Supplementary Information

Observable-Level Benchmarking of 15-Qubit Tilted-Field Ising Dynamics on a Single IBM Heron Processor

Samarth Lamba and Aleksandra A. Ziolkowska

Contents

1 Scope and saved data products **2**

2 Complete exact-versus-product-formula validation **2**

3 Fixed-H timestep convergence **3**

4 Reflection symmetry and indexing verification **4**
4.1 Hardware Reflection-Asymmetry Diagnostic 4

5 Full zero-baseline comparison **5**

6 Contraction-model details **6**
6.1 Inverse-Rescaling Diagnostic 6

7 Pearson component spread and serial-dependence sensitivity **7**

8 Complete total-magnetization trajectories **8**

9 All signed connected-correlation maps **9**

10 Logical circuit resources and missing physical resources **9**

11 Hardware uncertainty and Runtime-setting limitations **10**

12 Backend context **11**

13 Four-times-parameter stress test **12**

14 Agreement-breakdown sensitivity analysis **12**

15 Reproducibility details **14**

16 Quantum-information foundations **15**

17 Integrability, chaos, and operator growth **15**

18 Hamiltonian-scale regimes **16**

19 Quantum circuits and digital simulation **16**

20 Hardware-noise mechanisms **16**

21 Extended Trotterization derivation **17**

22 Expanded observable definitions **18**

23 Metric definitions and uncertainty detail **19**

24 Reproducibility and archived resources **19**

25 Extended limitations and future-work narrative **20**

26 Code and Data Availability **20**

1 Scope and saved data products

This Supplement reports analyses that support, qualify, or test the robustness of the main manuscript. All hardware calculations use the saved file

`data/hardware/ibm_graphs_3_4_5_every_0p5s_to_35s.npz`.

It contains six case keys, 71 times from 0 to 35 in increments of 0.5, and an array of shape $(6, 71, 29)$. The 29 values at each time are 15 local Z expectations and 14 Z_8Z_j expectations. Total magnetization and connected correlations are reconstructed deterministically as

$$M_z = \sum_{i=1}^{15} \langle Z_i \rangle, \tag{S1}$$

$$C_{8j}^{\mathrm{conn}} = \langle Z_8 Z_j \rangle - \langle Z_8 \rangle \langle Z_j \rangle. \tag{S2}$$

The matching simulation cache contains exact sparse matrix-exponential-action and ideal second-order product-formula values at the same 71 simulation times. No interpolation is used in hardware–simulation metrics. File-integrity tests verify case order, array shape, finite values, monotonically increasing time, and exact time-grid alignment.

The hardware cache does not contain raw counts or bitstrings, returned Estimator standard errors, individual job metadata, repeated runs, physical layouts, transpiled circuits, or path-specific calibration snapshots. Those quantities are listed as unavailable rather than reconstructed. Conditional trajectory-resampling intervals in this Supplement quantify sensitivity to serial structure in the saved trajectories; they do not replace experimental shot uncertainty or run-to-run uncertainty.

2 Complete exact-versus-product-formula validation

Table S1: Exact-versus-second-order-Trotter RMSE at $\Delta t = 0.1$ over the archived hardware sampling times.

Case	Local	Total	Connected
$h_z = 0, J/h_x = 0.9$	1.069×10^{-4}	1.192×10^{-3}	5.214×10^{-4}
$h_z = 0, J/h_x = 1.0$	1.155×10^{-4}	1.285×10^{-3}	4.739×10^{-4}
$h_z = 0, J/h_x = 1.4$	4.152×10^{-4}	2.877×10^{-3}	7.498×10^{-4}
$h_z = h_x, J/h_x = 0.9$	9.902×10^{-4}	8.335×10^{-3}	4.424×10^{-4}
$h_z = h_x, J/h_x = 1.0$	9.957×10^{-4}	8.027×10^{-3}	4.863×10^{-4}
$h_z = h_x, J/h_x = 1.4$	1.069×10^{-3}	8.049×10^{-3}	4.641×10^{-4}

Table S1 verifies all three observable families in every base-parameter case, not only central-site magnetization. The ideal second-order product-formula error is much smaller than the hardware discrepancy reported in the main manuscript. The largest local-magnetization RMSE is 1.069×10^{-3}, the largest total-magnetization RMSE is 8.335×10^{-3}, and the largest connected-correlation RMSE is 7.498×10^{-4}.

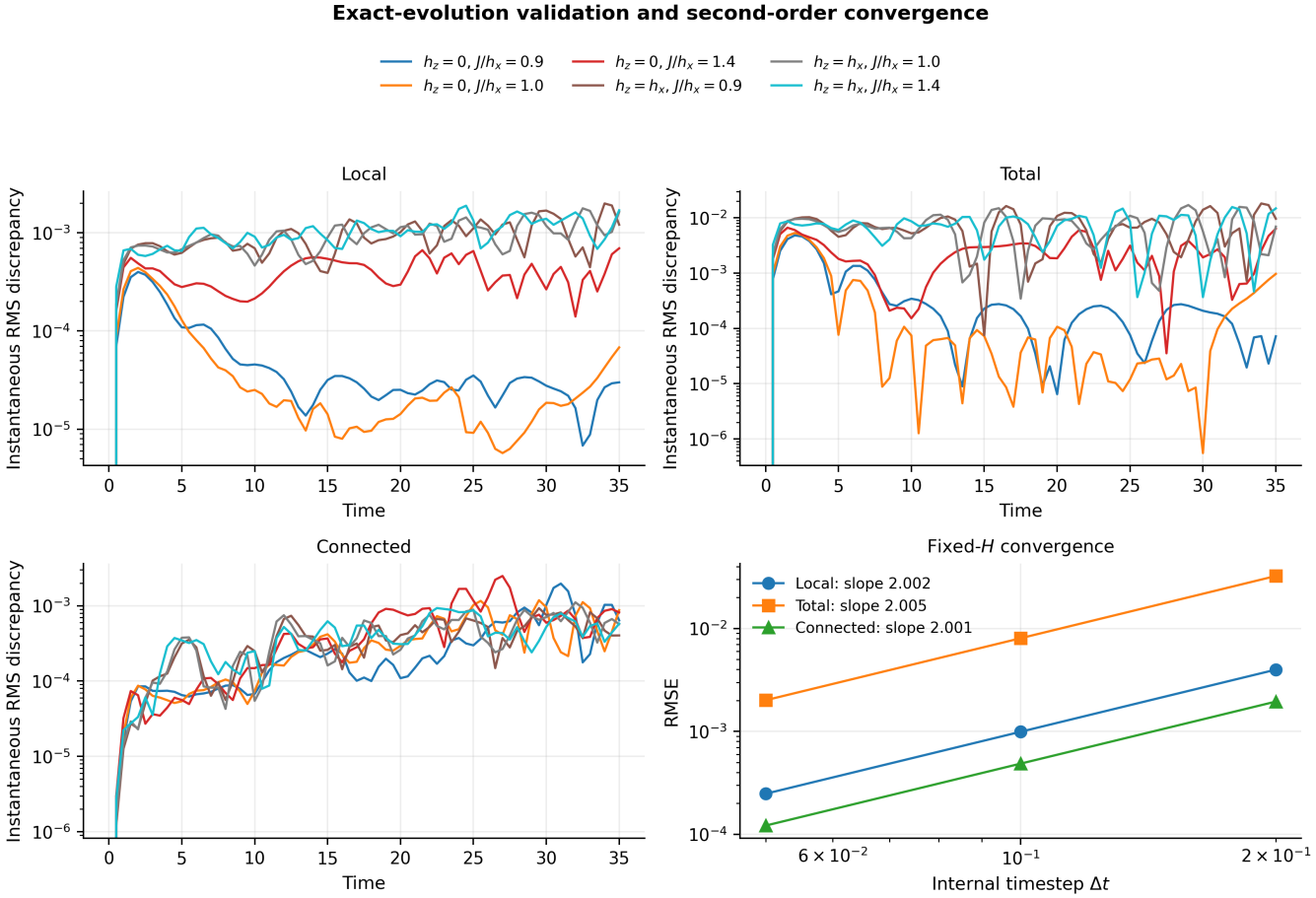


Figure S1: Exact-minus-product-formula residuals and fixed-H timestep convergence. The residual panels show all six base-parameter cases for local magnetization, total magnetization, and center-to-site connected correlations. The convergence panel holds the Hamiltonian and physical endpoint fixed while varying Δt. The fitted slopes are consistent with global second-order convergence.

3 Fixed-H timestep convergence

The timestep study holds the Hamiltonian, initial state, physical endpoint, and comparison grid fixed. The representative case is $h_z = h_x$, $J = h_x$. Exact evolution is sampled every 0.2 time units. Product-formula trajectories at $\Delta t = 0.20$, 0.10, and 0.05 are restricted to that common grid before RMSE is calculated.

Table S2: Fixed-H timestep convergence for $h_z = h_x$, $J = h_x$, and $0 \le t \le 35$.

Δt	Local RMSE	Total RMSE	Connected RMSE
0.20	0.003984	0.032350	0.001950
0.10	0.000993	0.008044	0.000487
0.05	0.000248	0.002008	0.000122
Slope	2.002	2.005	2.001

Each factor-of-two reduction in timestep reduces RMSE by approximately a factor of four. This is the appropriate test of global second-order convergence. Scaling all Hamiltonian parameters while holding Δt fixed is not a convergence study because it changes the dimensionless step size.

4 Reflection symmetry and indexing verification

For an open uniform chain with the center spin as the only flipped spin, the ideal dynamics are invariant under $i \mapsto 16-i$. Table S3 reports the largest mirror-pair discrepancy over all sampled times.

Table S3: Maximum ideal reflection-symmetry residuals.

Case	Method	Local	Connected
$h_z = 0, J/h_x = 0.9$	Exact	7.99e-15	8.10e-15
$h_z = 0, J/h_x = 0.9$	Trotter	6.33e-15	1.17e-14
$h_z = 0, J/h_x = 1.0$	Exact	7.99e-15	1.07e-14
$h_z = 0, J/h_x = 1.0$	Trotter	6.77e-15	8.55e-15
$h_z = 0, J/h_x = 1.4$	Exact	2.20e-14	1.50e-14
$h_z = 0, J/h_x = 1.4$	Trotter	3.77e-15	8.66e-15
$h_z = h_x, J/h_x = 0.9$	Exact	1.39e-14	1.60e-14
$h_z = h_x, J/h_x = 0.9$	Trotter	6.44e-15	8.77e-15
$h_z = h_x, J/h_x = 1.0$	Exact	2.23e-14	1.57e-14
$h_z = h_x, J/h_x = 1.0$	Trotter	7.99e-15	5.44e-15
$h_z = h_x, J/h_x = 1.4$	Exact	2.23e-14	2.38e-14
$h_z = h_x, J/h_x = 1.4$	Trotter	9.66e-15	6.22e-15

All exact and ideal product-formula residuals are at floating-point numerical precision. Consequently, a visible asymmetry in an ideal heatmap would indicate plotting or indexing error, not a physical effect of the uniform longitudinal field. Hardware mirror sites are not averaged because physical noise and the unknown transpiler mapping need not preserve the ideal symmetry.

4.1 Hardware Reflection-Asymmetry Diagnostic

The ideal Hamiltonian and centered initial state are reflection symmetric about site 8. This symmetry provides a reference-independent diagnostic of the archived hardware trajectories. For mirror distance $d = 1, \ldots, 7$, the hardware reflection difference was defined as

$$\Delta_d(t) = \langle Z_{8-d}(t) \rangle_{\mathrm{hw}} - \langle Z_{8+d}(t) \rangle_{\mathrm{hw}}. \tag{S3}$$

The corresponding ideal differences remain at floating-point precision, with maximum residuals between 3.77×10^{-15} and 9.66×10^{-15} across the six cases. Nonzero hardware values of $\Delta_d(t)$ therefore quantify execution asymmetry without requiring comparison with the ideal trajectory.

To summarize the seven mirror-pair trajectories, the root-mean-square reflection asymmetry was defined as

$$A_{\mathrm{RMS}} = \sqrt{\frac{1}{7N_t} \sum_{d=1}^{7} \sum_{k=1}^{N_t} \Delta_d(t_k)^2}, \tag{S4}$$

where $N_t = 71$. The resulting A_{RMS} values range from 0.117 to 0.167 across the six benchmark cases, showing a numerical left–right asymmetry in the archived hardware expectation values despite the numerical reflection symmetry of the ideal trajectories.

Table S4: Hardware reflection-asymmetry diagnostic for local magnetization. A_{RMS} is calculated across all seven mirror pairs and 71 archived hardware times. Early, intermediate, and late values use $0 \le t \le 5$, $5 < t \le 15$, and $15 < t \le 35$, respectively. The final column gives the RMS contribution from the $d = 4$ mirror pair, corresponding to logical sites 4 and 12.

h_z/h_x	J/h_x	A_{RMS}	Early RMS	Mid RMS	Late RMS	$d = 4$ RMS
0	0.9	0.1665	0.3229	0.1642	0.0840	0.2745
0	1.0	0.1549	0.2926	0.1574	0.0816	0.2775
0	1.4	0.1350	0.2604	0.1273	0.0748	0.2669
1	0.9	0.1344	0.2582	0.1257	0.0764	0.2709
1	1.0	0.1314	0.2591	0.1144	0.0752	0.2395
1	1.4	0.1170	0.2152	0.1046	0.0782	0.2143

The asymmetry does not increase with evolution time. In all six cases, the early-time RMS is larger than the intermediate- and late-time values, with the late-time RMS reduced to approximately 26–36% of the corresponding early-time value. The largest individual asymmetry in every case occurs at $t = 1.0$ for the $d = 4$ mirror pair, corresponding to logical sites 4 and 12. Pairwise RMS values also show that this mirror pair provides the largest contribution in all six cases, accounting for approximately 39–58% of the total squared asymmetry, although other pairs, particularly $d = 2$ and in some cases $d = 7$, also contribute.

At each matched value of J/h_x, the archived $h_z = h_x$ trajectory has a smaller overall reflection asymmetry than the corresponding $h_z = 0$ trajectory. This pattern is reported descriptively and is not assigned to the longitudinal field or to any particular hardware mechanism. Because the historical physical-qubit layout and path-specific calibration information were not retained, the observed mirror asymmetry cannot be attributed to specific physical qubits, couplers, routing choices, readout errors, or other device-level causes.

5 Full zero-baseline comparison

The zero-baseline skill is

$$S_0 = 1 - \frac{\mathrm{RMSE}(H, T)}{\mathrm{RMSE}(0, T)}. \tag{S5}$$

The full 18-row table is reproduced below. Positive values indicate that the hardware trajectory is closer to the ideal product-formula trajectory than an all-zero output.

Table S5: Hardware RMSE and skill relative to an all-zero trajectory. Positive skill means that hardware outperforms zero under RMSE.

Case	Observable	Hardware RMSE	Zero RMSE	S_0 (%)
$h_z = 0$, $J/h_x = 0.9$	Local	0.1791	0.2116	15.4
$h_z = 0$, $J/h_x = 0.9$	Total	1.7072	2.6927	36.6
$h_z = 0$, $J/h_x = 0.9$	Connected	0.1254	0.1238	-1.3
$h_z = 0$, $J/h_x = 1.0$	Local	0.1731	0.2159	19.8
$h_z = 0$, $J/h_x = 1.0$	Total	1.5601	2.7367	43.0
$h_z = 0$, $J/h_x = 1.0$	Connected	0.1269	0.1259	-0.8
$h_z = 0$, $J/h_x = 1.4$	Local	0.2810	0.3373	16.7
$h_z = 0$, $J/h_x = 1.4$	Total	1.7375	3.6673	52.6
$h_z = 0$, $J/h_x = 1.4$	Connected	0.1919	0.1977	2.9
$h_z = h_x$, $J/h_x = 0.9$	Local	0.5385	0.7098	24.1
$h_z = h_x$, $J/h_x = 0.9$	Total	7.2714	10.0272	27.5
$h_z = h_x$, $J/h_x = 0.9$	Connected	0.0875	0.0857	-2.2
$h_z = h_x$, $J/h_x = 1.0$	Local	0.5494	0.7209	23.8
$h_z = h_x$, $J/h_x = 1.0$	Total	7.4497	10.2053	27.0
$h_z = h_x$, $J/h_x = 1.0$	Connected	0.0916	0.0899	-1.9
$h_z = h_x$, $J/h_x = 1.4$	Local	0.5443	0.7540	27.8
$h_z = h_x$, $J/h_x = 1.4$	Total	7.4404	10.7275	30.6
$h_z = h_x$, $J/h_x = 1.4$	Connected	0.1001	0.1009	0.8

All six local- and total-magnetization cases beat zero. Connected correlations fail to beat zero for $h_z = 0$, $J/h_x = 0.9$; $h_z = 0$, $J/h_x = 1.0$; $h_z = h_x$, $J/h_x = 0.9$; and $h_z = h_x$, $J/h_x = 1.0$. Improvements for the other two cases

are 2.9% and 0.8%, respectively. These results do not establish robust hardware preservation of connected-correlation dynamics.

6 Contraction-model details

The empirical model

$$H_i(t) = e^{-\Gamma t} T_i(t) + \epsilon_i(t) \tag{S6}$$

is fitted by pooled least squares across all 15 sites and 71 hardware times within each case. A circular moving-block residual bootstrap uses blocks of seven samples (3.5 time units) and 1000 replicates. Residual blocks are resampled after subtracting their site-wise means and are added to the fitted deterministic trajectory before refitting. This interval is conditional on the single saved trajectory.

Table S6: Empirical local-magnetization contraction fits. Intervals are conditional block-bootstrap intervals, not shot uncertainty.

Case	Γ [95% interval]	Half-life	Raw RMSE	Model RMSE
$h_z = 0, J/h_x = 0.9$	1.022 [0.822, 1.272]	0.678	0.179	0.142
$h_z = 0, J/h_x = 1.0$	0.851 [0.732, 1.204]	0.814	0.173	0.141
$h_z = 0, J/h_x = 1.4$	0.572 [0.450, 0.932]	1.212	0.281	0.145
$h_z = h_x, J/h_x = 0.9$	0.126 [0.096, 0.196]	5.481	0.538	0.177
$h_z = h_x, J/h_x = 1.0$	0.128 [0.100, 0.191]	5.426	0.549	0.175
$h_z = h_x, J/h_x = 1.4$	0.104 [0.079, 0.152]	6.697	0.544	0.173

The model does not include a time-zero scale factor, site-dependent readout offsets, coherent phase shifts, or correlated gate errors. Its ability to reduce RMSE therefore identifies amplitude contraction as an important visible component, not as a unique depolarizing rate. The forward fit is descriptive. The inverse-rescaling diagnostic below explicitly tests, and rejects, using $e^{+\Gamma t}$ to reconstruct the ideal trajectories from the hardware values.

6.1 Inverse-Rescaling Diagnostic

As an additional diagnostic of the empirical contraction model, the fitted decay constants were applied in the inverse direction to the archived hardware local-magnetization trajectories. For each site,

$$\langle Z_i(t) \rangle_{\text{rescaled}} = e^{+\Gamma t} \langle Z_i(t) \rangle_{\text{hw}}. \tag{S7}$$

The previously fitted full-precision values of Γ were retained without refitting, and the transformation was applied to all 15 sites at all 71 archived hardware sampling times. No rescaled values were clipped to the physical expectation-value interval $[-1, 1]$, because the purpose of this calculation was to test the scalar-contraction hypothesis rather than to construct an error-mitigated estimator.

Table S7 compares the raw and inverse-rescaled hardware trajectories with the same ideal second-order Trotter reference. Inverse rescaling worsens agreement in every parameter case. For $h_z = 0$, the effect is particularly severe because the fitted decay constants are large: the maximum amplification factors range from approximately 4.92×10^8 to 3.39×10^{15} over the sampled time window. The rescaled NRMSE consequently increases from 0.0895–0.1405 to between 3.44×10^6 and 1.87×10^{13}.

For the $h_z = h_x$ cases, the fitted decay constants are smaller, but inverse rescaling still substantially worsens the trajectories. Raw NRMSE values of 0.269–0.275 increase to 0.940–1.692. The Fisher-z mean temporal correlations also change from positive raw values of 0.402–0.459 to negative values between -0.301 and -0.260. Thus, the deterioration is visible both in amplitude-sensitive error and in temporal trajectory shape.

Table S7: Inverse-rescaling diagnostic for local magnetization. NRMSE is RMSE divided by the local-expectation span of 2 and is computed relative to the same ideal second-order Trotter trajectory before and after applying $e^{+\Gamma t}$ to the archived hardware values. The reported correlation is the Fisher-z mean of the 15 site-wise temporal Pearson coefficients. The percentages of rescaled values outside $[-1, 1]$ are reported in the following paragraph.

h_z/h_x	J/h_x	Γ	Raw NRMSE	Rescaled NRMSE	Raw r	Rescaled r
0	0.9	1.0217	0.0895	1.873×10^{13}	0.724	-0.037
0	1.0	0.8512	0.0865	4.965×10^{10}	0.746	-0.061
0	1.4	0.5718	0.1405	3.445×10^{6}	0.694	-0.070
1	0.9	0.1265	0.2692	1.654	0.402	-0.286
1	1.0	0.1278	0.2747	1.692	0.425	-0.260
1	1.4	0.1035	0.2721	0.940	0.459	-0.301

The inverse transformation also produces widespread nonphysical amplitudes. The fraction of rescaled values outside $[-1, 1]$ is 96.6%, 95.1%, and 91.9% for the three $h_z = 0$ cases and 59.2%, 60.3%, and 54.9% for the three $h_z = h_x$ cases, respectively. The maximum absolute rescaled expectation value reaches approximately 4.58×10^{14} in the most strongly amplified case.

These results show that the fitted exponential contraction should not be interpreted as a recoverable physical decay channel. Although applying $e^{-\Gamma t}$ to the ideal trajectory reduces the forward-model RMSE, applying the inverse factor to the hardware data amplifies late-time fluctuations and does not recover the ideal temporal dynamics. The contraction model is therefore retained as a descriptive measure of the observed amplitude reduction rather than as an error-mitigation or reconstruction procedure.

7 Pearson component spread and serial-dependence sensitivity

Pearson coefficients describe temporal shape but can be misleading when amplitudes contract or trajectories are serially correlated. For each multicomponent observable, component coefficients are transformed with Fisher's $z = \operatorname{atanh}(r)$, averaged in z space, and transformed back. The table reports the component minimum, median, and maximum rather than only the aggregate.

Table S8: Fisher-z mean Pearson coefficients, component spread, and primary conditional block-bootstrap interval (seven samples per block). Intervals are trajectory diagnostics, not shot-level confidence intervals.

Case	Observable	Mean	Min.	Median	Max.	95% interval
$h_z = 0, J/h_x = 0.9$	Local	0.724	0.233	0.693	0.959	[-0.080, 0.896]
$h_z = 0, J/h_x = 0.9$	Total	0.844	0.844	0.844	0.844	[-0.784, 0.946]
$h_z = 0, J/h_x = 0.9$	Connected	-0.095	-0.529	-0.036	0.146	[-0.162, 0.015]
$h_z = 0, J/h_x = 1.0$	Local	0.746	0.214	0.786	0.954	[0.015, 0.912]
$h_z = 0, J/h_x = 1.0$	Total	0.868	0.868	0.868	0.868	[-0.632, 0.962]
$h_z = 0, J/h_x = 1.0$	Connected	-0.102	-0.542	-0.027	0.093	[-0.182, -0.002]
$h_z = 0, J/h_x = 1.4$	Local	0.694	0.274	0.671	0.922	[-0.212, 0.835]
$h_z = 0, J/h_x = 1.4$	Total	0.830	0.830	0.830	0.830	[-0.693, 0.911]
$h_z = 0, J/h_x = 1.4$	Connected	-0.031	-0.294	-0.009	0.210	[-0.103, 0.045]
$h_z = h_x, J/h_x = 0.9$	Local	0.402	0.055	0.337	0.773	[0.119, 0.592]
$h_z = h_x, J/h_x = 0.9$	Total	0.322	0.322	0.322	0.322	[-0.447, 0.697]
$h_z = h_x, J/h_x = 0.9$	Connected	-0.023	-0.241	-0.037	0.210	[-0.086, 0.035]
$h_z = h_x, J/h_x = 1.0$	Local	0.425	0.120	0.333	0.733	[0.177, 0.595]
$h_z = h_x, J/h_x = 1.0$	Total	0.343	0.343	0.343	0.343	[-0.368, 0.710]
$h_z = h_x, J/h_x = 1.0$	Connected	0.004	-0.470	0.081	0.237	[-0.045, 0.065]
$h_z = h_x, J/h_x = 1.4$	Local	0.459	-0.026	0.469	0.715	[0.226, 0.597]
$h_z = h_x, J/h_x = 1.4$	Total	0.407	0.407	0.407	0.407	[-0.323, 0.738]
$h_z = h_x, J/h_x = 1.4$	Connected	-0.139	-0.423	-0.106	0.020	[-0.194, -0.063]

The primary conditional interval uses seven samples per circular block. Sensitivity runs use block lengths 4 and 10; all values are available in `outputs/revision/tables/pearson_block_bootstrap.csv`. Connected-correlation intervals include zero in four of the six cases. The $h_z = 0, J/h_x = 1.0$ interval is narrowly negative at $[-0.182, -0.002]$, while the $h_z = h_x$, $J/h_x = 1.4$ interval is negative at $[-0.194, -0.063]$. None of the six cases therefore provides evidence of positive connected-correlation temporal agreement under this diagnostic. Local-magnetization coefficients are positive on average but heterogeneous across sites. Site-8 coefficients, reported in the machine-readable table, range from 0.294 to 0.578.

These intervals do not quantify shot noise, IBM Estimator uncertainty, calibration drift, or repeatability. They answer a narrower question: how sensitive is a trajectory-level shape statistic to serially dependent time segments in the single hardware execution?

8 Complete total-magnetization trajectories

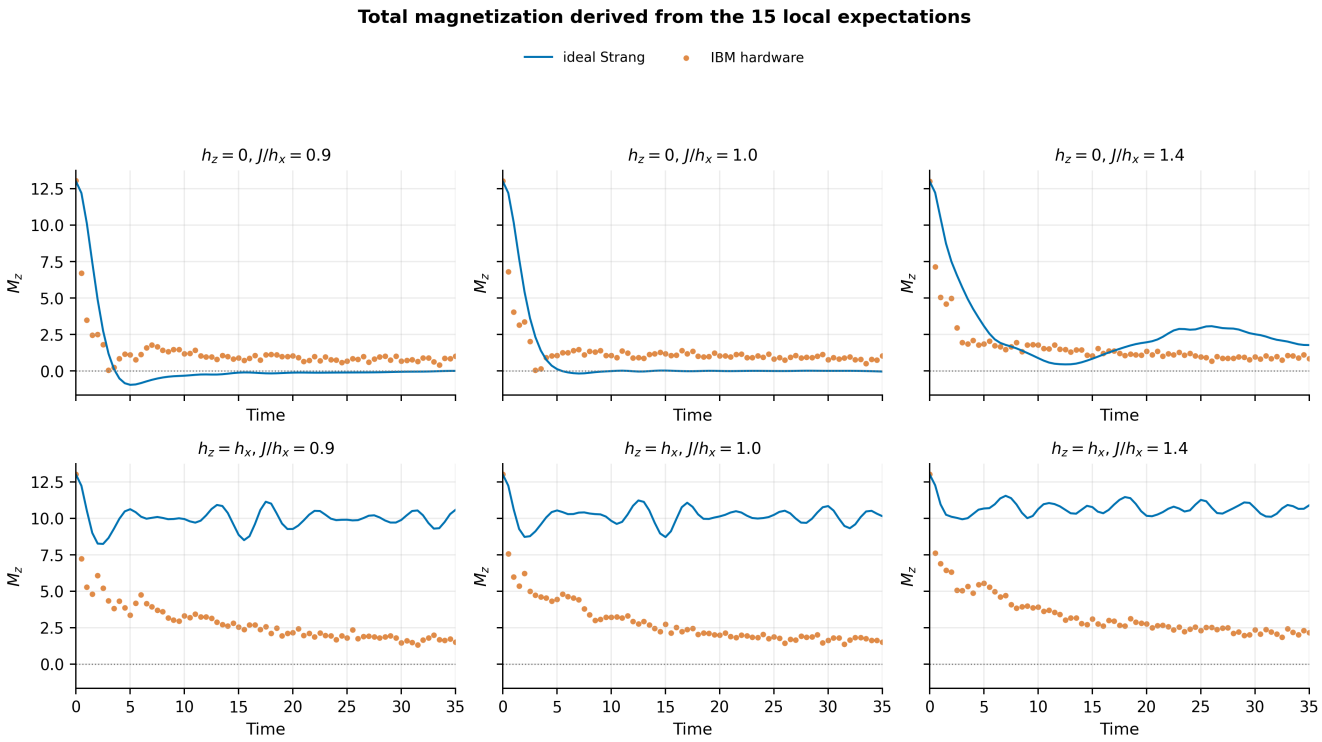


Figure S2: Total magnetization for all six parameter cases. Curves are ideal second-order product-formula values and markers are hardware values obtained by summing the same 15 local expectations used in the local-magnetization benchmark. Total magnetization is therefore a derived aggregation, not a separate hardware circuit family.

The ideal initial value is $M_z(0) = 13$. Total-magnetization RMSE is numerically larger because the observable sums 15 components and has a different scale from local magnetization. Cross-observable comparisons use zero-baseline skill rather than direct comparison of raw RMSE values.

9 All signed connected-correlation maps

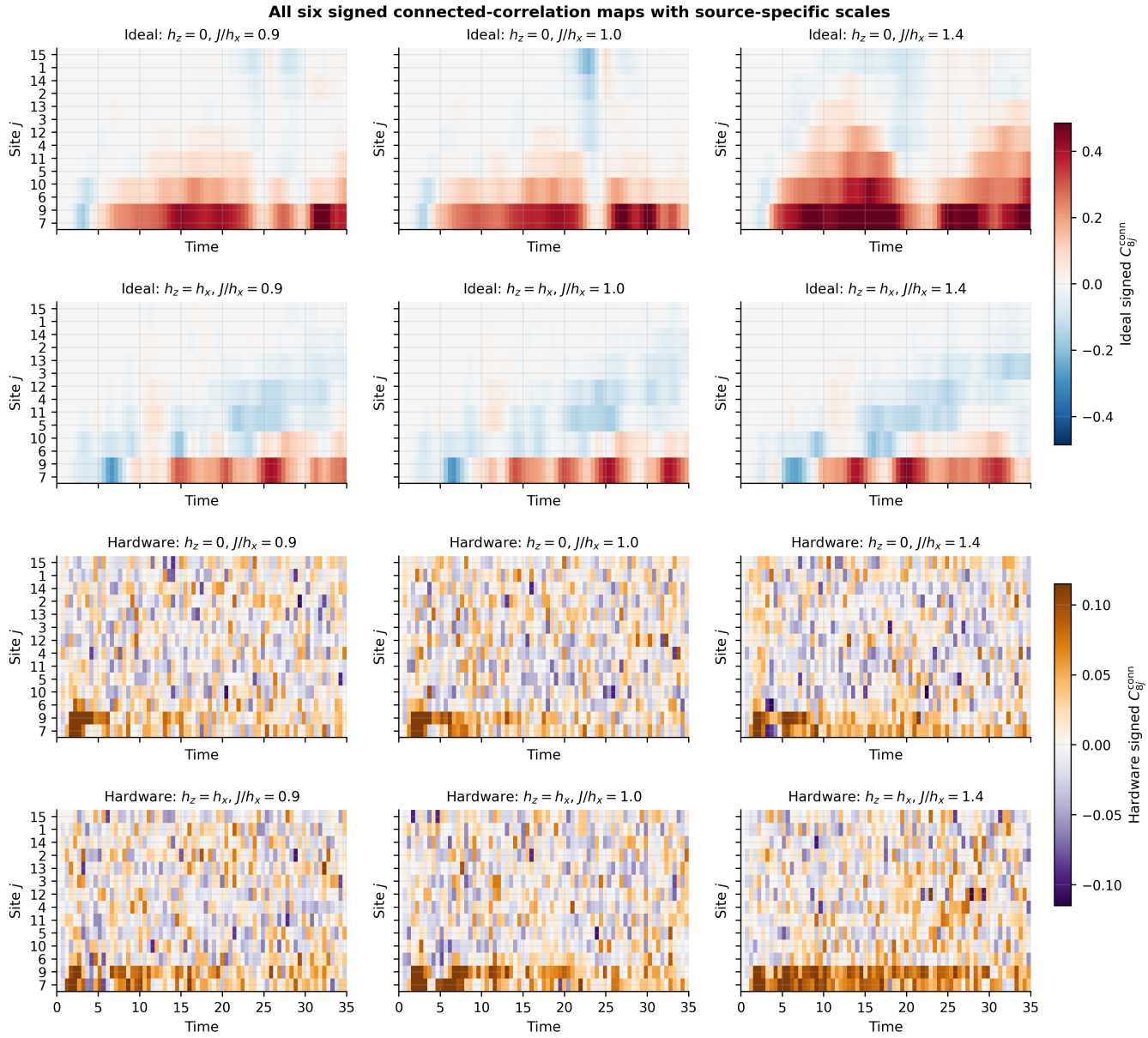


Figure S3: All six signed center-to-site connected-correlation maps. Ideal and hardware groups use separate, explicitly labeled diverging scales so that the smaller hardware range remains visible without implying equal amplitude. Site 8 is absent because self-correlation is excluded.

The maps retain sign information. The hardware color limits use the 99th percentile of absolute finite hardware values, with a minimum limit of 0.05; colors saturate beyond those limits. The ideal and hardware groups use their own scales, as labeled. Absolute-value maps are unsuitable for a propagation claim because they turn zero-centered fluctuations into positive intensity. A hardware first-arrival analysis would require a threshold based on the uncertainty of the connected estimator, including covariance between $\langle Z_8 Z_j \rangle$, $\langle Z_8 \rangle$, and $\langle Z_j \rangle$. Those data were not saved. No hardware front velocity is therefore reported.

10 Logical circuit resources and missing physical resources

Table S9: Logical resources implied by the unmerged circuit construction and by algebraically combining adjacent diagonal half-layers. These are not actual transpiled native-gate resources.

t	Steps	R_{ZZ} raw	R_Z raw	R_X	R_{ZZ} merged
0.0	0	0	0	0	0
0.1	1	28	30	15	28
0.5	5	140	150	75	84
1.0	10	280	300	150	154
5.0	50	1400	1500	750	714
10.0	100	2800	3000	1500	1414
20.0	200	5600	6000	3000	2814
35.0	350	9800	10500	5250	4914

The raw logical counts follow directly from the circuit builder. At $t = 35$, 350 internal steps produce 9800 logical R_{ZZ} gates, 5250 R_X gates, and, for $h_z = h_x$, 10500 R_Z gates before transpilation. Adjacent diagonal half-layers can be combined algebraically between steps; the final column illustrates that idealized merge for R_{ZZ}. The physical transpiler output may differ because routing, basis translation, cancellation, and hardware scheduling are backend- and layout-dependent. Since the QPY and transpiled circuits were not saved, no native two-qubit depth, physical path, or calibration-weighted accumulated error is reported.

More generally, for an evolution containing $n = t/\Delta t$ symmetric Trotter steps, the unmerged logical construction contains $28n$ nearest-neighbor R_{ZZ} rotations and $15n$ R_X rotations. When $h_z \neq 0$, it additionally contains $30n$ R_Z rotations. Algebraically merging adjacent diagonal half-layers gives $14(n + 1)$ logical R_{ZZ} operations for $n \geq 1$. At $t = 35$, 350 internal steps therefore produce 9800 logical R_{ZZ} gates, 5250 R_X gates, and, for $h_z = h_x$, 10500 R_Z gates before transpilation; the corresponding idealized merged R_{ZZ} count is 4914. These expressions provide the logical resource count at every sampled time. The physical transpiler output may differ because routing, basis translation, cancellation, and hardware scheduling are backend- and layout-dependent. Since the QPY and historical transpiled circuits were not saved, these logical counts are not used to infer native two-qubit depth, the physical qubit path, a calibration-weighted accumulated error, or a device-specific maximum usable circuit depth.

An offline recovery attempt was made using the original run window and backend identity. Historical job discovery could not begin because the original submitting account could not be authenticated. Consequently, zero historical jobs or PUB results were retrieved, and all 426 archived cache rows remain without matched historical Runtime metadata. No new IBM Quantum job was submitted during this recovery attempt.

11 Hardware uncertainty and Runtime-setting limitations

For a directly measured Pauli expectation P with $N_s = 2048$ independent shots,

$$\mathrm{SE}(\langle P \rangle) = \sqrt{\frac{1 - \langle P \rangle^2}{2048}} \leq 0.0221. \tag{S8}$$

This upper bound does not determine the uncertainty of connected correlations. If $C = Z_{8j} - Z_8 Z_j$, a covariance-aware delta method contains variances of all three estimators and covariance terms. Bitstring bootstrap would be preferable because all terms can be recomputed from each resample. Neither procedure can be applied to expectation values alone.

IBM Runtime EstimatorV2 was configured with `resilience_level=1` and XpXm dynamical decoupling. Complete returned job metadata were not saved, so these are reported as requested settings rather than verified per-job realizations. A future run should save job IDs, primitive result metadata, estimator standard errors, measurement counts, session and backend identity, QPY circuits, physical layouts, calibration snapshots, and transpiled resource counts for every timepoint.

12 Backend context

The following table records aggregate backend properties reported around the run period. These values are not the calibration of the unknown physical qubit path and are not used to calculate a circuit-specific noise budget.

Table S10: Run configuration and aggregate `ibm_marrakesh` context.

Backend property	Value
Backend name	`ibm_marrakesh`
Run date	May 16, 2026
Shots per circuit	2048
Hardware sampling interval	0.5 time units; internal $\Delta t = 0.1$
Hardware times and count	$0 \leq t \leq 35$; 71 per case; 426 total
Transpilation optimization level	0
Runtime settings requested	`resilience_level=1`; XpXm dynamical decoupling
User-implemented mitigation	None
Physical layout	Not fixed by the user and not saved
Python version	3.13.5
Qiskit version	2.3.0
qiskit-aer version	0.17.2; not used for the exact matrix-exponential-action reference
qiskit-ibm-runtime version	0.45.1
Qubits	156
Couplers	176
Processor and QPU version	Heron r2; 1.0.21
Region	Washington DC (us-east)
Basis gates	`cz`, `id`, `rx`, `rz`, `rzz`, `sx`, `x`
CLOPS	300K
Median two-qubit error	2.49×10^{-3}
Layered two-qubit error	4.16×10^{-3}
Best two-qubit error	1.32×10^{-3}
Median readout error	1.06×10^{-2}
Median CZ error	2.489×10^{-3}
Median SX error	3.273×10^{-4}
Median T_1 and T_2	180.53 μs; 89.13 μs

13 Four-times-parameter stress test

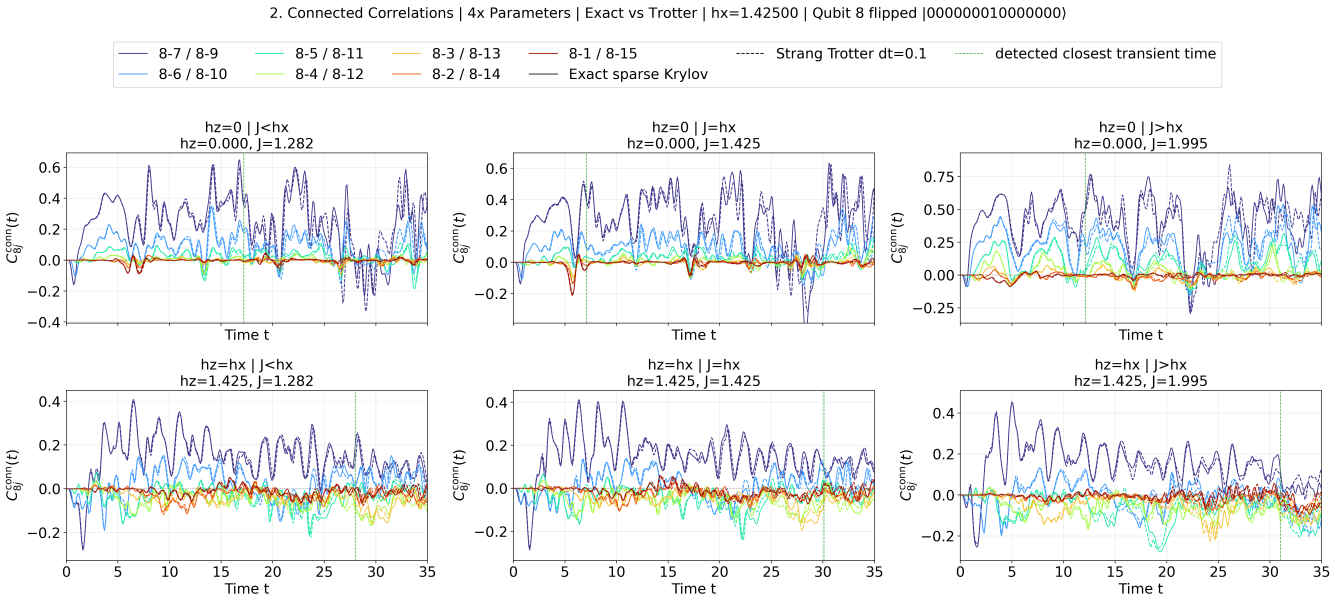


Figure S4: Exact-versus-product-formula connected-correlation stress test after multiplying J, h_x, and h_z by four while retaining $\Delta t = 0.1$. The resulting deviation is expected because the dimensionless quantities $J\Delta t$, $h_x\Delta t$, and $h_z\Delta t$ are four times larger. This figure is not a timestep-convergence test and is not used to validate the base-parameter run.

This stress test remains useful for illustrating the product formula's dependence on dimensionless step size. It does not show that the base circuit is inaccurate, nor can it be used to infer the order of convergence, because the Hamiltonian is changed.

14 Agreement-breakdown sensitivity analysis

The agreement-breakdown time t_B is reported only as a supplementary sensitivity statistic and is expressed in simulation time units. Because the archived hardware data were sampled every 0.5 time units, t_B can change only in increments of 0.5. It is defined as the first hardware sample at which a normalized discrepancy exceeds a threshold for a specified number of consecutive samples. Thresholds 0.05, 0.10, and 0.15 and persistence requirements 2, 3, and 4 were tested. Values > 35 are right-censored.

The agreement breakdown time requires careful interpretation. Under the original baseline definition, local and total magnetization crossed the threshold at $t_B = 0.5$ in every parameter case. Across the full sensitivity analysis, local magnetization produced $t_B = 0.5$ for every tested definition, while total magnetization was more sensitive to the persistence requirement.

Connected-correlation breakdown times depend strongly on the selected threshold. At a threshold of 0.05, breakdown occurred between $t = 2.0$ and $t = 24.5$, while at a threshold of 0.15 every case remained right-censored beyond $t = 35$. These later breakdown times do not indicate strong correlation fidelity, because the connected-correlation Pearson coefficients remain close to zero or negative. The threshold can remain uncrossed because the correlation signal and its absolute hardware–simulation discrepancy are small relative to the fixed normalization span.

Connected-correlation t_B therefore changes substantially with the threshold and persistence requirement because both the ideal signal and the absolute discrepancy are small relative to the normalization span. A late or censored t_B does not independently demonstrate faithful dynamics. The main analysis uses raw RMSE, zero-baseline skill, component-wise shape statistics, and signal-scale context instead of treating t_B as a headline fidelity measure.

Table S11: Sensitivity of the agreement-breakdown time t_B at normalized-discrepancy threshold 0.05. The columns give the required number of consecutive hardware samples. Right-censored values are reported as > 35.0.

Parameter regime	Observable	2 samples	3 samples	4 samples
$h_z = 0, J < h_x$	Local magnetization	0.5	0.5	0.5
$h_z = 0, J < h_x$	Total magnetization	0.5	0.5	0.5
$h_z = 0, J < h_x$	Connected correlations	2.0	2.0	11.5
$h_z = 0, J = h_x$	Local magnetization	0.5	0.5	0.5
$h_z = 0, J = h_x$	Total magnetization	0.5	0.5	0.5
$h_z = 0, J = h_x$	Connected correlations	2.0	2.0	9.0
$h_z = 0, J > h_x$	Local magnetization	0.5	0.5	0.5
$h_z = 0, J > h_x$	Total magnetization	0.5	0.5	0.5
$h_z = 0, J > h_x$	Connected correlations	4.0	4.0	4.0
$h_z = h_x, J < h_x$	Local magnetization	0.5	0.5	0.5
$h_z = h_x, J < h_x$	Total magnetization	0.5	0.5	0.5
$h_z = h_x, J < h_x$	Connected correlations	6.0	6.0	24.5
$h_z = h_x, J = h_x$	Local magnetization	0.5	0.5	0.5
$h_z = h_x, J = h_x$	Total magnetization	0.5	0.5	0.5
$h_z = h_x, J = h_x$	Connected correlations	5.5	5.5	5.5
$h_z = h_x, J > h_x$	Local magnetization	0.5	0.5	0.5
$h_z = h_x, J > h_x$	Total magnetization	0.5	0.5	0.5
$h_z = h_x, J > h_x$	Connected correlations	5.0	5.0	5.0

Table S12: Sensitivity of the agreement-breakdown time t_B at normalized-discrepancy threshold 0.10. The columns give the required number of consecutive hardware samples. Right-censored values are reported as > 35.0.

Parameter regime	Observable	2 samples	3 samples	4 samples
$h_z = 0, J < h_x$	Local magnetization	0.5	0.5	0.5
$h_z = 0, J < h_x$	Total magnetization	0.5	0.5	> 35.0
$h_z = 0, J < h_x$	Connected correlations	31.5	31.5	> 35.0
$h_z = 0, J = h_x$	Local magnetization	0.5	0.5	0.5
$h_z = 0, J = h_x$	Total magnetization	0.5	0.5	> 35.0
$h_z = 0, J = h_x$	Connected correlations	26.5	> 35.0	> 35.0
$h_z = 0, J > h_x$	Local magnetization	0.5	0.5	0.5
$h_z = 0, J > h_x$	Total magnetization	0.5	0.5	> 35.0
$h_z = 0, J > h_x$	Connected correlations	9.0	9.0	11.0
$h_z = h_x, J < h_x$	Local magnetization	0.5	0.5	0.5
$h_z = h_x, J < h_x$	Total magnetization	0.5	0.5	2.5
$h_z = h_x, J < h_x$	Connected correlations	> 35.0	> 35.0	> 35.0
$h_z = h_x, J = h_x$	Local magnetization	0.5	0.5	0.5
$h_z = h_x, J = h_x$	Total magnetization	0.5	0.5	2.5
$h_z = h_x, J = h_x$	Connected correlations	> 35.0	> 35.0	> 35.0
$h_z = h_x, J > h_x$	Local magnetization	0.5	0.5	0.5
$h_z = h_x, J > h_x$	Total magnetization	0.5	0.5	0.5
$h_z = h_x, J > h_x$	Connected correlations	> 35.0	> 35.0	> 35.0

Table S13: Sensitivity of the agreement-breakdown time t_B at normalized-discrepancy threshold 0.15. The columns give the required number of consecutive hardware samples. Right-censored values are reported as > 35.0.

Parameter regime	Observable	2 samples	3 samples	4 samples
$h_z = 0, J < h_x$	Local magnetization	0.5	0.5	0.5
$h_z = 0, J < h_x$	Total magnetization	0.5	0.5	> 35.0
$h_z = 0, J < h_x$	Connected correlations	> 35.0	> 35.0	> 35.0
$h_z = 0, J = h_x$	Local magnetization	0.5	0.5	0.5
$h_z = 0, J = h_x$	Total magnetization	0.5	0.5	> 35.0
$h_z = 0, J = h_x$	Connected correlations	> 35.0	> 35.0	> 35.0
$h_z = 0, J > h_x$	Local magnetization	0.5	0.5	0.5
$h_z = 0, J > h_x$	Total magnetization	0.5	> 35.0	> 35.0
$h_z = 0, J > h_x$	Connected correlations	> 35.0	> 35.0	> 35.0
$h_z = h_x, J < h_x$	Local magnetization	0.5	0.5	0.5
$h_z = h_x, J < h_x$	Total magnetization	0.5	3.5	3.5
$h_z = h_x, J < h_x$	Connected correlations	> 35.0	> 35.0	> 35.0
$h_z = h_x, J = h_x$	Local magnetization	0.5	0.5	0.5
$h_z = h_x, J = h_x$	Total magnetization	0.5	3.5	3.5
$h_z = h_x, J = h_x$	Connected correlations	> 35.0	> 35.0	> 35.0
$h_z = h_x, J > h_x$	Local magnetization	0.5	0.5	0.5
$h_z = h_x, J > h_x$	Total magnetization	2.5	2.5	2.5
$h_z = h_x, J > h_x$	Connected correlations	> 35.0	> 35.0	> 35.0

15 Reproducibility details

The complete analysis pipeline is `analysis/revision_analysis.py`. It regenerates:

- exact-versus-product-formula validation for all observables and cases;
- the fixed-H timestep-convergence study;
- zero-predictor RMSE and skill;
- contraction fits and moving-block residual-bootstrap intervals;
- Fisher-z trajectory correlations and block-length sensitivity;
- reflection-symmetry residuals;
- logical resource counts;
- the manuscript and supplementary figures and LaTeX tables.

Machine-readable outputs are written below `outputs/revision`. Automated tests verify data integrity, time alignment, reconstruction formulas, metric behavior, contraction recovery, block-index validity, and reflection symmetry. The repository also contains the simulation notebook and saved arrays so that every reported offline value can be independently regenerated without IBM credentials.

16 Quantum-information foundations

Quantum mechanics describes physical systems using states. A quantum state contains the information needed to calculate the probabilities of different measurement outcomes. Unlike a classical system, where a quantity may be assigned a definite value before measurement, a quantum system is generally described by probability amplitudes. These amplitudes are complex numbers, and their squared magnitudes give the probabilities of possible measurement results according to the Born rule[1].

For a two-level spin system, a general state can be written as a linear combination of two basis states:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle. \tag{S9}$$

Here, $|\psi\rangle$ represents the state of the system, $|0\rangle$ and $|1\rangle$ are basis states, and α and β are complex probability amplitudes. The probabilities of measuring the system in the states $|0\rangle$ and $|1\rangle$ are $|\alpha|^2$ and $|\beta|^2$, respectively. Since the total probability must be one, $|\alpha|^2 + |\beta|^2 = 1$.

The expression above is an example of superposition. Superposition means that, before measurement, the state is not described as being only one of the basis states. Instead, the state is a combination of the possible basis states, with each component contributing an amplitude to the final measurement probabilities. In the context of a quantum spin system, this is important because a spin does not have to remain fixed as simply "up" or "down" along one axis. The Hamiltonian can generate dynamics in which different spin configurations exist in quantum superposition with others.

The set of possible quantum states is called a Hilbert space. For a chain of N qubits, the Hilbert space has dimension 2^N, which is one reason many-body quantum systems are difficult to simulate even on the most powerful classical computers. In our study, we consider a 15-qubit 1D spin-chain, leading to a Hilbert space with dimension $2^{15} = 32768$.

Physical quantities in quantum mechanics are represented by operators. An operator is a mathematical object that acts on a quantum state and represents either a measurable quantity or a transformation of the state. A measurable physical quantity is called an observable, and its expectation value gives the average result expected from repeated measurements of identically prepared systems. For spin-$1/2$ systems, the relevant local operators are the Pauli operators X, Y, and Z[2]. The operator Z_i measures spin orientation along the z direction at site i, while X_i changes the spin state between the two Z-basis states.

An important feature of quantum mechanics is that some operators are non-commuting, meaning that applying them in different orders can produce different results. The commutator of two operators is defined as

$$[A, B] = AB - BA. \tag{S10}$$

If $[A, B] = 0$, the operators commute; if $[A, B] \neq 0$, the order of application matters. This becomes important in the tilted-field Ising model because the Hamiltonian contains both Z-type and X-type terms. These terms represent competing physical effects, and their non-commuting structure is one reason the model produces nontrivial quantum dynamics.

17 Integrability, chaos, and operator growth

Integrability means that a system has many conserved quantities that restrict its time evolution. A conserved quantity is a physical quantity whose value remains fixed as the system evolves. When a quantum many-body system has many conserved quantities, its dynamics are more constrained and often more regular. In the transverse-field Ising model, integrability is commonly understood through the Jordan-Wigner transformation, which maps the spin degrees of freedom into fermionic variables and allows the ideal model to be solved exactly[3]. Even in an integrable regime, interactions and information spreading can still occur; integrability constrains the dynamics through its conserved structure rather than preventing nontrivial evolution.

A non-integrable system does not have the same extensive set of constraints. In a non-integrable many-body system, the absence of many conserved quantities usually makes the dynamics harder to reduce to a small set of quasiparticles. As the system evolves, local observables can relax because information about the initial state becomes distributed across many degrees of freedom rather than remaining visible in a simple local measurement. This behavior is often

associated with quantum chaos and thermalization, although the full quantum state still evolves unitarily, meaning that information is redistributed[4,5].

The distinction above describes the structure of the ideal Hamiltonian. The observables used in the present benchmark are not level-spacing statistics, out-of-time-order correlators, or entanglement measures, and therefore are not used as direct diagnostics of quantum chaos or thermalization. The $h_z = 0$ and $h_z = h_x$ cases are instead compared through the magnetization and connected-correlation observables measured in this study.

18 Hamiltonian-scale regimes

The dynamics of the tilted-field Ising model also depend on the relative strengths of the interaction and field terms. A useful way to interpret the model is to compare the coupling strength J with the field strengths, collectively denoted by h. Here, h represents the relevant magnetic-field scale, such as the transverse field h_x or the combined effect of the applied fields.

In the field-dominated regime,

$$h \gg J, \tag{S11}$$

the field terms dominate the Hamiltonian. In this case, the dynamics are mainly controlled by single-qubit rotations produced by the magnetic fields. Since the spin-spin interaction is comparatively weak, neighboring spins influence each other less strongly, and correlations are expected to grow more slowly.

In the interaction-dominated regime,

$$J \gg h, \tag{S12}$$

the nearest-neighbor Ising interaction dominates the dynamics. In this regime, the behavior of each spin is strongly affected by the states of its neighboring spins. The dynamics are therefore more collective, because the system cannot be understood mainly as independent spin rotations.

The intermediate regime occurs when the interaction and field scales are comparable,

$$J \approx h. \tag{S13}$$

Here, the interaction and field terms have comparable strength. The interaction term tends to organize neighboring spins in the Z basis, while the field terms lead to spin rotations and fluctuations. Because neither effect is negligible, their competition can produce nontrivial correlation dynamics. This intermediate regime is especially useful for investigating how correlations form when the interaction and field terms compete.

The benchmark values used below all lie near this intermediate scale: $J/h_x = 0.9$, 1.0, and 1.4. They are therefore referred to as lower, matched, and higher coupling cases rather than as the asymptotic limits described above.

19 Quantum circuits and digital simulation

A quantum circuit is a sequence of quantum gates applied to qubits. A qubit, or quantum bit, is the basic unit of quantum information. It is a two-level quantum system that can be described using two basis states, usually written as $|0\rangle$ and $|1\rangle$. Unlike a classical bit, which has a definite value of either 0 or 1, a qubit can exist in a superposition of these basis states before measurement. A quantum gate is a controlled unitary operation that changes the state of one or more qubits[6]. In Hamiltonian simulation, the goal is to construct a circuit whose overall action approximates the time-evolution operator generated by a chosen Hamiltonian[7,8]. This method is called digital quantum simulation because continuous-time quantum evolution is approximated using discrete gate operations.

20 Hardware-noise mechanisms

In the ideal circuit, each gate is treated as an exact unitary operation. On quantum hardware, however, gates are applied through physical control pulses on real qubits, so the implemented operation can differ slightly from the intended unitary operation. This difference is called gate error and is one of the main limitations of noisy intermediate-scale

quantum devices[9]. In this study, two-qubit gate error is especially important because the ZZ layers implement the nearest-neighbor interaction term Z_iZ_{i+1} and create correlations between adjacent qubits. A small over-rotation, under-rotation, or unwanted phase shift in these gates can weaken the measured correlation structure.

The measured observables can also be affected by readout error, finite-shot noise, and decoherence. Readout assignment error occurs when a qubit in state $|0\rangle$ is recorded as 1, or a qubit in state $|1\rangle$ is recorded as 0. This directly affects expectation values such as $\langle Z_i\rangle$ and $\langle Z_iZ_j\rangle$, which are estimated from a finite number of measurements. Finite-shot sampling increases estimator variance but does not by itself imply systematic amplitude contraction. Decoherence means that the quantum state loses information to the surrounding hardware environment before the circuit finishes running: T_1 relaxation changes qubit populations, while T_2 dephasing reduces phase coherence. Crosstalk and residual coupling are additional possible contributions in multi-qubit circuits with repeated two-qubit layers[10]. Because calibration-matched noisy simulations and mechanism-isolating ablations were not archived, these processes are treated as possibilities rather than identified causes of the observed discrepancies.

21 Extended Trotterization derivation

Trotterization is a product-formula method for approximating the exponential of a sum of operators by a product of exponentials of simpler operators[11]. In Hamiltonian simulation, Trotterization is used to approximate continuous time evolution with a sequence of quantum-circuit layers, where each layer implements one part of the Hamiltonian separately. For the tilted-field Ising model used here, the Hamiltonian is separated into three non-commuting parts while preserving the sign convention used in the main-manuscript Hamiltonian:

$$H = H_{ZZ} + H_X + H_Z, \tag{S14}$$

$$H_{ZZ} = -J\sum_{i=1}^{N-1} Z_iZ_{i+1}, \quad H_X = -h_x\sum_{i=1}^{N} X_i, \quad H_Z = -h_z\sum_{i=1}^{N} Z_i, \tag{S15}$$

With this convention, the time-evolution operator for one timestep is

$$U(\Delta t) = e^{-iH\Delta t}. \tag{S16}$$

Because each Hamiltonian component carries a negative sign, the circuit layers implement positive phase evolution under the corresponding Pauli operators. In the IBM/Qiskit gate convention, $R_\alpha(\theta)$ implements a rotation of the form

$$R_\alpha(\theta) = e^{-i\theta\sigma_\alpha/2}. \tag{S17}$$

Therefore, a diagonal half-step uses Qiskit angles $\theta_{ZZ} = -J\Delta t$ and $\theta_{R_Z} = -h_z\Delta t$, while the full transverse step uses $\theta_{R_X} = -2h_x\Delta t$. These signs and factors match the Hamiltonian convention in the main-manuscript Hamiltonian.

Since the tilted-field Ising Hamiltonian is written as a sum of separate parts, one might first expect that the exponential of the full Hamiltonian could be separated into exponentials of each part. If the terms in the Hamiltonian all commuted, this would be exact. However, the interaction, transverse-field, and longitudinal-field terms generally do not commute. This is important for Trotterization because the method replaces one exponential of a sum of Hamiltonian terms with a product of exponentials of the separate terms.

The reason this replacement is only approximate can be seen using the Baker–Campbell–Hausdorff expansion[12]. For two operators A and B, the formula states that

$$e^Ae^B = \exp\left(A + B + \frac{1}{2}[A,B] + \frac{1}{12}[A,[A,B]] + \frac{1}{12}[B,[B,A]] + \cdots\right). \tag{S18}$$

Since the separated Hamiltonian terms do not generally commute, the exact time-evolution operator must be approximated by applying the separate pieces sequentially. In this study, second-order symmetric Trotterization was used:

$$S_2(\Delta t) = e^{-i(H_{ZZ}+H_Z)\Delta t/2}e^{-iH_X\Delta t}e^{-i(H_{ZZ}+H_Z)\Delta t/2}, \tag{S19}$$

which approximates $e^{-iH\Delta t}$ with local error $O(\Delta t^3)$. The H_{ZZ} and H_Z terms are grouped because they commute[13]. This is valid when Δt is small compared with the Hamiltonian scales set by J, h_x, and h_z, so that the neglected commutator and nested-commutator corrections remain small. Repeating this step many times approximates the full time evolution.

Each factor in the Trotter step corresponds to a circuit layer that can be implemented using available gates. The approximation becomes more accurate as the time step is made smaller, but smaller time steps require more circuit layers. This creates a trade-off between accuracy and circuit depth. In an ideal simulator, more layers increase computational cost; on real quantum hardware, more layers also increase exposure to noise and gate imperfections.

22 Expanded observable definitions

For any observable O, the exact value is $\langle O(t)\rangle = \langle\psi(t)|O|\psi(t)\rangle$, where $|\psi(t)\rangle = e^{-iHt}|\psi(0)\rangle$. The simulated observables instead use the Trotterized state defined below.

In this study, however, the circuit simulation uses a Trotterized approximation to the tilted-field Ising Hamiltonian. Therefore, the state used to calculate the simulated observables is

$$|\psi_{\text{Trot}}(t)\rangle = U_{\text{Trot}}(t)|\psi(0)\rangle, \tag{S20}$$

where $U_{\text{Trot}}(t)$ is the Trotterized approximation to the exact time-evolution operator. The difference between $U(t)$ and $U_{\text{Trot}}(t)$ is the source of Trotter error.

The local magnetization at site i is defined as

$$\langle Z_i(t)\rangle_{\text{Trot}} = \langle\psi_{\text{Trot}}(t)|Z_i|\psi_{\text{Trot}}(t)\rangle. \tag{S21}$$

This is the expectation value of Z_i in the Trotterized time-evolved state. It measures the average spin orientation at site i along the z direction. Local magnetization shows how each individual spin is affected by the Hamiltonian dynamics over time.

The total magnetization is defined as

$$M_z(t) = \sum_i \langle Z_i(t)\rangle_{\text{Trot}}. \tag{S22}$$

This is a measure of the spin polarization of the chain. While local magnetization describes individual sites, total magnetization describes the system as a whole. If total magnetization remains close to its initial value, the system retains more global memory of its initial spin imbalance. If it changes significantly, this suggests that the dynamics redistribute the initial spin information through the chain.

The two-point correlation between two sites is

$$\langle Z_i(t)Z_j(t)\rangle_{\text{Trot}} = \langle\psi_{\text{Trot}}(t)|Z_iZ_j|\psi_{\text{Trot}}(t)\rangle. \tag{S23}$$

This quantity measures how the spin orientations at sites i and j are related along the z direction. However, the raw two-point correlation includes both correlations between sites and the separate effects of the individual magnetizations. To remove the independent contribution, the connected correlation function is used:

$$C_{ij}^{\text{conn}}(t) = \langle Z_i(t)Z_j(t)\rangle_{\text{Trot}} - \langle Z_i(t)\rangle_{\text{Trot}}\langle Z_j(t)\rangle_{\text{Trot}}. \tag{S24}$$

If two spins behave independently, then the product $\langle Z_i(t)\rangle_{\text{Trot}}\langle Z_j(t)\rangle_{\text{Trot}}$ accounts for their separate behavior, and the connected correlation is small. If the Hamiltonian produces a relationship between the two spins, the connected correlation becomes nonzero. This makes connected correlations useful for studying how interactions build relationships between different sites in the chain[14].

The calculations include 15 local-magnetization components, one total magnetization derived by summing those same components, and 14 center-to-site connected correlations C_{8j}^{conn} with $j \neq 8$. Comparisons between $h_z = 0$ and $h_z = h_x$ show how these observables change after a longitudinal field is added, but they do not by themselves diagnose

integrability breaking or quantum chaos.

23 Metric definitions and uncertainty detail

Temporal shape was assessed separately for each site or center-to-site pair using Pearson's coefficient. Component coefficients were transformed using Fisher's $z = \operatorname{atanh}(r)$, averaged in z space, and transformed back. The minimum, median, maximum, and central-site value are also reported. Because the 71 time points are serially correlated, circular moving-block resampling with blocks of 4, 7, and 10 samples was used as a sensitivity analysis. These intervals are conditional trajectory diagnostics rather than experimental confidence intervals, because shot-level outcomes and repeated hardware runs were not retained.

The visible amplitude reduction in local magnetization was tested using the empirical contraction model

$$H_i(t) = e^{-\Gamma t} T_i(t) + \epsilon_i(t), \tag{S25}$$

fitted by pooled least squares over all 15 sites and 71 times in each case. The model measures how much of the discrepancy can be represented by scalar contraction, but it is not identified with a unique depolarizing, relaxation, readout, or mitigation rate.

With 2048 shots, the standard error of a directly sampled Pauli expectation P is bounded by

$$\mathrm{SE}(\langle P \rangle) = \sqrt{\frac{1 - \langle P \rangle^2}{2048}} \leq 0.0221. \tag{S26}$$

This bound does not determine the uncertainty of C_{8j}^{conn}, which combines three correlated estimators. The raw bitstrings, estimator variances, and covariance terms required for that calculation were not saved. The previously defined agreement-breakdown time is therefore retained only as a threshold-sensitivity diagnostic in the Supplementary Information rather than as a headline fidelity measure.

24 Reproducibility and archived resources

The ideal simulations and hardware circuit construction were performed using Python 3.13.5, Qiskit 2.3.0, qiskit-aer 0.17.2, qiskit-ibm-runtime 0.45.1, NumPy 2.3.5, SciPy 1.16.3, and Matplotlib 3.10.8. Qiskit Aer was installed but was not used to generate a calibration-matched noisy control. Exact reference trajectories were computed using `scipy.sparse.linalg.expm_multiply`, which evaluates the action of the sparse matrix exponential on the initial state[15]. SciPy does not expose a user-selected tolerance for this routine. Numerical reliability was checked through state-norm preservation and convergence of the reported observables. The exact code used to generate the circuits, run the simulations, process the measured observables, and produce the figures is provided in Appendix A.

Qubit observables were mapped using Qiskit's little-endian qubit convention. In the analysis code, physical spin site i was represented by the zero-based Qiskit qubit index $i - 1$. Pauli strings follow Qiskit's little-endian convention, so an operator acting on zero-based qubit index `site` was inserted at string position `N - 1 - site`. Hardware circuits were sampled every 0.5 units of evolution time from $t = 0$ to $t = 35$, giving 71 time points per parameter case and 426 hardware circuits across the six parameter cases. Each circuit was run with 2048 shots and transpiled with optimization level 0.

Each circuit at each sampled time was executed once. The stored hardware artifact has shape $(6, 71, 29)$: 15 local Z expectations and 14 center-to-site $Z_8 Z_j$ expectations for each case and time. Total magnetization and connected correlations were derived from these retained expectations. Raw bitstrings, Estimator standard errors, achieved precision, covariance terms, complete returned job metadata, job identifiers, historical transpiled circuits, initial and final physical layouts, and path-specific calibration snapshots were not archived. Consequently, experimental confidence intervals, run-to-run and calibration-window variability, physical-layout consistency, native gate counts and two-qubit depth, and a calibration-matched device-specific noise model cannot be reconstructed retrospectively. No claim in this manuscript is therefore tied to a specific physical-qubit path or to a path-specific hardware error.

At $t = 35$, the unmerged logical second-order construction contains 350 steps, 9,800 logical R_{ZZ} gates, 5,250

R_X gates, and 10,500 R_Z gates for $h_z = h_x$. Adjacent diagonal half-layers can be merged algebraically, but actual native counts and depth depend on routing, basis translation, and layout. These logical counts show why the late-time circuits are substantially deeper than the early-time circuits, but the missing transpiled artifacts prevent a defensible device-specific validity boundary from being drawn on the figures.

25 Extended limitations and future-work narrative

For this 15-qubit tilted-field Ising benchmark on `ibm_marrakesh`, magnetization-scale information remained partly visible in the archived hardware execution, while faithful center-to-site connected-correlation dynamics were not established. Exact sparse matrix-exponential-action validation and fixed-H timestep convergence show that the ideal second-order product-formula error is much smaller than the hardware discrepancy. Local and total magnetization outperform an all-zero trajectory in every case, with zero-baseline skill of 15.4–27.8% and 27.0–52.6%, respectively. Four connected-correlation cases do not outperform zero, and the other two improve by only 2.9% and 0.8%. The contraction analysis captures part, but not all, of the local-magnetization amplitude reduction and does not identify a unique physical noise mechanism. These results show that coarse collective observables can remain partially recoverable even when weaker multicomponent correlations are not.

These conclusions are deliberately limited to the archived execution. Although the submission code requested Runtime resilience level 1 and XpXm dynamical decoupling, the realized per-job settings cannot be independently verified from the retained data. Raw bitstrings, covariance information, repeated executions, physical layouts, historical transpiled circuits, and path-specific calibration snapshots are also unavailable. Device-specific error budgets, experimental confidence intervals, calibration drift, layout consistency, native resource counts, and causal mitigation or noise-channel effects therefore cannot be reconstructed retrospectively. Other limitations include the 15-qubit system size, one initial state, one backend, one execution period, and six selected parameter cases. The conclusions are consequently an execution-specific observable benchmark rather than a device-independent characterization of IBM Heron hardware.

Future work could repeat representative circuits across multiple execution periods and IBM backends while fixing the physical-qubit path and transpiler seed. It should retain raw counts, layouts, transpiled circuits, native resource counts, complete Runtime metadata, and calibration snapshots, and include explicit mitigation-off/on, dynamical-decoupling-off/on, and calibration-matched noisy-simulation controls. Smaller spin chains could help separate hardware noise from the intended dynamics. Additional diagnostics such as out-of-time-order correlators, entanglement entropy, and level-spacing statistics would be required to connect the results directly to quantum chaos or operator spreading[16,17].

For a future benchmark on an unfamiliar backend, local magnetization should be measured first because the same measurements also determine total magnetization. Connected correlations should be added only after verifying that their covariance signal exceeds the finite-shot and readout-noise background. Establishing a quantitative native-depth validity boundary would additionally require retaining the physical layout, transpiled circuits, scheduled duration, calibration snapshot, and repeated hardware executions.

Full quantum error correction would address the problem more directly, but it would require multiple physical qubits to encode each logical qubit, together with additional gates and measurements for error detection and correction[18]. For a many-body circuit, this overhead could move the simulation beyond the practical limits of the available device.

26 Code and Data Availability

The analysis code and archived processed data products used for the simulations, circuit construction, IBM quantum-hardware analysis, quantitative benchmarking, and figure generation reported in this manuscript are available in the following GitHub repository:

`https://github.com/samarth-lamba2009/1D-TFIM-CODE.git`

The Supplementary Information and its accompanying supplementary files provide the complete validation results, quantitative tables, diagnostic figures, backend information, and additional reproducibility material supporting the

main manuscript.

References

[1] D. J. Griffiths, D. F. Schroeter. *Introduction to quantum mechanics*. 3rd ed., Cambridge University Press, 2018.

[2] J. J. Sakurai, J. Napolitano. *Modern quantum mechanics*. 2nd ed., Cambridge University Press, 2017.

[3] P. Pfeuty. The one-dimensional Ising model with a transverse field. Annals of Physics. Vol. 57, pg. 79–90, 1970, https://doi.org/10.1016/0003-4916(70)90270-8.

[4] F. Haake, S. Gnutzmann, M. Kuś. *Quantum signatures of chaos*. 4th ed., Springer, 2018, https://doi.org/10.1007/978-3-319-97580-1.

[5] L. D'Alessio, Y. Kafri, A. Polkovnikov, M. Rigol. From quantum chaos and eigenstate thermalization to statistical mechanics and thermodynamics. Advances in Physics. Vol. 65, pg. 239–362, 2016, https://doi.org/10.1080/00018732.2016.1198134.

[6] M. A. Nielsen, I. L. Chuang. *Quantum computation and quantum information*. 10th anniversary ed., Cambridge University Press, 2010.

[7] S. Lloyd. Universal quantum simulators. Science. Vol. 273, pg. 1073–1078, 1996, https://doi.org/10.1126/science.273.5278.1073.

[8] A. M. Childs, D. Maslov, Y. Nam, N. J. Ross, Y. Su. Toward the first quantum simulation with quantum speedup. Proceedings of the National Academy of Sciences. Vol. 115, pg. 9456–9461, 2018, https://doi.org/10.1073/pnas.1801723115.

[9] J. Preskill. Quantum computing in the NISQ era and beyond. Quantum. Vol. 2, pg. 79, 2018, https://doi.org/10.22331/q-2018-08-06-79.

[10] P. Murali, D. C. McKay, M. Martonosi, A. Javadi-Abhari. Software mitigation of crosstalk on noisy intermediate-scale quantum computers. Proceedings of the Twenty-Fifth International Conference on Architectural Support for Programming Languages and Operating Systems. pg. 1001–1016, 2020, https://doi.org/10.1145/3373376.3378477.

[11] H. F. Trotter. On the product of semi-groups of operators. Proceedings of the American Mathematical Society. Vol. 10, pg. 545–551, 1959, https://doi.org/10.2307/2033649.

[12] M. Suzuki. General theory of fractal path integrals with applications to many-body theories and statistical physics. Journal of Mathematical Physics. Vol. 32, pg. 400–407, 1991, https://doi.org/10.1063/1.529425.

[13] A. M. Childs, Y. Su, M. C. Tran, N. Wiebe, S. Zhu. Theory of Trotter error with commutator scaling. Physical Review X. Vol. 11, pg. 011020, 2021, https://doi.org/10.1103/PhysRevX.11.011020.

[14] S. Sachdev. *Quantum phase transitions*. 2nd ed., Cambridge University Press, 2011.

[15] A. H. Al-Mohy, N. J. Higham. Computing the action of the matrix exponential, with an application to exponential integrators. SIAM Journal on Scientific Computing. Vol. 33, pg. 488–511, 2011, https://doi.org/10.1137/100788860.

[16] A. Nahum, S. Vijay, J. Haah. Operator spreading in random unitary circuits. Physical Review X. Vol. 8, pg. 021014, 2018, https://doi.org/10.1103/PhysRevX.8.021014.

[17] J. Maldacena, S. H. Shenker, D. Stanford. A bound on chaos. Journal of High Energy Physics. Vol. 2016, pg. 106, 2016, https://doi.org/10.1007/JHEP08(2016)106.

[18] A. G. Fowler, M. Mariantoni, J. M. Martinis, A. N. Cleland. Surface codes: Towards practical large-scale quantum computation. Physical Review A. Vol. 86, pg. 032324, 2012, https://doi.org/10.1103/PhysRevA.86.032324.