

# Quantum computation of molecular geometry via many-body nuclear spin echoes

C. Zhang<sup>#</sup>,<sup>1</sup> R. G. Cortiñas<sup>#</sup>,<sup>2</sup> A. H. Karamlou<sup>#</sup>,<sup>2</sup> N. Noll,<sup>2</sup> J. Provazza,<sup>3</sup> J. Bausch,<sup>4</sup> S. Shirobokov,<sup>4</sup> A. White,<sup>2,3</sup> M. Claassen,<sup>2,5</sup> S. H. Kang,<sup>1</sup> A. W. Senior,<sup>4</sup> N. Tomašev,<sup>4</sup> J. Gross,<sup>2</sup> K. Lee,<sup>2</sup> T. Schuster,<sup>2,6</sup> W. J. Huggins,<sup>2</sup> H. Celik,<sup>1</sup> A. Greene,<sup>2</sup> B. Kozlovskii,<sup>4</sup> F. J. H. Heras,<sup>4</sup> A. Bengtsson,<sup>2</sup> A. Grajales Dau,<sup>2</sup> I. Drozdov,<sup>2,7</sup> B. Ying,<sup>2</sup> W. Livingstone,<sup>2</sup> V. Sivak,<sup>2</sup> N. Yosri,<sup>2</sup> C. Quintana,<sup>2</sup> D. Abanin,<sup>2</sup> A. Abbas,<sup>2</sup> R. Acharya,<sup>2</sup> L. Aghababaie Beni,<sup>2</sup> G. Aigeldinger,<sup>2</sup> R. Alcaraz,<sup>2</sup> S. Alcaraz,<sup>2</sup> T. I. Andersen,<sup>2</sup> M. Ansmann,<sup>2</sup> F. Arute,<sup>2</sup> K. Arya,<sup>2</sup> W. Askew,<sup>2</sup> N. Astrakhantsev,<sup>2</sup> J. Atalaya,<sup>2</sup> B. Ballard,<sup>2</sup> J. C. Bardin,<sup>2,8</sup> H. Bates,<sup>2</sup> M. Bigdeli Karimi,<sup>2</sup> A. Bilmes,<sup>2</sup> S. Bilodeau,<sup>2</sup> F. Borjans,<sup>2</sup> A. Bourassa,<sup>2</sup> J. Bovaird,<sup>2</sup> D. Bowers,<sup>2</sup> L. Brill,<sup>2</sup> P. Brooks,<sup>2</sup> M. Broughton,<sup>2</sup> D. A. Browne,<sup>2</sup> B. Buchea,<sup>2</sup> B. B. Buckley,<sup>2</sup> T. Burger,<sup>2</sup> B. Burkett,<sup>2</sup> J. Busnaina,<sup>2</sup> N. Bushnell,<sup>2</sup> A. Cabrera,<sup>2</sup> J. Campero,<sup>2</sup> H.-S. Chang,<sup>2</sup> S. Chen,<sup>2</sup> Z. Chen,<sup>2</sup> B. Chiaro,<sup>2</sup> L.-Y. Chih,<sup>2</sup> A. Y. Cleland,<sup>2</sup> B. Cochrane,<sup>2</sup> M. Cockrell,<sup>2</sup> J. Cogan,<sup>2</sup> R. Collins,<sup>2</sup> P. Conner,<sup>2</sup> H. Cook,<sup>2</sup> W. Courtney,<sup>2</sup> A. L. Crook,<sup>2</sup> B. Curtin,<sup>2</sup> S. Das,<sup>2</sup> M. Damyanov,<sup>2</sup> D. M. Debroy,<sup>2</sup> L. De Lorenzo,<sup>2</sup> S. Demura,<sup>2</sup> L. B. De Rose,<sup>2</sup> A. Di Paolo,<sup>2</sup> P. Donohoe,<sup>2</sup> A. Dunsworth,<sup>2</sup> V. Ehimhen,<sup>2</sup> A. Eickbusch,<sup>2</sup> A. M. Elbag,<sup>2</sup> L. Ella,<sup>2</sup> M. Elzouka,<sup>2</sup> D. Enriquez,<sup>2</sup> C. Erickson,<sup>2</sup> V. S. Ferreira,<sup>2</sup> M. Flores,<sup>2</sup> L. Flores Burgos,<sup>2</sup> E. Forati,<sup>2</sup> J. Ford,<sup>2</sup> A. G. Fowler,<sup>2</sup> B. Foxen,<sup>2</sup> M. Fukami,<sup>2</sup> A. W. L. Fung,<sup>2</sup> L. Fuste,<sup>2</sup> S. Ganjam,<sup>2</sup> G. Garcia,<sup>2</sup> C. Garrick,<sup>2</sup> R. Gasca,<sup>2</sup> H. Gehring,<sup>2</sup> R. Geiger,<sup>2</sup> É. Genois,<sup>2</sup> W. Giang,<sup>2</sup> C. Gidney,<sup>2</sup> D. Gilboa,<sup>2</sup> J. E. Goeders,<sup>2</sup> E. C. Gonzales,<sup>2</sup> R. Gosula,<sup>2</sup> S. J. de Graaf,<sup>2</sup> D. Graumann,<sup>2</sup> J. Grebel,<sup>2</sup> J. Guerrero,<sup>2</sup> J. D. Guimaraes,<sup>2</sup> T. Ha,<sup>2</sup> S. Habegger,<sup>2</sup> T. Hadick,<sup>2</sup> A. Hadjikhani,<sup>2</sup> M. P. Harrigan,<sup>2</sup> S. D. Harrington,<sup>2</sup> J. Hartshorn,<sup>2</sup> S. Heslin,<sup>2</sup> P. Heu,<sup>2</sup> O. Higgott,<sup>2</sup> R. Hiltermann,<sup>2</sup> J. Hilton,<sup>2</sup> H.-Y. Huang,<sup>2</sup> M. Hucka,<sup>2</sup> C. Hudspeth,<sup>2</sup> A. Huff,<sup>2</sup> E. Jeffrey,<sup>2</sup> S. Jevons,<sup>2</sup> Z. Jiang,<sup>2</sup> X. Jin,<sup>2</sup> C. Joshi,<sup>2</sup> P. Juhas,<sup>2</sup> A. Kabel,<sup>2</sup> H. Kang,<sup>2</sup> K. Kang,<sup>2</sup> R. Kaufman,<sup>2</sup> K. Kechedzhi,<sup>2</sup> T. Khattar,<sup>2</sup> M. Khezri,<sup>2</sup> S. Kim,<sup>2</sup> R. King,<sup>2,9</sup> O. Kiss,<sup>2</sup> P. V. Klimov,<sup>2</sup> C. M. Knaut,<sup>2</sup> B. Kobrin,<sup>2</sup> F. Kostritsa,<sup>2</sup> J. M. Kreikebaum,<sup>2</sup> R. Kudo,<sup>2</sup> B. Kueffler,<sup>2</sup> A. Kumar,<sup>2</sup> V. D. Kurilovich,<sup>2</sup> V. Kutsko,<sup>2</sup> N. Lacroix,<sup>2</sup> D. Landhuis,<sup>2</sup> T. Lange-Dei,<sup>2</sup> B. W. Langley,<sup>2</sup> P. Laptev,<sup>2</sup> K.-M. Lau,<sup>2</sup> L. Le Guevel,<sup>2</sup> J. Ledford,<sup>2</sup> J. Lee,<sup>2</sup> B. J. Lester,<sup>2</sup> W. Leung,<sup>2</sup> L. Li,<sup>2</sup> W. Y. Li,<sup>2</sup> M. Li,<sup>2</sup> A. T. Lill,<sup>2</sup> M. T. Lloyd,<sup>2</sup> A. Locharla,<sup>2</sup> D. Lundahl,<sup>2</sup> A. Lunt,<sup>2</sup> S. Madhuk,<sup>2</sup> A. Maiti,<sup>2</sup> A. Maloney,<sup>2</sup> S. Mandra,<sup>2</sup> L. S. Martin,<sup>2</sup> O. Martin,<sup>2</sup> E. Mascot,<sup>2</sup> P. Masih Das,<sup>2</sup> D. Maslov,<sup>2</sup> M. Mathews,<sup>2</sup> C. Maxfield,<sup>2</sup> J. R. McClean,<sup>2</sup> M. McEwen,<sup>2</sup> S. Meeks,<sup>2</sup> K. C. Miao,<sup>2</sup> R. Molavi,<sup>2</sup> S. Molina,<sup>2</sup> S. Montazeri,<sup>2</sup> C. Neill,<sup>2</sup> M. Newman,<sup>2</sup> A. Nguyen,<sup>2</sup> M. Nguyen,<sup>2</sup> C.-H. Ni,<sup>2</sup> M. Y. Ni,<sup>2</sup> L. Oas,<sup>2</sup> R. Orosco,<sup>2</sup> K. Ottosson,<sup>2</sup> A. Pagano,<sup>2</sup> S. Peek,<sup>2</sup> D. Peterson,<sup>2</sup> A. Pizzuto,<sup>2</sup> E. Portoles,<sup>2</sup> R. Potter,<sup>2</sup> O. Pritchard,<sup>2</sup> M. Qian,<sup>2</sup> A. Ranadive,<sup>2</sup> M. J. Reagor,<sup>2</sup> R. Resnick,<sup>2</sup> D. M. Rhodes,<sup>2</sup> D. Riley,<sup>2</sup> G. Roberts,<sup>2</sup> R. Rodriguez,<sup>2</sup> E. Ropes,<sup>2</sup> E. Rosenberg,<sup>2</sup> E. Rosenfeld,<sup>2</sup> D. Rosenstock,<sup>2</sup> E. Rossi,<sup>2</sup> D. A. Rower,<sup>2</sup> M. S. Rudolph,<sup>2</sup> R. Salazar,<sup>2</sup> K. Sankaragomathi,<sup>2</sup> M. C. Saruhan,<sup>2</sup> K. J. Satzinger,<sup>2</sup> M. Schaefer,<sup>2,10</sup> S. Schroeder,<sup>2</sup> H. F. Schurkus,<sup>2</sup> A. Shahingohar,<sup>2</sup> M. J. Shearn,<sup>2</sup> A. Shorter,<sup>2</sup> N. Shutty,<sup>2</sup> V. Shvarts,<sup>2</sup> S. Small,<sup>2</sup> W. C. Smith,<sup>2</sup> D. A. Sobel,<sup>2</sup> R. D. Somma,<sup>2</sup> B. Spells,<sup>2</sup> S. Springer,<sup>2</sup> G. Sterling,<sup>2</sup> J. Suchard,<sup>2</sup> A. Szasz,<sup>2</sup> A. Sztein,<sup>2</sup> M. Taylor,<sup>2</sup> J. P. Thiruraman,<sup>2</sup> D. Thor,<sup>2</sup> D. Timucin,<sup>2</sup> E. Tomita,<sup>2</sup> A. Torres,<sup>2</sup> M. M. Torunbalci,<sup>2</sup> H. Tran,<sup>2</sup> A. Vaishnav,<sup>2</sup> J. Vargas,<sup>2</sup> S. Vdovichev,<sup>2</sup> G. Vidal,<sup>2</sup> C. Vollgraff Heidweiller,<sup>2</sup> M. Voorhees,<sup>2</sup> S. Waltman,<sup>2</sup> J. Waltz,<sup>2</sup> S. X. Wang,<sup>2</sup> B. Ware,<sup>2</sup> J. D. Watson,<sup>2</sup> Y. Wei,<sup>2</sup> T. Weidel,<sup>2</sup> T. White,<sup>2</sup> K. Wong,<sup>2</sup> B. W. K. Woo,<sup>2</sup> C. J. Wood,<sup>2</sup> M. Woodson,<sup>2</sup> C. Xing,<sup>2</sup> Z. J. Yao,<sup>2</sup> P. Yeh,<sup>2</sup> J. Yoo,<sup>2</sup> E. Young,<sup>2</sup> G. Young,<sup>2</sup> A. Zalcman,<sup>2</sup> R. Zhang,<sup>2</sup> Y. Zhang,<sup>2</sup> N. Zhu,<sup>2</sup> N. Zobrist,<sup>2</sup> Z. Zou,<sup>2</sup> G. Bortoli,<sup>2</sup> S. Boixo,<sup>2</sup> J. Chen,<sup>2</sup> Y. Chen,<sup>2</sup> M. Devoret,<sup>2,10</sup> M. Hansen,<sup>2</sup> C. Jones,<sup>2</sup> J. Kelly,<sup>2</sup> P. Kohli,<sup>4</sup> A. Korotkov,<sup>2</sup> E. Lucero,<sup>2</sup> J. Manyika,<sup>11</sup> Y. Matias,<sup>11</sup> A. Megrant,<sup>2</sup> H. Neven,<sup>2</sup> W. D. Oliver,<sup>2</sup> G. Ramachandran,<sup>2</sup> R. Babbush,<sup>2</sup> V. Smelyanskiy,<sup>2</sup> P. Roushan,<sup>2</sup> D. Kafri,<sup>2</sup> R. Sarpong,<sup>1</sup> D. W. Berry,<sup>12</sup> C. Ramanathan,<sup>13</sup> X. Mi,<sup>2,\*</sup> C. Bengs,<sup>1,14,15,†</sup> A. Ajoy,<sup>1,14,16,‡</sup> Z. K. Minev,<sup>2,§</sup> N. C. Rubin,<sup>2,¶</sup> and T. E. O'Brien<sup>2,\*\*</sup>

<sup>1</sup>Department of Chemistry, University of California, Berkeley, Berkeley, CA, USA

<sup>2</sup>Google Quantum AI, Santa Barbara, CA, USA

<sup>3</sup>Quantum Simulation Technologies Inc., Boston, MA, USA

<sup>4</sup>Google Deepmind, London, United Kingdom

<sup>5</sup>Department of Physics and Astronomy, University of Pennsylvania, PA, USA

<sup>6</sup>California Institute of Technology, Pasadena, CA, USA

<sup>7</sup>Department of Physics, University of Connecticut, Storrs, CT, USA

<sup>8</sup>USA Department of Electrical and Computer Engineering,

University of Massachusetts, Amherst, MA, USA

<sup>9</sup>Simons Institute for the Theory of Computing, Berkeley, CA, USA

<sup>10</sup>Department of Physics, University of California, Santa Barbara, CA, USA

<sup>11</sup>Google Research, Mountain View, CA, USA

<sup>12</sup>School of Mathematical and Physical Sciences, Macquarie University, Sydney, Australia

<sup>13</sup>Department of Physics and Astronomy, Dartmouth College, Hanover, NH, USA

<sup>14</sup>Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

<sup>15</sup>School of Chemistry, University of Southampton, Southampton, UK

<sup>16</sup>CIFAR Azrieli Global Scholars Program, Toronto, ON, Canada

Quantum-information-inspired experiments in nuclear magnetic resonance spectroscopy may yield a pathway towards determining molecular structure and properties that are otherwise challenging to learn. We measure out-of-time-ordered correlators (OTOCs) [1–4] on two organic molecules suspended in a nematic liquid crystal, and investigate the utility of this data in performing structural learning tasks. We use OTOC measurements to augment molecular dynamics models, and to correct for known approximations in the underlying force fields. We demonstrate the utility of OTOCs in these models by estimating the mean ortho-meta H-H distance of toluene and the mean dihedral angle of 3',5'-dimethylbiphenyl, achieving similar accuracy and precision to independent spectroscopic measurements of both quantities. To ameliorate the apparent exponential classical cost of interpreting the above OTOC data, we simulate the molecular OTOCs on a Willow superconducting quantum processor, using AlphaEvolve-optimized [5] quantum circuits and arbitrary-angle fermionic simulation gates. We implement novel zero-noise extrapolation techniques based on the Pauli pathing model of operator dynamics [4], to repeat the learning experiments with root-mean-square error 0.05 over all circuits used. Our work highlights a computational protocol to interpret many-body echoes from nuclear magnetic systems using low resource quantum computation.

Accurate molecular structure determination is key to probing structure-function relationships in many areas of chemistry and biology. Nuclear Magnetic Resonance (NMR) spectroscopy provides access to structurally rich information embedded within couplings between pairs of spins [6]. Dipolar couplings in solid-like systems provide the most direct access to geometric information. However, the complex spin dynamics generated by these terms increases the challenge of observing [7] and interpreting [8, 9] an information-containing signal. Current techniques to overcome these challenges [10–14] focus on reducing the effective system size to one or a few spins, where signals remain large and interpretable. This has been pivotal in structural elucidation efforts, ranging from model systems in solid state [15, 16] and liquid crystal [17, 18] to complex, biologically relevant targets such as amyloid fibrils [19, 20] and SARS-CoV-2 [21]. A methodological gap remains however, as these techniques limit the maximum measurable distance between pairs of spins (e.g. C-C distances are limited to around  $\sim 6\text{\AA}$ . [22]), and longer-range distance constraints remain more challenging to estimate.

Efficient simulation of many-body spin dynamics, such as in NMR [23–29], has been suggested as an application for quantum computers. For quantum applications, one targets observables that do not concentrate [30–32] and that are sensitive to microscopic system details [4]. In this work, we propose that this sensitivity is relevant to the NMR practitioner, as quantum-information-inspired experiments could provide access to long range structural information. Recent efforts have demonstrated quantum advantage in the estimation of out-of-time-ordered correlators (OTOCs) [1, 2, 33] on superconducting quantum hardware [4]. The OTOC experiment is based on a many-body echo, in which polarization initially localized on a

target spin migrates through the spin network, before a Hamiltonian-engineered time-reversal refocuses to the initial state. This refocusing is sensitive to perturbations on distant butterfly spins, which allows one to measure the extent of polarization propagation through the spin network [Fig. 1(a-b)]. We suggest this may help to fill the aforementioned NMR methodological gap.

Herein, we demonstrate a pathway towards determining otherwise challenging-to-compute molecular structure and properties via a hybrid technique of NMR spectroscopy and digital quantum simulation [Fig. 1(c)]. We measure OTOCs generated by Hamiltonian-engineered pulse sequences on  $^{13}\text{C}$ -labeled organic molecules, [ $4\text{-}^{13}\text{C}$ ]-toluene and [ $1\text{-}^{13}\text{C}$ ]-3',5'-dimethylbiphenyl (DMBP), which we suspend in liquid crystal solvents to suppress intermolecular couplings while partially retaining intramolecular through-space dipolar terms. We compare experimentally obtained OTOCs from toluene to classical simulations using reference data, and illustrate the sensitivity of OTOCs to molecular structure by simulating stretching the molecule between the ortho and meta carbon atoms, yielding an interpretable NMR signal. To alleviate the exponential-scaling simulation cost of this approach, we repeat the above analysis on superconducting quantum hardware. We compile a Trotterized approximation of the all-to-all dipolar Hamiltonian evolution to a swap network of arbitrary-angle fermionic simulation (fSim) gates, and develop novel error mitigation techniques based on the Pauli-path picture of operator dynamics to recreate the learning experiment with high accuracy. Next, we demonstrate a DMBP structure learning protocol by combining NMR OTOC data with a realistic model obtained through classical molecular dynamics to estimate the distribution of the dihedral angle between the two phenyl rings. We use the AlphaEvolve coding agent to optimize product formula generation, and implement the resulting circuits on the quantum device. We compare the resulting estimates of the dihedral angle distribution against multiple quantum coherence (MQC) spectroscopy data from an independent DMBP sample with deuterated methyl groups, experimentally validating the OTOC learning protocol through a secondary

---

\* mixiao@google.com

† c.bengs@soton.ac.uk

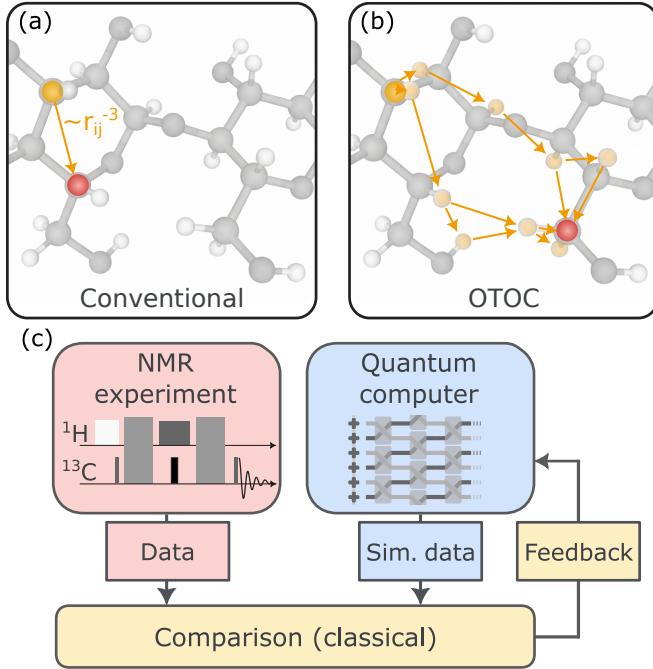
‡ ashokaj@berkeley.edu

§ zminev@google.com

¶ nickrubin@google.com

\*\* teobrien@google.com

# These authors contributed equally to this work.



**FIG. 1. Making a longer molecular ruler with an out-of-time-ordered correlator (OTOC).** a,b) A comparison between conventional spin transport measurements that infer distance restraints from single couplings, and OTOC measurements, which probe the growth of large quantum coherences through the H spin network. By utilizing all the couplings in the spin network, the OTOC is not limited in distance by the  $1/r^3$  scaling that limits the distances measurable by conventional techniques. c) Our proposal to use a quantum computer to assist in processing OTOC (or other challenging-to-classically-simulate) data from a large spin cluster. Following nuclear magnetic resonance (NMR) data collection, the quantum computer provides an artificial system that is iteratively tuned—via classical feedback—until it matches experiment.

unscalable spectroscopic approach. The combination of careful hardware error-mitigation, algorithmic compilation strategies, and physical chemistry model construction demonstrates the potential of scalable quantum-information-inspired experiments to augment traditional NMR data used for structure determination.

## RESULTS

We illustrate in Fig. 2(a-b) the encoding of long-range correlations in an NMR experiment using a sample of [4- $^{13}\text{C}$ ]-toluene in N-(4-Ethoxybenzylidene)-4-butylaniline (EBBA) liquid crystal as a benchmark system with externally determinable molecular structure. We perform experiments using the newly-developed Time-Accurate Reversal of Dipolar InteractionS (TARDIS) pulse scheme, that engineers an effective double-quantum Hamiltonian to propagate information between the  $x$ -polarized  $^{13}\text{C}$  “measurement” spin

and the methyl proton “butterfly” spins. Due to symmetries in the double-quantum Hamiltonian, the TARDIS sequence is approximately invertible, allowing us to refocus the propagated information back to the  $^{13}\text{C}$  spin, which is reflected in the amplitude of the free induction decay (FID). In the absence of any perturbation between forward and backward evolution this sequence performs a Loschmidt Echo (LE) on the system [Fig. 2(c), yellow]. In principle this LE sequence should yield a constant signal, corresponding to perfect refocusing. The observed decay is partially due to higher order Hamiltonian terms in the static effective expansion that are imperfectly refocused, and partially due to inhomogeneities in the applied radio frequency (RF) field across the NMR sample (see Supplementary Information for detailed modeling).

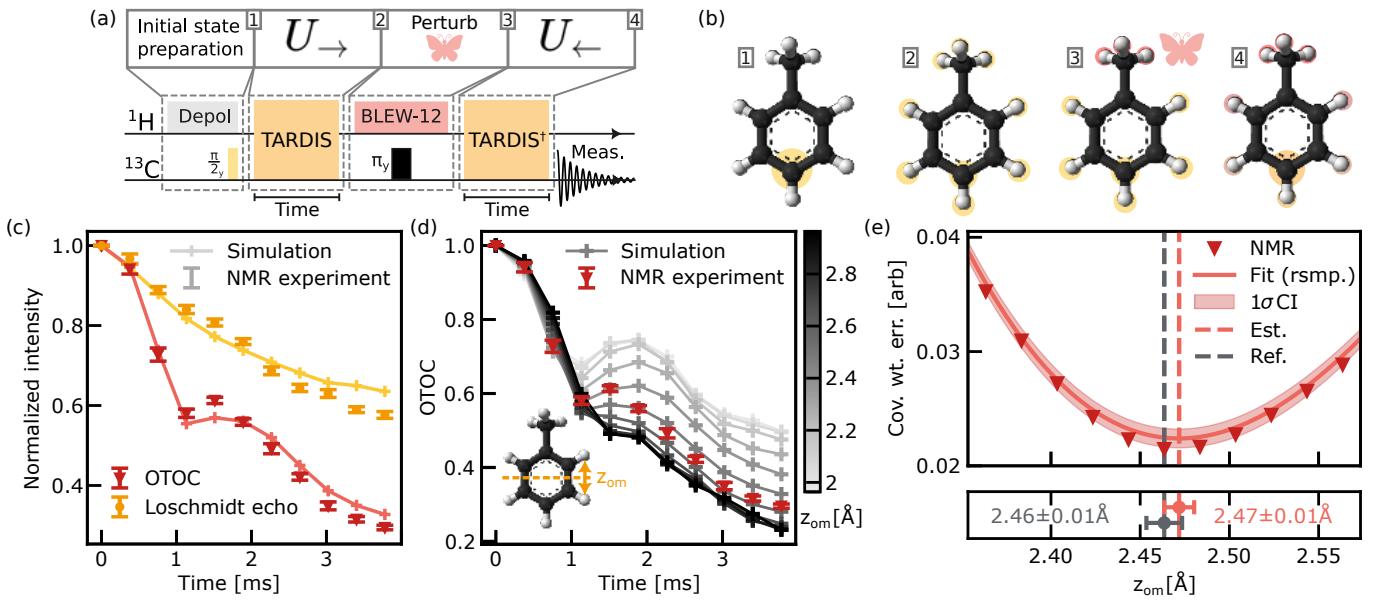
To determine the information spread through the toluene molecule during forward evolution, we apply a “butterfly” unitary operation on the methyl group before refocusing. This yields an OTOC [Fig. 2(c), red], which in the absence of error takes the functional form

$$C(t) = \text{Trace}[X_{13\text{C}}(t) B X_{13\text{C}}(t) B^\dagger], \quad (1)$$

where  $X_{13\text{C}}(t) = U(t)X_{13\text{C}}U^\dagger(t)$  is the forward-evolved measurement operator. At short times the  $^{13}\text{C}$  spin remains undisturbed by the distant butterfly; the OTOC matches the LE. However, as the spin spreads across the system the butterfly effect grows, and the OTOC decays faster than the LE. This decay is characteristic of all OTOCs: the information from the measurement spin propagates with a diffusive front, and causes decay upon reaching the butterfly.

The onset and rate of the OTOC decay depends on the dipolar couplings, enabling determination of a model molecular structure corresponding to the thermal average of all molecular configurations. Due to the simplicity and symmetry of toluene, this average structure can be well approximated by a rigid phenyl ring and a freely-rotating methyl group [34]. The liquid crystal environment can be accounted for by an orientational order parameter. We model OTOCs generated by the exact TARDIS pulse sequence, and include independently-measured RF inhomogeneity. Then, for illustrative purposes, we can simulate an artificial stretching of the molecule between the ortho- and meta- positions [Fig. 2(d), inset], measured by the ortho-meta H-H distance  $z_{om}$ . This affects the OTOC decay strongly [Fig. 2(d)]: we observe that a stretch or contraction of 0.5 Å shifts the OTOC by up to 20%. The direction of this shift is counterintuitive: decreasing  $z_{om}$  (which increases coupling strengths) slows the OTOC at later times. We explain this observation in the Supplementary Information.

To turn the OTOC sensitivity into a tool for learning molecular structure, we construct a cost function to minimize over our target parameter  $z_{om}$ . In Fig. 2(e) we plot the covariance-weighted error to the data from Fig. 2(c) and 4 other datasets generated by engineering an on-site field (see Supplementary Information). Performing a cubic fit and bootstrapping error bars, we obtain an



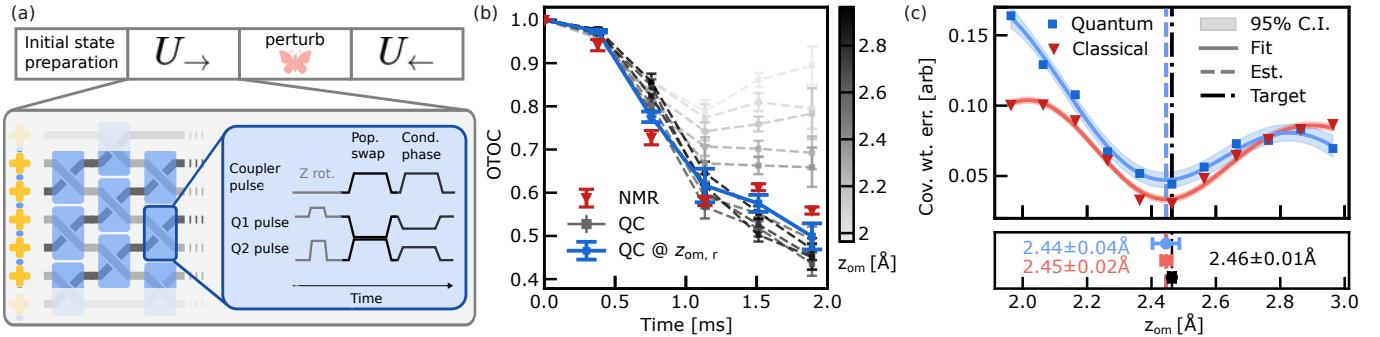
**FIG. 2. Benchmarking the structural sensitivity of out-of-time-ordered correlators (OTOCs) in  $[4-^{13}\text{C}]\text{-toluene}$ .** a) Description of the OTOC protocol (top) as implemented in a nuclear magnetic resonance (NMR) spectrometer (bottom). b) Cartoon showing the spread of the spin cluster through the molecule following the sequence in a). c) OTOC (red) and Loschmidt echo data (yellow) from the NMR experiment. Points show NMR experiment data with  $1\sigma$  confidence intervals (CI), compared to numerical simulations (lines). d) Sensitivity of the OTOC experiment to an artificial stretch of benzene between the ortho and meta carbon atoms (inset), with experimental data from b) overlaid (red). e) Simulated learning of the ortho-meta H-H distance  $z_{\text{om}}$  from OTOC data. (top) Red triangles show the covariance weighted mean error between simulation and NMR experiment, line shows a cubic fit with  $1\sigma$  CI shaded. (bottom) Estimate of  $z_{\text{om}}$  (red) compared to reference data from the literature (gray), with bootstrapped  $2\sigma$  CI.

estimate  $z_{\text{om},c} = 2.47 \pm 0.01 \text{ \AA}$ . This agrees with the reference value  $z_{\text{om},r} = 2.46 \pm 0.01 \text{ \AA}$  [34] to within error bars, and shows comparable precision to this reference data. We caution however that this fits only a single parameter of the entire toluene molecule, whilst Ref. [34] fits the entire molecular structure.

The classical hardness of OTOC simulation prevents scaling our structure learning protocol to larger systems of chemical and biological interest using classical post-processing alone. We ameliorate this issue by efficiently simulating the strongly correlated dipolar coupling Hamiltonian on superconducting quantum hardware, implementing the scheme envisaged in Fig. 1(c). Compared to established applications in quantum simulation of electronic structure, OTOC simulation is a relatively low-cost application for quantum computers due to short scrambling timescales, low measurement complexity, and large error tolerances. However, the all-to-all-coupled dipolar spin Hamiltonian is challenging to simulate compared to local spin models, due to the large number of terms. To solve this problem, we implement a first-order Trotterization [35] of the lowest-order Magnus expansion of the full TARDIS sequence, working in the interaction picture [36] on the  $^{13}\text{C}$ -para-H coupling, and neglecting all other C-H couplings. We compile the H-H interactions into a swap network [37], and implement the resulting fSim gate in an individually calibrated two-pulse scheme [Fig. 3(a)]. The resulting circuits use up to

1080 2-qubit pulses to simulate the first six time steps of the toluene OTOC curve. We further develop a physically motivated zero-noise extrapolation scheme [38, 39] based on the effect of noise in the Pauli-path OTOC picture [4, 26].

In Fig. 3, we repeat the sensitivity experiment across the range of  $z_{\text{om}}$  shown in Fig. 2(c), but replacing the classical computation with quantum simulation performed on a Willow device. We observe that the quantum device is able to qualitatively replicate the OTOC decay across the  $z_{\text{om}}$  range considered. Comparing our hardware to the classical simulation of Fig. 2(c), we find a root-mean-square error between the quantum and classical datasets of  $\epsilon_{\text{tot}} = 0.058$ . This is made up of algorithmic error (i.e. Trotter error) in the quantum circuit ( $\epsilon_{\text{alg}} = 0.035$ ) and residual experimental bias after mitigation ( $\epsilon_{\text{exp}} = 0.050$ ). In the Supplementary Information, we provide detailed budgeting of the algorithmic, NMR, and quantum computer error sources. Motivated by the accuracy in replicating the OTOC decay on quantum hardware, we reproduce the covariance-weighted error curve [Fig. 3(c), blue], and re-estimate the ortho-meta C-C bond length. We observe that the dip remains, and obtain an estimate  $z_{\text{om},q} = 2.44 \pm 0.04 \text{ \AA}$ , which suffers a small loss of precision, but continues to agree with the reference data within experimental error. To separate the effect of the reduced dataset from the algorithmic and experimental error, we repeat this plot using the ac-



**FIG. 3. Alleviating the exponential cost of out-of-time-ordered correlator (OTOC) simulation with a quantum computer.** a) Approximating nuclear spin evolution on superconducting quantum hardware. A Trotterized digital quantum simulation of the double quantum Hamiltonian is executed by compiling individual couplings with swap gates that permute spin indices. The compiled gate is executed using a pulse train divided into a Z rotation, partial population swap, and conditional phase gate. b) Error-mitigated simulations of the first five points of the OTOC curve on quantum hardware, swept over the same range of ortho-meta C-C bond lengths  $z_{\text{om}}$  as in Fig. 2(c). The target simulation at  $z_{\text{om},r} = 2.46 \text{ \AA}$  is emphasized (solid blue line), error bars are  $1\sigma$  confidence intervals (CI). c) (upper) Comparison of the learning experiment using quantum and classical data. Fits (lines) are from a bootstrapped Gaussian process regression, with  $1\sigma$  CI shaded. (lower) Comparison of two estimates of  $z_{\text{om}}$  to reference data, with bootstrapped  $2\sigma$  CI.

curate classical simulation from Fig. 2, but using only the first 5 points of the OTOC [Fig. 3(c), red], achieving an estimate  $z_{\text{om},c} = 2.45 \pm 0.02 \text{ \AA}$ . The loss of precision in the quantum calculation thus comes partially from the smaller dataset used, and partially from a mix of algorithmic and experimental error.

To demonstrate an OTOC based learning protocol on degrees of freedom common to chemical systems, we perform a learning experiment focusing on the biphenyl dihedral angle  $\phi$  of DMBP dissolved in 4-Cyano-4'-pentylbiphenyl (5CB). The energy barrier to this rotation is sufficient to prevent completely free spinning, but the barrier is insufficient to make the structure rigid at experimental temperatures and pressure, so we consider an ensemble of  $\phi$  angles. This distribution is not precisely captured by molecular dynamics with an approximate molecular mechanics force-field (MD/MM), nor with high-level vacuum electronic structure simulations. However, of the intramolecular degrees of freedom of the DMBP molecule, the phenyl and methyl bond distances are too rigid to be significantly affected by the interaction with the liquid crystal environment, whilst the methyl torsions are too unconstrained to be affected by this environment. We can leverage this by capturing the rigid rotations using MD/MM, and re-parameterizing our model such that only the dihedral angle contributions remain.

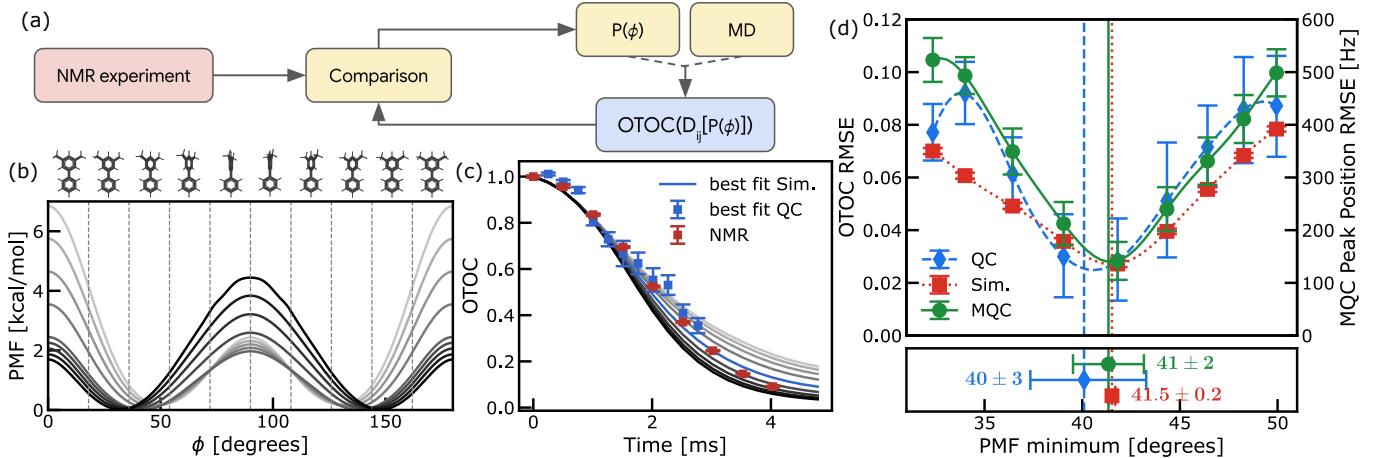
Our task is thus to learn the free energy surface, or potential of mean-force  $\text{PMF}(\phi)$  [Fig. 4(b)], along the dihedral angle  $\phi$ . This yields the dipolar couplings via Boltzmann's distribution:  $D_{ij} = \int d\phi \exp[-\beta \text{PMF}(\phi)] D_{ij}(\phi) / \int d\phi \exp[-\beta \text{PMF}(\phi)]$ , with  $\beta$  the inverse temperature. We capture the  $\phi$ -dependent couplings  $D_{ij}(\phi)$  by binning and averaging MD trajectories as separated by this collective variable. Then, we take a simple sinusoidal model with a few free parameters, and interpolate between estimates from

MD/MM, an artificial shifted double well potential, and a vacuum phase DFT torsion scan. This yielded 9 candidate PMFs (Fig. 4(b)), of which the MD/MM estimate corresponds to a PMF with minimum at  $\phi = 32.4^\circ$ , the DFT estimate corresponds to a PMF with minimum at  $\phi = 41.8^\circ$ , and the artificial double well with a corresponding PMF minimum at  $\phi = 50^\circ$ .

The learning task can be accomplished by simulating the OTOCs for these candidate PMFs on a Willow quantum chip and selecting the candidate PMF that minimizes the difference between the simulated and experimentally measured OTOC curves. To generate quantum circuits that match the numerical verification of TARDIS, we use exponentially-costing exact simulation and the AlphaEvolve coding agent to evolve a first-order Trotter formula generator and produce a novel product formula algorithm. This achieves a low mean error (0.0084) to the exact OTOC simulation across the entire PMF range, using no more than 792 CZ gates.

We show the learning experiment on the quantum device in Fig. 4(d), using the first 2.5 ms of NMR data for the quantum simulation (blue, QC), and the entire 4 ms decay curve for a comparative classical simulation (red, sim.) that includes accurate modeling of the RF inhomogeneity. The quantum computer achieves an RMSE of 0.05 to exact simulation of the input circuits across the entire range of data, matching the performance of the toluene example. The minima of the two cost functions agree within error bars,  $\phi_{\text{QC}} = 40^\circ \pm 3^\circ$ , whilst  $\phi_{\text{sim}} = 41.5^\circ \pm 0.2^\circ$ . In Fig. 4(c) we plot the OTOC for the QC simulation of the nearest PMF, with NMR experimental data and classical simulations overlaid.

To confirm the OTOC prediction of the dihedral angle model, we simulate MQC spectra from a partially deuterated DMBP sample [ $(1^{13}\text{C})\text{-}3',5'\text{-bis(methyl-d}_3\text{)}\text{biphenyl}$ ] using the 9 PMFs, and compare



**FIG. 4. Refining a molecular mechanics model of the torsion free energy of  $[1-^{13}\text{C}]\text{-}3',5'\text{-dimethylbiphenyl}$  (DMBP), validated by multiple quantum coherence (MQC) experiments.** a) Workflow diagram identifying whether a task is performed in the nuclear magnetic resonance spectrometer (red), on the quantum computer (blue), or classically (yellow). b) Plots of the 9 potential of mean-force (PMF) candidate functions of the dihedral angle of DMBP. c) Out of time ordered correlators (OTOC) of DMBP, comparing experimental NMR data (red points), OTOC simulations run on the Willow chip using the most likely PMF (blue points), and classical simulations with added RF inhomogeneity (lines). d) Root-mean-square error (RMSE) of the OTOC or MQC peak position for the 9 candidate free energy surfaces shown in (b); lines are a Gaussian process fit. Error bars in all plots are  $1\sigma$  CI except for the estimates of the PMF minimum which are shown with  $2\sigma$  CI from bootstrapping.

against experimental data. We plot [Fig. 4(d), green] the RMSE in the 6- and 7-quantum coherence spectra peak positions between experiment and the set of PMF candidates. The minima of the MQC, QC, and simulation RMSEs agree to within error bars, the error bars are similar in size, and the learned PMF improves the MQC RMSE by a factor 4 compared to a bare MD calculation. These results validate both the precision and the accuracy of our OTOC learning technique.

## OUTLOOK

This work introduces a new quantum-information-inspired framework for determining long-range distance constraints in NMR, with potentially broad applications. Long-range distance constraints have proven essential in many contexts in NMR [22], serving as “anchor points” in structural biology [19, 20, 40, 41]. In these works, they have been used to discriminate polymorphs or furnish critical restraints for intramolecular and inter-subunit geometries inaccessible to local probes. Beyond biomolecular systems, they have been critical in validating structural models and connectivity in materials such as metal–organic frameworks [42].

OTOCs probe long-range information via the propagation of local polarization across an extended spin network and its sensitivity to weak local perturbations under time-reversal. This approach offers advantages over traditional time-ordered correlation (TOC) experiments, i.e. protocols that do not involve time-reversal. Specifically, OTOCs may provide improved sensitivity to spa-

tial correlations and distances in systems where local control is limited [26], such as the  $^{13}\text{C}$ -proton networks studied within this work. While the size of the many-body correlations and the accessible distance measurements will be system-specific [43], these factors do not change the core applicability of the technique. Estimates (see Supplementary Information) suggest accessible distances of 20–60 Å for OTOC-based measurements, approaching the length scale of Förster Resonance Energy Transfer (FRET) [44]. This is beyond the reach of state of the art NMR techniques such as proton-driven spin diffusion (PDSD) [13], rotational echo double resonance (REDOR) [11], and radio-frequency-driven recoupling (RFDR) [12], all of which are TOC protocols.

Additionally, while the experiments presented here focus on single OTOCs, the framework naturally fits within the scope of multi-dimensional NMR. Multiple OTOCs, excited at distinct molecular sites or under effective Hamiltonians with varying symmetry-selection rules, could enable multi-anchor-point spectroscopy, supporting extensive extraction of structural parameters. Moreover, OTOCs need not be restricted to pulse-sequence excitation: molecular dynamics themselves may drive butterfly operations, providing access to distinct dynamical timescales and rendering the OTOC a direct probe of molecular motion [45]. The ability of quantum computers to simulate all of these potential experiments opens up future possibilities far beyond the studies undertaken in this work.

Despite the large body of work on the asymptotic analysis of product formulas [35], less work has gone into the constant factor analysis at beyond-classical system sizes.

The performance of product formula techniques [36, 46–49] can be highly system- and quantity-specific, and metrics such as unitary infidelity are only loose predictors of state- and operator-specific quantities such as OTOCs (see Supplementary Information). Estimates of Trotter errors suggest that naive methods would require  $10^5$ – $10^6$  gates to execute OTOCs on 50-spin systems (see Supplementary Information) — a large, but not astronomical gap to current hardware requirements given careful problem selection. This presents a challenge to the quantum algorithms community, as we do not expect this gap to be overcome by physical hardware alone.

Our application of the AlphaEvolve algorithm [5] presents an interesting new direction for such algorithm optimization research. Despite successfully overcoming the Trotterization challenge for DMBP, our AlphaEvolve optimization loop relies on evaluating candidate circuits against a complete, classically pre-computed dataset of OTOCs, which is not immediately scalable to beyond-classical system sizes. However, in this work we leveraged the inherent strengths of the LLM-based agent to write code that constructs Trotterized circuits, rather than generating circuits directly [50]. We argue that this has two key advantages. The first advantage is that the resulting code can generate circuits for times and landscape parameters outside its training data set, allowing some degree of generalization and protection against overfitting. However, while interpolation between the training data points works in limited cases, the generator commonly fails to extrapolate; improving on this is a key target for future research. The second advantage is that the resulting function and the prompts used can be understood and analysed by humans, which we attempt in the Supplementary Information. We find that the primary focus of the code (relative to a first order Trotter formula) is a mixture of light cone pruning, term and qubit ordering, adaptive time steps and a distance-based term rescaling, some of which have been suggested before in the literature [46, 48]. We suggest that such analysis can allow the ideas identified by AlphaEvolve to be scaled to system sizes inaccessible to exact simulation.

Molecular dynamics estimates are compromised by several sources of error, including finite size effects, force field inaccuracy, and insufficient sampling. While each of these issues may be addressed in principle, simultaneously addressing all of them is impractical. Instead, addressing these shortcomings through Hamiltonian learning represents a promising avenue for accurately assessing structural properties, enabled by scalable reproduction of many-body spin dynamics by quantum computers. Despite the limitations of the current scheme, in part due to experimental uncertainty, the learned mean value of the DMBP dihedral angle improves over predictions from condensed phase molecular mechanics. These results highlight that experimentally obtained quantum many-body echoes can parameterize highly tuned structural models by providing information that can correct the deficiencies of molecular dynamics simulations.

## MATERIALS AND METHODS

### Materials

Nematic liquid crystals 4'-Ethoxybenzylidene-4-butylaniline ( $\geq 99.0\%$ ) [EBBA] and 4-Cyano-4'-pentylbiphenyl ( $\geq 98.0\%$ ) [5CB] were purchased from Tokyo Chemical Industry America and used as received. Isotopically enriched [ $4\text{-}^{13}\text{C}$ ]-toluene ( $\geq 99.0\%$ ) was purchased from Sigma Aldrich and used without further purification. Isotopically labeled samples of [ $1\text{-}^{13}\text{C}$ ]-3',5'-dimethylbiphenyl and 3',5'-bis(methyl-d<sub>3</sub>)biphenyl were synthesized according to literature procedures using starting materials purchased from Sigma-Aldrich; full synthetic details are provided in the Supplementary Information. The purity of synthesized [ $1\text{-}^{13}\text{C}$ ]-3',5'-dimethylbiphenyl and [ $1\text{-}^{13}\text{C}$ ]-3',5'-bis(methyl-d<sub>3</sub>)biphenyl was confirmed by  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectroscopy, showing no detectable impurities.

### NMR sample preparation

[ $4\text{-}^{13}\text{C}$ ]-toluene was prepared by mixing 4.2 wt% of the compound with EBBA, heating to 358 K to reach the isotropic phase, and sonicating to homogenize. [ $1\text{-}^{13}\text{C}$ ]-3',5'-dimethylbiphenyl and 3',5'-bis(methyl-d<sub>3</sub>)biphenyl were prepared by mixing 2% v/v of each compound with 5CB, followed by heating to 313 K and sonication. Prior to each NMR experiment, all samples were homogenized by sonication above the nematic–isotropic transition temperature, injected into the NMR magnet, thermally cycled several times through the transition point, and then cooled slowly to the target experimental temperature.

### NMR experiments

NMR experiments were performed on a Bruker UltraShield 11.75 T (500 MHz for  $^1\text{H}$ ) magnet equipped with a Bruker Avance I console. A 5 mm double-resonance BBO smart probe was tuned to facilitate  $^1\text{H}$  and  $^{13}\text{C}$  experiments. The sample temperature was maintained at 295 K for [ $4\text{-}^{13}\text{C}$ ]-toluene@EBBA and 289 K for [ $1\text{-}^{13}\text{C}$ ]-3',5'-dimethylbiphenyl@5CB, [ $1\text{-}^{13}\text{C}$ ]-3',5'-bis(methyl-d<sub>3</sub>)biphenyl@5CB using the internal temperature control unit. Pulse powers of the  $^1\text{H}$  and  $^{13}\text{C}$  channels were synchronized to achieve a nutation amplitude of  $\sim 23.8$  kHz. All OTOC experiments were preceded by a preparation filter generating pure  $^{13}\text{C}$  magnetization via a combination of gradient pulses and Loschmidt echo elements, followed by the application of a TARDIS element for variable durations. The TARDIS element was designed to generate effective dipolar double-quantum evolution in the presence of homo- and heteronuclear dipolar couplings

( $H_{\text{eff}} \propto \sum_{i < j} d_{ij}^{II} I_i^+ I_j^+ + \sum_{i,j} d_{ij}^{IS} I_i^z S_j^z + \text{h.c.}$ ), taking into account finite pulse width and pulse amplitude effects (see Supplementary Information). A selective perturbation, in the form of a local rotation, was applied to the methyl protons in both target systems. This was achieved through the application of a BLEW12 [51] element applied to the proton channel, resonant with the methyl spins, followed by a second BLEW12 proton element sandwiched by two carbon  $\pi$ -pulses. Time-reversal was achieved through a second ( $\pi/2$ )-phase-shifted TARDIS element sandwiched by two  $\pi$  pulses on the carbon channel. Application of the ( $\pi/2$ )-phase-shift reversed the sign of the double-quantum part of the effective generator ( $R_z^I(\pi/2)I_i^+ I_j^+ R_z^I(-\pi/2) = -I_i^+ I_j^+$ ) [10], whereas the two  $\pi$  pulses on the carbon channel inverted the sign of the heteronuclear part similar to a spin echo. Finally, readout was performed on the  $^{13}\text{C}$  channel. MQC spectroscopy followed the general scheme outlined in [34]. Full details of the NMR experiments and all pulse sequence elements are given in the Supplementary Information.

### Processing of NMR and SC experimental data for learning experiments

Each NMR dataset shown is an average over 10 independent measurements, allowing us to calculate the standard error  $\sigma_t = [\sum_j (C(t) - C_j(t))^2]^{1/2}/N$ . Measurements of datapoints at different times were assumed to be independent. However, as the measured peak amplitudes are normalized by the OTOC at  $t = 0$ , the uncertainty in this amplitude was propagated to the other OTOC estimates, which led to a covariance matrix  $\Sigma_{\text{NMR}}$  with non-zero off-diagonal terms.  $\Sigma_{\text{NMR}}$  was then used to obtain the covariance weighted error plotted in Fig. 2(e):  $\text{Cov. wt. err} = [(C_{\text{NMR}} - C_{\text{sim}})^T \Sigma_{\text{NMR}} (C_{\text{NMR}} - C_{\text{sim}})]^{1/2} / \text{Trace}[\Sigma]^{1/2}$ . Here, we write  $C$  as shorthand for arrays containing the OTOC data indicated in the subscript. Error bars in Fig. 2(e) were obtained by bootstrapping over the initial 10 points for each NMR dataset. To combine this data with the output of the quantum computer for Fig. 3 and Fig. 4, we took the covariance matrix  $\Sigma_{\text{QC}}$  (which we assume to be diagonal), and combined this with the NMR covariance matrix to yield  $\Sigma_{\text{tot}} = [\Sigma_{\text{NMR}}^{-1} (\Sigma_{\text{NMR}} + \Sigma_{\text{QC}})^{-1} \Sigma_{\text{QC}}^{-1}]^{-1}$ . This was substituted for  $\Sigma_{\text{NMR}}$  in the above definition of the covariance weighted error to give the cost function in Fig. 3(c). Error bars in this figure were calculated by bootstrapping over the NMR datapoints, and re-sampling the quantum computer data from a normal distribution. For Fig. 4(d), we used the root-mean-square error  $\text{RMSE} = [\frac{1}{N} \|C_{\text{sim}} - C_{\text{NMR}}\|]^2$  as a cost function instead of the covariance weighted error.

### Molecular dynamics simulations details

Starting structures were generated with Packmol version 21.0.0 [52] at a concentration of 2.05% v/v 1,3-dimethyl-5-(phenyl-13C)benzene and a starting density of 1.01 g/ml. The liquid crystal molecules were approximately aligned along the x-axis of the simulation box with dimensions [65Å, 50Å, 50Å] by providing angle constraints in the Packmol input file. The initial structures were then relaxed by minimizing the system at constant volume and equilibrated with 500 ps of isothermal-isobaric (NPT) dynamics. All simulations used a 2 fs time step with hydrogen bond constraints and a hydrogen mass of 4 amu. NPT calculations used a Monte Carlo barostat with updates every 25 steps. All ensemble averages were evaluated with respect to 8 independent trajectories with production times of 1  $\mu\text{s}$  each. All molecular dynamics simulations were performed using OpenMM [53].

Non-bonded parameters from GAFF version 2.11 [54] were deployed for 5CB and DMBP in combination with bonded parameters from Grappa version 1.3.1 [55]. Although Grappa was designed for biomolecules, the training set includes a very large dataset of small organic molecules and we have observed empirically that the Grappa force field imparts additional flexibility to the aliphatic chains in the liquid crystal relative to GAFF. This is consistent with modifications made to GAFF for liquid crystal systems [56].

The order parameters of 5CB are known to be sensitive to the temperature [57, 58], and are difficult to model with molecular dynamics [59]. This necessarily impacts both the mean-value of the dihedral angle of DMBP, and the  $\phi$ -dependent coupling matrices  $D_{ij}(\phi)$ . To minimize this effect, we shift the temperature of the experiment to 289 K (as reported in the above section), where the order parameters most accurately match simulation, and use this throughout all DMBP experiments.

### Classical simulation of toluene

Toluene in EBBA was modeled as a static benzene ring with a freely rotating methyl group in a fixed plane, in a similar manner to Ref. [34]. The motion in the liquid crystal background was treated with the Saupe tensor formalism, assuming all molecular motion occurs on timescales  $\ll 1$  ms (see Supplementary Information). Data for the molecular co-ordinates and all J-couplings was taken from Ref. [34], except for the J-coupling between the  $^{13}\text{C}$  and the nearest proton (which is significantly larger than all other J-couplings). The Saupe tensor, chemical shifts, and the remaining J coupling were obtained by fitting to MQC data (see Supplementary Information). Data was separately validated by comparing to the molecular dynamics simulations of the toluene-EBBA system above.

## Trotterized Hamiltonian simulation and machine learning

### *AlphaEvolve setup*

AlphaEvolve [5] is an evolutionary coding agent that leverages large language models (LLMs) to discover novel and efficient solutions for a variety of scientific problems. In this work, AlphaEvolve is used to generate a set of Python functions that produce quantum circuits approximating Hamiltonian evolution across an entire parameter landscape of time and Hamiltonian inputs, as opposed to optimizing individual quantum circuits or circuit parameters directly. The process was seeded with a human-written program (a first-order Trotter formula) that served as the initial solution to a given task. Through a process of mutation, evaluation, and selection, AlphaEvolve ultimately generated a population of programs that produce simulation circuits with significantly lower approximation error against the reference (improving from a 10.4% mean error to 0.82%), while remaining below the gate budget enforced by the device (see Supplementary Information for further explanation).

The MAP-Elites-like setup of AlphaEvolve [60] was used as a base for this work, optimized for a single RMS error metric targeting the average approximation error across the entire set of landscape parameter and time values evaluated. By stochastically injecting domain-specific suggestions into the prompt, the optimization algorithm was steered towards granular changes relevant to the task. Here, task-specific modifications were introduced, such as a mixture of hand-crafted and LLM-generated task instructions specific to quantum simulation and circuit optimization problems. These led to increased performance of the algorithm, both in terms of achieved metric value and convergence speed, i.e. number of required LLM samples. The LLMs are pre-trained on large corpora of scientific literature and thus already possess knowledge of the most commonly used quantum algorithms, quantum simulation, and other relevant domain knowledge.

For each parent program, in addition to the program’s overall RMS error, an associated error matrix was included in a tabular format giving the OTOC error for each point in the parameter landscape and time. This feedback allows the LLM to perform in-context reinforcement learning: by focusing on regions with high error, the model can attempt to generate a more accurate circuit for these regions without sacrificing overall performance.

The algorithm’s parallelized nature was leveraged to redistribute resources. To cover a larger portion of the search space, multiple, initially independent AlphaEvolve runs were started. As the optimization progresses, the independent runs were regularly synced by re-seeding with the current best solution. This was either performed via full re-seeding or by eliminating only the worst-performing run. Empirically, we found that this strategy resulted in a much faster improvement of the

fitness function with respect to the number of LLM samples used.

### *Circuit optimization problem setup*

For processing by AlphaEvolve, a problem needs to be split into four well-defined pieces, as described here:

**Problem Specification:** A Python program implementing a first-order Trotter formula with a swap network for 15 qubits on a chain in a dense brick-wall fashion served as starting point for the evolutionary process. For this initial program, the input parameters of the function were the total time  $t$  and number of Trotter steps to be implemented, as well as the homogeneous/heterogeneous ( $j$ ) couplings and an on-site terms. Additionally, we provided the butterfly and measurement indices, the landscape parameter  $p$ , and a maximum number of allowed CZ gates, despite these being unused in the function initially. These additional parameters were included in the function signature to inform AlphaEvolve about the wider context of the program, and to give additional information could be useful for optimization.

**Prompt and sampling:** In each iteration, AlphaEvolve selected a program from its database to act as a parent. This parent program was then used to generate a prompt for an LLM, which was tasked with creating a child program that improves upon the parent. A fixed generic prompt was included, describing quantum simulation, Trotterization, NMR, and OTOCs on a very high level. Per LLM call, a domain-specific or generic prompt was further added, as well as focus prompts that primarily target code simplification, refactoring and removal of dead code, or to focus on a particular task (lightcone, THRIFT [61], commutator scaling). In all cases, the signature of the function to be generated remained fixed.

**Evaluation:** The obtained function was used to generate quantum circuits for all choices of parameters and time steps. These circuits were then individually evaluated by averaging the OTOC over 250 randomly-sampled initial states. To ensure that this random evaluation yields only a small approximation error, this approximate OTOC was compared with the exact OTOC computation that uses the full set of  $2^{15}$  states. For the latter a new technique using sparse matrix computations on GPU was implemented to produce unitary matrices for large circuits instead of much slower regular einsum-based method. The seed for the state sampling varied for each time and landscape parameter choice, but remained fixed throughout the optimization run to avoid small random fluctuations influencing the evaluation of mutations while speeding up the pipeline. The root mean square error between the evaluated OTOCs and the reference data was used as an optimization metric.

**Database Integration:** Child programs that were both valid and exhibited a sufficiently high fitness score were added to the database of programs. Programs were discarded, for example, if the circuit they produced ex-

ceeded a predefined number of gates. Counting CZ gates as implemented in the Cirq framework with Cartan’s KAK decomposition [62] was the dominant bottleneck, with this calculation alone being the dominant part of the step time. To optimize this, a highly parallel KAK decomposition was implemented, effectively removing the contribution of the compilation step to the total time and compute budget. Consequently, the evolved program was provided with an option to use CZ counting as a tool (which it did by implementing a two-pass circuit generation, see Supplementary Information).

We provide an additional analysis on the efficiency of different instruction prompts, high-level intuition of what kinds of modifications AlphaEvolve suggests and generated codes and circuits in the Supplementary Information.

### Details of superconducting qubit devices and gate calibration

Quantum simulation was performed on a Willow quantum computer similar to that described in [4] (see also [63]). Processors in this lineage are based on a 2D grid of flux tunable superconducting transmon qubits connected by flux tunable couplers between nearest neighbors. In this work, the mean frequency of operation was  $\sim 6.2$  GHz with anharmonicity  $\sim 210$  MHz.

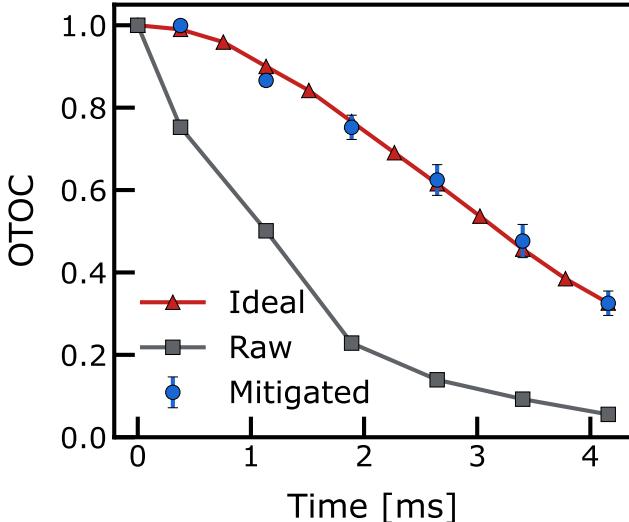
As the two molecules addressed in this work require nine and fifteen qubits only, we selected a sub-grid of the 105 qubit Willow with performance better than average in the relevant benchmark metrics. Qubits were arranged in a line, with the measurement qubit at one end. The average lifetime of the single photon excitation in the chosen line of qubits was  $T_1 = 114$   $\mu$ s. The average coherence time of the chosen qubits as measured by a Hahn-Echo sequence was  $T_{2E} = 130$   $\mu$ s. The performance of single qubit microwave gates was measured by Clifford randomized benchmarking (RB) error and the associated qubit impurity, which were found to be on average 0.00020 and 0.00015, respectively. The average cycle Pauli cross-entropy benchmarking (XEB) error, including error contributions from two single qubit gates and a single two-qubit gate, was for CZ 0.0015, and for  $\sqrt{i}$ SWAP 0.0014. The average readout error of the single measurement qubit was 0.00985. Since the data taking for the quantum simulation of the OTOC-NMR learning sequence took several days, these metrics drifted slightly before completion. While the processor requires recalibration during this time, these numbers are typical and representative of our experiments.

The all-to-all coupling was achieved by a swap network, a technique adapted from [37]. Here, one compiles swap gates through a depth  $N$  “brick wall” pattern of two-qubit interactions, such that the swap gates permute floating spin indices through the qubit array. This is optimal in terms of both the number of gates and the depth required to execute one interaction between each pair of

spins, as is needed for a single Trotter step of the all-to-all coupled dipolar Hamiltonian. In our quantum simulation of Toluene, we further approximated that the  $^{13}\text{C}$  spin only interacted with its nearest proton. This single coupling was treated in the interaction frame and the  $^{13}\text{C}$  spin was removed from the swap network, reducing the number of gates needed by 17%. This approximation was justified as all other  $^1\text{H}$ - $^{13}\text{C}$  couplings were two orders of magnitude smaller for the entire set of toluene geometries considered, and we observed that the interaction frame treatment slightly reduced the Trotter error.

The quantum simulation for Toluene was based on a natural decomposition of the target effective Hamiltonian into Fermionic simulation (fSim) gates [64]. These gates are parametrized by two two-qubit interaction angles: the swap angle and the conditional phase angle. Across the entire landscape, the Trotterization of the Toluene evolution required calibrating a set of 80 unique fSim gates. Previous work was able to operate with a single pulse fSim [3, 65–69]. However, achieving arbitrary combinations of swap and conditional phase angles with a single pulse often requires either a long gate duration, which degrades performance by decoherence, or a large interaction coupling strength between qubits, which increases leakage processes outside of the computational manifold. In this work, due to the range of coupling strengths in Toluene and the coarse-grained Trotterization, the swap angles that needed to be calibrated spanned the entire domain from 0 to  $\pi/2$ . To solve this issue, we adapted a two-pulse fSim approach from earlier work in a Sycamore architecture [64]. Here, a first base band interaction pulse sets the swap angle and induces a spurious  $< 100$  mrad conditional phase, while a second pulse enacts a conditional phase and a spurious  $\sim 30$  mrad swap angle. The two-pulse approach is calibration intensive, but provides higher performance and flexibility than a single pulse approach.

The two pulses for each fSim gate were calibrated together as a single gate with an iterative procedure adapting one of two high precision periodic “Floquet” calibrations previously developed, depending on the target swap angle. For small target swap angles ( $< 30$  mrad) we adapted the technique presented in [66]. For larger swap angles and for the conditional phase we adapted the technique presented in [65]. To achieve the desired precision in the full range we fine-tuned the number of gates used in the periodic calibration and fitting procedure depending on the value of the expected angle. The procedure accounts for the aforementioned spurious terms as part of the calibrated final gate, attaining the target interaction angles with a max tolerance of 20 mrad, and typical error  $< 5$  mrad. As in previous work [3, 4], single qubit phases of the fSim gates are measured by XEB fitting and removed by phase matching single qubit gates to yield the total gate. The median XEB error for the tuned fSim gates was measured at 0.0026 and max XEB error 0.0045. Compatible pairs of qubits were calibrated simultaneously, and therefore benchmarks include cross-



**Extended Data Fig. 1. Benchmark of Pauli Path ZNE error mitigation on the 15-spin molecule at odd time indices.** Out-of-time-order correlator (OTOC) measurements for DMBP at *odd* NMR time indices, taken during a different experimental run from that shown in the main text, on a different date, to provide a secondary benchmark. The smallest raw signal (gray squares) measured was  $0.055 \pm 0.003$  (data includes real-time readout mitigation, double-sided light cone filtering, twirling, and dynamical decoupling). Despite this small signal, the Pauli Path ZNE-mitigated data (blue circles) was successfully extracted, serving as a stringent test for the mitigation technique.

talk effects. More details on these experiments can be found in the Supplementary Information.

For the quantum simulation of DMBP, a 15 qubit linear chain with similar performance was used. In contrast to the Toluene experiments, the AlphaEvolve coding agent generated circuits using a CZ decomposition, which did not require us to use the fSim construction detailed above.

#### Error mitigation pipeline for noisy quantum circuits

The structured character of the NMR OTOC circuits prohibits the use of mitigation strategies used in earlier works that are tailored to random OTOC circuit ensembles [4]. The circuits used are also very deep: the deepest 9-spin circuits span 326 circuit moments and used 540 two-qubit composite fSim gates, while the deepest circuits for the 15-spin molecule span 484 moments and used 792 CZ gates. (These numbers count only relevant gates within the double light cones of the measurement and butterfly operators; we typically remove gates that will have no impact on the final OTOC measurement from the circuit.) The resulting damping of the noisy OTOC signal was observed in the worst case to be as small as 0.15.

To reliably and accurately extract data from these

noisy quantum circuits, we employ a composite, reconfigurable error mitigation framework designed to operate across different gate sets, using distinct configurations for the 9-spin (fSim-based) and 15-spin (CZ-based) molecular OTOC experiments. We developed a four-stage mitigation pipeline consisting of algorithmic and physical-level stages: (1) Circuit generation at the algorithmic level; (2) Algorithmic-level error mitigation, which involved double-sided light cone filtering [4], Pauli Path ZNE (described below) [70–73], algorithmic-level DD, and sub-Clifford algorithmic-block twirling [74, 75]; (3) Native gate decomposition and optimization; (4) Physical-layer mitigation, which involved hardware-adapted DD sequences for idle periods with Clifford twirling [74–77], and real-time randomized readout error mitigation [78]. This describes the full pipeline for the 9 spin fSim circuits. The 15 spin CZ-based circuit used the same pipeline, but without the need for algorithmic-level DD, and using full (rather than sub) Clifford group twirling for the two-qubit gates and idle periods.

As part of this work, we developed a novel Pauli-path zero-noise extrapolation (ZNE) scheme based on the evolution of the measurement operator in the Pauli path picture [3, 4, 30, 79]. This is based on the identification that the noisy expectation value  $C(\lambda)$  (with  $\lambda$  the noise scaling factor) is the Laplace transform of the underlying Pauli path total Hamming weight distribution  $c(H)$  [4], where  $H$  is the noise accumulation weight. Though this distribution  $c(H)$  can be highly oscillatory, the Laplace transform smooths out high-frequency components, allowing simple, low frequency approximations to  $c(H)$  to robustly extrapolate  $C(\lambda)$ . Taking a normal approximation of  $c(H)$ , we derive an analytical functional form for the observed decay which serves as our main model for extrapolation:

$$C(\lambda) = C(0) \frac{\left(1 + \text{erf}\left(\frac{\bar{H}_c - \lambda\sigma_c^2}{\sigma_c}\right)\right)}{\left(1 + \text{erf}(\bar{H}_c/\sigma_c)\right)} e^{-\lambda\bar{H}_c} e^{\lambda^2\sigma_c^2/2}. \quad (2)$$

The functional form is parameterized by three key quantities:  $C(0)$  is the noise-free extrapolated value, while  $\bar{H}_c$  and  $\sigma_c$  represent the mean and standard deviation respectively of the Pauli path Hamming weight distribution  $c(H)$ . These two parameters effectively characterize the noise-induced decay profile. We define the dimensionless parameter  $\beta$  as the ratio of the spread to the average noise accumulation  $\beta = \frac{\sigma_c}{\bar{H}_c}$ , where  $\beta > 0$  and typically  $\beta < 0.5$ . For deep circuits where  $\beta$  is non-negligible, the quadratic term  $e^{\lambda^2\sigma_c^2/2}$  and the error function (erf) correction become essential, as they account for the non-exponential, concave decay profile caused by the complex quantum dynamics, and the physical constraint of non-negative noise weight ( $H \geq 0$ ). This model is conjectured to work even in the regime where Pauli path numeric approximations fail (as is the case in these experiments).

The Pauli Path ZNE framework can be very sensitive to statistical noise when working with deep circuits with

low fidelities. To achieve robust extrapolation across the entire time series, we implemented a specialized Bayesian model-fitting update workflow. This time-correlated strategy utilizes data from shallower, less-noisy time steps—where the circuit fidelity is stronger—to establish a Bayesian prior. This prior then iteratively seeds and updates the model fit for deeper, noisier circuits, stabilizing the parameter estimation for the extrapolation. We employ a global-parameter optimization algorithm, which combines a local minimization routine with random jumps across the parameter space to overcome local optimization minima and improve robustness to noise, and a rare outlier rejection protocol. We employ bootstrap resampling over the mitigation circuit instances to evaluate both the mitigated values and their respective uncertainties. This combined with the Pauli path theory allows the system to overcome the challenges associated with signal damping and accurately estimate the zero-noise limit even in the deepest regions of the circuit.

## ACKNOWLEDGMENTS

The authors acknowledge useful discussions with L. Emsley, L.D. Nielsen, T. Polenova, R. Schurko, R. Tycko, G. Pantaleoni, and N. Wiebe. We thank Pines Magnetic Resonance Center’s Core NMR Facility at UC Berkeley for spectroscopic resources used in this study. D.W. Berry worked on this project under a sponsored research agreement with Google Quantum AI.

## AUTHOR CONTRIBUTIONS

C. Zhang prepared molecular samples and performed the NMR experiments. R.G. Cortiñas and A.H. Karamlou performed the superconducting quantum simulation experiments. C. Zhang, C. Bengs, C. Ramanathan, A. Ajoy, and T.E. O’Brien conceived of the NMR OTOC experiments. C. Zhang and C. Bengs designed the pulse sequence elements and selected the molecular systems. S.H. Kang and R. Sarpong synthesized isotopically-enriched samples of DMBP. J. Provazza, N.C. Rubin, and A. White performed molecular dynamics simulations of the 9Q and 15Q experiments and conceived of the learning experiment. N.C. Rubin, C. Zhang, C. Bengs, J. Provazza, A. White, and T.E. O’Brien performed noisy density matrix simulations of the toluene and DMBP experiments. R.G. Cortiñas, X. Mi, and T.E. O’Brien designed the circuits for the 9-qubit superconducting experiment. J. Bausch, S. Shirobokov, A. Senior, N. Tomašev, B. Kozlovskii, F.J.H. Heras, and P. Kohli designed, tested, and ran the AlphaEvolve optimization experiments. R.G. Cortiñas designed and implemented the pulse calibration sequence for the arbitrary-angle fSim gate, with assistance from X. Mi and J.A. Gross. Z.K. Minev, N. Noll, R.G. Cortiñas, A.H. Karamlou and T.E. O’Brien implemented the coding infrastructure for the

superconducting experiments. N. Noll, D. Kafri and Z.K. Minev performed accurate modeling of the superconducting qubit experiments. Z.K. Minev and N. Noll designed and implemented the error mitigation sequences with assistance from T.E. O’Brien and T. Schuster. The Google Quantum AI team designed, built, calibrated and maintained the superconducting qubit chip, along with the dilution refrigeration, the electronics stack, and the superconducting experiment codebase. All authors contributed to the writing and revisions of the manuscript.

## Supplementary Information: Quantum computation of molecular geometry via many-body nuclear spin echoes

### CONTENTS

I.	Outline	14
II.	Chemical Synthesis Details	14
A.	General considerations	14
1.	Reagents and solvents	14
2.	Experimental conditions	15
3.	Analytical Methods	15
B.	Synthesis of [1- <sup>13</sup> C]-3',5'-dimethylbiphenyl	16
C.	Synthesis of 3',5'-bis(methyl-d <sub>3</sub> )biphenyl	16
III.	NMR details	19
A.	Notation	19
B.	NMR Hamiltonian Overview	19
C.	OTOC Experiments	21
D.	The TARDIS pulse sequences	22
1.	TARDIS-1	24
2.	TARDIS-2	25
3.	TARDIS-2 sequences with on-site field	25
E.	Initial State Preparation and RF inhomogeneity	26
F.	Perturbation	27
G.	Readout	28
H.	Additional OTOC and learning experiments	28
I.	HETCOR Experiment	29
J.	MQC Experiments	31
1.	MQC spectra for Toluene	32
2.	MQC experiment of DMBP	33
K.	Higher order Magnus expansion terms	34
L.	Understanding features of the toluene OTOC decay	38
M.	Predicting ultimate decoherence distance limits in measuring OTOCs	39
IV.	Molecular Dynamics model	41
A.	15Q Simulation details	42
B.	15Q $P_2$ vs temperature Calibration	42
C.	Analysis of timescales for the 15Q system	44
D.	15Q Analysis of the effect of the non-isotropic environment	45
E.	15Q Analysis of the dipolar coupling matrix elements	45
F.	QM correction to the PMF	46
G.	15Q learning experiment candidate PMFs	47
H.	9Q Simulation details and results	49
V.	Trotter scaling estimates	50
A.	Analysis of Trotter accuracy	50
B.	Numerical tests of Trotter accuracy	53
1.	Precision and resources	54
C.	Extrapolating the cost of Trotterization in dipolar spin systems	56
D.	Correlation between different measures of Trotter error	57
VI.	AlphaEvolve circuit simulation	58
VII.	Classical resource scaling analysis on a simplified model	60
A.	MPO analysis	61
B.	Monte Carlo analysis	63

<b>VIII. The Superconducting Quantum Simulator</b>	64
A. Basic processor benchmarks	64
B. SWAP network: efficient all-to-all interaction from only nearest-neighbours coupling	64
C. Trotterization of a double-quantum Hamiltonian embedded in a SWAP network	66
D. Gate decomposition for composite fSim calibration	67
E. Composite fSim gate calibration	68
<b>IX. Accurate noise modeling of superconducting experiments</b>	70
A. Trajectory simulation of quantum circuits	70
B. Inhomogeneous noise sensitivity	70
C. Agreement of simulations to experiment	71
<b>References</b>	73

## I. OUTLINE

In this Supplementary Information, we cover additional details of the NMR and superconducting experiments, classical simulation, and AlphaEvolve optimization that went into this work. We begin in Sec. II with full details of the synthetic chemistry used in this work (excepting details of chemicals bought from external vendors). Then, in Sec. III, we detail the NMR theory required for this experiment, as well as the design of the Time-Accurate Reversal of Dipolar InteractionS (TARDIS) pulse sequences, details of the primary error mechanisms observed (explaining the noisy classical simulation that was used in the main text), and explicit description of all experiments performed in this work. This includes details of the construction of the reference toluene Hamiltonian used in this work from additional NMR experiments (other than OTOCs) and literature reference data. We also provide NMR data from experiments not shown in the main text (including in Fig. S14 the data that went into Fig. 2(e) of the main text, and in Fig. S16 a demonstration of learning two components of the toluene directional order parameter), and some physical insight into the observed features in the toluene OTOC decay. In Sec. IV, we provide details of the molecular dynamics simulations used for the DMBP simulation, how this was used to develop the dihedral-angle-dependent coupling matrices used in the main text. We also expand on various schemes used to correct the DMBP PMF with higher accuracy simulation, and results of MD simulations of Toluene. In Sec. V, we perform an analysis of lower-order Trotter methods for OTOCs in general, and specifically for the DMBP simulations performed in this work. We find that the error in the OTOC is significantly suppressed in first-order Trotter schemes, and that the DMBP OTOC curves can be replicated with only a few Trotter steps. Then, in Sec. VI, we provide extended analysis of the optimization and output of the AlphaEvolve coding agent applied to the DMBP results, in an attempt to make the output human-understandable. In Sec. VII, we test the performance of two classical computation methods at simulating OTOCs in a simplified 1D model to exemplify the scaling behaviour of both methods. Then, in Sec. VIII, we provide details of the superconducting chip used, the compilation of the time evolution to a SWAP network, and the calibration of the arbitrary angle FSim gate. And finally, in Sec. IX, we provide accurate noise modeling and budgeting of the superconducting qubit device, and a comparison between noisy simulations of the circuits run and the obtained experimental data.

## II. CHEMICAL SYNTHESIS DETAILS

### A. General considerations

#### 1. Reagents and solvents

Unless stated otherwise, commercial reagents were purchased from Sigma Aldrich, Acros Organics, Ambeed, Fisher Scientific, Matrix Scientific, Chem-Impex, Combi-blocks, Tokyo Chemical Industries, Oakwood Chemical, Strem Chemicals, Spectrum Chemical and/or Alfa Aesar, and used without additional purification. Solvents were purchased from Fisher Scientific, Acros Organics, Alfa Aesar, and Sigma Aldrich. Tetrahydrofuran (THF), diethyl ether ( $\text{Et}_2\text{O}$ ), benzene ( $\text{PhH}$ ), toluene ( $\text{PhMe}$ ), methanol ( $\text{MeOH}$ ), and triethylamine ( $\text{Et}_3\text{N}$ ) were sparged with argon and dried by passing through alumina columns using argon in a Glass Contour solvent purification system. Dichloromethane (DCM) was freshly distilled over calcium hydride under a nitrogen ( $\text{N}_2$ ) atmosphere prior to each use.

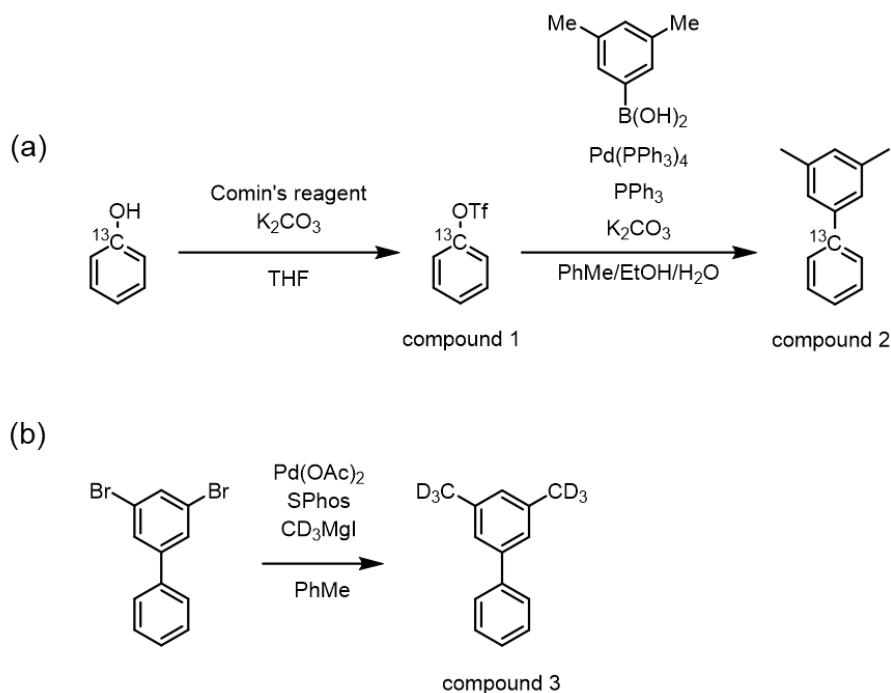


FIG. S1. (a) Synthesis of [1-<sup>13</sup>C]-3',5'-dimethylbiphenyl (compound 2) via Suzuki coupling. (b) Synthesis of 3',5'-bis(methyl-d<sub>3</sub>)biphenyl (compound 3) via Kumada cross coupling.

## 2. Experimental conditions

Unless otherwise noted, reactions were carried out in flame- or oven-dried glassware under a positive pressure of N<sub>2</sub> in anhydrous solvents using standard Schlenk techniques. Reaction temperatures above room temperature (22–23 °C) were controlled by an IKA® temperature modulator and monitored using liquid-in-glass thermometers. Reaction progress was monitored by thin-layer chromatography (TLC) on Macherey-Nagel TLC plates (60 Å, F254 indicator). TLC plates were visualized by exposure to ultraviolet light (254 nm), and/or stained by submersion in aqueous potassium permanganate solution (KMnO<sub>4</sub>), p-anisaldehyde or ceric ammonium molybdate (CAM) stains and heating with a heat gun. Organic solutions were concentrated under reduced pressure on an Heidolph temperature-controlled rotary evaporator equipped with a dry ice/isopropanol condenser. Flash column chromatography was performed with either glass columns using Silicycle SiliaFlash® P60 silica gel (40–63 μm particle size) using ACS grade solvents. All yields refer to chromatographically and spectroscopically (<sup>1</sup>H and <sup>13</sup>C NMR) pure material.

## 3. Analytical Methods

NMR spectra were recorded using deuterated solvents, obtained from Cambridge Isotope Laboratories, Inc. <sup>1</sup>H NMR and <sup>13</sup>C NMR data were recorded on Bruker AVQ-400, AVB-400, NEO-500, AV-600 and AV-700 spectrometers using deuterated chloroform (CDCl<sub>3</sub>) as solvent at ambient temperature. Chemical shifts ( $\delta$ ) are reported in ppm relative to the residual solvent signal (CDCl<sub>3</sub>:  $\delta_{\text{H}} = 7.26$  ppm,  $\delta_{\text{C}} = 77.16$  ppm). Data for <sup>1</sup>H and <sup>13</sup>C spectroscopy are reported as follows: chemical shift ( $\delta$  ppm), multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br = broad, app = apparent), coupling constant (Hz), integration. All raw free-induction-decays (FIDs) were processed, and spectra analyzed, using the program MestReNOVA 11.0 from Mestrelab Research S. L. Melting points were determined using a MEL-TEMPTM apparatus and are uncorrected. High-resolution mass spectra (HRMS) were obtained from the Mass Spectral Facility at the University of California, Berkeley, on a Finnigan/Thermo LTQ-FT instrument (ESI). Data acquisition and processing were performed using the XcaliburTM software.

### B. Synthesis of [1-<sup>13</sup>C]-3',5'-dimethylbiphenyl

In a 50 mL round-bottom flask, phenol-1-<sup>13</sup>C (250.00 mg, 1 equiv, 2.63 mmol), Comin's reagent (1.24 g, 1.2 equiv, 5.26 mmol) were dissolved in anhydrous THF (26 mL). The reaction was stirred under N<sub>2</sub> atmosphere overnight. Upon full conversion of the reaction, the crude mixture was concentrated in vacuo and redissolved in dichloromethane (100 mL). The organic phase was washed with H<sub>2</sub>O (100 mL) and the resulting aqueous phase was extracted with dichloromethane (3 × 100 mL). The combined organic layers were dried with sodium sulfate and concentrated in vacuo. The title compound 1 was obtained as yellow oil (534 mg, 89%) after purification with silica column chromatography (4:1 hexane:EtOAc).

In a 100 mL round-bottom flask, phenyl-<sup>13</sup>C trifluoromethanesulfonate compound 1 (543 mg, 1 equiv, 2.35 mmol) was dissolved in a solvent mixture of toluene (23 mL), ethanol (4 mL), and H<sub>2</sub>O (4 mL) with tetrakis(triphenylphosphine)-palladium(0) (135.83 mg, 0.05 equiv, 117.54 μmol), (3,5-dimethylphenyl)boronic acid (352.57 mg, 1 equiv, 2.35 mmol), triphenylphosphine (92.49 mg, 0.15 equiv, 352.61 μmol), and K<sub>2</sub>CO<sub>3</sub> (2.11 g, 6.5 equiv, 15.28 mmol). A reflux condenser was attached, and the reaction was stirred at reflux overnight. Upon completion of the reaction, H<sub>2</sub>O was added and the aqueous phase was extracted with dichloromethane (3 × 100 mL). The combined organic layers were dried with sodium sulfate and concentrated in vacuo. The title compound 2 was obtained as colorless oil (305 mg, 71%) after purification with silica column chromatography (100% hexane). R<sub>f</sub> (10% EtOAc/hexanes, CAM) 0.90; <sup>1</sup>H NMR (600 MHz, CDCl<sub>3</sub>) δ 7.59 (dq, J = 6.6, 1.0 Hz, 2H), 7.43 (q, J = 7.6 Hz, 2H), 7.35 (dd, J = 7.7, 1.4 Hz, 1H), 7.24 – 7.22 (m, 2H), 7.01 (t, J = 1.5 Hz, 1H), 2.40 (s, 6H); <sup>13</sup>C NMR (151 MHz, CDCl<sub>3</sub>) δ 141.63, 138.38, 129.03, 128.77, 127.51, 127.21, 127.12, 125.26, 21.55; HRMS (EI) Exact mass calcd for <sup>13</sup>CC<sub>13</sub>H<sub>14</sub> [M] [M]: 183.1129. Found 183.1128.

### C. Synthesis of 3',5'-bis(methyl-d<sub>3</sub>)biphenyl

In a 50 mL round-bottom flask, Pd(OAc)<sub>2</sub> (21.59 mg, 0.03 equiv, 96.15 μmol), 2-dicyclohexylphosphino-2,6-dimethoxy-1,1-biphenyl (52.63 mg, 0.04 equiv, 128.21 μmol), and 3,5-dibromo-1,1'-biphenyl (1.00 g, 1 equiv, 3.21 mmol) were dissolved in anhydrous toluene (25 mL). To this solution, (methyl-d<sub>3</sub>)magnesium iodide (1.19 g, 7.05 mL, 1.0 M, 2.2 equiv, 7.05 mmol) was added drop-wise. After stirring the reaction at room temperature for 30 min, the reaction was heated to 80 °C and stirred overnight. The reaction was quenched with a saturated aqueous solution of NH<sub>4</sub>Cl. The aqueous phase was extracted with ethyl acetate (3 × 100 mL). The combined organic layers were filtered over a plug of silica and concentrated in vacuo. The title compound 3 was obtained as colorless oil (510 mg, 85%) after purification with silica column chromatography (9:1 hexane:EtOAc).

R<sub>f</sub> (10% EtOAc/hexanes, CAM) 0.90; <sup>1</sup>H NMR (600 MHz, CDCl<sub>3</sub>) δ 7.66 – 7.61 (m, 2H), 7.49 – 7.44 (m, 2H), 7.41 – 7.35 (m, 1H), 7.27 (dd, J = 7.0, 2.5 Hz, 2H), 7.05 (dt, J = 6.5, 1.7 Hz, 1H); <sup>13</sup>C NMR (151 MHz, CDCl<sub>3</sub>) δ 141.64, 141.42, 138.27, 129.04, 128.77 (d, J = 2.0 Hz), 127.34 (d, J = 1.7 Hz), 127.21 (d, J = 2.2 Hz), 125.27, 20.72 (dt, J = 39.2, 19.7 Hz); HRMS (EI) exact mass calcd for C<sub>14</sub>H<sub>8</sub>D<sub>6</sub> [M]: 188.1472, found 188.1472.

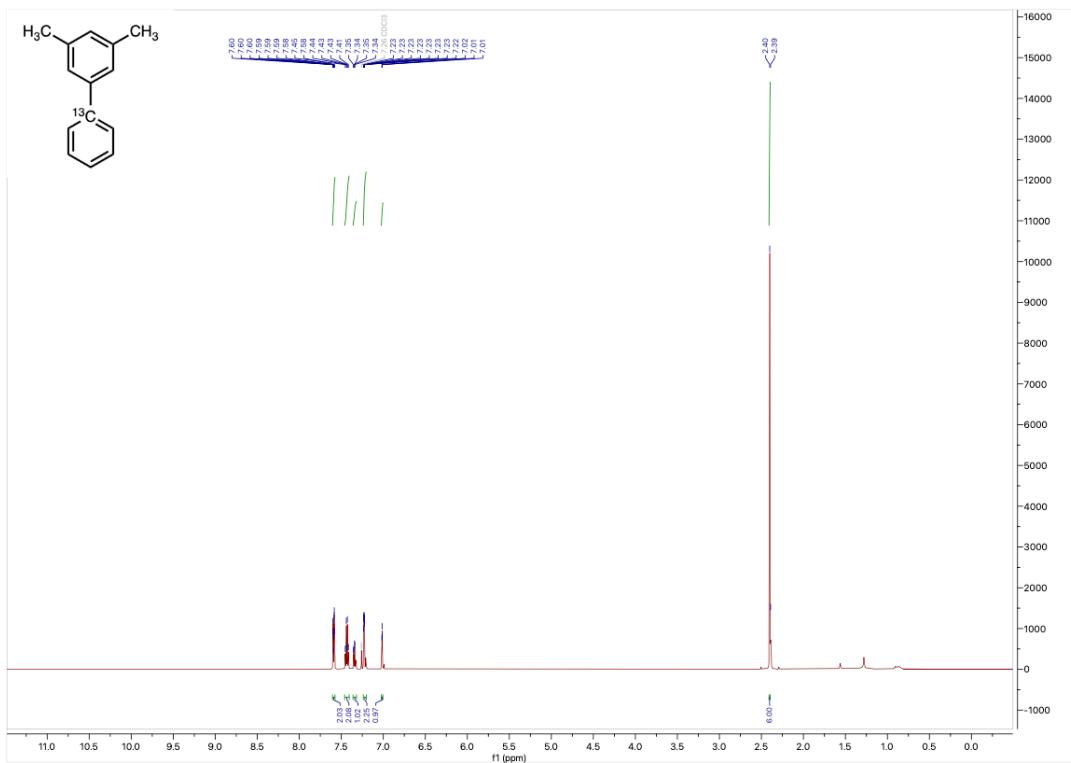


FIG. S2.  $^1\text{H}$  NMR (600 MHz,  $\text{CDCl}_3$ ) for [1- $^{13}\text{C}$ ]-3',5'-dimethylbiphenyl.

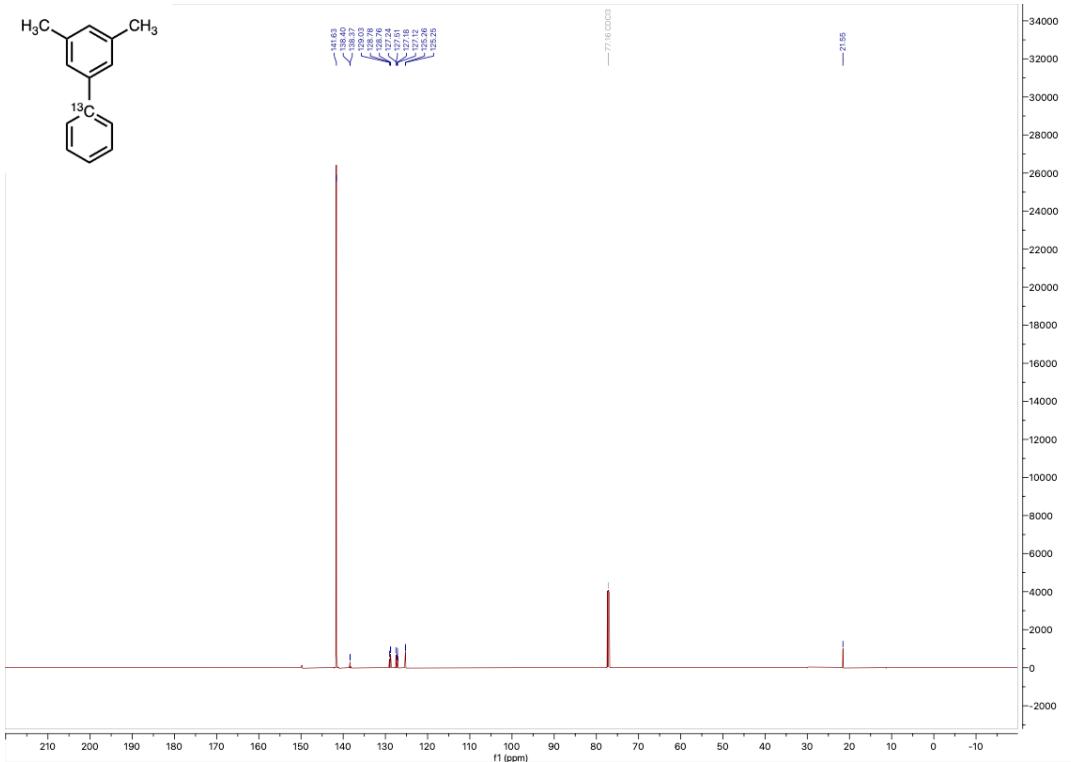
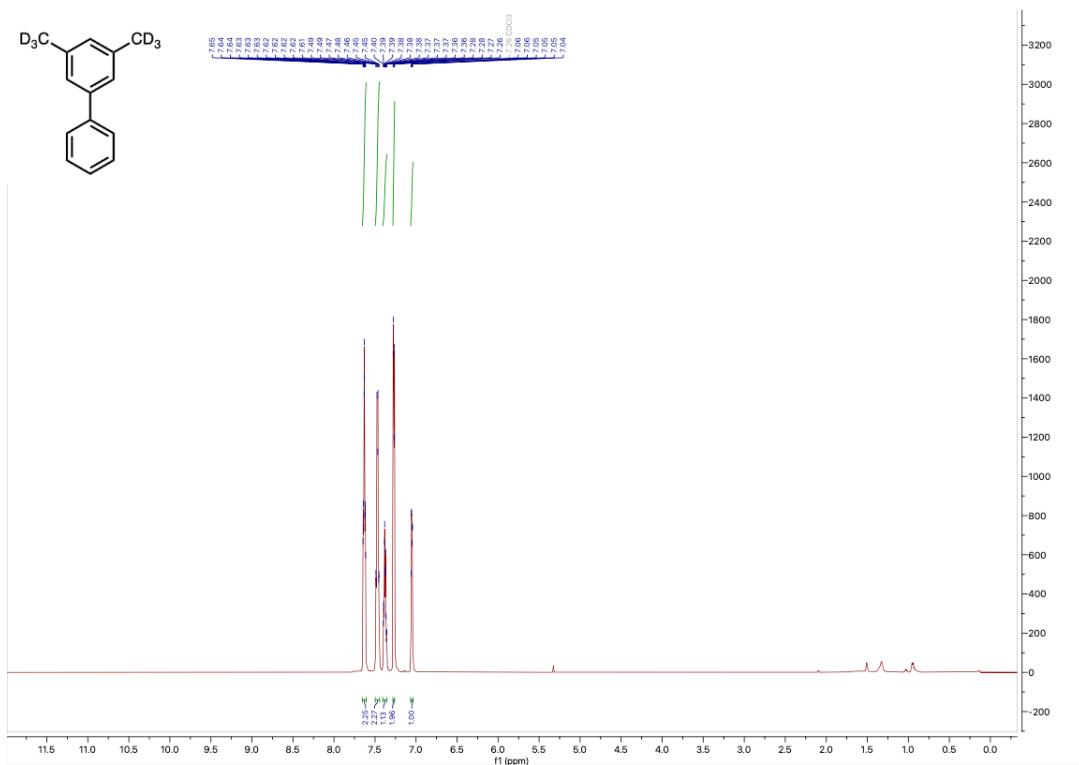
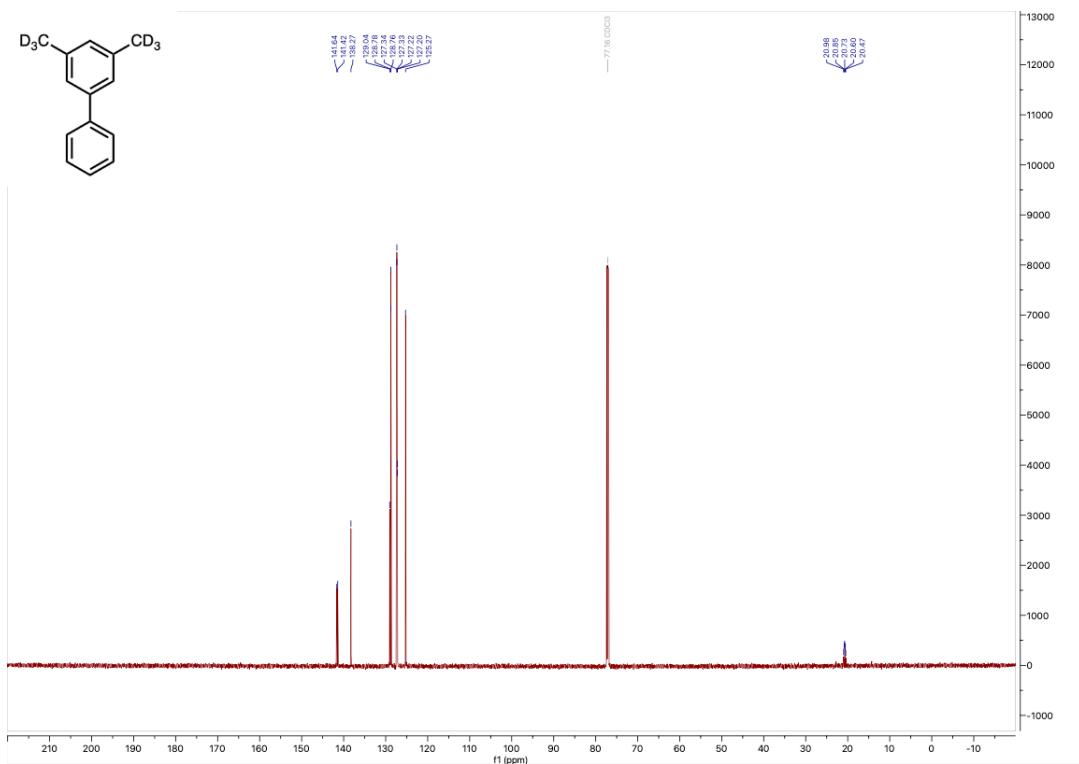


FIG. S3.  $^{13}\text{C}$  NMR (151 MHz,  $\text{CDCl}_3$ ) for [1- $^{13}\text{C}$ ]-3',5'-dimethylbiphenyl.

FIG. S4.  $^1\text{H}$  NMR (600 MHz, CDCl<sub>3</sub>) for 3',5'-bis(methyl-d<sub>3</sub>)biphenyl.FIG. S5.  $^{13}\text{C}$  NMR (151 MHz, CDCl<sub>3</sub>) for 3',5'-bis(methyl-d<sub>3</sub>)biphenyl.

### III. NMR DETAILS

In this section, we detail the NMR experiments performed in this work and present additional data not shown in the main text. We begin with a quick primer noting the differences between common notation in NMR and quantum information. Next, we review the NMR Hamiltonian, describing how it can be derived from molecular structure at finite temperature, where molecular motion cannot be neglected, and explain the Saupe tensor formalism that was used for modeling toluene in this work. We then expand on the OTOC experiment shown in Fig. 2(a) of the main text, providing details of the TARDIS-1 and TARDIS-2 pulse sequences used for forward and backward evolution, the filtering sequence used to prepare a high-fidelity initial state and mitigate RF inhomogeneity, and the BLEW12 pulse used as the butterfly excitation. We also calculate the effective Hamiltonians of all models to lowest order in average Hamiltonian theory. Subsequently, we present additional NMR OTOC datasets not included in the main text: additional datasets exploring the reproducibility of the method, and datasets with an on-site field added (that were used in the construction of Fig. 2(d) in the main text), and datasets showing learning of the two components of the Saupe tensor (see Eq. S9). We then detail the HETCOR experiment performed to determine chemical shift terms in DMBP, which are used in learning experiment. The MQC experiments are subsequently presented, which determine the reference interaction parameters of toluene and validate the DMBP results. Finally, we continue the Magnus expansion of the TARDIS sequence to third order, and show how higher order corrections affect the toluene OTOC.

#### A. Notation

To clarify notation, we summarize the small differences between NMR and QIS. In NMR, it is conventional to use spin angular momentum operators, while in QIS, the Pauli convention is typically preferred. The corresponding associations are as follows:

$$I_x^j = \frac{1}{2}X_j, \quad I_y^j = \frac{1}{2}Y_j, \quad I_z^j = \frac{1}{2}Z_j. \quad (\text{S1})$$

Additionally, in NMR  $I$  is typically reserved for proton spins, and  $S$  denotes the heteronuclear species ( $^{13}\text{C}$  in our case). QIS, on the other hand, adopts a linear indexing scheme (e.g. qubit 1, qubit 2, etc.), irrespective of nuclear species (where applicable). The conversion is as follows:

$$\{I_x^1, \dots, I_x^{N_I}, S_x^1, \dots, S_x^{N_S}\} \rightarrow \{X_1, \dots, X_{N_I}, X_{N_I+1}, \dots, X_{N_I+N_S}\}, \quad (\text{S2})$$

and similarly for  $Y$  and  $Z$ . For NMR it is further convenient to introduce collective spin angular momentum operators

$$I_\mu = \sum_{j=1}^{N_I} I_\mu^j, \quad S_\mu = \sum_{j=1}^{N_S} S_\mu^j. \quad (\text{S3})$$

Both QIS and NMR have a commonly used pictorial representation of their experiments, which are similar but slightly different. In quantum information the representation is the circuit diagram [80], where each qubit is given a line, and symbols are placed across lines to represent quantum gates that typically act on 1 or 2 qubits. However, in NMR one does not have the ability to target single spins. Instead, pulses are typically applied on all spins of the same species, with some possibility for frequency selectivity via pulse schemes such as BLEW12. Thus, a pulse diagram in NMR has one line per species: pulses are represented by boxes sitting on this line, with the understanding that this is applied to all spins of this species.

#### B. NMR Hamiltonian Overview

NMR experiments are performed on the molecular systems shown in Fig. S6, which we refer to  $^{13}\text{C-Tol}$  (or 9Q),  $^{13}\text{C-DMPB}$  (or 15Q) and  $\text{d}_6\text{-DMPB}$ , respectively. All experiments are performed in a nematic liquid crystal solvent leading to a partial alignment. Due to partial alignment, molecular tumbling is not fully isotropic resulting in the presence of residual intramolecular dipolar interactions [6]. The internal system Hamiltonian in rotating frame is then given by

$$H_{\text{int}} = H_{\text{CS}} + H_J + H_{\text{DD}}. \quad (\text{S4})$$

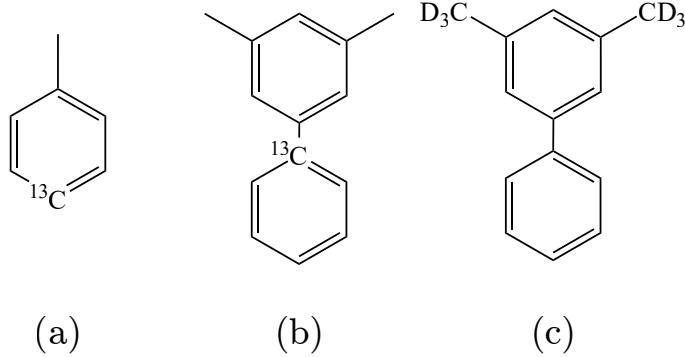


FIG. S6. Compounds used in OTOC and MQC experiments. (a) <sup>13</sup>C-Tol, (b) Compound 2 or <sup>13</sup>C-DMPB, (c) Compound 3 or d<sub>6</sub>-DMPB.

The chemical shift term  $H_{CS}$  describes the Zeeman interaction with the external magnetic field

$$H_{CS} = \sum_{i \in ^1\text{H}} \omega_i I_z^i, \quad (\text{S5})$$

where  $\omega_i$  is the chemically shifted resonance frequency relative to the <sup>1</sup>H channel carrier frequency. For the carbon nucleus, we assume the carrier frequency to be matched to its resonance frequency. Within the secular approximation the homo- and heteronuclear scalar coupling term  $H_J$  is given by

$$H_J = 2\pi \sum_{i,j \in ^1\text{H}, i < j} J_{ij}(I_x^i I_x^j + I_y^i I_y^j + I_z^i I_z^j) + 2\pi \sum_{i \in ^1\text{H}} J_{Ci} I_z^i S_z, \quad (\text{S6})$$

where  $J_{ij}$  represents the coupling constant between the  $i$ 'th and  $j$ 'th proton spin, whereas  $J_{Ci}$  represents the coupling constant between the <sup>13</sup>C and the  $i$ 'th proton spin. The dipolar interaction term  $H_{DD}$  represents through-space coupling of two spins. Within the secular approximation the dipolar interactions for homo- and heteronuclear spin pairs are given by

$$H_{DD} = 2\pi \sum_{i,j \in ^1\text{H}, i < j} D_{ij} \cdot (2I_z^i I_z^j - I_x^i I_x^j - I_y^i I_y^j) + 2\pi \sum_{i \in ^1\text{H}} D_{Ci} \cdot 2I_z^i S_z \quad (\text{S7})$$

Here,  $D_{ij}$  denote the dipolar coupling between the  $i$ 'th and the  $j$ 'th proton spin,  $D_{Ci}$  denotes the dipolar coupling between the <sup>13</sup>C and the  $i$ 'th proton spin. The effect of partial averaging is taken into account as follows

$$D_{ij} = -\frac{\mu_0 h}{4\pi} \gamma_i \gamma_j \left\langle \frac{\frac{3}{2}(\hat{r}_{ij} \cdot \hat{n})^2 - \frac{1}{2}}{r_{ij}^3} \right\rangle, \quad (\text{S8})$$

Here,  $\mu_0$  is the vacuum magnetic permeability and  $h$  is Planck's constant. The unit vector  $\hat{n}$  is aligned with the direction of the static external field, whereas  $\hat{r}_{ij}$  points from spin  $i$  to spin  $j$ ,  $r_{ij}$  represents the distance between the two spins. The angular brackets indicate averaging over the external and internal molecular motion.

For toluene, we assume that the aromatic ring tumbles rigidly about the liquid crystal director, while the methyl group rotates freely. The flat potential energy surface leads to rapid and isotropic methyl rotation, making the three protons equivalent and imposing approximate  $C_{2v}$  symmetry. With the ring rigidity and rapid methyl rotation decoupling external tumbling from internal motion, the residual dipolar couplings can be expressed in terms of the Saupe order tensor ( $S_{ab}$ ) [81]

$$\begin{aligned} D_{ij} &= -\frac{\mu_0 h}{4\pi} \gamma_i \gamma_j \left\langle \frac{\sum_{a,b=x,y,z} (\hat{r}_{ij} \cdot \hat{e}_a) \left[ \frac{3}{2}(\hat{n} \cdot \hat{e}_a)(\hat{n} \cdot \hat{e}_b) - \frac{1}{2}\delta_{ab} \right] (\hat{r}_{ij} \cdot \hat{e}_b)}{r_{ij}^3} \right\rangle \\ &= -\frac{\mu_0 h}{4\pi} \gamma_i \gamma_j \sum_{a,b=x,y,z} \left\langle \frac{(\hat{r}_{ij} \cdot \hat{e}_a)(\hat{r}_{ij} \cdot \hat{e}_b)}{r_{ij}^3} \right\rangle_{\text{internal}} \left\langle \frac{3}{2}(\hat{n} \cdot \hat{e}_a)(\hat{n} \cdot \hat{e}_b) - \frac{1}{2}\delta_{ab} \right\rangle_{\text{external}} \\ &= -\frac{\mu_0 h}{4\pi} \gamma_i \gamma_j \sum_{a,b=x,y,z} R_{ij,ab} S_{ab}, \end{aligned} \quad (\text{S9})$$

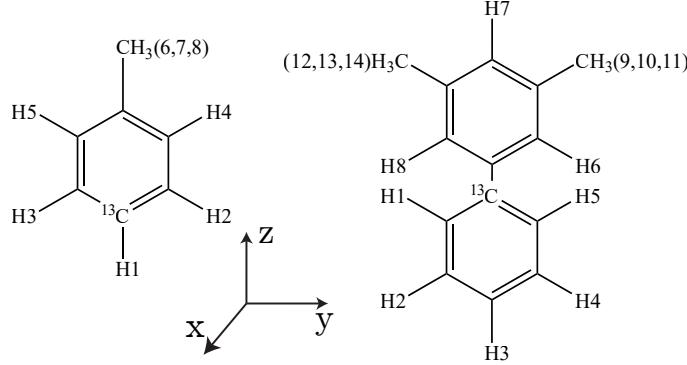


FIG. S7. Enumeration scheme for <sup>13</sup>C-Tol (left) and <sup>13</sup>C-DMPB (right). In both molecules, the  $z$ -axis of the molecular frame is defined as the long axis of the molecule. For <sup>13</sup>C-Tol, the  $x$ -axis is taken normal to the benzene ring plane. For <sup>13</sup>C-DMPB, the  $x$ -axis is taken normal to the unsubstituted phenyl ring.

The Saupe order captures the alignment of the molecular frame  $[\hat{e}_x, \hat{e}_y, \hat{e}_z]$  with respect to the magnetic field direction (see Fig. S7), whereas the quantities  $R_{ij,ab}$  represent the internal ordering of the molecule. Owing to the molecular symmetry, the off-diagonal elements of  $S_{ab}$  vanish. Given the tensor's traceless nature,  $S_{ab}$  is thus completely defined by its two independent diagonal components,  $S_{yy}$  and  $S_{zz}$ .

The tensor  $R_{ij,ab}$  for different spin pairs can be determined directly from the toluene structure. Between two spins on the aromatic ring,  $\mathbf{r}_{ij}$  is fixed in the molecular frame, making  $R_{ij,ab}$  straightforward to compute. For two methyl <sup>1</sup>H spins, the effect of isotropic rotation is included by averaging over all methyl orientations, simplifying the expression to

$$R_{ij,ab} = \frac{1}{r_{ij}^3} \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (\text{S10})$$

Here,  $r_{ij}$  represents the distance between two methyl <sup>1</sup>H spins. Similarly, for the dipolar coupling between a methyl <sup>1</sup>H spin and a spin on the ring,  $R_{ij,ab}$  can be computed numerically as

$$R_{ij,ab} = \left\langle \frac{(\hat{\mathbf{r}}_{ij} \cdot \hat{\mathbf{e}}_a)(\hat{\mathbf{r}}_{ij} \cdot \hat{\mathbf{e}}_b)}{r_{ij}^3} \right\rangle_{\text{rotation}} \quad (\text{S11})$$

where the angular brackets denote averaging over uniform methyl group orientations.

### C. OTOC Experiments

The formal definition of an out-of-time-order correlator (in the high-temperature limit) is given by [33]

$$C(t) = \text{Tr}\{B^\dagger(t)M^\dagger B(t)M\}. \quad (\text{S12})$$

With  $M$  representing a local operator for the “measurement” spin and  $B(t)$  representing a perturbation on the “butterfly” spins expressed in the Heisenberg picture ( $B(t) = U^\dagger(t)BU(t)$ ). If we identify  $M = S_x$  with the carbon spin angular momentum along  $x$  and  $B = V_H$  with a selective proton pulse sequence, we may explicitly connect the OTOC to a typical NMR experiment

$$C(t) = \text{Tr}\{S_x U^\dagger(t) V_H U(t) S_x U^\dagger(t) V_H^\dagger U(t)\}. \quad (\text{S13})$$

This equation now says the following: evolve the carbon magnetisation forward in time, apply a proton selective pulse sequence, reverse the initial evolution period, and measure the carbon FID (see Fig. S8). An outline of our experimental strategy is given below, explicit details of individual elements are given in the following:

1. Initialisation: After thermalization we selectively saturate all <sup>1</sup>H spins and initialize the state of the system in  $\rho_0 \propto S_x$  by applying a  $(\pi/2)_y$  pulse. The system is further subjected to a RF filter selecting a sub-ensemble with a tighter RF distribution. The second step is important to mitigate pulse error effects throughout the OTOC measurement.

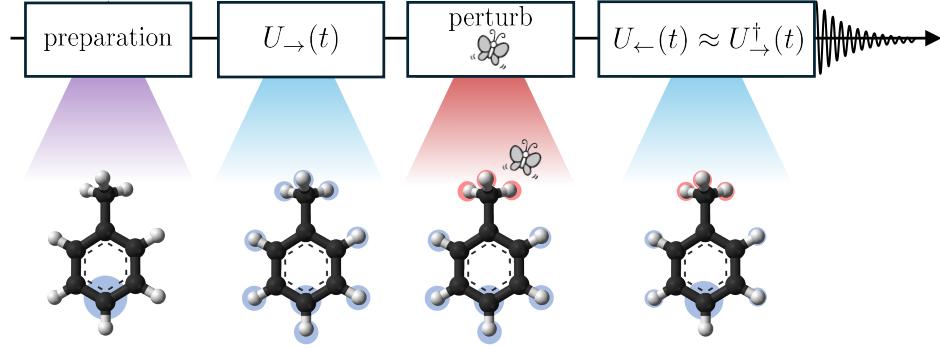


FIG. S8. Schematic protocol for OTOC-NMR illustrated for  $^{13}\text{C}$ -Tol. 0) a preparation period initializes polarization on a carbon spin (measurement spin). 1) An evolution period spreads the initial polarization across the carbon and proton network. 2) A selective perturbation of a subset of the protons acts as a butterfly. 3) The evolution period is approximately reversed aiming to return all polarization back to the initial carbon spin. 4) Measure the resulting carbon FID.

2. Forward evolution: To this end, we engineer periodic evolution under a dipolar double-quantum Hamiltonian utilizing a novel proton-carbon synchronized pulse element, denoted TARDIS described in detail below. To zeroth order the dynamics are governed by,  $U_{\rightarrow}(t_n) \approx \exp(-i\bar{H}_{\rightarrow}^{(0)}t_n)$ , where  $t_n = nt_c$ ,  $n$  is the cycles of the pulse sequence element and  $t_c$  the cycle time.
3. Local perturbation (butterfly): To locally perturb the system we devise a pulse sequence element that selectively rotates a subset of the  $^1\text{H}$  spins (we choose the methyl protons throughout):

$$V_{\text{H}} = \exp \left( -i\theta \vec{n} \cdot \sum_{i \in \text{methyl}} \vec{I}^i \right) \quad (\text{S14})$$

where  $\vec{n}$  is the rotation axis of the perturbation,  $\vec{I}^i = (I_x^i, I_y^i, I_z^i)$  is the angular momentum vector.

4. Backward evolution: To achieve backward evolution we engineer periodic evolution under a double-quantum Hamiltonian that to zeroth order satisfies

$$\bar{H}_{\leftarrow}^{(0)} = -\bar{H}_{\rightarrow}^{(0)}, \quad (\text{S15})$$

so that  $U_{\leftarrow}(t) \approx U_{\rightarrow}^{\dagger}(t)$ .

5. Readout: The OTOC is obtained as a modulation of the signal amplitude of the resulting  $^{13}\text{C}$  FID and can be directly obtained from the intensity of spectrum.

#### D. The TARDIS pulse sequences

We first describe the generation of dipolar double-quantum evolution blocks, which are vital components of the OTOC-NMR scheme. These blocks are suitable targets for OTOC experiments because they encode spatially meaningful information and offer time-reversal capabilities [82]. To simultaneously reverse both homo- and heteronuclear dipolar interactions—while also providing favorable compensation for finite pulse widths and pulse errors—we devised two novel double-resonance Floquet driving protocols, denoted as TARDIS-1 and TARDIS-2. The interaction between the spin system and the externally applied RF pulses is described by

$$H_{\text{rf}}(t) = \omega_{\text{rf}}^I(t) (I_x \cos \phi_I(t) - I_y \sin \phi_I(t)) + \omega_{\text{rf}}^S(t) (S_x \cos \phi_S(t) - S_y \sin \phi_S(t)) \quad (\text{S16})$$

Here  $\omega_{\text{rf}}^I$ ,  $\omega_{\text{rf}}^S$  are the nutation angular frequencies of the  $^1\text{H}$ ,  $^{13}\text{C}$  spins under the RF field. While  $\phi_I$ ,  $\phi_S$  denote the phase of pulses in  $^1\text{H}$ ,  $^{13}\text{C}$  channel respectively. Experimentally, the RF pulses are approximated as rectangular pulses, with a  $\pi/2$  pulse width of  $t_p = 10.5\mu\text{s}$ . The effect of the pulse sequence on the system is analyzed through an interaction picture decomposition

$$U(t) = U_{\text{rf}}(t) U_{\text{int}}(t) \quad (\text{S17})$$

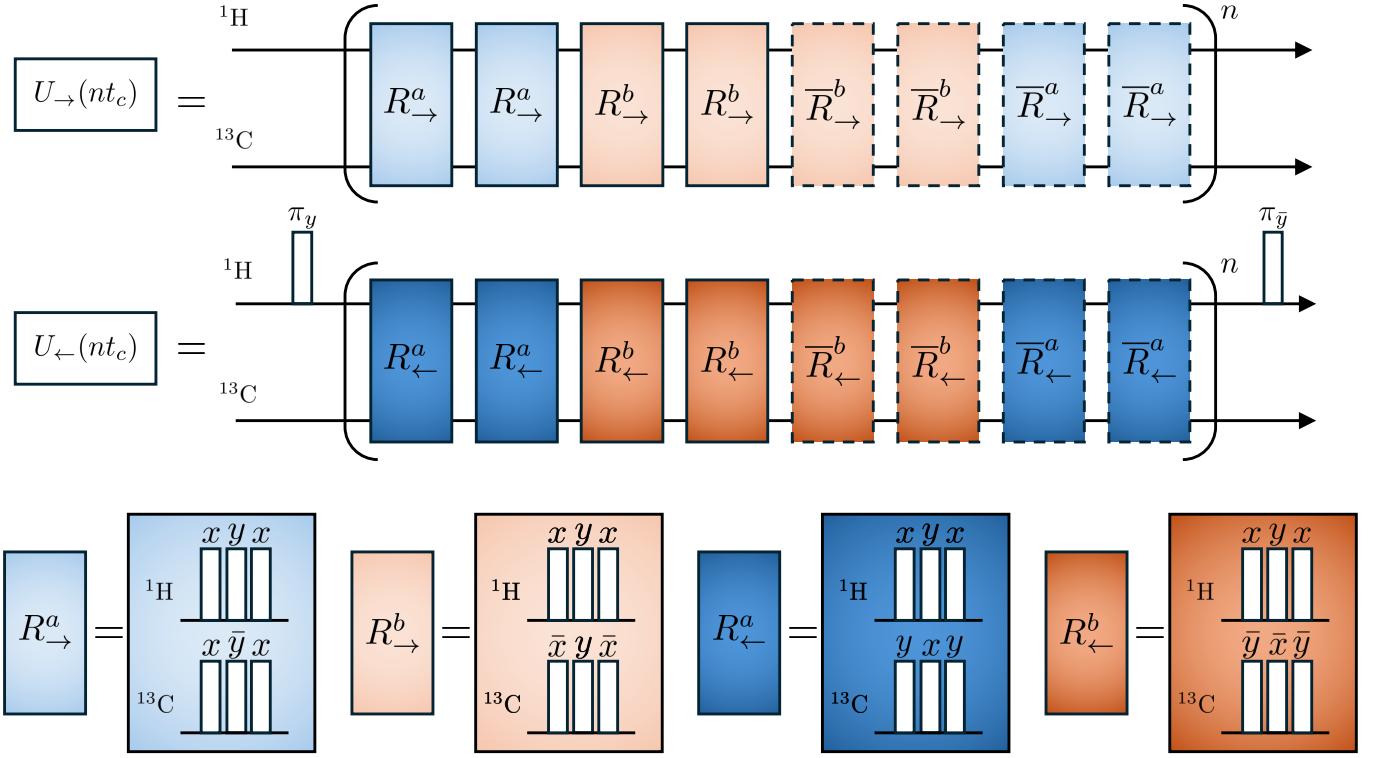


FIG. S9. TARDIS-1 pulse sequence. (a) TARDIS-1 forward- and backward-evolution pulse sequences. Each forward-evolution cycle consists of 8 synchronized composite double-resonance pulses applied on the  ${}^1\text{H}$  and the  ${}^{13}\text{C}$  channel:  $U_{\rightarrow}(t_c) = R_{\rightarrow}^a R_{\rightarrow}^a R_{\rightarrow}^b R_{\rightarrow}^b \bar{R}_{\rightarrow}^b \bar{R}_{\rightarrow}^a \bar{R}_{\rightarrow}^a$ . Reversal of TARDIS-1 is achieved by a  $\frac{\pi}{2}$  phase shift on the  ${}^{13}\text{C}$  pulses, combined with  ${}^1\text{H}$  channel  $\pi_y$  and  $\pi_{\bar{y}}$  pulses applied before and after the reverse cycles. (b) Definition of composite pulses  $R_{\rightarrow}^a$ ,  $R_{\rightarrow}^b$ ,  $R_{\leftarrow}^a$ , and  $R_{\leftarrow}^b$ . Each composite element consists of three synchronized  $\pi/2$  sub-pulses with alternating  $x/y$  phases on both channels, with matched nutation amplitudes  $\omega_{\text{rf}}^I = \omega_{\text{rf}}^S$  to ensure double-resonance conditions.

Where

$$U_{\text{rf}}(t) = \mathcal{T} \exp \left[ -i \int_0^t H_{\text{rf}}(t') dt' \right], \quad (\text{S18})$$

$$\tilde{U}_{\text{int}}(t) = \mathcal{T} \exp \left[ -i \int_0^t \tilde{H}_{\text{int}}(t') dt' \right]. \quad (\text{S19})$$

$\tilde{H}_{\text{int}}(t)$  represents the spin interactions in toggling frame,  $\tilde{H}_{\text{int}}(t) = U_{\text{rf}}^\dagger(t) H_{\text{int}} U_{\text{rf}}(t)$ . Under cyclicity we further have  $U_{\text{rf}}(nt_c) = \mathbf{1}$ , with  $t_c$  being the cycle period, and  $U(nt_c) = \tilde{U}_{\text{int}}^n(t_c)$ . According to Average Hamiltonian Theory (AHT) [83], the evolution operator over one period can be written as  $U(t_c) = \exp[-i\bar{H}t_c]$ , with  $\bar{H}$  being the effective or average Hamiltonian. Approximate expressions for  $\bar{H}$  may be generated via the Magnus expansion:  $\bar{H} = \bar{H}^{(0)} + \bar{H}^{(1)} + \bar{H}^{(2)} + \dots$ . The first three terms are given by:

$$\begin{aligned} \bar{H}^{(0)} &= \frac{1}{t_c} \int_0^{t_c} \tilde{H}_{\text{int}}(t') dt', \\ \bar{H}^{(1)} &= \frac{-i}{2t_c} \int_0^{t_c} dt_2 \int_0^{t_2} dt_1 \left[ \tilde{H}_{\text{int}}(t_2), \tilde{H}_{\text{int}}(t_1) \right], \\ \bar{H}^{(2)} &= \frac{-1}{6t_c} \int_0^{t_c} dt_3 \int_0^{t_3} dt_2 \int_0^{t_2} dt_1 \left\{ \left[ \tilde{H}_{\text{int}}(t_3), \left[ \tilde{H}_{\text{int}}(t_2), \tilde{H}_{\text{int}}(t_1) \right] \right] \right. \\ &\quad \left. + \left[ \tilde{H}_{\text{int}}(t_1), \left[ \tilde{H}_{\text{int}}(t_2), \tilde{H}_{\text{int}}(t_3) \right] \right] \right\}. \end{aligned} \quad (\text{S20})$$

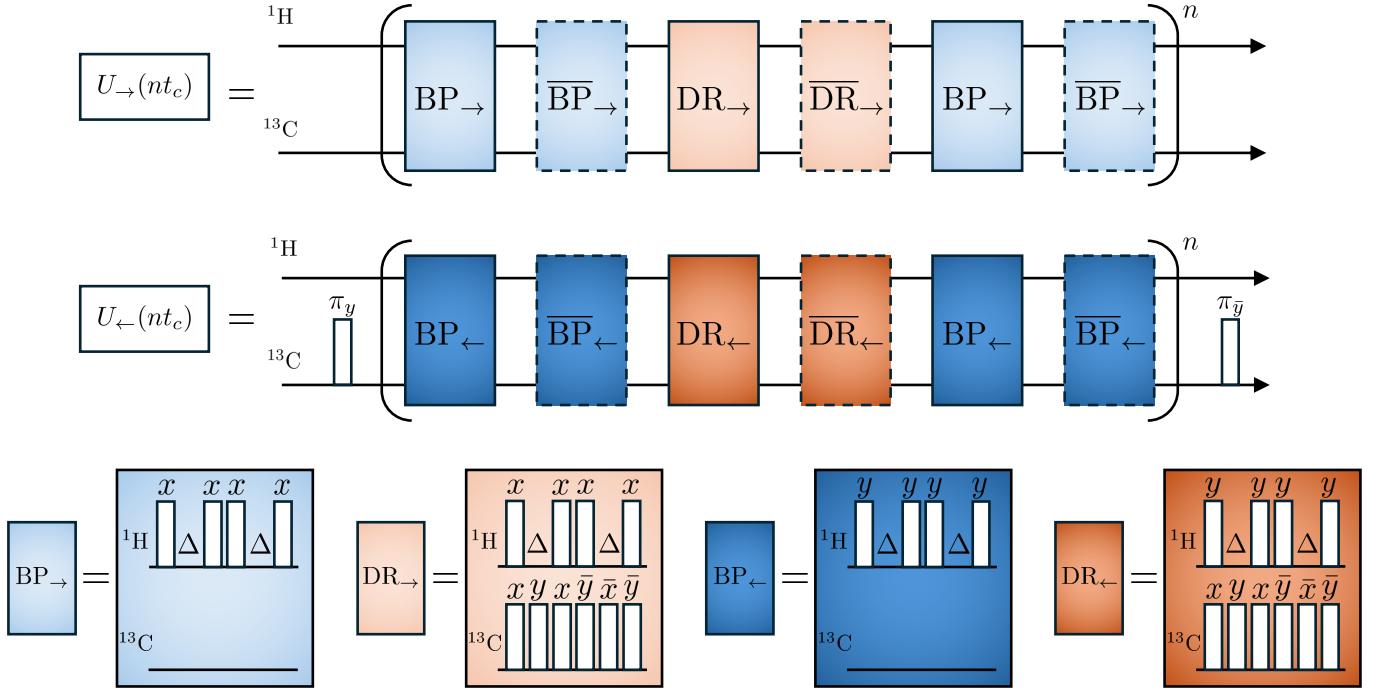


FIG. S10. TARDIS-2 pulse sequence. (a) TARDIS-2 forward- and backward-evolution pulse sequences. One forward-evolution cycle consists of six composite pulse blocks on the  $^1\text{H}$  and  $^{13}\text{C}$  channels:  $U_{\rightarrow}(t_c) = \text{BP}_{\rightarrow} \overline{\text{BP}}_{\rightarrow} \text{DR}_{\rightarrow} \overline{\text{DR}}_{\rightarrow} \text{BP}_{\rightarrow} \overline{\text{BP}}_{\rightarrow}$ . The BP block is adapted from the Baum-Pines sequence and acts only on  $^1\text{H}$ , while the DR block is a double-resonance element acting on both channels. Reversal of TARDIS-2 is obtained by a  $\pi/2$  phase shift on all  $^1\text{H}$  pulses and by applying  $^{13}\text{C}$  channel  $\pi_y$  and  $\pi_{\bar{y}}$  pulses before and after the reverse cycles. (b) Definition of the pulse blocks  $\text{BP}_{\rightarrow}$ ,  $\text{DR}_{\rightarrow}$ ,  $\text{BP}_{\leftarrow}$ , and  $\text{DR}_{\leftarrow}$ . Each BP block consists of four synchronized  $\pi/2$  sub-pulses on the  $^1\text{H}$  channel with interleaved delays of the same length of a  $\pi/2$  pulse ( $\Delta = t_p$ ), while each DR block contains alternating  $\pi/2$  sub-pulses on both channels with  $x/y$  phase patterns. The nutation amplitudes  $\omega_{\text{rf}}^I$  and  $\omega_{\text{rf}}^S$  are matched to maintain double-resonance conditions.

### 1. TARDIS-1

Fig. S9 gives an overview of the TARDIS-1 sequence and its building blocks. Each forward-evolution cycle consists of 8 composite pulses applied in synchrony on the proton and carbon channel:

$$U_{\rightarrow}(t_c) = R_{\rightarrow}^a R_{\rightarrow}^a R_{\rightarrow}^b R_{\rightarrow}^b \bar{R}_{\rightarrow}^b \bar{R}_{\rightarrow}^b \bar{R}_{\rightarrow}^a \bar{R}_{\rightarrow}^a \quad (\text{S21})$$

As shown in Fig. S9(b), the building blocks  $R_{\rightarrow}^a$  and  $R_{\rightarrow}^b$  are given by the composition:  $(\pi/2)_x(\pi/2)_{y/\bar{y}}(\pi/2)_x$ . The result is a  $\pi$  rotation along  $(\mp 1, 1, 0)$  affecting all Cartesian components of the spin Hamiltonian simultaneously. “Barred” elements  $\bar{R}_{\rightarrow}^b$ ,  $\bar{R}_{\rightarrow}^a$  refer to  $\pi$ -phase-shifted counterparts (phase shift occurs on both channels). As indicated in Fig. S9(b), the  $\frac{\pi}{2}$  pulses for protons and carbons are tuned to the same length. In practice this is achieved by matching the nutation amplitudes ( $\omega_{\text{rf}}^I = \omega_{\text{rf}}^S$ ). A standard AHT calculation shows that the zeroth order average Hamiltonian is given by

$$\bar{H}_{\rightarrow}^{(0)}/2\pi = \frac{2}{3} \sum_{i \in ^1\text{H}} (D_{Ci} + \frac{1}{2} J_{Ci}) \cdot I_z^i S_z + \frac{1}{\pi} \sum_{i,j \in ^1\text{H}, i < j} D_{ij} \cdot (I_x^i I_y^j + I_y^i I_x^j), \quad (\text{S22})$$

indicating homonuclear double-quantum and heteronuclear  $ZZ$  evolution. Time-reversal is achieved by applying a  $\frac{\pi}{2}$  phase shift to all pulses on the carbon channel of  $U_{\rightarrow}$ , combined with  $\pi_y$  and  $\pi_{\bar{y}}$  pulses on the proton channel applied before and after the reverse cycles:

$$U_{\leftarrow}(t_c) = \pi_y^I [R_{\leftarrow}^a R_{\leftarrow}^a R_{\leftarrow}^b R_{\leftarrow}^b \bar{R}_{\leftarrow}^b \bar{R}_{\leftarrow}^b \bar{R}_{\leftarrow}^a \bar{R}_{\leftarrow}^a] \pi_{\bar{y}}^I \quad (\text{S23})$$

Again, a standard AHT calculation shows that to zeroth order  $\bar{H}_{\rightarrow}^{(0)} = -\bar{H}_{\leftarrow}^{(0)}$ .

## 2. TARDIS-2

Fig. S10 gives an overview of the TARDIS-2 sequence and its building blocks. The forward evolution is composed of the concatenation of 6 pulse elements:

$$U_{\rightarrow}(t_c) = \text{BP}_{\rightarrow} \overline{\text{BP}}_{\rightarrow} \text{DR}_{\rightarrow} \overline{\text{DR}}_{\rightarrow} \text{BP}_{\rightarrow} \overline{\text{BP}}_{\rightarrow} \quad (\text{S24})$$

The  $\text{BP}_{\rightarrow}$  element is adapted from the Baum-Pines sequence [82]. The  $\text{DR}_{\rightarrow}$  double-resonance block is summarized in Fig S10(b). Both the  $\text{BP}_{\rightarrow}\overline{\text{BP}}$  block and the  $\text{DR}_{\rightarrow}\overline{\text{DR}}$  block are cyclic and generate a zeroth order average Hamiltonian of the type:

$$\bar{H}^{(0)}/2\pi = \eta_{\text{heter}} \sum_{i \in {}^1\text{H}} (D_{Ci} + \frac{1}{2}J_{Ci}) \cdot I_z^i S_z + \eta_{\text{homo}} \sum_{i,j \in {}^1\text{H}, i < j} D_{ij} \cdot (I_y^i I_y^j - I_x^i I_x^j). \quad (\text{S25})$$

Similar to TARDIS-1, the TARDIS-2 average Hamiltonian generates homonuclear double-quantum and heteronuclear  $ZZ$  evolution. The respective scaling factor are  $\{\eta_{\text{heter}}, \eta_{\text{homo}}\} = \{0, 1\}$  for the  $\text{BP}_{\rightarrow}\overline{\text{BP}}$  block and  $\{\eta_{\text{heter}}, \eta_{\text{homo}}\} = \{\frac{2}{3}, 1\}$  for  $\text{DR}_{\rightarrow}\overline{\text{DR}}$  block. By interleaving these two blocks, the scaling factors are averaged, leading to the overall scaling factor for one TARDIS-2 forward cycle:  $\{\eta_{\text{heter}}, \eta_{\text{homo}}\} = \{\frac{2}{9}, 1\}$  (This may be seen by utilizing the methods developed in reference [84]). Compared to TARDIS-1, TARDIS-2 has favorable properties in systems where the  ${}^{13}\text{C}$  is directly bonded to a  ${}^1\text{H}$  (for example,  ${}^{13}\text{C}$ -Tol). In such cases a strong  ${}^{13}\text{C}$ - ${}^1\text{H}$  coupling can partially truncate the homonuclear dipolar coupling, leading to an undesirable OTOC slowdown.

Time-reversal is achieved by applying a  $\frac{\pi}{2}$  phase shift to all pulses on the proton channel of  $U_{\rightarrow}$ , combined with  $\pi_y$  and  $\pi_{\bar{y}}$  pulses on the carbon channel applied before and after the reverse cycles:

$$U_{\leftarrow}(t_c) = \pi_y^S [\text{BP}_{\leftarrow} \overline{\text{BP}}_{\leftarrow} \text{DR}_{\leftarrow} \overline{\text{DR}}_{\leftarrow} \text{BP}_{\leftarrow} \overline{\text{BP}}_{\leftarrow}] \pi_{\bar{y}}^S \quad (\text{S26})$$

A straightforward calculation shows that the zeroth order average Hamiltonians are related by:  $\bar{H}_{\rightarrow}^{(0)} = -\bar{H}_{\leftarrow}^{(0)}$ .

## 3. TARDIS-2 sequences with on-site field

A uniform on-site field can be introduced for the  ${}^1\text{H}$  by the following phase shifting scheme (adapted from [85]): In the  $m$ -th cycle of the forward TARDIS-2 sequence, the earlier  $\text{BP}_{\rightarrow}\overline{\text{BP}}_{\rightarrow}$  block is phase shifted by  $(3m-3)\phi$  on the  ${}^1\text{H}$  channel, the  $\text{DR}_{\rightarrow}\overline{\text{DR}}_{\rightarrow}$  block by  $(3m-2)\phi$ , and the later  $\text{BP}_{\rightarrow}\overline{\text{BP}}_{\rightarrow}$  block by  $(3m-1)\phi$ . The evolution operator for each cycle is given by:

$$\begin{aligned} U(t_c, 0) &= \exp(i \cdot 2\phi I_z) \exp\left(-i \bar{H}_{\text{BP}_{\rightarrow}\overline{\text{BP}}}^{(0)} \cdot t_c / 3\right) \exp(-i \cdot 2\phi I_z) \\ &\quad \cdot \exp(i\phi I_z) \exp\left(-i \bar{H}_{\text{BP}_{\rightarrow}\overline{\text{BP}}}^{(0)} \cdot t_c / 3\right) \exp(-i\phi I_z) \\ &\quad \cdot \exp\left(-i \bar{H}_{\text{BP}_{\rightarrow}\overline{\text{BP}}}^{(0)} \cdot t_c / 3\right) \\ &\approx \exp(i \cdot 3\phi I_z) \exp\left[-i \left(\bar{H}_{\rightarrow}^{(0)} + \frac{3\phi}{t_c} I_z\right) t_c\right] \\ U(2t_c, t_c) &= \exp(i \cdot 3\phi I_z) U(t_c, 0) \exp(-i \cdot 3\phi I_z) \\ &\quad \dots \\ U(nt_c, (n-1)t_c) &= \exp(i \cdot 3(n-1)\phi I_z) U(t_c, 0) \exp(-i \cdot 3(n-1)\phi I_z) \end{aligned} \quad (\text{S27})$$

Hence the  $n$ -cycle propagator is:

$$U(nt_c, 0) = U(nt_c, (n-1)t_c) \dots U(2t_c, t_c) U(t_c, 0) \approx \exp(i \cdot 3n\phi I_z) \left[ \exp\left[-i \left(\bar{H}_{\rightarrow}^{(0)} + \frac{3\phi}{t_c} I_z\right) t_c\right] \right]^n \quad (\text{S28})$$

$$= \exp(i \cdot 3n\phi I_z) \exp\left(-i \bar{H}_{\rightarrow}^{(0)\prime} \cdot nt_c\right) \quad (\text{S29})$$

with the effective zero-order average Hamiltonian  $\bar{H}^{(0)\prime} = \bar{H}^{(0)} + \frac{3\phi}{t_c} I_z$ . The prefactor  $\exp(i \cdot 3n\phi I_z)$  can be compensated by phase shifting the perturbation sequence by  $3n\phi$ . An inverse on-site field can be introduced for the backward

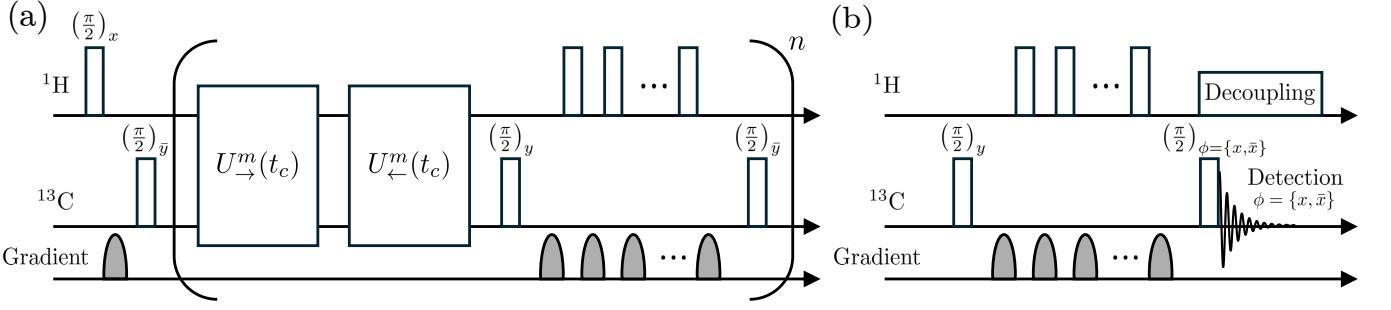


FIG. S11. Pulse sequences for (a)Initial state preparation and (b)Readout. (a) The sequence prepares the system in an initial state  $\rho_0 \propto S_x$  by rotating the  $^{13}\text{C}$  thermal magnetization and destroying the  $^1\text{H}$  magnetization. The  $^{13}\text{C}$  magnetization is then subjected to an RF pre-selection filter consisting of  $m = 10$  TARDIS forward/backward cycles. This filtering step selects ensemble members with small RF inhomogeneity deviations and enhances homogeneity across the sample. Subsequently, a series of gradient and random proton pulses selects pure  $^{13}\text{C}$   $x$ -magnetization. The process is repeated for  $n = 2$ . (b) Before detection, the density matrix is filtered to retain the  $S_x$  component using the same scheme as the second half of the initialization filter. The  $^{13}\text{C}$  FID is then recorded under proton decoupling.

evolution using an analogous phase-shifting scheme. Finally, the zero-order average Hamiltonian resulting from the on-site field TARDIS-2 sequence is given by

$$\bar{H}_{\rightarrow}^{(0)}/2\pi = \frac{2}{9} \sum_{i \in ^1\text{H}} (D_{Ci} + \frac{1}{2}J_{Ci}) \cdot I_z^i S_z + \sum_{i,j \in ^1\text{H}, i < j} D_{ij} \cdot (I_y^i I_y^j - I_x^i I_x^j) + h I_z \quad (\text{S30})$$

With  $h = \frac{3\phi}{2\pi t_c}$ .

### E. Initial State Preparation and RF inhomogeneity

The system initially starts out at thermal equilibrium,  $\rho_{\text{eq}} = \epsilon_{^1\text{H}} I_z + \epsilon_{^{13}\text{C}} S_z$ , where  $\epsilon_{^1\text{H}}$ ,  $\epsilon_{^{13}\text{C}}$  is the polarization of  $^1\text{H}$  and  $^{13}\text{C}$  respectively. From here, we want to prepare the system in  $\rho_0 \propto M = S_x$ . Moreover, there is a relatively large RF inhomogeneity across our sample (i.e. molecules at different physical locations experience a different strength of the applied RF pulses, which results in miscalibrated gates). To account for this, we want to further pre-select a smaller sample sub-region with a relatively consistent RF strength, by decohering the spin population outside this region to the identity.

We achieve both of the above goals via the pulse sequence scheme shown in Fig. S11(a). An initial  $\pi/2$  pulse on the proton channel followed by a gradient pulse spoils the proton magnetization. Afterwards the  $^{13}\text{C}$  magnetization is rotated onto the  $x$ -axis and the whole system subjected to an RF pre-selection filter. The filter consists of  $m = 10$  cycles of TARDIS (TARDIS-1 for  $^{13}\text{C}$ -DMPB experiments and TARDIS-2 for  $^{13}\text{C}$ -Tol experiments) forward evolution followed by  $m = 10$  cycles of backward evolution. This echo-based filter selectively saturates ensemble members with large RF deviations from the nominal reference. Subsequently, a string of gradient and random proton pulses selects pure  $^{13}\text{C}$   $x$ -magnetization. The combination of gradient and proton pulses is used to ensure destruction of all  $^1\text{H}$  coherences after the TARDIS filter. The process is repeated twice [ $n = 2$  in Fig. S11(a)] to further enhance RF homogeneity.

To model the filtered RF inhomogeneity in the OTOC experiment, we explicitly measure the RF distribution profile produced by the NMR coil, and model the pre-selection sequence. To determine the RF amplitude distribution of the  $^1\text{H}$  coil and the  $^{13}\text{C}$  coil, we prepare 12.5% v/v toluene- $^{13}\text{C}$  in benzene- $d_6$  of the same volume as other samples used in OTOC experiment. The RF amplitude inhomogeneity was obtained by Fourier transform of the nutation curve, following the method in Ref. [86]. For both the  $^{13}\text{C}$  channel and the  $^1\text{H}$  channel, we observe that the profile peaks at around 25kHz, with a FWHM around 15% [Fig. S12(top)]. We thus approximate the pre-filtered fractional RF error  $\epsilon$  for both  $^{13}\text{C}$  and  $^1\text{H}$  by a Gaussian with standard deviation 0.064, i.e.  $p_{\text{coil}}(\epsilon) = \mathcal{N}(\mu = 0, \sigma = 0.064)$ . To model the effect of filtering, we calculate an  $\epsilon$  dependent damping factor  $F(\epsilon)$ , corresponding to the return probability of the  $X$  polarization on the  $^{13}\text{C}$  spin after a single block of the sequence in Fig. S11(a)

$$F(\epsilon) = \text{tr} [S_x(U_{\leftarrow}(\epsilon))^{10}(U_{\rightarrow}(\epsilon))^{10}S_x(U_{\rightarrow}^{\dagger}(\epsilon))^{10}(U_{\leftarrow}^{\dagger}(\epsilon))^{10}] \quad (\text{S31})$$

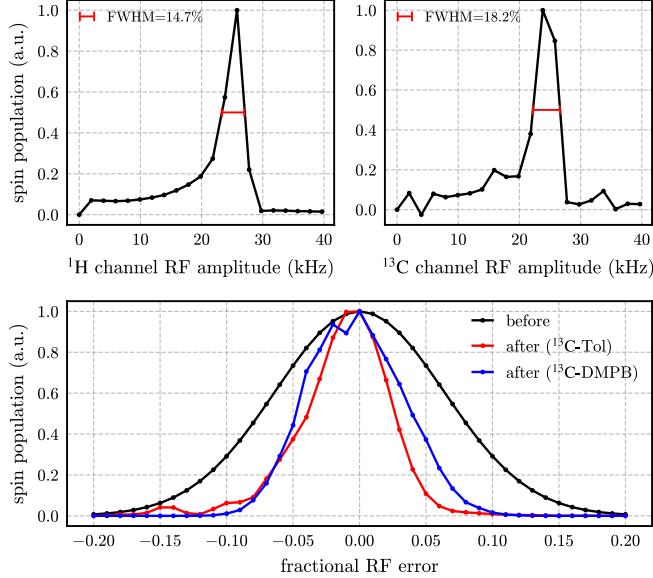


FIG. S12. Upper left: RF inhomogeneity profile of <sup>1</sup>H coil. Upper right: RF inhomogeneity profile of <sup>13</sup>C coil. Bottom: fractional RF error for <sup>13</sup>C-Tol and <sup>13</sup>C-DMPB before and after the pre-selection sequence. For <sup>13</sup>C-DMPB , the filtered RF error distribution is calculated using the dipolar couplings of the  $u_4$  PMF surface.

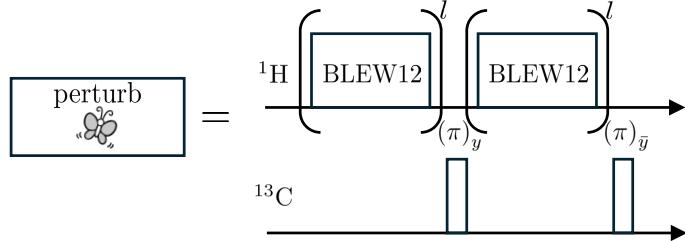


FIG. S13. Perturbation pulse sequence. The perturbation sequence is based on the homonuclear dipolar decoupled pulse sequence BLEW12. Two <sup>13</sup>C  $\pi$  pulses before and after the later BLEW12 block serves to refocus the heteronuclear dipolar coupling.  $l = 2$  across all experiments.

Here  $U_{\rightarrow}(\epsilon)/U_{\leftarrow}(\epsilon)$  represent TARDIS forward/backward evolution with RF error of  $\epsilon$ . The 'filtered' RF error distribution  $p(\epsilon)$  can then be written as:

$$p(\epsilon) = p_{\text{coil}}(\epsilon) \times F^2(\epsilon) \quad (\text{S32})$$

Note that  $F(\epsilon)$  is squared since the filtering cycle is executed twice. The 'filtered' RF error distribution is shown in Fig. S12(bottom).

## F. Perturbation

For both <sup>13</sup>C-Tol and <sup>13</sup>C-DMPB we choose the methyl protons as the butterfly spins. To selectively address the methyl group we utilize their chemical shift difference with respect to the remaining ring protons. As perturbation we choose a local rotation of the methyl protons, which may be achieved through the application of a BLEW12 [87] pulse sequence element (see Fig. S13). The BLEW12 elements on the <sup>1</sup>H channel removes their dipolar couplings and generates the following zeroth order average Hamiltonian:

$$\bar{H}_{\text{BLEW12}}^{(0)} = \eta \left[ 2\pi \sum_{i \in ^1\text{H}} \left( D_{Ci} + \frac{1}{2} J_{Ci} \right) (\vec{n} \cdot \vec{I}^i) S_z + \sum_{i \in ^1\text{H}} \omega_i (\vec{n} \cdot \vec{I}^i) \right] \quad (\text{S33})$$

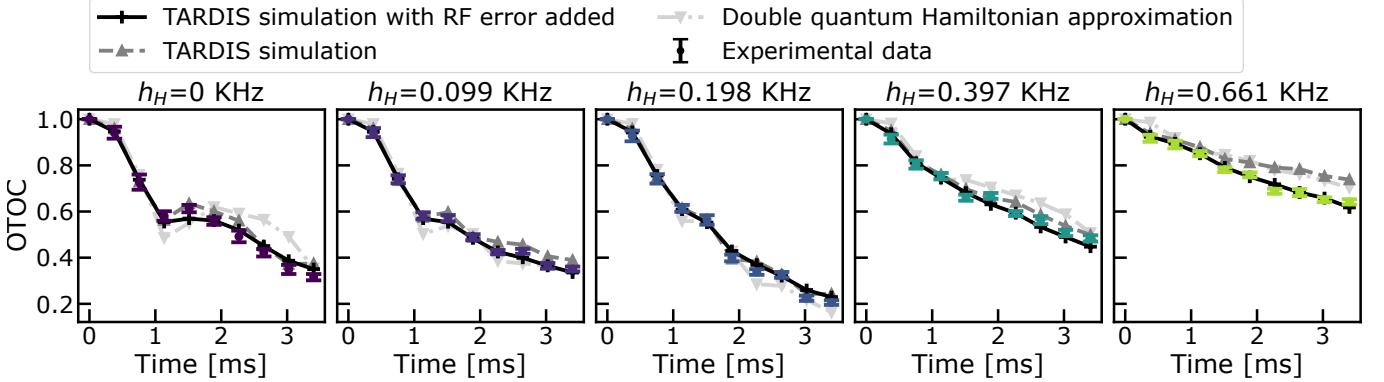


FIG. S14. Extended data showing OTOCs taken with five different on-site fields, using the experimental protocol defined in Sec. III D 3. This data was used in the generation of Fig. 2(e) of the main text. Colored points denote experimental data (error bars are  $2\sigma$  confidence intervals). Lines denote three levels of simulation: the double quantum Hamiltonian approximation (i.e. zeroth-order Magnus expansion), a full simulation of the TARDIS-2 sequence (and accurate simulation of BLEW12 butterfly pulse) without RF inhomogeneity, and the same, averaged over the externally-determined RF inhomogeneity given in Fig. S12(bottom, blue).

Where  $\eta = \frac{2\sqrt{5}}{3\pi}$  is the scaling factor and  $\vec{n} = \frac{1}{\sqrt{5}}(0, 2, 1)$  is the effective rotation axis. The sandwich of  $^{13}\text{C}$   $\pi$  pulses before and after the second BLEW12 block refocus the heteronuclear dipolar couplings. In total we then find

$$V_{\text{H}} = \exp \left( -i\eta \times 2l \sum_{i \in ^1\text{H}} \omega_i (\vec{n} \cdot \vec{\sigma}_i) \cdot 12t_p \right), \quad (\text{S34})$$

where  $t_p$  is the duration of a single pulse, and  $12t_p$  is the length of one BLEW12 cycle. The factor  $2l$  corresponds to the total number of BLEW12 cycles applied, and  $l$  is set to 2 in all experiments. If the carrier frequency is set to approximately match the ring proton resonances, equation S14 follows.

#### G. Readout

Before detection, the density matrix is once again filtered to isolate its  $S_x$  component. The filtering scheme is identical to the second half of the initialization filter. The carbon FID is recorded under proton decoupling (see Fig. S11(b)). Since each molecule contains only one  $^{13}\text{C}$ , proton decoupling leads to a single peak in the carbon spectrum. After phase-correction, the magnitude of the carbon magnetization is extracted by fitting a Lorentzian line shape function to the spectrum.

#### H. Additional OTOC and learning experiments

Here we give additional experimental datasets that were not added to the main text due to space constraints. In Fig. S14 we show an extended set of OTOCs with different onsite potentials, and plot behind this classical approximations of these curves at different levels of theory. In Fig. S15, we expand on this by showing an error budget for the curves in Fig. S14, by dividing the error into different levels of theory. We observe that our classical modelling reproduces the experimental data well across the entire curve; indeed the dataset shown in the main text has the largest residual error of the datasets shown here. Finally, in Fig. S16, we demonstrate using either a single or five OTOC curves to learn the two free components of the Saupe tensor (Eq. (S9)), which were otherwise determined as part of the MQC fitting procedure described in Sec. III J. We observe that when using a single dataset our cost function is relatively insensitive to a correlated change between the two components, which indicates that the data is low in quantity. However, this disappears when we use all datasets, and we obtain 95% confidence intervals that are only a factor 2 larger than the reference data.

In addition, to test the reproducibility of our experiments, we prepared a new  $^{13}\text{C}$ -Tol sample and repeated the HMQC experiment. Fitting of the new HMQC spectrum shows that, compared to the original sample, the components of the Saupe order tensor differ by less than 0.005, while the chemical shift terms and  $J_{C1}$  vary by less than 30 Hz. We attribute these small differences to variations in the toluene concentration: since toluene is highly volatile and the

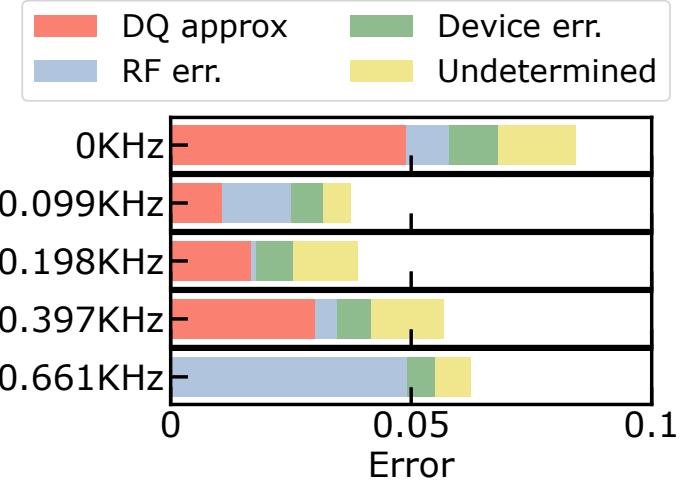


FIG. S15. Budgeting the error in the simulations of Fig. S14. For each of the five experiments, we split the observed error between the DQ approximation and the experimental data into components (calculated by measuring the improved RMSE as the accuracy of the simulation increases). After subtracting the observed error bars from the experimental data, the residual error is marked as undetermined.

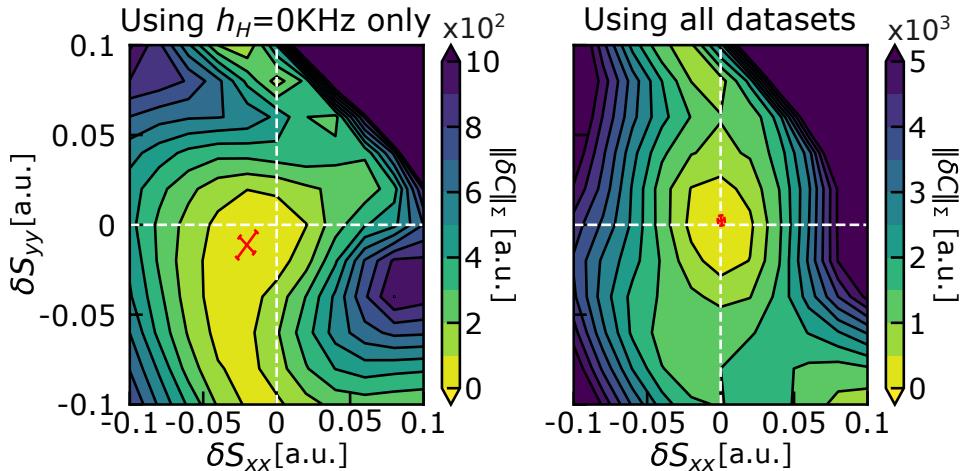


FIG. S16. Learning the two free components of the Saupe tensor (Eq. (S9)) from either (left) a single dataset, or (right) all data shown in Fig. S14. We show contour plots of the covariance-weighted squared error as a function of the two free components of Eq. (S9),  $S_{xx}$  and  $S_{yy}$ , in terms of their shifts from the reference values given in Tab. III (white dashed lines). The minima of the two surfaces is found by fitting a quadratic function and marked in red. The size of the red bars give  $2\sigma$  intervals along the principal components of the covariance matrix, calculated by resampling the experimental data and reperforming the fit.

sample preparation involves mixing toluene and EBBA at 358 K, different degrees of toluene loss may occur during the preparation of separate samples. Subsequently, we remeasured the dataset shown in Fig. 2(c) of the main text using the new sample, and the comparison is presented in Fig. S17. Although the two datasets were acquired a year apart—during which drift in the magnetic field, amplifier performance, and other experimental conditions could have occurred—the agreement between experiment and simulation remains consistent after a simple recalibration of the spectrometer. These results demonstrate the high reproducibility of our OTOC experiment.

## I. HETCOR Experiment

$^1\text{H}$  chemical shift information is a prerequisite for  $^{13}\text{C}$ -DMPB OTOC-learning experiment. Because of large number of transitions, no resolved peak is observed in  $^1\text{H}$  single quantum spectrum, all the readout needs to be performed in  $^1\text{H}$  decoupled  $^{13}\text{C}$  single quantum spectrum. Thus, we performed heteronuclear correlation (HETCOR) experiment, with a sequence adapted from [88], shown in Fig. S20(c). The obtained spectrum is shown in Fig. S18.

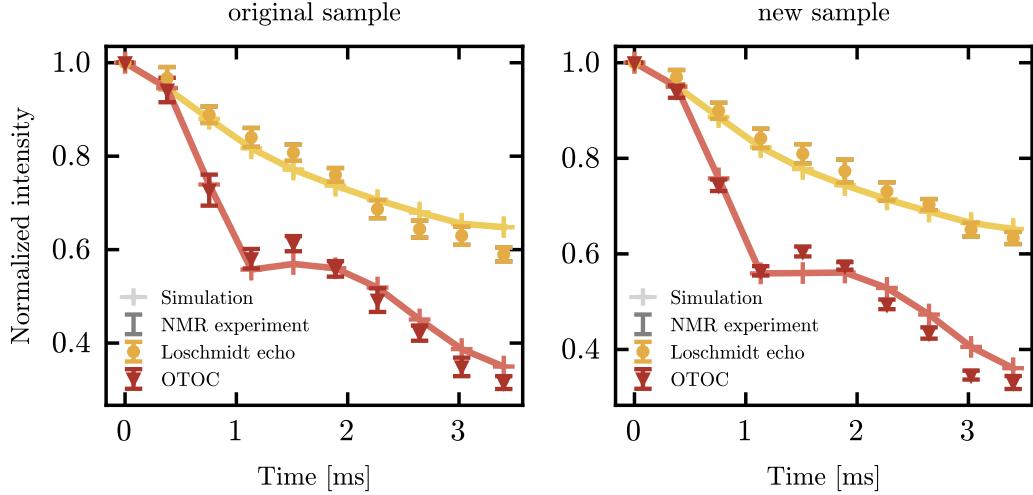


FIG. S17. Retaken dataset shown in Fig. 2(c) of the main text on a new sample (right), compared with the dataset from the original sample (left, replotted from the main text), demonstrating reproducibility. Experimental NMR data points are shown with  $2\sigma$  confidence intervals (CIs) and compared to numerical simulations (lines).

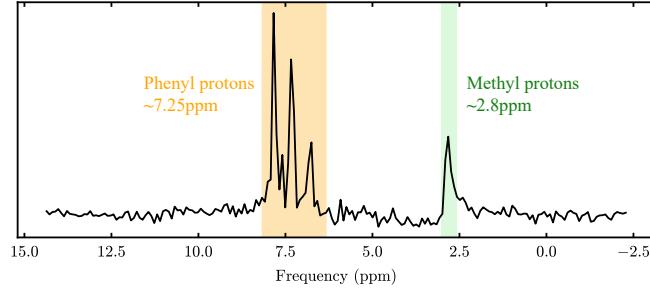


FIG. S18. HETCOR spectrum of  $^{13}\text{C}$ -DMPB. CP time = 2.5 ms; 100  $t_1$  increments, one scan each.  $\pi/2$  pulse = 10.5  $\mu\text{s}$ , giving an effective  $F_1$  width of 8363 Hz (BLEW12-scaled).  $F_2$  width = 15 kHz. The 1D projection was obtained by integrating the  $F_2$  peak, zero-filling to 200 points, and Fourier transforming along  $F_1$ . The peak at 2.8 ppm (shaded in green) is assigned to the methyl protons, whereas the peaks around 7.25 ppm (shaded in yellow) are attributed to the ring protons.

The peak at 2.8 ppm was assigned to methyl protons, while the three resolved peaks around 7.25 ppm correspond to phenyl protons. Assignment of these peaks to individual protons was not successful. We attempted HETCOR experiment with varying cross polarization buildup times to obtain additional information for peak assignment. However, since only the enriched carbon is observable in the  $^{13}\text{C}$  spectrum, insufficient information was obtained for reliable peak assignment. Therefore, we approximate that all the phenyl protons have chemical shift of 7.25 ppm. In OTOC experiment, We chose the carrier frequency to be on resonance with the phenyl protons. The relative chemical shift for  $^{13}\text{C}$ -DMPB is summarized in Table I.

TABLE I. Relative chemical shift (in Hz) for  $^{13}\text{C}$ -DMPB, determined from the HECTOR experiment,

Parameter	Value
$\omega_1/2\pi, \omega_2/2\pi, \dots, \omega_8/2\pi$	0
$\omega_9/2\pi, \omega_{10}/2\pi, \dots, \omega_{14}/2\pi$	-2225

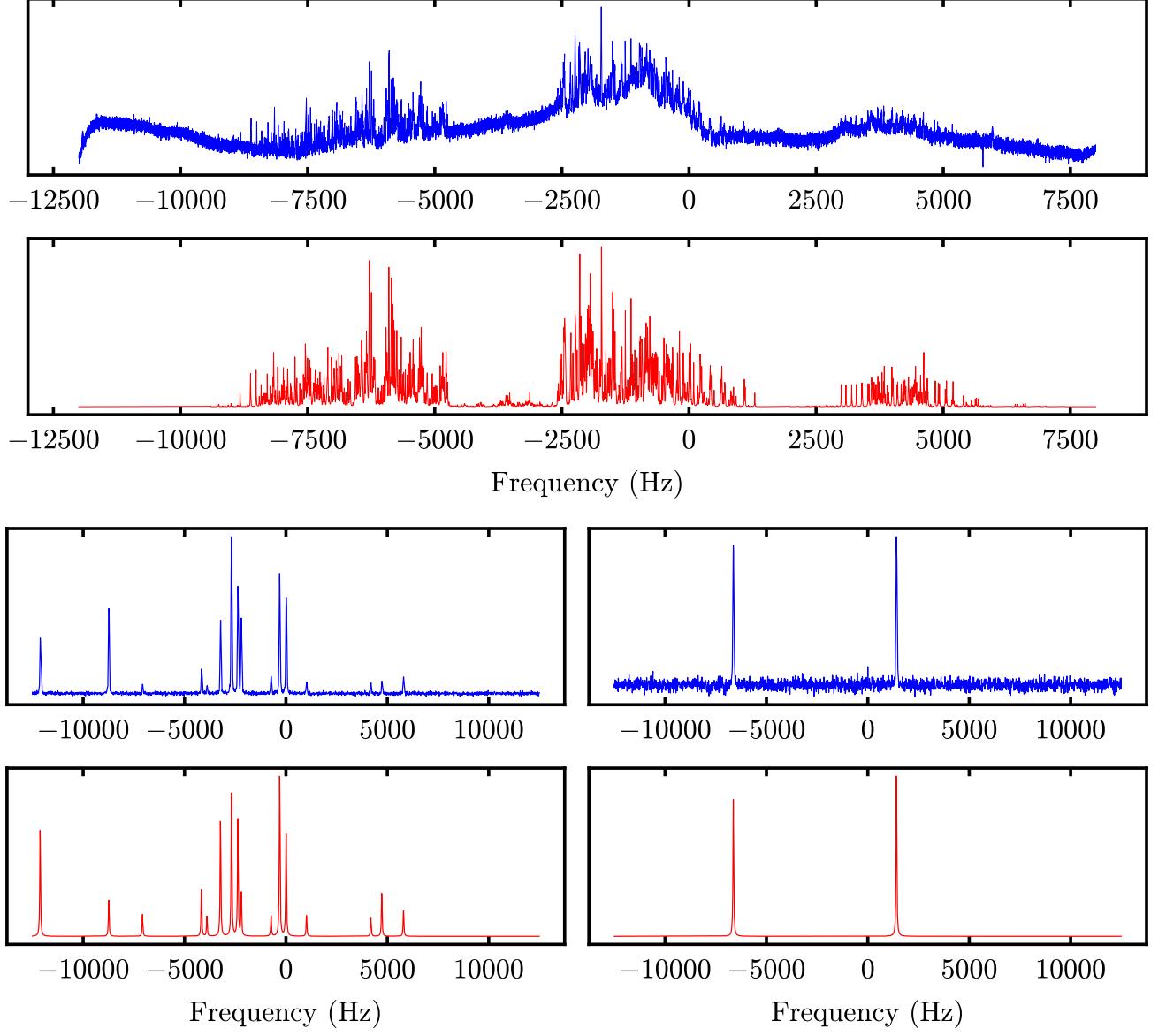


FIG. S19. Experimental (blue) and simulated (red) single-quantum spectrum (top), and HMQC spectra for  $\Delta m_S = 0$ ,  $\Delta m_I = 7$  (bottom left), 8 (bottom right). Single-quantum spectrum was taken to be one-shot. 32k data points are acquired and the spectral width is 20kHz. For HMQC experiments,  $\tau = 10\text{ ms}$ , 2500  $t_1$  increments, 6 scans per increment. Spectrum width is 25kHz for  $F_1$  and 20kHz for  $F_2$ . The magnitude spectrum of  $F_1$  is obtained by summing the absolute values of the Fourier transforms across different points in  $F_2$ . The average linewidth is 60Hz.

### J. MQC Experiments

To cross-validate the parameters inferred from OTOC-NMR Hamiltonian learning and to establish full or partial ground-truth reference data we employ a spectral estimation strategy based on multiple-quantum-coherence (MQC) NMR [34, 89]. In general spectral estimation via a standard pulse-acquire spectrum is difficult due to the large number of observable NMR transitions fulfilling the selection rule  $\Delta M_z = 1$ . To simplify spectral complexity and reduce spectral crowding multiple-quantum techniques aim to indirectly observe higher-order NMR transitions with  $\Delta m_I \neq 1$ .

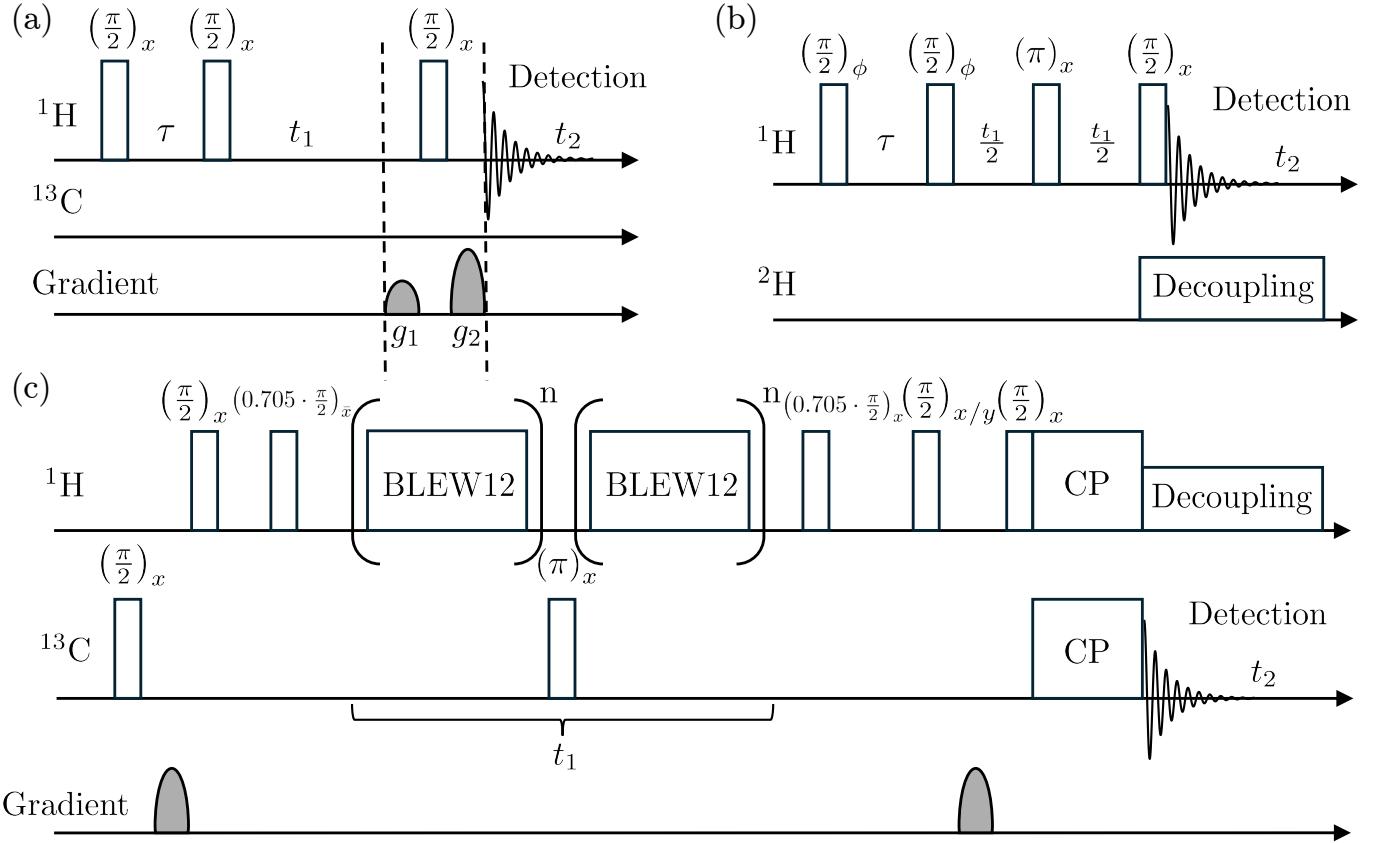


FIG. S20. Pulse sequences for (a) HMQC experiment performed on <sup>13</sup>C-Tol. (b) Proton MQC experiment performed on d<sub>6</sub>-DMPB. (c) HETCOR experiment performed on <sup>13</sup>C-DMPB. (a) High coherence orders are generated using two  $\pi/2$  pulse with a free evolution in between. Free evolution during the  $t_1$  period leads to phase-encoding onto individual quantum coherence terms. A subsequent  $\pi/2$  pulse converts coherence order operators back into observable 1Q terms which are read out during the  $t_2$  period. Two gradient pulses are applied to filter out signal components originating from a specific coherence order operator. The gradient strength may be adjusted according to the selection rule  $g_2 : g_1 = \Delta m_I + \frac{\gamma_S}{\gamma_I} \Delta m_S$ . (b) Similar to (a), high coherence orders are generated using two  $\pi/2$  pulse separated by a free-evolution period. The method of time proportional phase incrementation (TPPI) [91] is employed, in which the phase  $\phi$  is incremented proportionally with  $t_1$ , introducing an artificial offset term into the  $t_1$  evolution period, enabling separation of spectrum from different coherence orders. The central  $\pi$  pulse within  $t_1$  refocuses both <sup>1</sup>H chemical shifts and the <sup>1</sup>H – <sup>2</sup>H dipolar couplings evolution. To enhance the <sup>1</sup>H signal energy during detection, double-quantum decoupling is applied via continuous-wave irradiation at the center of <sup>2</sup>H spectrum, effectively suppressing heteronuclear dipolar couplings. (c) During the  $t_1$  evolution period, the <sup>1</sup>H magnetization evolves under chemical shift interactions. The BLEW-12 multiple pulse sequence removes <sup>1</sup>H-<sup>1</sup>H homonuclear coupling, while a <sup>13</sup>C  $\pi$  pulse refocuses the <sup>13</sup>C-<sup>1</sup>H heteronuclear dipolar coupling. Two  $0.705 \cdot \frac{\pi}{2}$  pulses sandwiching the BLEW-12 multiple-pulse train effectively rotate the precession plane into the  $xy$  plane. Following a gradient z-filter that separately selects the cosine- and sine-modulated component, the magnetization is transferred to <sup>13</sup>C via cross-polarization and detected during  $t_2$ .

### 1. MQC spectra for Toluene

Spectral estimation of the residual dipolar coupling parameters and chemical shifts for <sup>13</sup>C-Tol was performed using the heteronuclear MQC (HMQC) experiments shown in Fig. S20(a). Using the analysis in section III B and taking the molecular equilibrium positions from [34] (summarized in Table II), the residual dipolar coupling are completely determined by the molecular Saupe order tensor (see equation S9). Except for the strong heteronuclear J coupling  $D_{C1}$ , all other J couplings are taken from solution-state NMR data ignoring small J anisotropies [34, 90]. As a result only 6 free parameters remain for spectral estimation.

It may be shown that the combined information contained in the  $\Delta m_I = 7$ ,  $\Delta m_S = 0$  and  $\Delta m_I = 8$ ,  $\Delta m_S = 0$  HMQC spectra is sufficient to determine all free parameters. We then fit the free parameters to minimize the RMS deviation between the experimental and simulated peak positions. The estimated interaction parameters are summarized in Table III. A comparison of numerically simulated HMQC spectra (red), generated using the estimated parameters, with the experimental data (blue) is shown in Fig. S19.

TABLE II. Atom coordinate for Toluene-4-<sup>13</sup>C.

Nucleus	x/Å	y/Å	z/Å
H1	0.0	0.0	0.0
H2	0.0	-2.1486	1.2417±0.005
H3	0.0	2.1486	1.2417±0.005
H4	0.0	-2.1290±0.001	3.7051±0.009
H5	0.0	2.1290±0.001	3.7051±0.009
H6	-0.5150±0.0005	-0.8919±0.0005	5.7510±0.008
H7	-0.5150±0.0005	0.8919±0.0005	5.7510±0.008
H8	1.0299±0.0005	0.0	5.7510±0.008
<sup>13</sup> C <sup>a</sup>	0.0	0.0	1.0866

<sup>a</sup> Proton coordinates are taken from [34], whereas the <sup>13</sup>C coordinate obtained from PubChem database [92] is less precise and may bias  $D_{C1}$ . This bias is absorbed by the free parameter  $J_{C1}$ , since only the combination  $D_{C1} + \frac{1}{2}J_{C1}$  is relevant.

## 2. MQC experiment of DMBP

Since MQC intensity decays rapidly with increasing coherence order [93], direct MQC experiments on <sup>13</sup>C-DMPB are impractical due to the large number of <sup>1</sup>H spins (15). To address this, we synthesized d<sub>6</sub>-DMPB, in which the methyl <sup>1</sup>H are replaced by <sup>2</sup>H, and performed proton MQC experiments using the experiment scheme reported in [94](Fig. S20(b)). In this sequence, a central  $\pi$  pulse in the  $t_1$  evolution refocuses both <sup>1</sup>H chemical shifts and <sup>1</sup>H – <sup>2</sup>H dipolar couplings. Three spectra were acquired with  $\tau = 0.8, 1.0$ , and 1.2 ms and subsequently magnitude-averaged. The resulting full spectrum containing all coherence orders is shown in Fig. S21. The  $\Delta m = 6$  and  $\Delta m = 7$  spectra (Fig. S22) exhibit peaks symmetrically distributed around 0Hz, which were used for spectral estimation. We observed significant peak broadening, possibly due to the large quadrupolar splitting of <sup>2</sup>H and strong heteronuclear dipolar couplings, which lead to imperfect refocusing during the  $t_1$  evolution period. In addition, the relatively low solute concentration and limited deuterium decoupling power of our system result in a poor signal-to-noise ratio.

TABLE III. HMQC-estimated interaction parameters (in Hz) and Saupe order tensor of <sup>13</sup>C-Tol. Dipolar couplings are calculated from the estimated Saupe order tensor and the molecular structure (Table II)<sup>a</sup>

Parameter	Value	Parameter	Value
$S_{yy}$	-0.0176±0.0005	$J_{45}^c$	1.97
$S_{zz}$	0.1758±0.0003	$D_{12}, D_{13}$	-242.25
$\omega_1/2\pi$	19±9	$D_{14}, D_{15}$	-196.73
$\omega_2/2\pi, \omega_3/2\pi$	-72±14	$D_{16}, D_{17}, D_{18}$	-100.95
$\omega_4/2\pi, \omega_5/2\pi$	-111±14	$D_{23}$	26.64
$\omega_6/2\pi, \omega_7/2\pi, \omega_8/2\pi$	-2348±6	$D_{45}$	27.38
$J_{C1}^b$	559±16	$D_{24}, D_{35}$	-1412.40
$J_{C2}^c, J_{C3}$	1.07	$D_{25}, D_{34}$	-30.53
$J_{C4}^c, J_{C5}$	7.54	$D_{26}, D_{27}, D_{28}, D_{36}, D_{37}, D_{38}$	-126.91
$J_{C6}^c, J_{C7}, J_{C8}$	0.84	$D_{46}, D_{47}, D_{48}, D_{56}, D_{57}, D_{58}$	-405.22
$J_{12}^c, J_{13}$	7.48	$D_{67}, D_{68}, D_{78}$	1860.23
$J_{14}^c, J_{15}$	1.27	$D_{C1}$	-4139.45
$J_{23}^c$	1.49	$D_{C2}, D_{C3}$	50.16
$J_{24}^c, J_{35}$	7.68	$D_{C4}, D_{C5}$	-77.68
$J_{25}^c, J_{34}$	0.62	$D_{C6}, D_{C7}, D_{C8}$	-45.33

<sup>a</sup> All 18 peaks in  $\Delta m_I = 7$ ,  $\Delta m_S = 0$  and  $\Delta m_I = 8$ ,  $\Delta m_S = 0$  HMQC spectra were used in fitting, yielding in an r.m.s error of 7.2Hz

<sup>b</sup> In the fitting,  $J_{C1}$  appears larger than its typical isotropic value. This reflects compensation for the imprecise <sup>13</sup>C coordinate rather than a physical deviation, and does not affect the analysis since only  $D_{C1} + \frac{1}{2}J_{C1}$  enters the Hamiltonian.

<sup>c</sup> Non-direct bonded  $J$ -couplings were fixed to their isotropic values from the literature [34, 90].

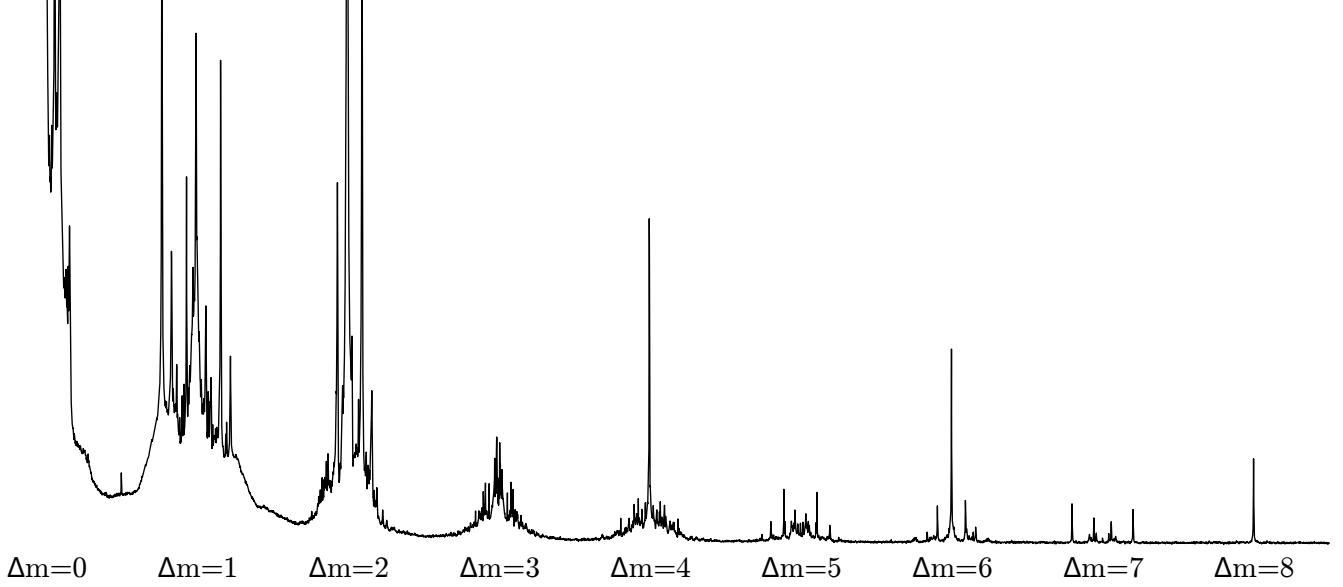


FIG. S21. Proton MQC spectrum of  $d_6$ -DMPB. The individual multiple-quantum coherence orders were separated by time proportional phase instrumentation (TPPI). The spectrum is obtained by magnitude-averaging of three spectra with  $\tau = 0.8$ , 1.0, and 1.2 ms. For all experiments, 16k  $t_1$  increments, 2 scans per increment. The offset introduced by TPPI is 25kHz. Spectrum width is 450kHz for  $F_1$  and 10kHz for  $F_2$ . The magnitude of deuterium decoupling rf field is about 3kHz in frequency unit.

Consequently, only 11 pairs of peaks in  $\Delta m = 6$  and  $\Delta m = 7$  spectra could be reliably identified and assigned, which are insufficient to determine all 14 unique dipolar couplings. Therefore, we directly evaluated the RMS deviation between the experimental peak positions and those predicted from the dipolar couplings obtained by MD using various candidate PMF surfaces, to identify the best-matching surface (result shown in Fig. 4(d) of the main text). Since J-couplings are much smaller than the spectra linewidth, they are ignored in simulations.

### K. Higher order Magnus expansion terms

Our goal of this section is to define a family of average Hamiltonians that closely approximates the full pulse sequence used in the experiment. The goal of this approximation is to make emulation of the NMR experiment feasible with reasonable depth quantum circuits. The NMR Hamiltonian, including coupling of the proton and carbon spins to the external field, is given by

$$H(t) = H_{\text{rf}}^{\text{H}}(t) + H_{\text{rf}}^{\text{C}}(t) + H_{\text{int}}. \quad (\text{S35})$$

The external field Hamiltonians take the form

$$\begin{aligned} H_{\text{rf}}^{\text{H}}(t) &= (1 + \epsilon_{\text{H}})H_{\text{nom}}^{\text{H}}(t), \\ H_{\text{rf}}^{\text{C}}(t) &= (1 + \epsilon_{\text{C}})H_{\text{nom}}^{\text{C}}(t), \end{aligned} \quad (\text{S36})$$

where  $\epsilon_{\text{H}}$  and  $\epsilon_{\text{C}}$  account for pulse amplitude imperfections due to  $B_1$ -inhomogeneity, and the nominal radio-frequency Hamiltonians are defined by

$$\begin{aligned} H_{\text{nom}}^{\text{H}}(t) &= \omega_{\text{nom}}^{\text{H}} \{ \cos(\phi_{\text{H}}(t))I_x - \sin(\phi_{\text{H}}(t))I_y \}, \\ H_{\text{nom}}^{\text{C}}(t) &= \omega_{\text{nom}}^{\text{C}} \{ \cos(\phi_{\text{C}}(t))S_x - \sin(\phi_{\text{C}}(t))S_y \}. \end{aligned} \quad (\text{S37})$$

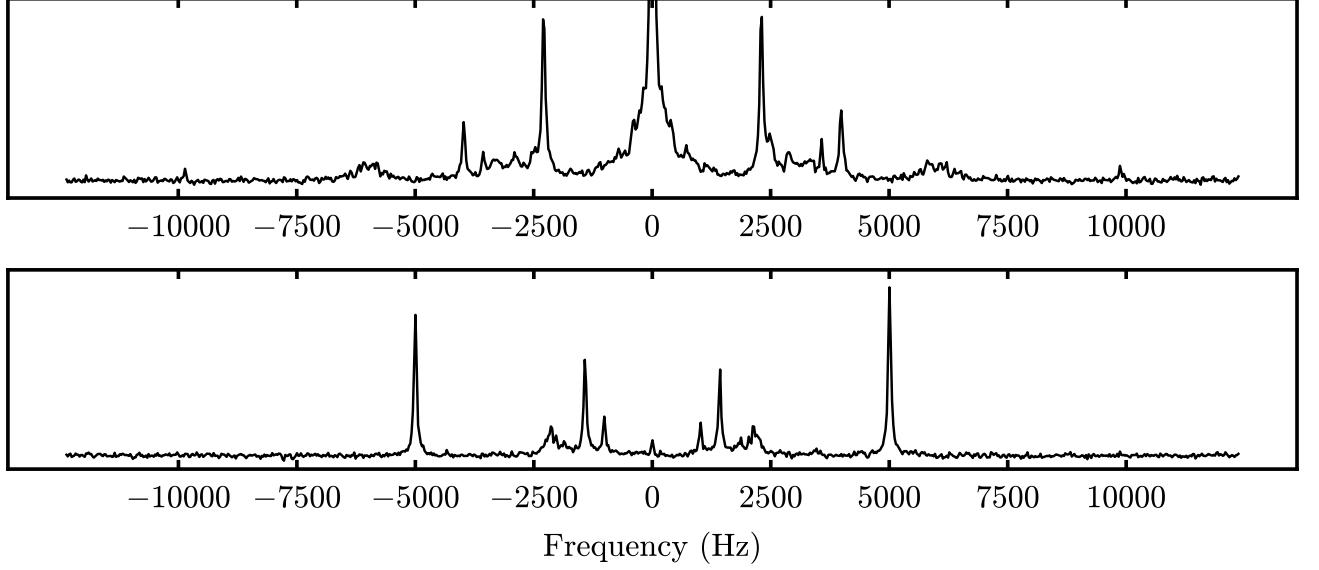


FIG. S22.  $\Delta m = 6$  (top) and  $\Delta m = 7$  (bottom) proton MQC spectrum of  $d_6$ -DMPB. Peaks are symmetrically distributed around 0Hz. 11 pairs of peaks are identified and assigned. The average linewidth is about 70Hz.

The nominal pulse amplitude  $\omega_{\text{nom}}$  is defined such that  $\omega_{\text{nom}}\tau_{\pi/2} = \pi/2$ , the phase  $\phi(t)$  is pulse sequence and channel dependent. Both  $H_{\text{rf}}^{\text{H}}(t)$  and  $H_{\text{rf}}^{\text{C}}(t)$  produce simple rotation operations

$$\begin{aligned} U_{\text{rf}}^{\text{H}}(t) &= \mathcal{T} \exp \left( -i \int_0^t H_{\text{rf}}^{\text{H}}(t') dt' \right) \in \text{SU}(2), \\ U_{\text{rf}}^{\text{C}}(t) &= \mathcal{T} \exp \left( -i \int_0^t H_{\text{rf}}^{\text{C}}(t') dt' \right) \in \text{SU}(2), \end{aligned} \quad (\text{S38})$$

and admit a decomposition in terms of Euler rotations

$$\begin{aligned} U_{\text{rf}}^{\text{H}}(t) &= R(\Omega^{\text{H}}(t)) = R_z(\alpha^{\text{H}}(t))R_y(\beta^{\text{H}}(t))R_z(\gamma^{\text{H}}(t)), \\ U_{\text{rf}}^{\text{C}}(t) &= R(\Omega^{\text{C}}(t)) = R_z(\alpha^{\text{C}}(t))R_y(\beta^{\text{C}}(t))R_z(\gamma^{\text{C}}(t)). \end{aligned} \quad (\text{S39})$$

The propagator for the full TARDIS-1 OTOC experiment may be represented as a concatenation of 6 basic elements

$$U_{\text{TARDIS-1,OTOC}} = \{U_{\leftarrow}(nt_c)\}\{\Pi_y^{\text{H}}\Pi_{-y}^{\text{C}}\}\{U_{\text{BLEW12}}^2\}\{\Pi_y^{\text{C}}\}\{U_{\text{BLEW12}}^2\}\{U_{\rightarrow}(nt_c)\} \quad (\text{S40})$$

Where  $\Pi$  denotes the propagator of the  $\pi$  pulse. By design,  $U_{\leftarrow}$ ,  $U_{\rightarrow}$  and  $U_{\text{BLEW12}}$  are so-called *cyclic* pulse sequence elements. As a consequence, the *nominal* part of the radio-frequency propagator, when considered in isolation, produces no net rotation upon completion of a period

$$U_{\rightarrow}^{\text{nom}} = \pm \mathbb{1}, \quad U_{\leftarrow}^{\text{nom}} = \pm \mathbb{1}, \quad U_{\text{BLEW12}}^{\text{nom}} = \pm \mathbb{1}. \quad (\text{S41})$$

For the non-trivial unitaries we then utilise an interaction frame decomposition factoring out the nominal part of the radio-frequency Hamiltonian

$$\begin{aligned} U_{\rightarrow}(nt_c) &= U_{\rightarrow}^{\text{nom}} \tilde{U}_{\rightarrow}(nt_c) = \pm \tilde{U}_{\rightarrow}(nt_c), \\ U_{\leftarrow}(nt_c) &= U_{\leftarrow}^{\text{nom}}(nt_c) \tilde{U}_{\leftarrow} = \pm \tilde{U}_{\leftarrow}(nt_c), \\ U_{\text{BLEW12}} &= U_{\text{BLEW12}}^{\text{nom}} \tilde{U}_{\text{BLEW12}} = \pm \tilde{U}_{\text{BLEW12}}, \end{aligned} \quad (\text{S42})$$

The corresponding interaction frame propagator  $\tilde{U}$  is generally given by

$$\tilde{U}(t_c) = \mathcal{T} \exp \left( -i \int_0^{t_c} \tilde{H}(t') dt' \right), \quad (\text{S43})$$

with

$$\begin{aligned}\tilde{H}(t') &= U^{\text{nom}\dagger}(t')\{H(t') - (H_{\text{nom}}^{\text{H}}(t') + H_{\text{nom}}^{\text{C}}(t'))\}U^{\text{nom}}(t') \\ &= \epsilon_{\text{H}}\tilde{H}_{\text{nom}}^{\text{H}}(t') + \epsilon_{\text{C}}\tilde{H}_{\text{nom}}^{\text{C}}(t') + \tilde{H}_{\text{int}}(t'),\end{aligned}\quad (\text{S44})$$

and

$$U^{\text{nom}}(t') = \mathcal{T} \exp\left(-i \int_0^{t'} H_{\text{nom}}(s) ds\right). \quad (\text{S45})$$

The propagator for the full TARDIS-1 sequence is then alternatively given by

$$U_{\text{TARDIS-1,OTOC}} = \{\tilde{U}_{\leftarrow}(nt_c)\}\{\Pi_y^{\text{H}}\Pi_{-y}^{\text{C}}\}\{\tilde{U}_{\text{BLEW12}}^2\}\{\Pi_y^{\text{C}}\}\{\tilde{U}_{\text{BLEW12}}^2\}\{\tilde{U}_{\rightarrow}(nt_c)\}. \quad (\text{S46})$$

$$U_{\text{TARDIS-1,LE}} = \{\tilde{U}_{\leftarrow}(nt_c)\}\{\Pi_y^{\text{H}}\}\{\tilde{U}_{\rightarrow}(nt_c)\}. \quad (\text{S47})$$

This representation is suitable for average Hamiltonian theory (AHT) approximation schemes, so that the interaction frame propagators  $\tilde{U}$  may be estimated by

$$\begin{aligned}\tilde{U}_{\rightarrow}(nt_c) &= \exp(-i\bar{H}_{\rightarrow}nt_c) \simeq \exp\left(-i(\bar{H}_{\rightarrow}^{(0)} + \bar{H}_{\rightarrow}^{(1)} + \dots)nt_c\right), \\ \tilde{U}_{\leftarrow}(nt_c) &= \exp(-i\bar{H}_{\leftarrow}nt_c) \simeq \exp\left(-i(\bar{H}_{\leftarrow}^{(0)} + \bar{H}_{\leftarrow}^{(1)} + \dots)nt_c\right), \\ \tilde{U}_{\text{BLEW12}} &= \exp(-i\bar{H}_{\text{BLEW12}}T_{\text{BLEW12}}) \simeq \exp(-i(\bar{H}_{\text{BLEW12}}^0 + \bar{H}_{\text{BLEW12}}^1 + \dots)T_{\text{BLEW12}}).\end{aligned}\quad (\text{S48})$$

The first few terms of this approximation may be calculated via the Magnus expansion (Eq. S20). The Euler decomposition given by Eq. S39 provides a convenient method for the calculation of the average Hamiltonian terms. To this end, the spin Hamiltonian is expanded in terms of irreducible tensor operators  $Q_{lm}$  of the rotation group SO(3)

$$R(\Omega)Q_{lm}R^{\dagger}(\Omega) = \sum_{n=-l}^{+l} Q_{ln}D_{nm}^l(\Omega), \quad (\text{S49})$$

where  $D_{nm}^l(\Omega)$  are Wigner matrix elements of spherical rank  $l$ . In the absence of  $B_1$ -inhomogeneities ( $\epsilon = 0$ ), the spin Hamiltonian in the interaction frame of the nominal rf part may be decomposed as follows

$$\tilde{H}(t) = \tilde{H}_{\text{DD}}^{\text{HH}} + \tilde{H}_{\text{CS}}^{\text{H}} + \tilde{H}_{\text{DD}}^{\text{HC}}, \quad (\text{S50})$$

$$\begin{aligned}\tilde{H}_{\text{DD}}^{\text{HH}} &\leftarrow \sum_{m=-2}^{+2} Q_{2m}D_{m0}^2(\Omega^{\text{H}}(t)), \\ \tilde{H}_{\text{CS}}^{\text{H}} &\leftarrow \sum_{m=-1}^{+1} Q_{1m}D_{m0}^1(\Omega^{\text{H}}(t)), \\ \tilde{H}_{\text{DD}}^{\text{HC}} &\leftarrow \sum_{m,n=-1}^{+1} Q_{1m}K_{1n}D_{m0}^1(\Omega^{\text{H}}(t))D_{n0}^1(\Omega^{\text{C}}(t)).\end{aligned}\quad (\text{S51})$$

The decomposition for  $H_{\text{DD}}^{\text{HC}}$  follows from the fact that pulses may be applied individually to the protons and the carbons, so technically speaking these operators span representations for  $\text{SO}(3) \otimes \text{SO}(3)$ . The basis tensors are explicitly given by

$$[Q]_{2m} = \sum_{i < j, i,j \in {}^1\text{H}} D_{ij} \begin{pmatrix} \sqrt{\frac{3}{2}}(I_x^i I_x^j - i(I_x^i I_y^j + I_y^i I_x^j - i(I_y^i I_y^j))) \\ \sqrt{\frac{3}{2}}(I_x^i I_z^j - i(I_y^i I_z^j + I_z^i I_y^j + i(I_z^i I_x^j))) \\ -I_x^i I_x^j - I_y^i I_y^j + 2I_z^i I_z^j \\ -\sqrt{\frac{3}{2}}(I_x^i I_z^j + i(I_y^i I_z^j + I_z^i I_y^j - i(I_z^i I_x^j))) \\ \sqrt{\frac{3}{2}}(I_x^i I_x^j + i(I_x^i I_y^j + I_y^i I_x^j + i(I_y^i I_y^j))) \end{pmatrix} \quad (\text{S52})$$

$$[Q]_{1m} = \sum_{i \in {}^1\text{H}} \omega_i \begin{pmatrix} \frac{1}{\sqrt{2}}(I_x^i - iI_y^i) \\ I_z^i \\ -\frac{1}{\sqrt{2}}(I_x^i + iI_y^i) \end{pmatrix} \quad (\text{S53})$$

$$[Q]_{1m}[K]_{1n} = \sum_{i \in {}^1\text{H}} (D_{Ci} + J_{Ci}/2) \begin{pmatrix} I_x^i S_x - iI_x^i S_y - iI_y^i S_x - I_y^i S_y \\ \sqrt{2}(I_x^i S_z - iI_y^i S_z) \\ -I_x^i S_x - iI_x^i S_y + iI_y^i S_x - I_y^i S_y \\ \sqrt{2}(I_z^i S_x - iI_z^i S_y) \\ 2I_z^i S_z \\ -\sqrt{2}(I_z^i S_x + iI_z^i S_y) \\ -I_x^i S_x + iI_x^i S_y - iI_y^i S_x - I_y^i S_y \\ -\sqrt{2}(I_x^i S_z + iI_y^i S_z) \\ I_x^i S_x + iI_x^i S_y + iI_y^i S_x - I_y^i S_y \end{pmatrix} \quad (\text{S54})$$

If we move to a linear indexing scheme

$$([Q]_{2m}, [Q]_{1m}, [Q]_{1m}[K]_{1n}) \rightarrow (C_1, C_2, \dots), \quad (\text{S55})$$

the interaction frame Hamiltonian takes the simple form

$$\tilde{H}(t) = \sum_{j=1}^N c_j(t) C_j, \quad (\text{S56})$$

where  $c_j(t)$  are the appropriate Wigner matrix elements. If we further define a commutator basis

$$\begin{aligned} C_{ij} &= [C_i, C_j], \\ C_{ijk} &= [C_i, [C_j, C_k]], \end{aligned} \quad (\text{S57})$$

the expansion terms of the Magnus series are given by

$$\begin{aligned} \bar{H}^{(0)} &= \frac{1}{t_c} \sum_{i=1}^N C_i w_i, \\ \bar{H}^{(1)} &= \frac{1}{2it_c} \sum_{i,j=1}^N C_{ij} w_{ij}, \\ \bar{H}^{(2)} &= -\frac{1}{6t_c} \sum_{i,j,k=1}^N C_{ijk} (w_{ijk} + w_{kji}), \end{aligned} \quad (\text{S58})$$

with the weights

$$\begin{aligned} w_i &= \int_0^{t_c} c_i(t_1) dt_1, \\ w_{ij} &= \int_0^{t_c} \int_0^{t_1} c_i(t_1) c_j(t_2) dt_2 dt_1 = \int_0^{t_c} c_i(t_1) w_j(t_1) dt_1, \\ w_{ijk} &= \int_0^{t_c} \int_0^{t_1} \int_0^{t_2} c_i(t_1) c_j(t_2) c_k(t_3) dt_3 dt_2 dt_1 = \int_0^{t_c} c_i(t_1) w_{jk}(t_1) dt_1. \end{aligned} \quad (\text{S59})$$

To include rf errors it is convenient to augment the basis tensors by

$$[F^{\text{H}}]_{1m} = \omega_{\text{nom}}^{\text{H}} \begin{pmatrix} \frac{1}{\sqrt{2}}(I_x - iI_y) \\ I_z \\ -\frac{1}{\sqrt{2}}(I_x + iI_y) \end{pmatrix} \quad (\text{S60})$$

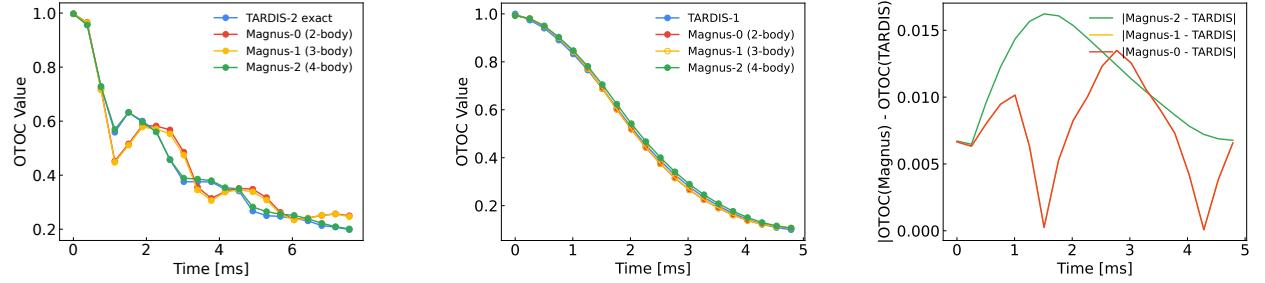


FIG. S23. (left) OTOC versus experiment time demonstrating the convergence of the Magnus expansion unitaries (Eq. (S20)) up to second order. For each order in the Magnus expansion correction we indicate the many-body order of the correction Hamiltonian. The similarity of green (Magnus-2) and blue (full TARDIS-2 sequence) indicates convergence of the average Hamiltonian expansion. Average Hamiltonian theory is applied to the forward, backwards, and BLEW12 sequence. The  $\pi$  and  $-\pi$  pulses on the  $^{13}\text{C}$  channel, which contain the full dipolar Hamiltonian background evolution for  $2t_p = 21\mu\text{s}$ , is simulated explicitly. (center) OTOC versus time demonstrating the convergence of the Magnus expansion unitaries towards the full TARDIS-1 unitary for the 15-spin system. (right) Difference in OTOC values for the 15-spin system at different orders in Magnus expansion and the TARDIS-1 OTOC value. Magnus-1 is indistinguishable from Magus-0, by design.

$$[F^C]_{1m} = \omega_{\text{nom}}^C \begin{pmatrix} \frac{1}{\sqrt{2}}(S_x - iS_y) \\ S_z \\ -\frac{1}{\sqrt{2}}(S_x + iS_y) \end{pmatrix} \quad (\text{S61})$$

The rf Hamiltonian may then be written as

$$\begin{aligned} H_{\text{rf}}^{\text{H}}(t) &= -(1 + \epsilon_{\text{H}}) \frac{\omega_{\text{nom}}^{\text{H}}}{\sqrt{2}} \sum_{m=-1}^{+1} m F_{1m}^{\text{H}} e^{-im\phi_{\text{H}}(t)}, \\ H_{\text{rf}}^{\text{C}}(t) &= -(1 + \epsilon_{\text{C}}) \frac{\omega_{\text{nom}}^{\text{C}}}{\sqrt{2}} \sum_{m=-1}^{+1} m F_{1m}^{\text{C}} e^{-im\phi_{\text{C}}(t)}, \end{aligned} \quad (\text{S62})$$

The rf error term in the interaction frame of the nominal pulse propagator is then given by

$$\begin{aligned} \tilde{H}_{\epsilon}^{\text{H}}(t) &= -\epsilon_{\text{H}} \frac{\omega_{\text{nom}}^{\text{H}}}{\sqrt{2}} \sum_{m,n=-1}^{+1} m F_{1n}^{\text{H}} D_{nm}(\Omega^{\text{H}}(t)) e^{-im\phi_{\text{H}}(t)} = -\epsilon_{\text{H}} \frac{\omega_{\text{nom}}^{\text{H}}}{\sqrt{2}} \sum_{n=-1}^{+1} F_{1n}^{\text{H}} f_n^{\text{H}}(t), \\ \tilde{H}_{\epsilon}^{\text{C}}(t) &= -\epsilon_{\text{C}} \frac{\omega_{\text{nom}}^{\text{C}}}{\sqrt{2}} \sum_{m,n=-1}^{+1} m F_{1n}^{\text{C}} D_{nm}(\Omega^{\text{C}}(t)) e^{-im\phi_{\text{C}}(t)} = -\epsilon_{\text{C}} \frac{\omega_{\text{nom}}^{\text{C}}}{\sqrt{2}} \sum_{n=-1}^{+1} F_{1n}^{\text{C}} f_n^{\text{C}}(t), \end{aligned} \quad (\text{S63})$$

where the time-dependent part is independent of the fractional error and may be pre-computed once.

## L. Understanding features of the toluene OTOC decay

In Fig. 2(c) of the main text we showed the observed NMR OTOC for toluene, which in addition to the typical decay profile of an OTOC contained a notable bump around 1.5 ms. We further observed in Fig. 2(d) that around this bump the OTOC had an inversion feature; increasing the ortho-meta HH coupling (which should typically cause faster decay) caused the OTOC to increase, rather than decrease. In this section, we outline our understanding of the physical mechanism causing this phenomenon, and provide further numerical simulations to back up our theory.

To understand this feature, we first note that we can separate the toluene spins into three clusters:

1. The  $^{13}\text{C}$  and para- $^1\text{H}$  spin.
2. The ortho- and meta- $^1\text{H}$  spins
3. The methyl- $^1\text{H}$  spins.

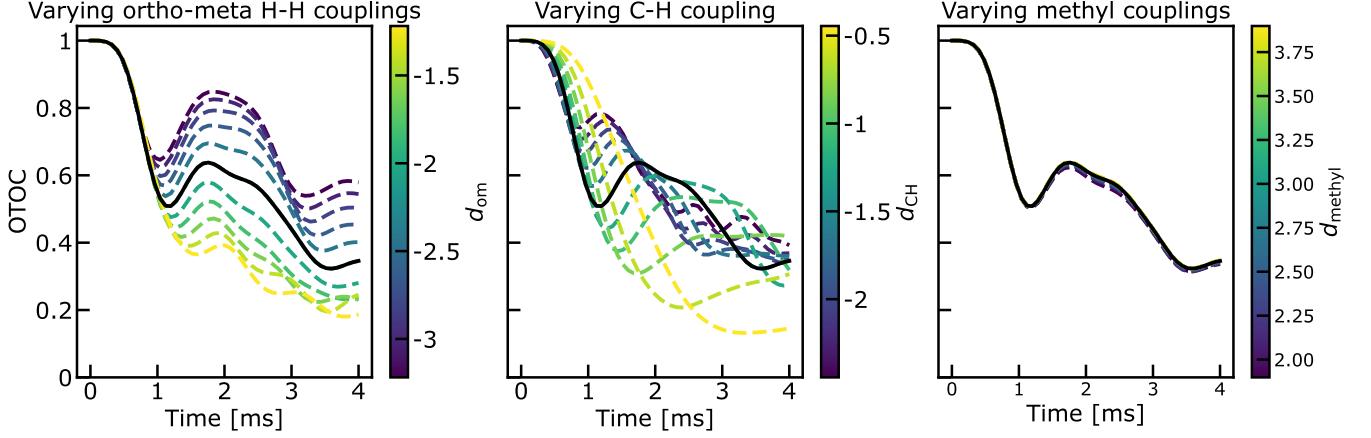


FIG. S24. Plots showing the change in the toluene OTOC as different couplings in the molecule are changed. All simulations are done to lowest order in average Hamiltonian theory (i.e. using the DQ Hamiltonian directly). For each plot, the black line shows the expected OTOC decay using the reference parameters described in the text, while the coloured dashed lines show the effect of changing a single coupling (as described in title).

The intra-cluster dipolar couplings are significantly larger ( $3 - 10 \times$ ) than the inter-cluster couplings, and so they form bound states. We hypothesize that these bound states cause two effects to the OTOC:

1. The operators within the C-H bound state oscillate quickly, which directly causes the OTOC oscillations.
2. The H-H bound state acts as a barrier, preventing the decay of the C-H bound state into the rest of the system.

We evidence this hypothesis in Fig. S24, where we vary the large couplings within each cluster simultaneously<sup>2</sup>. The crucial observation here is that changing the ortho-meta H-H coupling changes the height of the OTOC fluctuations but not the frequency, whilst changing the C-H coupling changes the frequency but has a lesser effect on the height, which qualitatively matches what we would expect from the above hypothesis. By contrast, we observe that the OTOC decay is almost completely independent of the intra-methyl couplings, which serve as a baseline.

In order to observe the feature detailed above, we needed the presence of bound states, but we also required they be isolated - i.e. cut off from the rest of the molecule save for coupling through another bound state. In a larger molecule we expect to continue to have bound states, however we would expect cases of isolated bound states to be more rare due to the increased number of spins in the system. Thus, we expect that the features observed here will be less frequent in larger system sizes, and the only characteristic feature of large-scale OTOCs will be their decay.

#### M. Predicting ultimate decoherence distance limits in measuring OTOCs

In this section, we estimate the ultimate length limit of an OTOC in a nuclear magnetic spin system due to the decoherence alone. Though decoherence was not the primary source of Loschmidt echo decay observed in Fig. 2(c) of the main text, the sources that were observed (higher-order corrections to the Magnus expansion and RF inhomogeneity) can be fixed in turn by going to higher pulse power spectrometers and by using a smaller sample holder respectively. Thus, we argue that the ultimate limit we should consider comes from the decoherence of the large spin cluster formed by forward evolution. We note here that the size of this spin cluster can in principle be reduced by selective deuteration, and that decoherence times in NMR systems can be reduced by lowering temperature. However, this additional overhead increases the complexity of the OTOC experiment, so we do not consider these (or other) experimental techniques for this simple calculation.

Our calculation proceeds by imagining the growth of a cluster starting from a single spin, assuming that all spins within the cluster decohere at a rate  $\tau_d$ . We can perform an OTOC using any spin (and observe a signal) in this cluster as a butterfly. We assume that the experiment lasts a time  $t_{\text{exp}}$ ; in this time the cluster grows to a volume

<sup>2</sup> Note that changing the ortho-meta coupling achieves a similar result to the stretching performed in Fig. 2(d) of the main text. The results here are not identical because a) here we do not

change couplings other than the single H-H coupling, and b) here we take the zeroth order of the Magnus expansion.

$V_s$ , which entangles  $N_s$  spins depending on the density of the protein:  $\rho_s = N_s/V_s$ . As a rough back-of-the-envelope calculation, we can estimate typical values for our physical constants as

$$\tau_d \approx 1 \text{ s}, \quad t_{\exp} \approx 1 \text{ ms}, \quad \rho_s \approx 1.38 \text{ g protein cm}^{-3} \approx 10^{-25} \text{ g H \AA}^{-3} \approx 0.1 \text{ spin \AA}^{-3}, \quad (\text{S64})$$

where  $1.38 \text{ g cm}^{-3}$  is the mean density of protein, of which about  $6 - 10\%$  is hydrogen (and we recall that the mass of a hydrogen atom is around  $10^{-24} \text{ g}$ ), and our approximation for  $t_{\exp}$  is based on previous numerical simulations of OTOCs in solid state proteins (which yielded  $t_{\exp} \in [0.5 \text{ ms}, 2 \text{ ms}]$  at distances of  $10 - 12 \text{ \AA}$ ).

We need to consider two complications in developing an accurate model of spin cluster growth:

1. Not every spin stays within the cluster for the same time.
2. The growth of the spin cluster is not spherically symmetric, due to the distance dependence of the dipolar interaction.

Let us first assume spherically symmetric growth, and attempt to capture the time dependence accurately. We can assign each spin  $s$  a time  $t_s$  that corresponds to how long the spin is active in the OTOC. Then, our decoherence criteria is

$$\sum_s t_s = \tau_d. \quad (\text{S65})$$

Without the directional dependence on the dipolar coupling, Ref. [95] suggests that the operator light cone for a  $1/r^3$  coupling in a 3-dimensional system should grow as  $e^{\sqrt{\ln(x)}}$ . However, the directional dependence will likely reduce the effective dimension below 3, in which case we expect a linear or superlinear growth of the operator. This implies that spins at distance  $x_s$  from the measurement spin will become entangled in a time

$$t(x_s) = ax^\zeta \rightarrow t_s = t_{\exp} - t(x_s) = a(r_c^\zeta - x_s^\zeta), \quad (\text{S66})$$

for some  $\zeta \in [0, 1]$ . We can thus write our decoherence criteria as

$$\tau_d = \sum_s a(r_c^\zeta - x_s^\zeta) = \int_0^{r_c} dx A(r_c^\zeta - x^\zeta) n(x), \quad (\text{S67})$$

where  $n(x)$  counts the number of spins in a layer  $dx$  thick around the radius  $x$ . Ignoring the directional dependence for now, we have  $n(x) = 4\pi x^2 \rho_s$ , and so

$$\tau_d = \int_0^{r_c} dx 4\pi \rho_s A(r_c^\zeta x^2 - x^{2+\zeta}) = 4\pi a \rho_s \left[ \frac{1}{3} r_c^\zeta x^3 - \frac{1}{3+\zeta} x^{3+\zeta} \right]_0^{r_c} = 4\pi a \rho_s r_c^{3+\zeta} \frac{\zeta}{9+3\zeta} = 4\pi \rho_s t_{\exp} r_c^3 \frac{\zeta}{9+3\zeta} \quad (\text{S68})$$

This can be rearranged to give

$$r_c = \left[ \frac{10000}{4\pi} \frac{9+3\zeta}{\zeta} \right]^{1/3} \text{\AA}. \quad (\text{S69})$$

We plot this as the blue line in Fig. S25, and observe that this yields a result between  $20 \text{ \AA}$  and  $40 \text{ \AA}$  depending on the value of  $\zeta$ .

Next, let us try to consider the effect of directional dependence. As a simplest first option, we can fix our measurement spin  $M$  at some co-ordinate in space, and assume that the spin cluster evolves following the equipotential lines of the  $1/r^3(1 - 3\cos^2(\theta))$  dipolar interaction. In Fig. S26 we determine this volume via numerical integration and find it to scale as  $V = 0.2r_c^3$ , which implies that along the  $z$ -direction, we have  $n(x) = 0.6x^2 \rho_s$ . Substituting this into the above yields

$$r_c = \left[ \frac{10000}{0.6} \frac{9+3\zeta}{\zeta} \right]^{1/3} \text{\AA}. \quad (\text{S70})$$

We plot this as the orange line in Fig. S25, and observe that distances of over  $60 \text{ \AA}$  could be measured in an OTOC under the correct circumstances. Note that in this expression we would expect the exponent of the OTOC growth  $\zeta$  to likely be 1 due to the dimensional reduction in the system. The above is furthermore a vast simplification of the operator growth dynamics in dipolar-coupled spin systems; understanding this in more detail is a clear target for future work.

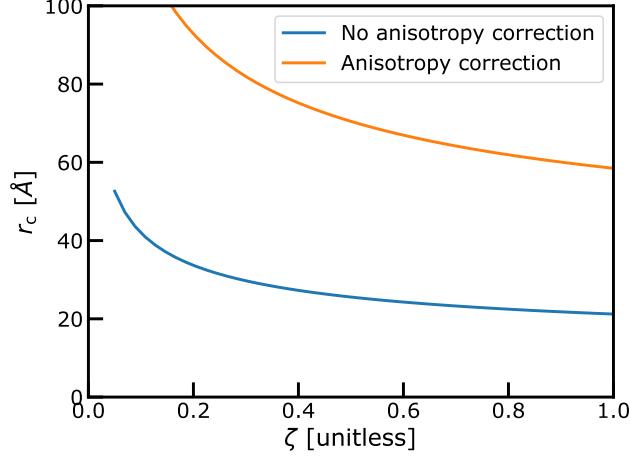
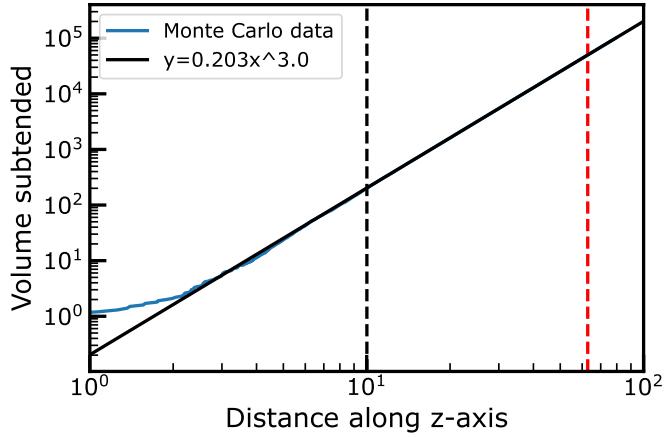


FIG. S25. Plot of Eq. (S69) (blue) and Eq. (S70) (orange)

FIG. S26. Determination of the volume enclosed by the (absolute) equipotential lines of the dipolar interaction, as a function of the direction along the  $z$ -axis (the axis of quickest coupling).

#### IV. MOLECULAR DYNAMICS MODEL

To simulate the structure of the solute in the liquid crystal environment, we ran molecular dynamics with a molecular mechanics force field. We use non-bonded parameters of GAFF version 2.11 [54], but we replace the bonded parameters with Grappa version 1.3.1 [55]. Although Grappa was designed for biomolecules, the training set includes a very large dataset of small organic molecules and we have observed empirically that the grappa force field makes the aliphatic chains in the liquid crystal more flexible which is consistent with modifications made to GAFF for liquid crystal systems [56]. For reference, we provide the GAFF force-field structure

$$\begin{aligned}
 E_{\text{pair}} = & \sum_{\text{bonds}} K_r (r - r_{\text{eq}})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{\text{eq}})^2 \\
 & + \sum_{\text{dihedrals}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] \\
 & + \sum_{i < j} \left[ \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]
 \end{aligned} \tag{S71}$$

where non-bonding parameters are  $A_{ij}$  and  $B_{ij}$ . In the following section we outline the construction and parameterization of the molecular dynamics model and the dihedral learning problem described in the main text.

### A. 15Q Simulation details

Starting structures were generated with Packmol version 21.0.0 [52] at a concentration of 2.05% v/v 1,3-dimethyl-5-(phenyl-<sup>13</sup>C)benzene (DMPB) and a starting density of 1.01 g/ml. The liquid crystal molecules were approximately aligned along the x-axis of the simulation box with dimensions [65Å, 50Å, 50Å] by providing angle constraints in the Packmol input file. The initial structures were relaxed by minimizing the system at constant volume and then equilibrated with 500 ps of NPT dynamics. All simulations use a 2 fs time step with hydrogen bond constraints and a hydrogen mass of 4 amu. NPT calculations used a Monte Carlo barostat with updates every 25 steps. All ensemble averages were evaluated with respect to 8 independent trajectories with production times of 1  $\mu$ s each. All molecular dynamics simulations were performed using OpenMM [53].

### B. 15Q $P_2$ vs temperature Calibration

The second-rank orientational order parameter of a molecule,  $P_2$ , characterizes the degree of molecular alignment and is defined as

$$P_2 = \left\langle \frac{3 \cos^2 \theta - 1}{2} \right\rangle \quad (\text{S72})$$

Here, the angle bracket denotes an ensemble average and  $\theta$  is the angle between the molecular reference (long) axis and the liquid crystal director. For flexible molecules such as 5CB, the reference axis is chosen as the principal axis of its inertia tensor.

As reported in Ref. 96, for biphenyl dissolved in nematic liquid crystal 5CB, the biphenyl  $P_2$  is directly proportional to  $P_2$  of 5CB ( $P_2^{5\text{CB}}$ ). Therefore, for a similar system like ours, it is essential to first match  $P_2^{5\text{CB}}$  in both experiment and MD in order for MD to accurately reproduce the experimental orientational order of the 15Q molecule. The  $P_2^{5\text{CB}}$  from MD can differ from experiment in two ways: (1) the critical temperature can be different and (2) the behavior of  $P_2$  relative to the critical temperature can be different. As pointed out in Ref. 96, even with a well-optimized force field,  $P_2^{5\text{CB}}$  can still differ by  $\sim 10\%$  between MD and experiment at the same reduced temperature. Furthermore, even determining the critical temperature accurately from MD can be difficult since  $P_2^{5\text{CB}}$  becomes hard to converge near the critical point. However, since we are not concerned with the temperature dependence of orientational order itself, rather than attempting to further optimize MD, we adopted a simpler strategy: we experimentally measured  $P_2^{5\text{CB}}$  as a function of temperature, and used this calibration curve to determine the experimental temperature that reproduces the same  $P_2^{5\text{CB}}$  value as in MD ( $T_{\text{match}}$ ).  $T_{\text{match}}$  was determined to be 289K, and was used for subsequent OTOC experiments. This approach was later validated by MQC experiments. Though the MD value of  $P_2^{5\text{CB}}$  at 300K is likely not sufficiently converged, the calculations at 294K are well converged and the predicted  $P_2$  value thus serves as the reference point for calibration relative to experiment.

The simulated  $P_2$  values at each temperature are obtained by averaging expectation values over 8 independent 1  $\mu$ s trajectories, as described in Section IV A. In Figure S27, we present the cumulative average and standard deviation of  $P_2$  value along each trajectory. The cumulative average for each trajectory was calculated according to

$$P_2(t) = \frac{1}{t} \int_0^t dt' p_2(t'), \quad (\text{S73})$$

where  $p_2(t')$  is the instantaneous orientational order parameter.

While the cumulative averages for each trajectory run at  $T = 294$  K have largely stabilized toward the end of the runtime, several of the trajectories run at  $T = 300$  K are not converged, as can be seen by the relatively large changes in the cumulative average combined with the increasing standard deviations over the second half of the trajectory.

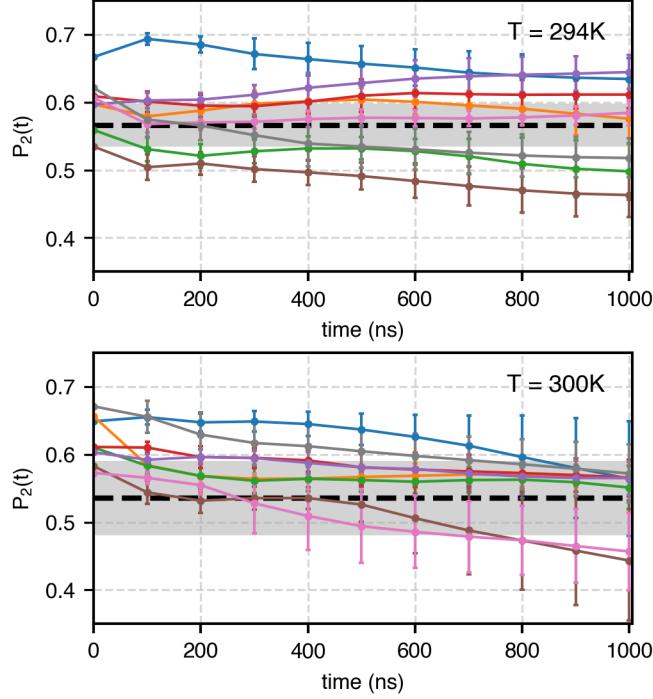


FIG. S27. Cumulative average and standard deviation of  $P_2$  along each trajectory for the simulations run at  $T = 294$  K and  $T = 300$  K. The black dashed line and shaded gray regions correspond to the overall mean and error for the full ensemble. The ensemble error is evaluated as the square root of the mean variance.

Experimentally,  $P_2^{5\text{CB}}$  was determined by measuring the  $^{13}\text{C}$  chemical shifts of 5CB [97]. It has been shown that there exists a simple semi-empirical relationship between the  $^{13}\text{C}$  chemical shifts and the order parameter:

$$\Delta\sigma = aP_2^{5\text{CB}} + b \quad (\text{S74})$$

where  $\Delta\sigma = \sigma_{\text{n}} - \sigma_{\text{iso}}$ , with  $\sigma_{\text{n}}$ ,  $\sigma_{\text{iso}}$  being the  $^{13}\text{C}$  chemical shifts in nematic and isotropic phases. The empirical coefficients  $a$  and  $b$  are listed in [97]<sup>3</sup>

<sup>3</sup> The empirical coefficients  $a$ ,  $b$  for C – 3' listed in [97] are incorrect, as pointed out in [98]. Therefore, all carbons in the 5CB

biphenyl cores except C – 3' were used to calculate  $P_2^{5\text{CB}}$ . [98] also identified several additional errors.

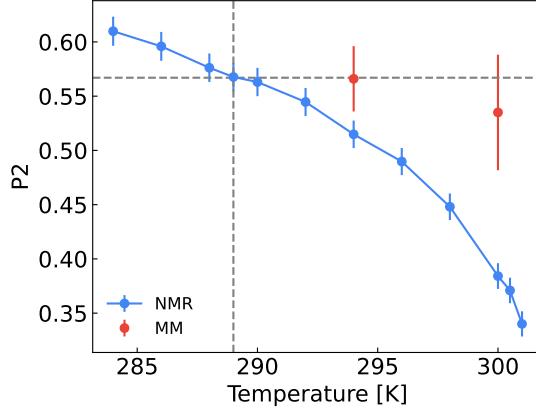


FIG. S28. Calibration curve of experimental  $P_2^{5\text{CB}}$  vs. temperature, overlaid with MD-simulated  $P_2^{5\text{CB}}$ . MD simulation at both 294K and 300K yields  $P_2^{5\text{CB}} \sim 0.566$ , which corresponds to  $T_{\text{match}} = 289\text{K}$ . Error bars on experimental  $P_2^{5\text{CB}}$  are obtained by propagating the reported uncertainties in the empirical coefficients  $a$  and  $b$  [97]

### C. Analysis of timescales for the 15Q system

For the 15Q system we verify that the dominant time-scales impacting the OTOC indicating there is no observable low-frequency mode on the scale of the experiment. We analyze the timescales associated with the methyl rotation angle,  $\chi$ , the inter-phenyl torsion angle ( $\phi$ ) and the overall alignment with the liquid crystal director, ( $\theta$ ). We can analyze the timescale associated with these angular quantities by computing the mean-squared displacement in each coordinate as a function of time. For example, we compute,

$$\langle(\phi(t_0) - \phi(t_0 + t))^2\rangle = \left\langle \frac{1}{N} \sum_{i=1}^N (\phi_i(t_0) - \phi_i(t_0 + t))^2 \right\rangle_{t_0}, \quad (\text{S75})$$

where  $N$  is the number of solute molecules in the simulation. Here, the expectation values are evaluated by averaging over time origins,  $t_0$ . In the long-time limit, the two angles should become uncorrelated so that

$$\lim_{t \rightarrow \infty} \langle(\phi(t_0) - \phi(t_0 + t))^2\rangle = 2\langle\phi^2\rangle - 2\langle\phi\rangle^2 = 2\text{Var}[\phi]. \quad (\text{S76})$$

For torsions, such as  $\phi$  and  $\chi$ , which would be uniformly distributed on  $[0, 2\pi]$  in the absence of interactions, this means that

$$\lim_{t \rightarrow \infty} \langle(\phi(t_0) - \phi(t_0 + t))^2\rangle \approx \frac{(2\pi)^2}{6} \quad (\text{S77})$$

The same is true for the polar angle,  $\theta$ , except that it is not uniformly distributed because of the non-uniform density of states. In the absence of any correlations, the distribution will be  $p(\theta) = 1/2 \sin \theta$ . This means that, in the long time limit, the mean squared displacement should be approximately

$$\lim_{t \rightarrow \infty} \langle(\theta(t_0) - \theta(t_0 + t))^2\rangle \approx 2\text{Var}[\theta] = \pi^2/2 - 4. \quad (\text{S78})$$

Given these expressions for the limiting behavior at long times, we can fit the mean-squared displacement to a function of the form

$$\text{MSD}[X](t) = C_X [1 - e^{-D_X t}] \quad [X = \phi, \theta], \quad (\text{S79})$$

where  $C_X$  is fixed according to the empirically observed long-time limit, which differs from the uncorrelated case, and  $D$  is like a diffusion constant with units of inverse time so that we can define a relaxation time,  $\tau_X \equiv 1/D_X$ . Plots of the mean-square displacement are shown in Figure S29. The inverses of the fitted constants,  $D_X$ , provide estimates of the timescale associated with motion in these coordinates.

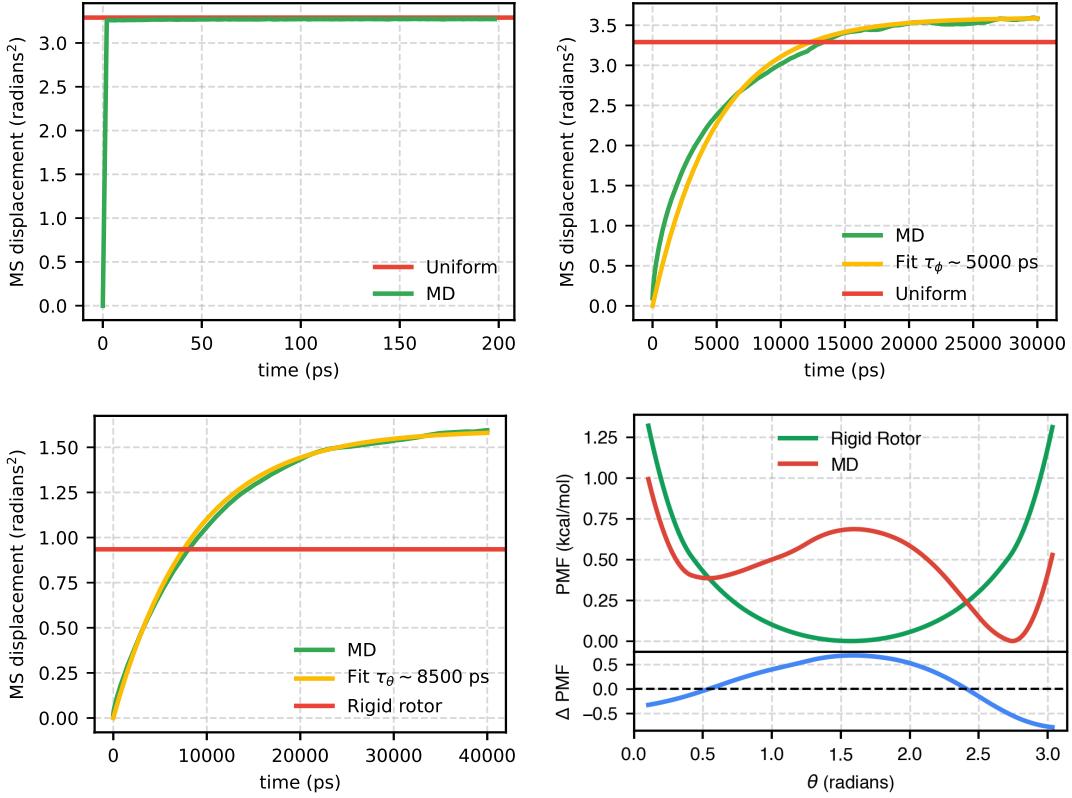


FIG. S29. The mean-square displacement as a function of simulation time for the methyl-rotation ( $\chi$ ), the inter-ring torsion ( $\phi$ ) and the overall alignment with the liquid crystal director ( $\theta$ ). The diffusion constants are  $D_\chi^{-1} < 1$  ps,  $D_\phi^{-1} = 900$  ps and  $D_\theta^{-1} = 4$  ns. We also show the PMF along the alignment with the liquid crystal director from MD, assuming a rigid rotor, and the difference between the two which reveals that interactions with the liquid crystal destabilize perpendicular orientations and stabilize parallel and antiparallel orientations.

#### D. 15Q Analysis of the effect of the non-isotropic environment

From molecular dynamics simulations, we can extract the free energy surface, or potential of mean force (PMF), along the coordinate that represents the alignment of solute molecules with the liquid crystal director,  $\theta$ . Specifically, the alignment is obtained by evaluating the angle between the para axis of the biphenyl core of the DMBP molecule and the instantaneous liquid crystal director. As discussed in Section IV C, the density of states is not uniform along this coordinate, so the non-isotropic behavior is most easily seen by subtracting the PMF of a non-interacting rigid rotor as shown in Figure S29. As expected, we observe that configurations that are aligned (either parallel or antiparallel) with the director of the liquid crystal are stabilized by the interaction with the liquid crystal environment (negative  $\Delta$ PMF), while perpendicular orientations are destabilized.

#### E. 15Q Analysis of the dipolar coupling matrix elements

To further analyze the dipolar couplings in particular, we plot the matrix elements of the  $D_{ij}$  matrix as a function of the torsion angle,  $\phi$ , in Figure S30. The indexing is consistent with the atom labels in Figure S7 with the labeled carbon as index zero. As expected, the couplings between hydrogens 1 and 6 (or any other equivalent pair) are most dramatically correlated with  $\phi$ . For other elements of  $D_{ij}$  the effect is more subtle, but still significant.

To assess the uncertainty in predicted dipolar coupling matrix elements, we perform a block averaging analysis on the dipolar coupling time series data for each solute molecule. We note that the dipolar coupling matrix elements are largely determined by the inter-ring torsion,  $\phi$ , and the alignment with the liquid crystal director,  $\theta$ . We define the blocks to be proportional to the  $\theta$  relaxation time, and choose 20 nanosecond blocks (where the associated mean squared displacement is roughly representative of no correlation), resulting in 50 blocks per solute molecule

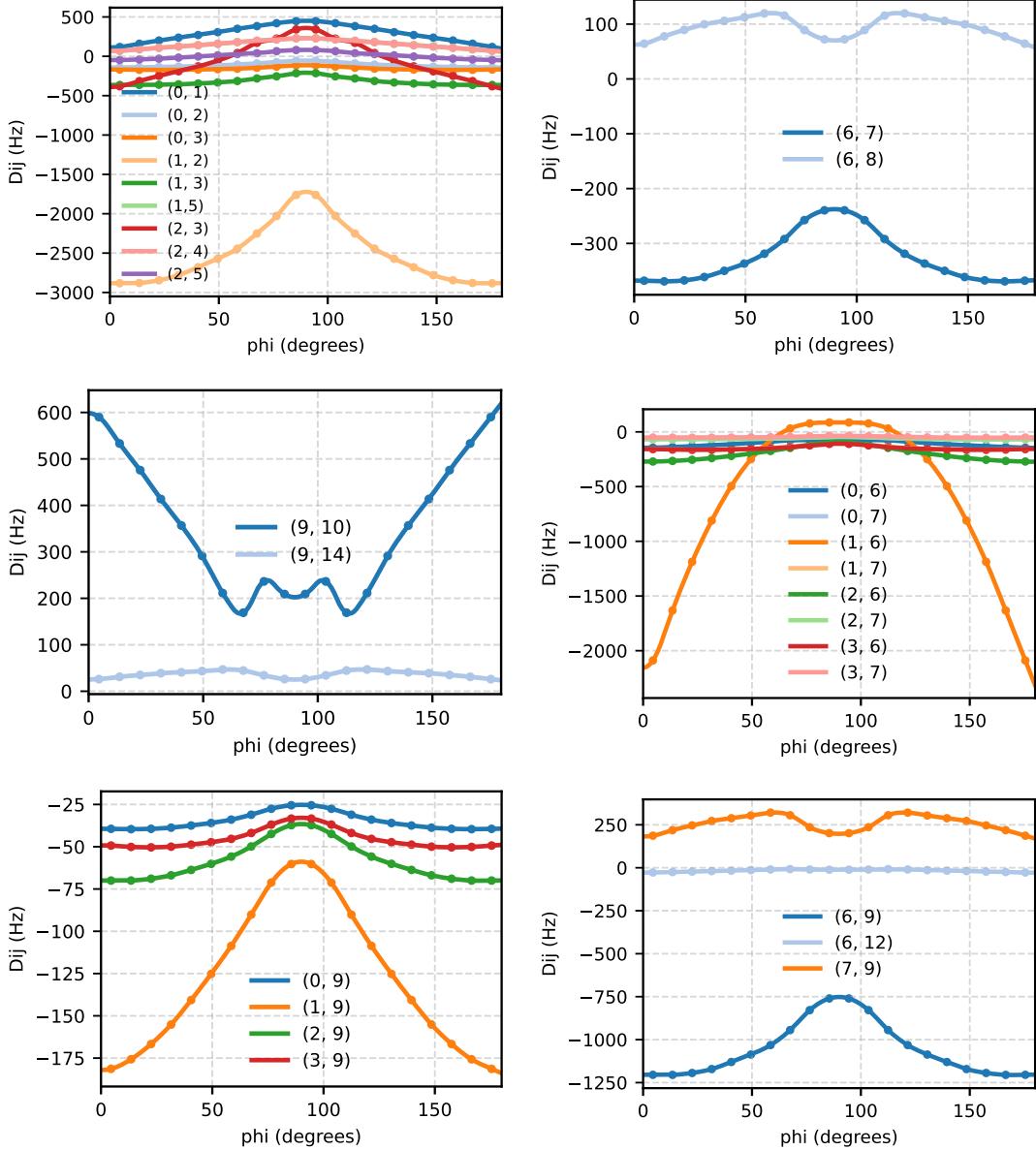


FIG. S30. All unique elements of  $D_{ij}$  as a function of the torsion angle. The first row shows the intra-ring elements, the second row shows the intra-methyl and inter-ring elements, and the third row shows the ring-methyl elements.

per trajectory. We subsequently define the standard error as  $S = \frac{\sigma_b}{\sqrt{M_b}}$  where  $\sigma_b$  is the standard deviation of block averages, and the  $M_b$  is the total number of blocks ( $M_b = 20 \times 50 \times 8 = 8000$ ). The resulting bounds on the 95% confidence interval for a given dipolar coupling matrix element is then evaluated as  $D_{ij} \pm 1.96S_{ij}$ .

#### F. QM correction to the PMF

A simple correction to the PMF along the DMBP dihedral angle can be evaluated by approximating the expectation value of Boltzmann reweighting factor. In order to evaluate this correction, we consider each DMBP solute molecule, whose coordinates are denoted by  $x$ , as a QM region. We then consider the solvent and all other solute molecules as an MM region, and their coordinates are denoted by  $R$ . The QM/MM energy is then given by

$$V_{QM/MM}(x, R) = V_{MM}(x, R) + V_{QM}(x) - V_{MM}(x), \quad (S80)$$

TABLE IV. Computed dipolar coupling values (in Hz) and their 95% confidence intervals for the 15Q system.

Parameter	Value	Parameter	Value
$D_{C,1}$	$237.27 \pm 22.49$	$D_{C,2}$	$-129.40 \pm 2.81$
$D_{C,3}$	$-170.46 \pm 4.85$	$D_{C,6}$	$-121.01 \pm 4.81$
$D_{C,7}$	$-67.43 \pm 1.92$	$D_{C,9}$	$-38.30 \pm 0.85$
$D_{1,2}$	$-2753.91 \pm 78.43$	$D_{1,3}$	$-352.21 \pm 10.13$
$D_{1,4}$	$-21.41 \pm 7.90$	$D_{1,5}$	$125.77 \pm 11.66$
$D_{1,6}$	$-844.83 \pm 30.18$	$D_{1,7}$	$-160.86 \pm 4.31$
$D_{1,9}$	$-152.64 \pm 3.97$	$D_{2,3}$	$-184.78 \pm 41.98$
$D_{2,6}$	$-235.79 \pm 6.27$	$D_{2,9}$	$-66.03 \pm 1.75$
$D_{3,6}$	$-161.29 \pm 4.25$	$D_{3,7}$	$-52.18 \pm 1.48$
$D_{3,9}$	$-49.42 \pm 1.32$	$D_{6,7}$	$-357.45 \pm 8.46$
$D_{6,8}$	$97.92 \pm 12.61$	$D_{6,9}$	$-1158.80 \pm 31.29$
$D_{6,12}$	$-19.38 \pm 2.65$	$D_{7,9}$	$268.22 \pm 34.05$
$D_{9,10}$	$411.47 \pm 49.54$	$D_{9,12}$	$38.31 \pm 4.89$

such that the PMF as a function of  $\phi$  for each solute molecule can be expressed as

$$\begin{aligned} F_{QM/MM}(\phi) &= -\beta^{-1} \log \left( \int dx dR e^{-\beta V_{QM/MM}(x,R)} \delta(\phi'(x) - \phi) \right) + C \\ &= -\beta^{-1} \log \left( \left\langle e^{-\beta(V_{QM}(x) - V_{MM}(x))} \delta(\phi'(x) - \phi) \right\rangle_W \right) + C'. \end{aligned} \quad (\text{S81})$$

Here, we have used Eq. (S80) for the energy and have defined the solvation free energy surface of the solute in the liquid crystal as

$$W(x) = -\beta^{-1} \log \left( \int dR \exp(-\beta V_{MM}(x,R)) \right). \quad (\text{S82})$$

In this simple scheme, the QM/MM PMF can then be obtained by evaluating the conditional expectation value of a Boltzmann reweighting factor that is proportional to the difference between the vacuum QM and vacuum MM energies using solute configurations from the ensemble generated by a molecular mechanics potential.

To simplify the analysis and circumvent statistical error, we then approximate the exponential to first order. After expanding the logarithm, we find that the QM/MM PMF can be approximated by a correction to the MM PMF,

$$F_{QM/MM}(\phi) \approx F_{MM}(\phi) + \Delta E(\phi), \quad (\text{S83})$$

where the correction is given by

$$\Delta E(\phi) = \frac{\int dR dx e^{-\beta V_{MM}(x,R)} (V_{QM}(x) - V_{MM}(x)) \delta(\tilde{\phi}(x) - \phi)}{\int dR dx e^{-\beta V_{MM}(x,R)} \delta(\tilde{\phi}(x) - \phi)}. \quad (\text{S84})$$

The correction is computed by binning the structures from all MM trajectories as a function of  $\phi$  and then computing the average of  $V_{QM} - V_{MM}$  within each bin. The results of this analysis are shown in Figure S31. The QM correction has the effect of lowering the barrier at 90 (and -90) degrees and raising the barrier at 0 (and 180) degrees.

#### G. 15Q learning experiment candidate PMFs

The 15Q learning experiment involves generating 9 candidate PMFs and then using NMR data to confirm which PMF generates  $D_{ij}$  coupling coefficients that give the lowest RMSE or generates MQC spectra with the lowest RMSE in peak position. Our strategy was to identify possible PMF curves and interpolate between them as a more efficient substitute to gradient descent on the parameter space of the model for the PMFs. This greatly simplifies the learning experiment to identifying which PMF matches the experiment the best. To generate the 9 candidate PMFs we interpolate between the PMF obtained from MD/MM, a vacuum torsion scan of the DMBP with the wb97X-D3BJ

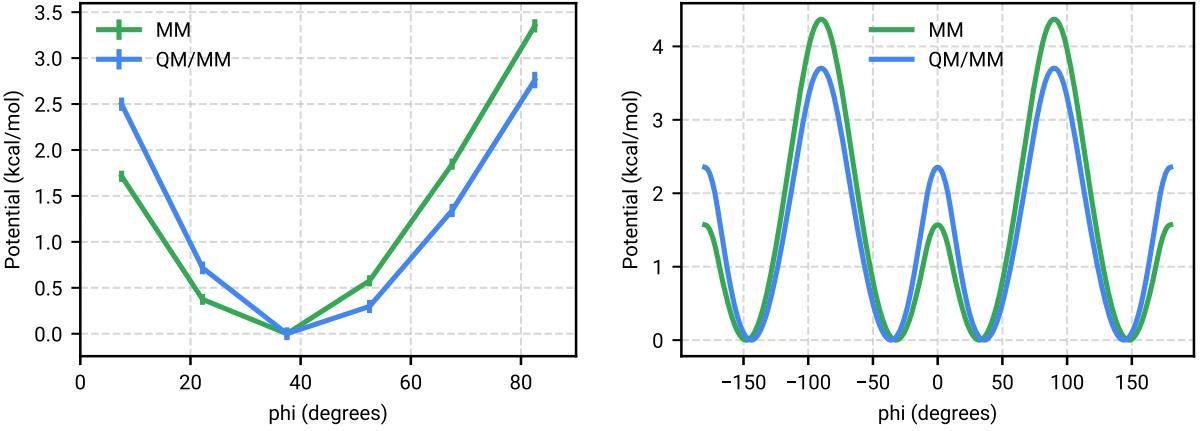


FIG. S31. The average MM and estimated QM/MM potential energy as a function of the torsion angle computed in the ensemble generated from MM trajectories (left). The QM-corrected PMF (all data is mapped to  $(-180, 180]$  and interpolated) (right).

functional using a def2-tzvp basis, and an artificial shifted double well potential. The artificial double well potential is obtained by fitting the parameters of a double-quantum 15Q Hamiltonian Eq. (S22) to an OTOC measured using the TARDIS-1 gate sequence on a sample of DMBP in 5CB liquid crystal at 300 Kelvin. The model used to fit the NMR data (Fig. S32(left) blue) is

$$U'(c, a, m, n) = c((1-a)\cos(x)^n + a\sin(x)^m) \quad (\text{S85})$$

where the optimal parameters are found to be  $c = 9.271505070720735$ ,  $a = 0.2620267312627955$ ,  $m = 12$ ,  $n = 10$ . The optimal parameters were determined by Gaussian Process optimization implemented in the Bayesian Optimization Python Package [99]. The optimization compared the L2-norm of the difference in OTOC decay points computed at times  $[2, 4, 6, 8, 9, 10, 12, 13]$  (index starting at zero) of the model (Eq. (S85)) and NMR data. Shown in Fig. S32 along with the NMR TARDIS-1 OTOC decay curve is the simulated double-quantum OTOC decay curve and the simulated TARDIS-1 OTOC decay curve using the optimal parameters. The right side of Fig. S32 plots Eq. (S85) at the optimal parameters (green) compared to the MD/MM PMF.

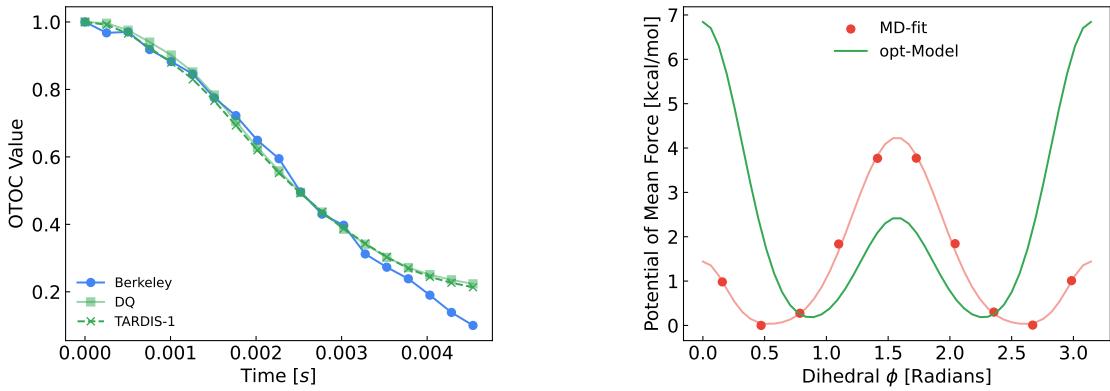


FIG. S32. (left) Comparison of OTOC decay curves simulated using a double-quantum Hamiltonian (green squares) and the full TARDIS-1 pulse sequence (green cross) and the NMR OTOC decay curve generated using the TARDIS-1 pulse sequence (blue). (right) PMFs for MD/MM (red) compared to the PMF determined by fitting the double-quantum OTOC to the 300K NMR data (green).

Following a similar double well fitting protocol as Ref. 96 the first seven even Fourier coefficients can be fit to approximate a double well between  $[0, \pi]$  for the dihedral angle of DMBP. We first fit the coefficients of Eq. (S86)

for MD/MM PMF, vacuum DFT torsion scan, and the aforementioned artificially fit at higher temperature PMF. The Fourier coefficients for these three curves labeled as candidate PMF 8, 4, 0 in Table V, respectively. We then linearly interpolate between the three curves producing curves [1, 2, 3] and curves [5, 6, 7]. All Fourier coefficients for all 9 curves are shown in Table V. Figure S33 shows all interpolated PMFs along with the dihedral angle with the minimum free energy.

$$U(\phi) = \sum_n u_n \cos(2n\phi) \quad (\text{S86})$$

Candidate PMF Index	0	1	2	3	4	5	6	7	8
$u_0$	2.2318	1.9306	1.6294	1.3282	1.0270	1.1817	1.3364	1.4910	1.6457
$u_2$	1.8669	1.4057	0.9445	0.4834	0.0222	-0.3806	-0.7834	-1.1861	-1.5889
$u_4$	2.1908	1.9118	1.6329	1.3539	1.0750	1.1241	1.1732	1.2224	1.2715
$u_6$	0.3404	0.3026	0.2647	0.2269	0.1890	0.1828	0.1766	0.1704	0.1642
$u_8$	0.2119	0.1809	0.1499	0.1189	0.0879	0.0962	0.1045	0.1127	0.1210
$u_{10}$	-0.0009	0.0063	0.0134	0.0205	0.0276	0.0283	0.0290	0.0297	0.0304
$u_{12}$	0.0012	0.0059	0.0105	0.0152	0.0199	0.0216	0.0233	0.0250	0.0267

TABLE V. Coefficients for each candidate PMF interpolated from the yellow to red to blue curves of Fig S33 where the interpolating function is defined in Eq. (S86).

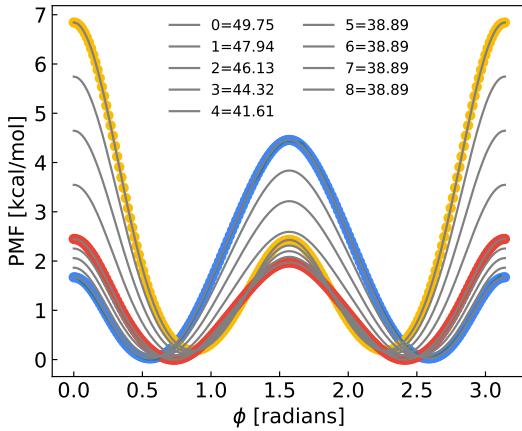


FIG. S33. 9 candidate PMFs (grey) interpolated between the artificial double well (yellow), vacuum DFT curve (red), and MD/MM (blue). The legend indicates the minimum of each PMF. The  $U_0$  PMF is the artificial double well, the  $U_4$  PMF is the vacuum DFT curve, and  $U_8$  PMF is MD/MM.

Finally, we note that though one of these candidate PMFs curves are obtained from a vacuum DFT torsion scan of DMBP the selection of this curve giving the lowest RMSE compared to NMR TARDIS-1 is purely coincidence. This coincidence does not mean vacuum DFT is sufficient for accurately modeling free energy surfaces in liquid crystal environments. In this work, the vacuum DFT curve is merely used as a candidate interpolation point. A full PMF computed with density functional theory would require *ab initio* MD with an appropriate density functional energy for forces. Though *ab initio* MD on the systems used in this work may be possible it would be a heroic computational effort and difficult to justify experimental expense.

#### H. 9Q Simulation details and results

Simulations of the 9Q (Toluene) system in EBBA liquid crystal were performed using molecular mechanics parameters from GAFF version 2.11 [54]. The initial structures were generated with Packmol version 21.0.0 [52] at a

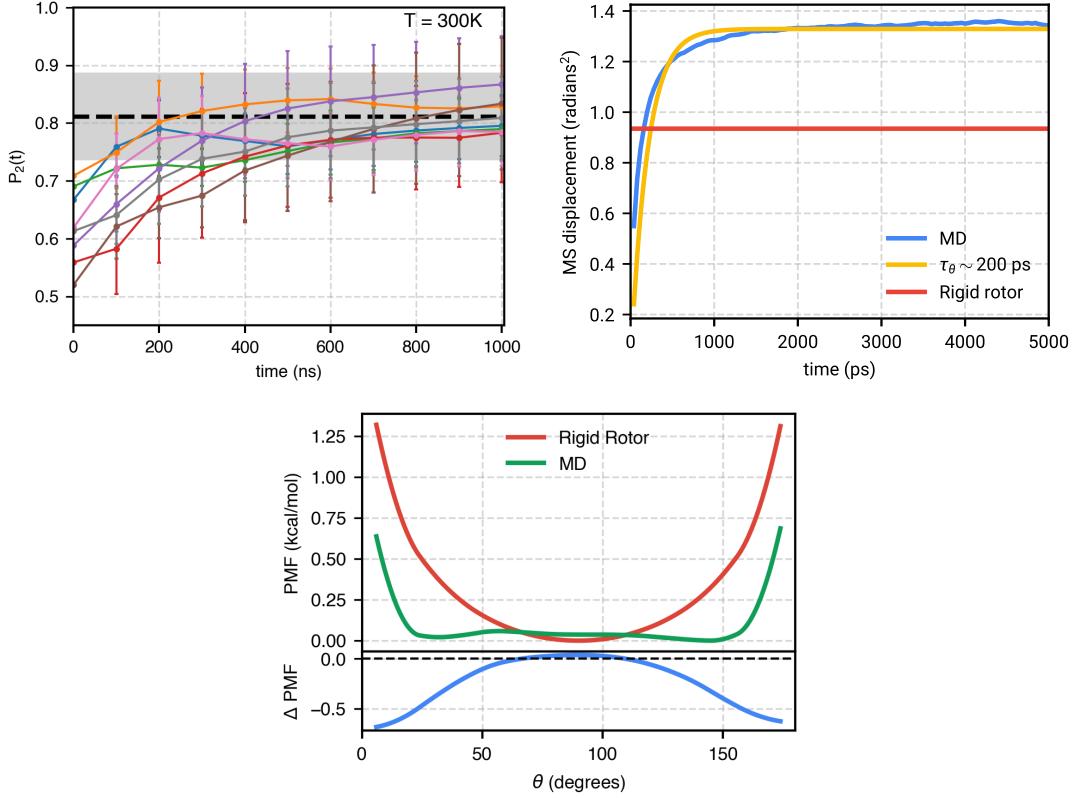


FIG. S34. The cumulative average of the orientational order parameter for an EBBA liquid crystal for 8 independent  $1\ \mu\text{s}$  trajectories simulated at  $T=300\ \text{K}$  (top-left). The mean-square displacement as a function of simulation time for the overall alignment of Toluene with the liquid crystal director (top-left). The potential of mean force along the liquid crystal alignment angle,  $\theta$  from the simulation (green) and assuming a rigid rotor (red). The difference between these (blue) reveals that interactions with the liquid crystal stabilize the parallel and anti-parallel configurations (bottom).

concentration of 0.33% wt% Toluene and a starting density of 1.00 g/ml. The EBBA molecules were approximately aligned along the x-axis of the simulation box with dimensions [52Å, 30Å, 30Å] by providing angle constraints in the Packmol input file. The initial structures were relaxed by minimizing the system at constant volume and then equilibrated with 500 ps of NPT dynamics at a temperature of 300 K. All simulations use a 2 fs time step with hydrogen bond constraints and a hydrogen mass of 4 amu. NPT calculations used a Monte Carlo barostat with updates every 25 steps. After equilibration, expectation values were evaluated from 8 independent from  $1\ \mu\text{s}$  production runs.

At a temperature of 300K, the EBBA liquid crystal becomes highly ordered, and multiple trajectories were observed to have (at least transiently) crystallized. This is reflected by the high  $P_2$  value and the large variance associated with each trajectory in Figure S34.

## V. TROTTER SCALING ESTIMATES

We will now show that we can use low order Trotter formulas to simulate the time evolution generated by the average Hamiltonians described in Section IIIK to obtain circuits of reasonable depth that are capable of simulating the OTOC experiment to sufficiently high accuracy. Given the error bars on the experiment, we consider an error in the OTOC value that is  $< 0.02$  to be acceptable. Unless otherwise noted, all simulations reported in this section are noiseless with the dipolar Hamiltonian derived from the  $u = 4$  PMF of the 15Q system.

### A. Analysis of Trotter accuracy

When using a product formula to simulate the time evolution for short times, the error in the OTOC will be one order higher in general, and two orders higher for the first-order Trotter. To show this result, the error due to using

an order- $p$  product formula can be expressed using the effective Hamiltonian

$$H_{\text{eff}} = H + \delta H t^p + \mathcal{O}(t^{p+1}), \quad (\text{S87})$$

where  $H$  is the true Hamiltonian, and  $\delta H$  is an error term in the form of multi-commutators of the terms in  $H$ . Now let  $M(t)$  be  $M$  after time evolution under  $H$ , so

$$M(t) = e^{-iHt} M e^{iHt}. \quad (\text{S88})$$

Then, if  $M_{\text{prod}}(t)$  is  $M$  with the time evolution under the product formula, the error can be given as

$$M_{\text{prod}}(t) - M(t) = -i[\delta H, M]t^{p+1} + \mathcal{O}(t^{p+2}). \quad (\text{S89})$$

The OTOC value is given by

$$C(t) = \text{Tr}(B^\dagger M(t) B M(t)), \quad (\text{S90})$$

so the error from using the product formula is

$$\begin{aligned} \Delta C(t) &= -it^{p+1} \text{Tr}(B^\dagger [\delta H, M] B M(t)) - it^{p+1} \text{Tr}(B^\dagger M(t) B [\delta H, M]) + \mathcal{O}(t^{p+2}) \\ &= -it^{p+1} \text{Tr}(B^\dagger [\delta H, M] B M) - it^{p+1} \text{Tr}(B^\dagger M B [\delta H, M]) + \mathcal{O}(t^{p+2}). \end{aligned} \quad (\text{S91})$$

In the second line we have used the fact that  $M$  only differs by  $M(t)$  by  $\mathcal{O}(t)$ , and so the difference can be bundled into  $\mathcal{O}(t^{p+2})$ . Next, we can simplify the error to

$$\begin{aligned} \Delta C(t) &= -it^{p+1} \text{Tr}([\delta H, M] B M B^\dagger) - it^{p+1} \text{Tr}([\delta H, M] B^\dagger M B) + \mathcal{O}(t^{p+2}) \\ &= -2it^{p+1} \text{Tr}([\delta H, M] M) + \mathcal{O}(t^{p+2}) \\ &= -2it^{p+1} (\text{Tr}(\delta H M M) - \text{Tr}(M \delta H M)) + \mathcal{O}(t^{p+2}) \\ &= -2it^{p+1} (\text{Tr}(\delta H M M) - \text{Tr}(\delta H M M)) + \mathcal{O}(t^{p+2}) \\ &= \mathcal{O}(t^{p+2}). \end{aligned} \quad (\text{S92})$$

The first line uses the cyclic permutation property, the second uses the fact that  $M$  and  $B$  commute, the third expands the commutator, and the fourth uses cyclic permutation again.

This increase in the order is for the total evolution time, rather than for later time steps. For approximation of time evolution using product formulae, the total time  $t$  is broken up into  $r$  steps, and there is error  $\mathcal{O}((t/r)^{p+1})$  for each step, for a total of  $\mathcal{O}(r(t/r)^{p+1})$ . The above reasoning is not exact for each of the  $r$  time steps (so these do not give error order  $\mathcal{O}((t/r)^{p+2})$ ). However, because the total error is  $\mathcal{O}(t^{p+1})$  in the total evolution time, for short overall time the above reasoning shows that the error is reduced.

In the case where the product formula is first order we can go further and show that one more term in the error cancels, giving the error term of order  $t^4$ . The conditions to be used in this derivation are that

1.  $B$  and  $M$  commute,
2.  $M M = I$ ,
3.  $M H M = -H_{\text{het}} + H_{\text{hom}}$ ,
4.  $H_{\text{het}}$  is a sum of commuting terms,
5. the terms in  $H_{\text{hom}}$  have Pauli operators on pairs of qubits excluding qubit 1,
6. the terms in  $H_{\text{het}}$  have Pauli operators on pairs of qubits *including* qubit 1, and
7.  $B$  is a tensor product of single-qubit operations.

Here,  $H_{\text{het}}$  and  $H_{\text{hom}}$  are the heterogeneous and homogeneous terms in the Hamiltonian, respectively. The first two conditions listed are those already used above to show that the error is one order higher.

The effective Hamiltonian for the first-order product formula can be written in the form

$$H_{\text{eff}} = H + H_1 t + H_2 t^2 + \mathcal{O}(t^3), \quad (\text{S93})$$

where  $H_1$  and  $H_2$  are in the form of commutators. Then when we consider the time-evolved  $M$ , we obtain error up to third order

$$\begin{aligned} M_{\text{prod}}(t) - M(t) &= -i[H_1 t + H_2 t^2, M]t - [H, [H_1, M]]t^3/2 - [H_1, [H, M]]t^3/2 + \mathcal{O}(t^4) \\ &= -i[H_1, M]t^2 - i[H_2, M]t^3 - [H, [H_1, M]]t^3/2 - [H_1, [H, M]]t^3/2 + \mathcal{O}(t^4). \end{aligned} \quad (\text{S94})$$

Here, the first two terms come from  $n = 1$  in the time evolution, whereas the other two come from  $n = 2$ . Substituting that into the OTOC gives error

$$\begin{aligned} \Delta C(t) &= \text{Tr}(B^\dagger(-i[H_1, M]t^2)BM) + \text{Tr}(B^\dagger MB(-i[H_1, M]t^2)) \\ &\quad + \text{Tr}(B^\dagger(-i[H_2, M]t^3)BM) + \text{Tr}(B^\dagger MB(-i[H_2, M]t^3)) \\ &\quad + \text{Tr}(B^\dagger(-i[H_1, M]t^2)B(-i[H, M]t)) + \text{Tr}(B^\dagger(-i[H, M]t)B(-i[H_1, M]t^2)) \\ &\quad + \text{Tr}(B^\dagger(-\frac{1}{2}[H, [H_1, M]]t^3)BM) + \text{Tr}(B^\dagger MB(-\frac{1}{2}[H, [H_1, M]]t^3)) \\ &\quad + \text{Tr}(B^\dagger(-\frac{1}{2}[H_1, [H, M]]t^3)BM) + \text{Tr}(B^\dagger MB(-\frac{1}{2}[H_1, [H, M]]t^3)) + \mathcal{O}(t^4). \end{aligned} \quad (\text{S95})$$

These lines correspond to the four terms in Eq. (S94), except the third line corresponds to cross terms between the first term in Eq. (S94) and the first-order term for the time-evolution in  $M$ . The first and second terms in the third line correspond to considering error in the first  $M(t)$  versus the second  $M(t)$  in the OTOC.

The first two lines of Eq. (S95) are zero using exactly the same reasoning as above in Eq. (S92). To show the remaining terms are zero, we simplify using cyclic permutation within the trace, and condition 1 that  $M$  and  $B$  commute, to give  $BMB^\dagger = M$ . This gives

$$\begin{aligned} \Delta C(t) &= -\text{Tr}([H_1, M]B[H, M]B^\dagger)t^3 - \text{Tr}([H_1, M]B^\dagger[H, M]B)t^3 \\ &\quad - \text{Tr}([H, [H_1, M]]M)t^3 \\ &\quad - \text{Tr}([H_1, [H, M]]M)t^3 + \mathcal{O}(t^4). \end{aligned} \quad (\text{S96})$$

For line one we have used a cyclic permutation. For lines two and three we have used  $BMB^\dagger = M$  and cyclic permutations, as well as combining the two factors of  $1/2$ .

Expanding these commutators then gives

$$\begin{aligned} \Delta C(t) &= -\text{Tr}(H_1 MBHMB^\dagger - MH_1 BHMB^\dagger - H_1 MBMH B^\dagger + MH_1 BMH B^\dagger)t^3 \\ &\quad - \text{Tr}(H_1 MB^\dagger HMB - MH_1 B^\dagger HMB - H_1 MB^\dagger MHB + MH_1 B^\dagger MHB)t^3 \\ &\quad - \text{Tr}(HH_1 MM - HMH_1 M - H_1 MHM + MH_1 HM)t^3 \\ &\quad - \text{Tr}(H_1 HMM - H_1 MHM - HMH_1 M + MHH_1 M)t^3 + \mathcal{O}(t^4) \\ &= -2\text{Tr}(H_1 BMHMB^\dagger - H_1 BHB^\dagger)t^3 - 2\text{Tr}(H_1 B^\dagger MHMB - H_1 B^\dagger HB)t^3 \\ &\quad - 4\text{Tr}(H_1 H - H_1 MHM)t^3 + \mathcal{O}(t^4). \end{aligned} \quad (\text{S97})$$

We have made several simplifications using the condition that  $B$  and  $M$  commute, as well as that  $MM = I$  (condition 2). Now conditions 3 means that  $MHM$  flips the sign of the heterogeneous part, and leaves the homogeneous part unchanged, so gives

$$H - MHM = 2H_{\text{het}}. \quad (\text{S98})$$

The error therefore simplifies to

$$\Delta C(t) = 4\text{Tr}(H_1 BH_{\text{het}}B^\dagger)t^3 + 4\text{Tr}(H_1 B^\dagger H_{\text{het}}B)t^3 - 8\text{Tr}(H_1 H_{\text{het}})t^3 + \mathcal{O}(t^4). \quad (\text{S99})$$

We can prove the traces are zero by expanding the expressions inside the trace as sums of Pauli strings. The only Pauli string that can give nonzero trace is the identity. The term  $H_1$  in the error corresponds to commutators of terms in the Hamiltonian. Because of condition 4,  $H_1$  will not contain any commutators between terms in  $H_{\text{het}}$ . We can then consider commutators between terms within  $H_{\text{hom}}$ , or between those in  $H_{\text{hom}}$  and those in  $H_{\text{het}}$ . Condition 5 means that commutators between terms in  $H_{\text{hom}}$  must have the identity on qubit 1. Condition 6 means that  $H_{\text{het}}$  has a non-identity Pauli operator on qubit 1, and so any product between terms in  $H_{\text{het}}$  and commutators between terms in  $H_{\text{hom}}$  also has a non-identity Pauli on qubit 1. As a result its trace is zero.

This means the remaining terms in  $H_1$  we need to consider are commutators of terms in  $H_{\text{hom}}$  and  $H_{\text{het}}$ . If a term in  $H_{\text{het}}$  does not commute with a term in  $H_{\text{hom}}$ , then the two terms must have Pauli operators at a matching location.

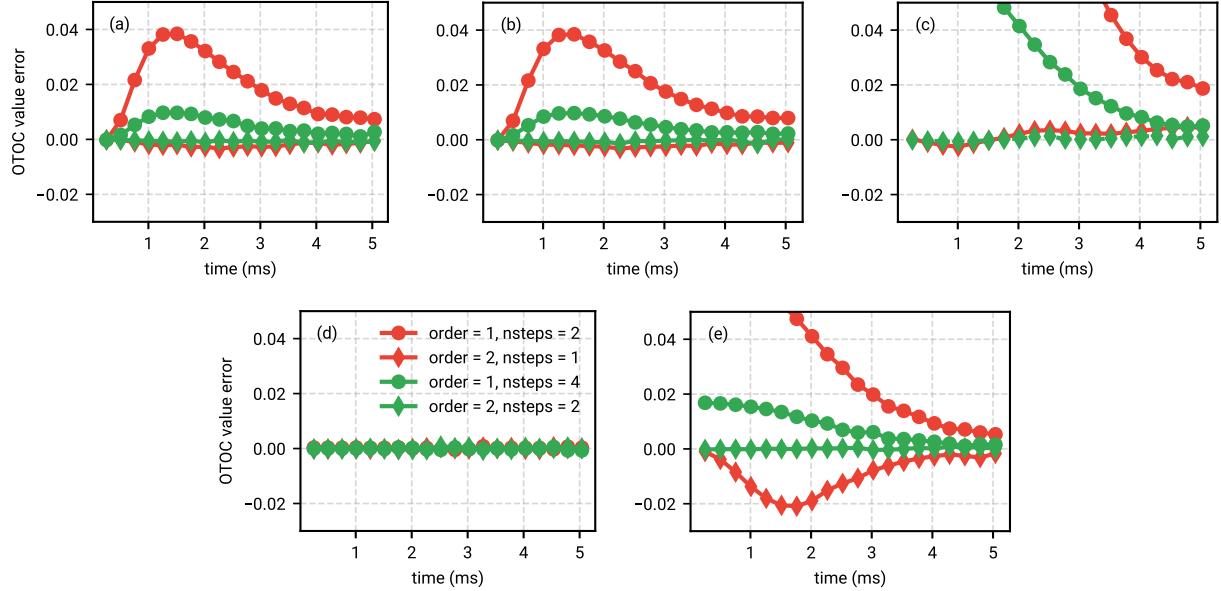


FIG. S35. The error in the OTOC value as a result of individually approximating each unitary with 1st or 2nd order Trotter. In particular, we show the effect of Trotterizing the forward evolution ( $\tilde{U}_\rightarrow$ )(a), the backward evolution,  $\tilde{U}_\leftarrow$ , (b), the BLEW12 term,  $\tilde{U}_{BL12}$ , (c), the initial  $\pi$  pulse,  $\Pi_y^C$ , (d) and the final  $-\pi$  pulse,  $\Pi_y^H \Pi_y^C$ , (e). Though these terms have different error profiles, we find that in all cases 2nd order Trotter outperforms the 1st order formula for roughly the same circuit depth and only a very small number of steps are needed for an error that is small relative to a target accuracy of 0.02 in the OTOC curve.

(They cannot have two matching locations because of conditions 5 and 6.) As a result, the commutator must not be equal to the identity at *three* locations; qubit 1 and the two qubits where the term from  $H_{hom}$  has Pauli operators. Because terms from  $H_{het}$  have Pauli operators at only two locations, the product of these terms and the commutator must again have at least one non-identity Pauli operator. This reasoning therefore implies that  $\text{Tr}(H_1 H_{het}) = 0$ .

We also have terms in Eq. (S99) with  $BH_{het}B^\dagger$  and  $B^\dagger H_{het}B$ . Condition 7 is that  $B$  only acts locally on qubits, and therefore leaves identity operators in the Pauli strings unchanged. As a result,  $BH_{het}B^\dagger$  and  $B^\dagger H_{het}B$  are still sums of terms with non-identity operators at only two locations. Hence, the same reasoning as for  $\text{Tr}(H_1 H_{het})$  holds, and the remaining traces in Eq. (S99) are equal to zero. This shows that the order  $t^3$  terms all cancel, and the OTOC error must be at least order  $t^4$  when using the first-order product formula.

Again, the reasoning is not exact for time steps that do not start at time 0, so we do not obtain error  $\mathcal{O}(r(t/r)^4)$  when using  $r$  time steps. The primary condition that is violated for steps with a nonzero initial time is that  $B$  and  $M$  commute, because  $M$  is replaced with  $M(t_0)$ . The time-evolved  $M$  no longer commutes with  $B$ , though for short times it will approximately commute. For first-order product formulae, condition 3 is violated as well, but again will be approximately true for short times. These considerations mean that the error is still significantly reduced when using first-order product formulae over multiple time steps.

## B. Numerical tests of Trotter accuracy

Within the AHT framework, we simulate the TARDIS-1 pulse sequence given in Eq. (S46) using a low-order Magnus expansion for each unitary. For the 15Q system, we use only the first term (Magnus-0), and we order terms within each Trotter step by grouping them into commuting partitions which are then applied in series. The Trotter error for each term is shown in Figure S35. Given the discussion in Section V B, it is not surprising to see that for 1st and 2nd order Trotter, only a very small number of time steps (1 or 2) is needed for accurate propagation. In Figure S36 we show the Trotter error using one 2nd order Trotter step for each unitary with the exception of  $-\pi$ -pulse for which we use 2 steps. The numerical error for this scheme is shown in Figure S36.

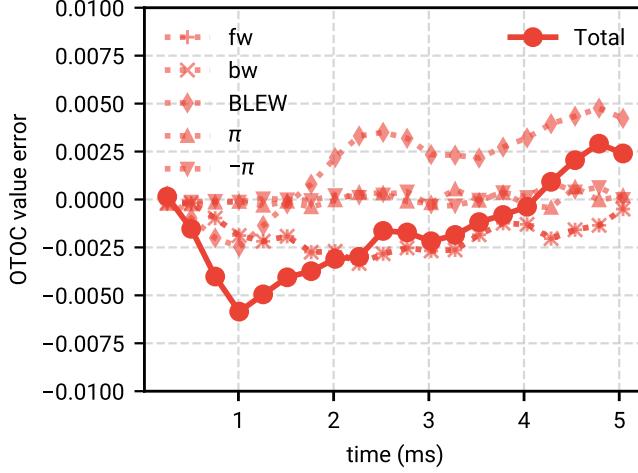


FIG. S36. The error in the final OTOC value using 2nd order Trotter for all unitaries. A single time step is used for each unitary except for the  $-\pi$  pulse which uses two Trotter steps. This scheme results in an error that is significantly less than our target accuracy of 0.02.

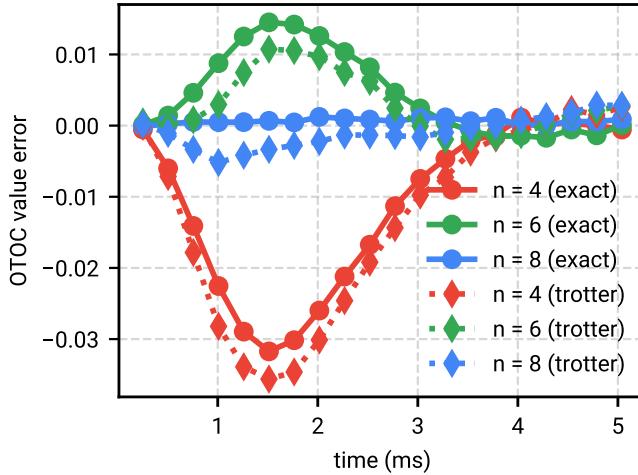


FIG. S37. The error in the simulated OTOC curves for different truncations of the rotation angles. Here  $n$  is the number of bits of decimal precision and the results are shown for both exact evolution and for the Trotter scheme described in Section V B.

### 1. Precision and resources

In Figure S37 we show that truncating the angles in the Trotterized propagators to 6 fractional bits of precision (this roughly corresponds to an allowed error of  $2^{-7} \approx 0.0078$  in the rotation angles) provides sufficient accuracy in the simulated OTOC curve. Given these precision requirements, we compute the resources needed to simulate the OTOC curves using the FLuid Allocation of Surface code Qubits (FLASQ) cost model for early fault-tolerant quantum computing (manuscript in preparation). The results are shown in Figure S38. These resource estimates indicate that these simulations are feasible on realistic models of early fault-tolerant devices. Note that we have not performed a careful optimization of these circuits and the FLASQ model inherently neglects certain optimizations, so we expect that it is possible to improve on these cost estimates.

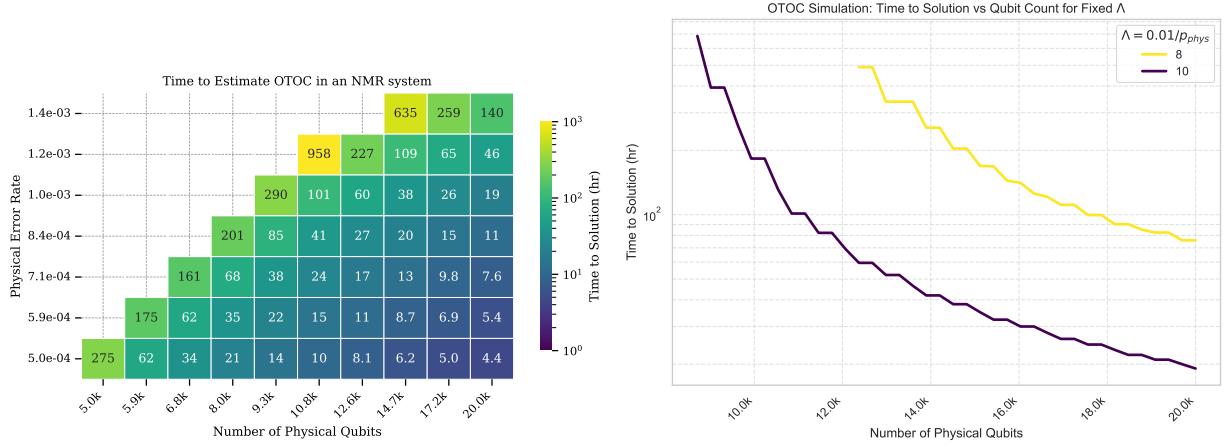


FIG. S38. Resource estimates for the longest circuit (the final time point) needed to simulate the 15Q system. (left) Total runtime in the FLASQ model for different error rates and numbers of physical qubits. (right) Time to solution as a function of the number of physical qubits for different fixed values of  $\Lambda$ .

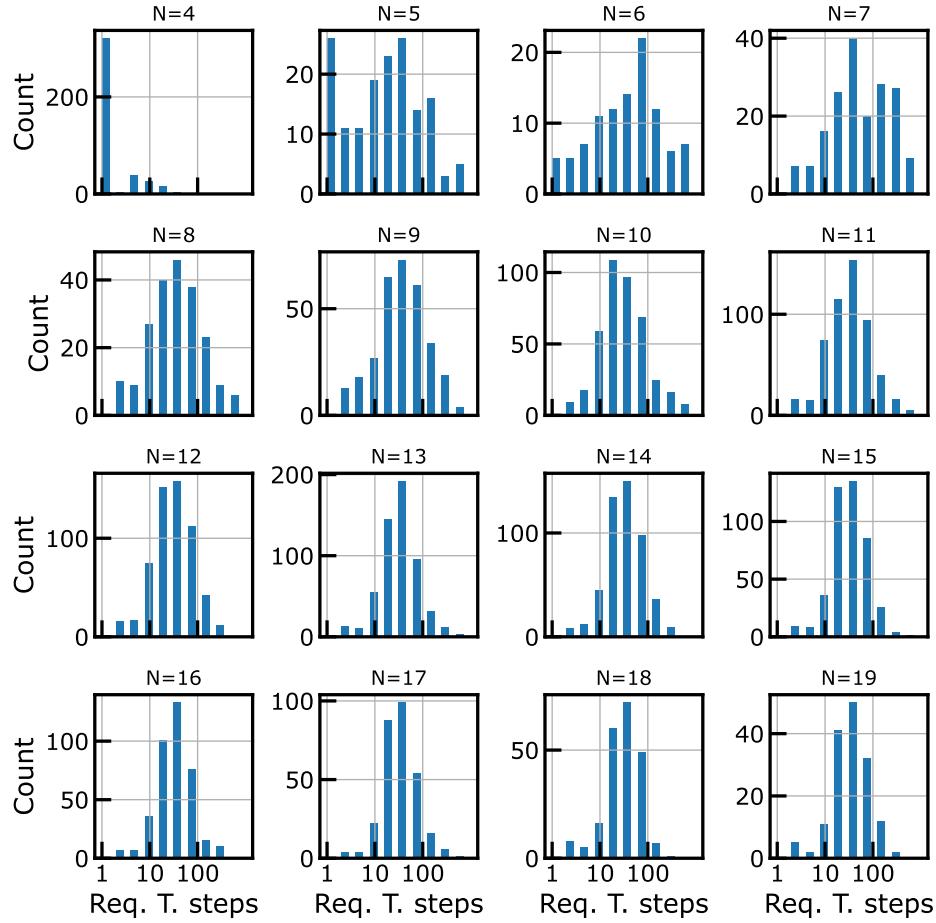


FIG. S39. Histograms of the number of Trotter steps required to converge OTOCs on artificial dipolar spin systems (see text) to an error of 0.02 at times when  $C(t) \in [0.4, 0.5]$ . The x-axis is plotted in log space.

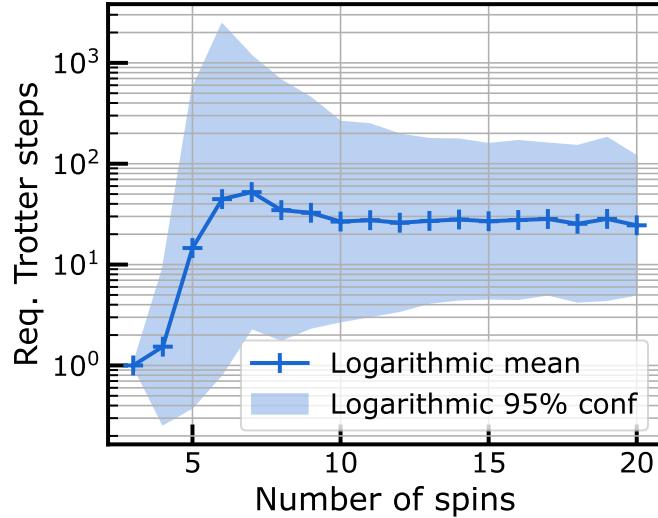


FIG. S40. Plot of the the logarithmic mean and 95% confidence interval of the data in Fig. S39.

### C. Extrapolating the cost of Trotterization in dipolar spin systems

One of the most important questions left open in our work is the quantum resource requirements to extend these methods to a practical demonstration of quantum advantage. Here, we make a brief attempt at an order-of-magnitude estimate using a naive first-order Trotterization with no additional optimization. The key challenge lies in estimating the number  $K$  of all-to-all-coupled Trotter steps required to converge the OTOC. Without further optimization, executing a Trotter step in both the forward and backward pass on  $N$  qubits using a swap network requires  $N(N - 1)$  interactions, or  $G(N) = 3N(N - 1)$  CZ gates. For  $N = 50$  ( $N = 60$ ), we have  $G(50) = 7350$  ( $G(60) = 10620$ ) CZ gates per Trotter step; the total number of gates per circuit is  $K \times G$ .

The required number  $K$  of Trotter steps is highly system and quantity specific. Here we define a simplified model system of the nuclear spin systems studied in this work, and then outline a protocol to estimate an ensemble of Trotterized OTOCs throughout this system. We define a model system by scattering spin positions  $\mathbf{r}_i$  randomly in a large 3D box at a fixed density  $\rho_s = 0.1 \text{ \AA}^{-3}$ , and with a minimum distance between spins of  $1.5 \text{ \AA}$ . This density is chosen to match the density of spins in a protein (Eq. (S64)), and the minimum distance is chosen to be slightly smaller than the  $1.7 \text{ \AA}$  distance between methyl spins (larger distances created packing issues due to the random scattering of spins). From this, we generate a set of dipolar couplings  $d_{ij}$  using Eq. (S8), assuming the spins are protons fixed in space and the magnetic field points in the  $\hat{z}$  direction.

Next, we choose pairs of measurement and butterfly spins  $(m, b)$  at random separated by between  $4 \text{ \AA}$  and  $12 \text{ \AA}$ , sufficiently within the edges of the box to prevent observing the boundaries. (For efficiency reasons, we generate 20 unique pairs within cubic regions of width  $20 \text{ \AA}$  plus an additional gap to the system boundary; we expect this volume to be large enough to prevent significant correlations between different problem instances.) For efficiency, we desire to simulate OTOCs between these spin pairs including only relevant spectator spins  $j$  (i.e. spins that contribute to the OTOC). To achieve this, we construct a classical estimate of the time-of-flight  $t_j^{(c)}$  for information to spread from the measurement spin through a spectator spin to the butterfly. This estimate  $t_j^{(c)}$  is calculated by treating the spins as a weighted graph with adjacency matrix  $s_{ij} = d_{ij}^{-1}$ , which we interpret as the time taken for information to propagate from spin  $i$  to spin  $j$ . Under this interpretation, the time taken for information to propagate from the measurement spin  $m$  through spin  $j$  to the butterfly spin  $b$  is equal to the minimum weight of a path  $m \rightarrow j \rightarrow b$ , which can be efficiently calculated using Dijkstra's algorithm.

Once the above time-of-flight calculation is complete, we order spins  $j$  by their time-of-flight, and added them from the shortest  $t_j^{(c)}$  to the largest until the OTOC converges to a value below  $C(t) = 0.5$ . (We approximate the infinite temperature OTOC by averaging over initial states until the standard error drops below 0.004, which we judge to be sufficiently small to be irrelevant.) As we add each spin, we first converge the Trotterization of the  $n$ -spin system by repeatedly doubling the number of Trotter steps  $k$  and estimating  $c_{n,k}(t)$  until we find  $|c_{n,k} - c_{n,k}| < 0.02$ , yielding an estimate  $K(N) = k/2$ ,  $C_n(t) = c_{n,k}(t)$ . We then test whether  $C_n(t) = C_{n-1}(t)$ ; if so we report the pair  $(N, K) = (n - 1, K(n - 1))$  as the last steps were used to confirm convergence. When we have not converged, we

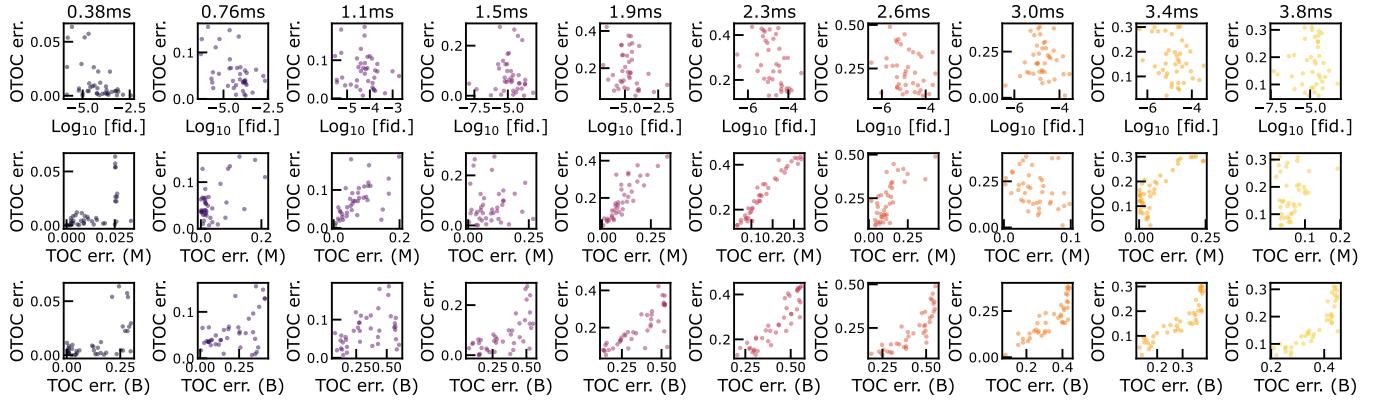


FIG. S41. Plots of Trotter error as measured via OTOC error, unitary fidelity, or TOC error, for a range of different Trotterizations of the toluene DQ Hamiltonian differing by time, number of Trotter steps per timestep, and the ordering within the swap network. Points are coloured to separate different time instances.

update the time  $t$  to ensure the OTOC is near but below 0.5, add a new spin, and restart the process. (Due to a large chosen time gradation, this results in a final OTOC value  $C(t) \in [0.4, 0.5]$ ; we reject outlying experiments that fall below 0.4.)

In Fig. S39, we plot histograms of the number of steps to converge, on a logarithmic axis, separated for different values of  $N$ . We observe that the distribution is relatively symmetric within this log space, with a mode around  $K = 32$  for all  $N > 11$ . Beyond this surprisingly small value the distribution appears to be converging; in Fig. S40 we plot the mean and  $2\sigma$  interval of our obtained estimates of  $K$  as a function of  $N$ . We observe that the resulting mean appears to have converged quickly to a number around  $K = 30$ . This would correspond to needing  $2.2 \times 10^5$  ( $3.2 \times 10^5$ ) CZ gates for a 50 (60) qubit simulation. The standard deviation trend is slower: at  $N = 20$  we observe the population spread is over roughly an order of magnitude. Extrapolating this over three additional decades is highly unreliable, and significantly more work needs to be done to determine the cost of Trotterization at larger system sizes. However, we are optimistic that these results may suggest instances of OTOC estimation at 50 – 60 qubits requiring  $10^5 – 10^6$  two-qubit gates, and this work forms our justification for this claim in the main text.

#### D. Correlation between different measures of Trotter error

Meaningfully capturing the error in Trotter approximations is a challenging task. Traditionally, this has been estimated by holistic metrics such as unitary fidelity, or via matrix norms on the multiplicative or additive error term [35], which provide upper bounds on other desirable quantities. However, ultimately one is only concerned with the accuracy of the data extracted at the end of the computation. As we demonstrated in Sec. VA, this can be even asymptotically better than unitary fidelity metrics, due to error terms cancelling out. Given the vast freedom in designing product formulas, we would like to optimize even further, but this runs into two problems. The first is a spoofing problem: we can achieve arbitrary single expectation values using short-depth unitaries that bear no physical resemblance to the original Hamiltonian, so the error in such an expectation value is a completely unreliable product formula metric. We have attempted to mitigate this concern in our AlphaEvolve-optimized circuits by testing across a landscape of times and Hamiltonians, and by inspecting the output product formula code (see the main text and Sec. VI for further details); further investigation here is needed. The second is a prediction problem: we desire the ability to estimate the error in an expectation value without being able to estimate the expectation value itself, which requires finding correlated error metrics.

In this section, we briefly investigate the challenge in solving the prediction problem, by comparing different error metrics on a range of product formula approximations to the evolution of toluene under the double quantum Hamiltonian approximation (Eq. (S110)). We generate approximate unitaries  $U_{\text{pf}}(t) \approx e^{-iHt}$  using swap networks with randomly chosen qubit orderings and a range of Trotter step sizes  $dt$ . In Fig. S41, we measure the error in these approximations with four metrics:

1. The OTOC error

$$\text{OTOC err.} = \frac{1}{2^N} |\text{Trace}[MU_{\text{pf}}^\dagger(t)BU_{\text{pf}}(t)MU_{\text{pf}}^\dagger(t)B^\dagger U_{\text{pf}}(t)] - \text{Trace}[Me^{iHt}Be^{-iHt}Me^{iHt}B^\dagger e^{-iHt}]|, \quad (\text{S100})$$

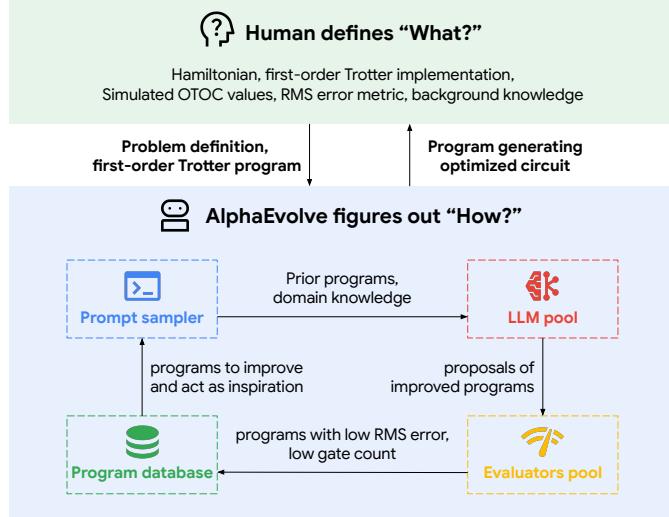


FIG. S42. Overview of AlphaEvolve optimization cycle as presented in [5]

where  $M = X_{13}\text{C}$ , and  $B$  is the single-qubit approximate butterfly pulse to the BLEW-12 sequence (Eq. (S14)) that acts on the methyl protons.

2. The unitary fidelity,

$$\text{fid.} = \frac{1}{2^N} \text{Trace}[U_{\text{pf}}^\dagger e^{iHt}]. \quad (\text{S101})$$

3. The TOC error on the measurement operator (using the same definitions as above),

$$\text{TOC err. (M)} = \frac{1}{2^N} |\text{Trace}[MU_{\text{pf}}^\dagger(t)MU_{\text{pf}}(t)] - \text{Trace}[Me^{iHt}Me^{-iHt}]|. \quad (\text{S102})$$

4. The TOC error on the butterfly operator (using the same definitions as above),

$$\text{TOC err. (B)} = \frac{1}{2^N} |\text{Trace}[B^\dagger U_{\text{pf}}^\dagger(t)BU_{\text{pf}}(t)] - \text{Trace}[B^\dagger e^{iHt}Be^{-iHt}]|. \quad (\text{S103})$$

Overall, we observe that the predictive power of any of the other three metrics for the error in an OTOC is somewhat limited. The unitary fidelity is a particularly bad metric: despite the OTOC being often well approximated by various product formulas shown here, this never rises above  $10^{-2}$  (optimally the unitary fidelity should be 1), and we observe no correlation between this metric and the OTOC. This should not be surprising given the results in Sec. V A, as the leading contributions to the error in the unitary fidelity cancel out in the OTOC error. The result here implies that there is little correlation between the contributions that cancel and those that remain. We observe some correlation between the two TOC errors and the OTOC error, though this correlation is curiously inconsistent across different times, and not terribly strong for the most part (with a notable exception at 2.3 ms). This demonstrates the need for further investigation into the mechanisms behind Trotter error to improve our predictive power.

## VI. ALPHAEVOLVE CIRCUIT SIMULATION

We employed AlphaEvolve, an evolutionary coding agent that uses LLMs, to discover Python functions which generate quantum circuits for approximating Hamiltonian evolution over a landscape of time and Hamiltonian parameters. The evolutionary process was seeded with a human-written, first-order Trotter formula for a 15-qubit chain and iteratively improved programs through an LLM-driven mutation, evaluation, and selection loop, as shown in Fig S42. The optimization targeted the root mean square error (RMSE) of the OTOC averaged across the entire parameter landscape, with the LLM's mutation prompts guided by a mixture of domain-specific instructions and a tabular error matrix feedback to perform in-context learning on high-error regions.

To accelerate convergence and enhance search space exploration, we ran multiple independent AlphaEvolve instances that were periodically synchronized with the best-performing solution. We also overcame a key computational bottleneck in the evaluation step by implementing a highly parallelized KAK decomposition for efficient CZ gate counting, a tool which the final evolved programs learned to utilize for more efficient circuit generation.

In order to assess the efficacy of prompting AlphaEvolve with problem- and domain-specific knowledge, we analysed the learning curve of the program evolution. To get a high-level overview, we group prompt instructions given to AlphaEvolve into five general categories presented in Table VI. The generic prompts (GENERIC) are simply general guidelines to improve the program. The prompts grouped under domain knowledge, topology, and gate count span quantum computing knowledge (THRIFT [61], higher order Trotter formulas), molecular structure, and circuit compilation respectively. Parameter tuning then explicitly instructs AlphaEvolve to adjust coupling strengths and time steps based on the given parameters.

To understand how AlphaEvolve proceeds and get a sense of which of the instruction prompts are most beneficial, Figure S43(left) shows an optimization curve with the improvement in RMSE against evolution steps. The colored background annotates which category of prompt was used to obtain the improvements. In addition, in Figure S43(right) we provide a histogram of cumulative importance of the prompt categories used. We note that we see a lot of variation from run to run, so the lineage of this particular program is not a prescription of how to proceed optimally; repeated runs will produce a completely different result. We do note however that the performance (in terms of RMSE reached by the end of the optimization) can be reliably achieved.

The parameter tuning prompts yield most of the improvements, as seen in Figure S43(left), followed by instructions to utilize the molecular topology. Domain knowledge, generic and gates-specific prompts less frequently yield improvements. When summing the net contributions of each prompt towards the overall aggregated min, mean, and RMSE scores, we find a similar picture: the overwhelming improvement comes from parameter-specific instructions, followed by topology. With respect to the max error, domain- and gate-specific prompts seem detrimental; however, we note that many of the parameter- and topology-specific improvements only make sense after the domain-specific improvements have been applied; the order in which mutations take place naturally matters a lot.

To further give a high-level intuition of what kinds of modifications AlphaEvolve suggests, we present a translation of the full, optimized Trotter generator function into pseudocode in Figure S44. As the full evolved program has a length of roughly 1300 lines of code and thus is hard to interpret directly, we prompt an LLM to convert the final AlphaEvolve optimized program into pseudocode as a starting point, and then manually try to match the extracted concepts with the given code. We highlight that the overall program structure and concepts used are highly specific to the simulation target at hand, and that some of the concepts invoked may be implemented in an incomplete, wrong

TABLE VI. Additional prompt instructions by category

Domain Knowledge	Topology	Parameter tuning	Gate count	Generic
■ DOMAIN	■ TOPO	■ PARAM	■ GATES	□ GENERIC
Implement THRIFT [61] by identifying a strongly coupled subset of atoms	Adjust couplings based on vicinity to “unimportant” atoms	Adjust step dt based on parameter $p$	Reduce the circuit depth based on $t$ and $p$	Suggest an idea to prove the code
Implement a 2nd order Suzuki-Trotter formula	Adjust frequency of some couplings based on shift parameter $p$	Adjust effective time step with an envelope that may depend on atom, $p$ , butterfly/measurement indices	Increase the circuit depth based on $t$ and $p$	Suggest an idea to optimize the code
Implement a 4th order Suzuki-Trotter formula	Handle components of interaction graph independently	Adjust couplings based on $p$	Build a multi-pass approach, filtering out unimportant gates/adding further steps	
Invent new gates based on the Weyl piece (use Cartan decompositions of $SU(4)$ )	Lightcone based on but- terfly and measurement operator support	Derive scaling exponents from $p$ and Suzuki-Trotter formula	Implement a 4 <sup>th</sup> order exponentiate couplings by it	
Implement fractional Trotter steps/partial layers	Keep the $^{13}\text{C}$ atom state/partial layers			

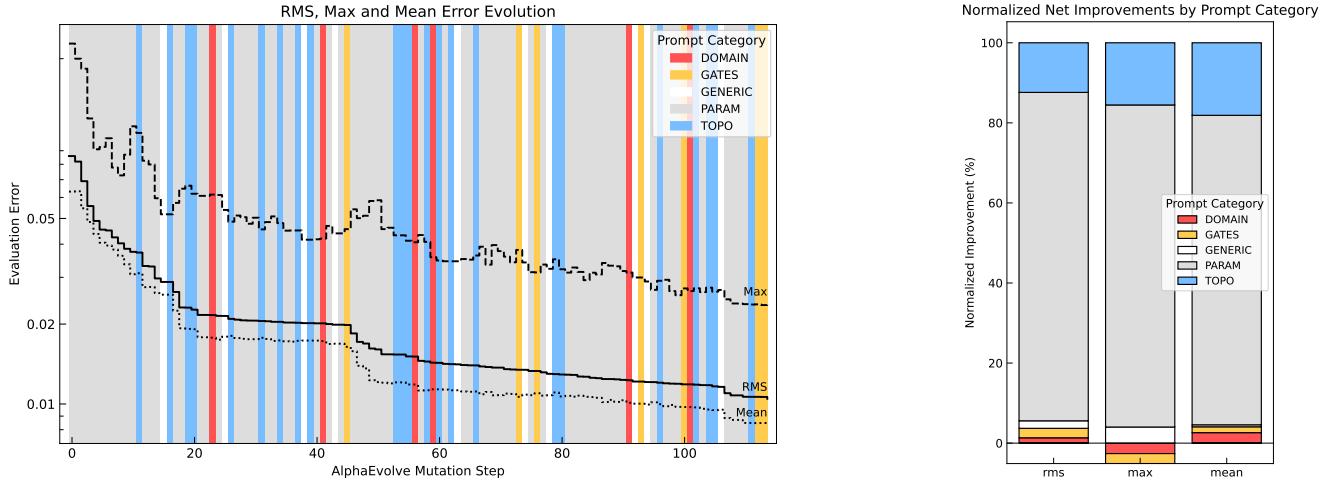


FIG. S43. Prompt category contributions. (left) Evaluation error as function of mutation number. Bold line represents optimized metric (rms error), dashed lines correspond auxiliary tracked metrics (mean, min). The colored stripes indicate the type of prompt category used that resulted in this score change, as listed in Table VI. (right) Histogram of a cumulative prompt category importance. The importance is calculated as a weighted sum over all score improvements for a given category.

or obscure way; if implemented from first principles, these concepts might individually work quite differently, and their efficacy could depend on the order in which the successive mutations were applied.

The concepts AlphaEvolve used in the final program in Figure S44 are outlined below. We note that all of them find their origin in at least one of the prompt instructions from Table VI, such as THRIFT scaling, or utilizing the time and landscape parameters to attenuate the parameters in a nonlinear fashion. The two-pass approach was also translated into a major component, as well as light cone filtering as a final pruning step.

- *Adaptive Hamiltonian Scaling.* The Hamiltonian couplings are scaled, by modulating them with an exponent dependent on the parameter  $p$ , via  $J \mapsto \text{sign}(J)|J|^{a(p)}$  for some polynomial  $a(p)$ .
- *Hamiltonian Decomposition.* The Hamiltonian is decomposed based on its interaction graph.
- *Dynamic Budgeting.* In a two-step approach we first estimate the CZ cost of a single, pruned Trotter step across all components. The number of Trotter steps is then the minimum of a heuristic desired  $N$  (which depends on  $t$  and  $p$ ) and the maximum  $N$  allowed by the CZ budget.
- *Time Step Modulation.* Scale time steps via  $t \mapsto \text{sign}(t)|t|^{e(p,t)}$  for some polynomial  $e(p,t)$ . This implements a time-dependent envelope, global THRIFT scaling, and scaling dependent on the importance of the qubits (vicinity to measurement indices, vicinity to important atoms, e.g. the carbon atom).
- *Adaptive Symmetric Second-Order Trotterization.* Uses a symmetric decomposition within each layer based on the formula  $\exp(-i(A+B)\Delta t) \approx \exp(-iAx\Delta t) \exp(-iB\Delta t) \exp(-iA(1-x)\Delta t)$  for some  $x = x(p,t)$ ; the Hamiltonian is split into  $A$  (background) and  $B$  (perturbation) in a THRIFT-inspired way.
- *Lightcone Pruning.* Gates that do not causally affect the final measured observables are removed.

Overall, while it remains unclear to what extent the above instructions were correctly implemented as intended and whether the concepts could function well independently, we see this experiment as a promising first step towards automatically translating scientific knowledge into optimized domain-specific tools.

## VII. CLASSICAL RESOURCE SCALING ANALYSIS ON A SIMPLIFIED MODEL

As a preliminary assessment of the classical resources required to simulate out-of-time-ordered correlators (OTOCs) for nuclear spin ensembles, we consider a spin-1/2 double-quantum Hamiltonian with long-ranged interactions on an  $N$ -site chain

$$H = \sum_{i < j} \frac{J}{4|i-j|} (X_i X_j - Y_i Y_j) \quad (\text{S104})$$

FIG. S44. Pseudocode describing the AlphaEvolve-generated algorithm to generate Trotter circuits.

```

1 Function GenerateTrotterNetwork_Improved(
2     Time t
3     CZ budget MaxCZ
4     Hamiltonian couplings H
5     Parameter p
6     Butterfly/Measurement indices I
7     ):
8     // Concept a: Adaptive Hamiltonian Scaling
9     H_scaled := ApplyAdaptiveCouplingScaling(H, p)
10    // Concept b: Hamiltonian Decomposition
11    Components := DecomposeHamiltonian(H_scaled, I)
12    // Concept c: Dynamic Budgeting (Two-Pass Approach)
13    TotalCostPerStep := 0
14    For C in Components Do
15        DummyCircuit := TrotterStep(C, t, 1, p)
16        TotalCostPerStep += CountCZGates(DummyCircuit)
17    End For
18    N := CalculateTrotterSteps(t, p, MaxCZ, TotalCostPerStep)
19    OverallGateSequence := []
20    For C in Components Do
21        ComponentCircuit := TrotterStep(C, t, N, p)
22        OverallGateSequence += ComponentCircuit
23    End For
24    Return OverallGateSequence
25 End Function
26
27 Function TrotterStep(
28     ComponentData C
29     Time t
30     Steps N
31     Parameter p
32     ):
33     ComponentGateSequence := []
34     For k in 0..N-1 Do
35         // Concept d: Time Step Modulation.
36         dt := CalculateEffectiveTimeSteps(t, N, p, k)
37         CurrentStepGates := []
38         For LayerIdx in 0..C.LayersPerStep-1 Do
39             // Concept e: Adaptive Symmetric Second-Order Trotterization
40             CurrentStepGates += Adaptive2ndOrderLayer(dt, p, t, C)
41             CurrentStepGates += SwapNetwork(LayerIdx, C)
42         End For
43         // Concept f: Lightcone Pruning
44         ComponentGateSequence += LightconePruning(CurrentStepGates, C)
45     End For
46     Return ComponentGateSequence
47 End Function

```

where  $X_i, Y_i$  are Pauli operators on site  $i$ . This gives a simplified model with similar behavior to our 3 dimensional dipolar Hamiltonian with a  $1/r^3$  coupling, that we can use for scaling analyses of different classical methods.

### A. MPO analysis

Matrix product operator (MPO) evolution is a powerful method for simulating the operator dynamics and out-of-time-ordered correlators of such one-dimensional quantum systems. MPOs provide an efficient representation of operators that are local or have low operator entanglement. Starting from an MPO representation of the butterfly operators  $A$  and  $B$ , the computation of out-of-time-ordered correlators can be approximated via computing the Heisenberg evolution  $A(t)$  at truncated maximum bond dimension for the time-evolved MPO. Prior work has demonstrated that such MPO evolution can capture the early-onset behavior and wave front dynamics of OTOCs in one-dimensional systems, while appreciable OTOC growth and late-time regimes remain out of reach even for local Hamiltonians unless operator entanglement is bounded [100–103].

To estimate MPO simulation resources for OTOCs for the long-ranged double quantum Hamiltonian, we study a

Trotterized time evolution using the “ $U^{\text{II}}$ ” approximation [104] of the time evolution operator  $e^{i\hat{H}\delta t}$  at fixed time step  $\delta t = 0.1/J$ , and monitor the convergence of OTOCs (within this fixed time-step Trotter evolution) as a function of maximum bond dimension for the butterfly operator. As the required MPO bond dimension for representing the long-ranged double quantum Hamiltonian itself diverges polynomially with system size, we employ the MPO compression algorithm of Ref. [105] and truncate the Hamiltonian bond dimension at 42. At this fixed Trotter step and Hamiltonian bond dimension, the problem is representative of the complexity of simulating OTOCs for long-ranged interactions, but permits isolating the resource requirements (bond dimension) needed to represent the time-evolved butterfly operators.

We choose a pair of butterfly  $\hat{X}$  operators at positions  $n = N/3$  and  $m = 2N/3$  of the  $N$ -site chain. As the overall system is invariant under time translations (by the Trotter step  $\delta t$ ), the OTOC can be computed using either one-sided (1S) evolution of a single butterfly operator to time  $t$ , or two-sided (2S) time evolution of both butterfly operators forward/backward in time

$$\mathcal{O}_{1S}(t) = \frac{1}{\mathcal{N}} \langle X_n(t) X_m(0) X_n(t) X_m(0) \rangle , \quad (\text{S105})$$

$$\mathcal{O}_{2S}(t) = \frac{1}{\mathcal{N}} \langle X_n(\frac{t}{2}) X_m(-\frac{t}{2}) X_n(\frac{t}{2}) X_m(-\frac{t}{2}) \rangle , \quad (\text{S106})$$

where  $\mathcal{N}$  is the Hilbert space dimension. These quantities are equivalent in the exact limit. To assess the convergence of MPO simulations at truncated bond dimension, this therefore motivates a simple error metric

$$\epsilon_{\text{rel}}(t) = 2 \frac{|\mathcal{O}_{1S}(t) - \mathcal{O}_{2S}(t)|}{|\mathcal{O}_{1S}(t) + \mathcal{O}_{2S}(t)|} . \quad (\text{S107})$$

that is expected to converge approximately monotonically at large bond dimension.

Fig. S45 depicts the relative error incurred at three fixed OTOC decay levels (0.98, 0.95, 0.90) as a function of maximum MPO bond dimension for system sizes  $N = 20$  to  $N = 60$  for the long-ranged double-quantum Hamiltonian. The relative error decreases with increasing bond dimension, however the rate of this decrease slows rapidly with growing number of spins. Extrapolating the observed scaling behavior to larger bond dimensions suggests that computing appreciable OTOC growth, accessible in experiments, remains out reach using standard MPO-based methods even for spin chains with non-local interactions. We anticipate that similar scaling behaviors will be observed for three-dimensional spin ensembles with dipolar interactions, as the long-range nature of the interactions is expected to lead to rapid growth of operator entanglement and large required bond dimensions.

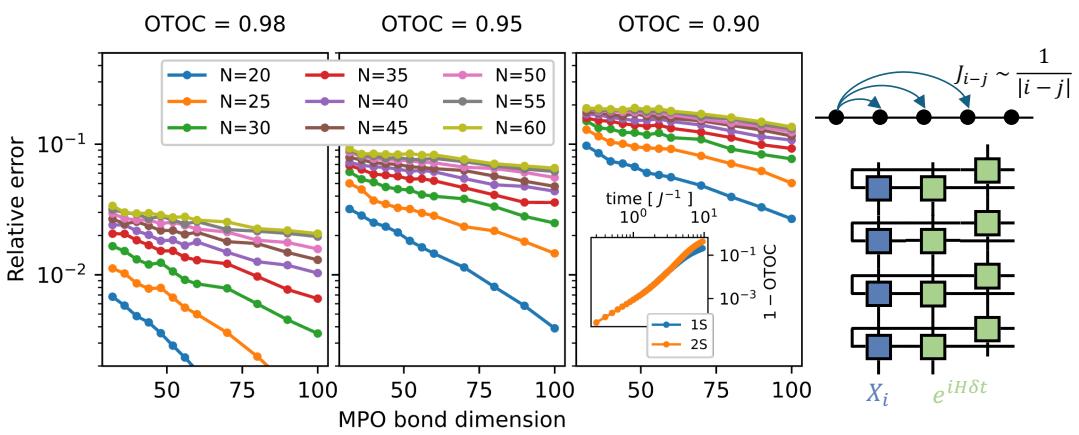


FIG. S45. Scaling of classical resources for simulating out-of-time-ordered correlators (OTOCs) via matrix product operator (MPO) evolution, for a spin-1/2 double-quantum Hamiltonian with long-ranged  $1/R$  interactions on an  $N$ -site chain. Relative simulation error as a function of MPO bond dimension is estimated by comparing the obtained OTOCs for one-sided (1S) MPO time evolution of a single butterfly operator  $[\langle X_n(t) X_m(0) X_n(t) X_m(0) \rangle / \mathcal{N}]$  to the two-sided (2S) MPO evolution of both butterfly operators forward/backward in time  $[\langle X_n(\frac{t}{2}) X_m(-\frac{t}{2}) X_n(\frac{t}{2}) X_m(-\frac{t}{2}) \rangle / \mathcal{N}]$ . The relative error between these approaches is plotted as a function of MPO bond dimension at fixed OTOC decay levels 0.98 (2%), 0.95 (5%), and 0.90 (10%). Inset depicts the early-growth behavior of the commutator OTOC.

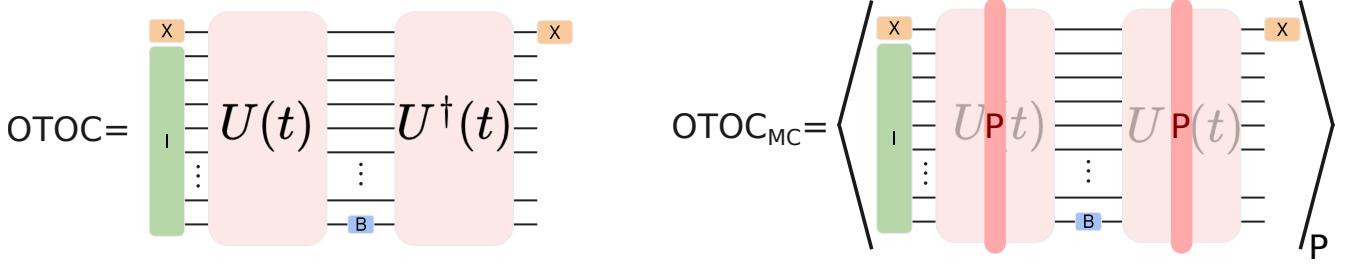


FIG. S46. Explanation of the TNMC algorithm as implemented in this work. (left) a typical OTOC circuit generated by a unitary  $U(t)$  (in our case  $U(t) = e^{iHt}$ ). (right) The TNMC algorithm inserts a single layer of Pauli operators in the middle of a circuit decomposition of  $U(t)$ , and an identical layer of Pauli operators in  $U^\dagger(t)$ . These are averaged over the inserted layer  $P$  (drawing  $I, X, Y, Z$  on each qubit with equal probability). The TNMC algorithm can estimate the resulting ensemble average, with a classical complexity scaling as the cost of simulating an OTOC using only half the circuit.

### B. Monte Carlo analysis

It is perhaps unsurprising that MPO techniques should find OTOC estimation challenging, due to the operator entanglement generated by OTOCs in even random circuits. However, not all classical methods face an entanglement barrier, and Monte Carlo techniques have shown reasonable prowess in estimating OTOCs in random circuits [4], due to the lack of large loop interference in the Pauli path picture. However, in our case interference between Pauli paths is generated not only by the OTOC (which connects Pauli paths at either end), but by Hamiltonian dynamics, which can be expected to cause path interference even within a single unitary.

To test this hypothesis, we apply the tensor-network Monte Carlo (TNMC) algorithm in Pauli space [4] on small systems using the model defined at the start of this section, and compare the output to exact dynamics. The TNMC algorithm works by “cutting” a circuit  $U$  into sub-circuits  $U_j$ , and propagating the butterfly operator between these blocks via Monte Carlo. (For Hamiltonian time-evolution, this requires that we choose a Trotterization or other simulation method to convert  $e^{iHt}$  into a circuit.) For each  $U_j$ , we calculate

$$U_j^\dagger B_{j-1} U_j = \sum_{P \in \mathbb{P}^N} b_P(j) P. \quad (\text{S108})$$

We then sample from the distribution  $b_P^2(j)$ , which by unitarity of  $U_j$  is normalized to 1. Repeating for each  $j$ , this generates an estimate  $B_{\text{MC}}(t)$  for the evolved butterfly operator  $B(t)$ , from which we calculate the OTOC as

$$\text{OTOC}_{\text{MC}} = \text{Trace}[\rho B_{\text{MC}}(t) M B_{\text{MC}}(t) M]. \quad (\text{S109})$$

The cost of classical estimation scales only with the cost of simulating the hardest  $U_j$ , which can be significantly smaller than the cost of simulating  $U$  if sufficient cuts are made.

It turns out that this is equivalent to estimating the ensemble average of a set of circuits where, before each  $U_j$ , we insert randomly chosen Pauli operators on each qubit, and apply the same Pauli operator after  $U_j^\dagger$ . (i.e. Pauli operators are chosen i.i.d for each qubit and each  $j$ , but are identical between the forward and inverse evolution.) Intuitively, the inserted random Pauli operators “dephase” the operator  $U_j^\dagger B_{j-1} U_j$  in the Pauli basis. This deletes constructive interference between the Pauli operators  $P$  at the boundary of  $U_j$ , which has an identical effect to Monte Carlo sampling the  $P$ . We show a cartoon of this process in Fig. S46 for a single cut. This connection gives a useful method to simulate the performance of the TNMC algorithm using circuit simulation alone.

In practice, the all-to-all circuit simulations needed for long-range Hamiltonians are much deeper than they are wide. Hence, we expect the TNMC algorithm to need to cut a 50 – 60 qubit circuit multiple times in order to obtain practically simulable pieces. In Fig. S47, we observe that even a single cut in these circuits yields a significant OTOC error. Here, we simulated the 1D chain described at the start of this section for a range of times, throughout which the OTOC is held to approximately 0.4. We observe that the TNMC relative error quickly increases to 20%, and grows with the system size. (This growth is in some part due to the OTOC decay with system size; the absolute error holds static at around 0.1.) This demonstrates that OTOC simulation methods that have simulated random circuit OTOCs to comparably high fidelity do not immediately extend to the Hamiltonian OTOCs considered here. Testing other methods, and especially custom methods targeting Hamiltonian OTOCs is a clear target for future work.

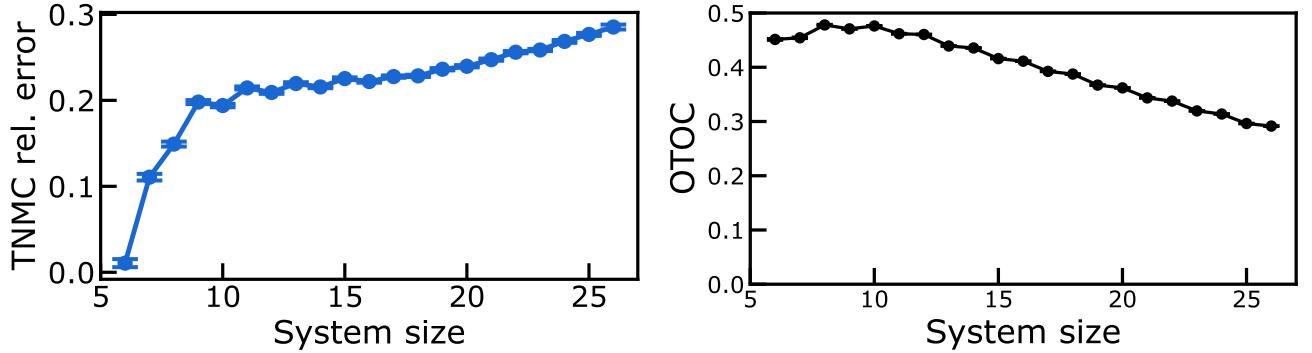


FIG. S47. Results of TNMC estimation on systems of up to 26 qubits with a single cut inserted in the middle of the circuit. (left) Relative error in TNMC estimation, normalized by the OTOC itself. (right) Target OTOC values used for (left).

## VIII. THE SUPERCONDUCTING QUANTUM SIMULATOR

### A. Basic processor benchmarks

Our quantum simulation is performed on a Willow quantum computer similar to that described in [4] (see also [63]). The processors in this lineage are based on a 2D grid of flux tunable superconducting transmon qubits connected by flux tunable couplers between nearest neighbours. In this work, the mean frequency of qubit operation is  $\sim 6.2$  GHz and their anharmonicity is  $\sim 210$  MHz.

Given that the quantum simulation of the two molecules addressed in this work require nine and fifteen qubits only, we can select a subsection of the 105 qubit Willow with performance better than average in the relevant benchmark metrics. The qubits are arranged in a line, and the unique measurement qubit in the quantum simulation is at one end of it. In this configuration, the all-to-all coupling is optimally achieved by a SWAP network as adapted from [37]. In fig. S48 (a) we show the cumulative distribution of the lifetime of the single photon excitation in the qubits used for this work, which is  $T_1 = 114$   $\mu$ s in average. In the same figure we also show the coherence time of these qubits as measured by a Hahn-Echo sequence and is in average  $T_{2E} = 130$   $\mu$ s. The performance of single qubit microwave gates, which we measure by Clifford randomized benchmarking (RB). In fig. S48 (b) we show the total RB error and purity which are  $\epsilon_{\text{RB}} = 0.00020$  and  $\epsilon_{\text{purity}} = 0.00015$  in average, respectively. The same figure shows the performance of controlled-Z (CZ) gates and square-root of iSWAP ( $\sqrt{i\text{SWAP}}$ ) gates as measured by the cross entropy benchmarking (XEB) error, a previously established method [3, 66, 106]. Here gates are calibrated in parallel and therefore include errors arising from cross-talk effects in the parallelization policy for the linear chain experiment, whose moments are divided between two (“even” and “odd”) layers. Here we chose to inform the cycle Pauli error because it is a previously reported metric [66] allowing comparison of performance as the technology rapidly develops. This metric includes error contributions from two single qubit gates and a single two-qubit gate. This reflects well the structure of the OTOC-NMR quantum simulation where, as detailed below, two-qubit gates always appear accompanied by two single qubit gates in each layer of the Trotterization. The XEB cycle errors are respectively  $\epsilon_{\text{CZ}} = 0.0015$  and  $\epsilon_{\sqrt{i\text{SWAP}}} = 0.0014$ , in average. In fig. S48 (b) we also show the readout performance of the single measure qubit in the quantum simulation. The probabilities of preparing and measuring a given computational state [ $p(0|0)$  and  $p(1|1)$ ] on the measurement qubit are measured repeatedly during a typical experimental run of several hours. The corresponding distribution in fig. S48 (b) shows the averaged readout error  $\epsilon_{\text{ro}} = 1 - p(0|0)/2 - p(1|1)/2$  during the quantum simulation and averages over six hours to 0.00985. Although in general these metrics vary with a timescale of several hours and require recalibration during the data taking sessions, the figures here reflect the typical operating performance of our experiments.

### B. SWAP network: efficient all-to-all interaction from only nearest-neighbours coupling

The Hamiltonian targeted by the superconducting quantum simulation,

$$\hat{H} = \sum d_{ij}^{XY} (X_i X_j - Y_i Y_j) + \sum d_{ij}^{ZZ} Z_i Z_j, \quad (\text{S110})$$

is known as the double-quantum Hamiltonian due to the term  $X_i X_j - Y_i Y_j = \sigma_i^+ \sigma_j^+ + \sigma_i^- \sigma_j^-$ . This is a squeezing

term that creates and destroys coherently excitations in pairs. It is therefore not excitation number conserving and typical of driven systems. Our faculty to implement its Trotter decomposition of this Hamiltonian relies on the ability to implement high fidelity SWAP network to simulate an all-to-all spin interaction with hardware with only nearest neighbour connectivity [37].

A circuit implementing a SWAP network Hamiltonian simulation of Toluene is shown in fig. S49. Each two qubit interaction is depicted as a box connecting two adjacent qubit rails. The different colored boxes represent unique Fermionic simulation (fSim) gates [3, 64, 66, 68, 69, 106] angle configuration encoding a spin-spin simulated interaction. The SWAP exchange is depicted as crossing wires. Superconducting qubits correspond to each of the horizontal rails and play the role of, i.e. simulate, different spins in the molecule. Each of the eight Hydrogen protons spins is assigned a color in this representation. The meandering color rails cross every other color rail in a single step of the network. This represents how each spin's quantum state is swapped across the superconducting qubits. So, the qubit-spin mapping changes in every step of the simulation. Importantly, this is optimal in the sense that the number of gates in a Trotter step is equal to the number of pairs [37]. That is, the number of gates is minimal and will not be reduced by an increased hardware connectivity: even if one would count with all-to-all connectivity, one would still require one gate per pair to simulate the all-to-all coupling. The technical challenge is to achieve state swapping efficiently and with sufficiently low error. This poses a metrological challenge over the implementation of tunable-angle fSim gates.

Due to the geometry of para-Carbon<sup>13</sup>-Toluene addressed in this work, the Carbon spin, labelled “C13” in fig. S49, only interacts with the Hydrogen proton labelled “H1”. Or rather, its interaction with every other spin is two

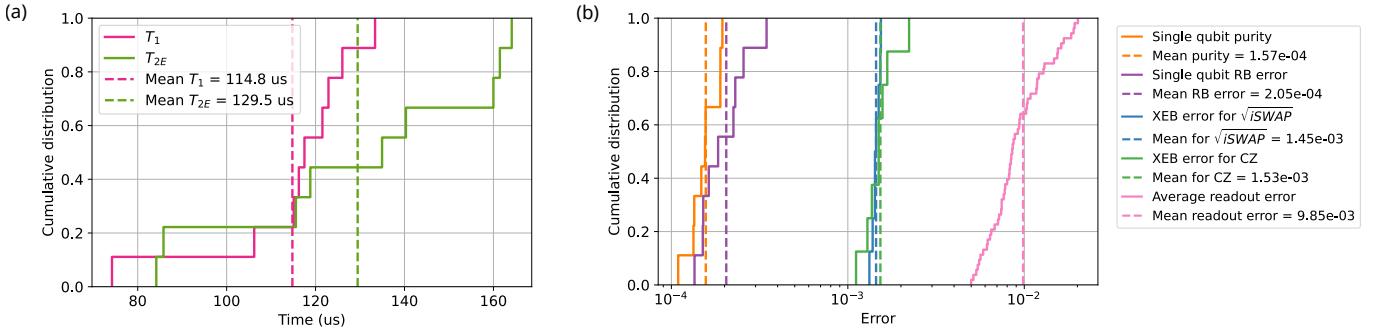


FIG. S48. Base qubit and gate performance metrics. In panel (a) we show a cumulative distribution of single photon excitation lifetime  $T_1$  for the qubits used in the quantum simulation. The qubits are arranged in a linear chain. The coherence time  $T_{2E}$  of each qubit as measured by a Hahn-Echo sequence is also shown. In panel (b) we show a cumulative distribution of single and two qubit gate performance for each qubit and qubit pair involved in the quantum simulation. The single qubit performance is assessed by randomized benchmarking (RB), which is bounded from below by the purity of the state at the end of the sequence. The two-qubit gate performance is assessed here in terms of the CZ and  $\sqrt{i}\text{SWAP}$  XEB cycle Pauli error. Finally, we also present a cumulative distribution of the average  $[1 - p(0|0)/2 - p(1|1)/2]$  readout error on the measurement qubit. The readout error distribution is over the quantum simulation time, about six hours.

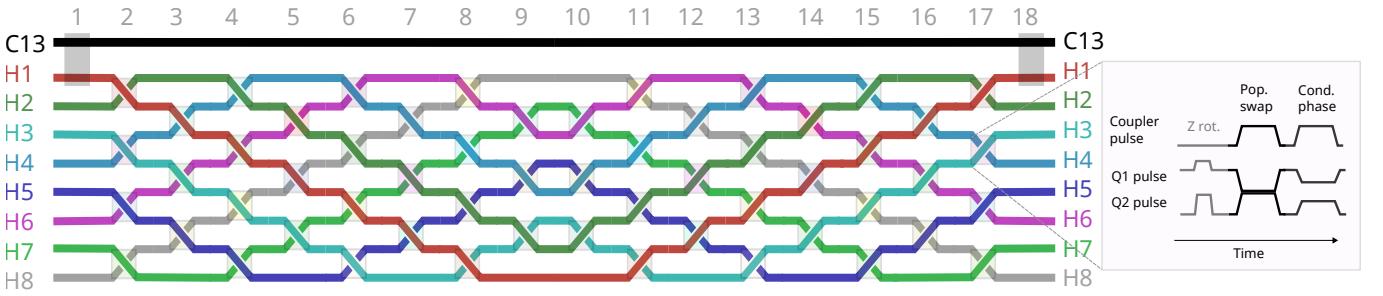


FIG. S49. fSim SWAP network for the quantum simulation of Toluene. The circuit illustrates a single Trotter step of the double-quantum Hamiltonian simulation for Toluene on nine superconducting qubits arranged in a line. Horizontal rails represent superconducting transmon qubits, while meandering colored rails track the state of the nine molecular spins [one Carbon (C13) and eight Hydrogen protons(H1 to H8)] as their quantum states are swapped across the physical qubits. The colored boxes represent the tunable fSim( $\theta, \phi$ ) two-qubit gates including single-qubit gate conjugations. The layers of non-overlapping gates are numbered from 1 to 18. The gates are implemented using a two-pulse composite sequence depicted in the right-most blow up. Note the exclusion of the Carbon spin from the main SWAP network, which reduces the gate count by 17%.

orders of magnitude smaller and therefore negligible in the timescale of the experiment. Spin C13 can be therefore extracted from the SWAP network to reduce the gate count by 17%, and therefore the error introduced by imperfect gates without otherwise affecting the learning experiment. We note also that the gates removed are full SWAP and therefore having the lowest performance. This technique is, in principle, general but does modify the ordinary Trotter formula. In particular, after one Trotter step the qubit-spin mapping is inverted leaving the qubits simulating C13 and H1 far apart and unable to interact (see mid layers 9 and 10 in fig. S49, red and black rails). A second Trotter step needs to be applied to bring them next to each other before a new Trotterized interaction can take place between them (see layer 18 in fig. S49), which corresponds to working in the interaction frame of the C-H coupling.

### C. Trotterization of a double-quantum Hamiltonian embedded in a SWAP network

The Trotter approximation for the time evolution generated by the double-quantum Hamiltonian eq. (S110) is constructed by sequentially applying the unitaries corresponding to the individual terms of the Hamiltonian. For a time increment  $\delta t$  these read

$$\text{SWAP} \times \exp[-id_{ij}^{XY}\delta t(X_iX_j - Y_iY_j)] = \begin{pmatrix} \cos 2d_{ij}^{XY}\delta t & 0 & 0 & -i \sin 2d_{ij}^{XY}\delta t \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ -i \sin 2d_{ij}^{XY}\delta t & 0 & 0 & \cos 2d_{ij}^{XY}\delta t \end{pmatrix},$$

and

$$\text{SWAP} \times \exp[-id_{ij}^{ZZ}\delta t Z_i Z_j] = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & e^{i2d_{ij}^{ZZ}\delta t} & 0 \\ 0 & e^{i2d_{ij}^{ZZ}\delta t} & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

As we will now discuss, these unitaries are equivalent, up to single qubit rotations, to the so-called (fSim) gates. These are the gates native to the Willow processor, and present a natural fit of this hardware to this application.

In a Willow processor, the effective two qubit gate dynamics can be modelled by the Hamiltonian

$$H(g, \Delta, \eta) = \begin{pmatrix} 0 & g & 0 & 0 & 0 \\ g & \Delta & 0 & 0 & 0 \\ 0 & 0 & \Delta & \sqrt{2}g & \sqrt{2}g \\ 0 & 0 & \sqrt{2}g & 2\Delta + \eta & 0 \\ 0 & 0 & \sqrt{2}g & 0 & \eta \end{pmatrix}. \quad (\text{S111})$$

It describes the evolution of two neighbouring transmon qubits, detuned by an amount  $\Delta$  and coupled with a single excitation resonant swap rate  $g$ , in the truncated oscillator basis  $\{|01\rangle, |10\rangle, |11\rangle, |20\rangle, |02\rangle\}$ . The ability to control both  $\Delta$  and  $g$  as a function of time allows the digital control of the interaction. The parameter  $\eta$  is the transmon anharmonicity, approximated in this expression to be the same for both oscillator and constant.

The integration of the Schrödinger equation yields a fSim unitary with five effective parameters of the form

$$\text{fSim}(\theta, \phi, \Delta_+, \Delta_-^{\text{diag}}, \Delta_-^{\text{off diag}}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{i(\Delta_+ + \Delta_-^{\text{diag}})} \cos \theta & -ie^{i(\Delta_+ - \Delta_-^{\text{off diag}})} \sin \theta & 0 \\ 0 & -ie^{i(\Delta_+ + \Delta_-^{\text{off diag}})} \sin \theta & e^{i(\Delta_+ - \Delta_-^{\text{diag}})} \cos \theta & 0 \\ 0 & 0 & 0 & e^{i(2\Delta_+ + \phi)} \end{pmatrix}. \quad (\text{S112})$$

By applying  $Z$ -phase gates on each qubit before and after the interaction it is possible to gauge out the single qubit phases  $\Delta_+$ ,  $\Delta_-^{\text{diag}}$ ,  $\Delta_-^{\text{off diag}}$  and bring the interaction unitary to the canonical fSim form

$$\text{fSim}(\theta, \phi) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & -i \sin \theta & 0 \\ 0 & -i \sin \theta & \cos \theta & 0 \\ 0 & 0 & 0 & e^{i\phi} \end{pmatrix}. \quad (\text{S113})$$

That is

$$\text{fSim}(\theta, \phi) = (Z(\varphi_1) \otimes Z(\varphi_2)) \text{fSim}(\theta, \phi, \Delta_+, \Delta_-^{\text{diag}}, \Delta_-^{\text{off diag}})(Z(\varphi_3) \otimes Z(\varphi_4)),$$

with  $\varphi_1 = \Delta_-^{\text{off diag}} + \Delta_-^{\text{diag}} + 2\Delta_+$ ,  $\varphi_2 = \Delta_-^{\text{off diag}} + \Delta_-^{\text{diag}} - 2\Delta_+$ ,  $\varphi_3 = \Delta_-^{\text{diag}} - \Delta_-^{\text{off diag}}$ , and  $\varphi_4 = -\Delta_-^{\text{diag}} - \Delta_-^{\text{off diag}}$  and  $Z(\varphi) = \text{diag}(1, e^{-i\varphi})$ .

We note that it is possible to obtain with a single pulse a fSim gate with arbitrary two-qubit angles [69, 107]. While previous work was able to operate in this regime, it was exploited that given sets of two-qubit angles are more natural than other. For piecewise constant dependence of the parameters, one can approximate the swap angle during the gate changes as  $\theta \approx t\sqrt{g^2 + \Delta^2}/4$  and the conditional phase changes as  $\phi \approx tg^2/\Delta_{12}$  where  $\Delta_{12}$  is the detuning between  $|11\rangle$  and  $|02\rangle$ . The different scaling with  $t$  and  $g$  allows to tune both angles independently. When the experimental exploration at hand allows to only tune one of the knobs, say  $g$ , keeping the other one ( $t$ ) fixed, this is the simple and most efficient way to proceed [3, 65–69]. On the other hand, imposing the single pulse fSim approach to produce arbitrary control of both  $\theta$  and  $\phi$  almost certainly forces too long gate times that degrades performance by decoherence, or too high coupling constant  $g$  which enhances leakage probability outside of the computational manifold. In our case, both two-qubit angles are simultaneously specified by the NMR Hamiltonian and are required with high precision and high fidelity for the simulation. We therefore chose not to proceed with the single pulse fSim approach and instead adapted a two-pulse approach from earlier work [64].

#### D. Gate decomposition for composite fSim calibration

To define the mapping between the Trotterization of the double-quantum Hamiltonian and the native gates on Willow it is convenient to introduce the composite gate unitary

$$U_{\text{comp}}(\varphi, \phi_{\text{cond}}, \theta_{\text{fSim}}, \phi_{\text{fSim}}) = Z_i(\varphi)Z_j(\varphi)\text{CPHASE}_{ij}(\phi_{\text{cond}})\text{fSim}_{ij}(\theta_{\text{fSim}}, \phi_{\text{fSim}})$$

Where CPHASE is a form of fSim gate defined by convention as  $\text{CPHASE}_{ij}(\phi_{\text{cond}}) = \text{fSim}_{ij}(\theta_{\text{fSim}} = 0, \phi_{\text{fSim}} = -\phi_{\text{cond}})$ . The motivation to implement a composite gate from a CPHASE gate and a fSim gate is that it allows one to use independent calibrations to grossly target a population swapping gate with the desired  $\theta_{\text{fSim}}$ , unencumbered by the associated spurious  $\phi_{\text{fSim}}$ . Thereafter, independent control of  $\phi_{\text{cond}}$  allows to obtain an unconstrained effective conditional phase for the composite unitary, therefore providing independent control over the effective two two-qubit angles of the composite gate. Experimental imperfections make all the effective angles of the composite gate differ by more than the target tolerance of 20 mrad from what one would predict from an ideal composition of the constituent gates. These errors are addressed on a second, fine tuned, calibration of the composite gate as a whole.

Experimental imperfection aside from now, this notation makes it is easy to see that

$$\text{SWAP} \times \exp[-id_{ij}^{XY}\delta t(X_iX_j - Y_iY_j)] = X_iU_{\text{comp}}X_j,$$

where  $U_{\text{comp}}$  is specified by  $\phi_{\text{cond}} = \phi_{\text{fSim}}$ ,  $\phi_{\text{cond}} = \pi - \phi_{\text{fSim}}$ ,  $\theta_{\text{fSim}} = \pi/2 + 2d_{ij}^{XY}\delta t$ ,  $\varphi = -(\phi_{\text{cond}} + \phi_{\text{fSim}})/2$ . While

$$\text{SWAP} \times \exp[-id_{ij}^{ZZ}\delta tZ_iZ_j] = U_{\text{comp}},$$

where  $U_{\text{comp}}$  is specified by  $\phi_{\text{cond}} = \phi_{\text{fSim}}$ ,  $\phi_{\text{cond}} = \pi - \phi_{\text{fSim}} + 4d_{ij}^{ZZ}\delta t$ ,  $\theta_{\text{fSim}} = \pi/2$ ,  $\varphi = -(\phi_{\text{cond}} + \phi_{\text{fSim}})/2$ .

Alternatively, considering together the  $(X_iX_j - Y_iY_j)$  term and the  $Z_iZ_j$  term, one can write

$$U_{\text{comp}}(\varphi, \phi_{\text{cond}}, \theta_{\text{fSim}}, \phi_{\text{fSim}}) = \text{fSim}(\theta_{\text{comp}}, \phi_{\text{comp}}, \Delta_+, \Delta_-^{\text{diag}}, \Delta_-^{\text{off diag}}),$$

with the mapping

$$\theta_{\text{comp}} = \pi/2 + 2d_{ij}^{XY}\delta t, \quad \phi_{\text{comp}} = -\pi - 4d_{ij}^{ZZ}\delta t,$$

$$\Delta_+ = -\phi/2, \quad \text{and} \quad \Delta_-^{\text{diag}} = \Delta_-^{\text{off diag}} = 0.$$

### E. Composite fSim gate calibration

For a small Trotter step  $\delta t$  (i.e.,  $d_{ij}^{XY}\delta t, d_{ij}^{ZZ}\delta t \ll \pi$ ) every two-qubit gate in the Trotterization at hand is a perturbation of the SWAP gate. This requires only a precise calibration of both  $\theta_{\text{comp}}$  and  $\phi_{\text{comp}}$  around a small range around the SWAP values. However, such a decomposition would require far too many gates degrading the circuit fidelity below what can be recovered by current error mitigation techniques. A back of the envelope calculation can be used to estimate the gate cost of such an approach. With our implementation in mind, taking  $d_{ij}^{XY}\delta t = 1/10$ , the total simulated circuit extending to  $50 \times \delta t$  (this corresponds to  $d_{ij}^{XY} \approx 2$  kHz and total simulated time evolution of 2.5 ms forward, and 2.5 ms backwards) and the nine qubit experiment requiring 36 gates to implement a SWAP network, the gate count estimate yields  $2 \times 50 \times 36 = 3600$ . With current two-qubit gate fidelity of at best 0.998, the smallest circuit fidelity is estimated to be  $0.998^{3600} \approx 0.0007$ .

In order to perform our quantum simulation with sufficient fidelity we perform longer Trotter steps to bound the two-qubit gate count to only a few hundreds. To match the NMR data we chose  $\delta t = 0.375$  ms for our simulation. In consequence, the gates calibrated do depart appreciably from the SWAP angles. Due to the rich scale of energies involved in Toluene the range of two-qubit angles that need to be calibrated span the entire domain from 0 to  $\pi$ . The conditional phase angles required for our SWAP network are all  $\pi$  while the cognitional phase for the C13-H1 interaction are either 1.02 mrad for the gates at the beginning and end of the Trotterized evolution and doubles to 2.04 rad for gates in the middle of the circuit where the end gate of one step in compiled with the first gate of the next (these are the gates in layers 1 and 18 in fig. S49). The swap angles involved in the full learning experiment are instead plotted in fig. S50 (a). The wide range of angles adds to the complexity of the calibration required for this experiments, as different calibration techniques are required for different angle ranges to achieve the required precision 20 mrad (see numerical section XI).

The hardware gate pulses used to engineer these interactions are implemented with a series of base band trapezoidal pulses (see fig. S49) with a fidelity-optimized pulse duration of 12 ns (8 ns hold time) for the physical phase-matching  $Z$  gates, 47 ns (45 ns hold time) for the population swap pulse, and 33 ns for the conditional phase pulse (31 ns hold time). The coupler pulses have a total duration of 26 ns (18 ns hold time) for the population swap pulse and 22 ns (16 ns hold time). In practice, each gate in the composite gate is calibrated independently to provide an initial guess for the implementation. Then, the composite gate is calibrated as a whole to obtain the effective angles, which normally differ from the predicted ones from the independent calibration of the pulses by a few tens of miliradians. An iterative produce is then put in place to obtained the desired precision.

The iterative calibration procedure develops as follows. Having calibrated the CPHASE pulse as in previous work [66] we move to calibrate the population swap pulse. The gate duration is optimized to implement a  $\sqrt{i}$ SWAP gate since with a swap angle of  $\theta = \pi/4$  lies in the centre of the distribution of typical gates we will calibrate for. For fix pulse duration, we use fast two-qubit tomography to fine tune the qubit frequencies to maximize the population transfer totalling a complete  $i$ SWAP. At the calibrated qubit detuning we use fast two-qubit tomography again to determine both the swap angle and the conditional phase that the pulse imprints over the qubit pair as a function of the max coupling amplitude  $g_{\max}$  in the trapezoidal coupler pulse. In agreement with a simple model, we fit the swap angle  $\theta_{\text{fsim}}$  by a linear dependence in  $g_{\max}$  and the conditional phase  $\phi_{\text{fsim}}$  by a quadratic dependent in  $g_{\max}$ . The variance of the fit errors is < 50 mrad for both angles consistently. This fit provides a first good coarse guess to build the composite pulse.

A composite pulse is initially built by combining a fSim pulse with the swap angle dictated by the targetted Hamiltonian simulation  $\theta_{\text{fsim}} = \pi/2 + 2d_{ij}^{XY}\delta t$  folded in the  $[0, \pi/2]$  interval, and a CPHASE gate. The fSim pulse interaction strength  $g_{\max}$  is dictated by the linear fit obtained for  $\theta_{\text{fsim}}$ . The CPHASE pulse is set by an independently calibrated angle of  $\phi_{\text{cond}} = -\pi - 4d_{ij}^{ZZ}\delta t - \phi_{\text{fsim}}$ , where  $\phi_{\text{fsim}}$  is the spurious phase of the fSim gate (previously obtained from the quadratic fit against the interaction strength  $g_{\max}$ ). We note that the CPHASE, introduce to compensate the spurious conditional phase of the population swapping fSim pulse, has itself a small spurious swap angle of  $\sim 30$  mrad that will add to the final effective swap angle. These two swap rotations are in general not collinear and are taken care of by the iterative calibration protocol provided the following technicality. Due to the spurious swap angle in the CPHASE pulse ( $\sim 30$  mrad), the order in which the two interaction pulses in the composite gate is applied matters, if one aims for sub 20 mrad precision. The arbitrary angle between this consecutive swap rotations makes them non-commuting rotations. Furthermore, the single qubit physical  $Z$  gates involved (see fig. S49) are physical flux pulses (as opposed to virtual phase update on the microwave drives) as these  $Z$  gates do not commute trough any of the swapping interaction. Even if it approximately commutes with the CPHASE, the finite spurious angle creates a non-commutativity error larger than our 20 mrad budget. These  $Z$  gates have a phase that depends on the context. They depend on the single qubit gates that have been applied just before, since consecutive single qubit gates get merged together for efficiency in the compilation. Due to phase tracking, they even depend on idling time durations. Having a physical  $Z$  gate between the two pulses in the composite gate, then, affects in an undesirable way the calibration and the final effective swap angle. It changes the relative angle of the two consecutive swap rotations

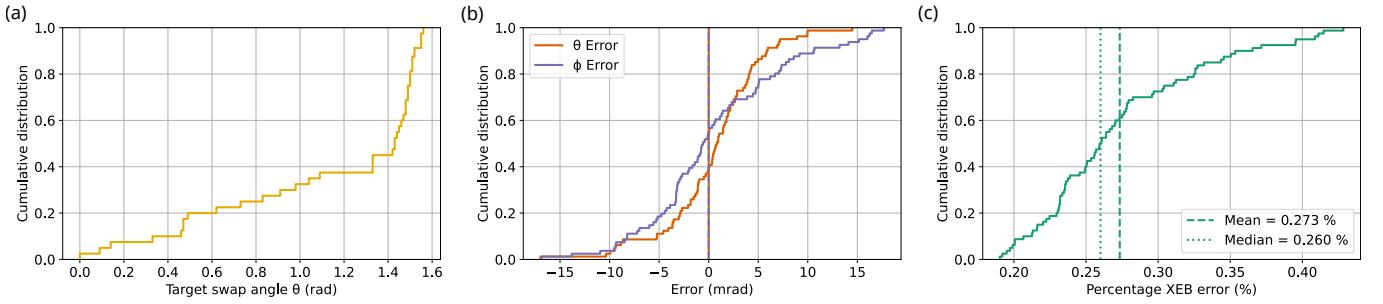


FIG. S50. (a) Histogram showing the distribution of the 80 unique target swap angles  $\theta$  required for the full learning quantum simulation of Toluene. They span the full range available  $[0, \pi]$ . (b) Distribution of two-qubit angle errors achieved by the iterative calibration protocol. All final errors are below the 20 mrad by construction. (c) Distribution of the Cross-Entropy Benchmarking (XEB) error for the 80 calibrated, phase-matched fSim gates as they are deployed in the quantum simulation circuit. This metric includes the error contributions from both the two-qubit pulse and the required single-qubit Z gates used for phase matching.

of the gate as a function of context, as it changes therefore the relative angle with which these two rotations interfere. This is easily avoided by ensuring the physical Z gates are not played between the interaction pulses but before them (or, alternatively, after).

With a first composite pulse composed, high precision Floquet calibrations, tailored to the target angle range, are used to measure the swap angle and the conditional phase angle of the composite pulse. For small swap angles ( $< 30$  mrad) we adapted to the two pulse fSim implementation the technique discussed in [66], while for larger angles we adapted the technique described in [65], and adjusted the number of gates used in the periodic calibration as well as the fitting procedure to the value of the expected angle. This first high precision measurements of both the swap and conditional phase angles of the composite gate are compared to the target angles. If any one of the two two-qubit angles differs from target by more than 20 mrad a correction proportional to the errors is applied to the control for compensation and a new high precision measurement of both angles is performed. That is, the error in the conditional phase feeds back into the control of the CPHASE pulse (i.e., the location of the detuning-coupling parameter along the minimum-leakage arc, see [64, 66]), while the error in the swap angle feeds back to the control of the fSim gate (i.e.,  $g_{\max}$ ). Now with two high precision measurements for each, for the swap angle and the conditional phase angle, we perform a local linear extrapolation targeting the control parameters for zero error. The procedure of local linear extrapolation repeats taking into account all previous high precision measurement until the target precisions is attained in both angles. This procedure is done in parallel for all gates compatible with simultaneous calibration (at most 4, see fig. S49), and always succeeds. It typically concludes after three rounds of simultaneous interpolation lasting  $\sim 10$  minutes total. The distribution of two-qubit angle errors is shown in fig. S50 (b) for our full Toluene learning experiment. For a given pulse shape and pair, periodic calibrations are able to reproducibly measure both angles with a shot-to-shot spread (precision) of  $< 5$  mrad.

For Toluene, 33 different fSim gates needed to be calibrated independently down to a precision of  $< 20$  mrad in order to simulate a single OTOC trajectory. Furthermore, for the learning experiment, nine different trajectories are needed. However, not all  $9 \times 33$  fSim gates are meaningfully different: rounding the interaction angles to 10 mrad we are left with only 80 unique gates to calibrate and still maintain an excellent approximation of the evolution. The unique-angle two-qubit gates appear in the quantum simulation in different combination, totalling 107 distinct moments. Calibrating this number of distinct moments is impractical, but we found that in-context calibration does not provide an advantage. While simultaneous calibration of different gates does degrade their XEB fidelity, this degradation is independent of the specific context. Importantly, due to the symmetries of the molecules, many of the distinct calibration moments differ by a single or only a few gates. We then shuffle the 80 distinct fSim gates needed for the full Hamiltonian learning experiment into calibration moments where each gate appears only once. This reduces the calibration overhead to only 28 moments with typical wall-clock time of 5hs of calibration.

Once the pulse parameter mapping to the target two-qubit angles is obtained, the gate's single qubit angles are measured by XEB fitting. The single qubit angles  $\Delta_+$ ,  $\Delta_-^{\text{diag}}$ , and  $\Delta_-^{\text{off diag}}$  from the XEB fit are then used to phase-match the gates and build the desired composite canonical fSim( $\theta, \phi$ ) gate by single qubit Z-phase conjugation, as explained above. This produces the final composite fSim gates used in our simulation. The XEB fidelity of these gates is measured once again to take into account the phase-matching correcting as a final metric assessing their performance. In fig. S50 (c) we show the XEB error of the phase-matched gates. The median cycle Pauli infidelity for the composite-phase matched gate is 0.0026.

## IX. ACCURATE NOISE MODELING OF SUPERCONDUCTING EXPERIMENTS

### A. Trajectory simulation of quantum circuits

To validate our experimental results and accurately account for the impact of device noise, we performed comprehensive numerical simulations of the NMR OTOC protocol. These simulations were designed to faithfully reproduce the experimental conditions, incorporating a multi-layered noise model parameterized by the device benchmarks detailed in the main text (see Basic processor benchmarks). Our numerical simulations were performed using an in-house quantum trajectory simulator `kraus-sim`. Quantum channels are represented by Kraus-operators, defined by the respective action on the density matrix  $\rho$  as:

$$\mathcal{E}(\rho) = \sum_{j \geq 1} K_j \rho K_j^\dagger, \quad \text{where} \quad \sum_{j \geq 1} K_j^\dagger K_j = I$$

`kraus-sim` simulates the evolution of a pure state vector  $|\psi\rangle$  instead of the full density matrix  $\rho$ , offering a significant computational advantage at the cost of induced sampling error. Each trajectory represents one independent “shot” of the experiment. The simulation begins with a pure initial state  $|\psi\rangle$ . For each channel  $\mathcal{E}$  in the circuit, one of its Kraus operators  $K_j$  is randomly sampled and applied to the state. The probability  $p(j)$  of selecting a specific operator  $K_j$  is given by the Born rule:

$$p(j) = \langle \psi | K_j^\dagger K_j | \psi \rangle$$

After sampling, the state is updated and re-normalized to

$$|\psi'\rangle = \frac{K_j |\psi\rangle}{\sqrt{p(j)}}$$

This process, which is equivalent to the state back-action from a POVM measurement with elements  $M_j = K_j^\dagger K_j$ , is repeated for every channel in the circuit. The final experimental statistics are recovered by averaging the outcomes of many independent trajectories. A more detailed account of the quantum trajectories method can be found in Ref. [108]

### B. Inhomogeneous noise sensitivity

To understand the circuit’s susceptibility to errors, we performed a first-order sensitivity analysis by simulating the deterministic injection of a single error at all points in the circuit. This study allows us to map the spatio-temporal impact of a local error on the final OTOC observable, revealing the flow of information and the most sensitive regions of the quantum simulation. In particular, the methodology consists of running a series of ideal, noiseless simulations of the OTOC circuits. For each independent simulation, a single X gate was deterministically inserted at a unique location to emulate a bit flip error, defined by a specific qubit and a specific circuit moment (i.e., after a particular gate layer). We then computed the absolute deviation of the final measured OTOC value from the ideal, error-free case. The results of this analysis are visualized in the heatmap in Fig. S51.

The heatmap in Fig. S51 plots this OTOC deviation as a function of the error’s location. The vertical axis represents the physical qubit index while the horizontal axis represents the circuit moment index, or time. The circuit consists of a forward evolution  $U(t)$  followed by the butterfly operator and the backward evolution  $U^\dagger(t)$ . The bright yellow and white regions indicate locations where a single bit-flip has a large impact on the final OTOC, while the dark purple regions indicate low sensitivity. The most striking feature of this analysis is the clear, structured pattern of error sensitivity inherited from the structured nature of the Hamiltonian. The regions of highest impact are not randomly distributed but instead form distinct diagonal and V-shaped paths through the circuit’s space-time volume. This pattern is a direct visualization of the SWAP network’s operation. As highlighted, the logical state of the Carbon spin begins on the physical qubit indexed as 0. The bright path originating from C13 qubit at the start of the circuit traces the trajectory of this logical spin interaction with the Hydrogen qubits as it is swapped from one physical qubit to the next during the forward evolution  $U(t)$ , and then back during the backward evolution  $U^\dagger(t)$ .

This result demonstrates that the observed circuit’s infidelity is dominated by errors that occur within the effective “light cone” of the operator being measured. In the context of the OTOC, the measurement is sensitive to the evolution of the initially localized operator (on the Carbon spin). An error occurring on or nearby a qubit that is

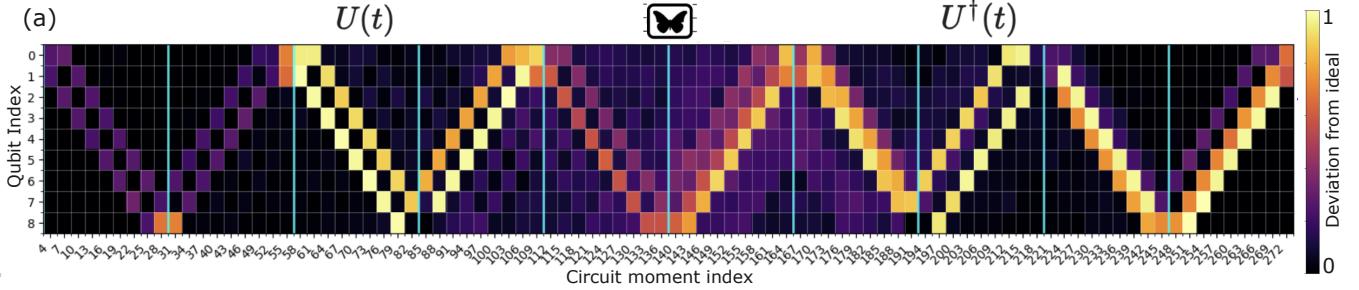


FIG. S51. Spatio-temporal sensitivity map of the 9-qubit OTOC circuit to a single bit-flip error. A single Pauli-X gate was deterministically inserted at each point in the circuit, defined by the qubit index (vertical axis) and the circuit moment index (horizontal axis). The color at each point indicates the magnitude of the deviation of the final OTOC value from the ideal, noiseless result. The circuit is composed of a forward time evolution  $U(t)$  followed by a backward time evolution  $U^\dagger(t)$ . The bright regions of high sensitivity clearly trace the path of the logical state of the Carbon spin (initially on Qubit 0) as it is swapped through the device via the SWAP network, illustrating the operator's "light cone."

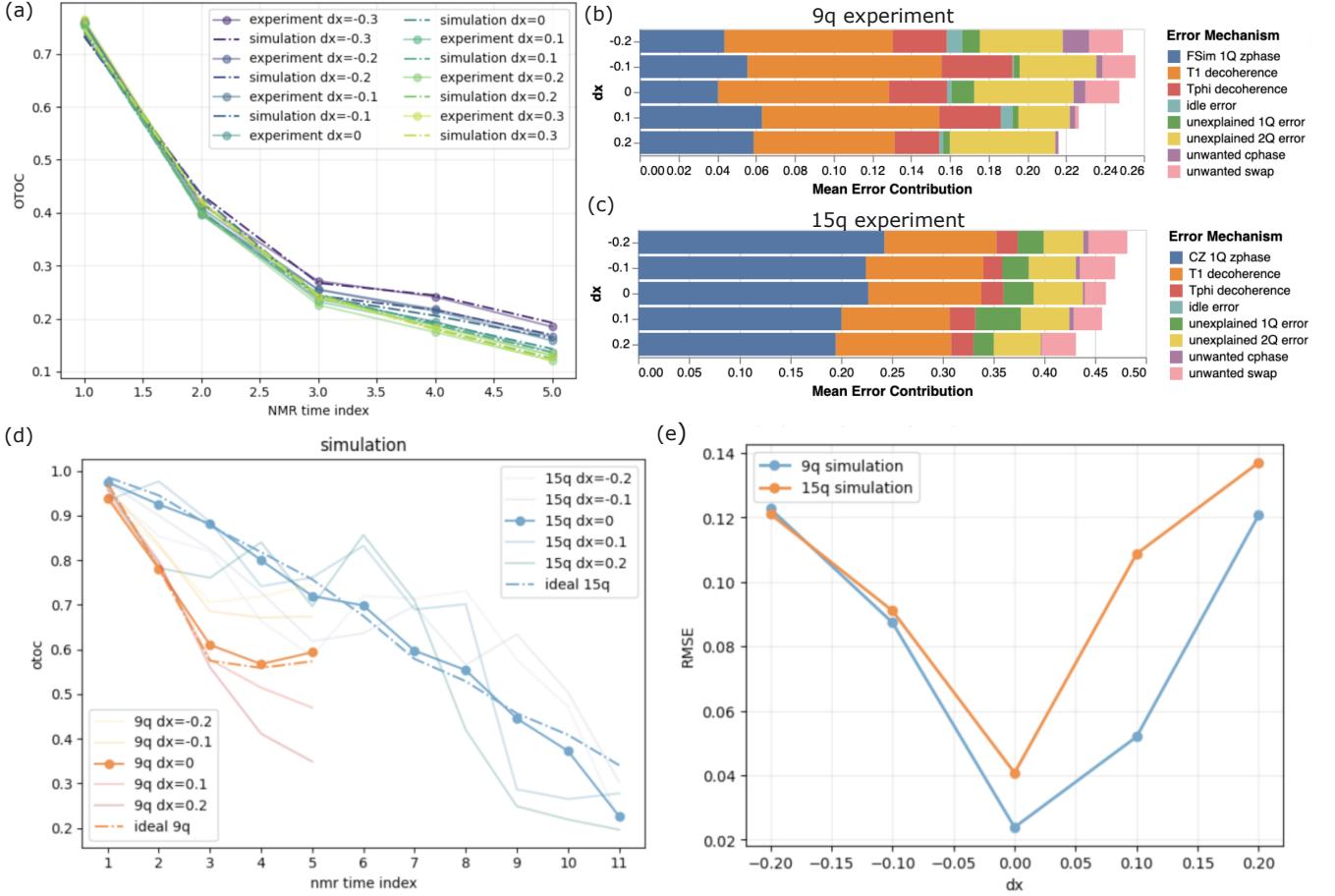
currently holding the logical state of the Carbon spin directly corrupts this information, leading to a large deviation in the final measurement. Conversely, an error on a distant qubit, outside this light cone, has a much smaller, higher-order effect on the final OTOC value, as its influence must propagate through many two-qubit gates to reach the operator of interest. This has important ramifications on calibration – fSim gates within the lightcone have more stringent precision requirements.

### C. Agreement of simulations to experiment

The specific Kraus operators for our error channels were parameterized using the calibrated benchmarks from the Willow processor. Incoherent errors were applied following the ideal gate execution, scaled by the physical gate time set by the waveform. In particular, qubit decay and dephasing were modeled as a composite amplitude and phase damping channels set by the median characterized values across all qubits in the grid,  $T1=114.8\ \mu s$  and  $T2=130\ \mu s$  respectively. Additionally, coherent errors were modeled as additional small angle fSim( $\theta, \phi$ ) following each ideal gate to capture imprecision in calibration of the swap and cphase angles for each composite gate. Furthermore, to model the imprecision in estimating the required physical Z gates to compensate for the composite gate, we additionally add an unwanted  $Z^{\alpha_{1,2}}$  gate on each qubit. Importantly, we assume these parameters are set by the calibration and thus are reproducible per unique calibrated gate, i.e. the same coherent error is applied on each successive gates. To this end, for each unique fSim gate we sampled  $\theta, \phi, \alpha_i$  parameters from a Gaussian distribution with zero mean and standard deviation set by the calibration characterization above. For the 15 qubit experiments, the gateset was not arbitrary angled fSim but rather utilized CZ gates and arbitrary single qubit microwave gates. However, the underlying coherent error channels were modeled equivalently, i.e. by a small angled fSim gate and physical Z gates following each CZ. Unique CZ gates were in turn specified simply by the qubit pair on which it acts. All taken together, coherent errors, amplitude and phase damping channels could not fully explain the single qubit fidelities as measured by randomized benchmarking error or the two qubit XEB cycle Pauli infidelities, originating from features not modeled such as low frequency noise, or leakage out of the computational subspace. One and two qubit depolarizing channels were appended to the noise channels following each gate to match experimental error rates. Readout error was modeled as a classical confusion matrix applied to the final measurements.

The resulting noise model is validated in Fig. S52(a), which shows the simulated OTOC decay plotted directly against the experimental data for the 9-qubit experiment. The solid lines represent the experimental data, while the dashed/dotted lines represent the output of our full noise-model simulation. The different colors correspond to experiments where a known, additional coherent error (parameterized by  $dx$ ) was applied, testing the model's predictive power. The simulation accurately recapitulates the experimental results across all time steps and error parameters, confirming the validity of our model.

To better understand the dominant sources of infidelity, we performed a detailed error budget analysis, shown in Fig. S52(b) and (c). These plots show the mean error contribution to the measured OTOC signal, averaged per gate, for the two different experimental sets of circuits. For the 9-qubit experiment (Fig. S52(b)), which uses the composite fSim gate set, the error budget is dominated by two-qubit errors. We note that a large contribution is the "unexplained 2Q error", accounting for approximately 0.25 of the mean error. This is followed by coherent errors: "unwanted swap" (approx. 0.10) and "unwanted cphase" (approx. 0.08). We note that precision calibration discussed



**FIG. S52. Simulation validation and error analysis.** (a) Comparison of experimentally measured OTOC (solid lines) versus full noise-model simulation (dashed/dotted lines) for the 9-qubit experiment, showing strong agreement. The parameter  $dx$  represents an additional, intentionally-added coherent error to test the model. (b, c) Mean error contribution per gate for the (b) 9q fSim-based experiment and (c) 15q CZ-based experiment, breaking down the sources of infidelity. (d, e) In-silico learning experiment for the (d) 9-qubit and (e) 15-qubit simulations. The calibrated noise model ( $dx = 0$ , solid lines) is compared against mismatched models ( $dx \neq 0$ ) and the ideal, noiseless evolution (dashed lines). (f) Root Mean Square Error (RMSE) between the simulated and ideal curves from (d) and (e), showing a distinct minimum at  $dx = 0$ .

above helped mitigate the impact of these error channels. Decoherence during gates (“T1 decoherence,” approx. 0.05) and “unexplained 1Q error” (approx. 0.04) are secondary, but non-negligible, contributions. For the 15-qubit experiment (Fig. S52(c)), which uses a CZ-based gate set, the error profile is qualitatively different. The coherent errors related to the fSim gate (“unwanted swap” and “unwanted cphase”) are negligible, but the uncalibrated Z-phase disproportionately impacts the OTOC fidelity. The error budget is also more evenly distributed. The “unexplained 2Q error” is still a dominant contributor (approx. 0.022), but it is followed closely by “T1 decoherence” (approx. 0.018) and “unexplained 1Q error” (approx. 0.015). This indicates that for the CZ-based experiment, single-qubit errors and decoherence play a much larger relative role in the circuit fidelity than in the fSim-based experiment.

Fig. S52(d) and (e) demonstrate an in-silico learning experiment based on our validated simulation framework. In these plots, the dashed lines represent the ideal, noiseless evolution of the OTOC for the 9q and 15q systems, respectively. The solid blue and orange curves ( $dx=0$ ) show the simulation using our calibrated noise model. The significant gap between the  $dx=0$  curve and the ideal curve highlights the substantial impact of the characterized device noise on the quantum simulation. The lighter-colored curves show the simulated dynamics when the coherent error model is intentionally mismatched (e.g.,  $dx = -0.2, 0.1$ , etc.). Qualitatively, the calibrated  $dx=0$  simulation tracks the shape of the ideal evolution more closely than the mismatched models. This is quantified in Fig. S52(f), which plots the Root Mean Square Error (RMSE) between each simulated  $dx$  curve and the ideal, noiseless simulation. For both the 9q and 15q cases, the RMSE is clearly minimized at  $dx = 0.00$ . This result is twofold: first, it demonstrates that our simulation of the experiment successfully recapitulates the results of the on-device learning experiment, and

second, it confirms that the calibrated noise model is an accurate representation of the hardware

- 
- [1] Igor L. Aleiner, Lara Faoro, and Lev B. Ioffe, “Microscopic model of quantum butterfly effect: Out-of-time-order correlators and traveling combustion waves,” *Ann. Phys.* **375**, 378–406 (2016).
- [2] Daniel A. Roberts, Douglas Stanford, and Leonard Susskind, “Localized shocks,” *J. High Energy Phys.* **2015**, 51 (2015).
- [3] Xiao Mi, Pedram Roushan, Chris Quintana, Salvatore Mandra, Jeffrey Marshall, Charles Neill, Frank Arute, Kunal Arya, Juan Atalaya, Ryan Babbush, Joseph C. Bardin, Rami Barends, Andreas Bengtsson, Sergio Boixo, Alexandre Bouassa, Michael Broughton, Bob B. Buckley, David A. Buell, Brian Burkett, Nicholas Bushnell, Zijun Chen, Benjamin Chiaro, Roberto Collins, William Courtney, Sean Demura, Alan R. Derk, Andrew Dunsworth, Daniel Eppens, Catherine Erickson, Edward Farhi, Austin G. Fowler, Brooks Foxen, Craig Gidney, Marissa Giustina, Jonathan A. Gross, Matthew P. Harrigan, Sean D. Harrington, Jeremy Hilton, Alan Ho, Sabrina Hong, Trent Huang, William J. Huggins, L. B. Ioffe, Sergei V. Isakov, Evan Jeffrey, Zhang Jiang, Cody Jones, Dvir Kafri, Julian Kelly, Seon Kim, Alexei Kitaev, Paul V. Klimov, Alexander N. Korotkov, Fedor Kostritsa, David Landhuis, Pavel Laptev, Erik Lucero, Orion Martin, Jarrod R. McClean, Trevor McCourt, Matt McEwen, Anthony Megrant, Kevin C. Miao, Masoud Mohseni, Wojciech Mruczkiewicz, Josh Mutus, Ofer Naaman, Matthew Neeley, Michael Newman, Murphy Yuezhen Niu, Thomas E. O’Brien, Alex Opremcak, Eric Ostby, Balint Pato, Andre Petukhov, Nicholas Redd, Nicholas C. Rubin, Daniel Sank, Kevin J. Satzinger, Vladimir Shvarts, Doug Strain, Marco Szalay, Matthew D. Trevithick, Benjamin Villalonga, Theodore White, Z. Jamie Yao, Ping Yeh, Adam Zalcman, Hartmut Neven, Igor Aleiner, Kostyantyn Kechedzhi, Vadim Smelyanskiy, and Yu Chen, “Information scrambling in computationally complex quantum circuits,” *Science* **374**, 1479 (2021).
- [4] Google Quantum AI and Collaborators, “Constructive interference at the edge of quantum ergodic dynamics,” *arXiv:2506.10191* (2025).
- [5] Alexander Novikov, Ngan Vuu, Marvin Eisenberger, Emilien Dupont, Po-Sen Huang, Adam Zsolt Wagner, Sergey Shirobokov, Borislav Kozlovskii, Francisco J. R. Ruiz, Abbas Mehrabian, M. Pawan Kumar, Abigail See, Swarat Chaudhuri, George Holland, Alex Davies, Sebastian Nowozin, Pushmeet Kohli, and Matej Balog, “Alphaevolve: A coding agent for scientific and algorithmic discovery,” (2025), *arXiv:2506.13131 [cs.AI]*.
- [6] Malcolm H Levitt, *Spin dynamics: basics of nuclear magnetic resonance* (John Wiley & Sons, 2008).
- [7] Hyungjoon Cho, Thaddeus D Ladd, Jonathan Baugh, David G Cory, and Chandrasekhar Ramanathan, “Multispin dynamics of the solid-state nmr free induction decay,” *Physical Review B—Condensed Matter and Materials Physics* **72**, 054427 (2005).
- [8] Mark C Butler, Jean-Nicolas Dumez, and Lyndon Emsley, “Dynamics of large nuclear-spin systems from low-order correlations in liouville space,” *Chemical Physics Letters* **477**, 377–381 (2009).
- [9] Jean-Nicolas Dumez, Meghan E Halse, Mark C Butler, and Lyndon Emsley, “A first-principles description of proton-driven spin diffusion,” *Physical Chemistry Chemical Physics* **14**, 86–89 (2012).
- [10] J. Baum, M. Munowitz, A. N. Garboway, and A. Pines, “Multiple-quantum dynamics in solid state NMR,” *The Journal of Chemical Physics* **83**, 2015–2025 (1985).
- [11] Terry Gullion and Jacob Schaefer, “Rotational-echo double-resonance nmr,” *Journal of Magnetic Resonance* (1969) **81**, 196–200 (1989).
- [12] Andrew E Bennett, Chad M Rienstra, Janet M Griffiths, Weiguo Zhen, Peter T Lansbury Jr, and Robert G Griffin, “Homonuclear radio frequency-driven recoupling in rotating solids,” *The Journal of chemical physics* **108**, 9463–9479 (1998).
- [13] M.G. Colombo, B.H. Meier, and R.R. Ernst, “Rotor-driven spin diffusion in natural-abundance <sup>13</sup>C spin systems,” *Chemical Physics Letters* **146**, 189–196 (1988).
- [14] Andreas Grommek, Beat H. Meier, and Matthias Ernst, “Distance information from proton-driven spin diffusion under mas,” *Chemical Physics Letters* **427**, 404–409 (2006).
- [15] D. P. Raleigh, F. Creuzet, S. K. Das Gupta, M. H. Levitt, and Robert G. Griffin, “Measurement of internuclear distances in polycrystalline solids. rotationally enhanced transfer of nuclear spin magnetization,” *Journal of the American Chemical Society* **111**, 4502–4503 (1989), <https://doi.org/10.1021/ja00194a057>.
- [16] F. Creuzet, A. McDermott, R. Gebhard, K. van der Hoef, M. B. Spijker-Assink, J. Herzfeld, J. Lugtenburg, M. H. Levitt, and R. G. Griffin, “Determination of membrane protein structure by rotational resonance nmr: Bacteriorhodopsin,” *Science* **251**, 783–786 (1991), <https://www.science.org/doi/pdf/10.1126/science.1990439>.
- [17] W. S. Warren and A. Pines, “Analogy of multiple-quantum nmr to isotopic spin labeling,” *Journal of the American Chemical Society* **103**, 1613–1617 (1981), <https://doi.org/10.1021/ja00397a001>.
- [18] James M. Polson and E. Elliott Burnell, “Conformational equilibrium and orientational ordering: <sup>1</sup>H-nuclear magnetic resonance of butane in a nematic liquid crystal,” *The Journal of Chemical Physics* **103**, 6891–6902 (1995).
- [19] Aneta T Petkova, Yoshitaka Ishii, John J Balbach, Oleg N Antzutkin, Richard D Leapman, Frank Delaglio, and Robert Tycko, “A structural model for alzheimer’s  $\beta$ -amyloid fibrils based on experimental constraints from solid state nmr,” *Proceedings of the National Academy of Sciences* **99**, 16742–16747 (2002).
- [20] Christian Wasmer, Adam Lange, Hélène Van Melckebeke, Ansgar B Siemer, Roland Riek, and Beat H Meier, “Amyloid fibrils of the het-s (218–289) prion form a  $\beta$  solenoid with a triangular hydrophobic core,” *Science* **319**, 1523–1526 (2008).
- [21] João Medeiros-Silva, Aurelio J. Dregni, Noah H. Somberg, Pu Duan, and Mei Hong, “Atomic structure of the open

- SARS-CoV-2 E viroporin," *Science Advances* **9**, eadi9007 (2023).
- [22] Alexander A. Shcherbakov, João Medeiros-Silva, Nhi Tran, Martin D. Gelenter, and Mei Hong, "From Angstroms to Nanometers: Measuring Interatomic Distances by Solid-State NMR," *Chemical Reviews* **122**, 9848–9879 (2022).
- [23] Dries Sels, Hesam Dashti, Samia Mora, Olga Demler, and Eugene Demler, "Quantum approximate bayesian computation for nmr model inference," *Nature machine intelligence* **2**, 396–402 (2020).
- [24] Thomas E O'Brien, Lev B Ioffe, Yuan Su, David Fushman, Hartmut Neven, Ryan Babbush, and Vadim Smelyanskiy, "Quantum computation of molecular structure using data from challenging-to-classically-simulate nuclear magnetic resonance experiments," *PRX Quantum* **3**, 030345 (2022).
- [25] Manuel G Algaba, Mario Ponce-Martinez, Carlos Munuera-Javaloy, Vicente Pina-Canelles, Manish J Thapa, Bruno G Taketani, Martin Leib, Inés De Vega, Jorge Casanova, and Hermanni Heimonen, "Co-design quantum simulation of nanoscale nmr," *Physical Review Research* **4**, 043089 (2022).
- [26] Thomas Schuster, Murphy Niu, Jordan Cotler, Thomas O'Brien, Jarrod R McClean, and Masoud Mohseni, "Learning quantum systems via out-of-time-order correlators," *Physical Review Research* **5**, 043284 (2023).
- [27] Kushal Seetharam, Debopriyo Biswas, Crystal Noel, Andrew Risinger, Daiwei Zhu, Or Katz, Sambuddha Chattopadhyay, Marko Cetina, Christopher Monroe, Eugene Demler, *et al.*, "Digital quantum simulation of nmr experiments," *Science Advances* **9**, eadh2594 (2023).
- [28] Keith R Fratus, Nicklas Enenkel, Sebastian Zanker, Jan-Michael Reiner, Michael Marthaler, and Peter Schmitteckert, "Can a quantum computer simulate nuclear magnetic resonance spectra better than a classical one?" *arXiv preprint arXiv:2508.06448* (2025).
- [29] Michael Marthaler, Peter Pinski, Pascal Stadler, Vladimir Rybkin, and Marina Walt, "What is a good use case for quantum computers?" *arXiv preprint arXiv:2506.15426* (2025).
- [30] Armando Angrisani, Alexander Schmidhuber, Manuel S Rudolph, M Cerezo, Zoë Holmes, and Hsin-Yuan Huang, "Classically estimating observables of noiseless quantum circuits," *arXiv preprint arXiv:2409.01706* (2024).
- [31] Kostyantyn Kechedzhi, Sergei V Isakov, Salvatore Mandrà, Benjamin Villalonga, Xiao Mi, Sergio Boixo, and Vadim Smelyanskiy, "Effective quantum volume, fidelity and computational cost of noisy quantum processing experiments," *Future Generation Computer Systems* **153**, 431–441 (2024).
- [32] Thomas Schuster, Chao Yin, Xun Gao, and Norman Y Yao, "A polynomial-time classical algorithm for noisy quantum circuits," *arXiv preprint arXiv:2407.12768* (2024).
- [33] Jun Li, Ruihua Fan, Hengyan Wang, Bingtian Ye, Bei Zeng, Hui Zhai, Xinhua Peng, and Jiangfeng Du, "Measuring out-of-time-order correlators on a nuclear magnetic resonance quantum simulator," *Physical Review X* **7**, 031011 (2017).
- [34] Leslie D Field and Saadallah A Ramadan, "Multiple quantum nmr spectra of toluene and p-bromotoluene partially aligned in a nematic phase," *Magnetic Resonance in Chemistry* **41**, 933–938 (2003).
- [35] Andrew M Childs, Yuan Su, Minh C Tran, Nathan Wiebe, and Shuchen Zhu, "Theory of trotter error with commutator scaling," *Physical Review X* **11**, 011020 (2021).
- [36] Guang Hao Low and Nathan Wiebe, "Hamiltonian simulation in the interaction picture," *arXiv preprint arXiv:1805.00675* (2018).
- [37] Ian D Kivlichan, Jarrod McClean, Nathan Wiebe, Craig Gidney, Alán Aspuru-Guzik, Garnet Kin-Lic Chan, and Ryan Babbush, "Quantum simulation of electronic structure with linear depth and connectivity," *Physical review letters* **120**, 110501 (2018).
- [38] Kristan Temme, Sergey Bravyi, and Jay M Gambetta, "Error mitigation for short-depth quantum circuits," *Physical review letters* **119**, 180509 (2017).
- [39] Ying Li and Simon C Benjamin, "Efficient variational quantum simulator incorporating active error minimization," *Physical Review X* **7**, 021050 (2017).
- [40] Ann McDermott, "Structure and dynamics of membrane proteins by magic angle spinning solid-state nmr," *Annual review of biophysics* **38**, 385–403 (2009).
- [41] Yun Han, Jinwoo Ahn, Jason Concel, In-Ja L Byeon, Angela M Gronenborn, Jun Yang, and Tatyana Polenova, "Solid-state nmr studies of hiv-1 capsid protein assemblies," *Journal of the American Chemical Society* **132**, 1976–1987 (2010).
- [42] Xueqian Kong, Hexiang Deng, Fangyong Yan, Jihan Kim, Joseph A Swisher, Berend Smit, Omar M Yaghi, and Jeffrey A Reimer, "Mapping of functional groups in metal-organic frameworks," *Science* **341**, 882–885 (2013).
- [43] Alexander Karabanov, Ilya Kuprov, G. T. P. Charnock, Anniek van der Drift, Luke J. Edwards, and Walter Köckenberger, "On the accuracy of the state space restriction approximation for spin dynamics simulations," *The Journal of Chemical Physics* **135**, 084106 (2011).
- [44] L Stryer and R P Haugland, "Energy transfer: a spectroscopic ruler." *Proceedings of the National Academy of Sciences* **58**, 719–726 (1967), <https://www.pnas.org/doi/pdf/10.1073/pnas.58.2.719>.
- [45] Alexey Krushelnitsky, Detlef Reichert, and Kay Saalwächter, "Solid-state nmr approaches to internal dynamics of proteins: from picoseconds to microseconds and seconds," *Accounts of chemical research* **46**, 2028–2036 (2013).
- [46] Andrew M Childs, Aaron Ostrander, and Yuan Su, "Faster quantum simulation by randomization," *Quantum* **3**, 182 (2019).
- [47] Earl Campbell, "Random compiler for fast hamiltonian simulation," *Physical review letters* **123**, 070503 (2019).
- [48] Hongzheng Zhao, Marin Bukov, Markus Heyl, and Roderich Moessner, "Making trotterization adaptive and energy-self-correcting for nisq devices and beyond," *PRX Quantum* **4**, 030319 (2023).
- [49] Refik Mansuroglu, Felix Fischer, and Michael J Hartmann, "Problem-specific classical optimization of hamiltonian simulation," *Physical Review Research* **5**, 043035 (2023).
- [50] Linus Jern, Valter Uotila, Cong Yu, and Bo Zhao, "Agent-q: Fine-tuning large language models for quantum circuit

- generation and optimization,” (2025), arXiv:2504.11109 [quant-ph].
- [51] D.P Burum, M Linder, and R.R Ernst, “Low-power multipulse line narrowing in solid-state nmr,” *Journal of Magnetic Resonance (1969)* **44**, 173–188 (1981).
- [52] Leandro Martínez, Ricardo Andrade, Ernesto G Birgin, and José Mario Martínez, “Packmol: A package for building initial configurations for molecular dynamics simulations,” *Journal of computational chemistry* **30**, 2157–2164 (2009).
- [53] Peter Eastman, Raimondas Galvelis, Raúl P Peláez, Charlles RA Abreu, Stephen E Farr, Emilio Gallicchio, Anton Gorenko, Michael M Henry, Frank Hu, Jing Huang, *et al.*, “Openmm 8: molecular dynamics simulation with machine learning potentials,” *The Journal of Physical Chemistry B* **128**, 109–116 (2023).
- [54] Junmei Wang, Romain M Wolf, James W Caldwell, Peter A Kollman, and David A Case, “Development and testing of a general amber force field,” *Journal of computational chemistry* **25**, 1157–1174 (2004).
- [55] Leif Seute, Eric Hartmann, Jan Stühmer, and Frauke Gräter, “Grappa—a machine learned molecular mechanics force field,” *Chemical Science* **16**, 2907–2930 (2025).
- [56] Jianguo Zhang, Jiaye Su, and Hongxia Guo, “An atomistic simulation for 4-cyano-4'-pentylbiphenyl and its homologue with a reoptimized force field,” *The Journal of Physical Chemistry B* **115**, 2214–2227 (2011).
- [57] Roger G Horn, “Refractive indices and order parameters of two liquid crystals,” *Journal de physique* **39**, 105–109 (1978).
- [58] PL Sherrell and DA Crellin, “Susceptibilities and order parameters of nematic liquid crystals,” *Le Journal de Physique Colloques* **40**, C3–211 (1979).
- [59] Ivo Cacelli, Silvio Campanile, Giacomo Prampolini, and Alessandro Tani, “Stability of the nematic phase of 4-n-pentyl-4'-cyanobiphenyl studied by computer simulation using a hybrid model,” *The Journal of chemical physics* **117**, 448–453 (2002).
- [60] Jean-Baptiste Mouret and Jeff Clune, “Illuminating search spaces by mapping elites,” (2015), arXiv:1504.04909 [cs.AI].
- [61] Jan Lukas Bosse, Andrew M Childs, Charles Derby, Filippo Maria Gambetta, Ashley Montanaro, and Raul A Santos, “Efficient and practical hamiltonian simulation from time-dependent product formulas,” *Nat. Commun.* **16**, 2673 (2025).
- [62] Robert R. Tucci, “An introduction to cartan’s kak decomposition for qc programmers,” (2005), arXiv:quant-ph/0507171 [quant-ph].
- [63] Rajeev Acharya, Laleh Aghababaie-Beni, Igor Aleiner, Trond I. Andersen, *et al.*, “Quantum error correction below the surface code threshold,” *Nature* **638**, 920–926 (2025).
- [64] Brooks Foxen, Charles Neill, Andrew Dunsworth, Pedram Roushan, Ben Chiaro, Anthony Megrant, Julian Kelly, Zijun Chen, Kevin Satzinger, Rami Barends, *et al.*, “Demonstrating a continuous set of two-qubit gates for near-term quantum algorithms,” *Physical Review Letters* **125**, 120504 (2020).
- [65] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, *et al.*, “Observation of separated dynamics of charge and spin in the Fermi-Hubbard model,” *Science Advances* **6**, eaay5901 (2020), arXiv:2010.07965 [quant-ph].
- [66] Xiao Mi, Matteo Ippoliti, Roderich Moessner, S. L. Sondhi, *et al.*, “Observation of time-crystalline eigenstate order on a quantum processor,” *Nature* **601**, 53–57 (2022), arXiv:2107.13571 [quant-ph].
- [67] Chris Neill, Tom McCourt, Xiao Mi, Zhang Jiang, Muqing Y. Niu, *et al.*, “Accurately computing the electronic properties of a quantum ring,” *Nature* **594**, 508–512 (2021), arXiv:2012.00921 [quant-ph].
- [68] Xiao Mi, Marius Sonner, Muqing Y. Niu, Kang W. Lee, Benjamin Foxen, Rajeev Acharya, Igor Aleiner, Trond I. Andersen, Frank Arute, *et al.*, “Noise-resilient edge modes on a chain of superconducting qubits,” *Science* **378**, 785–790 (2022), arXiv:2204.11372 [quant-ph].
- [69] Alexis Morvan, Trond I. Andersen, Xiao Mi, Charles Neill, *et al.*, “Formation of robust bound states of interacting microwave photons,” *Nature* **612**, 240–245 (2022), arXiv:2206.05254 [quant-ph].
- [70] Kristan Temme, Sergey Bravyi, and Jay M. Gambetta, “Error mitigation for short-depth quantum circuits,” *Physical Review Letters* **119** (2016), 10.1103/PhysRevLett.119.180509.
- [71] Ying Li and Simon C. Benjamin, “Efficient variational quantum simulator incorporating active error minimization,” *Physical Review X* **7**, 021050 (2017).
- [72] Zhenyu Cai, Ryan Babbush, Simon C. Benjamin, Suguru Endo, William J. Huggins, Ying Li, Jarrod R. McClean, and Thomas E. O’Brien, “Quantum error mitigation,” *Reviews of Modern Physics* **95**, 045005 (2023), arXiv:2210.00921 [quant-ph].
- [73] Zlatko K. Minev, Khadijeh Najafi, Swarnadeep Majumder, Juven Wang, Ady Stern, Eun-Ah Kim, Chao-Ming Jian, and Guanyu Zhu, “Realizing string-net condensation: Fibonacci anyon braiding for universal gates and sampling chromatic polynomials,” *Nature Communications* **16**, 6225 (2025).
- [74] Charles H. Bennett, Gilles Brassard, Sandu Popescu, Benjamin Schumacher, John A. Smolin, and William K. Wootters, “Purification of noisy entanglement and faithful teleportation via noisy channels,” *Physical Review Letters* **76**, 722–725 (1995).
- [75] E. Knill, “Fault-tolerant postselected quantum computation: Threshold analysis,” arXiv e-prints , quant-ph/0404104 (2004), arXiv:quant-ph/0404104 [quant-ph].
- [76] Lorenza Viola and Seth Lloyd, “Dynamical suppression of decoherence in two-state quantum systems,” *Phys. Rev. A* **58**, 2733–2744 (1998), quant-ph/9803057.
- [77] Lorenza Viola, Emanuel Knill, and Seth Lloyd, “Dynamical Decoupling of Open Quantum Systems,” *Phys. Rev. Lett.* **82**, 2417–2421 (1999), quant-ph/9809071.
- [78] Ewout van den Berg, Zlatko K. Minev, and Kristan Temme, “Model-free readout-error mitigation for quantum expectation values,” *Phys. Rev. A* **105**, 032620 (2022).
- [79] Manuel S Rudolph, Tyson Jones, Yanting Teng, Armando Angrisani, and Zoë Holmes, “Pauli propagation: A computational framework for simulating quantum systems,” arXiv preprint arXiv:2505.21606 (2025).

- [80] Michael A Nielsen and Isaac L Chuang, *Quantum computation and quantum information* (Cambridge university press, 2010).
- [81] Alfred Saupe, “Kernresonanzen in kristallinen flüssigkeiten und in kristallinflüssigen lösungen. teil i,” *Zeitschrift für Naturforschung A* **19**, 161–171 (1964).
- [82] Jean Baum, Michael Munowitz, Allen N Garroway, and Alex Pines, “Multiple-quantum dynamics in solid state nmr,” *The Journal of chemical physics* **83**, 2015–2025 (1985).
- [83] Ulrich Haeberlen and John S Waugh, “Coherent averaging effects in magnetic resonance,” *Physical Review* **175**, 453 (1968).
- [84] M Hohwy and NC Nielsen, “Systematic design and evaluation of multiple-pulse experiments in nuclear magnetic resonance spectroscopy using a semi-continuous baker–campbell–hausdorff expansion,” *The Journal of chemical physics* **109**, 3780–3791 (1998).
- [85] Ken Xuan Wei, Chandrasekhar Ramanathan, and Paola Cappellaro, “Exploring localization in nuclear spin chains,” *Physical review letters* **120**, 070501 (2018).
- [86] Marc Guenneugues, Patrick Berthault, and Hervé Desvaux, “A method for determining b1field inhomogeneity. are the biases assumed in heteronuclear relaxation experiments usually underestimated?” *Journal of Magnetic Resonance* **136**, 118–126 (1999).
- [87] DP Burum, M Linder, and RR Ernst, “Low-power multipulse line narrowing in solid-state nmr,” *Journal of Magnetic Resonance* (1969) **44**, 173–188 (1981).
- [88] Mei Hong, Alexander Pines, and Stefano Caldarelli, “Measurement and assignment of long-range c- h dipolar couplings in liquid crystals by two-dimensional nmr spectroscopy,” *The Journal of Physical Chemistry* **100**, 14815–14822 (1996).
- [89] DP Weitekamp, JR Garbow, and A Pines, “Determination of dipole coupling constants using heteronuclear multiple quantum nmr,” *The Journal of Chemical Physics* **77**, 2870–2883 (1982).
- [90] Ted Schaefer, James Peeling, and Glenn H Penner, “The spin–spin coupling mechanism for 5 j (c-4, ch $\alpha$ ) in toluene derivatives and its conformational applications,” *Canadian Journal of Chemistry* **61**, 2773–2776 (1983).
- [91] G Drobny, A Pines, S Sinton, D Weitekamp, and D Wemmer, “Faraday div. c’hem,” in *SW. Symp*, Vol. 13 (1979) p. 49.
- [92] Sunghwan Kim, Paul A Thiessen, Evan E Bolton, Jie Chen, Gang Fu, Asta Gindulyte, Lianyi Han, Jane He, Siqian He, Benjamin A Shoemaker, *et al.*, “Pubchem substance and compound databases,” *Nucleic acids research* **44**, D1202–D1213 (2016).
- [93] Christian Bengs, Chongwei Zhang, and Ashok Ajoy, “Fundamental bounds on many-body spin cluster intensities,” *The Journal of Chemical Physics* **162** (2025).
- [94] SW Sinton, DB Zax, JB Murdoch, and A Pines, “Multiple-quantum nmr study of molecular structure and ordering in a liquid crystal,” *Molecular Physics* **53**, 333–362 (1984).
- [95] Tianci Zhou, Shenglong Xu, Xiao Chen, Andrew Guo, and Brian Swingle, “The operator lévy flight: light cones in chaotic long-range interacting systems,” *Phys. Rev. Lett.* **124**, 180601 (2020).
- [96] Antonio Pizzirusso, Maria Enrica Di Pietro, Giuseppina De Luca, Giorgio Celebre, Marcello Longeri, Luca Muccioli, and Claudio Zannoni, “Order and conformation of biphenyl in cyanobiphenyl liquid crystals: a combined atomistic molecular dynamics and 1h nmr study,” *ChemPhysChem* **15**, 1356–1367 (2014).
- [97] Wen Guo and BM Fung, “Determination of the order parameters of liquid crystals from carbon-13 chemical shifts,” *The Journal of chemical physics* **95**, 3917–3923 (1991).
- [98] Dick Sandström, Andrei V Komolkin, and Arnold Maliniak, “Orientational order in a liquid crystalline mixture studied by molecular dynamics simulation and nmr,” *The Journal of chemical physics* **104**, 9620–9628 (1996).
- [99] Fernando Nogueira, “Bayesian Optimization: Open source constrained global optimization tool for Python,” (2014–).
- [100] Kévin Hémery, Frank Pollmann, and David J. Luitz, “Matrix product states approaches to operator spreading in ergodic quantum systems,” *Phys. Rev. B* **100**, 104303 (2019).
- [101] Shenglong Xu and Brian Swingle, “Accessing scrambling using matrix product operators,” *Nature Physics* **16**, 199–204 (2020).
- [102] Javier Lopez-Piqueres, Brayden Ware, Sarang Gopalakrishnan, and Romain Vasseur, “Operator front broadening in chaotic and integrable quantum chains,” *Phys. Rev. B* **104**, 104307 (2021).
- [103] Shenglong Xu and Brian Swingle, “Scrambling dynamics and out-of-time-ordered correlators in quantum many-body systems,” *PRX Quantum* **5**, 010201 (2024).
- [104] Michael P. Zaletel, Roger S. K. Mong, Christoph Karrasch, Joel E. Moore, and Frank Pollmann, “Time-evolving a matrix product state with long-ranged interactions,” *Phys. Rev. B* **91**, 165112 (2015).
- [105] Daniel E. Parker, Xiangyu Cao, and Michael P. Zaletel, “Local matrix product operators: Canonical form, compression, and control theory,” *Phys. Rev. B* **102**, 035147 (2020).
- [106] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando G. S. L. Bradao, *et al.*, “Quantum supremacy using a programmable superconducting processor,” *Nature* **574**, 505–510 (2019).
- [107] Elliott Rosenberg, TI Andersen, Rhine Samajdar, Andre Petukhov, JC Hoke, Dmitry Abanin, Andreas Bengtsson, IK Drozdov, Catherine Erickson, PV Klimov, *et al.*, “Dynamics of magnetization at infinite temperature in a heisenberg spin chain,” *Science* **384**, 48–53 (2024).
- [108] Sergei V. Isakov, Dvir Kafri, Orion Martin, Catherine Vollgraff Heidweiller, Wojciech Mruczkiewicz, Matthew P. Harrigan, Nicholas C. Rubin, Ross Thomson, Michael Broughton, Kevin Kissell, Evan Peters, Erik Gustafson, Andy C. Y. Li, Henry Lamm, Gabriel Perdue, Alan K. Ho, Doug Strain, and Sergio Boixo, “Simulations of quantum circuits with approximate noise using qsim and cirq,” (2021), arXiv:2111.02396 [quant-ph].