_i^z S_j^z \rangle \tag{B11}$$

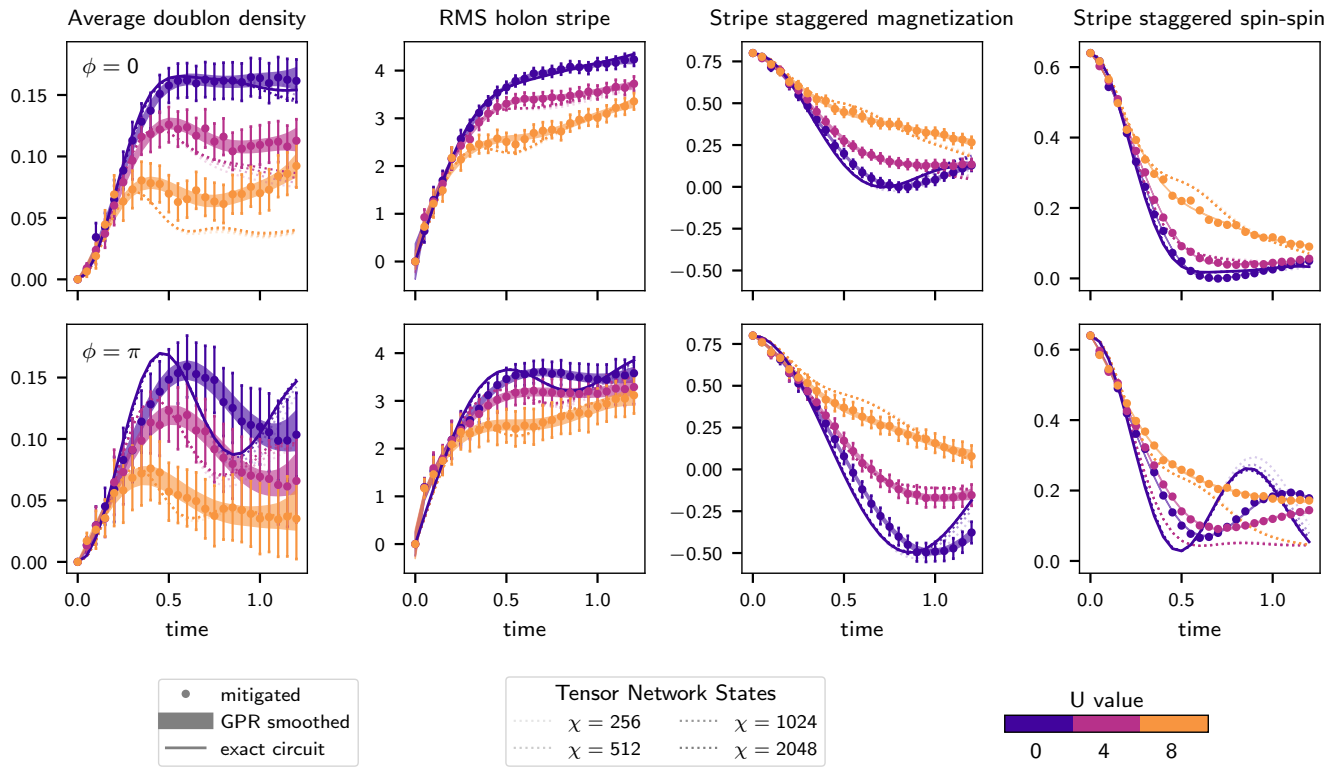


FIG. 11: Average doublon density B4, RMS holon stripe B9, stripe staggered magnetisation B10 and stripe staggered spin-spin as a function of time for $\phi = 0$ (top row) and $\phi = \pi$ (bottom row) for the 5×5 lattice with a holon stripe down the middle. Using 51 qubits and 2626 two-qubit gates. Here we also plot the results obtained using tensor network simulations at different bond dimensions, and exact circuit and TDS evolution when possible ($U = 0$).

$M^{\text{stripe}}_{\text{stag}}$ measures the magnetisation, taking into account the initial state configuration, where the type of bipartition in the Néel order flips crossing the holon stripe. $(S^z S^z)_{\text{stag}}$, on the other hand, captures the correlation between spins, with the initial staggered order considered between all pairs. We see that these signals dilute in time as expected due to the presence of mobile particles. A sketch of this process is shown in Fig. 12 (left).

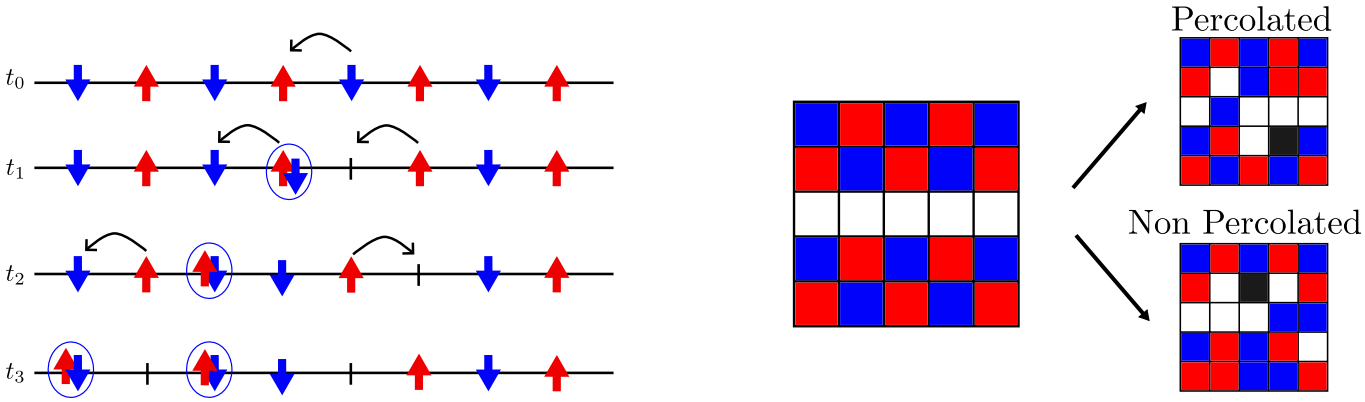


FIG. 12: Left: Illustration of the melting of antiferromagnetic order in the presence of mobile charges. The hopping of electrons in the Néel ordered background produces doublon-holon pairs of net zero spin. Their movement reduces the value of the staggered magnetisation and spin-spin correlations. Right: The evolution of the initial holon-stripe generates states that, when sampled in the computational basis, can be qualified as percolated or non-percolated. We use the percolated subset of shots to analyse the breakdown of the mean-field description of the stripe dynamics. Black sites denote the presence of a doublon.

We want to investigate the departure from the mean field description of the stripe dynamics, which assumes a stationary Néel background and only movement of the stripe such that it fluctuates around its initial position. The rationale behind this is based on the strong interaction picture, where the creation of doublons is forbidden and breaking the Néel pattern is very costly energetically. Under strict enforcement of these conditions, the only soft mode of the stripe is oscillation around its "centre of mass" – preserving the bipartition of the Néel pattern. We have already observed that both conditions are broken in practice for the interactions that we discussed. Motivated by this result, we study how likely it is for a configuration of holons to percolate from one edge of the lattice to the other with non-zero background magnetisation. Assuming the mean-field description, all sampled configurations should satisfy these criteria. In particular, we define the percolation fraction as the number of samples where a continuous line of holons crosses the lattice from one side to the opposite, divided by the total number of samples at a given time. A path is continuous if any holon in the path has a neighbour on its side or along the diagonal. An illustration of a percolated and a non-percolated path is shown in Fig. 12 (right). In Fig. 13, we show the percolation fraction postselected by number of doublons and by antiferromagnetic order. The first filters the obfuscation introduced by doublon-holon pairs appearing in the dynamics, while the second selects the states with a stronger Néel pattern. We defined the antiferromagnetic order in this case as

$$\mathrm{AFM} = \frac{1}{N_{\mathrm{shots}}(\mathcal{D})} \sum_{\substack{s\in\mathrm{shots} \\ N_d\in\mathcal{D}}} \sum_{\langle i,j\rangle} \frac{S_i^z(s)S_j^z(s)}{n_{\mathrm{pairs}}(s)} \tag{B12}$$

where shots($\mathcal{D}$) is the number of experimental shots with doublons in the range $\mathcal{D}$, and $n_{\mathrm{pairs}}(s)$ is the number of nearest neighbour pairs in the shot s with nonzero correlation $S_i^z(s)S_j^z(s)$. This observable essentially measures the antiferromagnetic correlation of the background of a postselected shot, according to the number of doublons present, and is normalised such that AFM $\in [-1, 1]$. In Fig. 13, we show the evolution of the percolation fraction in time for zero and π flux. In this figure, we also include the result of sampling from the MPS produced by TDVP at different bond dimensions. Since the initial state is Gaussian, we can efficiently sample from the $U = 0$ experiment to compute the true expected percolation fraction and then apply TFLO corrections to our hardware data. We observe some interesting behaviour. Apart from the non-interacting case, the effect of flux is not important. Moreover, the late time (postselected) percolation fraction *decreases* with interaction, in a completely opposite trend to the one expected by the mean-field description of the stripe dynamics. A likely explanation is provided by the disordering effect of the moving doublon-holon pairs which, although they are being postselected to a low number (in Fig. 13 $\mathcal{D} = [0, 1, 2, 3]$ doublons), are still present in the dynamics. Consider a virtual transition of a state with zero doublons, where a doublon-holon pair is created, they move and then they annihilate. In this process, the background could be perturbed away from the exact bipartite symmetry, reducing the antiferromagnetic order.

To gain an understanding of this phenomenon, we analyse the evolution of the magnetisation $\sum_{\langle ij\rangle} S_i^z S_j^z$ in time, over postselected shots with a doublon number in Fig. 14, where we employ MESR as the error mitigation technique. For more information on the FLO sampler, see Appendix E 6. For the TNS, we can also exactly and efficiently sample from the MPS state in the TeNPy TDVP simulations.

We observe that indeed a lower number of doublons leads to a slower melting of the antiferromagnetic order as the interaction increases. We plot the comparison between this signal obtained from sampling the classical tensor network state at the largest bond dimension considered $\chi = 2048$, against the results obtained from the device, where the experimental shots have been mitigated according to the shot mitigation procedure outlined in Appendix E 10. We see qualitative agreement between the TN and the experimental pictures.

For similar plots for other system sizes, see Appendix K.

3. Spin-charge separation as a function of doping and magnetic field

Spin charge separation is a phenomenon wherein the electron's charge and spin fractionalise into independent degrees of freedom. In one dimension, this separation was predicted from the exact solution [81] and it has since been observed in many scenarios [28, 82]. In higher dimensions, it has been conjectured that above the ground state and around half filling, the electron fractionalises into spinons and chargons (holons and doublons) of fermionic and bosonic character respectively, where the bosons are the carriers of charge [83, 84]. We study this fractionalisation by looking at the dual description of the Fermi-Hubbard model as a dual gauge theory in terms of the chargons and spinons. We do this using the (modified) Kotliar-Ruckenstein representation [85] of the electron operator

$$c_{i\sigma}^\dagger = f_{i\sigma}^\dagger h_i + \sigma f_{i\bar{\sigma}} d_i^\dagger, \tag{B13}$$

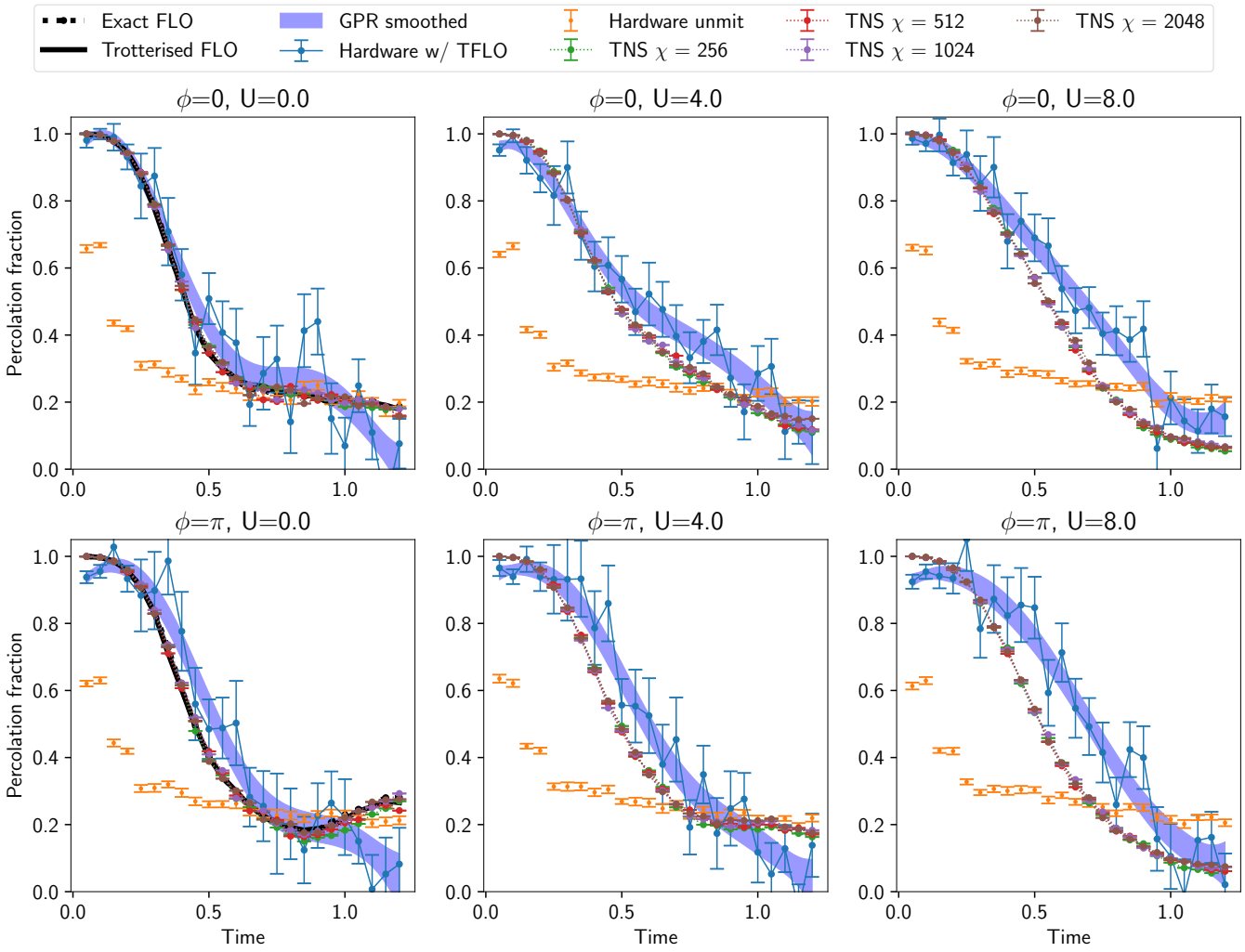


FIG. 13: Percolation fraction for a 5×5 system. Note that unmitigated hardware data is postselected by antiferromagnetic order measured by AFM $\in [-1, 0]$ (see Eq. (B12), number of doublons $\mathcal{D} = [0, 1, 2, 3]$ present and particle number conservation. TFLO is then applied to this set of shots to produce the Hardware w/o TFLO line, and finally GPR is applied to give the 68% one sigma region (for more details, see Appendix E 2, Appendix E 7 and Appendix E 3).

where $\sigma = \{\uparrow, \downarrow\} = \{1, -1\}$ is the spin index. In this representation, the doublon $d_i^\dagger := \frac{1}{2}(X_i^d + iY_i^d)$ (holon $h_i^\dagger := \frac{1}{2}(X_i^h + iY_i^h)$ satisfies the usual algebra inherited from the Pauli matrices (X, Y, Z). These quasiparticles commute at different sites. The spinon operator, on the other hand, satisfies fermionic anticommutation relations $\{f_{i\sigma}, f_{j\sigma'}\} = 0$ and $\{f_{i\sigma}, f_{j\sigma'}^\dagger\} = \delta_{ij}\delta_{\sigma\sigma'}$. As the representation Eq. (B13) enlarges the local Hilbert space, a local constraint is needed to recover the physical subspace. The constraint of exactly one quasiparticle per site

$$C_i := d_i^\dagger d_i + f_{i\uparrow}^\dagger f_{i\uparrow} + f_{i\downarrow}^\dagger f_{i\downarrow} + h_i^\dagger h_i = 1 \tag{B14}$$

defines the physical states and fixes the correct commutation relations of the physical fermion operators, i.e for all states satisfying $C_i|\text{phys}\rangle = |\text{phys}\rangle \ \forall i$, the operator $c_{i\sigma}$ satisfies

$$\{c_{i\sigma}, c_{j\sigma'}^\dagger\}|\text{phys}\rangle = |\text{phys}\rangle \tag{B15}$$

thus defining a fermionic operator. The projection $\mathcal{P} = \prod_j \frac{C_j(2-C_j)(3-C_j)(4-C_j)}{6}$ into the physical space commutes with $c_{i\sigma}$ as $[C_i, c_{j\sigma}] = 0$. This implies that the action of $c_{j\sigma}$ in a physical state produces another physical state.

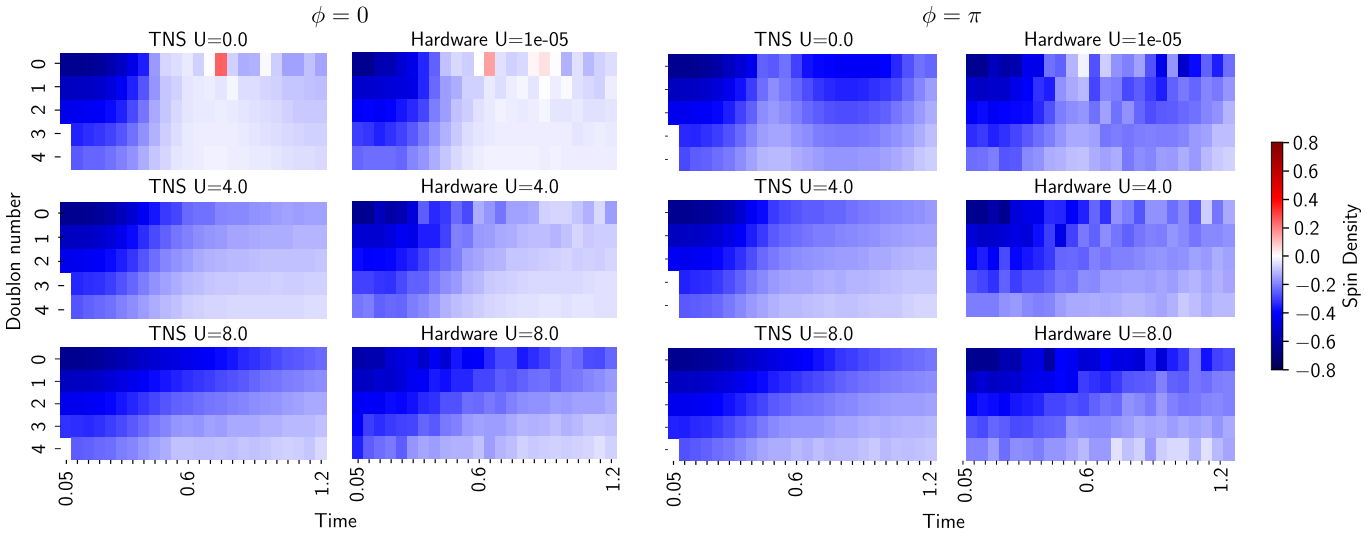


FIG. 14: Evolution of antiferromagnetic order, defined as $\sum_{\langle ij\rangle} S_j^z S_j^z$ for different fluxes (left two columns $\phi = 0$, right two columns $\phi = \pi$, for AFM range $\in [-1, 0]$, doublons included $\in [0, 1, 2, 3]$. TNS represent classical results using tensor network simulation (see Appendix F), while hardware results employ the MESR error mitigation discussed in appendix E 10).

The Hamiltonian acting on the physical space becomes $H = H_n + H_s + H_{\text{int}}$ with

$$H_n = -J \sum_{\sigma\langle ij\rangle} e^{i\phi_{ij}} (f_{i\sigma}^\dagger (h_i h_j^\dagger) f_{j\sigma} + f_{i\sigma} (d_i^\dagger d_j) f_{j\sigma}^\dagger) + \text{h.c.}, \tag{B16}$$

$$H_s = -J \sum_{\langle ij\rangle} e^{i\phi_{ij}} [(f_{i\uparrow}^\dagger f_{j\downarrow}^\dagger - f_{i\downarrow}^\dagger f_{j\uparrow}^\dagger) h_i d_j + (f_{i\downarrow} f_{j\uparrow} - f_{i\uparrow} f_{j\downarrow}) d_i^\dagger h_j^\dagger] + \text{h.c.}, \tag{B17}$$

$$H_{\text{int}} = U \sum_i d_i^\dagger d_i. \tag{B18}$$

where $e^{i\phi_{ij}}$ represents flux through the system, as considered in some of our experiments. This extended Hamiltonian has a local $U(1)$ gauge symmetry generated by $e^{i\theta C_j}$ as $e^{i\theta C_j} f_{i,\sigma}^\dagger e^{-i\theta C_j} = e^{i\theta} f_{i\sigma}^\dagger$, and similarly for the doublon and holon quasiparticles operators. This local gauge symmetry motivates the definition of two extended operators, the Wilson loop

$$\mathcal{W}_d(\mathcal{C}) := \prod_{j\in\mathcal{C}} (1 - n_{j,\uparrow} n_{j,\downarrow}), \tag{B19}$$

that measures the deconfinement transition of the gauge field mediating the interaction between spinons, and the Wilson lines between chargons (holons and doublons)

$$V_{ab}(M) := \sum_{\substack{\text{sites } i,j: \\ |i-j|=M}} \frac{n_i^a n_j^b}{N_{\text{pairs}}} \sum_{\substack{\text{paths } \gamma: \\ \text{from } i \text{ to } j}} \frac{\sum_{(m,n)\in\gamma} S_m^z S_n^z}{N_{\text{paths}}(i,j)} \tag{B20}$$

where in the inner sum m, n are consecutive positions along the path γ. The labels a and b can be either a holon or a doublon label, with $n_j^d = n_{j\uparrow} n_{j\downarrow}$ and $n_j^h = (1 - n_{j\uparrow})(1 - n_{j\downarrow})$.

In Fig. 15, we show the initial configuration we consider, namely a half-filled lattice with 2 holes and a singlet covering. Some examples of Wilson lines with fixed Manhattan distance between holon-doublon pairs entering the computation of $V_{hd}(M)$ are shown in Fig. 15b.

Figs. 16 and 17 show results for $V_{hd}(M)$ for a 5×5 lattice initialised in a state with two holes and singlet covering with and without magnetic flux, respectively. Figs. 18 and 19 show the same for a 6×6 lattice. In the presence of confinement between chargons, we expect $V_{hd}(M)$ to grow with the (Manhattan) distance between the holon and the doublon of a given pair. In the absence of magnetic flux, both Figs. 16 and 18 show that after an initial transient, the value of $V_{hd}(M)$ for $U = 0$ decreases faster than in the presence of interactions. This suggests the absence of any

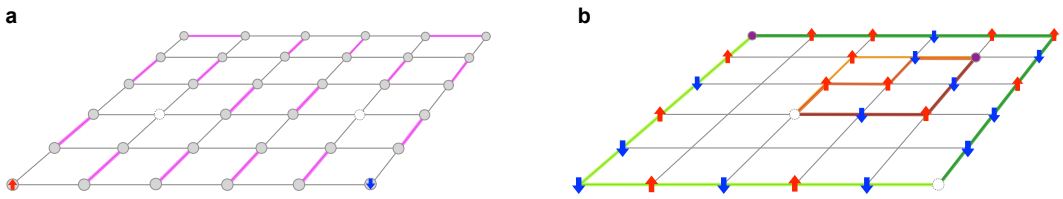


FIG. 15: **a** Sketch of the initial state with two holes (empty circles) and singlet covering (magenta links) for a 6×6 lattice. **b** Various open Wilson lines between pairs with one holon (empty circle) and one doublon (purple circle). Here, the green paths highlight the two only possible paths with Manhattan distance $M = 10$, while the red paths show three of the six possible paths with Manhattan distance $M = 6$. To compute the value of $V_{hd}(M)$ on a single sample, one has to average the spin correlations between neighbours along all the possible paths between all the possible doublon-holon pairs with Manhattan distance M.

confinement between chargons in the non-interacting regime with $\phi = 0$. We ascribe any residual value to a lack of thermalisation (see also Appendix B 4). Interestingly, despite the MESR and TDVP signals agreeing well at short times, the experimental data predict a faster decay of $V_{hd}(M)$ at longer times.

On the other hand, as shown in Figs. 17 and 19, a magnetic flux makes $V_{hd}(M)$ non-vanishing even for $U = 0$, at least up to the largest time we can access in our experiment. This suggests that a magnetic flux induces a stronger confinement between chargons. In this case as well, MESR and TDVP signals agree well at short times and start to deviate from each other at intermediate times. However, in contrast with the case with $\phi = 0$, at longer times the agreement between the two signals improves again. From the global observables, like the staggered magnetisation shown in Fig. 20 below, we observe a similar revival in the non-interacting case, signalling that in this case the system is far away from equilibration and the signal is showing transient behaviour.

In all the figures discussed above, we observe that $V_{hd}(M)$ saturates at a value of M which depends on the strength of the interactions. While a decay at larger M can be ascribed to combinatorics effects (e.g., as shown in Fig. 15b, there are only a few holon-doublon configurations with Manhattan distance $M = 10$), we defer the analysis of this phenomenon to future work.

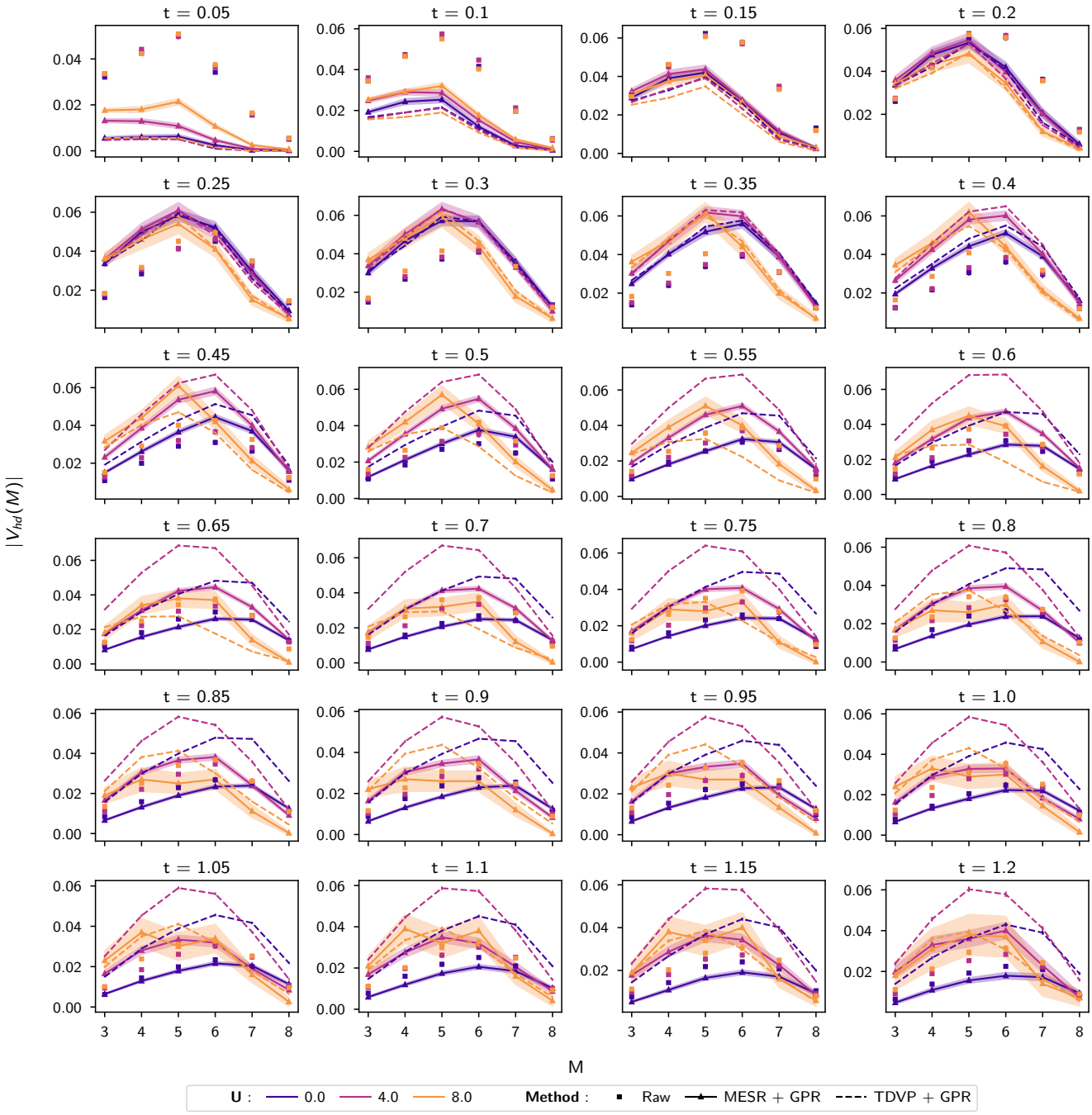


FIG. 16: Expectation value of open Wilson lines $V_{hd}(M)$ for a 5×5 instance initialised in the state with singlets covering and two holes with zero flux ($\phi = 0$) as a function of the Manhattan distance M between holon-doublon pairs for different values of time. Dashed lines have been obtained using 10000 samples from a TDVP simulation with $\chi = 2048$. Solid lines have been obtained using the maximum entropy shot-reweighting (MESR) procedure described in Appendix E 10. Both experimental and TDVP data have been smoothed using GPR on the time-dependent signal.

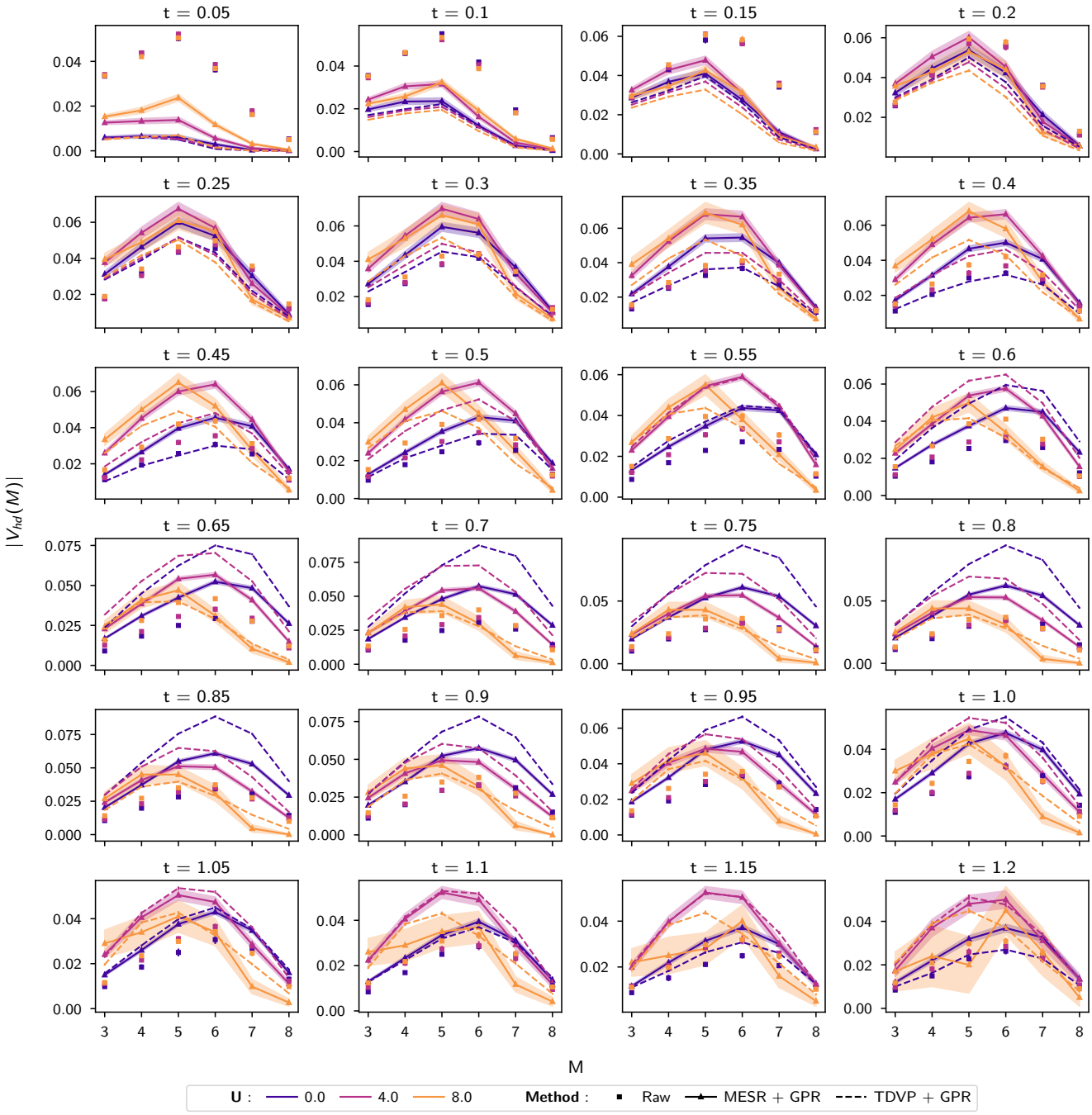


FIG. 17: Same as Fig. 16 for $\phi = \pi$.

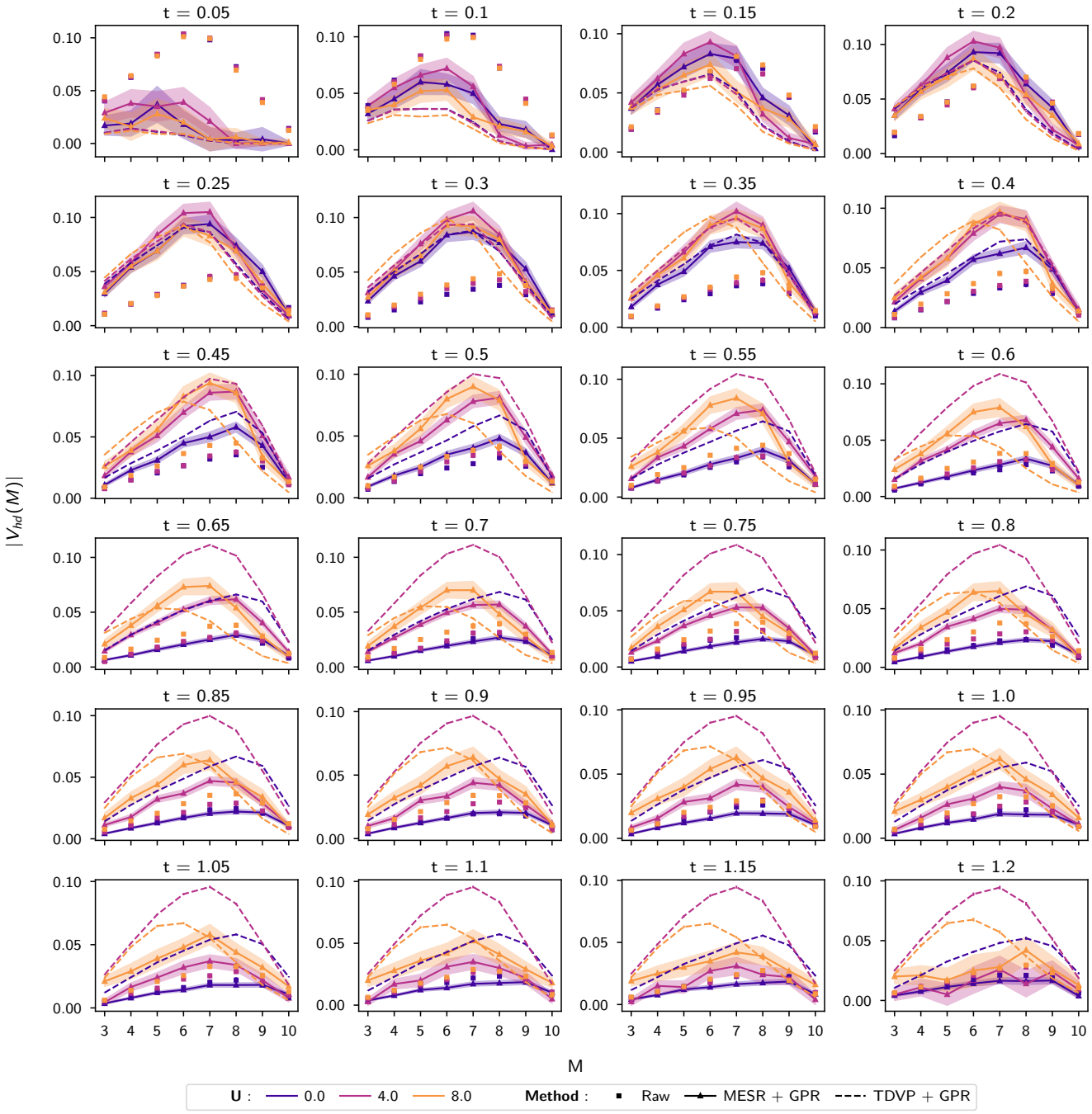


FIG. 18: Expectation value of open Wilson lines $V_{hd}(M)$ for a 6×6 instance initialised in the state sketched in Fig. 15a with zero flux ($\phi = 0$) as a function of the Manhattan distance M between holon-doublon pairs for different values of time. Dashed lines have been obtained using 10000 samples from a TDVP simulation with $\chi = 2048$. Solid lines have been obtained using the maximum entropy shot-reweighting (MESR) procedure described in Appendix E 10. Both experimental and TDVP data have been smoothed using GPR on the time-dependent signal.

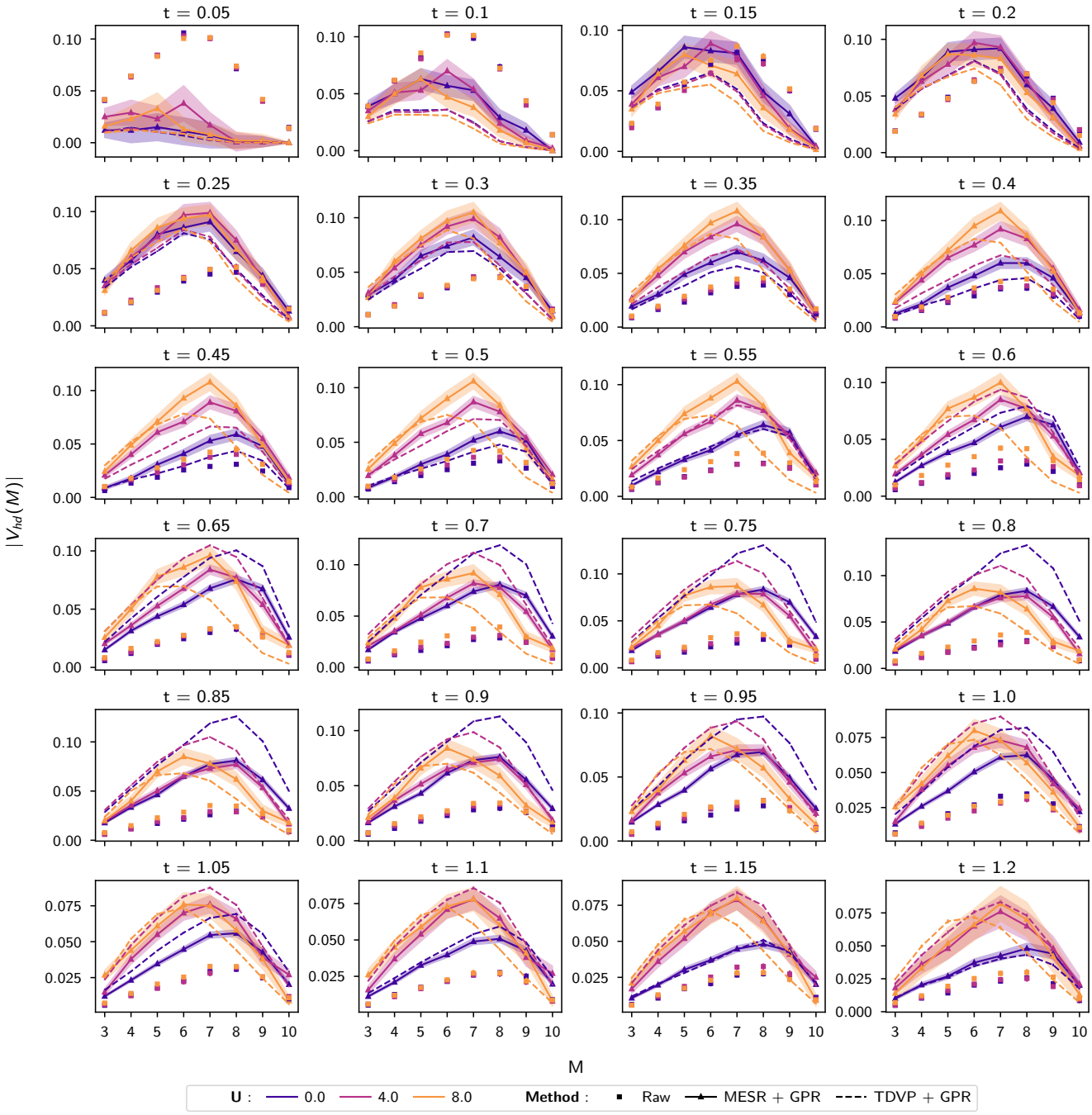


FIG. 19: Same as Fig. 18 for $\phi = \pi$.

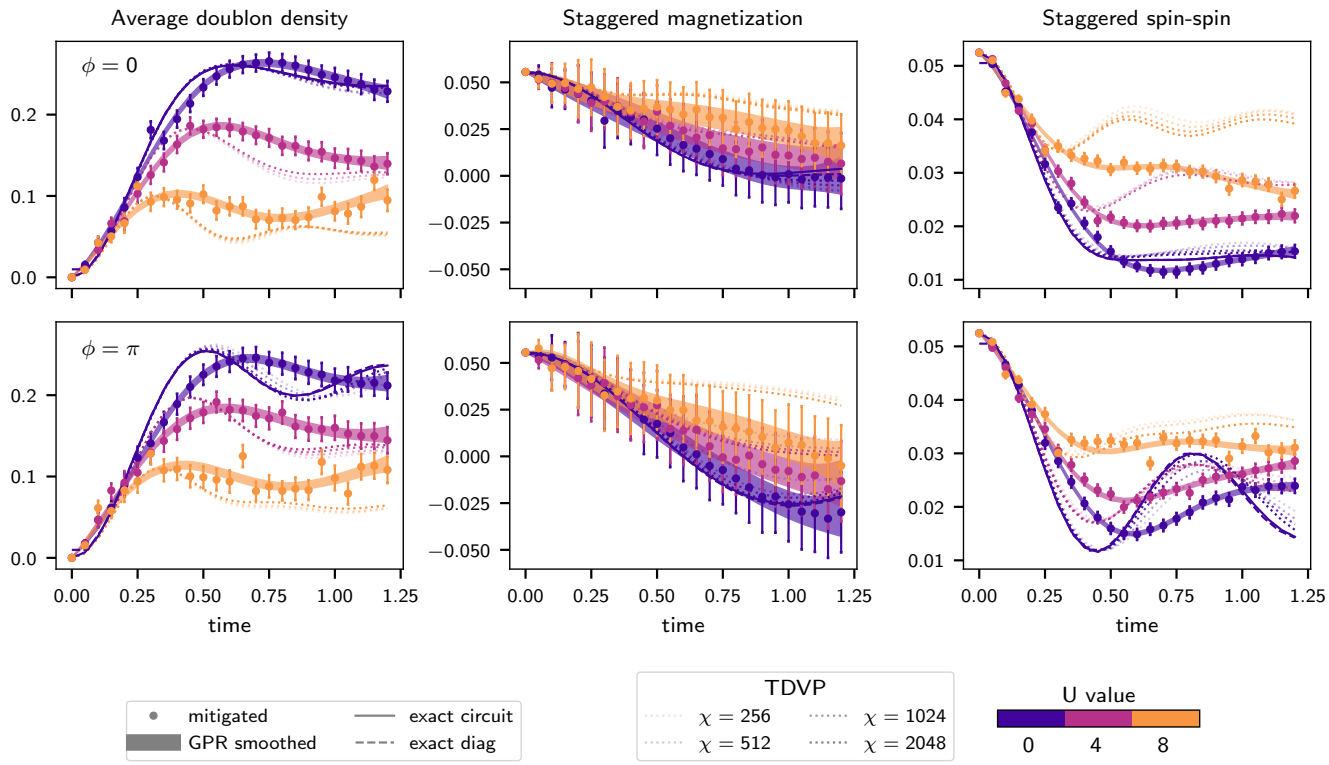


FIG. 20: Global observables evolved from the 6×6 initial state with two holons. Note that the tiny values of the staggered magnetisation are expected due to the initial singlet configuration. The revival seen in the staggered spin-spin observable indicates that the system is far away from a steady state (equilibrated).

4. Thermalisation

As we have seen so far from the global observables on different initial states, the behaviour of signals between them is qualitatively similar, showing, e.g. an increase in the number of doublons as time progresses, seemingly approaching an asymptotic value depending on the interaction strength and the initial occupation. Under generic conditions, it is expected [86–88] that the nonequilibrium dynamics relaxes to states where macroscopic quantities are stationary, universal with respect to initial conditions sharing the same global quantum numbers, and predictable through (quantum) statistical mechanics. The eigenstate thermalisation hypothesis [5, 89] provides a mechanism for this relaxation to happen. It implies that expectation values of all local observables reach their asymptotic values given by the *diagonal ensemble* (in the eigenstate basis) for that initial state and this value is given by the expectation in the microcanonical ensemble on a narrow energy range around the energy of the initial state.

The timescale that we consider in our evolution is not enough to see full thermalisation, as several observables have nontrivial time dependence even at the largest time explored. Nevertheless, the trend of approaching the thermal asymptotic value can be distinguished in terms of different initial quantum numbers. In this section, we explore the approach to thermalisation in the Fermi-Hubbard model as we change the number of electrons. We do this by looking at observables obtained by averaging five different initial states with the same number of holons, ranging from one to six, placed on random positions on an otherwise Néel state with zero flux.

a. *Doublon Number*

In Fig. 21a, we see the evolution of the doublon number (given by the density Eq. (B4) times the system's size, $N_d = \langle \sum_j n_{j\uparrow} n_{j\downarrow} \rangle$) in time, for different hole-doped states in a 5×5 system. It shows the expected dependence of $N_d(t)$ on U (slower dynamics for larger repulsion) and doping (more holes in the initial state result in fewer doublons). The asymptotic long-time values of $N_d(t)$ for the non-interacting case are given by $N_\uparrow N_\downarrow / L$, and are shown on the plot with grey lines. The shape of the curves for $U = 4$ and $U = 8$ suggests the possibility of collapse upon rescaling, which has been investigated in Fig. 21b, where the $U = 4$ curve is shown in its original form while the $U = 8$ curve is rescaled along both axes, with coefficients dependent on the problem size.

We can understand this scaling as follows. Consider the Fermi-Hubbard Hamiltonian $H = H_0 + UH_I$, where H_0 is the non-interacting Hamiltonian and H_I parametrises the interactions. Here we make explicit the dependence on the interaction parameter U. An operator $\mathcal{O}$ made of densities commutes with the interactions, so going to the interaction picture with respect to H_I gives

$$\langle \mathcal{O}(t) \rangle = \left\langle e^{it(H_0+UH_I)} \mathcal{O} e^{-it(H_0+UH_I)} \right\rangle = \left\langle \mathcal{T} e^{i \int_0^t H_0(\tau)} \mathcal{O} \mathcal{T} e^{-i \int_0^t H_0(\tau)} \right\rangle \tag{B21}$$

where $H_0(\tau) = e^{iU\tau H_I} H_0 e^{-iU\tau H_I}$. Now using the fermion commutation relations we have

$$e^{iU\tau H_I} c_{j\sigma} e^{-iU\tau H_I} = e^{iUt n_{j\sigma} n_{j\bar{\sigma}}} c_{j\sigma} e^{-iUt n_{j\sigma} n_{j\bar{\sigma}}} = c_{j\sigma} e^{-iUt n_{j\bar{\sigma}}} \tag{B22}$$

so the time dependent Hamiltonian is $H_0(\tau) = \sum_{\langle jk \rangle} c^\dagger_{k\sigma} c_{j\sigma} e^{-iUt(n_{j\bar{\sigma}} - n_{k\bar{\sigma}})}$. The expectation value takes the form

$$\langle \mathcal{O}(t) \rangle_U = \left\langle \mathcal{T} e^{i \int_0^t \sum_{\langle jk \rangle} c^\dagger_{k\sigma} c_{j\sigma} e^{-iU\tau(n_{j\bar{\sigma}} - n_{k\bar{\sigma}})} d\tau} \mathcal{O} \mathcal{T} e^{-i \int_0^t \sum_{\langle jk \rangle} c^\dagger_{k\sigma} c_{j\sigma} e^{-iU\tau(n_{j\bar{\sigma}} - n_{k\bar{\sigma}})} d\tau} \right\rangle \tag{B23}$$

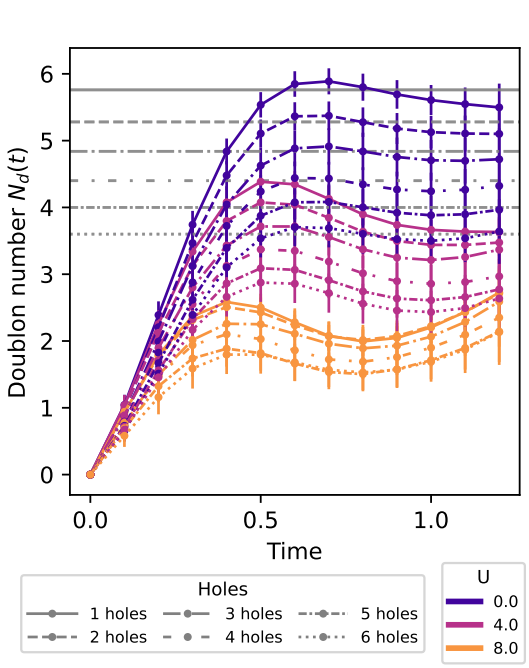
where we use the subscript U to emphasise that this is an expectation of a time evolved operator where the interaction has strength U. The integral in the time-ordered product can be written as

$$\int_0^t \sum_{\langle jk \rangle} c^\dagger_{k\sigma} c_{j\sigma} e^{-iU\tau(n_{j\bar{\sigma}} - n_{k\bar{\sigma}})} d\tau = \frac{1}{U} \int_0^{Ut} \sum_{\langle jk \rangle} c^\dagger_{k\sigma} c_{j\sigma} e^{-is(n_{j\bar{\sigma}} - n_{k\bar{\sigma}})} ds \tag{B24}$$

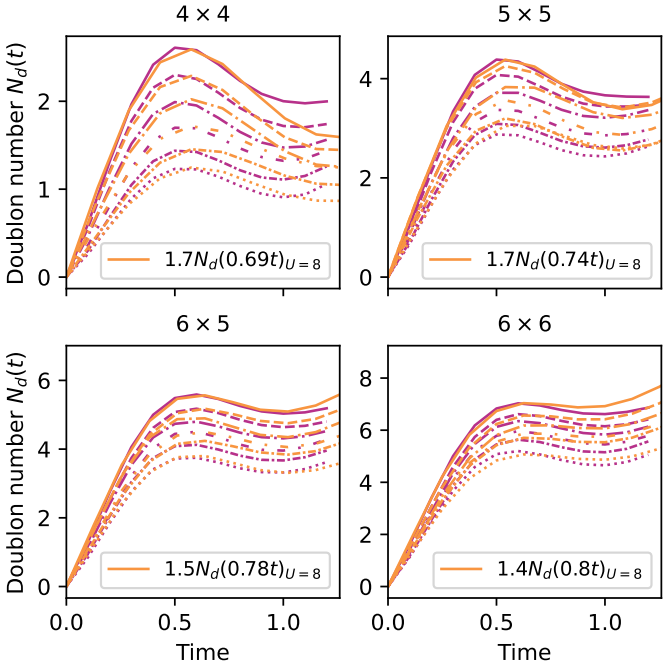
so expanding in powers of $1/U$, we have

$$\langle \mathcal{O}(t) \rangle_U = \langle \mathcal{O}(0) \rangle + i \frac{1}{U} \sum_{\langle jk \rangle, \sigma} \int_0^{Ut} e^{-is(n_{j\bar{\sigma}} - n_{k\bar{\sigma}})} ds \left\langle \left[c^\dagger_{k\sigma} c_{j\sigma}, \mathcal{O} \right] \right\rangle \tag{B25}$$

$$+ \frac{1}{U^2} \sum_{\langle jk \rangle, \langle lm \rangle, \sigma} \int_0^{Ut} e^{-is(n_{j\bar{\sigma}} - n_{k\bar{\sigma}})} e^{-is'(n_{l\bar{\sigma}} - n_{m\bar{\sigma}})} ds ds' \left(\left\langle c^\dagger_{k\sigma} c_{j\sigma} \mathcal{O} c^\dagger_{m\sigma} c_{l\sigma} \right\rangle - \frac{1}{2} \left\langle \left\{ c^\dagger_{k\sigma} c_{j\sigma} c^\dagger_{m\sigma} c_{l\sigma}, \mathcal{O} \right\} \right\rangle \right) + O(t^3)$$



(a) Doublon number for the 5×5 system. Grey lines show the long-time asymptotic values of doublon number for the free fermion system, $N_\uparrow N_\downarrow / L$.



(b) Doublon number collapse for 4×4, 5×5, 6×5, and 6×6 system. Shown are the original $U = 4$ line and the rescaled $U = 8$ line.

FIG. 21: Doublon number. All lines are obtained by averaging over five initial configurations with a given number of randomly placed holons on top of the Néel background.

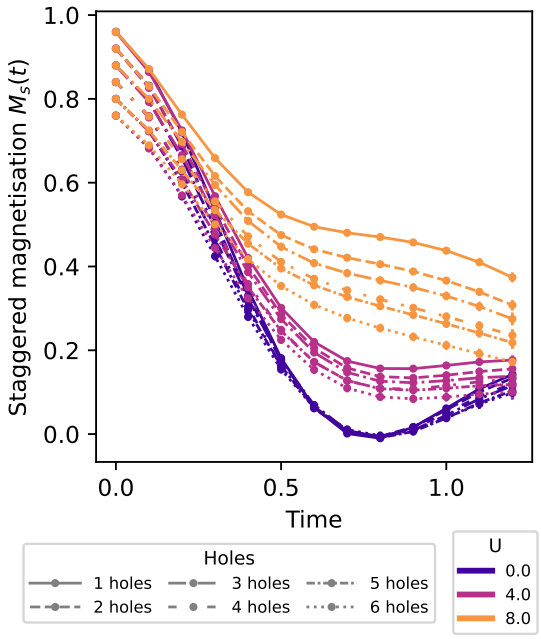
For every operator $\mathcal{O}$ made of densities, $\mathcal{O}|\psi\rangle = n_\mathcal{O}|\psi\rangle$ where $n_\mathcal{O}$ is an integer which implies that the $O(t)$ term is always zero as $\langle[c^\dagger_{k\sigma}c_{j\sigma}, \mathcal{O}]\rangle = \langle c^\dagger_{k\sigma}c_{j\sigma}\mathcal{O}\rangle - \langle \mathcal{O}c^\dagger_{k\sigma}c_{j\sigma}\rangle = n_\mathcal{O}(\langle c^\dagger_{k\sigma}c_{j\sigma}\rangle - \langle c^\dagger_{k\sigma}c_{j\sigma}\rangle) = 0$. This equation can be further simplified for our case, where the initial state is a product state with the number of doublons $\mathcal{O} = N_d = 0$. The expectation of the anti-commutator also becomes zero since $n_\mathcal{O} = 0$ and following the same argument as the commutator. Additionally, since the initial state is a product state, the only non-zero terms occur when $m = j$ and $l = k$. So, for early times, the doublon number satisfies the scaling relation

$$\begin{aligned}
\langle N_d(t)\rangle_{U_1} &= \frac{1}{U_1^2}\sum_{\langle jk\rangle,\sigma}\int_0^{U_1 t} e^{-i(s-s')(n_{j\bar\sigma}-n_{k\bar\sigma})}dsds'\left\langle c^\dagger_{k\sigma}c_{j\sigma}N_d c^\dagger_{j\sigma}c_{k\sigma}\right\rangle + O(t^3)\\
&= \left(\frac{U_2}{U_1}\right)^2\frac{1}{U_2^2}\sum_{\langle jk\rangle,\sigma}\int_0^{U_2(U_1/U_2)t} e^{-i(s-s')(n_{j\bar\sigma}-n_{k\bar\sigma})}dsds'\left\langle c^\dagger_{k\sigma}c_{j\sigma}N_d c^\dagger_{j\sigma}c_{k\sigma}\right\rangle + O(t^3)\\
&= \left(\frac{U_2}{U_1}\right)^2\left\langle N_d\left(\frac{U_1}{U_2}t\right)\right\rangle_{U_2} + O(t^3).
\end{aligned}$$

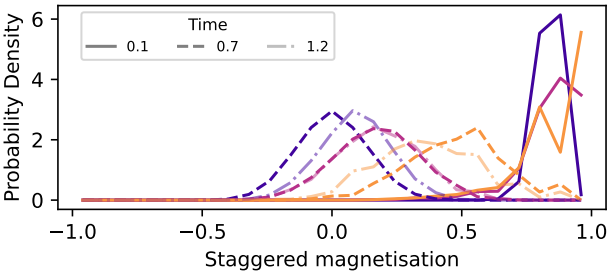
We observe this same scaling relation with renormalised parameters, shown in the insets of Fig. 21b. Note that the collapse deteriorates for larger times $t > 1$.

b. *Staggered Magnetisation*

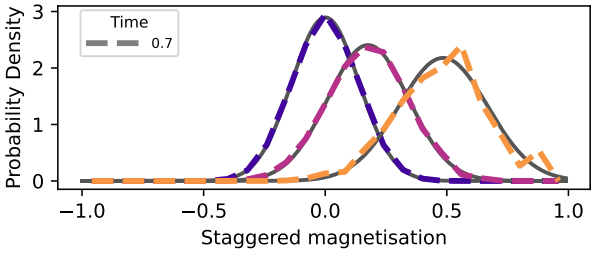
In the spin sector, we study the staggered magnetisation Eq. (B6) which is shown in Fig. 22a as a function of doping for different values of interactions $U = 0, 4, 8$ in a 5×5 system, from error-mitigated experimental results (see Appendix E). As we have observed before, increasing the interaction strength U slows down the dynamics of the system. The magnetic order generally decreases with time as the system evolves, while a lower number of electrons



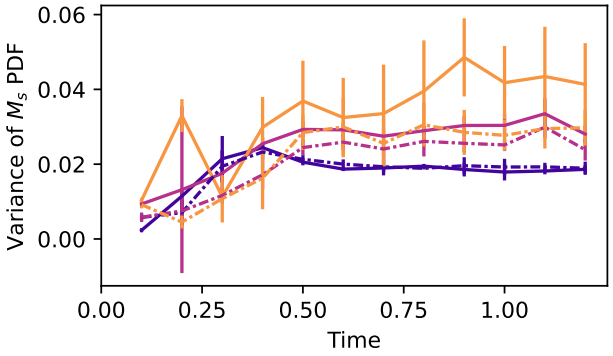
(a) Staggered magnetisation for the 5×5 system, shown for different interactions and dopings.



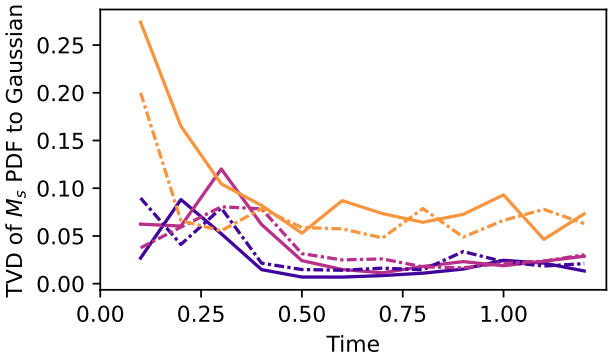
(b) Evolution of the staggered magnetisation distribution, averaged over all one hole states.



(c) Staggered magnetisation distribution quickly acquires nearly Gaussian shape (grey line), same initial state as in (b).



(d) Evolution of the variance of the staggered magnetisation distribution, shown for initial states with one and five holons.



(e) Total Variational Distance between the staggered magnetisation distribution and a normal distribution with the same mean and variance.

FIG. 22: Staggered magnetisation for 5×5 system, for various numbers of holes on top of the Néel background, averaged over five random states for each number of holes.

leads monotonically to a lower magnetisation. The degree to which the reduction in magnetisation happens at a fixed time as we vary the doping is controlled by the interaction.

All lines show non-trivial behaviour far from infinite-time thermal values ($\langle M_s \rangle_\infty = 0$ for the even number of sites L and $\langle M_s \rangle_\infty = (N_\uparrow - N_\downarrow)/L$ for odd system size $L = L_x \times L_y$).

A more refined observable allowing one to track the system's departure from the antiferromagnetic order is the probability distribution of M_s defined as [90]

$$P_{M_s}(m,t) = \langle \delta(M_s - m) \rangle \,, \tag{B26}$$

where m are the eigenvalues of M_s, whose details are presented in Fig. 22b. In particular, Fig. 22b shows the profile of $P_{M_s}(m,t)$ at different moments of time and for various values of U. These plots illustrate in more detail how larger values of U slow down the spin exchange by suppressing the doublon formation.

We approximate the evolution of $P_{M_s}(m,t)$ by the Gaussian distribution

$$P_{M_s}(m,t) = \frac{1}{\sqrt{2\pi\sigma(t)}} \exp\left\{-\frac{[m - M_s(t)]^2}{2\sigma(t)}\right\}. \tag{B27}$$

where the variance $\sigma(t)$ is a (time-dependent) fitting parameter.

The distributions $P_{M_s}(m,t)$ rapidly approach the Gaussian shape (B27), as is evident from Fig. 22c. The variances of distributions from Fig. 22b are shown in Fig. 22d from which the dependence on interaction and doping can be inferred. Stronger interactions maintain configurations with large staggered magnetization, leading to a broader distribution. Increasing the number of holes introduces mobile charge carriers that disrupt long-range spin correlations, accelerating relaxation. The convergence of experimental probability distributions $P_{M_s}(m,t)$ to the Gaussian shape can be characterised by their total variation distance, which is shown in Fig. 22e. Larger values of U lead to a slower convergence toward the Gaussian distribution, with $U = 8$ not showing signs of convergence, an observation that remains consistent across different system sizes. We observe that while the variance is not strongly dependent on the number of holes present, it does have a monotone dependence on the interaction strength.

c. *Inverse Participation Ratio*

So far, we have discussed the approach to thermalisation through the behaviour of local few-body observables. In this section, we discuss a more distributed measure of thermalisation, the inverse participation ratio (IPR) [89, 91]. This measure, first used in the study of Anderson localisation of non-interacting fermions [91–93], quantifies the degree to which the wavefunction is spread in some particular basis. For a state $|\psi\rangle = \sum_i c_i|i\rangle$ expanded in the basis $|i\rangle$, the IPR is defined as

$$\mathrm{IPR} = \sum_i |c_i|^4 = \sum_i p_i^2. \tag{B28}$$

where p_i is the probability of observing the outcome i measuring the state $|\psi\rangle$ in the basis $|i\rangle$. For a fully delocalised state, corresponding to an equal superposition over all possible basis states, the IPR becomes $1/D$ where D is the Hilbert space dimension. On the other hand, for a state with support on a single basis state, the IPR takes the value one. Since in a given basis $\mathrm{IPR} = \sum_i (\rho^2)_{ii}$, it can be viewed as a basis-dependent "classical" version of purity $\mathrm{Tr}(\rho^2)$, which always provides a lower bound on it.

Calculating the full IPR defined in Eq. (B28) requires a full characterisation of the probabilities p_i, which is unfeasible at the system sizes that we consider. Instead, we study a *marginal* version of the IPR, which for a state $|\psi\rangle = \sum_{ij} c_{ij}|i\rangle_A|j\rangle_B$ defined on a bipartition A and B corresponds to the usual IPR over the marginalised probabilities, i.e.

$$\mathrm{IPR}_A := \sum_i p_i^2, \qquad p_i := \sum_j |c_{ij}|^2. \tag{B29}$$

For a marginal on m sites, this quantity can be estimated experimentally with sample overhead that scales at most with the subsystem dimension 4^m, providing a measure of thermalisation in that particular basis as well as a lower bound on the subsystem's purity. As suggested by classical state vector simulations on smaller systems, the marginal IPR exhibits behaviour qualitatively similar to that of the full-system IPR: decay in time as the state becomes delocalised in the Fock basis, with decay rates that can be related to other observables like time-dependent holon number as described below.

For the Fermi–Hubbard model, we construct several versions of the marginal inverse participation ratio (IPR) using different single-site bases, each corresponding to a distinct assignment of spin and charge quantum numbers to the original local Fock states $\{|0\rangle, |\uparrow\rangle, |\downarrow\rangle, |\uparrow\downarrow\rangle\}$. The full single-site IPR employs the complete alphabet $\{0, \uparrow, \downarrow, \uparrow\downarrow\}$ to characterise the basis at each site. The charge version reduces this alphabet to $\{H, S, D\}$ by merging the spin-resolved singly occupied states into a single "singly-occupied" class S, while keeping the empty (hole, H) and doubly occupied (doublon, D) states distinct. Conversely, the spin version isolates the spin degrees of freedom by merging the empty and doublon states into a single "spinless" category $\varnothing$, while preserving the distinction between up (U) and down (D) singly occupied states. These mappings define coarse-graining functions $C(\cdot)$ and $S(\cdot)$ on the single-site states:

$$C(0) = H, \quad C(\uparrow) = C(\downarrow) = S, \quad C(\uparrow\downarrow) = D, \tag{B30a}$$

$$S(0) = S(\uparrow\downarrow) = \varnothing, \quad S(\uparrow) = U, \quad S(\downarrow) = D. \tag{B30b}$$

For a single-site marginal, the three IPR variants are:

$$\mathrm{IPR}_1^{\mathrm{full}} = p_0^2 \ + p_\uparrow^2 \ + p_\downarrow^2 \ + p_{\uparrow\downarrow}^2 \,, \tag{B31a}$$

$$\mathrm{IPR}_1^{\mathrm{charge}} = p_H^2 + p_S^2 + p_D^2 = p_0^2 + (p_\uparrow + p_\downarrow)^2 + p_{\uparrow\downarrow}^2 \,, \tag{B31b}$$

$$\mathrm{IPR}_1^{\mathrm{spin}} = p_\varnothing^2 + p_U^2 + p_D^2 = (p_0 + p_{\uparrow\downarrow})^2 + p_\uparrow^2 + p_\downarrow^2 \,. \tag{B31c}$$

These definitions naturally extend to m-site marginals. Let $p_{i_1 \ldots i_m}$ denote the joint probability of observing the configuration $(i_1, \ldots, i_m)$ in the original four-state Fock basis across m sites. The m-site IPRs are defined as the sum of squared probabilities of the relevant marginal distributions. For the full IPR, this uses the original four-state Fock basis; for the charge and spin variants, it uses coarse-grained probabilities obtained by summing over all fine-grained configurations consistent with the charge (C) or spin (S) mappings:

$$\mathrm{IPR}_m^{\mathrm{full}} = \sum_{i_1,\ldots,i_m \in \{0,\uparrow,\downarrow,\uparrow\downarrow\}} \left(p_{i_1 \ldots i_m}\right)^2 , \tag{B32a}$$

$$\mathrm{IPR}_m^{\mathrm{charge}} = \sum_{c_1,\ldots,c_m \in \{H,S,D\}} \left(\sum_{\substack{i_1,\ldots,i_m \\ C(i_k)=c_k \ \forall k}} p_{i_1 \ldots i_m} \right)^2 , \tag{B32b}$$

$$\mathrm{IPR}_m^{\mathrm{spin}} = \sum_{s_1,\ldots,s_m \in \{\varnothing,U,D\}} \left(\sum_{\substack{i_1,\ldots,i_m \\ S(i_k)=s_k \ \forall k}} p_{i_1 \ldots i_m} \right)^2 . \tag{B32c}$$

All three IPR variants are directly computable from (mitigated) experimental measurement shots: one first constructs the m-site marginal by counting the relative frequencies of observed configurations, and then – depending on the desired IPR – either uses the raw Fock outcomes (for the full IPR) or applies the charge (C) or spin (S) mapping to each site before aggregating probabilities and evaluating the sum of squared frequencies. The resulting IPRs are shown in Figs. 23a, 23b, and 23c for the 4-site marginal depicted in Fig. 23d.

The absolute lower bound is obtained when each mode on each site is occupied with probability 1/4, so for an m-site marginal IPR this floor is given by $(1/4)^m$ for the usual IPR on all modes and $(3/8)^m$ for both the charge and spin IPRs.

Under the independent and identically distributed (IID) -site approximation, the instantaneous doublon count $N_d(t)$ can be used to estimate all m-site probabilities required for the marginal IPRs. The number of sites occupied with spin up and down fermions (but not doublons), as well as holons at a given moment t can be found as

$$N_0(t) = L - N_\uparrow - N_\downarrow + N_d(t) \,, \quad N_\uparrow(t) = N_\uparrow - N_d(t) \,, \qquad N_\downarrow(t) = N_\downarrow - N_d(t) \,. \tag{B33}$$

where L is the system's size. Dividing by L gives the corresponding probabilities $p_\alpha(t) = N_\alpha(t)/L$ for $\alpha \in \{0, \uparrow, \downarrow, \uparrow\downarrow\}$ gives the full m-site marginal IPR estimate:

$$\mathrm{IPR}_{m,\mathrm{est}}^{\mathrm{full}}(t) := \left[\left(\tfrac{N_H(t)}{L}\right)^2 + \left(\tfrac{N_\uparrow(t)}{L}\right)^2 + \left(\tfrac{N_\downarrow(t)}{L}\right)^2 + \left(\tfrac{N_d(t)}{L}\right)^2 \right]^m . \tag{B34}$$

The same doublon-dependent probabilities can be coarse-grained to obtain the charge and spin variants. For the masked charge alphabet $\{H, S, D\}$, the counts are

$$N_H(t) = N_0(t), \quad N_S(t) = L - N_H(t) - N_D(t) = (N_\uparrow + N_\downarrow) - 2N_d(t), \quad N_D(t) = N_d(t) \,, \tag{B35}$$

leading to

$$\mathrm{IPR}_{m,\mathrm{est}}^{\mathrm{charge}}(t) := \left[\left(\tfrac{N_H(t)}{L}\right)^2 + \left(\tfrac{N_S(t)}{L}\right)^2 + \left(\tfrac{N_D(t)}{L}\right)^2 \right]^m . \tag{B36}$$

For the masked spin alphabet $\{\varnothing, U, D\}$, the empty and doublon sites are combined into a single spinless category with

$$N_\varnothing(t) = N_0(t) + N_d(t) = L - (N_\uparrow + N_\downarrow) + 2N_d(t) \,, \tag{B37}$$

while the singly occupied states remain spin-resolved as $N_U(t) = N_\uparrow - N_d(t)$ and $N_D(t) = N_\downarrow - N_d(t)$. This yields

$$\mathrm{IPR}^{\mathrm{spin}}_{m,\mathrm{est}}(t) := \left[\left(\frac{N_\varnothing(t)}{L}\right)^2 + \left(\frac{N_U(t)}{L}\right)^2 + \left(\frac{N_D(t)}{L}\right)^2\right]^m . \tag{B38}$$

IPRs for a 4-site marginal, consisting of four qubits forming a square including the centre of the 5×5 lattice, as well as their estimates obtained from the IID approximation, are shown in Fig. 23. As one can see from Figs. 23b and 23c, this very simple IID approximation provides a good estimate for the charge IPR but not for the full or spin IPR. Similar to other observables, marginal IPR approaches the asymptotic value more slowly for larger values of U.

The IPR on either the full system or on a marginal is also the probability that a pair of samples, each drawn independently from the same system, match as bit strings on all sites. We can express this as the probability of an event $\mathrm{IPR}_m = \Pr(\text{spins and charges match})$, with the understanding that it refers to a single region m of sites and the "match" is an event on two samples from the system. With this we can apply the definition of conditional probability $P(A|B) = P(A,B)/P(B)$ to interpret IPR ratios,

$$\Pr(\text{spins match}|\text{charges match}) = \frac{\mathrm{IPR_m}}{\mathrm{IPR}^{\mathrm{charge}}_m} , \tag{B39}$$

$$\Pr(\text{charges match}|\text{spins match}) = \frac{\mathrm{IPR}_m}{\mathrm{IPR}^{\mathrm{spin}}_m} . \tag{B40}$$

The conditional probabilities allow for examination of thermalisation in the various sectors of spin and charge separately. For example, the decay of Pr(spins match|charges match) quantifies the mixing of spins that is not due to mixing between modes of different charges, i.e. it isolates the thermalisation of spin states of single occupancy sites. Similarly, the decay of Pr(charges match|spins match) quantifies transport in the charge sector — a smaller value of Pr(charges match|spins match) means that a site with a charge carrier is equally likely to be a doublon or holon, indicating thermalisation in the positions of charge carriers. These interpretations are also supported by considering the additive contributions these IPR ratios make to the classical Rényi-2 entropy $S_2 = -\ln(\mathrm{IPR})$.

Results for the IPR ratios and the associated predictions are shown in Fig. 24. The first row demonstrates spin dynamics and thermalisation in the directions of spin on single occupancy sites. We see a relatively weak dependence on U, consistent with the prediction that spin dynamics proceeds even in the presence of strong interactions. The second row demonstrates the transport of charge — given that a site is a doublon or a holon, is it equally likely to be either of those, or does it show a high probability to be one or the other? We see at large U the charge carrier "remembers" its identity, i.e. if a holon is seen somewhere, then a doublon is unlikely to be seen on that site in another sample at the same time.

The ratio $\mathrm{IPR}^{(\mathrm{spin})}_m(t)/\mathrm{IPR}^{(\mathrm{charge})}_m(t)$, shown in Fig. 24c, can be interpreted using the IID approximation above. Since $\mathrm{IPR}^{(\mathrm{spin})}_m/\mathrm{IPR}^{(\mathrm{charge})}_m \approx (N_H N_D)/(N_\uparrow N_\downarrow)$, this ratio is larger than one when the number of holons and doublons is larger than the number of singly occupied sites and smaller than one in the opposite case. This ratio stays close to 1 in the non-interacting case and is less than one for positive values of U, agreeing with the fact that interactions reduce the formation of holon-doublon pairs.

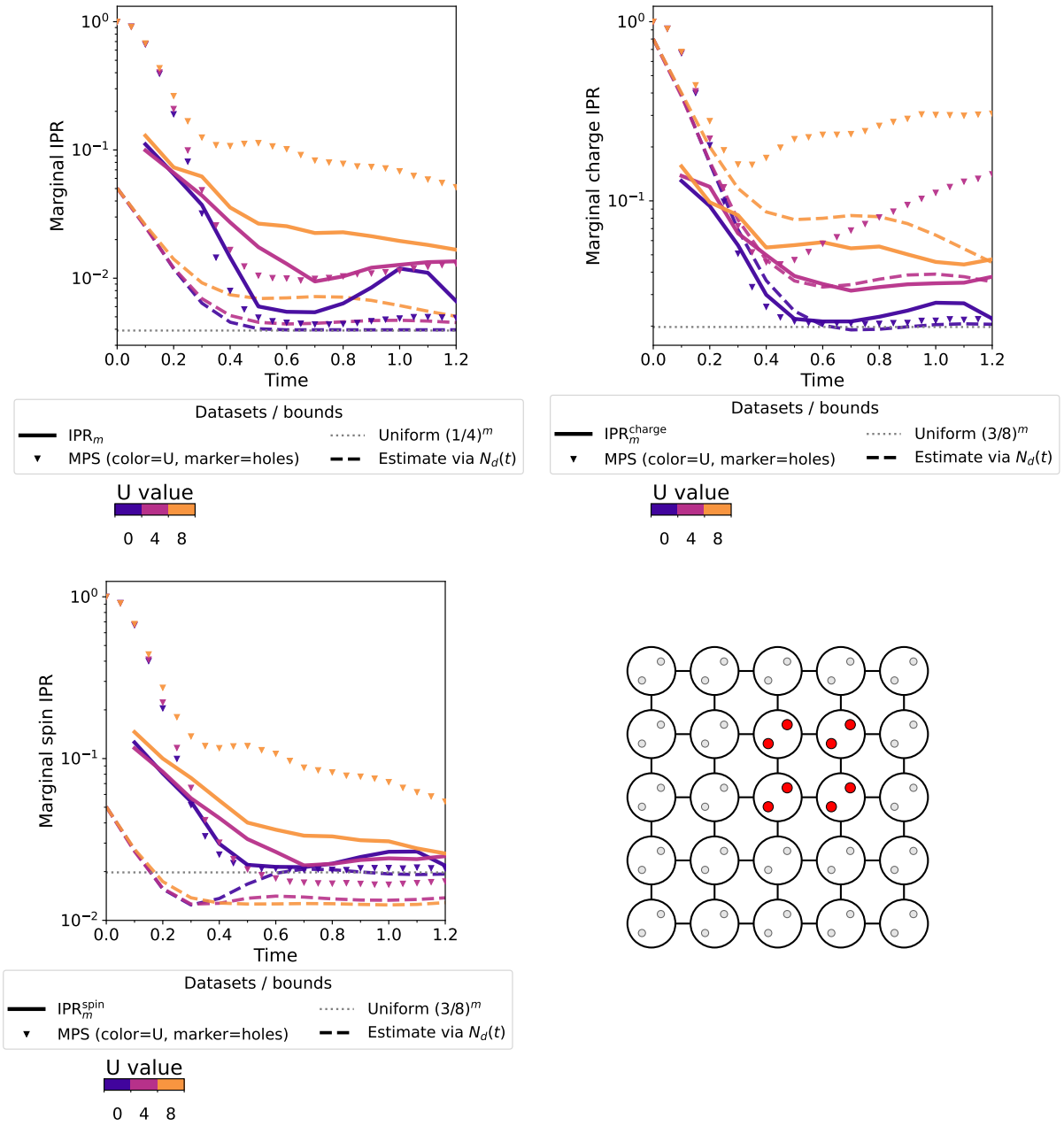
In addition to the hardware data, Figs. 23d and 24 also display results obtained using the TDVP algorithm, which deviate increasingly from the hardware results as the interaction strength U grows.

5. Holon Pair Dynamics

Of particular interest in Fermi-Hubbard type models is the pairwise behaviour of holons due to their joint interaction with the background spin-order [48, 94–97]. The anti-ferromagnetic order of the system is predicted to be disturbed by the motion of a holon, creating a region of spin disturbance around each holon, which is predicted to introduce interactions between holons mediated by this disturbance. Because of the proliferation of holon-doublon pairs in these simulations, it is difficult to observe the relative motion of individual pairs of holons. However, we can investigate the dynamics of the collective pairwise behaviour of holons.

We consider input states which contain a pair of separated holons. We observe the dynamics of the average pairwise distance between holons:

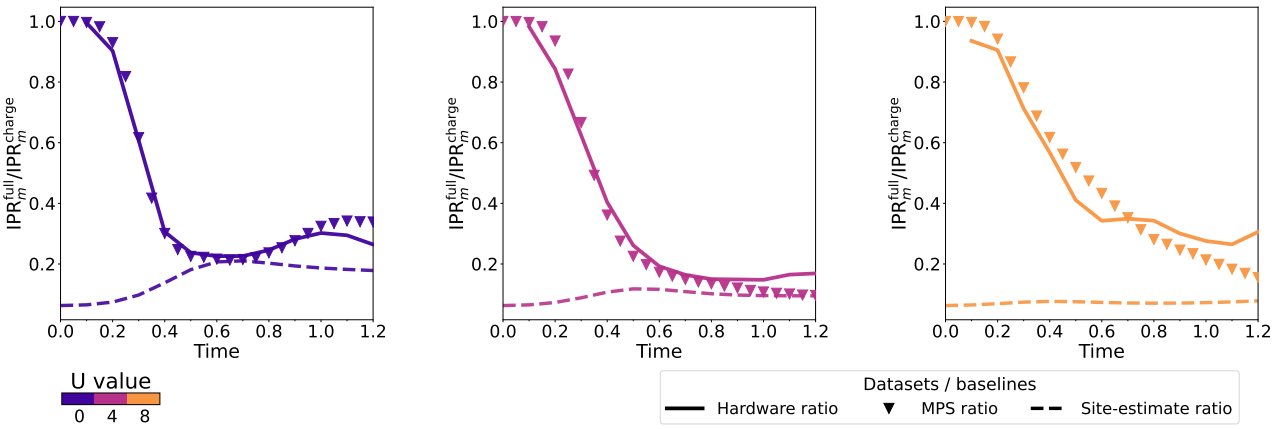
$$\langle r_h \rangle := \frac{\sum_{i<j} r_{ij} \langle h_i h_j \rangle}{\sum_{i<j} \langle h_i h_j \rangle}, \tag{B41}$$



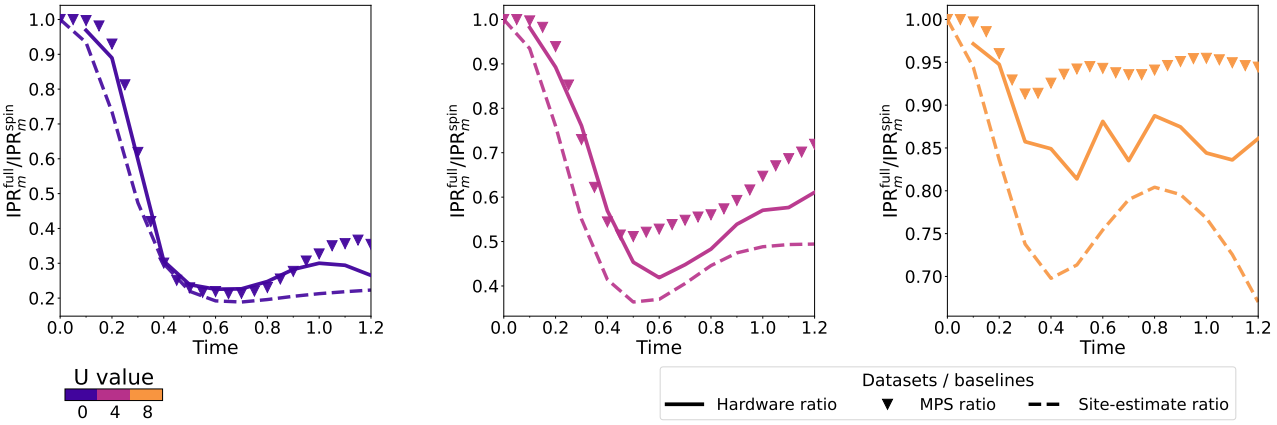
(c) Marginal spin IPR.

(d) Four marginal sites (eight qubits) used for the IPR calculations.

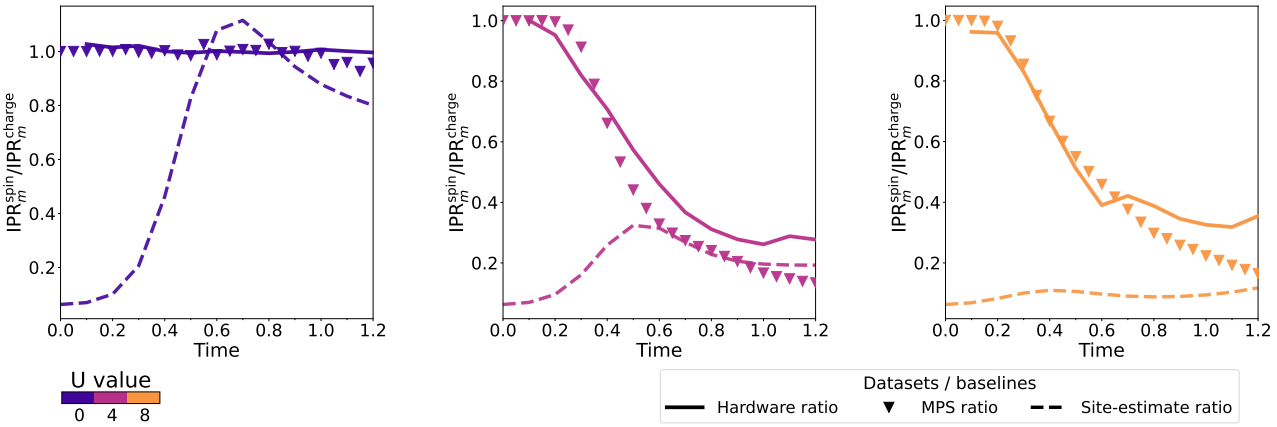
FIG. 23: Marginal inverse participation ratio (IPR) for an $m = 4$-site marginal on 5×5 system, averaged over all one hole states, as defined in Eq. (B32). Marginal IPR characterises the delocalisation of the wavefunction in a fixed basis: it reaches its maximum value of 1 for a wavefunction localised on a single basis state and reaches its minimum 2^{-m} in the maximally mixed state. For the marginal on the original Fock basis $\{0, \uparrow, \downarrow, \uparrow\downarrow\}$ (panel (a) or a spin basis $\{\varnothing, U, D\}$, the IID model used in Eq. (B34) provides a loose agreement at long times for zero and intermediate interaction $U = 0, 4$. For stronger interactions, it is not a good estimate. In panel (b), on the other hand, the marginal charge IPR, where we use the possible states $\{H, S, D\}$, follows very closely the simple estimate Eq. (B36).



(a) $\mathrm{IPR}_m^{(\mathrm{full})}(t)/\mathrm{IPR}_m^{(\mathrm{charge})}(t)$ ratio.



(b) $\mathrm{IPR}_m^{(\mathrm{full})}(t)/\mathrm{IPR}_m^{(\mathrm{spin})}(t)$ ratio.

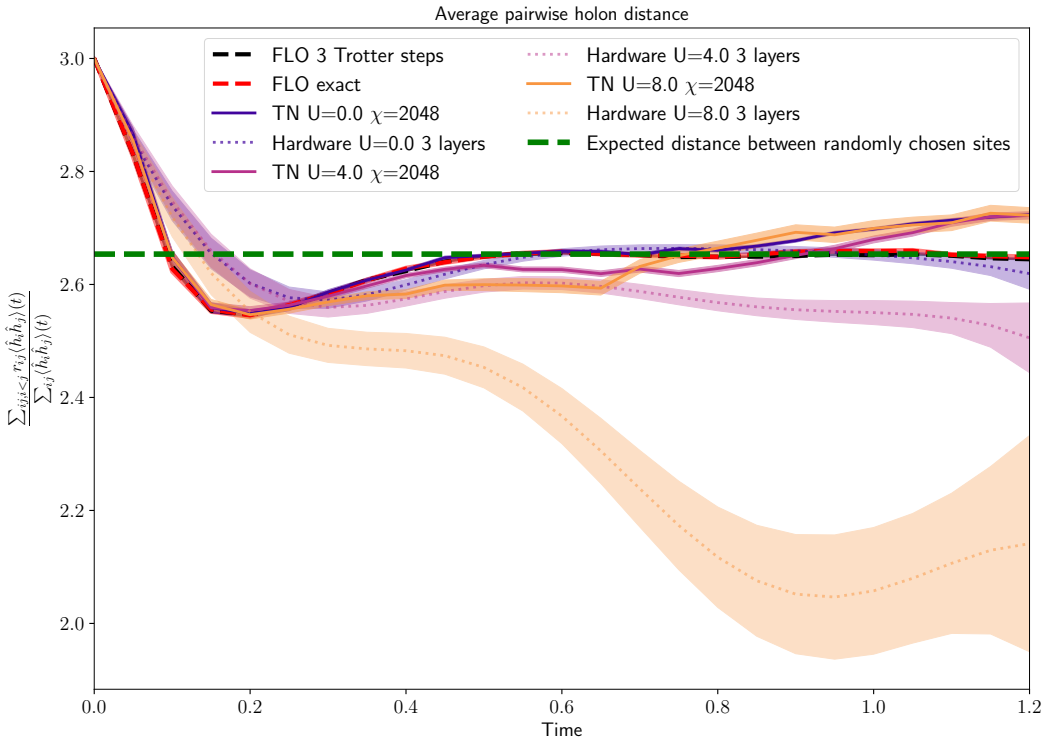


(c) $\mathrm{IPR}_m^{(\mathrm{spin})}(t)/\mathrm{IPR}_m^{(\mathrm{charge})}(t)$ ratio.

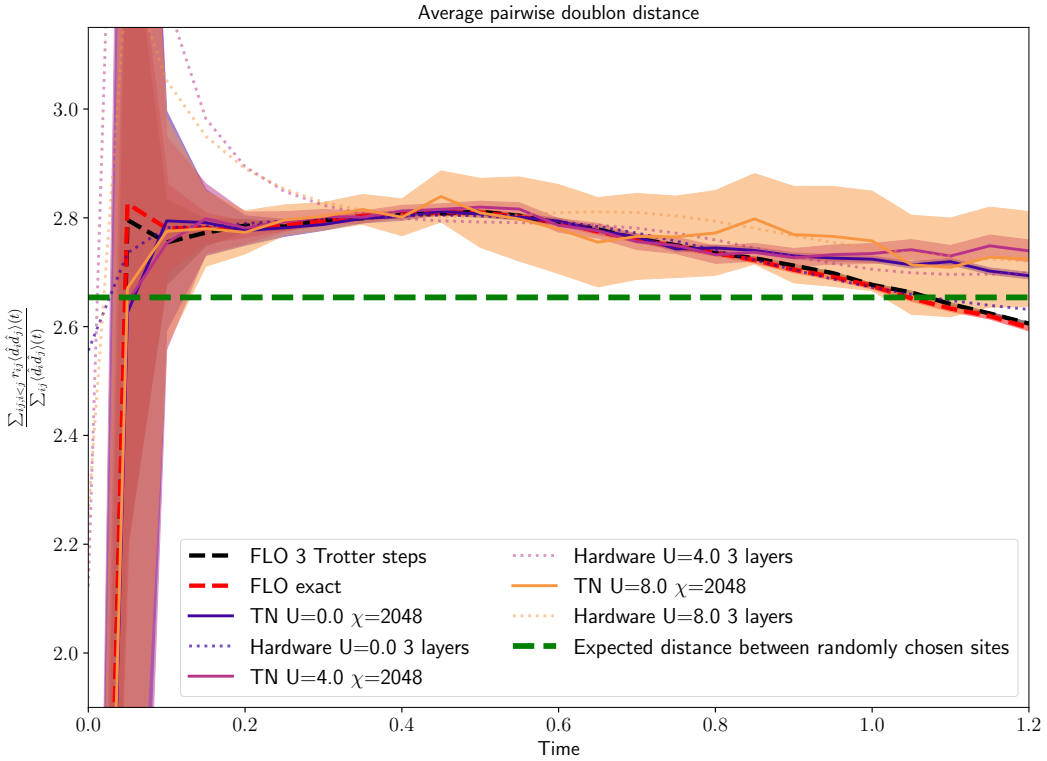
FIG. 24: Ratios of marginal inverse participation ratios defined in Eq. (B32) and shown in Fig. 23 for the 5×5 system.

where $h_i = (1-n_{i\uparrow})(1-n_{i\downarrow})$ and $r_{i,j}$ is the euclidean distance between site i and j, and similarly, the average pairwise distance between doublons:

$$\langle r_d \rangle := \frac{\sum_{i<j} r_{ij}\langle d_i d_j\rangle}{\sum_{i<j}\langle d_i d_j\rangle}, \tag{B42}$$



(a)



(b)

FIG. 25: Dynamics of the average pairwise holon distance $\langle r_h \rangle$ (a) and average pairwise doublon distance $\langle r_d \rangle$ (b) of a 5×5 Néel ordered state with a pair of holons placed $r = 3$ apart. TFLO is applied to all hardware curves; however, GPR is applied to $\langle r_h \rangle$ but not $\langle r_d \rangle$ as it performed poorly. We have retained the error bar style employed by GPR in both curves to improve legibility.

where $d_i = n_{i\uparrow}n_{i\downarrow}$. For an $n \times m$ square grid, the average Euclidean distance r between a pair of randomly chosen non-overlapping points is given by the average distance between distinct cells:

$$\mathbb{E}_{n,m}[r] := \frac{2\sum_{dx=1}^{n-1}\sum_{dy=1}^{m-1}(n-dx)(m-dy)\sqrt{dx^2+dy^2} \;+\; m\sum_{dx=1}^{n-1} dx\,(n-dx) \;+\; n\sum_{dy=1}^{m-1} dy\,(m-dy)}{\binom{nm}{2}}, \quad \text{(B43)}$$

where here the first term counts instances of pairwise distances not aligned along the x or y axis, and the second and third terms count all instances of pairwise distances aligned along an x or y axis.

Figs. 25a and 25b show the average pairwise distances between holons and doublons (resp.) when a pair of holons is introduced to a Néel ordered background on a 5×5 system. We see that for $\langle r_d \rangle$ there is a strong tendency, at every U, for the dynamics to approach $\mathbb{E}_{5,5}[r]$ approach the average distance between two randomly located points in a 5×5 lattice, $E_{5,5}[r] = 2.65$ – illustrated by a green dashed line in Figs. 25a and 25b. This suggests that holon-doublon pairs are appearing in a spatially homogeneous way.

The $\langle r_h \rangle$, on the other hand, behaves quite differently, with the hardware signal starting at 3, corresponding to the distance between the initial holon pair, and then trending away from $\mathbb{E}_{5,5}[r]$ by an increasing degree with the strength of U. The fact that this behaviour does not appear for the average pairwise doublon distance is suggestive that this trend is not a consequence of an inhomogeneity in the background holon proliferation, but rather due to the evolution of the average relative position of the pair of holons consistently introduced at $t = 0$. This implies that the hardware signal shows that, as U increases, the pair of holons has a stronger tendency to move towards one another. We note that the hardware signal and TDVP signal disagree quite dramatically, even at early times, and predict qualitatively different physical behaviour – the TDVP data suggests that U has little to no impact on the average holon distance, and that the average holon distance trends towards random placement of holons for all U. We note that for $U = 0$, the circuit dynamics match both the exact circuit and the exact underlying FLO dynamics at late times, whereas TDVP fails to do so for this signal. We note that the dramatic variation in size of the error bars on these signals is due to normalisation of $\langle r_h \rangle$ and $\langle r_d \rangle$, which can diverge when there is a small number of doublons or holons in the system. This sensitivity may contribute to the large divergence from TDVP – and we expect that, at mid to late times, Trotter error will play a significant role in the $U = 8$ curve. Nevertheless, it is not definitive which method (hardware or TDVP) gives a more accurate qualitative account of the physics.

Appendix C: Quantum Computing Hardware

All experimental results were produced using superconducting quantum computers provided by Google. All devices have degree four connectivity arranged in a diamond lattice. The native two-qubit gates employed are CZ gates using tunable couplers.

The experiments were run on 72-qubit and 105-qubit Willow quantum processors. The qubit layouts, the two-qubit gates calibrated, and typical gate fidelities for a given calibration are shown in Figs. 27 and 26. Gate errors are reported as Pauli errors as described in [74]. Both devices were accessed using Google's Quantum Engine. Device performance was comparable to that reported in [19].

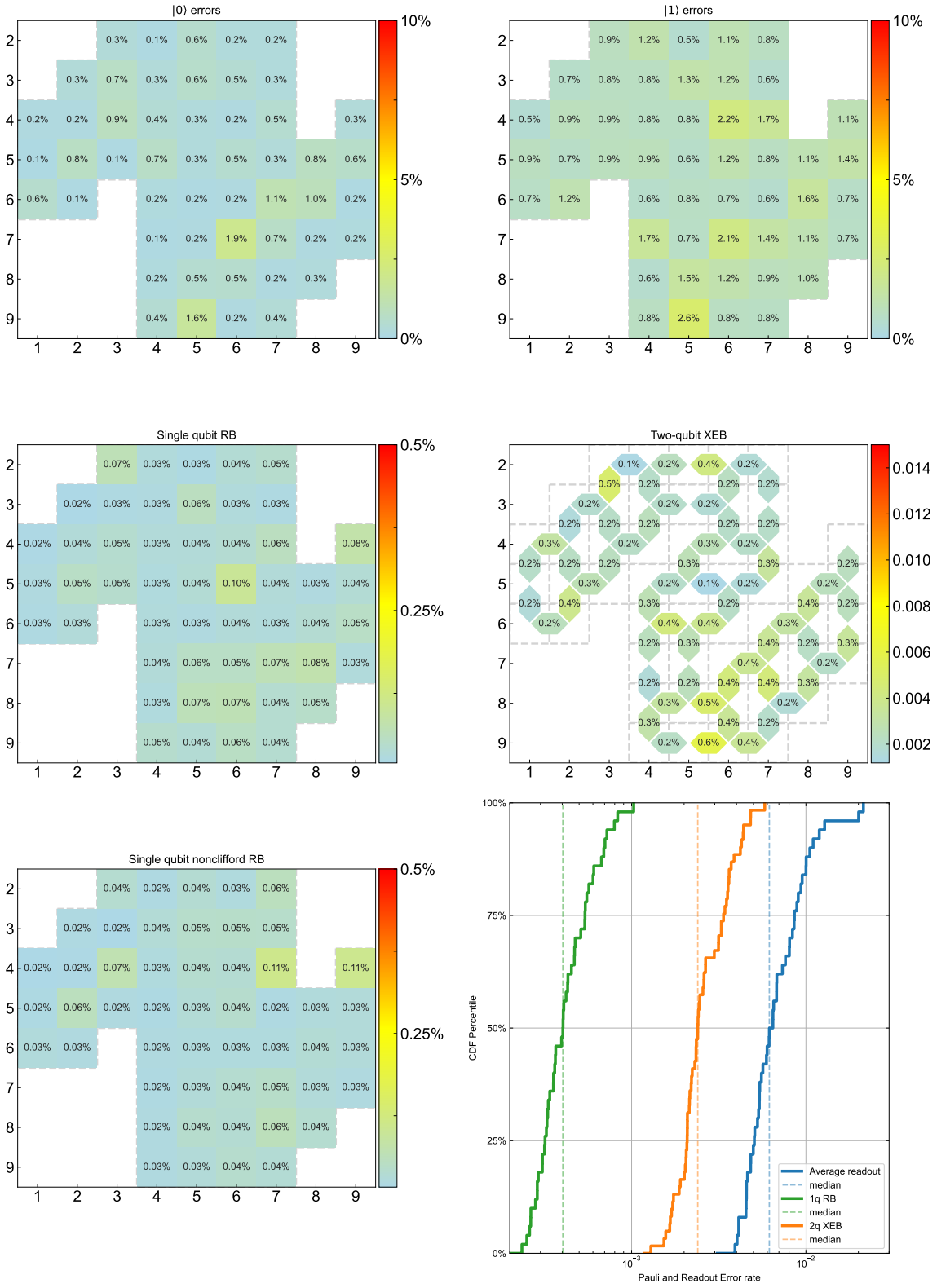


FIG. 26: Benchmark of the qubits and gates used on the 72 qubit superconducting quantum chip using randomised benchmarking (RB) and cross-entropy benchmarking (XEB).

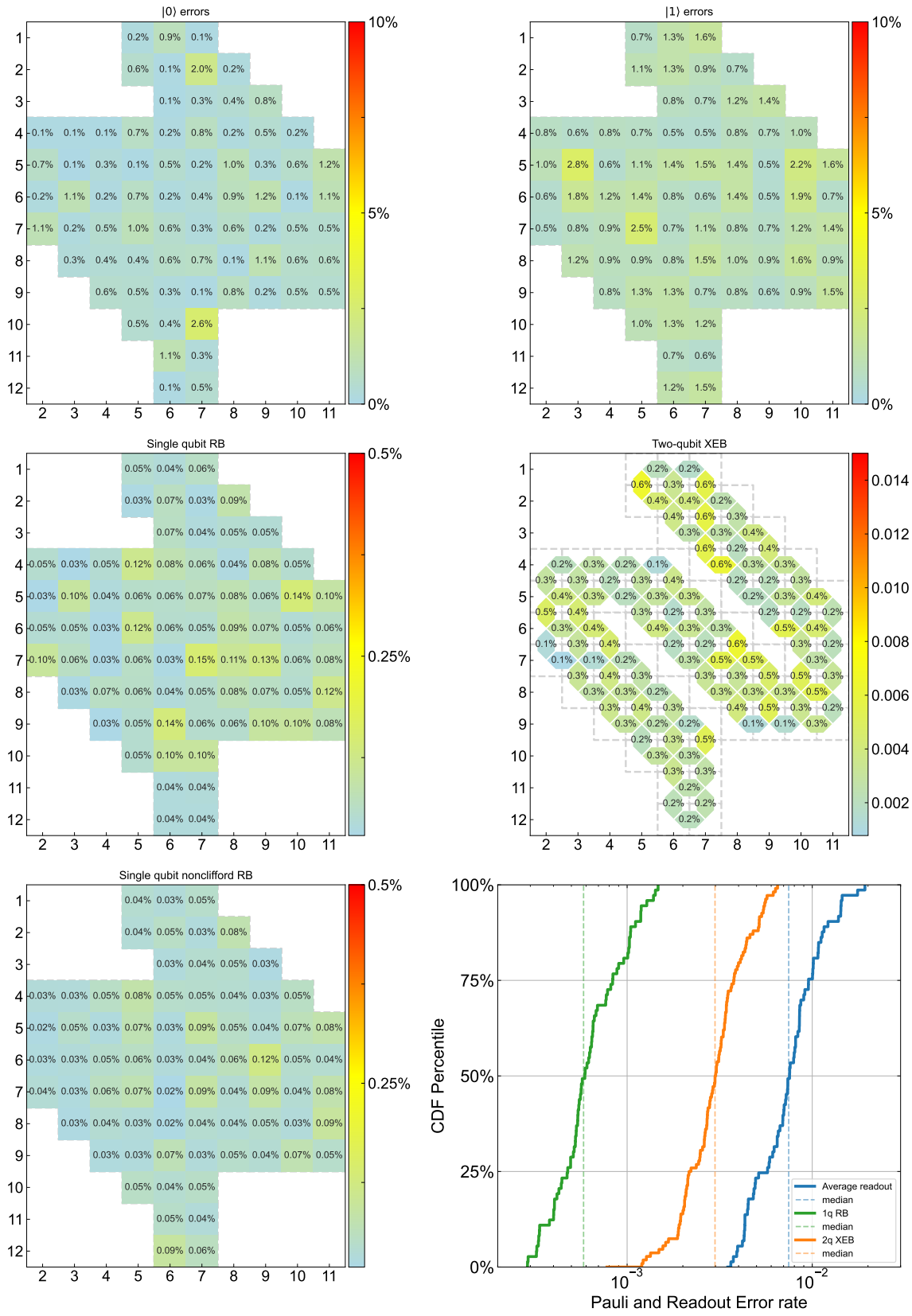


FIG. 27: Benchmark of the qubits and gates used on the 105 qubit superconducting quantum chip using randomised benchmarking (RB) and cross-entropy benchmarking (XEB).

Appendix D: Circuit Descriptions

1. Fermion to Qubit Mapping

We use the Jordan-Wigner (JW) mapping to encode the fermionic system. Each mode is assigned to a qubit and Fock states are represented by computational basis states with $|1\rangle$ ($|0\rangle$) denoting that a qubit's mode is occupied (unoccupied). The modes are assigned an ordering, the same applying to the qubits, and the creation/annihilation operators map to qubit operators

$$c_j^\dagger = \sigma_j^+ \prod_{i<j} Z_i, \quad c_j = \sigma_j^- \prod_{i<j} Z_i, \quad n_j = \frac{1}{2}(1 - Z_j), \tag{D1}$$

with $\sigma_j^\pm = (X \mp iY)/2$. This structure ensures that the encoded fermionic operators satisfy the canonical anticommutation relations. We use the "snake" ordering for modes on a square lattice where modes are ordered along rows from left-to-right and right-to-left in an alternating manner – see Fig. 28. As we do not include any simulations of hopping between spin sectors, one can equivalently think of the encoding of the spin-up and spin-down sectors one after the other on a single JW chain, or otherwise think of them as encoded on two separate JW representations – this does not practically change any of the circuits.

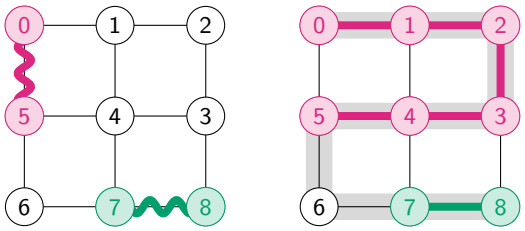


FIG. 28: How the support of hopping terms maps from fermionic modes (left) to qubits (right) under Jordan-Wigner with the snake ordering (highlighted on the qubit side). In the qubit picture, a hopping term is supported on the qubits corresponding to the modes and all qubits between them in the ordering. Horizontal hopping terms (green) are supported on 2 qubits, but vertical hops (magenta) can be supported on longer chains of qubits.

a. *Encoded Interactions*

Hopping interactions are represented as

$$c_i^\dagger c_j + c_j^\dagger c_i = \frac{1}{2}(X_i X_j + Y_i Y_j) \prod_{i<k<j} Z_k. \tag{D2}$$

The onsite Coulomb terms of the Fermi-Hubbard Hamiltonian are represented on qubits as

$$n_{i,\uparrow} n_{i,\downarrow} = \frac{1}{4}(1 - Z_{i,\uparrow} - Z_{i,\downarrow} + Z_{i,\uparrow} Z_{i,\downarrow}). \tag{D3}$$

When summed over every spin pair, $\sum_i (1 - Z_{i,\uparrow} - Z_{i,\downarrow})$ is a constant when restricted to a fixed particle number subspace, so this part may be ignored and it suffices to encode the Coulomb term as

$$n_{i,\uparrow} n_{i,\downarrow} \approx \frac{1}{4} Z_{i,\uparrow} Z_{i,\downarrow}. \tag{D4}$$

In the presence of a magnetic field, the hopping terms pick up phase factors as in Eq. (A1). We simulate a π flux perpendicular to the lattice, which corresponds to setting $\phi_{ij} = \pi$ for a single edge of every plaquette in the Fermi-Hubbard lattice, e.g. every other row of horizontal edges. This only adds a minus sign to the corresponding hopping terms, so they are treated the same as the unmodified hops except for a sign flip on the t parameter in the circuits in Appendix D 3.

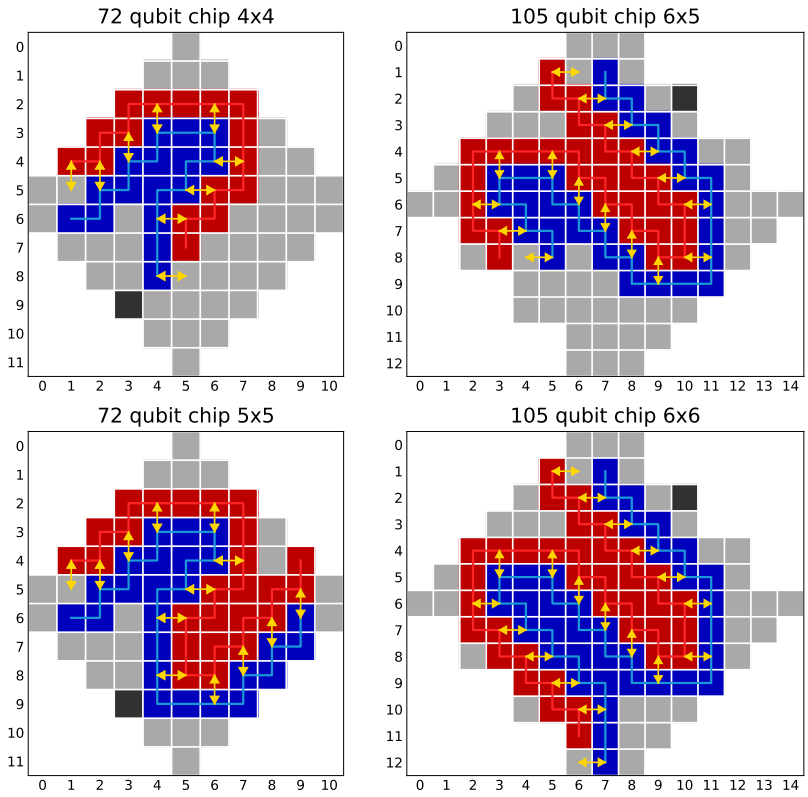


FIG. 29: Layout of qubits on the chips used. Qubits highlighted in red (blue) correspond to encoded spin-up (spin-down) fermionic modes. Dead qubits are highlighted in black. The Jordan-Wigner ordering is indicated by the lines running through each spin sector. The yellow arrows show where FSWAPs are applied to bring together qubits corresponding to the spin-up and spin-down parts of a site in preparation for evolving under the Coulomb interactions. FSWAPs suffice instead of SWAPs, as the encoded Coulomb interaction is a ZZ Pauli on the two qubits.

b. *Layout On Device*

Despite the grid structure of the simulated system, the operations for hopping between neighbouring sites (discussed in Appendix D 2) only act between qubits adjacent in the JW ordering. This means that the qubits for each spin sector need only be connected to their neighbours in the ordering, with the additional restriction that qubit pairs corresponding to spin pairs be nearby to minimise the cost of bringing them together for the 2-qubit Coulomb interaction. Fig. 29 shows the on-chip qubit layouts used in this work.

2. Second-Order Trotter Circuit

Our second-order Trotter step of time t has the basic structure

$$S_2(t) = \prod_{h\in\mathcal{H}}^{\rightarrow} e^{-ih\frac{t}{2}} \prod_{o\in\mathcal{O}} e^{-iot} \prod_{h\in\mathcal{H}}^{\leftarrow} e^{-ih\frac{t}{2}}, \tag{D5}$$

where $\mathcal{H}$ is the set of hopping interactions in the Hamiltonian and $\mathcal{O}$ is the set of Coulomb interactions. The first round of hopping interactions is applied using an FSWAP network, which will be explained below and the second round of hopping interactions is applied in reverse order as indicated by the arrows. The Coulomb terms are implemented in parallel, see Appendix D 3 for the circuit. Due to the layout, not all spin pairs are adjacent with respect to the device connectivity, so swaps must be applied to bring them together (see Fig. 29). It suffices to use FSWAPs to do this because all that is required is the transformation of a ZZ on two non-adjacent qubits to one on two adjacent qubits. This reduces the gate cost as the FSWAP can be applied with fewer gates than a qubit SWAP. After this, the Coulomb terms are applied and the FSWAPs are undone.

a. Swap Network

As noted in Fig. 28, horizontal hopping terms are supported on only two qubits and can be implemented efficiently (see Appendix D 3) while vertical hopping terms may involve long strings of Z operators, which can be costly. One way around this is fermionic swap networks [98], in which FSWAP operations

$$\text{FSWAP} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} = \text{SWAP} \cdot \text{CZ}, \tag{D6}$$

are used to rearrange fermionic modes to shorten these Z strings.

FSWAP transforms Paulis under conjugation like

$$XI \leftrightarrow ZX, \quad YI \leftrightarrow ZY, \quad ZI \leftrightarrow IZ, \tag{D7}$$

and in the fermionic picture, when acting on qubits adjacent in the ordering, it transforms encoded fermionic operators like

$$c_j \leftrightarrow c_{j+1}, \quad c_j^\dagger \leftrightarrow c_{j+1}^\dagger, \quad n_j \leftrightarrow n_{j+1}, \tag{D8}$$

swapping the positions of the encoded modes. From this, one sees that if two modes are made adjacent via FSWAPs, then the encoded hopping term between them is transformed into a two-qubit term with no Z string.

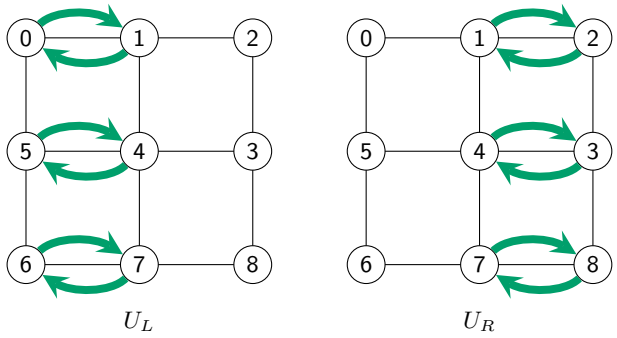


FIG. 30: Action of U_L and U_R on a 3×3 grid. The green arrows indicate FSWAPs.

Given a lattice of dimensions L_x and L_y, we define as U_L (U_R) (see Fig. 30) the circuit that FSWAPs even (odd) columns with those on the right, starting with column 0 (1). As discussed in Appendix D 3, where a hopping term needs to be applied to the pairs of modes being swapped, this can be done by simultaneously applying both hop and FSWAP without increasing the circuit cost. For any given Trotter step size t, the hopping terms are implemented — for time $t/2$ — according to the following:

Even L_x:

1. Implement all vertical hopping terms between adjacent modes.
2. Apply U_L
3. Implement all vertical hopping terms between adjacent modes.
4. Apply U_R
5. Repeat from 2. terminating when all hopping terms have been implemented.

Odd L_x:

1. Implement all vertical hopping terms between adjacent modes.
2. Apply U_L.
3. Apply U_R.

4. Repeat from 1. terminating when all hopping terms have been implemented.

In both cases, the FSWAPs in the U_L and U_R circuits should be replaced with a merged FSWAP and hop if the hopping term between the mode pair has not yet been implemented.

After the Coulomb interactions, the second round of hopping terms is implemented by applying the same gates as the first round, only in reverse order. See Fig. 31 for an example illustration of a full step.

This schedule results in fewer swaps than a typical swap network [98], which restores the modes to their original positions at the end. Here, the modes are left in a jumbled state and the second round of hops is applied by reversing the schedule of the first round, restoring the initial layout before the following step. This jumbling does not affect the Coulomb interactions, as every spin-pair moves together.

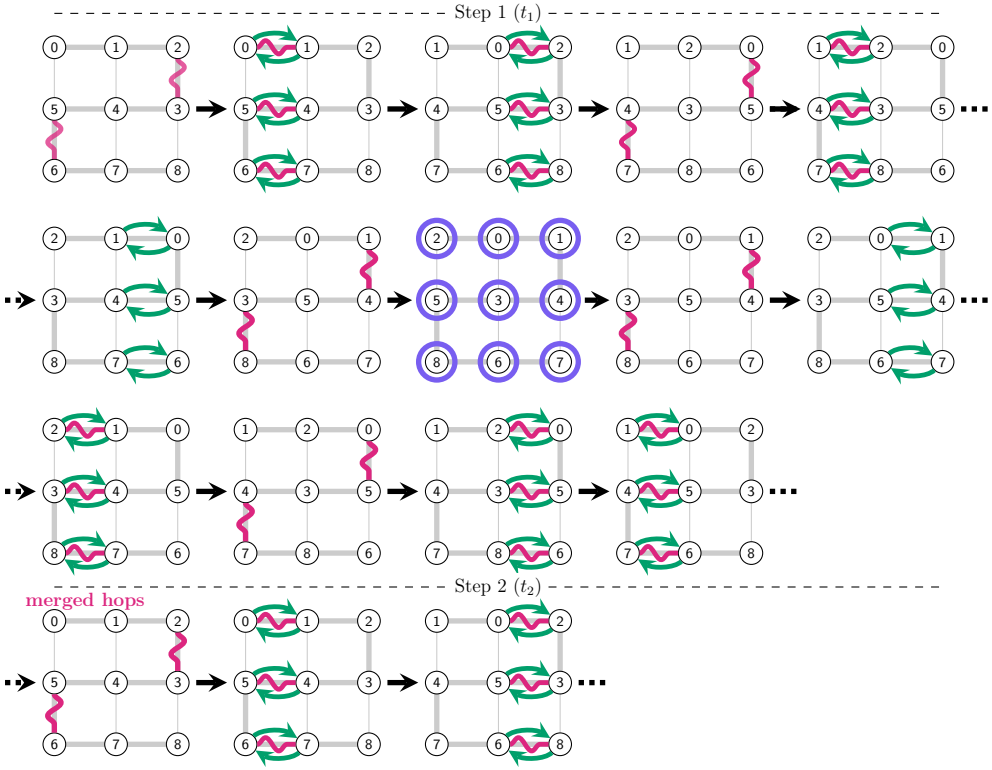


FIG. 31: Illustration of the first Trotter step and the start of the second on a 3×3 lattice. Identical operations are applied to both spin sectors, so only spin-up is shown. Green arrows indicate FSWAPs, magenta wavy lines indicate hopping interactions. All hopping terms are for time $t_1/2$ ($t_2/2$) in step 1 (2), except when merging is noted. Hops and FSWAPs on the same lattice indicate simultaneous implementation and the violet circles indicate Coulomb interactions for time t_1 (t_2) with the corresponding spin-down modes. Labels indicate the location of the encoded fermionic modes at each stage. The final stage of step 1 and the first stage of step 2 are merged, so the hopping interactions are applied for time $(t_1 + t_2)/2$.

b. Trotter Step Merging

The symmetric structure of a second-order Trotter step offers an opportunity to reduce gate cost by merging repeated actions across steps. In particular, a Trotter step ends with the application of all vertical hopping terms adjacent in the ordering, which is then immediately followed by the same action at the start of the following step. In a circuit with N steps with step $j \in (1, \ldots, N)$ running for time t_j we omit these from the end of every step $j < N$ and at the beginning of each step $k > 1$ we have these terms evolve for time $(t_k + t_{k-1})/2$ rather than $t_k/2$, reducing the number of gates required.

This is illustrated in Fig. 31.

3. Gate Decompositions

All operations required for the Trotter circuit can be implemented with only single-qubit Pauli rotations and CZ gates, which are native to the hardware. We use the following notation for single-qubit rotations:

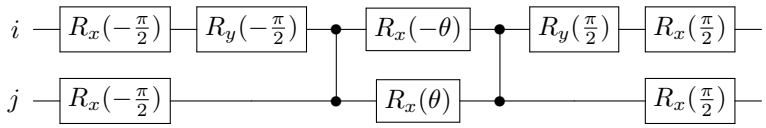
$$R_\sigma(\theta) = e^{-i\sigma\frac{\theta}{2}}. \tag{D9}$$

a. *Interactions*

We use the "fsim" gate [99] as a primitive in our circuits:

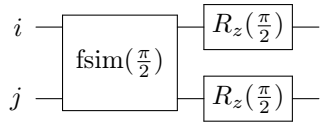
$$\mathrm{fsim}(\theta) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta) & -i\sin(\theta) & 0 \\ 0 & -i\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

which decomposes into the native gate set, acting on qubits i and j, as

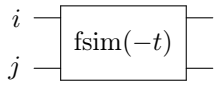


The operation is symmetric under the exchange of qubits.

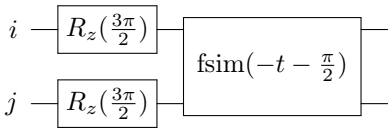
FSWAP is implemented with the circuit



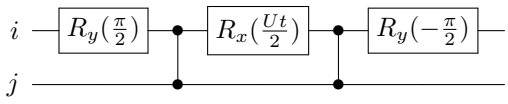
evolution under a hopping term for time t between two modes adjacent in the JW ordering is implemented with



and a hopping term merged with an FSWAP

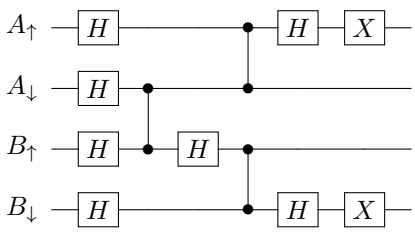


in cases where a magnetic field flips the sign of a hopping term, we use the same circuit, only with the sign of t flipped. Evolution under the Coulomb term for time t with strength U is implemented with



b. *Singlet Initial States*

Singlets between lattice sites A and B are constructed with the following circuit



	This Work		Standard Swap Network		3 Trotter steps		
System Size	2q Gate Count	Depth	2q Gate Count	Depth	2q Gate count	Depth	# qubits
4×4	$400N + 20$	$64N + 6$	$516N + 32$	$72N + 6$	1220	198	34
5×5	$862N + 40$	$88N + 6$	$1022N + 40$	$96N + 6$	2626	270	51
6×5	$1040N + 52$	$88N + 6$	$1236N + 48$	$96N + 6$	3172	270	62
6×6	$1440N + 52$	$96N + 6$	$1756N + 72$	$104N + 6$	4372	294	74

TABLE I: Total number of 2-qubit quantum gates and gate depth for a circuit with N Trotter layers and concrete numbers for our deepest circuits with 3 Trotter steps. Costs are given after an automated optimisation pass, which may remove additional gates from our manually optimised circuits. This optimisation compiles the circuit into alternating layers of 1 and 2-qubit gates, so 2-qubit gate depth can be taken as half the depth. Measurement at the end of the circuit requires a final layer of single-qubit gates, which is included in these costs. Note that these numbers are the same for all initial states except for the singlet initial state, which requires two more layers of 2-qubit gates.

Where $A_\uparrow$ is the qubit representing spin-up for the site A and so on. Due to the arrangement of the up and down qubits on the device, the qubits required for some singlets are not necessarily adjacent on the quantum processor. As a result, SWAP or FSWAP gates must be inserted. FSWAPS can only be used when any adjacent singlets have not already been entangled and all qubits involved in the swap are in the $|0\rangle$ state.

The Hadamard gate H decomposes into single qubit rotations as

$$-\boxed{H}- = -\boxed{R_y(\tfrac{\pi}{2})}-\boxed{R_x(\pi)}-$$

and the X gate decomposes as

$$-\boxed{X}- = -\boxed{R_x(\pi)}-$$

4. Total Gate Costs

Shortening the swap network by exploiting the symmetry of the second-order Trotter step leads to significant savings in gate costs versus a standard swap network, which returns modes to their original positions as in [98]. Table I compares 2-qubit gate count and total gate depth for these two methods. For fairness, any gate savings from merging between Trotter steps as described in Appendix D 2 b are included when costing the standard swap network as well as our circuit.

5. Comparison with Compact Encoding and native fsim

We include here a comparison between the circuit costs of our experiment against potential alternatives for simulating 3 second-order Trotter layers of a square Fermi-Hubbard model. Specifically, we consider the comparative cost of employing a different fermionic encoding – the compact encoding [100] – and also the potential cost of using variable-angle FSim gates.

One of the major circuit overheads of our experiment is deploying the fermionic swap networks on the JW encoding which causes circuit depth and gate count for our circuits to respectively scale linearly and quadratically with system size. There exist local fermionic encodings whose operator mappings – an example being the compact encoding. Here we demonstrate that the JW encoding remains the more cost-effective choice at the scales of these experiments. However, we note that we are right at the cusp of a cross-over point for 6×6 in terms of both two-qubit gate cost and circuit depth when using CZ gates, as illustrated in the right-hand insets of Figs. 32 and 33. We note, however, as illustrated in Fig. 34, that were we to use the compact encoding for 6×6, we would require more qubits than were available to us on the device at the time of this work.

The native gateset employed for this experiment was the CZ gate. While the Willow quantum processor can, in principle, have a variable-angle fsim gate tuned up we chose to keep the CZ gate due to its high fidelity. However, looking towards future experiments, we include here an illustration of the potential improvements an arbitrary angle fsim gate can produce in terms of gate depth and count, as seen in the left-hand insets of Figs. 32 and 33. In this case, the crossover point for gate count is an 8×8 FH model, and for gate depth a 7×7 FH model – requiring 200 and 150 qubits respectively. We note that these figures do not change dramatically even if the variable angles are

kept to a small number of fixed values, such as, for example, a native FSWAP and CZ. This motivates work targeting a high-quality tune-up of a small number of different angles of the fsim gate on the device.

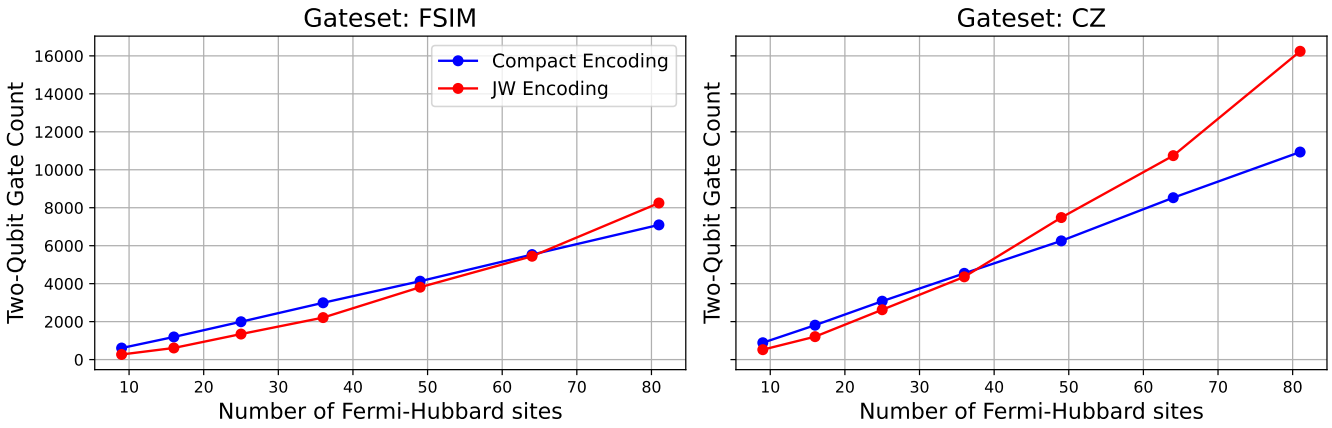


FIG. 32: Two-Qubit Gate Counts for square lattices with 3 second-order Trotter steps, comparing compact and JW encodings.

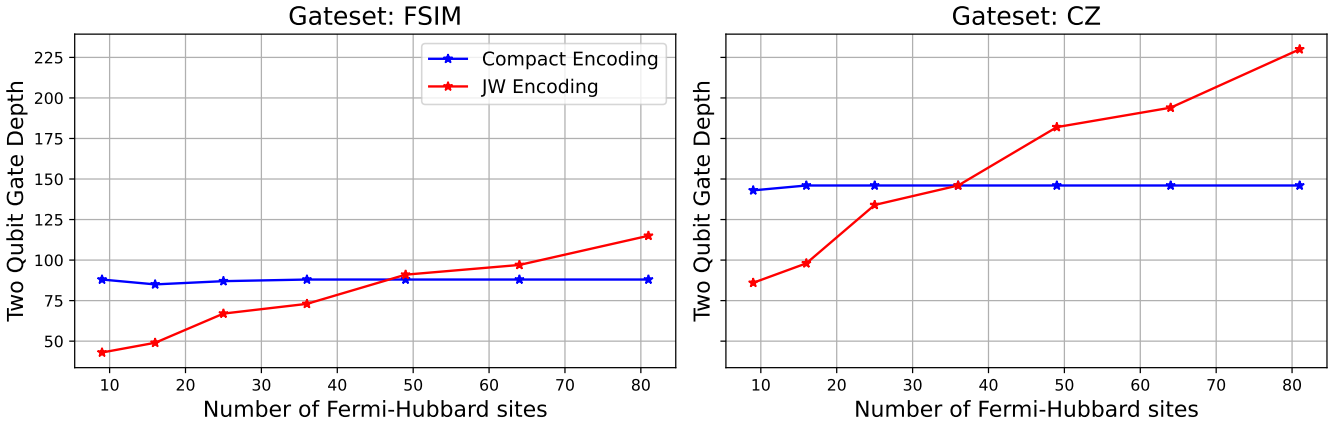


FIG. 33: Two-qubit gate depth for square lattices with 3 Trotter steps, comparing compact and JW encodings.

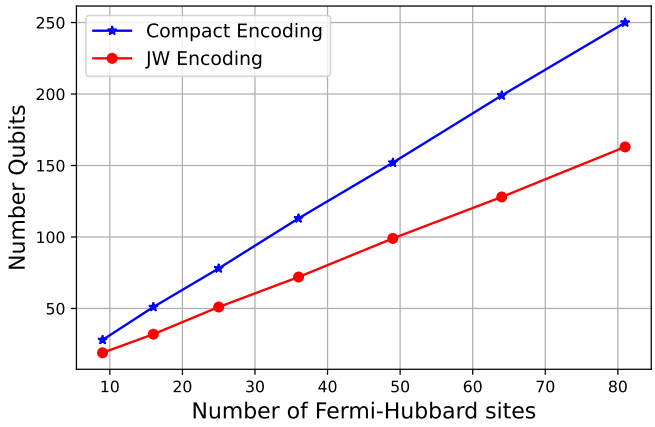


FIG. 34: Number of qubits required for square lattices with 3 Trotter steps, comparing compact and JW encodings.

Appendix E: Error Mitigation Techniques

A number of known error mitigation techniques were tested and compared to optimise the performance of the experiment. These included Pauli and readout twirling, post-selection, training on fermionic linear optics (TFLO), Gaussian process regression (GPR), averaging over symmetries, and probabilistic error amplification / zero-noise extrapolation (PEA+ZNE). We also introduce a novel error-mitigation method based on reweighting samples, called Maximum Entropy Shot-Reweighting. In this section, we outline these methods and present a comparative analysis of their performance. In the final experiments, we used all of these methods except for PEA+ZNE, which we found performed less well than TFLO and used significantly more resources. For most experiments, unless otherwise stated, we used Pauli and readout twirling, post-selection, TFLO, GPR, and averaging over symmetries. We used sample reweighting when TFLO and GPR were not available (e.g. when not computing an observable).

The overall effect of the various methods that we implemented is shown in Fig. 35. We find that error mitigation is less effective at higher times, which we attribute to the observables being less polarised, so the noisy observables are closer to the true values than at lower times. Nevertheless, in all cases, the combination of TFLO and GPR is able to significantly reduce the worst-case errors.

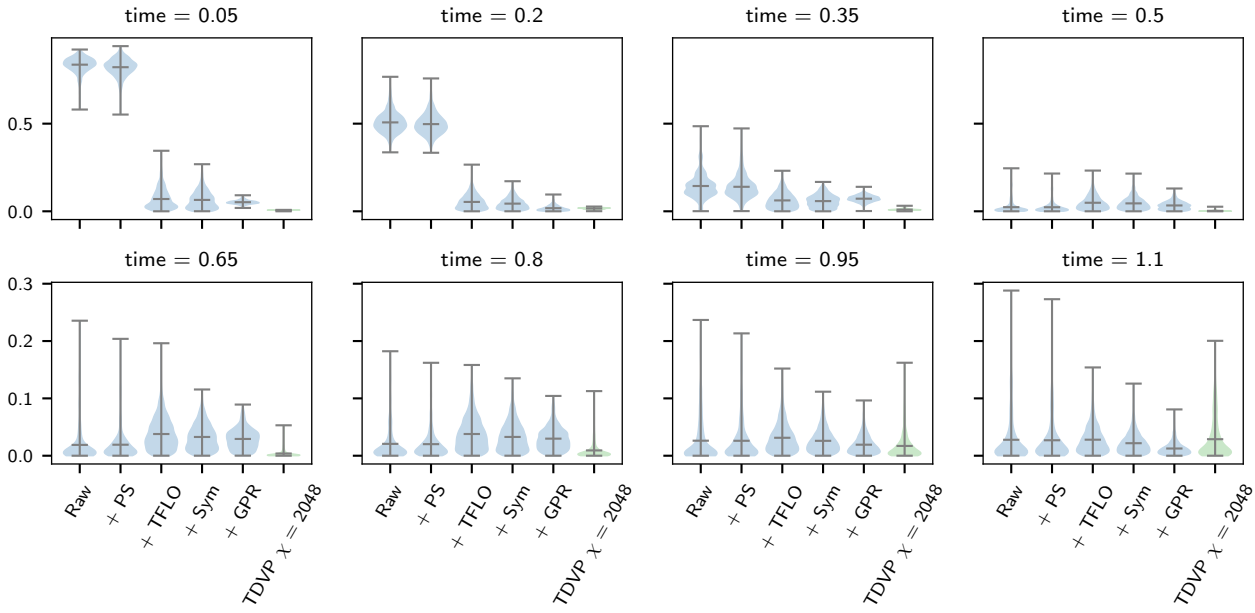


FIG. 35: Error profile across all single and two qubit Z expectation values across different stages of the error mitigation pipeline and for TDVP at $U = 0$ for a 6×5 system. PS: postselection; Sym: symmetry averaging.

1. Noise averaging and suppression

To simplify the noise model for further error mitigation and also suppress it as much as possible on the hardware level, we implemented different noise averaging and suppression techniques; namely, Pauli twirling, dynamical decoupling and read-out twirling.

Pauli twirling [101] is a simple and effective technique to suppress and simplify error models in quantum circuits. The main assumption going into it is that the main source of errors are two-qubit gates and that these are all Clifford gates (such as the CNOT or CZ gate). Each such Clifford C gate gets pre-multiplied with a randomly chosen 2-qubit Pauli gate P and post-multiplied by an appropriate 2-qubit Pauli gate Q such that the ideal gate C remains unchanged, i.e. such that $PCQ = C$. However, in the presence of noise, this sandwiching has two advantageous effects: It simplifies the noise model by making the transfer matrix of the noise gate diagonal in the Pauli basis, and it turns coherent errors into incoherent ones such that their effect accumulates with the square root of the number of noisy gates instead of linearly.

Besides Pauli twirling, we also implemented dynamical decoupling by inserting X gates at the beginning and end of all idling sequences. This has the effect of flipping the qubit state and thus causing phase errors accumulated during one half of the idle period to cancel those from the other half. This refocusing mechanism is analogous to a spin echo in nuclear magnetic resonance.

For the read-out stage, Pauli twirling gets modified to read-out twirling with random X gates [102]. This is particularly advantageous because readout on superconducting qubit hardware is affected by biased errors. Instead of pre-multiplying the readout operation with randomly chosen Pauli gates, we pre-multiply it with randomly chosen X gates and save a bitstring corresponding to the bitflip pattern effected by the X gates. The effect of the random X gates can then be undone in post-processing by xor-ing the measurement results with the saved bitstring.

The Pauli and readout twirling leave the two- and single-qubit structure of our circuits invariant—they only insert new single-qubit gates adjacent to pre-existing single-qubit gates. This enables simple compression by merging all runs of adjacent single qubits into one general, single-qubit `cirq.PhasedXZGate` such that the final circuits are alternating layers of CZ gates and `cirq.PhasedXZGate`s. This fixed-layer structure further enables the use of parametric circuits. Instead of submitting one circuit per value of U, t and twirling instance to the Google Quantum Cloud API we can submit only one parametric circuit for each number of Trotter layers and system size and all information about the twirling protocol, evolution time t and interaction strength U enters only into the parameters of the `PhasedXZGate`s and can be submitted via the API for parametric circuits. We found that this massively sped up our experimental runs and brought down run-times from 10+ hours for a full run down to less than an hour.

2. Post-Selection

The symmetries present in the Hamiltonian imply that particle number is conserved in each spin sector of the model. For all input states in the simulations, the particle number in each spin sector is a good quantum number, allowing post-selection on these conserved quantities. We note that deviation from the true sector follows a Gaussian profile, as shown in Fig. 36. This hints that errors have a local effect on the desired output state, as expected, and hence that local observables will be less affected by noise than if the error were well modelled by a completely depolarising channel.

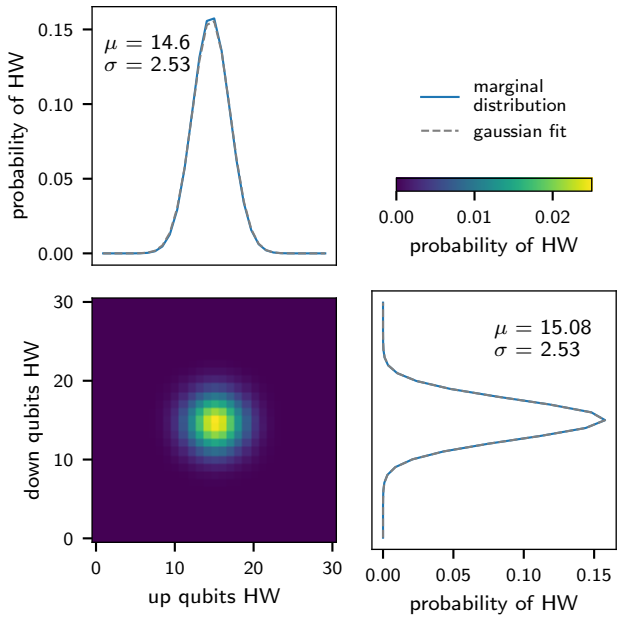


FIG. 36: The distribution of Hamming weights on the up and down qubits is well described by a Gaussian centred at $(14.6, 15)$ (the correct Hamming weight in the absence of hardware errors). 2.4% of shots have exactly Hamming weight $(14, 15)$. The data presented here is from the same 6×5 run also presented in Fig. 70. For 6×6 runs the probability of a sample being in the initial Hamming weight sector further drops to 1.9%.

For different system sizes and initial states, we observe differing performance when using the post-selection to improve the signal in calculated expectation values. The differences become particularly notable when using these

expectation values in the rest of our error mitigation pipeline. Generally, we see that full post-selection on the correct particle number per spin sector is useful for the smallest system size 4×4, whereas for the larger system sizes, the signal obtained improves when keeping shots with up to 8 Hamming weight errors, or more. Nevertheless, it is insightful to see how the acceptance fraction of shots retained during post-selection changes as we increase the number of Trotter layers and system size. This is explored in Fig. 37.

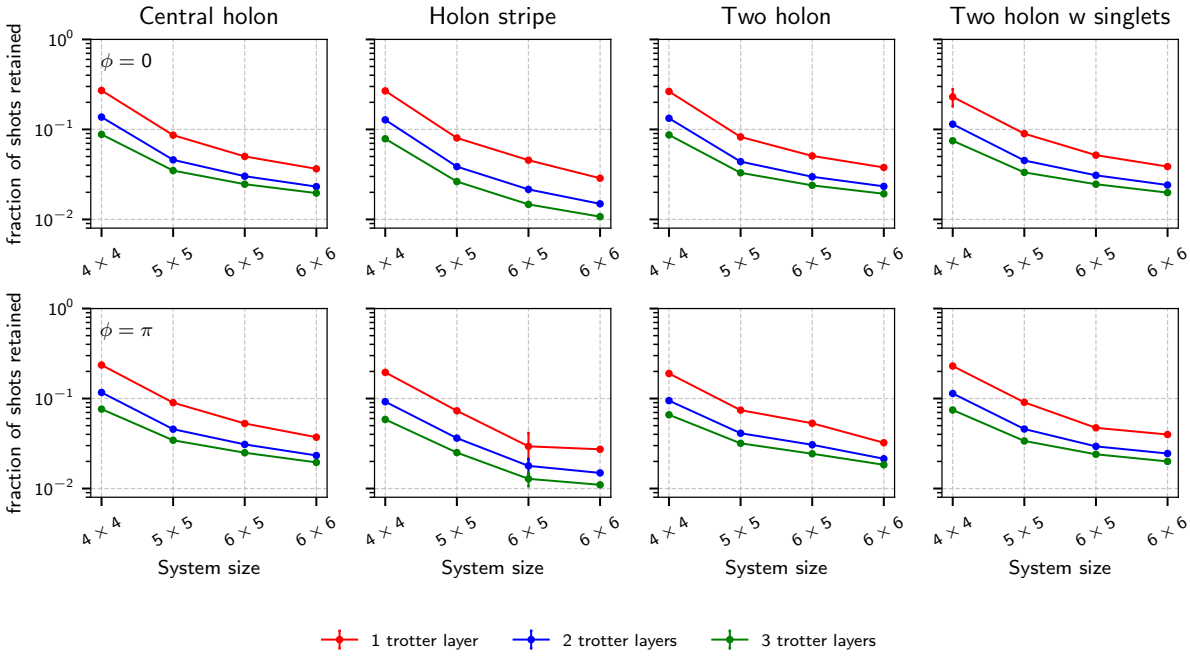


FIG. 37: The median fraction of accepted shots after post-selecting on the correct Hamming weight per spin sector. The median is taken across all times and all values of U and demonstrates how the fraction of retained shots changes as we increase the number of Trotter layers and the system size.

The best post-selection hyper-parameter, the maximum Hamming weight error to accept per shot, is estimated by benchmarking the performance of the TFLO error-mitigated expectation values on the $U = 0$ data for all initial states and system sizes. In Fig. 39, we show how the mean, median and maximum absolute error change as we vary the post-selection hyper-parameter. We see that the training error ($U \approx 0$) shows a similar trend to the test error ($U = 4, 8$), which motivates using the training error in other systems as a proxy for expected performance overall when the test error is not accessible.

3. Training on Fermionic Linear Optics (TFLO)

The dynamics of the non-interacting ($U = 0$) Fermi-Hubbard model can be efficiently simulated by fermionic linear optics (FLO) [57]. The TFLO error mitigation method [20] leverages this fact by training a noise model on the output of noisy quantum simulations of the non-interacting Fermi-Hubbard model against the noiseless output of an efficient FLO simulation. The arguably simplest model to use is an affine linear fit of the form

$$O_{\text{exact}} = mO_{\text{noisy}} + b \tag{E1}$$

which is the ansatz suggested in [20] and found to be effective in [67] in the context of mitigating VQE expectation values. In simulable small-scale experiments, we additionally found a significant dependence of the noisy value on the evolution time t (see Fig. 38 below). We hence modified the fitting ansatz to use the form

$$O_{\text{exact}} = mO_{\text{noisy}} + ct + b \tag{E2}$$

which we found to work slightly better in the context of time dynamics simulations (see Figs. 42 and 39). We call the former method "linear" and the latter method "linear-linear".

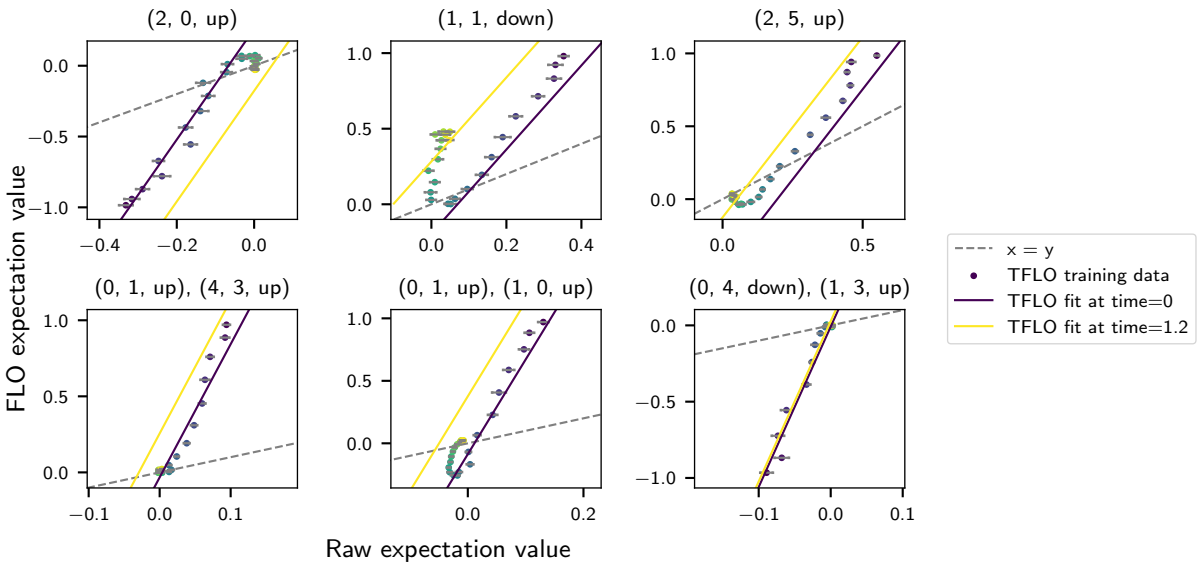


FIG. 38: A few examples of the linear-linear fits performed for TFLO error mitigation for 6×5. The top row shows a selection of single-qubit Z expectation values, while the bottom row shows some two-qubit Z expectation values. We see that the fit slope for the two-qubit fits is larger than for the single-qubit fits, and that the product of the single-qubit slopes often matches the two-qubit slope reasonably well.

To do error mitigation with this model, we performed a linear least squares fit with Eq. (E2) using the device data at $U = 0$ as O_{noisy} and the FLO simulation data as O_{exact}. For each observable, the optimal fit parameters c^*, m^* and b^* are then used together with the $U \neq 0$ data to predict mitigated values according to Eq. (E2). The TFLO training set is generated by using the exact same time dynamics simulation parameters as in the non-interacting cases, while setting $U \approx 0$ in the circuits, such that the compiler does not optimise away the gate.

As discussed above, TFLO relies on the assumption that we can fit the exact observables to the output of a noiseless FLO simulation. Whilst this can often be done using a controlled least-squares fit, the case we are considering is unsuitable for this approach because of the relative stability of the observables over time, which leads to a signal-to-noise ratio that is too small. We therefore chose to adopt a more sophisticated, three-stage fitting procedure:

1. We determine the best TFLO rescaling parameters via a standard least-square fit procedure on those observables with a sufficiently large SNR.

2. After grouping the observables by (i) the number of qubits they act on, and (ii) whether they act in multiple Fermi-Hubbard sites, we compute a prior distribution over the optimal fit parameters.

3. For each of the groups defined in 2., we use the prior distribution computed above to perform a linear least-square fit with regularisation.

The process above allows us to obtain good fit parameters even in the presence of low SNR in the time series.

4. Non-Gaussian circuits - classical simulation

One type of experiment highlighted in the main text considers an initial singlet covering state. This is a non-Gaussian or magic state [103–109] because it cannot be generated a FLO unitary [57], and in general the simulation cost grows exponentially with the number of singlets present.

Since fermionic Gaussian states are those of non-interacting fermions corresponding to the $U = 0$ regime, this is implicitly assumed for the rest of this section.

The singlet creation operator on vacuum is given by $\frac{c^\dagger_{a,\uparrow} c^\dagger_{b,\downarrow} - c^\dagger_{a,\downarrow} c^\dagger_{b,\uparrow}}{\sqrt{2}}$ for physical sites a and b. However, we note that the initial state can be written as a product state across N_{singlets} singlets and N_{lonely} lonely sites (which can be

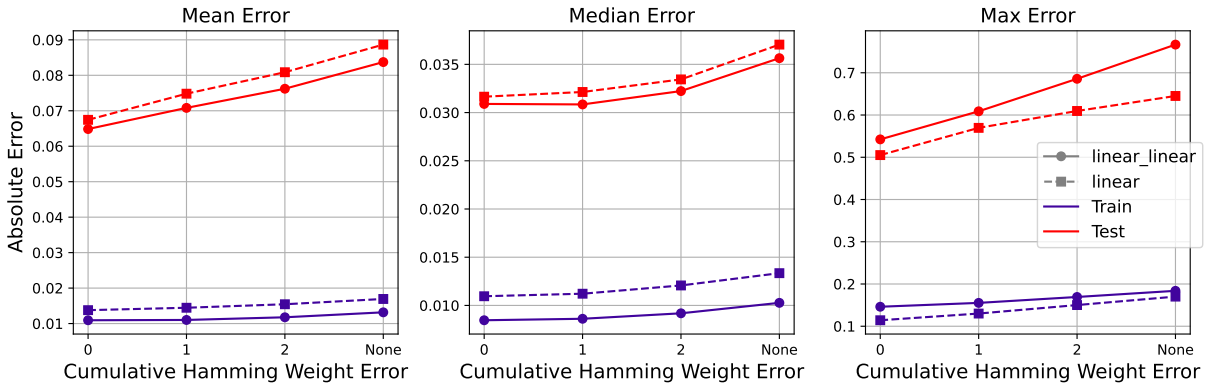


FIG. 39: The mean, median and maximum training error ($U \approx 0$) and test error ($U = 4, 8$) over all single body and two body Pauli-Z expectation values mitigated with different levels of post-selection, and additional TFLO, GPR and symmetry averaging. This data is produced by the 4×4 experiment with the central hole initial state. A comparison of the linear-linear and linear ansatz deployed in TFLO is shown, while varying the cumulative hamming weight errors accepted in the shots used to give the expectation values.

empty, fully ($\uparrow\downarrow$) or partially filled with $\uparrow$ or $\downarrow$) according to Eq. (E3)

$$|\psi\rangle_{\text{init}} = \bigotimes_{k=1}^{N_{\text{singlets}}} |\text{singlet}\rangle_{s_k} \bigotimes_{l=1}^{N_{\text{lonely}}} |\text{lonely}\rangle_{s_l} \tag{E3}$$

where s_k is the pair of sites from the k^{th} singlet and s_l a lonely site. The observables needed for TFLO error mitigation are linear combinations of $\langle Z_i \rangle$ and $\langle Z_i Z_j \rangle$ correlators $\forall i, j$ modes, which are enough to perform TFLO corrections for all physical observables considered in the main text except for the Wilson lines. These require higher weight Z strings, and this computation is omitted since the exponential cost with respect to N_{singlets} becomes cumbersome.

Densities

The observables $\langle \hat{n}_i \rangle, \forall i$ fermionic modes is

$$\langle \psi_{\text{init}}(t)| c_i^\dagger c_i |\psi_{\text{init}}(t)\rangle = \langle \psi_{\text{init}}| e^{iHt} c_i^\dagger c_i e^{-iHt} |\psi_{\text{init}}\rangle \tag{E4}$$

$$= \sum_{\alpha,\beta} R_{\alpha,i} R^*_{\beta,i} \langle \psi_{\text{init}}| c_\alpha^\dagger c_\beta |\psi_{\text{init}}\rangle \tag{E5}$$

$$= \sum_{\alpha} |R_{\alpha,i}|^2 \tag{E6}$$

where H is the Hamiltonian defined in Eq. (1) and $R = e^{-iht}$ where h is the lattice adjacency matrix. Because of the initial state structure, for a non-vanishing overlap in the sum, the coefficients α and β must be equal, since flipping two qubits within or across physical sites results in an orthogonal state.

Density-Density correlators

Here, we are concerned with

$$\langle \psi_{\text{init}}(t)| c_i^\dagger c_i c_j^\dagger c_j |\psi_{\text{init}}(t)\rangle = \langle \psi_{\text{init}}| e^{iHt} c_i^\dagger c_i c_j^\dagger c_j e^{-iHt} |\psi_{\text{init}}\rangle \tag{E7}$$

$$= \sum_{\alpha,\beta,\gamma,\eta} R_{\alpha,i} R^*_{\beta,i} R_{\gamma,j} R^*_{\eta,j} \langle \psi_{\text{init}}| c_\alpha^\dagger c_\beta c_\gamma^\dagger c_\eta |\psi_{\text{init}}\rangle \tag{E8}$$

Since ψ_{init} is not a Gaussian state, we need to consider the $(2L)^P$ mode combinations in the sum with operator weight P. For this special case of $P = 4$, the computational complexity can be reduced to scale as $O(L^2)$ due to the structure of the initial state.

Note that $c_\alpha^\dagger c_\beta c_\gamma^\dagger c_\eta$ flips four qubits and the structure of a singlet state across sites a and b is

$$\frac{1}{\sqrt{2}}(|1_{i,\uparrow} 0_{i,\downarrow} 0_{j,\uparrow} 1_{j,\downarrow}\rangle + |0_{i,\uparrow} 1_{i,\downarrow} 1_{j,\uparrow} 0_{j,\downarrow}\rangle). \tag{E9}$$

The non-vanishing overlap of the singlet state with itself is to flip all four qubits. Similarly, any lonely site has non-vanishing self-overlap when nothing is flipped. Therefore, we only need to consider the cases where α, β, γ and η land on a single or a pair of physical sites that are part of the same singlet.

Four distinct indices

Using this argument, one concludes that all non-trivial combinations of distinct α, β, γ and η must be within a single singlet. For this case, the correlator becomes

$$\langle n_i n_j \rangle_{4 \text{ distinct}}(t) = \sum_{(i,j)}^{N_{\text{singlets}}} \left(\sum_{\alpha,\beta,\gamma,\eta \in \text{perm}[i\uparrow, i\downarrow, j\uparrow, j\downarrow]} R_{\alpha,i} R^*_{\beta,i} R_{\gamma,j} R^*_{\eta,j} \right.$$
$$\left. \times \frac{1}{2} \langle \Omega | \, (c^\dagger_{i,\uparrow} c^\dagger_{j,\downarrow} - c^\dagger_{i,\downarrow} c^\dagger_{j,\uparrow})^\dagger c^\dagger_\alpha c_\beta c^\dagger_\gamma c_\eta (c^\dagger_{i,\uparrow} c^\dagger_{j,\downarrow} - c^\dagger_{i,\downarrow} c^\dagger_{j,\uparrow}) | \Omega \rangle \right) \tag{E10}$$

for the vacuum state $|\Omega\rangle$. 'Perm' specifies permutations of a list and (i, j) are site pairs of the singlet covering.

Two distinct indices

There are two cases to consider here, $\alpha = \beta, \gamma = \eta$ and $\alpha = \eta, \beta = \gamma$. For both, the distinct indices are either

1. within the same singlet or lonely site.
2. across two singlets or two lonely sites.
3. between a singlet and a lonely site.

(E11)

The two-body correlator is therefore

$$\langle n_i n_j \rangle_{2 \text{ distinct}}(t) = \sum_{(i,j)} \left(\sum_{\alpha,\gamma \in \text{perm}[i\uparrow, i\downarrow, j\uparrow, j\downarrow]} |R_{\alpha,i}|^2 |R_{\gamma,j}|^2 \langle \Omega | \, \mathcal{C}^\dagger_{i,j} n_\alpha n_\gamma \mathcal{C}_{i,j} | \Omega \rangle \right.$$
$$\left. + \sum_{\alpha,\beta \in \text{perm}[i\uparrow, i\downarrow, j\uparrow, j\downarrow]} R_{\alpha,i} R^*_{\beta,i} R_{\beta,j} R^*_{\alpha,j} \langle \Omega | \, \mathcal{C}^\dagger_{i,j} n_\alpha (1 - n_\beta) \mathcal{C}_{i,j} | \Omega \rangle \right) \tag{E12}$$

where (i, j) is a site index pair across all the conditions outlined in Eq. (E11) and $\mathcal{C}_{ij}$ is the creation operator of the type of initial state inhabiting sites i and j.

Single distinct index This is similar to the case of the number operator discussed in Appendix E 4, i.e.

$$\langle n_i n_j \rangle_{1 \text{ distinct}}(t) = \sum_\alpha \sum_i^{2L} |R_{\alpha,i}|^2 |R_{\alpha,j}|^2 \langle \psi_{\text{init}} | \, n_i | \psi_{\text{init}} \rangle \tag{E13}$$

where i is the mode index.

Finally, the case where $c^\dagger_\alpha c_\beta c^\dagger_\gamma c_\eta$ has three unique indices is trivially zero since there is either a double creation/annihilation on the same mode. Therefore,

$$\langle n_i n_j \rangle(t) = \langle n_i n_j \rangle_{1 \text{ distinct}}(t) + \langle n_i n_j \rangle_{2 \text{ distinct}}(t) + \langle n_i n_j \rangle_{4 \text{ distinct}}(t) \tag{E14}$$

which scales as $O(L^2)$.

5. Sampled bitstring probabilities at zero interaction strength

In this section, we discuss the classical computations of overlap probabilities of the form $|\langle z|U|0^{\otimes n}\rangle|^2$ for some bitstring z (sampled in the computational basis, from the device, TN MPS state or exact FLO-sampler), which are required for cross-entropy benchmarking analyses discussed in Appendix I. Here, U denotes both the creation of the initial state and subsequent time evolution. Consider any initial state considered in the main text in the most general form,

$$|\psi_{\text{init}}\rangle = \sum_m \alpha_m \left(\prod_{i=1}^W c^\dagger_{m_i} \right) |\Omega\rangle \tag{E15}$$

where index m_i runs over the modes making up the Fock state (indexed m) with complex coefficient α_m and $|\Omega\rangle$ is the vacuum state. W is the hamming weight of all the Fock states making up $|\psi_{\text{init}}\rangle$. For any FLO initial state considered (e.g. Néel state with central hole) there is a single term in the summation and the overlap inherits the efficient classical simulatability of conventional fermionic linear optics. The initial state with singlet covering has exponentially many terms in the summation with respect to the number of singlets present.

We can write a sampled bitstring z as a single Slater determinant

$$|z\rangle = \prod_{i=1}^{W} c^\dagger_{m'_i} |\Omega\rangle \tag{E16}$$

such that the overlap probability with the time-evolved state is

$$\left|\langle z|\psi(t)\rangle\right|^2 = \left|\sum_m \alpha_m \langle\Omega| \left(\prod_{i=1}^{W} c^\dagger_{m'_i}\right)^\dagger e^{-iHt} \left(\prod_{i=1}^{W} c^\dagger_{m_i}\right)|\Omega\rangle\right|^2 \tag{E17}$$

$$= \left|\sum_m \alpha_m \left(\sum_{q_1,\ldots,q_W} \left(\prod_{i=W}^{1} R^*_{q_i,m'_i}\right)\right) \langle\Omega|\, c_{q_W} \ldots c_{q_1} c^\dagger_{m_1} \ldots c^\dagger_{m_W} |\Omega\rangle\right|^2 \tag{E18}$$

$$= \left|\sum_m \alpha_m \left(\sum_{q_1,\ldots,q_W} \left(\prod_{i=W}^{1} R^*_{q_i,m'_i}\right)\right) \text{sgn}\left(\sigma^{m_1,\ldots,m_W}_{q1,\ldots,q_W}\right)\right|^2 \tag{E19}$$

$$= \left|\sum_m \alpha_m \det \begin{pmatrix} (m_1) & R^*_{m_1,m'_1} & R^*_{m_1,m'_2} & \ldots & R^*_{m_1,m'_W} \\ \vdots & \ldots & \ldots & \vdots & \\ (m_W) & R^*_{m_W,m'_1} & R^*_{m_W,m'_2} & \ldots & R^*_{m_W,m'_W} \end{pmatrix}\right|^2 \tag{E20}$$

where $R = e^{-iht}$. The values m_1 and m_W in grey denote the row indices; to help visualise their positions relative to each other, and do not take part in the overall determinant. $\text{sgn}(\sigma^B_A)$ is equal to $(-1)^{P_{AB}}$, where P_{AB} is the number of swaps needed to permute the set of indices A into that of B.

For the singlet covering case, $|\alpha_m| = (\frac{1}{\sqrt{2}})^{N_{\text{singlets}}}\ \forall m$ so the main bottleneck for computing this probability is the exponentially growing number of Fock states with respect to N_{singlets}.

6. Sampling from Trotterised and exact FLO simulations

For several experiments, such as that of the holon stripes in Appendix B 2, holon/doublon pairwise distance dynamics in Appendix B 5 and crucially the cross-entropy benchmarking analyses in Appendix I, FLO samples were required to provide the ground truth for the $U = 0$ experiments. These samples were generated via the Bravyi-König sampling algorithm [110] developed in the FLOYao package [111].

7. Gaussian Process Regression (GPR)

In the ideal case, all time series signals considered in this work are smooth functions of t, but hardware, twirling and shot noise lead to discontinuous time series—even after all other error mitigation is done. To further use the prior information about the smoothness of our time series, we use GPR [112, 113] and fit a Gaussian process model to the noisy time series. Concretely, we used a squared exponential kernel with length scale $l = 0.4$ and noise strength $\sigma = 1.5$. These values of l and σ were found by maximizing the log-marginal-likelihood of these hyper-parameters on a wide selection of observable time series and noticing that the result of this maximisation clustered around these values of the hyper-parameters. In all the figures in this work that include GPR, we show the 68% one-sigma region as a shaded region.

8. Symmetry averaging

The symmetries of the system provide a natural avenue to mitigate errors in the calculation. First of all, we expect the measured expectation values to be invariant under some symmetries, depending on the initial state,

lattice structure and observable of interest. Moreover, the Hamiltonian itself will be invariant under both the spatial symmetries of the lattice and the spin-reflection symmetry $Rc_{i,\uparrow}R^\dagger = c_{i,\downarrow}$. Up to Trotter error, this invariance will carry through to the circuits used to simulate the Hamiltonian. As a consequence of this, we can increase the effective shot-count for the expectation values of multi-qubit Z strings (and any expectation value derived from them) by averaging each multi-qubit Z expectation value with its symmetry-related partners. Moreover, we can exploit the fact that some observables need to be identical due to lattice symmetries to reduce Trotter error. Indeed, since said observables might be affected differently by Trotter error, by averaging across them, we can expect to reduce the Trotter error itself.

9. Error propagation

The time series Figs. 3, 9, 70, 73, etc. show error bars computed using Gaussian error propagation. To that end, the bars on the initial raw expectation values obtained from the shots are calculated using

$$\sigma^2_{\text{raw}} = \frac{1}{N_t(N_t-1)}\sum_{i=1}^{N_t}(\hat{\mu}_i - \mu)^2 + \frac{1}{N_t^2 N_s}\sum_{i=1}^{N_t}\sigma_i^2 \tag{E21}$$

where N_t is the number of twirling instances, N_s is the number of shots taken per twirling instance, μ_i and σ_i^2 are the sample mean and sample variance per twirling instance. Finally, $\mu = \frac{1}{N_t}\sum_{i=1}^{N_t}\mu_i$ is the total mean.

a. *Pauli twirling and error bars*

Determining error bars in conjunction with Pauli twirling requires some care, in order to correctly account for the iterated expectations. In particular, one can show that the naïve approach of collecting all $N_t \times N_s$ into a single set and then computing the expectation of an observable O as

$$\hat{\mu}_O = \frac{1}{N_t \times N_s}\sum_{i,j=1}^{N_t,N_s} O(x_{i,j}) \tag{E22}$$

and the variance of $\hat{\mu}_O$ in the usual way as

$$\hat{\sigma}^2_{\hat{\mu}_O} = \frac{1}{N_t \times N_s(N_t \times N_s - 1)}\sum_{i,j=1}^{N_t,N_s}(O(x_{i,j}) - \hat{\mu}_O)^2 \tag{E23}$$

is inaccurate, since in non-trivial cases (such as when $N_s = 1$ or all twirling instances are the same) it will underestimate the overall error. This can be easily demonstrated by considering the case in which we take a large number of shots (N_s) taken from a small set of N_t twirling instances that produce very different distributions. In this case, the error computed via Eq. (E23) will be very small, despite the fact that the data has only been collected from very few twirling instances, which all give very different estimates for μ_O. There are two key contributions that need to be correctly accounted for to get an accurate estimate $\hat{\sigma}^2_{\hat{\mu}_O}$ for the variance of $\hat{\mu}_O$. First, the variance coming from sampling different twirling instances (given by the first term in Eq. (E21)), and second, the contribution coming from the shot noise in each circuit (the second term in Eq. (E21)). By employing the law of total variance and going through the derivations, we then obtain the formula Eq. (E21) used in this work.

b. *Verification via bootstrapping*

We chose to use the `uncertainties` Python package [114] to compute error propagation. Whilst the package implements automatic Gaussian error propagation for many basic mathematical functions, the validity of its propagation calculations depends on the assumption that 1. the input distributions are all Gaussian 2. all data transformation functions are (or at least well-approximated by) affine linear functions. Moreover, it is important to note that the `uncertainties` package only tracks and propagates the individual uncertainty of each variable, but does not take

correlations into account. In order to verify the accuracy of Gaussian error propagation, we have compared the results to those obtained via the alternative method of bootstrapping [115, Chap. 8]. Bootstrapping is a technique used to estimate the distribution (or properties thereof) of an estimator by replacing the full distribution of the in-going (unknown) random variable with the empirical distribution obtained from the measured samples. In our work, the estimator is the mitigated and processed observable values. When implementing the bootstrap method in conjunction

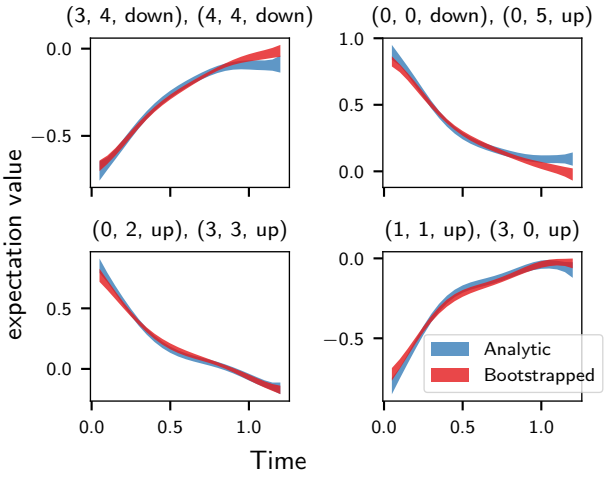


FIG. 40: Comparison of error bars computed via bootstrapping vs Gaussian error propagation for a random choice of 4 observables. Again, we use the same 6×5 data set also presented in Fig. 70 for this example.

with pseudo-twirling, a similar argument to that discussed in Appendix E 9 a holds, meaning that one first needs to sample the N_t circuits with replacement from the pool of twirling instances and then sample N_s shots with replacement from each of the sampled circuits. Only in this way can one obtain a meaningful dataset that can be processed through the whole error mitigation and data analysis pipeline. This procedure needs to be repeated multiple times to allow us to estimate the standard deviations of the final mitigated observables. As shown in Fig. 40, we find that – for the systems under consideration – the error bars obtained via bootstrapping are in remarkably good agreement with those obtained via Gaussian error propagation. Considering that bootstrapping is significantly more computationally expensive and does not appear to give any meaningful benefit, we have opted to use Gaussian error propagation for all error bars reported here.

10. Maximum entropy shot-reweighting

In this section, we apply the principle of maximum entropy [116] to smoothly reweight samples according to the minimal adjustment in information gained by knowing some corrected/improved local expectation values. We consider a list of sampled bit strings $\{x_1, ..., x_{n_{shots}}\}$, where each $x_i \in \{0,1\}^n$, which we use to estimate some observables $\{O_j\}_{j=1}^{n_{observables}}$,

$$\tilde{O}_j = \frac{1}{n_{shots}} \sum_{i=1}^{n_{shots}} O_j(x_i)$$

and suppose for this class of observables we have a correction procedure for these estimates: $\tilde{O}_j \to \hat{O}_j$. For example, the observables $\{\hat{O}_j\}$ could be 1-local and 2-local Paulis obtained from TFLO.

A natural question is: can we *reconcile* our list of samples with the corrected expectation values? Here, we propose to smoothly reweight the samples according to the principle of maximum entropy. Introduce a reweighting function p, which is a normalised $\sum_i p(x_i) = 1$ probability mass function on the set $\{x_1, ..., x_{n_{shots}}\}$ and is intended to satisfy:

$$\sum_{i=1}^{m} p(x_i) O_j(x_i) = \hat{O}_j \tag{E24}$$

for each corrected observable in the specified set $\{O_j\}_{j=1}^{n_{observables}}$. Defining $\vec{O} = (O_1, ..., O_{n_{observables}})$, this equation Eq. (E24) has a solution if and only if $\hat{\vec{O}}$ is a normalised convex combination of the vectors $\{\vec{O}(x_i)\}_{i=1}^{n_{shots}}$. To specify a unique solution, we consider the principle of maximum entropy, and seek a reweighting distribution p that maximises entropy $H(\{p\}) = -\sum_{i=1}^{m} p(x_i) \ln p(x_i)$ subject to the linear constraints from the corrected observables.

When exact solutions for p in Eq. (E24) exist (for a particular set of samples and corrected observables) the maximum entropy problem is feasible, and if all $p(x)$ are bounded away from the extreme points $0 < p(x) < 1$, the optimisation problem has a solution in terms of Lagrange multipliers $\{\lambda_j\}_{j=1}^{n_{observables}}$, one λ_j for each observable O_j

$$p(x) = \frac{1}{Z(\vec{\lambda})} e^{\vec{\lambda}\cdot\vec{O}(x)} \quad , \quad Z(\lambda) = \sum_{i=1}^{n_{shots}} e^{\vec{\lambda}\cdot\vec{O}(x_i)}$$

The condition $0 < p(x) < 1$ for all x is necessary for the $\vec{\lambda}$ to be finite. This is typical for the case $n_{shots} > n_{observables}$, and then this solution can greatly reduce the number of parameters involved in the distribution, down from the number of samples to the number of observables that are being corrected.

An important distinction is that by only reweighting samples from the device, we are not replacing the quantum measurement distribution with a maximum entropy approximation based on local observables, rather we are "tilting" the empirical distribution with importance weights that reflect the corrected local observables,

$$p_{\text{mitigated}} \propto e^{\vec{\lambda}\cdot\vec{O}(x)} p_{\text{device}}(x). \tag{E25}$$

Therefore, the reweighted samples can be regarded as samples from the distribution above, up to the fact that the solutions for $\vec{\lambda}$ are also estimated from a finite set of samples. For fixed $\vec{\lambda}$ (assumed to converge on the sample set), we can compute the standard error in the usual way for samples with importance weights,

$$\langle F \rangle = \sum_{i=1}^{n_{shots}} p(x_i) F(x_i) \quad , \quad \text{SE}^2 = \frac{\sum_{i=1}^{n_{shots}} p(x_i)(F(x_i) - \langle F \rangle)^2}{n_{\text{eff}} - 1} \quad , \quad n_{\text{eff}} = \frac{1}{\sum_{i=1}^{n_{shots}} p(x_i)^2}$$

The effective number of mitigated shots n_{eff} accounts for the fact that concentrating most of the weight on a small number of samples increases the uncertainty in computing further expectation values. When the corrected observable values are far from the predictions of the unmitigated data, the reweighting procedure concentrates the weight on relatively few of the samples. This is determined by the magnitude of the additive correction, so in practice the concentration of weights (large $||\vec{\lambda}||$) is most apparent when the expectation values of (traceless) local observables are large (e.g. at early times in the quantum circuit).

Lastly, the optimal $\{\lambda_j\}$ satisfying the constraints Eq. (E24) can be found by gradient descent iterations, with

$$\nabla\vec{\lambda} = \hat{\vec{O}} - \langle \vec{O} \rangle_{\vec{\lambda}}$$

with iterations $\lambda_j \to \lambda_j + \eta\, \nabla\lambda_j$ where η is a positive parameter. Note that the expectation in the current model $\langle \vec{O} \rangle_{\vec{\lambda}}$ can be evaluated efficiently since it only involves a sum over n_{shots} samples. To justify this, consider the quantity which is analogous to the free energy in statistical mechanics,

$$F = \log Z(\lambda),$$

which has the property that its derivatives with respect to $\{\lambda_j\}$ are exactly the expectation values of the O_j,

$$\frac{\partial F(\vec{\lambda})}{\partial \lambda_j} = \frac{1}{Z(\vec{\lambda})} \sum_{i=1}^{n_{shots}} O_j(x_i) e^{\vec{\lambda}\cdot\vec{O}(x_i)} = \langle O_j \rangle_{\vec{\lambda}}$$

so $\nabla F(\vec{\lambda}) = \langle \vec{O} \rangle_{\vec{\lambda}}$. Consider the entropy

$$H(\vec{\lambda}) = -\sum_{i=1}^{n_{shots}} p(x_i) \ln p(x_i) = -\sum_{i=1}^{n_{shots}} \frac{e^{\vec{\lambda}\cdot\vec{O}(x_i)}}{Z(\vec{\lambda})} \left(\vec{\lambda} \cdot \vec{O}(x_i) - \ln Z(\vec{\lambda}) \right) = F(\vec{\lambda}) - \vec{\lambda} \cdot \vec{O}(x_i)$$

For any $\vec{\lambda}$ define

$$\hat{H}(\vec{\lambda}) = F(\vec{\lambda}) - \vec{\lambda} \cdot \hat{\vec{O}} \tag{E26}$$

which satisfies

$$\nabla \hat{H}(\vec{\lambda}) = \langle \vec{O} \rangle_{\vec{\lambda}} - \hat{\vec{O}} \tag{E27}$$

so that $\nabla \hat{H}(\vec{\lambda}) = 0$ implies the expectation value constraints are satisfied. If $\vec{\lambda}^*$ is the optimal solution, then $\hat{H}(\vec{\lambda}) - \hat{H}(\vec{\lambda}^*)$ is the KL divergence $D_{\mathrm{KL}}(p_{\vec{\lambda}^*} || p_{\vec{\lambda}})$ since

$$D_{\mathrm{KL}}(p_{\vec{\lambda}^*} || p_{\vec{\lambda}}) = \sum_{i=1}^{n_{shots}} p_{\vec{\lambda}^*} \ln \left(\frac{p_{\vec{\lambda}^*}}{p_{\vec{\lambda}}} \right) = (\vec{\lambda}^* - \vec{\lambda}) \cdot \hat{\vec{O}} + F(\vec{\lambda}) - F(\vec{\lambda}^*) = \hat{H}(\vec{\lambda}) - \hat{H}(\vec{\lambda}^*)$$

therefore $\hat{H}(\vec{\lambda})$ is also the cross entropy to the optimal distribution. We can explicitly verify that $\hat{H}$ is convex by relating the Hessian of $\hat{H}$ to the covariance matrix,

$$\frac{\partial^2 \hat{H}(\vec{\lambda})}{\partial \lambda_k \partial \lambda_j} = \langle O_k O_j \rangle_{\vec{\lambda}} - \langle O_k \rangle_{\vec{\lambda}} \langle O_j \rangle_{\vec{\lambda}}$$

therefore $\nabla^2 \hat{H} = \mathrm{Cov}_{p_{\vec{\lambda}}}(\vec{O}) \succeq 0$ and so $\hat{H}$ is convex. This ensures that the gradient descent will reduce the cross entropy $\hat{H}(\vec{\lambda})$ to the ideal solution at each step, $\hat{H}\left(\vec{\lambda} - \eta \nabla \hat{H}\right) < \hat{H}(\vec{\lambda})$, and so the gradient iterations in (E27) converge efficiently.

In our implementation, we used all single-qubit Z expectation values, together with certain two-qubit ZZ expectation values (those across adjacent sites in the lattice in the same spin sectors, and between spin sectors on each site), as the known exact expectation values $\{\hat{O}_j\}$. These are either computed exactly, when $U = 0$, or via TFLO when $U \neq 0$. The optimal Lagrange multipliers λ_j are found by performing gradient descent on the entropy $\hat{H}(\lambda)$ given by Eq. (E26) with the gradient of Eq. (E27). As mentioned, when the expectation values are strongly polarised (e.g. at early times), the norm $\|\vec{\lambda}\|$ can become large, and so we introduce an ad-hoc regularisation by adding $+c\|\vec{\lambda}\|^2$ to the cost function and modifying the gradient accordingly. Another challenge arises because the TFLO-corrected local observables are not exact, so they are not guaranteed to be globally consistent. In that case, no feasible solution to the optimisation problem exists, and the regularisation leads to an approximate solution.

11. Probabilistic Error Amplification (PEA) and Zero Noise Extrapolation (ZNE)

We implemented a variant of PEA introduced in [117] in which a local Pauli noise model (under Pauli twirling) is learned by fitting the decay rates of tomographic signals [118]. An amplified variant of this noise model is then simulated in the circuit by the stochastic inclusion of Pauli operators. Zero-noise extrapolation of the amplified signal is then applied via either a linear or an exponential fit. Only the CZ gate noise is modelled. We assume local noise confined to the support of the target gates. We identify all instances of distinct layers of parallel CZ gates in the target circuit and learn the local noise model for each such distinct layer.

The protocol for learning the noise is to assume that a *Pauli-twirled* noise channel Λ is associated with a layer of gates. The Pauli twirling is done with the same protocol described in Appendix E 1. The noise model is the formal exponentiation of a sparse Lindbladian $\Lambda = e^{\mathcal{L}}$ and

$$\mathcal{L}(\rho) = \sum_{k \in \mathcal{K}} \lambda_k (P_k \rho P_k - \rho) \tag{E28}$$

$$\Lambda(\rho) = \prod_{k \in \mathcal{K}} (\omega_k \cdot + (1 - \omega_k) P_k \cdot P_k^\dagger) \rho, \tag{E29}$$

where $\mathcal{K}$ indexes the set of Pauli operators generating the Lindbladian and $\omega_k = (1 + e^{-2\lambda_k})/2$. In line with the approach of [117, 118], we assume that the generators of the model are local with respect to the device topology, so $\mathcal{K}$ is chosen to contain all single-qubit Pauli operators and all two-qubit Pauli operators for each pair of qubits sharing an edge in the hardware graph.

The model parameters λ_k are estimated by measuring Pauli fidelities for a chosen measurement basis $\{B_j\}_{j \in \mathcal{B}}$ and solving a linear system of equations. The fidelities

$$f_j = \frac{1}{2^n} \mathrm{Tr}[B_j^\dagger \Lambda(B_j)], \tag{E30}$$

are estimated through standard benchmarking techniques [118], which extract them from an exponential decay.

The measurement bases $\{B_j\}_{j\in\mathcal{B}}$ are used to define a matrix $M = \mathcal{M}(\mathcal{B}, \mathcal{K})$ with entries

$$M_{i,j} = \begin{cases} 0, & [B_i, P_j] = 0 \\ 1, & [B_i, P_j] \neq 0. \end{cases} \tag{E31}$$

We then estimate the λ_k by solving

$$\lambda(f) := \underset{\lambda\geq 0}{\operatorname{argmin}} \frac{1}{2}\|M\lambda + \log(f)/2\|_2^2, \tag{E32}$$

where the logarithm here is entrywise and λ and f are vectors of the noise parameters and learned fidelities. To ensure that the matrix M is non-singular, it suffices to choose $\mathcal{B} = \mathcal{K}$, and, as shown in [118], fidelities for this set of local observables can be measured using just 9 distinct measurement settings.

To amplify the effect of noise and produce the ZNE fitting data, we simulate increasing the noise rates on the device as $\lambda_k \to G\lambda_k$ with gain factors $G \in [1.0, 1.1, 1.34, 1.58]$, by randomly inserting Pauli gates according to equation Eq. (E29) with appropriately modified probabilities ω_k. To produce the ZNE fitting data, we sample 500 random instances per logical circuit, where the randomisation incorporates both twirling and error amplification, and take 1024 samples per circuit. The number of instances was chosen because we found it sufficient to get well-converged expectation values, while the number of shots was chosen to match the overhead incurred by changing the circuit parameters that we run at. To ensure that all circuits are run within a reasonable time frame from when the noise profiling takes place and to reduce the compilation overhead, we employ parameterised circuits as already described in Appendix E 1 as well as the same Pauli twirling protocol.

12. Performance Comparisons

In order to assess the best performing error mitigation strategy, we benchmark variants of TFLO and ZNE on the 4×4 system. For these circuits, we have access to exact expectation values, enabling us to directly measure the performance of each error mitigation strategy. As our benchmark experiment, we chose the setup consisting of a central holon and evolution under the Hamiltonian excluding a magnetic flux.

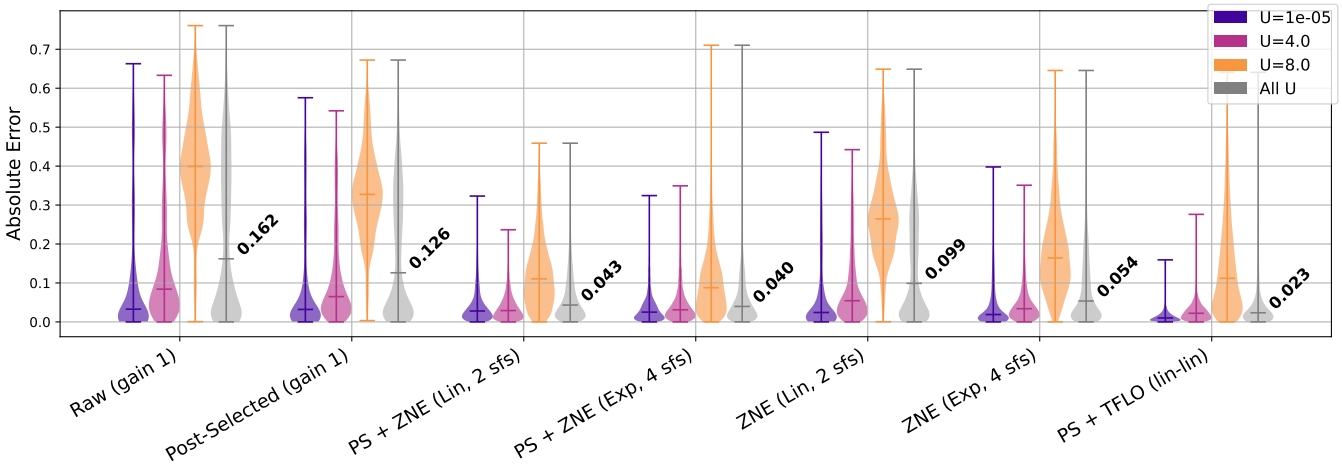


FIG. 41: Comparison of error distributions across different ZNE implementations benchmarked against TFLO. The distribution of absolute errors is shown for the exponential and linear ansatz deployed in ZNE. The number of scale factors (sfs) was chosen to give the best ZNE performance. Coloured violins correspond to individual interaction strengths U, while the grey violins show the pooled error over all U values. Medians over the pooled U distributions are annotated. The different ZNE fitting procedures are described in the surrounding text.

In Fig. 41, we compare different ZNE fitting procedures against each other, and also against TFLO and no error mitigation or only post-selection. "PS + ZNE" always refers to ZNE applied to post-selected shots, "Lin." and "Exp." refer to extrapolation fits with a linear and exponential function, respectively, while "$< n >$ sfs" refers to the number

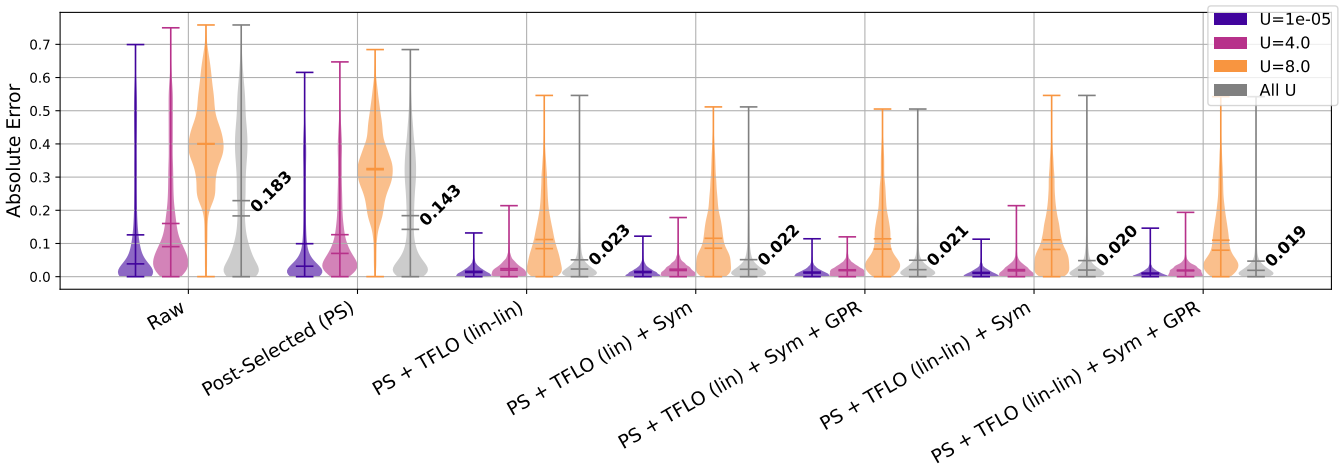


FIG. 42: Comparison of error distributions across different TFLO implementations when applied to mitigating all single and two-body Pauli Z expectation values for the dynamics of the 4×4 model with a central holon on a Néel background. The distribution of absolute errors is shown for each method, highlighting the effect of post-selection, symmetrisation, and TFLO mitigation. Coloured violins correspond to individual interaction strengths U, while the grey violins show the pooled error over all U values. Medians over the pooled U distributions are annotated.

of gain factors included in the extrapolation fit. It is clear that TFLO outperforms all variants of ZNE that we tried in terms of median error (by almost a factor of 2), and reduces the raw error by almost a factor of 8.

In Fig. 42, we show the distribution of absolute errors over all single-body Z and all two-body ZZ expectation values, for a variety of error mitigation methods, and their combinations. As before, "+" refers to an error mitigation method being applied on top of the preceding ones, "lin-lin." to the TFLO fit being done according to Eq. (E2) and "lin." according to Eq. (E1), respectively. Small differences between apparently identical conditions in Fig. 41 and 42 are due to results corresponding to different experimental runs. We see that the use of TFLO achieves a significant reduction in error (by up to more than 9x compared with raw data), including for $U = 4, 8$. In addition, GPR provides a further small improvement. The lin-lin variant of TFLO was the most effective.

It is also worth noting that PEA+ZNE comes with a significant circuit overhead: To get reasonably well-converged twirling and error amplification at the higher scaling factors, we found that ~ 500 twirling instances are necessary. Multiplying this by the four scaling factors used to get stable extrapolation fits, we end up with a $100\times$ circuit overhead over the 20 Pauli twirling instances that we used at gain level 1 for TFLO. Given the minor performance differences between the best-performing TFLO and the best-performing PEA+ZNE implementation, we hence opted to use TFLO as the much cheaper error mitigation method for all our experiments.

Appendix F: Tensor Network Simulations

We implemented state-of-the-art tensor network simulations using TeNPy [63], employing the `TwoSiteTDVPEngine` for the time-dependent variational principle (TDVP) algorithm, which can capture the precise time evolution with sufficient resources.

We studied alternative tensor network routines before concluding that TDVP was the most suitable for the system under study. These studies will be presented in future work [62], but in brief, they contained: direct contraction of the quantum circuit [119]; simulation of the time-dependent density matrix renormalisation group within the MPS/MPO framework [120, 121]; and an explicit fourth-order Runge–Kutta integrator applied to the fermionic MPS representation with compression at each step [70].

TDVP can simulate the real- or imaginary-time evolution of quantum many-body systems represented by Matrix Product States (MPS). MPSs are characterised by their bond dimension, χ, which limits the size of indices connecting the tensors within, which effectively constrains the correlations that can be expressed, i.e., the entanglement the network can represent is limited, and may be less than that of the target system [122]. TDVP progresses by iteratively evolving the Schrödinger equation for time step Δt, and at each step projecting the state onto the tangent space of the MPS manifold, i.e., the subset of states which may be represented within the given bond dimension. Smaller Δt yield more precise evolution, at the expense of higher runtime [69, 123]. For a more extensive review of time dynamics methods for MPS states, see the review by Paeckel *et al.* [124]. In addition to this cornerstone algorithm, there is an extensive, ever-evolving tensor network literature, often tailored to specific studies. For example, tensor networks combined with Monte Carlo methods [125] have been applied to highly disordered systems. Tensor networks combined with machine learning have been used to study molecular open quantum systems [126]. Finally, PEPS [127] and MERAs [128] hold promise for capturing more entanglement in higher-dimensional systems, but also pose many numerical challenges. For instance, contracting a PEPS is computationally #P-complete [129]. In this context, approximate PEPS contraction algorithms based on Belief Propagation/Simple Update have proven particularly effective in simulating large 2D systems with low coordination number [59, 72, 73]. More recently, a variety of tensor network methods for simulating (noisy) quantum circuits have emerged and have shown the great promise [25, 71, 130–132].

The two most common variants of TDVP are the one- and two-site evolution, which differ in how much of the MPS is updated at each time step. One-site TDVP evolves each tensor individually, while two-site TDVP evolves pairs of neighbouring sites by forming local clusters of enlarged bond dimension, before truncating the bond to the given χ through the singular value decomposition (SVD) algorithm. Two-site TDVP is known to be more accurate at the expense of higher runtime and memory requirements.

The computational cost of TDVP is cubic in the bond dimension, and also depends on the physical dimension d of each of the N sites in the system, i.e., the runtime of each time step scales like $\mathcal{O}(Nd^3\chi^3)$. To ensure the tensor network simulations are as accurate as possible, we use two-site TDVP throughout this work and choose a cautious value of $\Delta t = 0.01$. However, we note that the tensor network simulations could be made more accurate with a smaller Δt and a larger χ than was feasible to investigate here. However, investigations on small-sized systems with accessible ground truth found that using $\Delta t = 0.001$ did not substantially affect accuracy. In Fig. 43 we report the maximum requirements of computing the TDVP instances reported throughout, noting the considerable increase in runtime with respect to bond dimension, as expected.

To achieve exact evolution of a system with N sites, TDVP requires $\chi = d^{\frac{N}{2}}$, where d is the dimension of each site [133]. In our case for fermions, each site can be unoccupied, doubly occupied, or occupied by either an up or down spin, so $d = 4$. For the largest systems studied in this work, we have a lattice of size 6×6, i.e. $N = 36$, which demands $\chi = 4^{18} = 2^{36}$, which is intractable in both memory and runtime requirements. However, we can still retrieve meaningful approximate simulations with a much smaller bond dimension.

For all hardware data reported in the main text, we implement corresponding instances of TDVP with $\chi = 2^n$ for $n \in [8, 11]$. The largest system we study is of size 6×6 with $\chi = 2^{11}$, which took 6.5 days to run and required 13 GB of RAM. In all simulations, we used a time step $\Delta t = 0.01$ to evolve the dynamics incrementally, balancing the algorithm's accuracy with practical considerations of compute runtime. We also fix the truncation threshold within the singular value decomposition subroutine to 10^{-10}. We performed our simulations on the Google Cloud Platform with virtual machines of type `E2-highmem-4`, i.e., on 2.25 GHz Intel processors with four vCPUs and 32 GB RAM per instance [134].

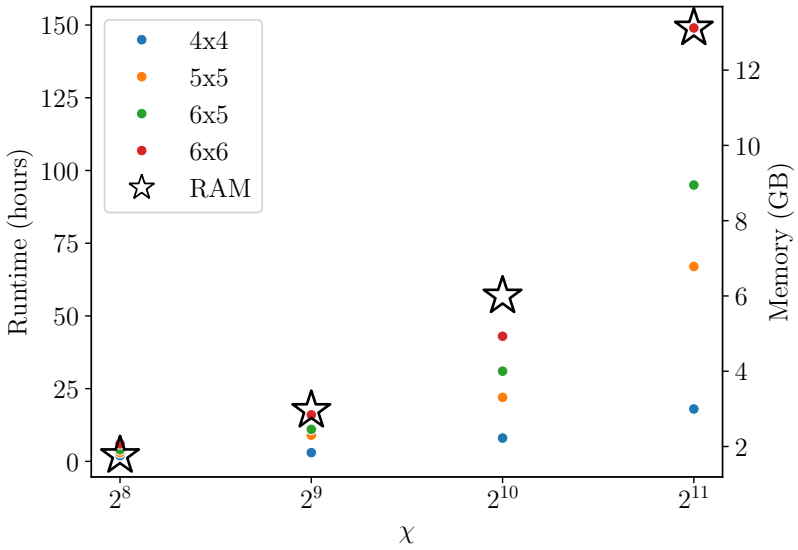


FIG. 43: Time and memory requirements for TDVP simulations corresponding to experimental data. The left-hand y-axis shows the wallclock runtime corresponding to coloured points in the chart, while the right-hand y-axis indicates the RAM requirements of the largest system size (6×6) at each bond dimension, shown by the black stars.

Appendix G: Majorana Propagation Simulations

In order to further test the ability of classical methods to simulate the system we are considering accurately, we implemented the Majorana propagation (MP) algorithm [26]. This method relies on implementing Heisenberg evolution through the circuit on a set of polynomials of Majorana operations that represent the operators of interest: the expectation values are then given by the overlap between the Heisenberg-evolved observable and the initial state.

It is important to note that (i) the majority of gates in our circuits (FSWAP and hopping) leave the Majorana weight unchanged, and (ii) that many observables of physical interest (including spins, charges and related quantities) have low Majorana weight. This suggests that the Majorana propagation algorithm should be expected to perform well in the simulation of the time series shown in Fig. 3, 9, 70 and 73.

The MP algorithm increases the number of terms to track with each gate application. To keep the number of terms feasible, it is necessary to regularly truncate the Majorana polynomial during the computation. In order to control this, we truncate the Majorana polynomial by discarding those terms for which either the Majorana weight – the number of single Majorana operators in the monomial – exceeds a certain weight threshold, or the coefficient is below a given coefficient threshold.

We implement this algorithm following the implementation of the related Pauli propagation algorithm [135, 136] in [137] and building upon the `Yao.jl` quantum simulation framework [138].

Naively, the inclusion of singlets in the initial state would increase the simulation cost exponentially with the number of singlets, because each circuit that prepares the singlet term doubles the number of Majorana monomials we need to track. But by exploiting the fact that – at least in the Heisenberg picture – the state preparation circuits are the last ones to act on the qubits involved, we can discard all terms orthogonal to the zero-state immediately after applying a singlet preparation circuit. This keeps the number of Majorana monomials we need to track constant and implies that including singlets in the initial state does not increase the hardness of the MP simulation. We hence expect that simulations with the other initial states considered in this work can be done to the same accuracy as those shown in Fig. 44.

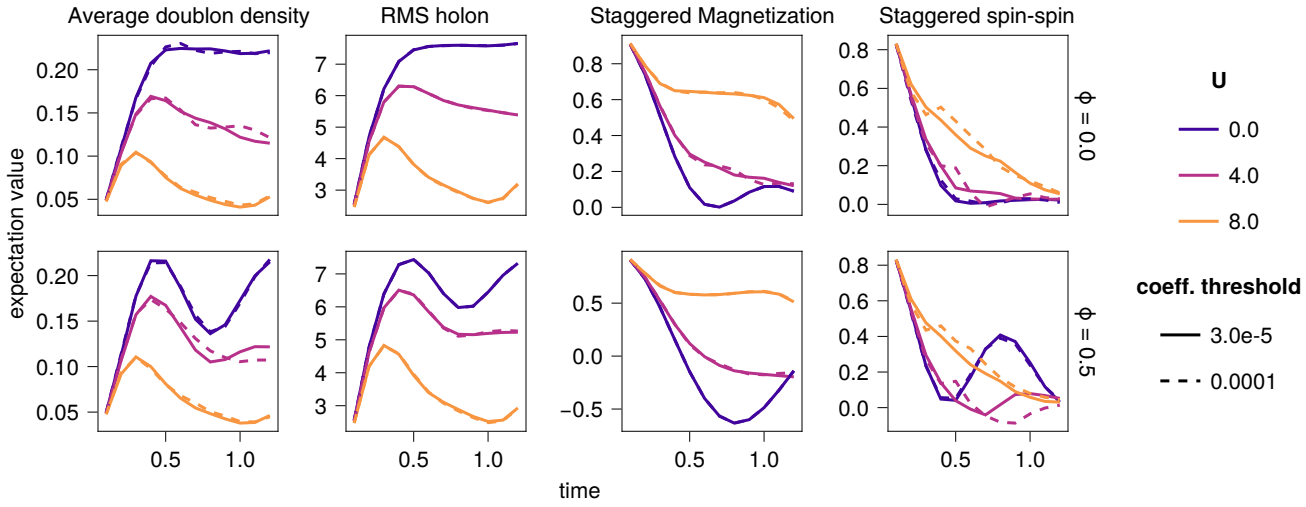


FIG. 44: Majorana propagation simulations time series for coefficient truncation thresholds for the 6×6 system with the Néel state with a central hole. The data with the coefficient threshold 3×10^{-5} is also shown in Fig. 73 and we obtained very similar results for the 5×5 and 6×5 lattices

Fig. 44 shows the performance of our Majorana propagation simulations for different coefficient and weight thresholds, up to a weight threshold of 10 and a coefficient threshold of 3×10^{-5}. It is apparent that the simulations for the average number of doublons, RMS holon and staggered magnetisation are not significantly affected by the change in coefficient threshold, suggesting that the time series presented in the main text are converged – a necessary condition for quantitative accuracy. This is further exemplified by fig. 45 where we show the loss of ℓ^2-norm incurred by the truncation throughout the circuit for the same time series shown in fig. 44. We see that, except for the staggered spin-spin correlator, we lose only relatively little ℓ^2-norm due to the truncation, hinting that the Majorana propagation simulation results for all observables except the staggered spin-spin correlator are fairly accurate.

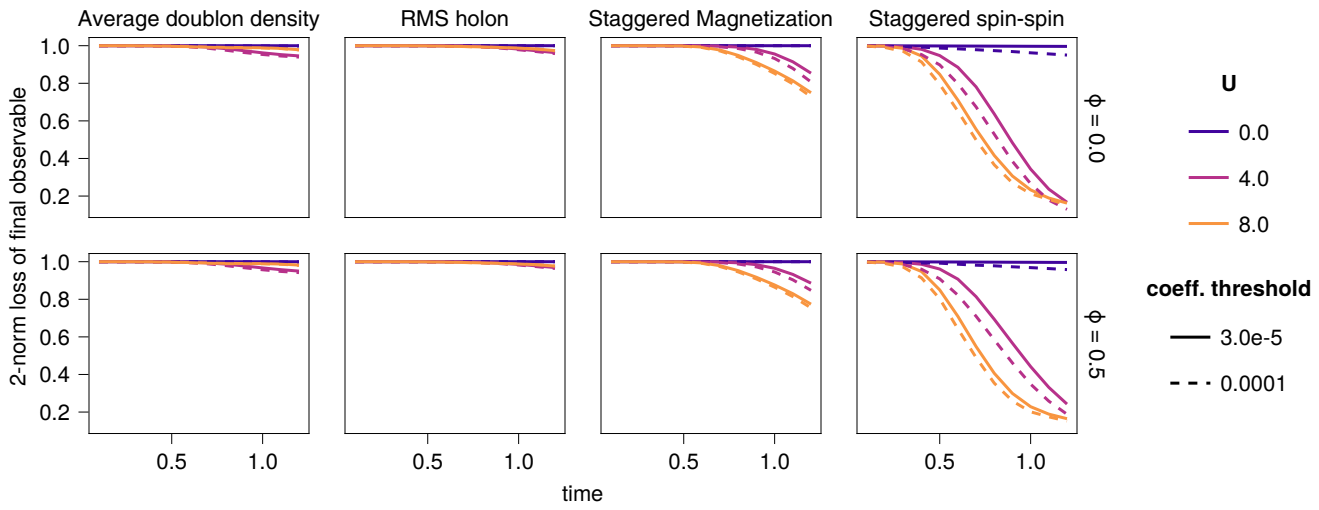


FIG. 45: Ratio of the ℓ^2-norm of the observable before and after Heisenberg evolving it through the circuit for the same time series shown in Fig. 44. In the absence of truncation, the unitary evolution would preserve the ℓ^2-norm of the observable perfectly, making the ℓ^2-norm loss shown here a loose proxy for the accuracy of the simulation.

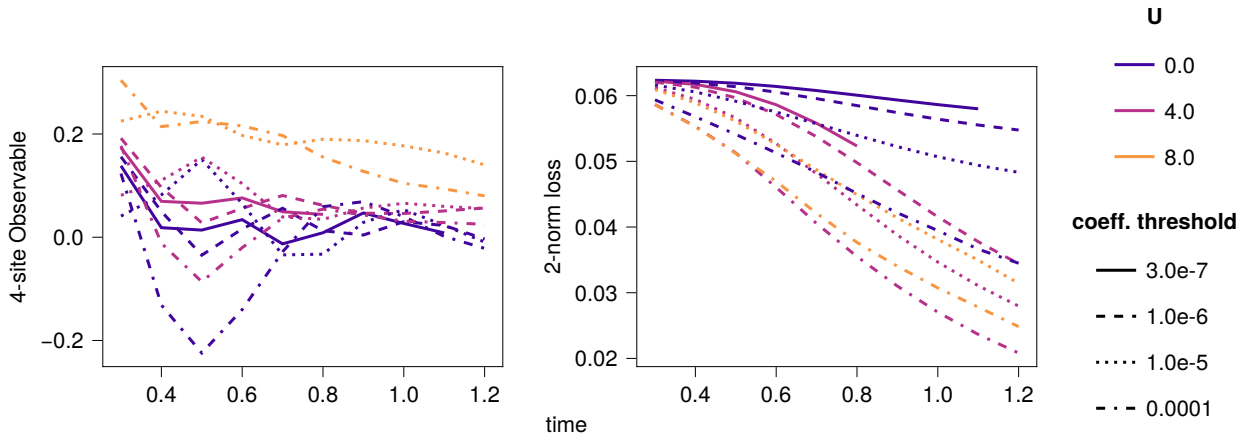
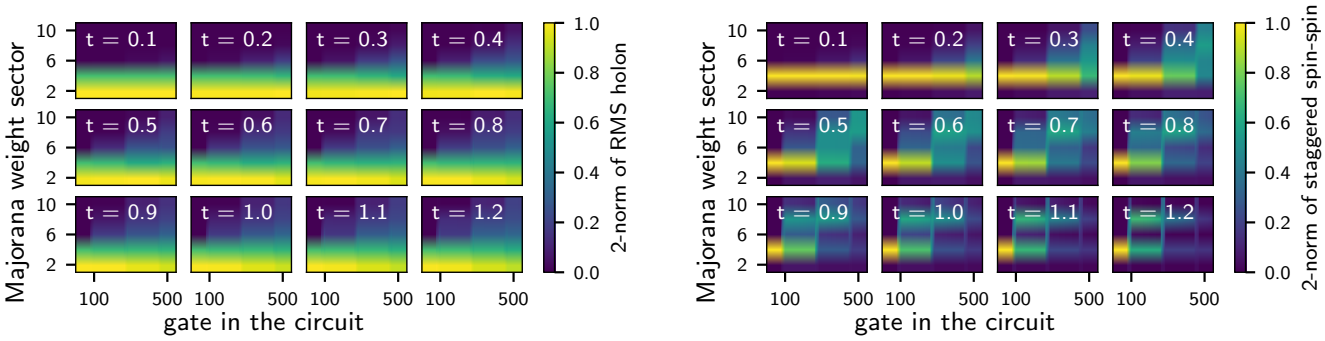


FIG. 46: (left) Time series of the optimised 4-site observable shown in the first column Fig. 50 for different coefficient truncation thresholds and values of U together with (right) the ℓ^2-norm loss incurred by truncation in these observables.

This is in stark contrast to fig. 46, where we show the time series and loss of ℓ^2-norm for Majorana propagation simulations of the optimised 4-site observable shown in the first column of fig. 50. There we see how, even despite a modest ℓ^2-norm loss at time ≈ 0.5, the time series vary widely as a function of the coefficient truncation threshold. And in fig. 50 we also see that even at $U = 0$ the coeff. threshold $= 10^{-6}$ curve does not match the ground truth. Overall, we take this to mean that the convergence with the truncation thresholds is a better proxy for the accuracy of the Majorana propagation simulations than the ℓ^2-norm loss is. That we have such different behaviour between the staggered spin-spin correlator —significant ℓ^2-norm loss but still reasonably converged simulations—and the optimised 4-site observables—less and slower ℓ^2-norm loss, but unconverged Majorana propagation simulations —can be explained with a sign problem. In the case of the staggered spin-spin correlator, each term in the Majorana sum has positive overlap with the initial Néel state. This implies that discarding terms "randomly" by truncation has a controlled effect on accuracy. In the case of the optimised 4-site observable, on the other hand, roughly half of the terms have positive overlap with the initial state, while the other half have negative overlap. This means that

discarding terms "randomly" by truncation leads to much faster loss of accuracy.



(a) (b)

FIG. 47: Spreading of the observables over the Majorana weight sectors as they Heisenberg evolve through the circuit for RMS holon (left) and the spin-spin correlator (right) at all simulated times for $U = 8$ and $\phi = 0$. Note that for the MP simulations we wrote the circuit entirely in the fermionic language, i.e. as a sequence of hopping and onsite evolutions instead of a hardware native sequence of single qubit and CZ gates. This means the gate count for MP simulations is much lower than for the hardware circuits, even though they describe equivalent unitaries.

Fig. 47 shows the impact of the Heisenberg evolution through the circuit of the spread of Majorana polynomials over the different Majorana weights. It is important to note that the weight 2 and 4 sectors can be simulated exactly without truncation: this implies that, for all observables of interest, in the $U = 0$ case, the simulations are essentially intact. For higher values of U, however, this changes depending on the circuit of interest. In particular, while the number of RMS holons remains mostly supported in the weight-2 sector throughout the evolution, for the staggered spin-spin correlation function, this changes dramatically, with support rapidly spreading from the weight-4 sector to higher-weight sectors.

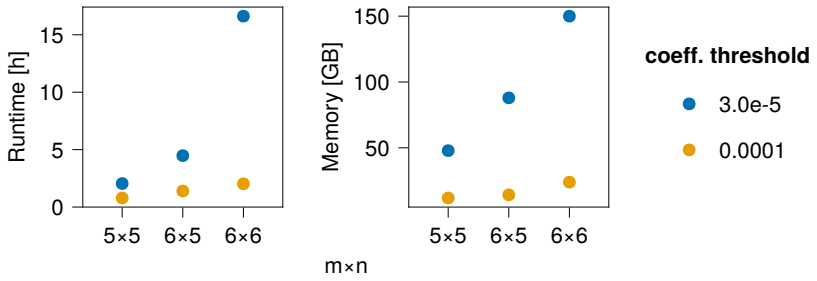


FIG. 48: Maximum memory and runtime requirements across all observables, and time steps for the Majorana propagation simulations .

Appendix H: Many-body quantum correlations and observables

Here, we investigate the many-body entanglement and quantum correlations that build up in the time-evolved quantum states produced by this experiment, and compare the quantum simulation with classical methods through this lens.

1. Background and motivation

We expect the entanglement entropy of time-evolved states of beyond-classical many-body quantum Hamiltonians to tend towards a volume-law scaling over time. Meanwhile, tensor network states necessarily have area-law scaling of entanglement entropy, by construction. However, probing the complex many-body entanglement and correlation structure in these time-evolved states is challenging.

Indeed, we expect that most small subsystems of highly entangled states will have reduced density matrices that are close to maximally mixed. The measurement outcome distributions of arbitrary measurements on maximally mixed states are easily reproducible classically, as are those produced by simple tensor network states. E.g. the GHZ state, which can be represented efficiently classically as a matrix product state (MPS) with a bond-dimension of just 2, already has all 1-body reduced density matrices maximally mixed. Thus, we expect that most single- or few-body observables will not be good candidates to distinguish time-evolved many-body states from tensor network simulations and other leading classical methods. Fortunately, thanks to its programmability, digital quantum computing hardware opens up the possibility of evaluating more complex, genuinely many-body observables and correlations, both in the computational basis and in complementary bases.

2. Localizable distinguishability measures

Tensor network and other classical simulations of many-body quantum dynamics face more stringent demands than simply finding a good approximation to the final output state: they must track the evolution of the state under the Hamiltonian dynamics throughout the evolution, as well as capture the many-body quantum correlations that build up during that evolution.

Similarly to states with volume-law scaling, genuine many-body entanglement and correlations (such as those in the GHZ or cluster states) cannot be detected by 1-body local observables. It can, however, be detected through global quantities. A motivating example of such a quantity is the "localizable entanglement", defined as the expected bipartite entanglement of two qudits, averaged over the outcomes of product measurements on all the other qudits [139]:

$$E^{\mathrm{loc}(i,j)}(\psi) := \mathbb{E}_x E^{(i,j)}(\mathbb{1}_{\sigma_i,\sigma_j} \otimes \Pi^{(x)}_{\mathrm{rest}}|\psi\rangle) \tag{H1}$$

on an n-qudit system, where $x \in \{0,1\}^{n-2}$ is an $(n-2)$-bit string, $\Pi^{(x)} = \bigotimes_{k\in[n]\setminus\{i,j\}} \Pi_k^{(x_k)}$ is a product of local dichotomous projective measurements on all but the i and j'th qudits, and E is some entanglement measure on bipartite pure-states. (The original definition of localizable entanglement [139] additionally maximises over the bases of the local projective measurements, which need not be dichotomous, or even over all local POVMs. But we will restrict here to two-outcome projective measurements in some fixed basis for simplicity.)

Note that the localizable entanglement itself is not a good measure for distinguishing from tensor network states. Not only is it not directly experimentally accessible (at least, not without exponential sampling overhead), but it can also be maximised by tensor network states. E.g. the GHZ state has maximum localizable entanglement on any pair of qubits, despite being a simple MPS with bond-dimension 2. Nonetheless, localizable entanglement is a quantity that is not reducible to few-body local observables, as it depends on the outcomes of measurements across the entire system, yet is able to probe the many-body entanglement structure via outcomes of local 1-body measurements.

This motivates examining the structure of many-body correlations in local measurement outcomes on time-evolved states produced in digital quantum simulations, to assess to what extent the underlying time-evolved states can be reproduced by classical simulation methods, in particular, tensor network methods.

3. Distinguishability measures on marginals

The most fundamental such quantity is the total variation distance between the probability distribution over local measurement outcomes on the quantum hardware and the probability distribution over outcomes of the same

measurements predicted by classical methods. Estimating the total variation distance requires a number of samples that scales exponentially in the number of qubits being measured, so this comparison cannot be done across all of the qubits in the quantum computer. But on hardware with fast clock speeds, such as superconducting circuits, it is practical to estimate the total variation distance of the marginal probability distributions on subsystems of up to around 10 qubits, which requires on the order of 1000s of samples to estimate accurately. (In Appendix I, we study a different method of assessing the difference between the measurement outcome distributions, via cross-entropy benchmarking.)

The total variation distance between two probability distributions over measurement outcomes, $\vec{p}$ and $\vec{q}$, is given mathematically by the ℓ_1-distance: $TVD(\vec{p},\vec{q}) = \frac{1}{2}\|\vec{p}-\vec{q}\|_1$.

As is well-known, the total variation distance is lower-bounded by the difference in expectation value of any other quantity computed over the same pair of probability distributions, with equality attained on a quantity that assigns some sign value ± 1 to each outcome. Thus, we can equivalently express the total variation distance between probability distributions of local dichotomous quantum measurement outcomes on a subset of n qubits, in terms of an optimisation over all many-body observables on those n qubits, with equality attained by some ± 1-valued observable that is diagonal in the local measurement bases. (Or, equivalently, an optimisation over ± 1-valued Boolean functions on the n-bit strings of measurement outcomes.) This well-known fact implies that any observable on some subset of n-qubits lower-bounds the total variation distance between the measurement outcome distributions on those qubits, so could potentially witness a difference between the quantum and classical results.

More generally, we can quantify the difference between two probability distributions via the ℓ_p norms: $\|\vec{p}-\vec{q}\|_p$ for any $1 \leq p \leq \infty$, with different values of p having different operational interpretations. We have seen that the ℓ_1-distance (or total variation distance) quantifies the *best-case* probability of distinguishing between the two distributions, optimised over all possible observables. Whereas the ℓ_2-distance is related to the *average* distinguishing probability, which is arguably more meaningful in an experimental setting:

4. Localizable quantum correlation measures

The total variation distance over measurement outcomes in a fixed basis itself lower-bounds the trace distance between the underlying quantum states. In this case, between the quantum state on the quantum computer and, e.g. the quantum state described by the tensor network produced by classical computation. On digital quantum computers, we have access to measurements in arbitrary local bases by performing just one additional layer of single-qubit gates before the final measurement. Quantum correlations and entanglement (as opposed to classical correlations) cannot be witnessed by measurements in one fixed set of local bases. But access to measurements in complementary bases does, as famously demonstrated by Bell inequalities. Therefore, access to measurements in multiple, complementary bases on a digital quantum computer affords more opportunity to probe the quantum many-body entanglement structure of the state.

With access to measurements in arbitrary bases, we can, in principle, calculate the trace distance itself. But this amounts to performing full state tomography, which requires not only an exponential number of *samples*, but also measurements in an exponential number of different local bases. This becomes intractable in practice beyond a much smaller number (3 or 4, say) of qubits. Thus, we must instead look for quantities that use measurements in a small number of different local bases judiciously to have something implementable in practice.

One such quantity, motivated by localizable entanglement, is the "localizable ℓ_p-distance":

$$D_p^{\mathrm{loc}(i,j)}(|\psi\rangle,|\varphi\rangle) := \mathbb{E}_x \, \frac{1}{2} \left\| (\mathbb{1}_{\sigma_i,\sigma_j} \otimes \Pi_{\mathrm{rest}}^{(x)})(|\psi\rangle - |\varphi\rangle) \right\|_p . \tag{H2}$$

This is a quantity that can be obtained from local measurement outcomes, yet cannot be reduced to expectations of few-body local observables, nor to the computational basis distribution.

Why might we expect this quantity to discriminate between tensor network states and states with volume-law entanglement entropy scaling and genuinely many-body correlations? For simplicity, consider an MPS state (1D tensor network state), and assume qudits i, j are some distance apart along the MPS chain. It is helpful here to think of the MPS in the equivalent projected-entangled-pair picture, where the state is constructed by Bell pairs between neighbouring virtual qudits, which are locally "projected" (mapped by a not-necessarily-invertible linear operator) to the physical Hilbert space. To produce the correct two-qudit quantum state conditioned on all the intermediate local measurement outcomes, the quantum correlations will have to be faithfully propagated all along the chain by entanglement swapping along the virtual Bell pairs. Unless the MPS "projectors" implement perfect entanglement swapping (as in the MPS for the GHZ state), we might expect to lose fidelity as the correlations are propagated along each link of the MPS, resulting in poor overall state fidelity by the time we reach all the way from qudit i to qudit j.

Of course, we could choose a different ordering for the MPS chain so that i, j are now close together in the MPS chain ordering. But then some other pair i', j' will be far apart. Typically, to accurately track the time evolution, the MPS ordering needs to be fixed according to the interaction graph of the Hamiltonian being simulated. Thus, only a handful of sites $\{(i)\}$ need to be considered to ensure that at least some pair of sites in the set are far apart in the MPS ordering. Although we have sketched the argument for 1D MPS, a similar argument goes through for 2D PEPS, albeit with a square-root reduction in the distance that the quantum correlations have to be propagated by entanglement-swapping along the PEPS.

The localizable ℓ_p-distance can straightforwardly be related directly to many-body correlations in the outcomes of single-qubit measurements, which are readily accessible in digital quantum simulations.

$$D_1^{\mathrm{loc}(x,y)}(|\psi_Q\rangle, |\psi_{TN}\rangle) \tag{H3}$$

$$= \sum_{x\in\{0,1\}^{n-2}} \Pr(x) \sum_{\sigma_i,\sigma_j\in\{\mathbb{1},X,Y,Z\}} \left|\mathrm{tr}\left(\sigma_i\otimes\sigma_j(\psi_Q^{(x)} - \psi_{TN}^{(x)})\right)\right| \tag{H4}$$

where $\psi_Q^{(x)} = (\mathbb{1}_{\sigma_i,\sigma_j}\otimes\Pi_{\mathrm{rest}}^{(x)})\,|\psi_Q\rangle\langle\psi_Q|\,(\mathbb{1}_{\sigma_i,\sigma_j}\otimes\Pi_{\mathrm{rest}}^{(x)})$ is the state conditioned on outcome x, and similarly for $\psi_{TN}^{(x)}$,

$$= \sum_x \Pr(x) \max_{S(x,\sigma_i,\sigma_j)\in\{0,1\}} \sum_{\sigma_i,\sigma_j} (-1)^{S(x,\sigma_i,\sigma_j)}\,\mathrm{tr}\left(\sigma_i\otimes\sigma_j(\psi_Q^{(x)} - \psi_{TN}^{(x)})\right) \tag{H5}$$

$$= \sum_x \Pr(x) \max_{S(x,\sigma_i,\sigma_j)} \sum_{\sigma_i,\sigma_j} (-1)^{S(x,\sigma_i,\sigma_j)} \sum_{a,b\in\{0,1\}} \mathrm{tr}\left(\lambda_{ab}^{(\sigma_i,\sigma_j)}\Pi_{ab}^{(\sigma_i\sigma_j)}(\psi_Q^{(x)} - \psi_{TN}^{(x)})\right) \tag{H6}$$

where we have taken the spectral decomposition of the Pauli $\sigma_i\otimes\sigma_j = \sum_{ab}\lambda_{ab}\Pi_{ab}$ with $\lambda_{ab} = \pm 1$,

$$= \sum_x \Pr(x) \max_{S(x,\sigma_i,\sigma_j)} \sum_{\sigma_i,\sigma_j} (-1)^{S(x,\sigma_i,\sigma_j)} \sum_{a,b\in\{0,1\}} \lambda_{ab}^{(\sigma_i,\sigma_j)} \left(\Pr_Q(ab|x\sigma_i\sigma_j) - \Pr_{TN}(ab|x\sigma_i\sigma_j)\right) \tag{H7}$$

$$= \max_{S(x,\sigma_i,\sigma_j)} \sum_{x,\sigma_i,\sigma_j,a,b} (-1)^{S(x,\sigma_i,\sigma_j)}\lambda_{ab}^{(\sigma_i,\sigma_j)}\big(\Pr_Q(xab|\sigma_i,\sigma_j) - \Pr_{TN}(xab|\sigma_i,\sigma_j)\big) \tag{H8}$$

$$= \lim_{\#\mathrm{samples}\to\infty} \max_{S(x,\sigma_i,\sigma_j)} \Big(\frac{1}{\#\text{Q samples}} \sum_{\sigma_i,\sigma_j}\sum_{x,a,b} (-1)^{S(x,\sigma_i,\sigma_j)}\lambda_{ab}^{(\sigma_i,\sigma_j)} - \frac{1}{\#\text{TN samples}} \sum_{\sigma_i,\sigma_j}\sum_{x,a,b} (-1)^{S(x,\sigma_i,\sigma_j)}\lambda_{ab}^{(\sigma_i,\sigma_j)}\Big) \tag{H9}$$

$\Pr_Q(xab|ij)$ can be estimated empirically as $\frac{\#xab\text{ outcomes}}{\#\text{ samples}}$ from the output bit-strings obtained by measuring the Pauli strings $\sigma_x^{(i)}ZZZ\ldots Z\sigma_y^{(j)}$ on $|\psi_Q\rangle$ on quantum hardware, and $\Pr_{TN}(ab)$ can similarly be estimated from the tensor network simulation. Concretely, this means summing the sign factors over all the sampled bit-strings and then normalising by the total number of samples (Eq. (H9)). This empirical estimate will converge to the true value in a central-limit way, so it can be estimated from however many samples are practical on the hardware.

Meanwhile, the sign factors $(-1)^{S(x,i,j)}$ can be optimised over in classical post-processing. A brute-force search over all possible sign factors will require exponential-time classical post-processing, which may be acceptable. On the other hand, as it is a maximisation, even a heuristic optimisation that does not attain the true maximum might suffice to give good discrimination from tensor network states.

As with the total variation distance, it is not practical to perform this on very large subsets of qubits, due to the exponential sample-overhead. But subsets of around 10 qubits are feasible with superconducting circuit clock speeds.

5. Results

First, we show that for many relatively small marginals, the experiment achieves a better approximation of the known exact marginal at $U = 0$ than that achieved by TDVP, as measured by the total variation distance, while also achieving different results to TDVP for $U = 4$ (where an exact solution is not available). We focus on 5×5 instances with a central hole and Néel starting state, and make three comparisons:

1. Average total variation distance between experiment and FLO over marginals on k sites, for $U = 0$;

2. Average total variation distance between TDVP and FLO over marginals on k sites, for $U = 0$;

3. Average total variation distance between experiment and TDVP over marginals on k sites, for $U = 4$.

In each case, we compute total variation distances based on sampling from the experiment (~ 7000 samples per data point) / FLO simulator (1000 samples) / MPS (10,000 samples). In the case of the quantum algorithm, we use error mitigation on shots as described in Appendix E 10. It is important to note that the $U = 0$ FLO simulation here are not simulations of the quantum circuit run on device, but rather of much higher-depth (20) Trotterized FLO circuits for the $U = 0$ Trotter dynamics, where the Trotter error is much smaller. Therefore, the FLO simulations we are comparing to are essentially the underlying, continuous-time evolution under the Fermi-Hubbard Hamiltonian itself. This comparison therefore accounts for all sources of approximation error in the quantum hardware, TDVP and Majorana propagation results: the various sources of algorithmic, truncation and approximation errors in all approaches, and (in the case of the quantum simulation) hardware noise.

The results are shown in Fig. 49. We observe that the error achieved by TDVP increases over time, as expected. By contrast, the error of the quantum algorithm remains relatively flat over time, except for short-time fluctuations, which we attribute to the behaviour of our error mitigation procedure. In particular, for the longest times we consider, the error achieved by the quantum algorithm is lower than the error achieved by TDVP for marginals on up to 4 sites. For 5 sites, TDVP achieves lower errors on average. In addition, we see that for $U = 4$, there is a significant difference between the experimental results and TDVP results throughout.

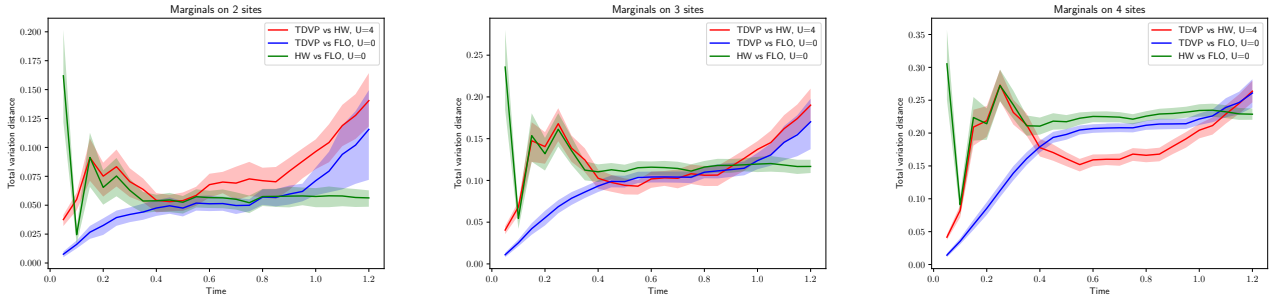


FIG. 49: Total variation distances for marginals on 2–4 sites between output distributions achieved by various methods. Plots show mean TVD and upper and lower quartiles for each pair of methods considered.

Next, we seek an *observable* that achieves significant separations between the quantum algorithm and TDVP. To find a good local observable for separating these two conditions, we use the following procedure, motivated by the concept of localisable correlations discussed above, albeit restricted, for now, to computational basis measurements.

A (computational basis) observable on a marginal of k qubits can be specified by a vector $v \in \mathbb{R}^{2^k}$, where v_x is the value taken by v on computational basis state $|x\rangle$, $x \in \{0,1\}^k$. The expectation of that observable with respect to a given marginal distribution $m : \{0,1\}^k \to \mathbb{R}$ is just the inner product $v \cdot m$. We aim to find an observable that simultaneously maximises the difference between the marginal produced by the experiment and the marginal produced by TDVP, over the entire range $t = 0, \ldots, 1.2$, and $U \in \{0, 4\}$. It is computationally advantageous to maximise this difference with respect to the ℓ_2 norm. Then, abstractly, this is the problem of finding a vector v such that $\sum_i (v \cdot (a^i - b^i))^2$ is as large as possible, for some vectors a^i, b^i. Writing $\delta^i = a^i - b^i$, we want to compute $\sum_i (v \cdot \delta^i)^2$, which is just $\|Dv\|^2$ for some matrix D that has the vectors δ^i as rows. The solution that maximises this (relative to the Euclidean norm of v) is the principal right singular vector of D, which can be found efficiently using the singular value decomposition. One final technical aspect is that we only compute this difference over times in the second half of the range of times t, where genuine differences are expected to be found, rather than statistical fluctuations in the short-time regime where TDVP is accurate.

For any marginal (subset of qubits), the above method produces an observable. We illustrate in Fig. 50 some examples of marginals where large differences are obtained between the observable values produced by experiment and TDVP, and in one case, Majorana propagation. In each case, we produce a marginal on $2k$ qubits by choosing a subset of k sites and considering both spin sectors, for $k = 4, \ldots, 7$. We also show the values of this observable calculated by exact simulation at $U = 0$. We see that the experiment closely matches the exact simulation for $U = 0$, despite this not being built into the definition of the observable, whereas TDVP significantly differs from it. For $U = 4$, we see significant divergence between TDVP and the experiment at large times.

Error bars are computed using the method described in Appendix E 10. The exact dynamics for $U = 0$ are again computed by high-depth (20) FLO circuit simulations, which accurately approximate the underlying continuous-time

Hamiltonian dynamics. We observe that these error bars are much larger for short times, which we attribute to the effective number of samples being low. This, in turn, can be caused by the local observables used for shot reweighting being highly polarised, the overall entropy being low, and the weights effectively "zooming in" on a small number of samples.

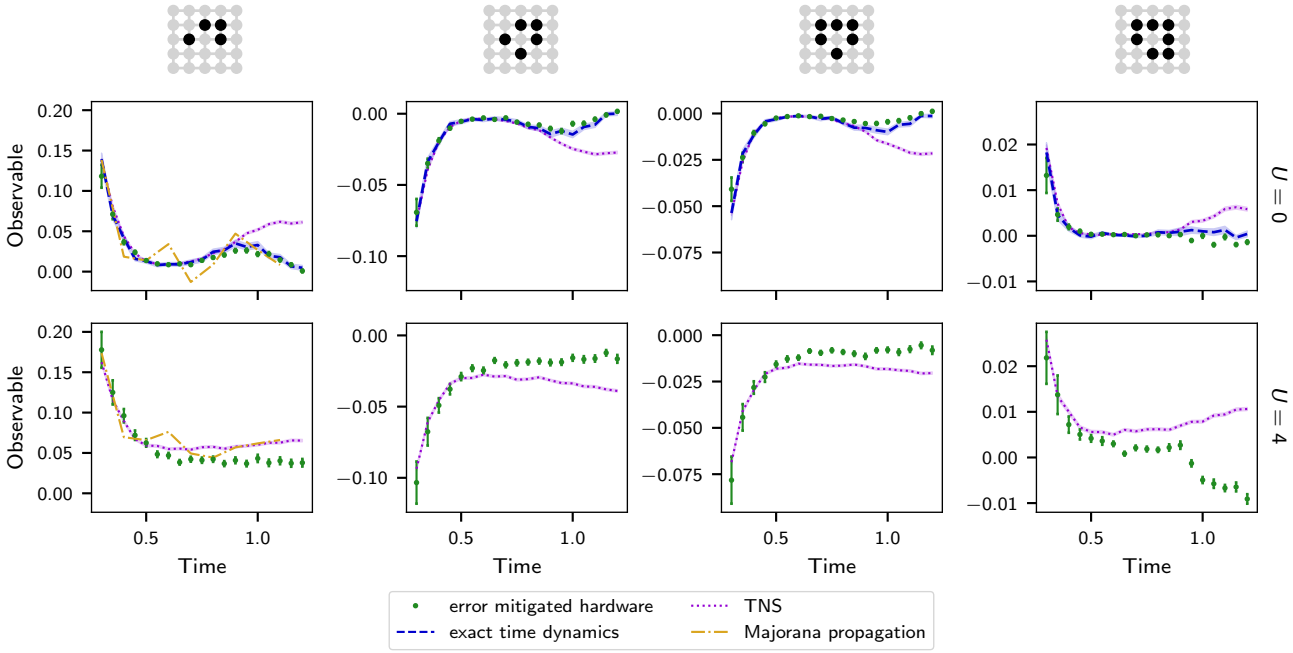


FIG. 50: Optimised observables for marginals on 4, 5, 6 and 7 sites in the 5×5 lattice to distinguish between computational methods. Each plot shows TDVP (TNS) results using $\chi = 2048$ and experimental (error-mitigated hardware) results with MESR shot error mitigation, for two values of the Fermi-Hubbard parameter $U = 0$ and $U = 4$. The exact results are also shown for $U = 0$ computed by FLO, and Majorana propagation results for the 4-site observable.

Fig. 50 compares the expectation value of the observable computed directly on the state produced by the TDVP simulation, with the expectation value estimated from the error-mitigated bit-strings obtained on the quantum hardware. However, we can also apply our novel MESR shot-reweighting error-mitigation method to bit-strings sampled from the *classical* TDVP simulation. Figures 51 and 52 show that applying the same error-mitigation method to the TDVP simulations allows us to recover the correct dynamics of these observables for $U = 0$, where exact computations are feasible. Fig. 51 shows the results obtained on the identical optimised observables as in Fig. 50, which are obtained by optimising against the TDVP simulation itself. Fig. 51 shows the results obtained when the observables are optimised against the error-mitigated TDVP data. In both cases, for $U = 4$, where exact classical computation is not possible, the quantum simulation and TDVP simulation results differ. As both show comparable agreement with exact results and with one other on $U = 0$, it is not possible to determine for these many-body observables which of the simulations, quantum or TDVP, is more accurate for the interacting $U = 4$ Fermi-Hubbard model.

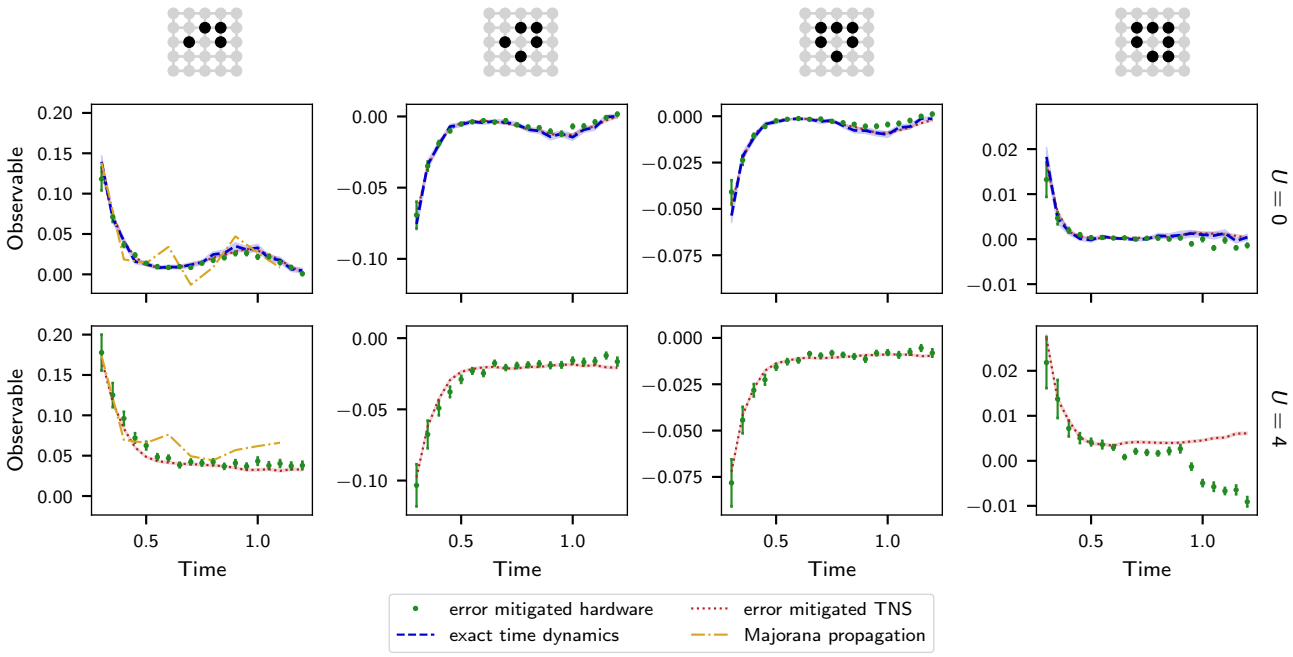


FIG. 51: Optimised observables with respect to mitigated experimental and unmitigated MPS data for marginals on 4, 5, 6 and 7 sites in the 5×5 lattice to distinguish between computational methods. Each plot shows mitigated TDVP (error-mitigated TNS) results using $\chi = 2048$ and experimental (error-mitigated hardware) results with MESR shot error mitigation, for two values of the Fermi-Hubbard parameter $U = 0$ and $U = 4$. The exact results are also shown for $U = 0$ computed by FLO, and Majorana propagation results for the 4-site observable.

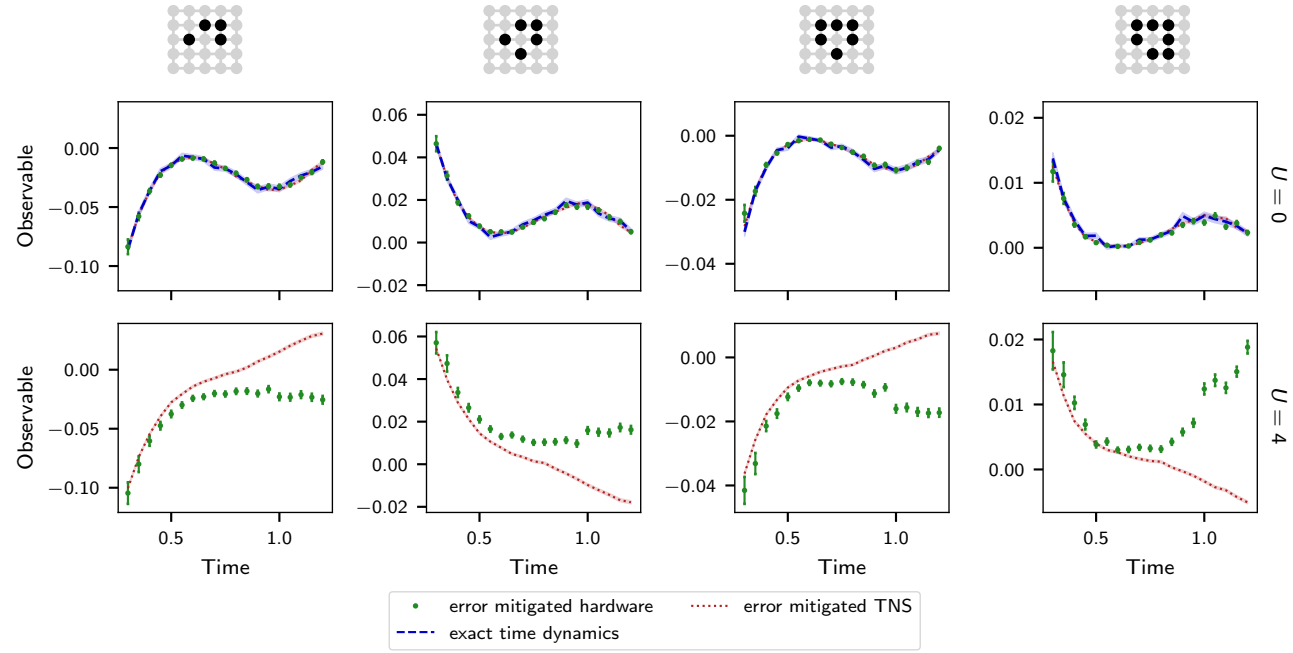


FIG. 52: Optimised observables with respect to mitigated experimental and mitigated MPS data for marginals on 4, 5, 6 and 7 sites in the 5×5 lattice to distinguish between computational methods. Each plot shows mitigated TDVP (error-mitigated TNS) results using $\chi = 2048$ and experimental (error-mitigated hardware) results with MESR shot error mitigation, for two values of the Fermi-Hubbard parameter $U = 0$ and $U = 4$. The exact results are also shown for $U = 0$ computed by FLO.

Appendix I: Cross-Entropy Benchmarking

Cross-entropy benchmarking is by now a well-established method for validation of NISQ devices [74–78], using either exponential classical resources to compute ideal amplitudes or efficient classical simulation of special ensembles of circuits (such as Clifford circuits [140] or number-conserving circuits with $\mathcal{O}(1)$ [141]). Existing applications of this method rely on approximate Haar random states as output, e.g. produced through ergodic circuit ensembles [77] or time averaging in an analogue simulator [142]. Here, we adapt the statistical cross-entropy toolkit to highly entangled (but not at all random) output states of the Fermi-Hubbard simulation results described in the main text. We extend linear XEB to the setting of free fermion evolutions by using FLO determinants to efficiently compute output amplitudes, along with related classical simulation methods to determine the time-dependent collision probability of the ideal circuit. This allows us to perform XEB verification without using exponential-time classical resources.

The goal of our benchmarking protocol is to determine whether the device samples are reflected with a higher-than-average probability in the ideal *untrotterised* output state of the circuit, leading to a fidelity estimate that combines the effects of device noise and Trotterisation to benchmark the entire digital implementation of the time dynamics.

1. Linear XEB for FLO circuits

The standard LXEB quantity for random circuit sampling is

$$F = 2^n \langle p_{\text{ideal}} \rangle_{\text{device}} - 1$$

in which $\langle p_{\text{ideal}} \rangle_{\text{device}}$ is estimated by summing the ideal circuit probabilities $p_{\text{ideal}}(z) = |\langle z|U|0^n\rangle|^2$ (for more details see Appendix E 5) of sufficiently many bit string sample z observed from the device. If the device were noiseless, then this expectation value would be the collision probability of the circuit output in the Z basis

$$\langle p_{\text{ideal}} \rangle_{\text{ideal}} = \sum_{z \sim p_{\text{ideal}}} p_{\text{ideal}}(z) = \sum_z p_{\text{ideal}}(z)^2 \equiv C_{\text{ideal}}$$

This computational basis collision probability C_{ideal} for approximately Haar random circuits in Hilbert space dimension D is tightly concentrated on $C_{\text{ideal}} = 2/D$, and so a noiseless implementation of random quantum circuits will find $F = 1$. In contrast, the collision probability for the uniform distribution is $C_{\text{uniform}} = 1/D$, which has $F = 0$. The main reason that F is scaled and shifted this way is so that it corresponds directly to an overall circuit fidelity under the global depolarising noise model,

$$\rho_{\text{device}} = F|\psi_{\text{ideal}}\rangle\langle\psi_{\text{ideal}}| + (1-F)(I/D)$$

The ansatz above implies $p_{\text{device}}(z) = F p_{\text{ideal}}(z) + (1-F)(1/D)$ and so the estimator becomes

$$\langle p_{\text{ideal}} \rangle_{\text{device}} = \sum_{z \sim p_{\text{ideal}}} p_{\text{ideal}} = \sum_z p_{\text{ideal}}(z)(F p_{\text{ideal}}(z) + (1-F)(1/D)) = F(2/D) + (1-F)(1/D) = (1+F)/D$$

which explains how the definition of F corresponds to a circuit fidelity. Summarising, the method requires (1) classical simulation of the ideal output probabilities to compute the estimator of $\langle p_{\text{ideal}} \rangle_{\text{device}}$, and (2) knowledge of the collision probability for the ideal output state, in order to compute the correct rescaling and shifting to relate $\langle p_{\text{ideal}} \rangle_{\text{device}}$ to a circuit fidelity F under the global depolarizing noise ansatz.

For $U = 0$ circuits with appropriate input states, we can compute amplitudes in polynomial time, so we can compute the estimator $\langle p_{\text{ideal}} \rangle_{\text{device}}$ using samples from the device. To understand how to scale it into a fidelity under the global depolarising model, we need to estimate the time-dependent collision probability

$$C(t) = \sum_z p_{\text{ideal}}(z,t)^2$$

which we can approximate at each point in time using a Monte Carlo estimator for $C = \langle p_{\text{ideal}} \rangle_{\text{ideal}}$ that combines efficient strong and weak simulation of FLO circuits. Now, if we assume the global depolarising noise ansatz with fidelity F,

$$\langle p_{\text{ideal}} \rangle_{\text{device}} = \sum_z p_{\text{ideal}}(z)(F p_{\text{ideal}}(z) + (1-F)/D) = FC + (1-F)/D$$

This motivates defining the FLO-LXEB fidelity to be

$$F = \frac{D\langle p_{\text{ideal}}\rangle_{\text{device}} - 1}{CD - 1}$$

Note that the collision probability is always $C \geq D^{-1}$, so the denominator is always sensible. Since the method is designed around the benchmarking of highly entangled states, we exclude short times when $C \gg D^{-1}$, and present results for times $0.8, ..., 1.2$, when the collision entropy of the ideal distribution is nearly saturated.

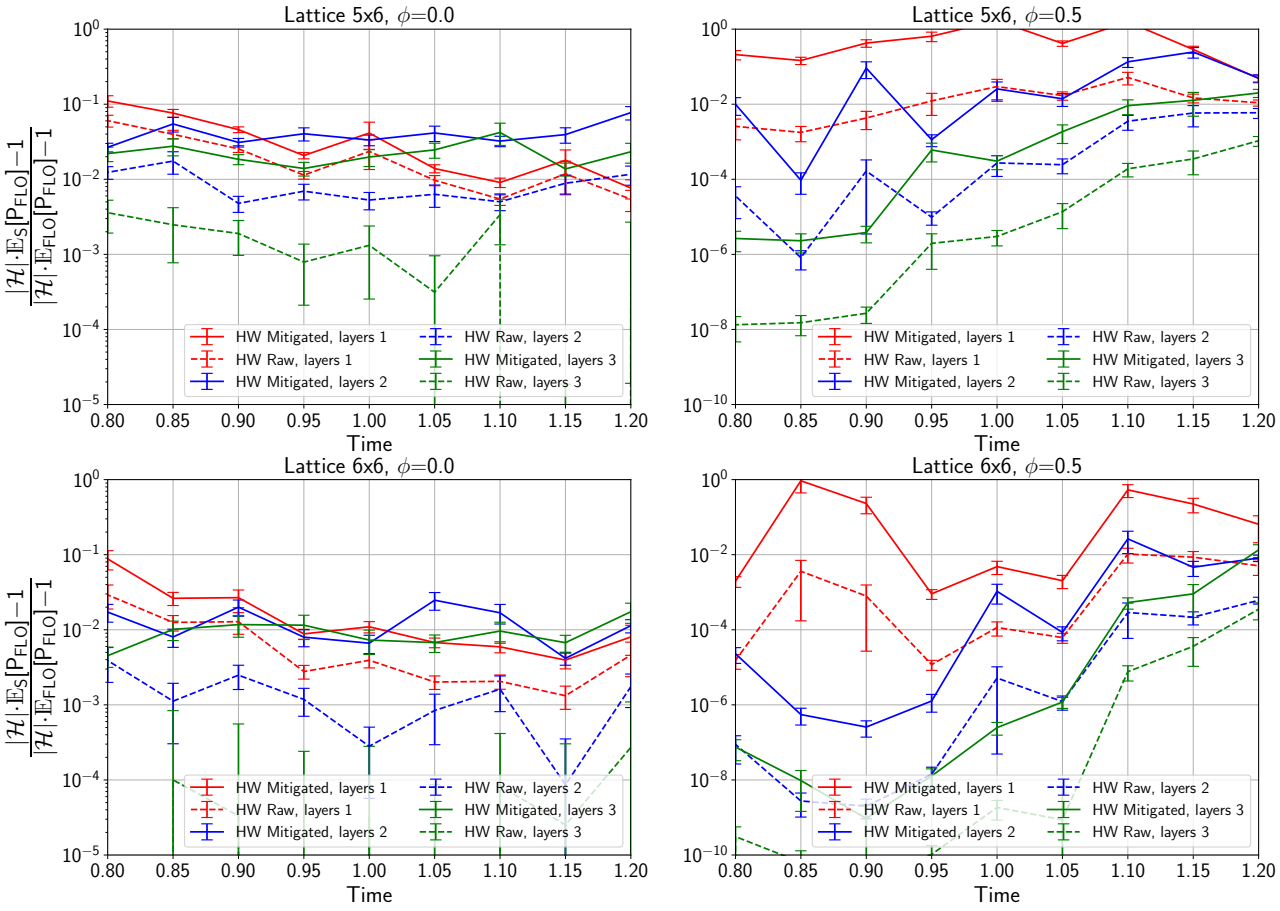
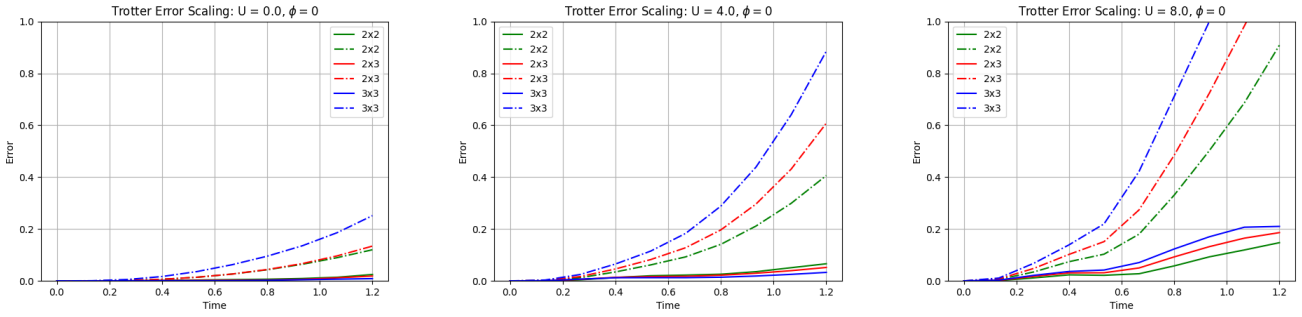


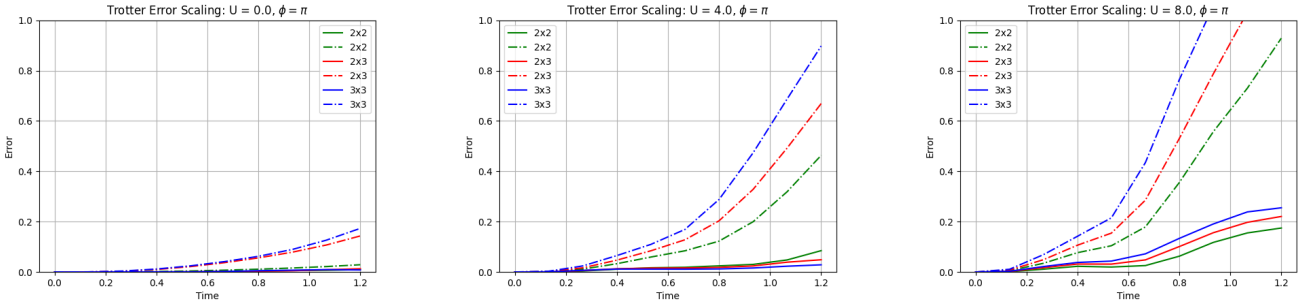
FIG. 53: Estimates of circuit fidelity under the global depolarising noise model using linear XEB applied to FLO circuits. In all cases, the samples from the device are compared to ideal amplitudes from untrotterised FLO, and so the fidelity represents the total overlap of the digital implementation with the ideal state, including device noise and Trotter error.

Linear XEB results for unmitigated samples, mitigated samples (produced by the maximum entropy shot reweighting of Appendix E 10) and different numbers of Trotter layers are shown in Fig. 53. The mitigated results show fidelities around the 1% level at late times, and also enable a comparison of simulations with different numbers of Trotter layers. In the $\phi = 0$ results, we can discern a negative slope for the $n_{\text{layers}} = 1$ results (indicating an exponential decay in fidelity with evolution time) while the $n_{\text{layers}} = 2, 3$ do not show this decay trend and achieve a higher fidelity at time $t = 1.2$.

Appendix J: Trotter Error



(a) $U = 0$, $\phi = 0$ (b) $U = 4$, $\phi = 0$ (c) $U = 8$, $\phi = 0$



(d) $U = 0$, $\phi = \pi$ (e) $U = 4$, $\phi = \pi$ (f) $U = 8$, $\phi = \pi$

FIG. 54: Scaling of average error in state infidelity (dashed) and observable expectation value (solid) with time for coupling strength $U = 0, 4, 8$ and magnetic field $\phi = 0, \pi$ for lattice sizes $L_x \times L_y = 2 \times 2, 2 \times 3$ and 3×3.

We evaluate the impact of Trotterisation on state infidelity and the expectation values of observables for small lattice sizes. We use three second-order Trotter steps and look at times between $t = 0$ and $t = 1.2$ and coupling strengths $U = 0, 4, 8$. We compute the exact time-evolved state $|\phi(t)\rangle = e^{-itH}|\phi(0)\rangle$ and the Trotterised time-evolved state, $|\phi'(t)\rangle = \prod_{j=1}^{3} S_2(t/3)|\phi(0)\rangle$. We then calculate

$$2\sqrt{1 - |\langle\phi(t)|\phi'(t)\rangle|^2} \tag{J1}$$

for a variety of initial states $|\phi(0)\rangle$. We look at a Néel ordered state, hole-doping in the centre and upper left corner of the lattice, an upper and lower side hole and a vertical holon stripe.

For all Pauli-Z observables between weight-1 and weight-4, we compute the error in expectation value

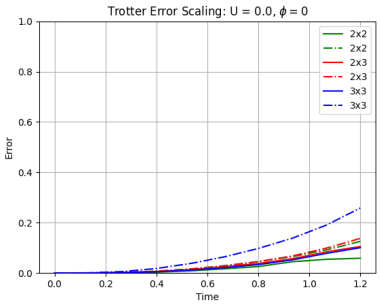
$$|\langle\phi(t)| P |\phi(t)\rangle - \langle\phi'(t)| P |\phi'(t)\rangle|. \tag{J2}$$

We then compute the average of both metrics over all states and observables at each time, the result of which is shown in Fig. 54 and also the maximum of both metrics over all states and observables, as shown in Fig. 54.

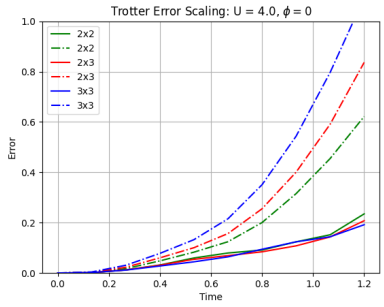
We find that for $U = 0$ and $U = 4$ with and without magnetic flux, the average Trotter error on up to weight 4 observables remains below 10% for all times, with no clear increasing trend as a function of system size. However, for $U = 8$, the Trotter error begins to become more dramatic, increasing to up to %25.

There are good reasons to suspect that Trotter error for this model should not dramatically increase with system size for a given fixed time. This is because Trotter error is upper bounded by local commutator relations [13, 143], which do not increase with system size for intensive observables. It has also been shown that this system size independence of Trotter error can hold for relatively large Trotter steps, based on the separate study of Trotterised dynamics as Floquet dynamics [144].

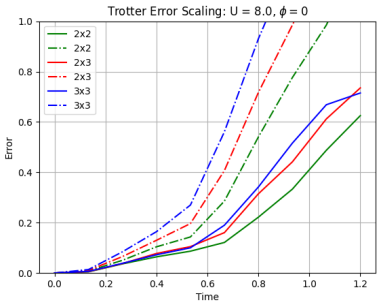
Additionally, we repeat this analysis for observables, which scale with system size. We compute the mean and maximum Trotter error at each time for all weight-4 observables for size 2×2, all weight-6 observables for size 2×3 and all weight-8 observables for size 3×3. We do this for the initial state, which corresponds to hole doping at the centre of the lattice. The results of this analysis can be found in Fig. 56 and Fig. 57.



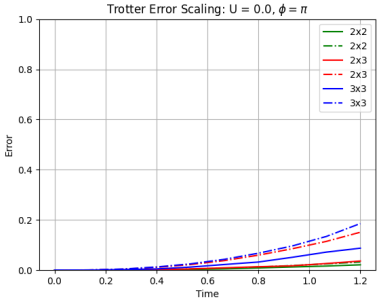
(a) $U = 0$, $\phi = 0$



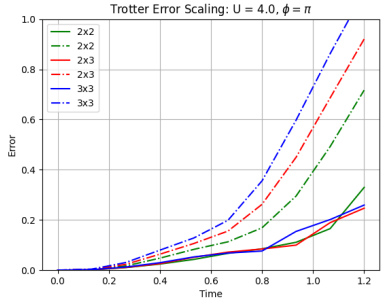
(b) $U = 4$, $\phi = 0$



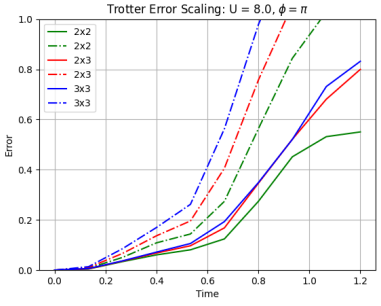
(c) $U = 8$, $\phi = 0$



(d) $U = 0$, $\phi = \pi$

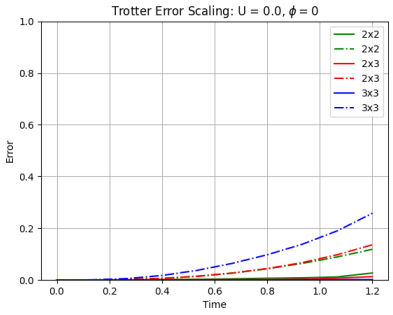


(e) $U = 4$, $\phi = \pi$

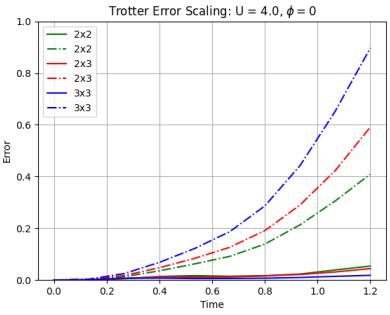


(f) $U = 8$, $\phi = \pi$

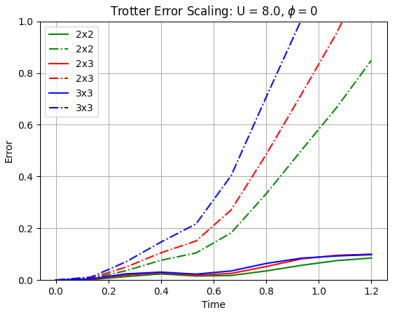
FIG. 55: Scaling of maximum error in state infidelity (dashed) and observable expectation value (solid) with time for coupling strength $U = 0, 4, 8$ and magnetic field $\phi = 0, \pi$ for lattice sizes $L_x \times L_y = 2 \times 2, 2 \times 3$ and 3×3.



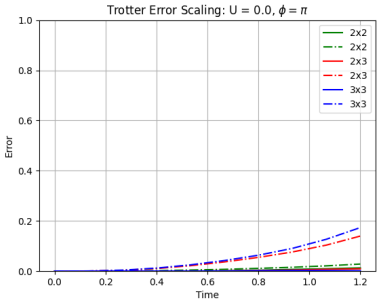
(a) $U = 0$, $\phi = 0$



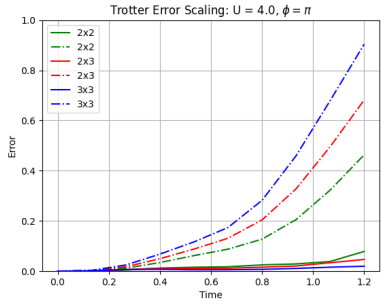
(b) $U = 4$, $\phi = 0$



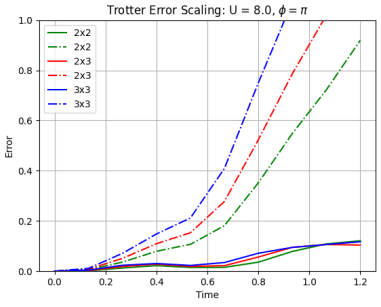
(c) $U = 8$, $\phi = 0$



(d) $U = 0$, $\phi = \pi$

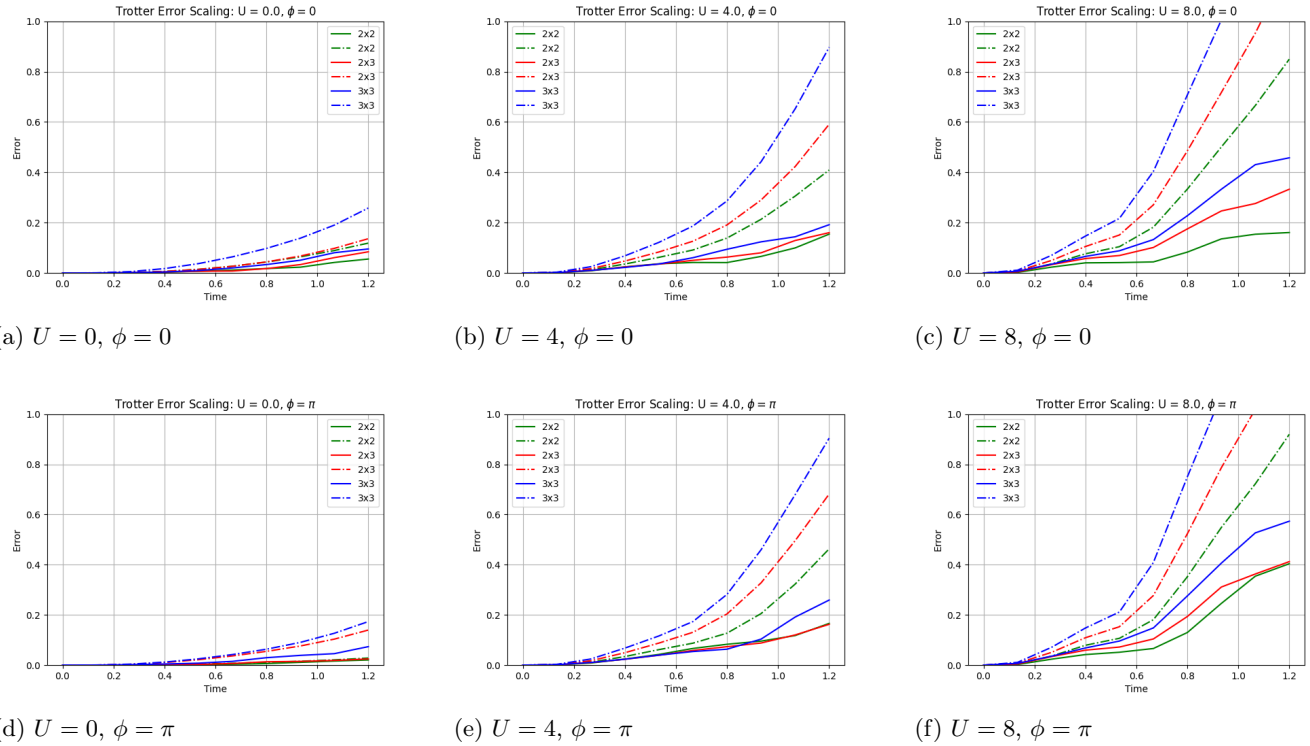


(e) $U = 4$, $\phi = \pi$



(f) $U = 8$, $\phi = \pi$

FIG. 56: Scaling of state infidelity (dashed) and of maximum error in observable expectation value (solid) with time for coupling strength $U = 0, 4, 8$ and magnetic field $\phi = 0, \pi$ for lattice sizes and observable weights: $L_x \times L_y = 2 \times 2$ and weight-4, 2×3 and weight-6, 3×3 and weight-8.



(a) $U = 0$, $\phi = 0$ (b) $U = 4$, $\phi = 0$ (c) $U = 8$, $\phi = 0$

(d) $U = 0$, $\phi = \pi$ (e) $U = 4$, $\phi = \pi$ (f) $U = 8$, $\phi = \pi$

FIG. 57: Scaling of state infidelity (dashed) and of maximum error in observable expectation value (solid) with time for coupling strength $U = 0, 4, 8$ and magnetic field $\phi = 0, \pi$ for lattice sizes and observable weights $L_x \times L_y = 2 \times 2$ and weight-4, 2×3 and weight-6, 3×3 and weight-8.

For $U = 0$ and $U = 4$, with and without magnetic flux, the average error Trotter error in observables remains below %10, and there is still no obvious trend in system size.

Although the above analyses are not reasonably possible for a 4×4 system, we can simulate both the exact circuit and target continuous dynamics of the Fermi-Hubbard model for a 4×4 system with some effort. We have also run our experiments on a 4×4 system. So we can compare the performance of the Trotter circuit to the target dynamics, and compare these to the hardware signal, TDVP and MP. This is shown for a few of the global target observables considered in this work in Fig. 58.

We see that for $U = 0$, the Trotter error is negligible at all times, and all signals approximately agree. For $U = 4$, Trotter error begins to play an important role at approximately $t = 0.9$, and for $U = 8$ Trotter error begins to play an important role at about $t = 0.5$

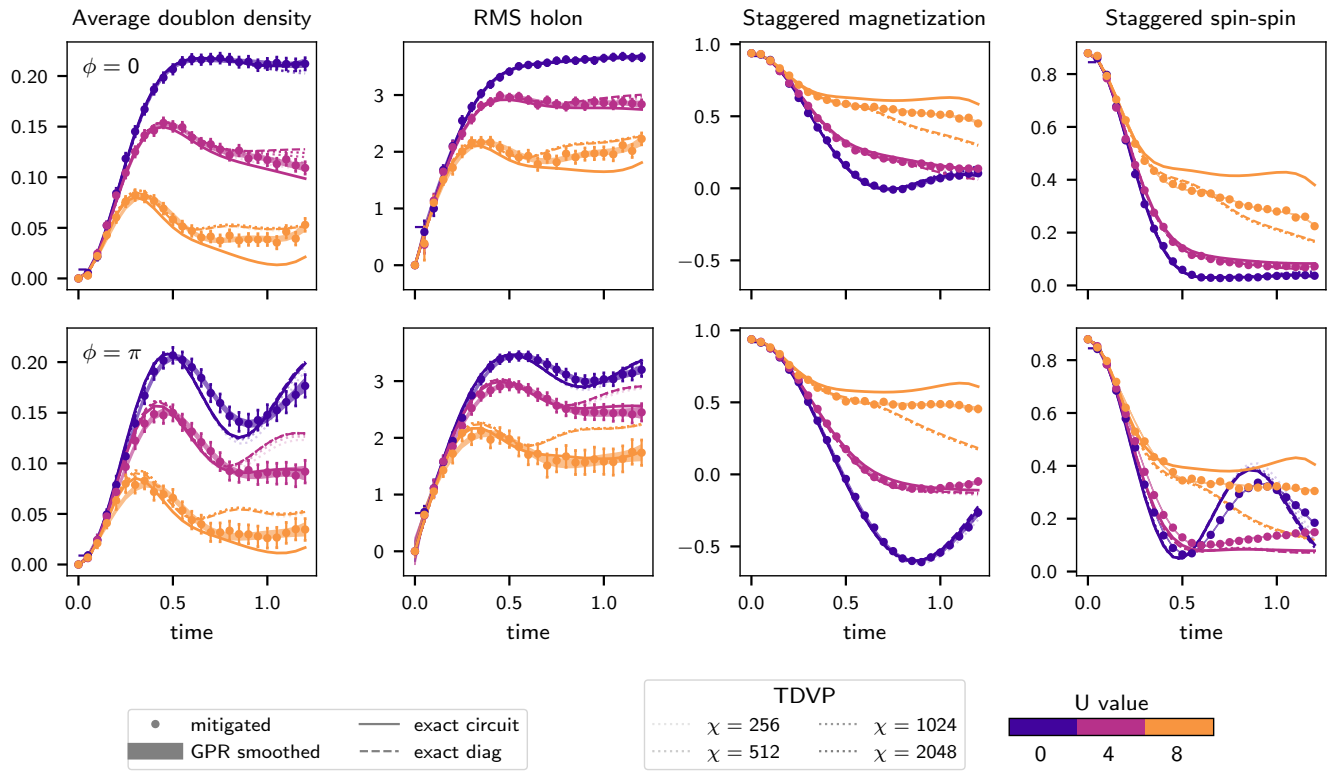


FIG. 58: Evolution of average doublon density, RMS holon, staggered magnetisation and staggered spin-spin correlation for the 4×4 lattice with a central hole for $\phi = 0$ (top row) and $\phi = \pi$ (bottom row), obtained from TFLO mitigated experimental results using 34 qubits and 1220 two-qubit gates. Here we also plot the results obtained using tensor network simulations at different bond dimensions, and exact circuit and TDS evolution when possible ($U = 0, 4, 8$).

Appendix K: Additional holon stripe discussion

In Appendix B 2, we presented the time evolution of the 5×5 lattice with an initial state of a holon stripe in a background which is exactly Néel in the squeezed space (i.e. when the holon stripe is removed) with post-selection on doublon number and antiferromagnetic-ness. We include a discussion here on the sensitivity of the percolation observable to post-selection of shots and noise. By comparing the results with and without post-selection, we see that the proliferation of background holons confounds the percolation signal as a consequence of this sensitivity. This is demonstrated by comparing Fig. 59 vs Fig. 60.

We also note that the performance trend between 4×4, 5×5 and 6×6 systems — given in Figs. 59, 13 and 61 respectively. We see that as our system size scales, the signal performance degrades rapidly, with TFLO failing for 6×6. In Appendix K 1, we include a discussion of the fragility of the percolation observable to noise by employing a simple measurement error model on the post-selected TDVP samples.

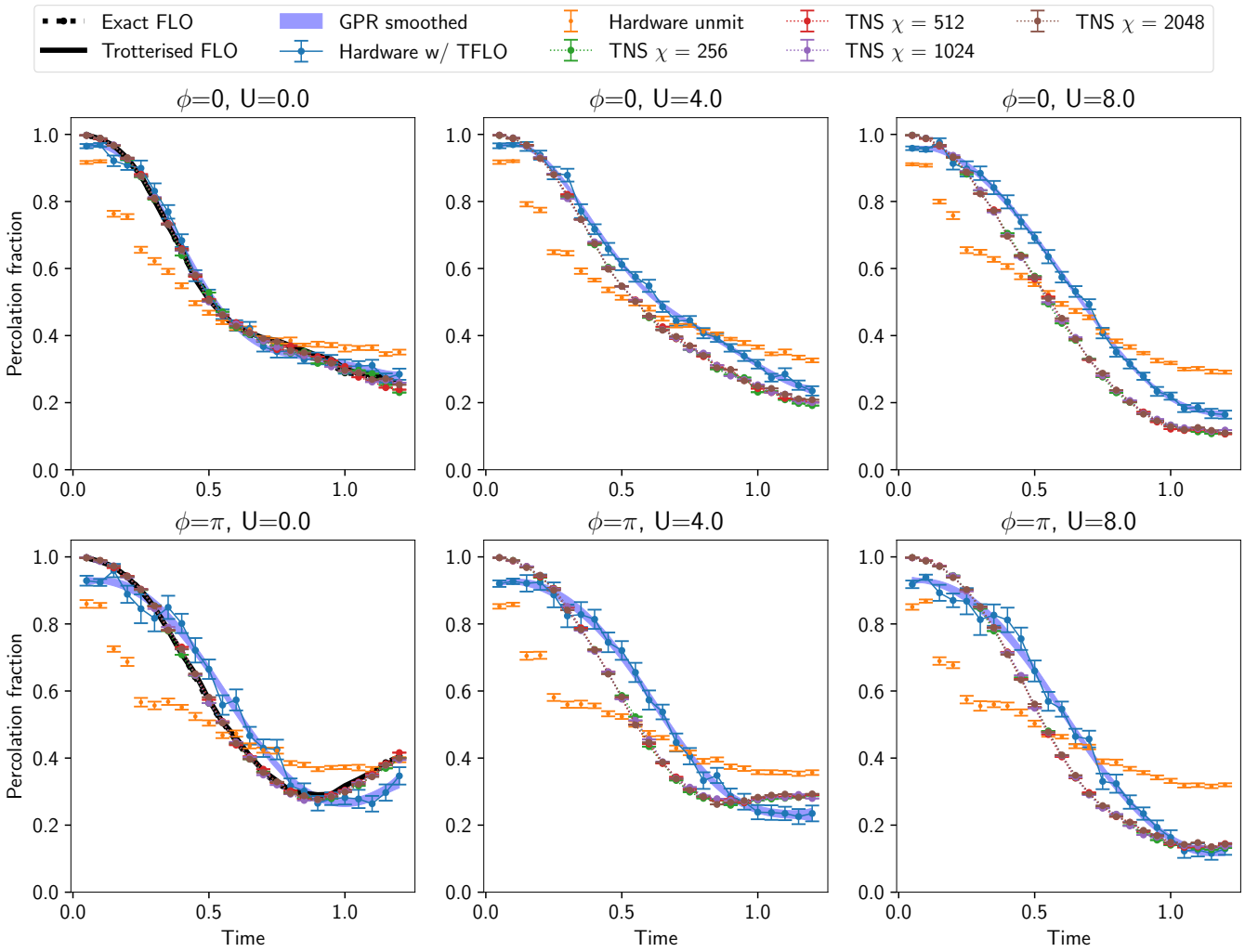


FIG. 59: 4×4, The unmitigated hardware data is postselected by antiferromagnetic order measured by AFM $\in [-1, 0]$ (see Eq. (B12), number of doublons $\mathcal{D} = [0, 1, 2]$ present and particle number conservation. TFLO is then applied to this set of shots to produce the Hardware w/ TFLO line, and finally GPR is applied to give the 68% one sigma region (for more details, see Appendix E 2, Appendix E 7 and Appendix E 3).

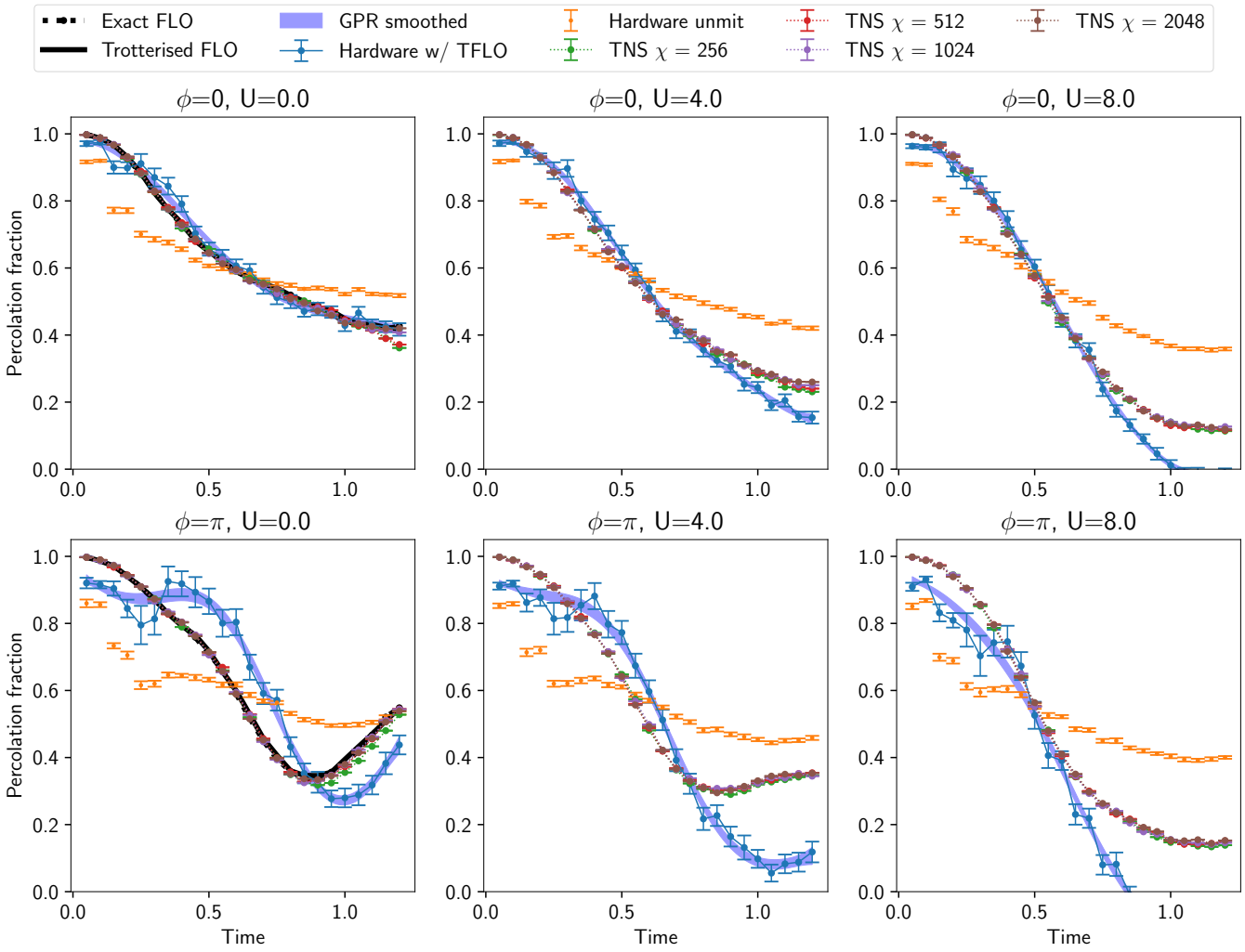


FIG. 60: 4×4, The unmitigated hardware data is postselected by only particle number conservation. TFLO is then applied to this set of shots to produce the Hardware w/ TFLO line, and finally GPR is applied to give the 68% one sigma region (for more details, see Appendix E 2, Appendix E 7 and Appendix E 3).

1. Effects of a simple bitflip noise model on percolation

In the previous sections discussing percolations, the hardware results are visually much flatter than those of TNS and FLO. Here, we demonstrate that a very simple bitflip error model can introduce dramatic flattening of the percolation curve, thereby showing that the non-local percolation observable is very fragile to basic SPAM noise. To implement this, for the TNS computational basis samples, we simply applied a bitflip with probability p and recomputed the percolation fraction. The results are shown in Fig. 62 and Fig. 63 with $p = 0.1$, $p = 0.5$ respectively. The results uncover a dramatic change in the TNS results, as the lines flatten to become comparable to those of the unmitigated hardware results.

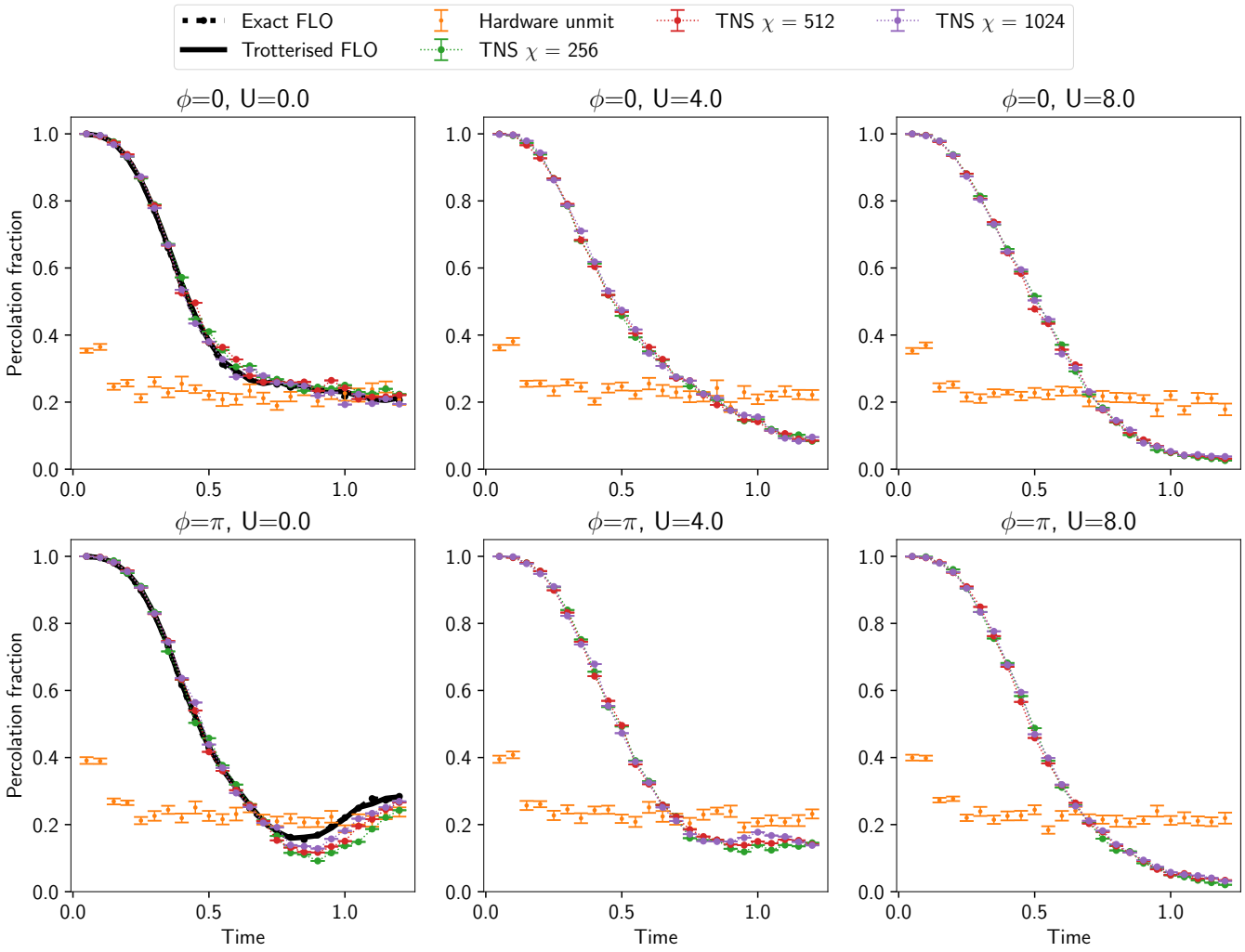


FIG. 61: 6×6 percolation fraction. The unmitigated hardware data is postselected by antiferromagnetic order measured by AFM $\in [-1, 0]$ (see Eq. (B12)), number of doublons $\mathcal{D} = [0, 1, 2, 3, 4, 5, 6]$ present and particle number conservation. At this system size we can clearly see a deviation between the $U = 0$ free fermion FLO simulation vs TNS, most prominent for $\phi = \pi$ flux experiment, where the TNS converges slower towards exact FLO with increasing χ compared to smaller system sizes.

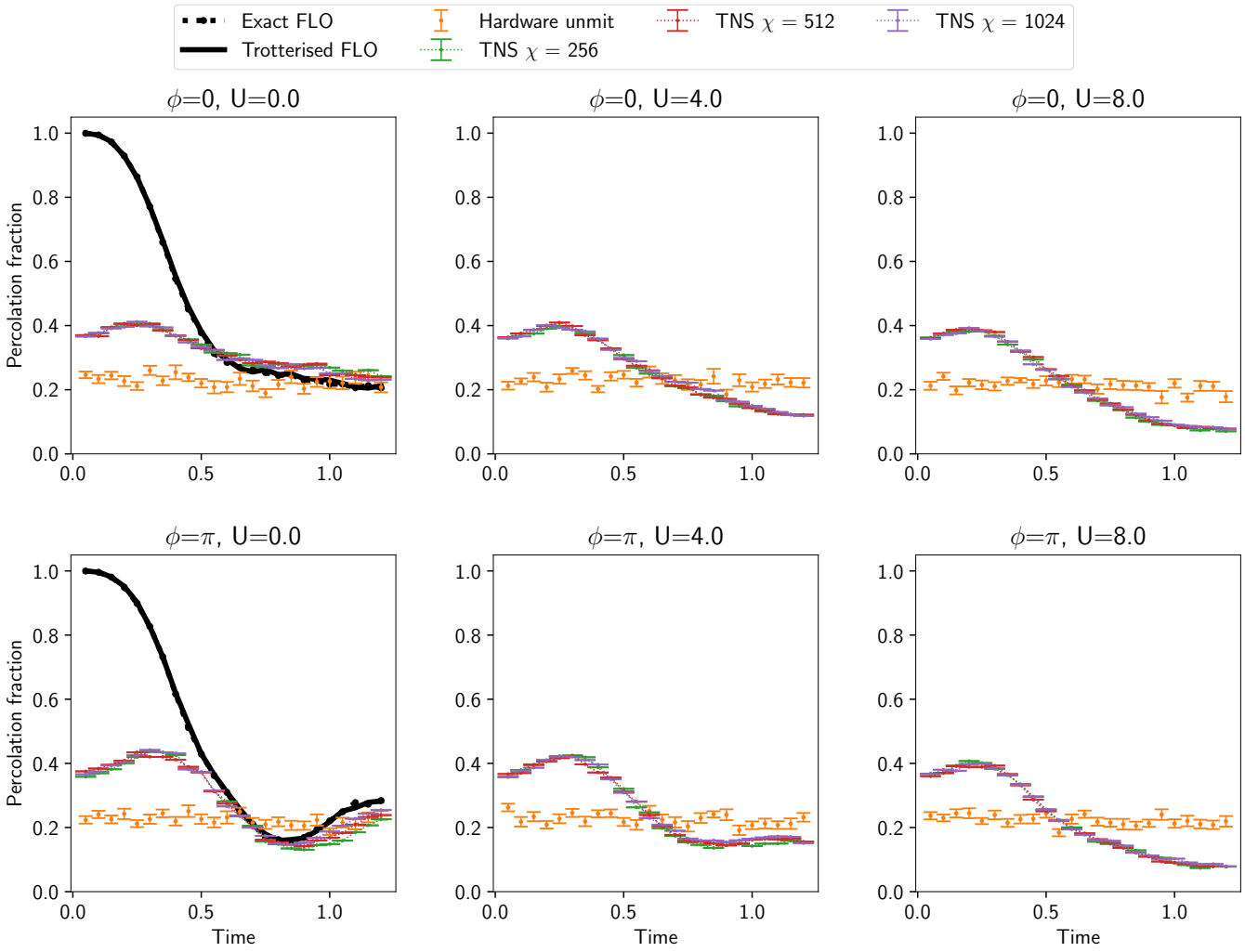


FIG. 62: 6×6 lattice percolation fraction. The TNS results are changed with the application of a bitflip error model with probability $p = 0.1$ to TNS samples. Shots are post-selected to AFM range [-1,0] and doublon numbers $\mathcal{D} = [0, 1, 2, 3, 4, 5, 6]$.

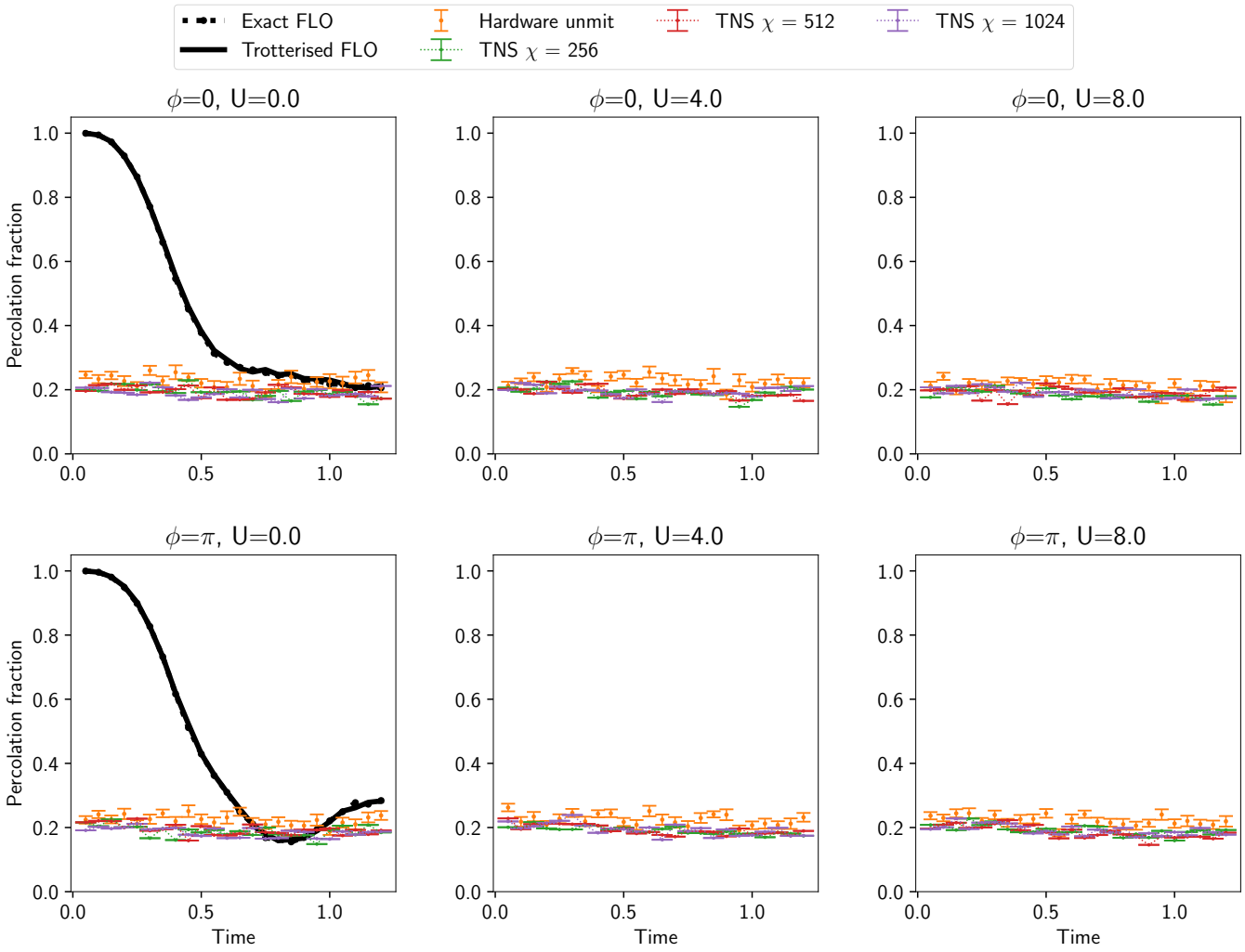


FIG. 63: 6×6 lattice percolation fraction. The TNS results are changed with the application of a bitflip error model with probability $p = 0.5$ to TNS samples. Shots are post-selected to AFM range [-1,0] and doublon numbers $\mathcal{D} = [0, 1, 2, 3, 4, 5, 6]$.

Appendix L: Illustrative Plots of Performance and Scaling

For completeness, here we collate further data retrieved from the quantum hardware. These plots are meant to illustrate the general behaviour of the simulations at various sizes. The plots herein all relate to initial states with a central hole and the dynamics with and without a magnetic flux. In all cases, the data refers to experimental data, which has then been mitigated with TFLO and GPR. The plots show a collection of global observables: average doublon density (Eq. (B4)); RMS holon (Eq. (B5)); staggered magnetisation (Eq. (B6)); and staggered spin-spin (Eq. (B7)). We also show the dynamics of spin densities, the correlation between nearest neighbours, and the doublon density deviation from the system mean.

1. 4×4

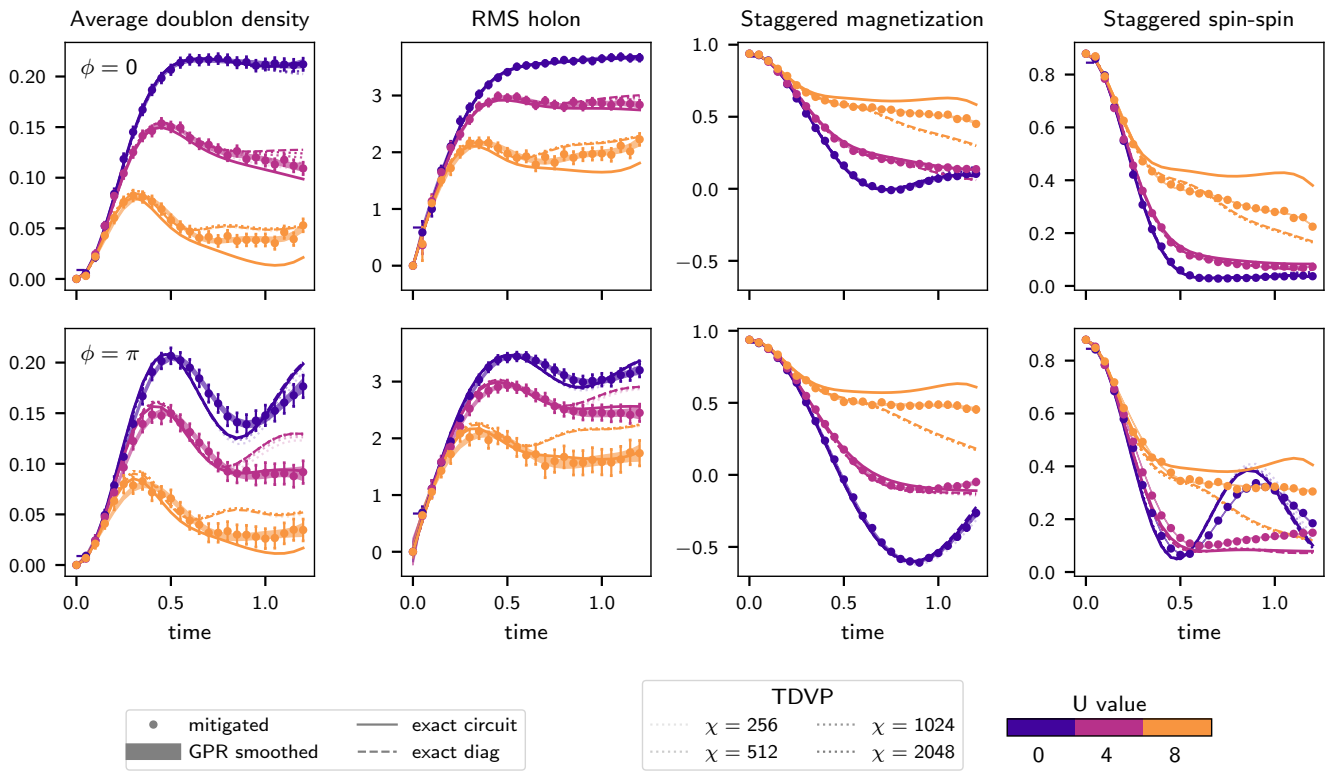


FIG. 64: Evolution of average doublon density, RMS holon, staggered magnetisation and staggered spin-spin correlation for the 4×4 lattice with a central hole for $\phi = 0$ (top row) and $\phi = \pi$ (bottom row), obtained from TFLO mitigated experimental results using 34 qubits and 1220 two-qubit gates. Here we also plot the results obtained using tensor network simulations at different bond dimensions, and exact circuit and TDS evolution when possible ($U = 0, 4, 8$).

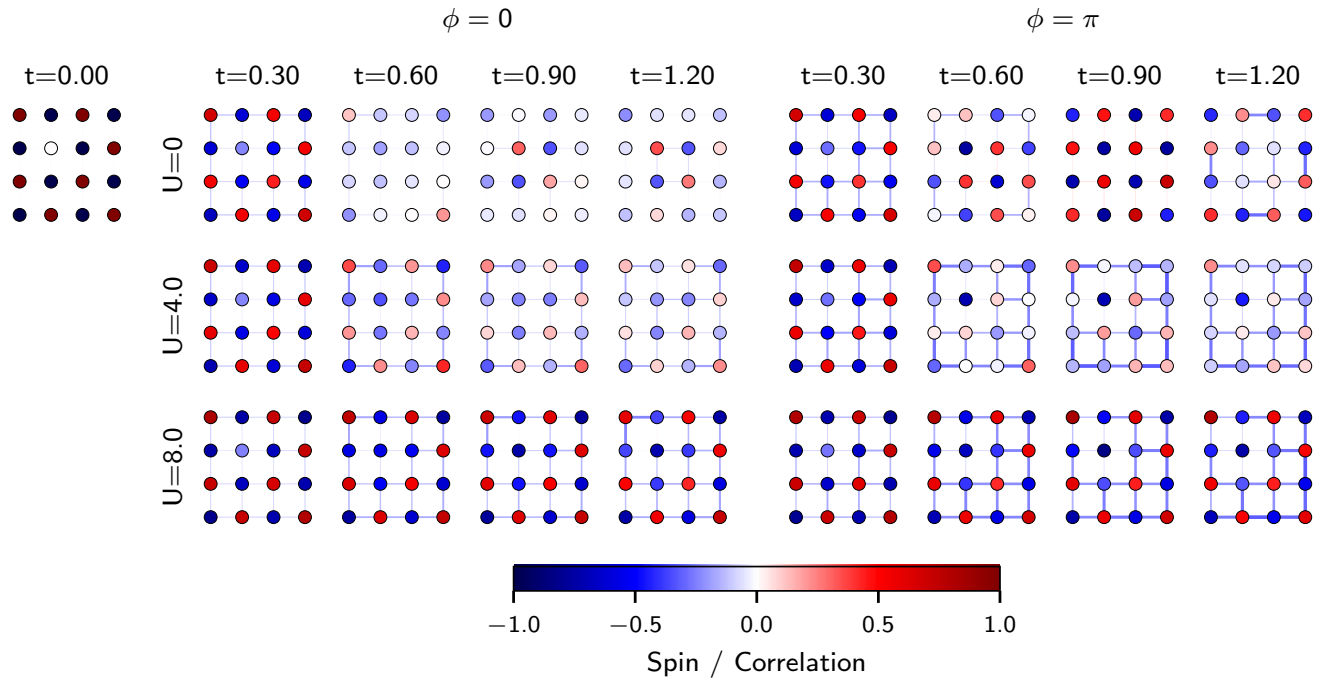


FIG. 65: Spin densities at each site (circles) and correlations between nearest neighbours (links) at selected times for the 4×4 lattice with a central hole for $\phi = 0$ (left) and $\phi = \pi$ (right). Observe the enhancement of spin correlations in the Dirac semimetal regime ($\phi = \pi$).

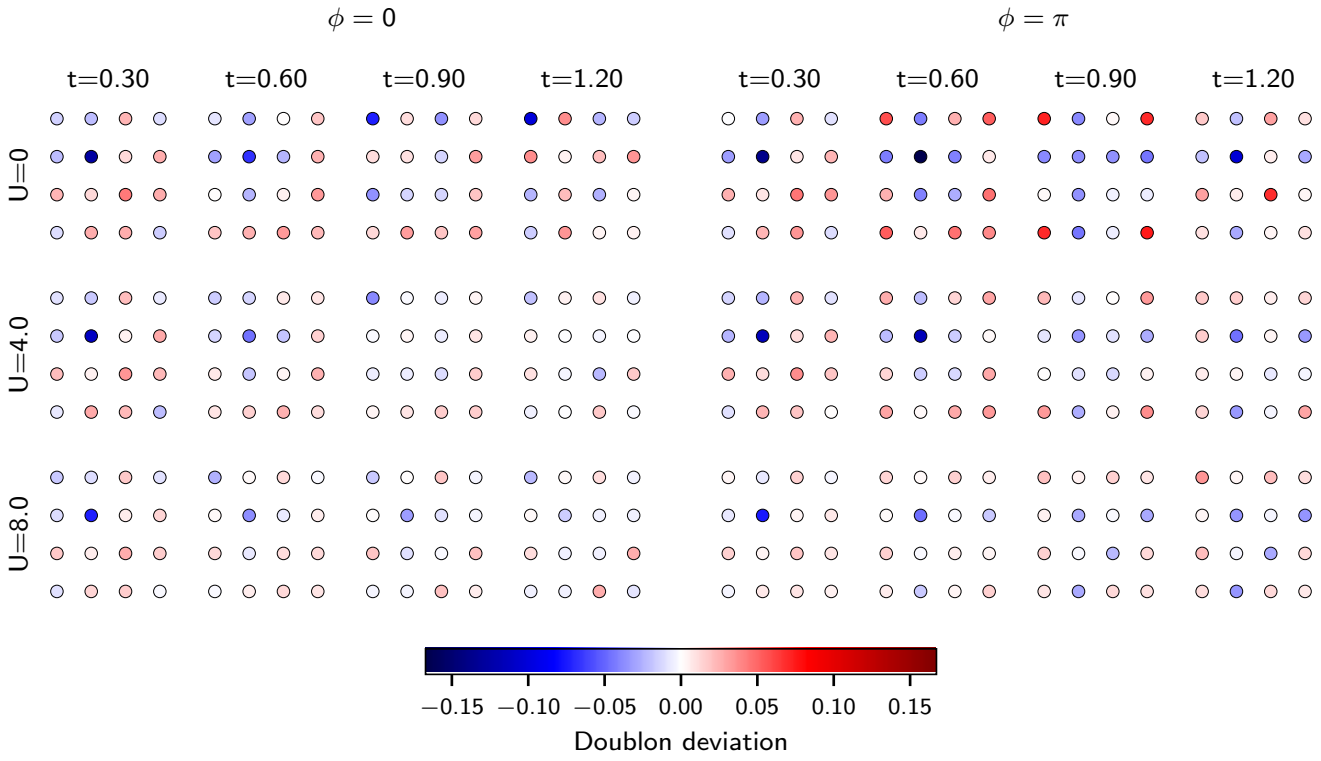


FIG. 66: Doublon density deviation from the mean at selected times for the 4×4 lattice with a central hole for $\phi = 0$ (left) and $\phi = \pi$ (right).

2. 5×5

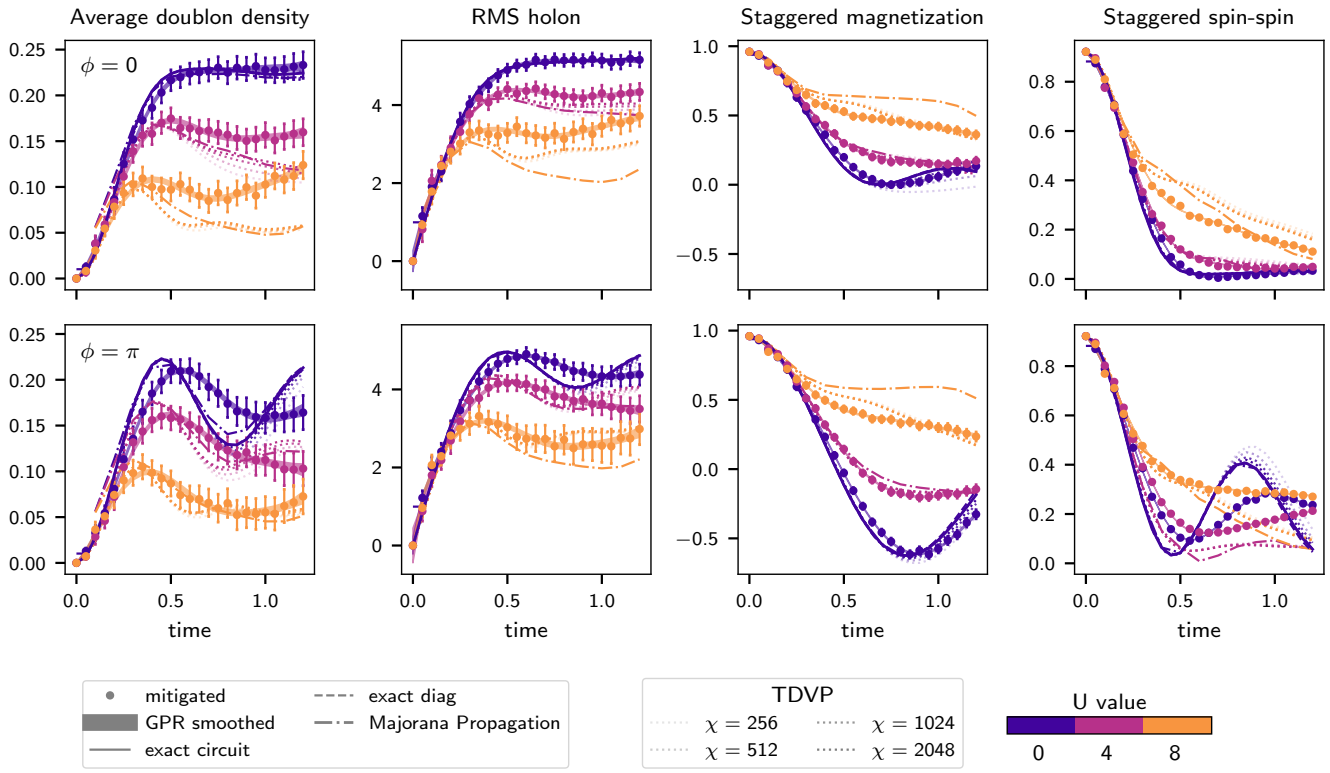


FIG. 67: Evolution of average doublon density, RMS holon, staggered magnetisation and staggered spin-spin correlation for the 5×5 lattice with a central hole for $\phi = 0$ (top row) and $\phi = \pi$ (bottom row), obtained from TFLO mitigated experimental results using 51 qubits and 2626 two-qubit gates. Here we also plot the results obtained using tensor network simulations at different bond dimensions, and exact circuit and TDS evolution when possible ($U = 0, 4, 8$).

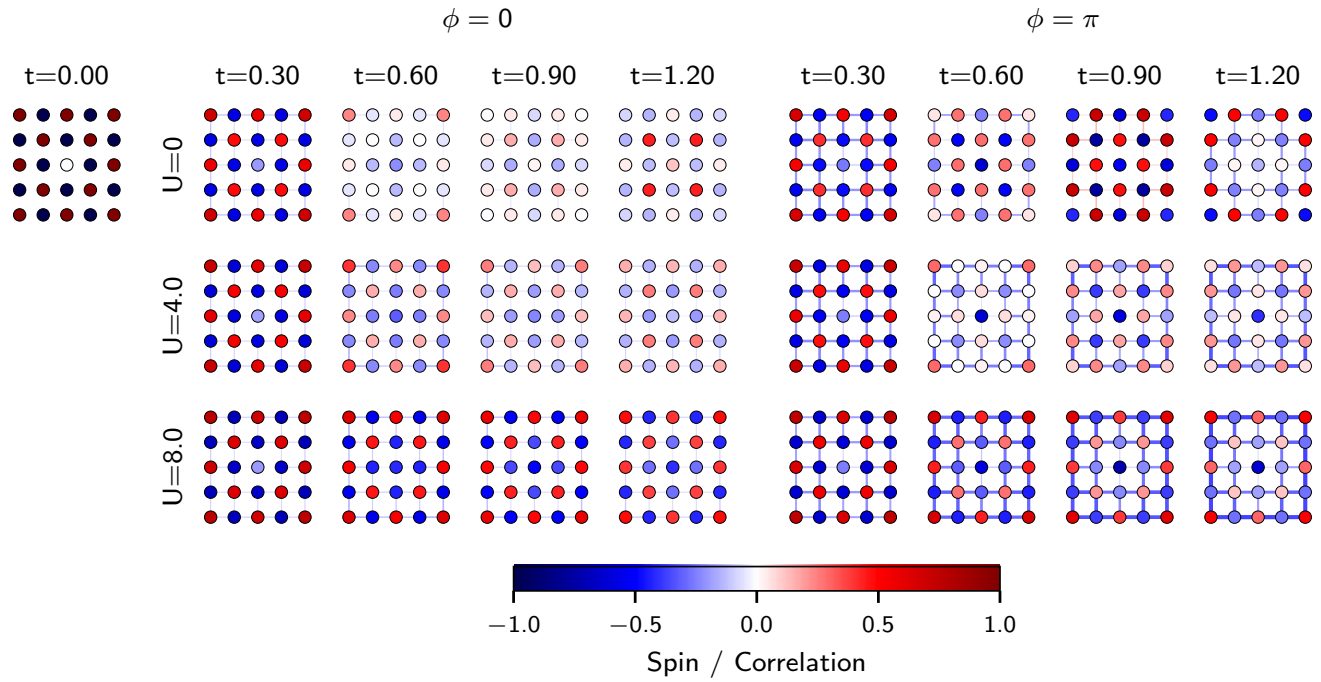


FIG. 68: Spin densities at each site (circles) and correlations between nearest neighbours (links) at selected times for the 5×5 lattice with a central hole for $\phi = 0$ (left) and $\phi = \pi$ (right). Observe the enhancement of spin correlations in the Dirac semimetal regime ($\phi = \pi$).

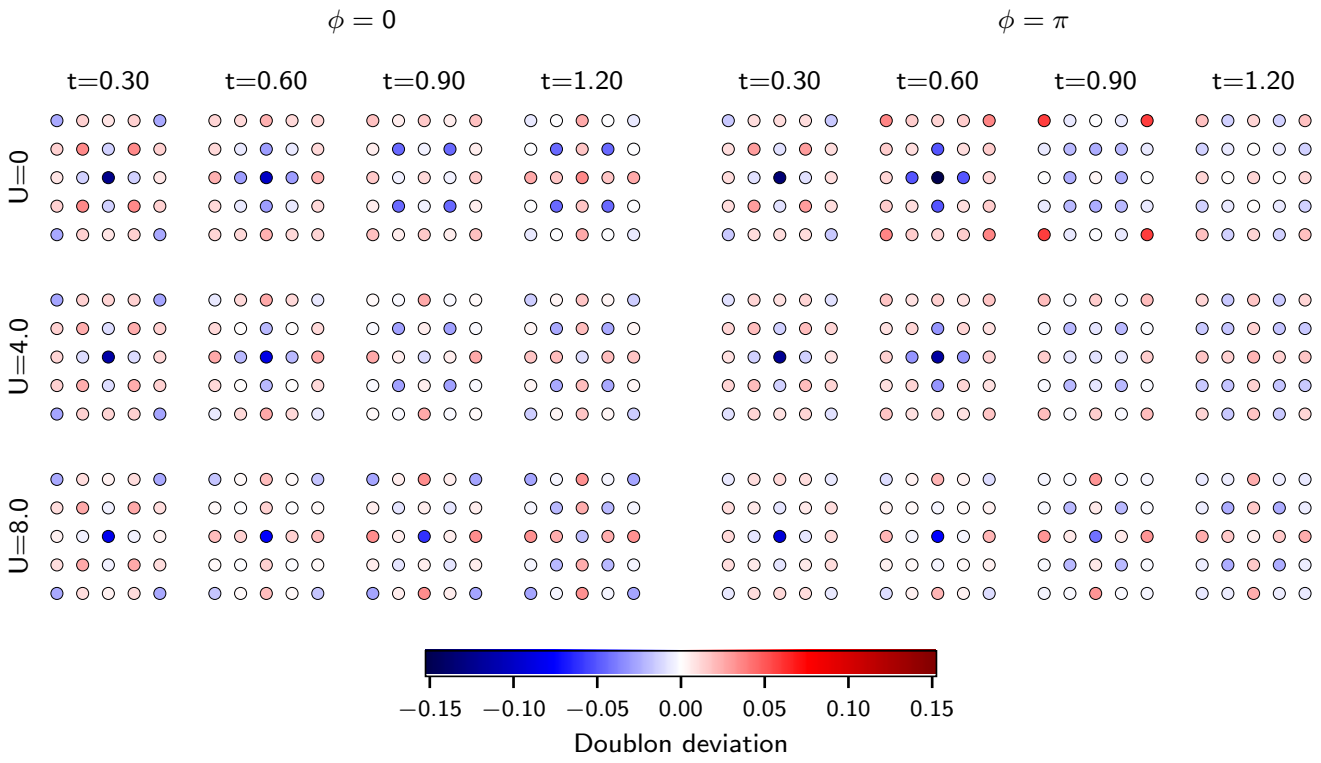


FIG. 69: Doublon density deviation from the mean at selected times for the 5×5 lattice with a central hole for $\phi = 0$ (left) and $\phi = \pi$ (right).

3. 6×5

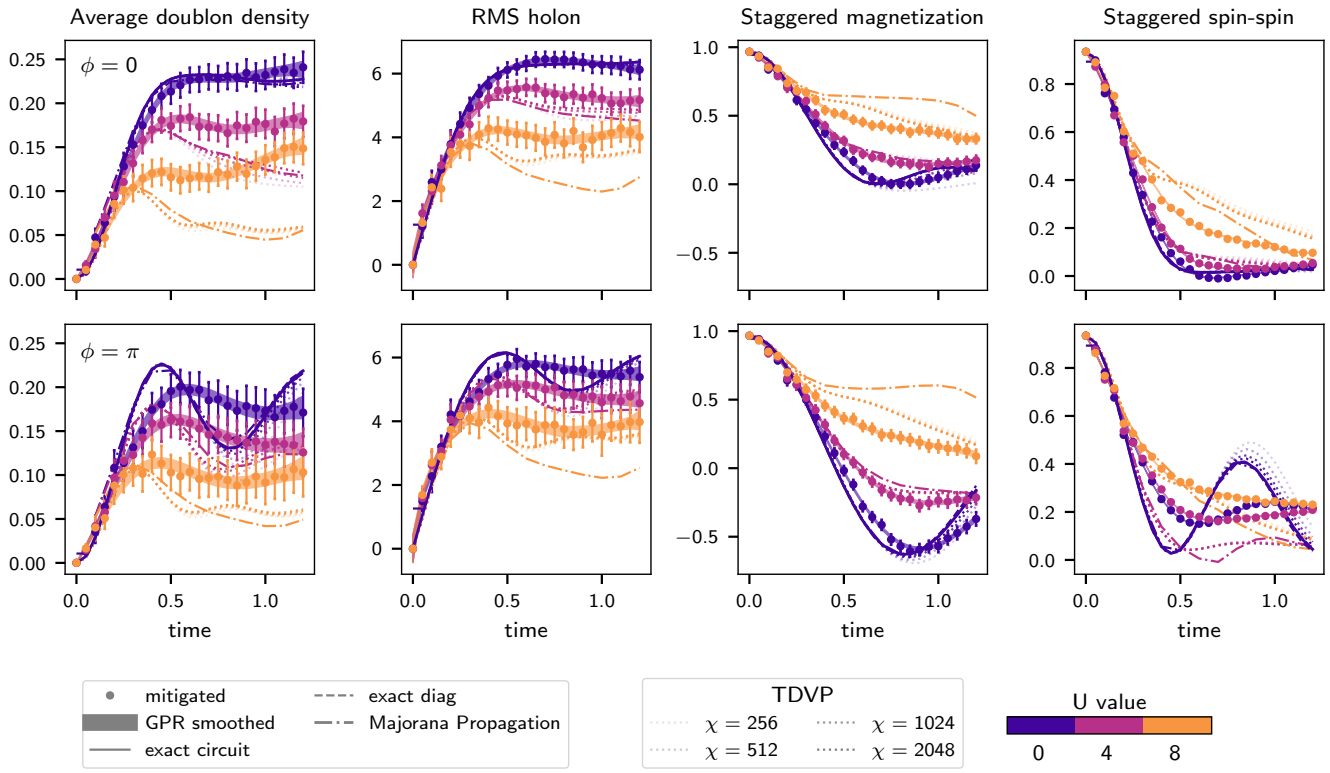


FIG. 70: Evolution of average doublon density, RMS holon, staggered magnetisation and staggered spin-spin correlation for the 6×5 lattice with a central hole for $\phi = 0$ (top row) and $\phi = \pi$ (bottom row), obtained from TFLO mitigated experimental results using 62 qubits and 3172 two-qubit gates. Here we also plot the results obtained using tensor network simulations at different bond dimensions, and exact circuit and TDS evolution when possible ($U = 0, 4, 8$).

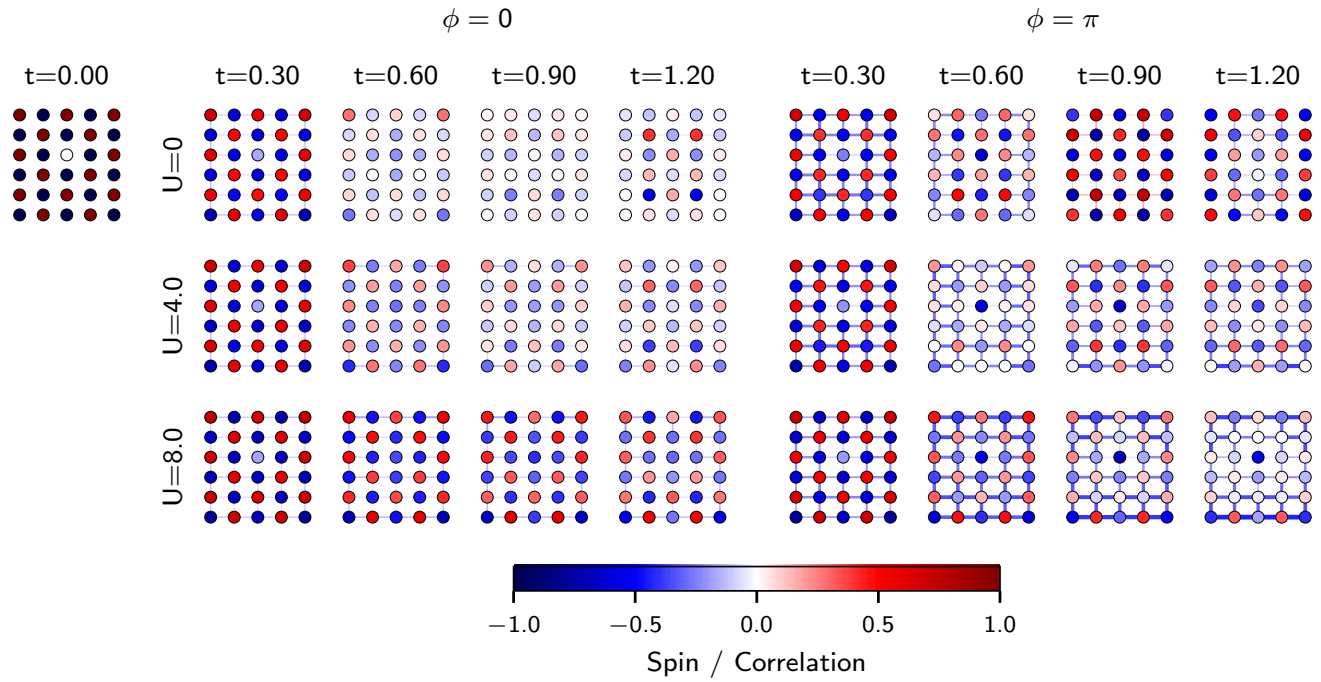


FIG. 71: Spin densities at each site (circles) and correlations between nearest neighbours (links) at selected times for the 6×5 lattice with a central hole for $\phi = 0$ (left) and $\phi = \pi$ (right). Observe the enhancement of spin correlations in the Dirac semimetal regime ($\phi = \pi$).

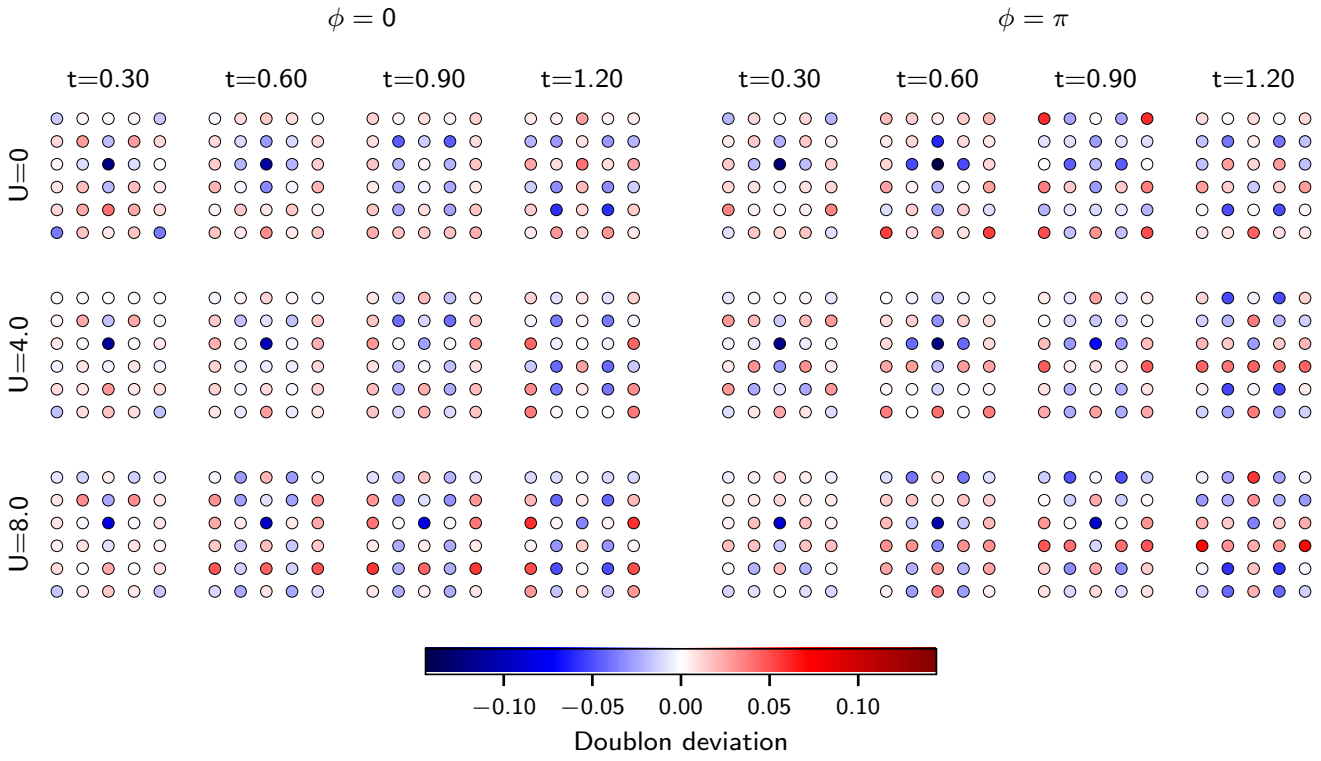


FIG. 72: Doublon density deviation from the mean at selected times for the 6×5 lattice with a central hole for $\phi = 0$ (left) and $\phi = \pi$ (right).

4. 6×6

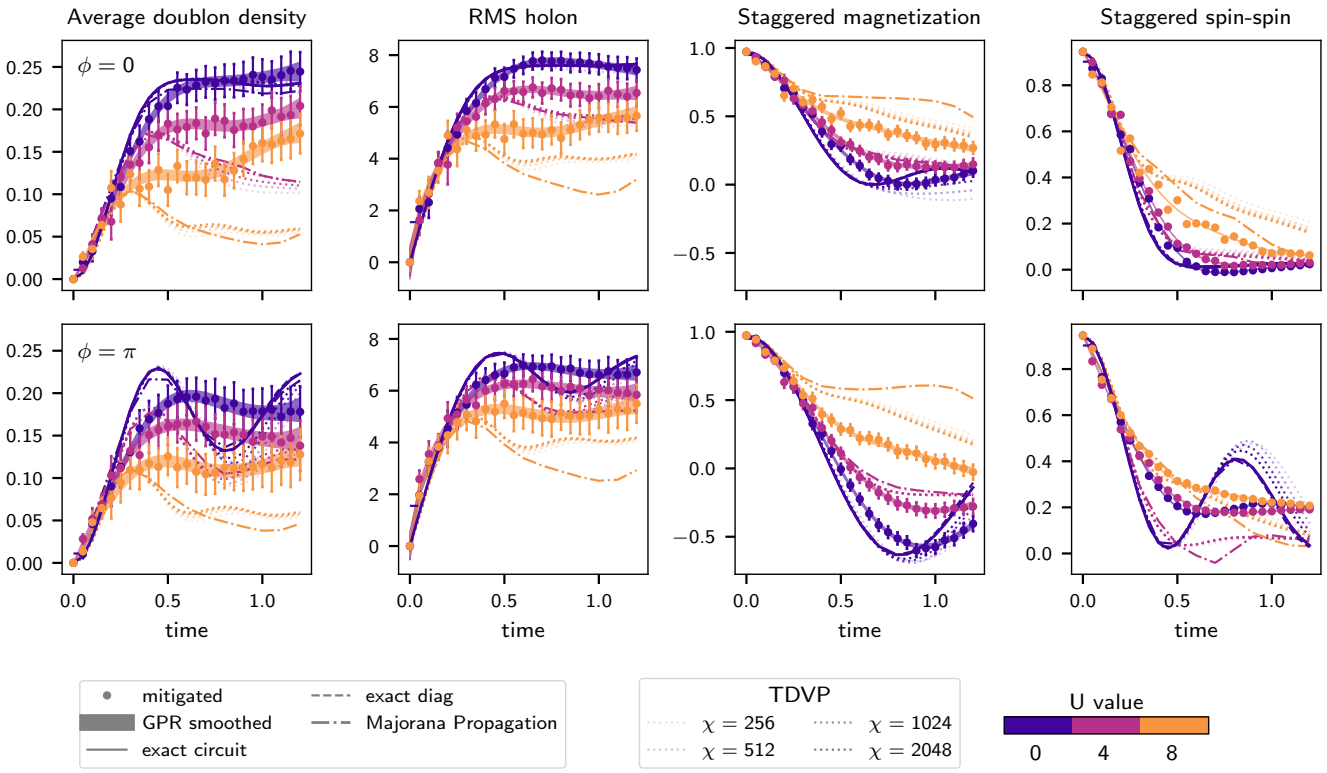


FIG. 73: Evolution of average doublon density, RMS holon, staggered magnetisation and staggered spin-spin correlation for the 6×6 lattice with a central hole for $\phi = 0$ (top row) and $\phi = \pi$ (bottom row), obtained from TFLO mitigated experimental results using 74 qubits and 4372 two-qubit gates. Here we also plot the results obtained using tensor network simulations at different bond dimensions, and exact circuit and TDS evolution when possible ($U = 0, 4, 8$).

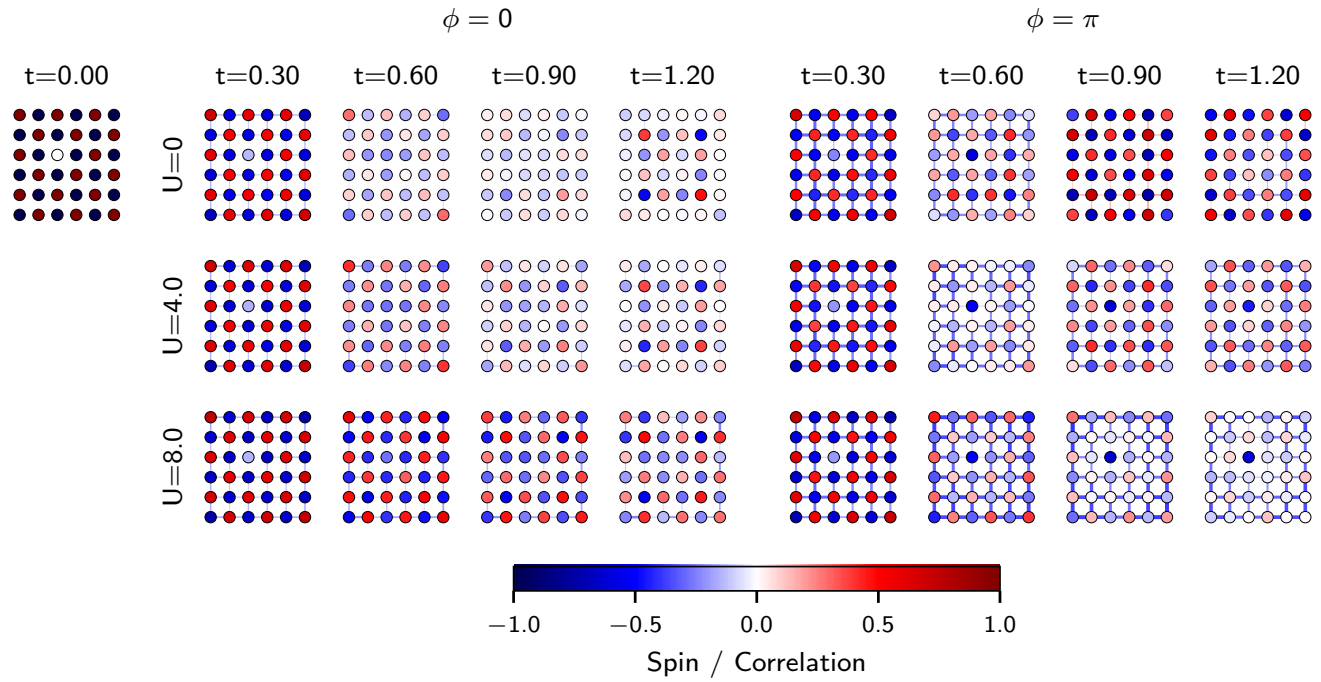


FIG. 74: Spin densities at each site (circles) and correlations between nearest neighbours (links) at selected times for the 6×6 lattice with a central hole for $\phi = 0$ (left) and $\phi = \pi$ (right). Observe the enhancement of spin correlations in the Dirac semimetal regime ($\phi = \pi$).

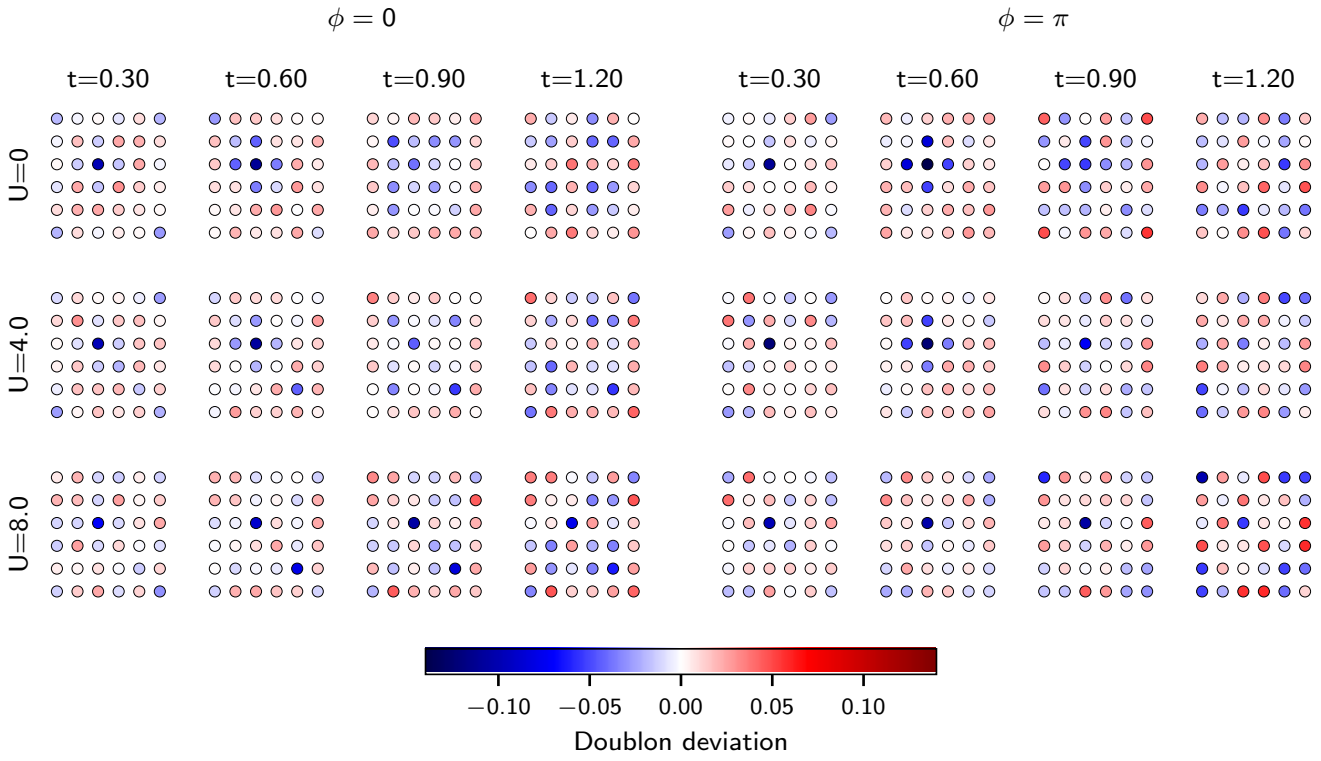


FIG. 75: Doublon density deviation from the mean at selected times for the 6×6 lattice with a central hole for $\phi = 0$ (left) and $\phi = \pi$ (right).

[1] Feynman, R. P. Simulating physics with computers. *International Journal of Theoretical Physics* **21**, 467–488 (1982).

[2] Alexandrov, A. S. & Devreese, J. T. *Advances in Polaron Physics*. Springer Series in Solid-State Sciences (Springer, Berlin, Germany, 2009), 2010 edn.

[3] Zaanen, O. Y., Osman, H. V. & Kruis, Z. J. The geometric order of stripes and Luttinger liquids. *Philosophical Magazine B* **81**, 1485–1531 (2001). URL http://dx.doi.org/10.1080/13642810110051791.

[4] Anderson, P. Resonating valence bonds: A new kind of insulator? *Materials Research Bulletin* **8**, 153–160 (1973). URL https://www.sciencedirect.com/science/article/pii/0025540873901670.

[5] Rigol, M., Dunjko, V. & Olshanii, M. Thermalization and its mechanism for generic isolated quantum systems. *Nature* **452**, 854–858 (2008). URL https://www.nature.com/articles/nature06838.

[6] Nasajpour-Esfahani, N. *et al.* Comprehensive review of lithium-ion battery materials and development challenges. *Renewable and Sustainable Energy Reviews* **203**, 114783 (2024). URL https://www.sciencedirect.com/science/article/pii/S1364032124005094.

[7] Adebanjo, I. T. *et al.* A comprehensive review of lithium-ion battery components degradation and operational considerations: a safety perspective. *Energy Adv.* **4**, 820–877 (2025). URL http://dx.doi.org/10.1039/D5YA00065C.

[8] Booth, S. G. *et al.* Perspectives for next generation lithium-ion battery cathode materials. *APL Materials* **9**, 109201 (2021). URL https://doi.org/10.1063/5.0051092. _eprint: https://pubs.aip.org/aip/apm/article-pdf/doi/10.1063/5.0051092/19674936/109201_1_5.0051092.pdf.

[9] Urban, A., Seo, D.-H. & Ceder, G. Computational understanding of Li-ion batteries. *npj Computational Materials* **2**, 16002 (2016). URL https://www.nature.com/articles/npjcompumats20162. Publisher: Nature Publishing Group.

[10] Park, N.-G. Perovskite solar cells: an emerging photovoltaic technology. *Materials today* **18**, 65–72 (2015). URL https://www.sciencedirect.com/science/article/pii/S1369702114002570.

[11] Lloyd, S. Universal quantum simulators. *Science* **273**, 1073–1078 (1996). URL https://www.science.org/doi/10.1126/science.273.5278.1073.

[12] Gilyén, A., Su, Y., Low, G. H. & Wiebe, N. Quantum singular value transformation and beyond: exponential improvements for quantum matrix arithmetics. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC '19, 193–204 (ACM, 2019). URL http://dx.doi.org/10.1145/3313276.3316366.

[13] Childs, A. M., Su, Y., Tran, M. C., Wiebe, N. & Zhu, S. Theory of Trotter error with commutator scaling. *Physical Review X* **11** (2021). URL http://dx.doi.org/10.1103/PhysRevX.11.011020.

[14] Jordan, S. P., Krovi, H., Lee, K. S. & Preskill, J. BQP-completeness of scattering in scalar quantum field theory. *Quantum* **2**, 44 (2018). URL https://quantum-journal.org/papers/q-2018-01-08-44/.

[15] Cao, Y. *et al.* Quantum chemistry in the age of quantum computing. *Chemical reviews* **119**, 10856–10915 (2019). URL https://pubs.acs.org/doi/10.1021/acs.chemrev.8b00803.

[16] McArdle, S., Endo, S., Aspuru-Guzik, A., Benjamin, S. C. & Yuan, X. Quantum computational chemistry. *Reviews of Modern Physics* **92**, 015003 (2020). URL https://journals.aps.org/rmp/abstract/10.1103/RevModPhys.92.015003.

[17] Clinton, L. *et al.* Towards near-term quantum simulation of materials. *Nature Communications* **15** (2024). URL http://dx.doi.org/10.1038/s41467-023-43479-6.

[18] Auerbach, A. *Interacting electrons and quantum magnetism* (Springer Science & Business Media, 2012).

[19] AI, G. Q. & Collaborators. Quantum error correction below the surface code threshold. *Nature* **638**, 920–926 (2025). URL https://www.nature.com/articles/s41586-024-08449-y.

[20] Montanaro, A. & Stanisic, S. Error mitigation by training with fermionic linear optics (2021). URL https://arxiv.org/abs/2102.02120. arXiv:2102.02120.

[21] Jiang, Y.-F., Devereaux, T. P. & Jiang, H.-C. Ground-state phase diagram and superconductivity of the doped Hubbard model on six-leg square cylinders. *Physical Review B* **109**, 085121 (2024). URL https://journals.aps.org/prb/abstract/10.1103/PhysRevB.109.085121.

[22] Sun, F. & Xu, X. Y. Boosting determinant quantum Monte Carlo with submatrix updates: Unveiling the phase diagram of the 3D Hubbard model. *SciPost Physics* **18**, 055 (2025). URL https://scipost.org/SciPostPhys.18.2.055.

[23] Häner, T. & Steiger, D. S. 0.5 petabyte simulation of a 45-qubit quantum circuit. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, SC '17, 1–10 (ACM, 2017). URL http://dx.doi.org/10.1145/3126908.3126947.

[24] Baruffa, F., Lougovski, P. & Jones, T. Simulating 44-qubit quantum circuits using aws parallelcluster. URL https://aws.amazon.com/blogs/hpc/simulating-44-qubit-quantum-circuits-using-aws-parallelcluster/.

[25] Thompson, A. P., Soeteman, A., Cade, C. & Niesen, I. Non-zero noise extrapolation: accurately simulating noisy quantum circuits with tensor networks (2025). URL https://arxiv.org/abs/2501.13237. arXiv:2501.13237.

[26] Miller, A. *et al.* Simulation of Fermionic circuits using Majorana propagation (2025). URL https://arxiv.org/abs/2503.18939. arXiv:2503.18939.

[27] Bespalova, T. A., Delić, K., Pupillo, G., Tacchino, F. & Tavernelli, I. Simulating the Fermi-Hubbard model with long-range hopping on a quantum computer. *Physical Review A* **111**, 052619 (2025). URL https://journals.aps.org/pra/abstract/10.1103/PhysRevA.111.052619.

[28] Arute, F. *et al.* Observation of separated dynamics of charge and spin in the Fermi-Hubbard model (2020). URL https://arxiv.org/abs/2010.07965. arXiv:2010.07965.

[29] Vilchez-Estevez, L., Santos, R. A., Wang, S. Y. & Gambetta, F. M. Extracting the spin excitation spectrum of a fermionic system using a quantum processor. *Physical Review B* **112** (2025). URL http://dx.doi.org/10.1103/ydfw-k83n.

[30] Chowdhury, T. A., Korepin, V., Pascuzzi, V. R. & Yu, K. Quantum utility in simulating the real-time dynamics of the Fermi-Hubbard model using superconducting quantum computers (2025). URL https://arxiv.org/abs/2509.14196. arXiv:2509.14196.

[31] Evered, S. J. *et al.* Probing the Kitaev honeycomb model on a neutral-atom quantum computer. *Nature* **645**, 341–347 (2025). URL https://www.nature.com/articles/s41586-025-09475-0.

[32] Browaeys, A. & Lahaye, T. Many-body physics with individually controlled Rydberg atoms. *Nature Physics* **16**, 132–142 (2020). URL https://www.nature.com/articles/s41567-019-0733-z.

[33] Monroe, C. *et al.* Programmable quantum simulations of spin systems with trapped ions. *Reviews of Modern Physics* **93**, 025001 (2021). URL https://journals.aps.org/rmp/abstract/10.1103/RevModPhys.93.025001.

[34] Shao, H.-J. *et al.* Antiferromagnetic phase transition in a 3D fermionic Hubbard model. *Nature* **632**, 267–272 (2024). URL https://www.nature.com/articles/s41586-024-07689-2.

[35] Hartke, T., Oreg, B., Turnbaugh, C., Jia, N. & Zwierlein, M. Direct observation of nonlocal fermion pairing in an attractive Fermi-Hubbard gas. *Science* **381**, 82–86 (2023). URL https://www.science.org/doi/10.1126/science.ade4245.

[36] Xu, M. *et al.* A neutral-atom Hubbard quantum simulator in the cryogenic regime. *Nature* **642**, 909–915 (2025). URL https://doi.org/10.1038/s41586-025-09112-w.

[37] Guardado-Sanchez, E. *et al.* Subdiffusion and heat transport in a tilted two-dimensional Fermi-Hubbard system. *Physical Review X* **10**, 011042 (2020). URL https://link.aps.org/doi/10.1103/PhysRevX.10.011042.

[38] Brown, P. T. *et al.* Bad metallic transport in a cold atom Fermi-Hubbard system. *Science* **363**, 379–382 (2019). URL https://www.science.org/doi/abs/10.1126/science.aat4134.

[39] Nichols, M. A. *et al.* Spin transport in a Mott insulator of ultracold fermions. *Science* **363**, 383–387 (2019). URL https://www.science.org/doi/abs/10.1126/science.aat4387.

[40] Bakr, W. S., Ba, Z. & Prichard, M. L. Microscopy of ultracold fermions in optical lattices (2025). URL https://arxiv.org/abs/2507.04042. arXiv:2507.04042.

[41] Bohrdt, A., Homeier, L., Reinmoser, C., Demler, E. & Grusdt, F. Exploration of doped quantum magnets with ultracold atoms. *Annals of Physics* **435**, 168651 (2021). URL http://dx.doi.org/10.1016/j.aop.2021.168651.

[42] Andersen, T. I. *et al.* Thermalization and criticality on an analog-digital quantum simulator (2024). URL https://arxiv.org/abs/2405.17385. arXiv:2405.17385.

[43] Fava, S. *et al.* Magnetic field expulsion in optically driven $YBa_2Cu_3O_{6.48}$. *Nature* **632**, 75–80 (2024). URL https://www.nature.com/articles/s41586-024-07635-2.

[44] LeBlanc, J. P. F. & Gull, E. Equation of state of the fermionic two-dimensional hubbard model. *Phys. Rev. B* **88**, 155108 (2013). URL https://link.aps.org/doi/10.1103/PhysRevB.88.155108.

[45] Young, S. M. & Kane, C. L. Dirac semimetals in two dimensions. *Physical Review Letters* **115**, 126803 (2015). URL https://link.aps.org/doi/10.1103/PhysRevLett.115.126803.

[46] Koepsell, J. *et al.* Imaging magnetic polarons in the doped Fermi-Hubbard model. *Nature* **572**, 358–362 (2019). URL https://www.nature.com/articles/s41586-019-1463-1.

[47] Koepsell, J. *et al.* Microscopic evolution of doped Mott insulators from polaronic metal to Fermi liquid. *Science* **374**, 82–86 (2021). URL http://dx.doi.org/10.1126/science.abe7165.

[48] Ji, G. *et al.* Coupling a mobile hole to an antiferromagnetic spin background: Transient dynamics of a magnetic polaron. *Physical Review X* **11**, 021022 (2021). URL https://journals.aps.org/prx/abstract/10.1103/PhysRevX.11.021022.

[49] Mihailovic, D. Insights into the polaron physics of high temperature superconductors from optical and time-domain spectroscopy. *Physica C: Superconductivity and its Applications* **615**, 1354385 (2023). URL https://www.sciencedirect.com/science/article/pii/S0921453423001764.

[50] Bennemann, K. & Ketterson, J. *Novel Superfluids: Volume 2*. International Series of Monographs on Physics (OUP Oxford, 2014). URL https://books.google.co.uk/books?id=v0xVBQAAQBAJ.

[51] Zheng, B.-X. *et al.* Stripe order in the underdoped region of the two-dimensional Hubbard model. *Science* **358**, 1155–1160 (2017). URL https://www.science.org/doi/10.1126/science.aam7127.

[52] Liu, W.-Y., Zhai, H., Peng, R., Gu, Z.-C. & Chan, G. K.-L. Accurate simulation of the Hubbard model with finite fermionic projected entangled pair states. *Physical Review Letters* **134** (2025). URL http://dx.doi.org/10.1103/r4q9-4yvj.

[53] Senthil, T., Vishwanath, A., Balents, L., Sachdev, S. & Fisher, M. P. A. Deconfined quantum critical points. *Science* **303**, 1490–1494 (2004). URL https://www.science.org/doi/abs/10.1126/science.1091806.

[54] Senthil, T., Balents, L., Sachdev, S., Vishwanath, A. & Fisher, M. P. A. Quantum criticality beyond the Landau-Ginzburg-Wilson paradigm. *Physical Review B* **70**, 144407 (2004). URL https://link.aps.org/doi/10.1103/PhysRevB.70.144407.

[55] Wang, F., Kivelson, S. A. & Lee, D.-H. Nematicity and quantum paramagnetism in FeSe. *Nature Physics* **11**, 959–963 (2015). URL https://www.nature.com/articles/nphys3456.

[56] Wilson, K. G. Confinement of quarks. *Physical Review D* **10**, 2445–2459 (1974). URL https://link.aps.org/doi/10.1103/PhysRevD.10.2445.

[57] Terhal, B. M. & DiVincenzo, D. P. Classical simulation of noninteracting-fermion quantum circuits. *Physical Review A* **65**, 032325 (2002). URL https://link.aps.org/doi/10.1103/PhysRevA.65.032325.

[58] Berezutskii, A. *et al.* Tensor networks for quantum computing. *Nature Reviews Physics* 1–13 (2025). URL https://www.nature.com/articles/s42254-025-00853-1.

[59] Patra, S., Jahromi, S. S., Singh, S. & Orús, R. Efficient tensor network simulation of IBM's largest quantum processors. *Physical Review Research* **6**, 013326 (2024). URL https://journals.aps.org/prresearch/abstract/10.1103/PhysRevResearch.6.013326.

[60] Hémery, K., Pollmann, F. & Luitz, D. J. Matrix product states approaches to operator spreading in ergodic quantum systems. *Physical Review B* **100**, 104303 (2019). URL https://journals.aps.org/prb/abstract/10.1103/PhysRevB.100.104303.

[61] Leviatan, E., Pollmann, F., Bardarson, J. H., Huse, D. A. & Altman, E. Quantum thermalization dynamics with matrix-product states (2017). URL https://arxiv.org/abs/1702.08894. arXiv:1702.08894.

[62] Phasecraft. Manuscript in preparation. *forthcoming* (2025).

[63] Hauschild, J. *et al.* Tensor network Python (TeNPy) version 1. *SciPost Phys. Codebases* 41 (2024). URL https://scipost.org/SciPostPhysCodeb.41.

[64] Rall, P., Liang, D., Cook, J. & Kretschmer, W. Simulation of qubit quantum circuits via Pauli propagation. *Physical Review A* **99**, 062337 (2019). URL https://journals.aps.org/pra/abstract/10.1103/PhysRevA.99.062337.

[65] Jordan, P. & Wigner, E. Über das paulische äquivalenzverbot. *Zeitschrift für Physik* **47**, 631–651 (1928).

[66] Cade, C., Mineh, L., Montanaro, A. & Stanisic, S. Strategies for solving the Fermi-Hubbard model on near-term quantum computers. *Physical Review B* **102**, 235122 (2020). URL https://journals.aps.org/prb/abstract/10.1103/PhysRevB.102.235122.

[67] Stanisic, S. *et al.* Observing ground-state properties of the Fermi-Hubbard model using a scalable algorithm on a quantum computer. *Nature communications* **13**, 5743 (2022). URL https://www.nature.com/articles/s41467-022-33335-4.

[68] Vidal, G. Efficient simulation of one-dimensional quantum many-body systems. *Physical Review Letters* **93**, 040502 (2004). URL https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.93.040502.

[69] Haegeman, J. *et al.* Time-dependent variational principle for quantum lattices. *Physical Review Letters* **107**, 070601 (2011). URL https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.107.070601.

[70] Provazza, J. *et al.* Fast Emulation of Fermionic Circuits with Matrix Product States. *Journal of Chemical Theory and Computation* **20**, 3719–3728 (2024). URL https://pubs.acs.org/doi/10.1021/acs.jctc.4c00200.

[71] Ayral, T. *et al.* Density-matrix renormalization group algorithm for simulating quantum circuits with a finite fidelity. *PRX Quantum* **4** (2023). URL http://dx.doi.org/10.1103/PRXQuantum.4.020304.

[72] Tindall, J., Fishman, M., Stoudenmire, E. M. & Sels, D. Efficient tensor network simulation of IBM's eagle kicked ising experiment. *PRX Quantum* **5**, 010308 (2024). URL https://link.aps.org/doi/10.1103/PRXQuantum.5.010308.

[73] Begušić, T., Gray, J. & Chan, G. K.-L. Fast and converged classical simulations of evidence for the utility of quantum computing before fault tolerance. *Science Advances* **10** (2024). URL http://dx.doi.org/10.1126/sciadv.adk4321.

[74] Arute, F. *et al.* Quantum supremacy using a programmable superconducting processor. *Nature* **574**, 505–510 (2019). URL https://www.nature.com/articles/s41586-019-1666-5.

[75] Morvan, A. *et al.* Phase transitions in random circuit sampling. *Nature* **634**, 328–333 (2024). URL https://www.nature.com/articles/s41586-024-07998-6.

[76] Zlokapa, A., Villalonga, B., Boixo, S. & Lidar, D. A. Boundaries of quantum supremacy via random circuit sampling. *npj Quantum Information* **9**, 36 (2023). URL https://www.nature.com/articles/s41534-023-00703-x.

[77] Cheng, B., Meng, F., Zhang, Z.-J. & Yung, M.-H. Generalized cross-entropy benchmarking for random circuits with ergodicity. *Chip* 100127 (2025). URL https://www.sciencedirect.com/science/article/pii/S2709472325000012.

[78] DeCross, M. *et al.* Computational power of random quantum circuits in arbitrary geometries. *Physical Review X* **15**, 021052 (2025). URL https://journals.aps.org/prx/abstract/10.1103/PhysRevX.15.021052.

[79] Lieb, E. H. Flux phase of the half-filled band. *Physical Review Letters* **73**, 2158–2161 (1994). URL https://link.aps.org/doi/10.1103/PhysRevLett.73.2158.

[80] Luttinger, J. M. Fermi surface and some simple equilibrium properties of a system of interacting fermions. *Physical Review* **119**, 1153–1163 (1960). URL https://link.aps.org/doi/10.1103/PhysRev.119.1153.

[81] Ogata, M. & Shiba, H. Bethe-ansatz wave function, momentum distribution, and spin correlation in the one-dimensional strongly correlated Hubbard model. *Physical Review B* **41**, 2326 (1990). URL https://journals.aps.org/prb/abstract/10.1103/PhysRevB.41.2326.

[82] Vijayan, J. *et al.* Time-resolved observation of spin-charge deconfinement in fermionic Hubbard chains. *Science* **367**, 186–189 (2020). URL https://doi.org/10.1126/science.aay2354.

[83] Anderson, P. W. & Khveshchenko, D. Charge-spin separation in 2D Fermi systems: Singular interactions as modified commutators, and solution of the 2d Hubbard model in the bosonized approximation. *Physical Review B* **52**, 16415–16419 (1995). URL https://link.aps.org/doi/10.1103/PhysRevB.52.16415.

[84] Lee, P. A., Nagaosa, N. & Wen, X.-G. Doping a Mott insulator: Physics of high-temperature superconductivity. *Reviews of Modern Physics* **78**, 17–85 (2006). URL https://link.aps.org/doi/10.1103/RevModPhys.78.17.

[85] Kotliar, G. & Ruckenstein, A. E. New functional integral approach to strongly correlated Fermi systems: The Gutzwiller approximation as a saddle point. *Physical Review Letters* **57**, 1362–1365 (1986). URL https://link.aps.org/doi/10.1103/PhysRevLett.57.1362.

[86] Deutsch, J. M. Quantum statistical mechanics in a closed system. *Physical Review A* **43**, 2046–2049 (1991). URL https://link.aps.org/doi/10.1103/PhysRevA.43.2046.

[87] Srednicki, M. Chaos and quantum thermalization. *Physical Review E* **50**, 888–901 (1994). URL https://link.aps.org/doi/10.1103/PhysRevE.50.888.

[88] Gogolin, C. & Eisert, J. Equilibration, thermalisation, and the emergence of statistical mechanics in closed quantum systems. *Reports on Progress in Physics* **79**, 056001 (2016). URL http://dx.doi.org/10.1088/0034-4885/79/5/056001.

[89] D'Alessio, L., Kafri, Y., Polkovnikov, A. & Rigol, M. From quantum chaos and eigenstate thermalization to statistical mechanics and thermodynamics. *Advances in Physics* **65**, 239–362 (2016). URL http://dx.doi.org/10.1080/00018732.2016.1198134.

[90] Collura, M. & Essler, F. H. How order melts after quantum quenches. *Physical Review B* **101**, 041110 (2020). URL https://journals.aps.org/prb/abstract/10.1103/PhysRevB.101.041110.

[91] Kramer, B. & MacKinnon, A. Localization: theory and experiment. *Reports on Progress in Physics* **56**, 1469 (1993). URL https://doi.org/10.1088/0034-4885/56/12/001.

[92] Thouless, D. J. Electrons in disordered systems and the theory of localization. *Physics Reports* **13**, 93–142 (1974). URL https://www.sciencedirect.com/science/article/pii/0370157374900295.

[93] Evers, F. & Mirlin, A. D. Anderson transitions. *Reviews of Modern Physics* **80**, 1355–1417 (2008). URL https://link.aps.org/doi/10.1103/RevModPhys.80.1355.

[94] Weng, Z., Sheng, D. & Ting, C. Mean-field description of the phase string effect in the t-J model. *Physical Review B* **59**, 8943 (1999). URL https://journals.aps.org/prb/abstract/10.1103/PhysRevB.59.8943.

[95] Wang, Y. Effective holon-holon interaction in the resonating-valence-bond state. *Physical Review B* **40**, 2698 (1989). URL https://journals.aps.org/prb/abstract/10.1103/PhysRevB.40.2698.

[96] Schrieffer, J., Wen, X. & Zhang, S. Dynamic spin fluctuations and the bag mechanism of high-Tc superconductivity. *Physical Review B* **39**, 11663 (1989). URL https://journals.aps.org/prb/abstract/10.1103/PhysRevB.39.11663.

[97] Nielsen, K. K., Pohl, T. & Bruun, G. Nonequilibrium hole dynamics in antiferromagnets: Damped strings and polarons. *Physical Review Letters* **129**, 246601 (2022). URL https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.129.246601.

[98] Kivlichan, I. D. *et al.* Quantum simulation of electronic structure with linear depth and connectivity. *Physical Review Letters* **120**, 110501 (2018). URL https://link.aps.org/doi/10.1103/PhysRevLett.120.110501.

[99] Foxen, B. *et al.* Demonstrating a continuous set of two-qubit gates for near-term quantum algorithms. *Physical Review Letters* **125**, 120504 (2020). URL https://link.aps.org/doi/10.1103/PhysRevLett.125.120504.

[100] Derby, C., Klassen, J., Bausch, J. & Cubitt, T. Compact fermion to qubit mappings. *Phys. Rev. B* **104**, 035118 (2021). URL https://link.aps.org/doi/10.1103/PhysRevB.104.035118.

[101] Bennett, C. H., DiVincenzo, D. P., Smolin, J. A. & Wootters, W. K. Mixed-state entanglement and quantum error correction. *Physical Review A* **54**, 3824–3851 (1996). URL https://link.aps.org/doi/10.1103/PhysRevA.54.3824.

[102] van den Berg, E., Minev, Z. K. & Temme, K. Model-free readout-error mitigation for quantum expectation values. *Physical Review A* **105**, 032620 (2022). URL https://link.aps.org/doi/10.1103/PhysRevA.105.032620.

[103] Hebenstreit, M., Jozsa, R., Kraus, B., Strelchuk, S. & Yoganathan, M. All pure fermionic non-Gaussian states are magic states for matchgate computations. *Physical Review Letters* **123**, 080503 (2019). URL https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.123.080503.

[104] Liu, Z.-W. & Winter, A. Many-body quantum magic. *PRX Quantum* **3**, 020333 (2022). URL https://journals.aps.org/prxquantum/abstract/10.1103/PRXQuantum.3.020333.

[105] Howard, M. & Campbell, E. Application of a resource theory for magic states to fault-tolerant quantum computing. *Physical Review Letters* **118**, 090501 (2017). URL https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.118.090501.

[106] Cudby, J. & Strelchuk, S. Gaussian decomposition of magic states for matchgate computations (2025). URL https://arxiv.org/abs/2307.12654. arXiv:2307.12654.

[107] Reardon-Smith, O., Oszmaniec, M. & Korzekwa, K. Improved simulation of quantum circuits dominated by free fermionic operations. *Quantum* **8**, 1549 (2024). URL https://quantum-journal.org/papers/q-2024-12-04-1549/.

[108] Dias, B. & Koenig, R. Classical simulation of non-Gaussian fermionic circuits. *Quantum* **8**, 1350 (2024). URL https://quantum-journal.org/papers/q-2024-05-21-1350/.

[109] Hakkaku, S., Tashima, Y., Mitarai, K., Mizukami, W. & Fujii, K. Quantifying fermionic nonlinearity of quantum circuits. *Physical Review Research* **4**, 043100 (2022). URL https://journals.aps.org/prresearch/abstract/10.1103/PhysRevResearch.4.043100.

[110] Bravyi, S. & König, R. Classical simulation of dissipative fermionic linear optics. *Quantum Information and Computation* **11**, 0925–0943 (2012). URL https://www.rintonpress.com/journals/doi/QIC12.11-12-2.html.

[111] Bosse, J. L. FLOYao (2022). URL https://doi.org/10.5281/zenodo.7303997. Version 1.2.0, Zenodo. DOI: 10.5281/zenodo.7303997.

[112] Rasmussen, C. E. Gaussian processes in machine learning. In *Advanced Lectures on Machine Learning* (Springer Berlin Heidelberg, 2004). URL https://doi.org/10.1007%2F978-3-540-28650-9_4.

[113] Bishop, C. M. *Pattern Recognition and Machine Learning (Information Science and Statistics)* (Springer-Verlag, Berlin, Heidelberg, 2006).

[114] Lebigot, E. O. `uncertainties` – a Python package for calculations with uncertainties. URL http://pythonhosted.org/uncertainties/.

[115] Wasserman, L. *All of Nonparametric Statistics*. Springer Texts in Statistics (Springer, New York, NY, 2007), 1 edn.

[116] Jaynes, E. T. Information theory and statistical mechanics. *Physical review* **106**, 620 (1957). URL https://journals.aps.org/pr/abstract/10.1103/PhysRev.106.620.

[117] Kim, Y. *et al.* Evidence for the utility of quantum computing before fault tolerance. *Nature* **618**, 500–505 (2023). URL https://www.nature.com/articles/s41586-023-06096-3.

[118] van den Berg, E., Minev, Z. K., Kandala, A. & Temme, K. Probabilistic error cancellation with sparse Pauli–Lindblad models on noisy quantum processors. *Nature Physics* **19**, 1116–1121 (2023). URL https://www.nature.com/articles/

s41567-023-02042-2.

[119] Gray, J. & Kourtis, S. Hyper-optimized tensor network contraction. *Quantum* **5**, 410 (2021). URL https://doi.org/10.22331/q-2021-03-15-410.

[120] White, S. R. & Feiguin, A. E. Real-time evolution using the density matrix renormalization group. *Physical Review Letters* **93**, 076401 (2004). URL https://link.aps.org/doi/10.1103/PhysRevLett.93.076401.

[121] Ronca, E., Li, Z., Jimenez-Hoyos, C. A. & Chan, G. K.-L. Time-Step Targeting Time-Dependent and Dynamical Density Matrix Renormalization Group Algorithms with ab Initio Hamiltonians. *Journal of Chemical Theory and Computation* **13**, 5560–5571 (2017). URL https://pubs.acs.org/doi/10.1021/acs.jctc.7b00682.

[122] Schollwöck, U. The density-matrix renormalization group in the age of matrix product states. *Annals of Physics* **326**, 96–192 (2011). URL http://dx.doi.org/10.1016/j.aop.2010.09.012.

[123] Haegeman, J., Lubich, C., Oseledets, I., Vandereycken, B. & Verstraete, F. Unifying time evolution and optimization with matrix product states. *Physical Review B* **94** (2016). URL http://dx.doi.org/10.1103/PhysRevB.94.165116.

[124] Paeckel, S. *et al.* Time-evolution methods for matrix-product states. *Annals of Physics* **411**, 167998 (2019). URL http://dx.doi.org/10.1016/j.aop.2019.167998.

[125] Chen, T., Guo, E., Zhang, W., Zhang, P. & Deng, Y. Tensor network Monte Carlo simulations for the two-dimensional random-bond Ising model. *Physical Review B* **111**, 094201 (2025). URL https://journals.aps.org/prb/abstract/10.1103/PhysRevB.111.094201.

[126] Schröder, F. A., Turban, D. H., Musser, A. J., Hine, N. D. & Chin, A. W. Tensor network simulation of multi-environmental open quantum dynamics via machine learning and entanglement renormalisation. *Nature communications* **10**, 1062 (2019). URL https://www.nature.com/articles/s41467-019-09039-7.

[127] Verstraete, F. & Cirac, J. I. Renormalization algorithms for quantum-many body systems in two and higher dimensions (2004). URL https://arxiv.org/abs/cond-mat/0407066. arXiv:cond-mat/0407066.

[128] Vidal, G. Class of quantum many-body states that can be efficiently simulated. *Physical Review Letters* **101**, 110501 (2008). URL https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.101.110501.

[129] Schuch, N., Wolf, M. M., Verstraete, F. & Cirac, J. I. Computational complexity of projected entangled pair states. *Physical Review Letters* **98**, 140506 (2007). URL https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.98.140506.

[130] Sander, A. *et al.* Quantum circuit simulation with a local time-dependent variational principle (2025). URL https://arxiv.org/abs/2508.10096. arXiv:2508.10096.

[131] Guo, C. *et al.* General-purpose quantum circuit simulator with projected entangled-pair states and the quantum supremacy frontier. *Physical Review Letters* **123** (2019). URL http://dx.doi.org/10.1103/PhysRevLett.123.190501.

[132] Noh, K., Jiang, L. & Fefferman, B. Efficient classical simulation of noisy random quantum circuits in one dimension. *Quantum* **4**, 318 (2020). URL http://dx.doi.org/10.22331/q-2020-09-11-318.

[133] Perez-Garcia, D., Verstraete, F., Wolf, M. M. & Cirac, J. I. Matrix product state representations (2007). URL https://arxiv.org/abs/quant-ph/0608197. arXiv:quant-ph/0608197.

[134] Google Cloud. *Compute Engine machine types*. Google LLC (2025). URL https://cloud.google.com/compute/docs/general-purpose-machines. Accessed: 2025-10-15.

[135] Angrisani, A. *et al.* Classically estimating observables of noiseless quantum circuits. *Physical Review Letters* **135**, 170602 (2025). URL https://link.aps.org/doi/10.1103/lh6x-7rc3.

[136] Fontana, E., Rudolph, M. S., Duncan, R., Rungger, I. & Cîrstoiu, C. Classical simulations of noisy variational quantum circuits. *npj Quantum Information* **11** (2025). URL http://dx.doi.org/10.1038/s41534-024-00955-1.

[137] Rudolph, M. S., Jones, T., Teng, Y., Angrisani, A. & Holmes, Z. Pauli propagation: A computational framework for simulating quantum systems (2025). URL https://arxiv.org/abs/2505.21606. arXiv:2505.21606.

[138] Luo, X.-Z., Liu, J.-G., Zhang, P. & Wang, L. Yao.jl: Extensible, efficient framework for quantum algorithm design. *Quantum* **4**, 341 (2020). URL http://dx.doi.org/10.22331/q-2020-10-11-341.

[139] Popp, M., Verstraete, F., Martín-Delgado, M. A. & Cirac, J. I. Localizable entanglement. *Physical Review A* **71** (2005). URL http://dx.doi.org/10.1103/PhysRevA.71.042306.

[140] Chen, J., Ding, D., Huang, C. & Kong, L. Linear cross-entropy benchmarking with Clifford circuits. *Physical Review A* **108**, 052613 (2023). URL https://journals.aps.org/pra/abstract/10.1103/PhysRevA.108.052613.

[141] Kaneda, T., Fujii, K. & Ueda, H. Large-scale quantum device benchmarking via lxeb with particle-number-conserving random quantum circuits (2025). URL https://arxiv.org/abs/2505.10820. arXiv:2505.10820.

[142] Mark, D. K., Choi, J., Shaw, A. L., Endres, M. & Choi, S. Benchmarking quantum simulators using ergodic quantum dynamics. *Physical Review Letters* **131**, 110601 (2023). URL https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.131.110601.

[143] Childs, A. M. & Su, Y. Nearly optimal lattice simulation by product formulas. *Physical Review Letters* **123** (2019). URL http://dx.doi.org/10.1103/PhysRevLett.123.050503.

[144] Heyl, M., Hauke, P. & Zoller, P. Quantum localization bounds Trotter errors in digital quantum simulation. *Science Advances* **5** (2019). URL http://dx.doi.org/10.1126/sciadv.aau8342.