

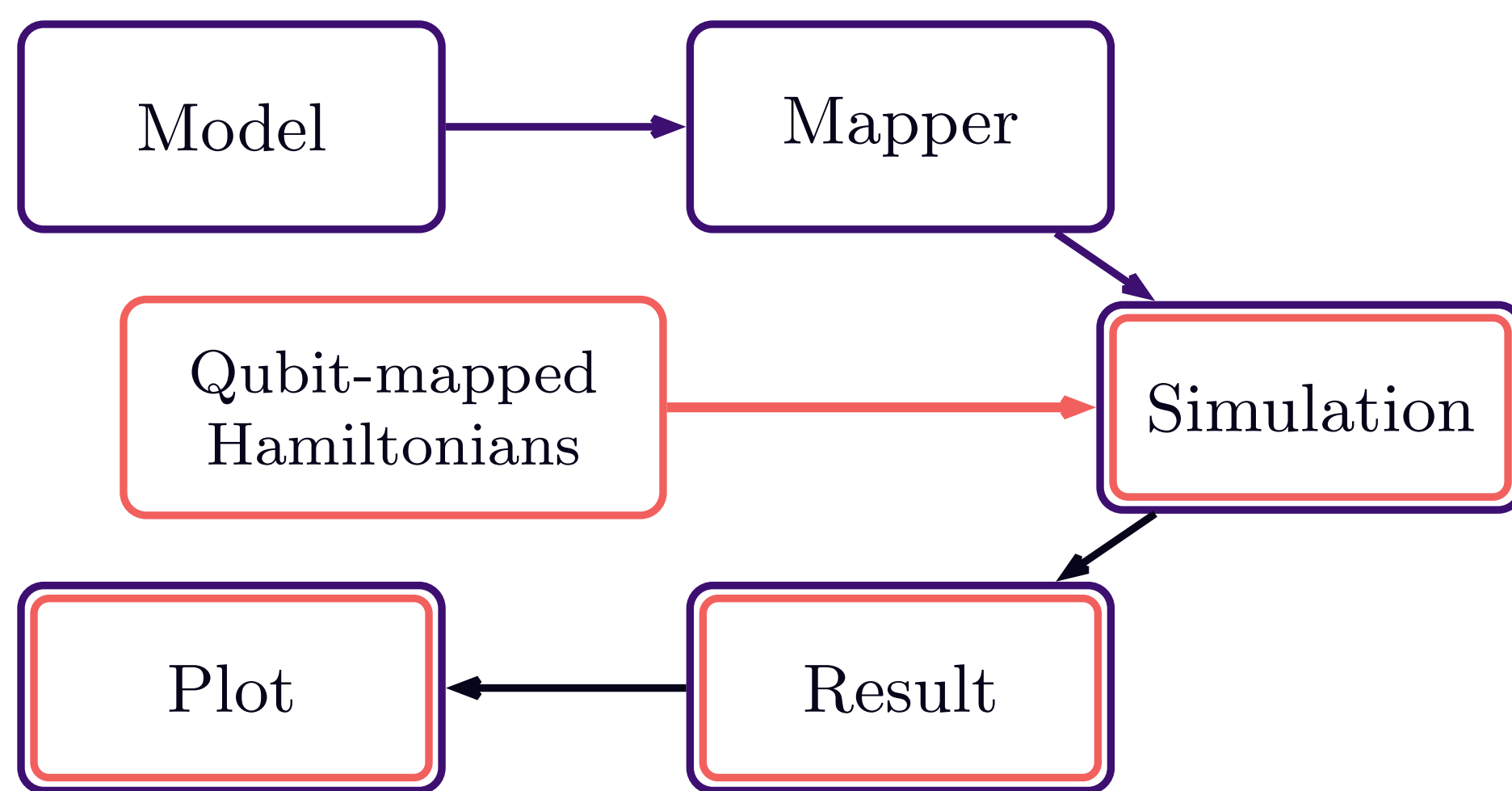
1 Why phase diagrams?

A Hamiltonian alone is not a benchmark

Benchmarks like **HamPerf** and **HamLib** make the Hamiltonian the unit of evaluation, but it must still be **mapped, swept, executed** and **reduced to an observable** through hand-built scripts.

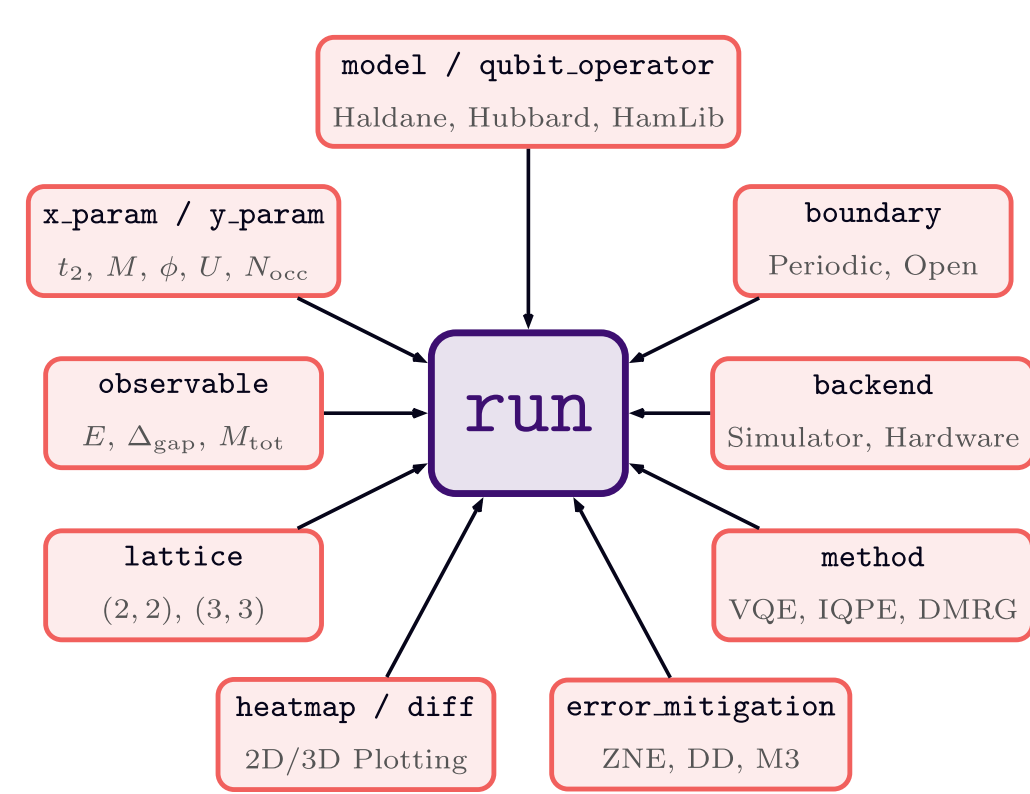
A **phase diagram** judges a method on whether it reproduces the *shape of a whole landscape*, including sharp boundaries, against **exact** and **DMRG** baselines.

2 One end-to-end pipeline



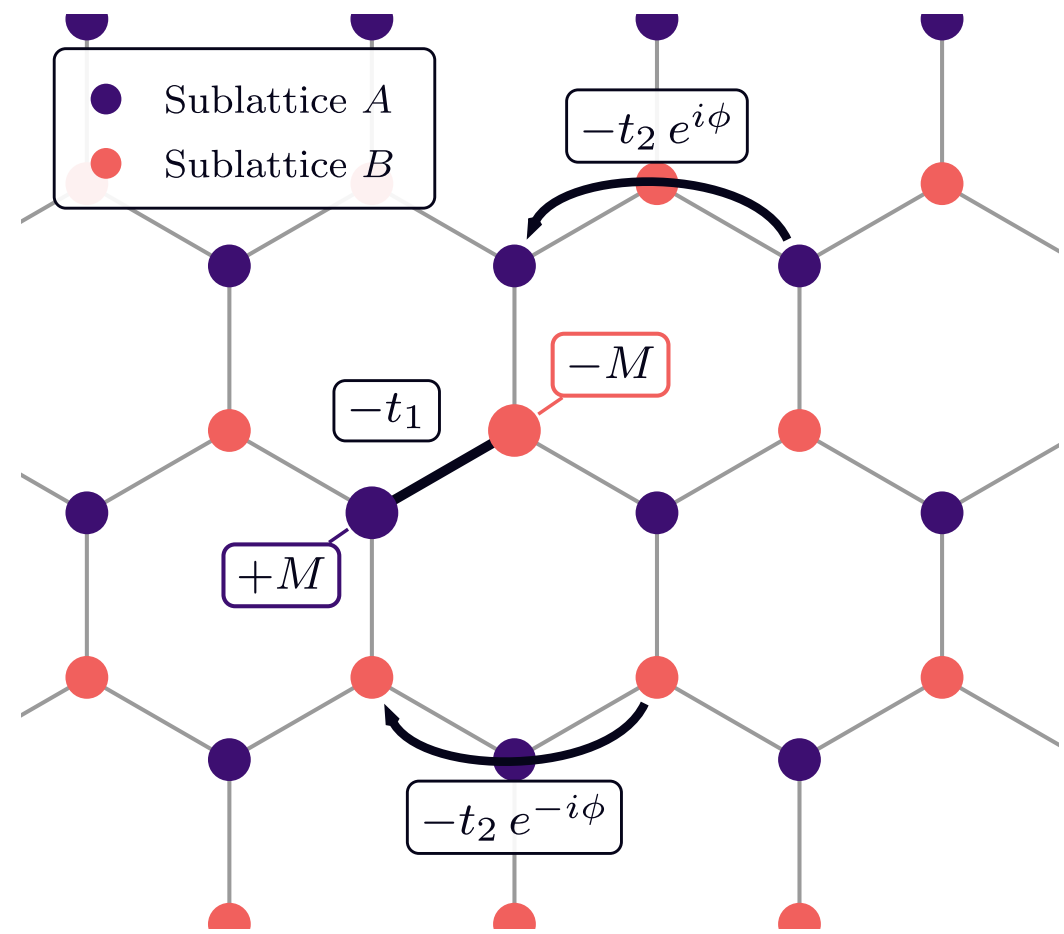
Built-in models pass through representation and simulation stages. Pre-mapped **HamLib** or custom qubit operators enter directly at *Simulation*; all paths produce logged results and plots.

One run specification



Choose the **Hamiltonian, lattice, boundary, sweep axes, observable, method, backend and mitigation independently**. New methods and investigations use modular extension interfaces.

Lattice-aware model encoding

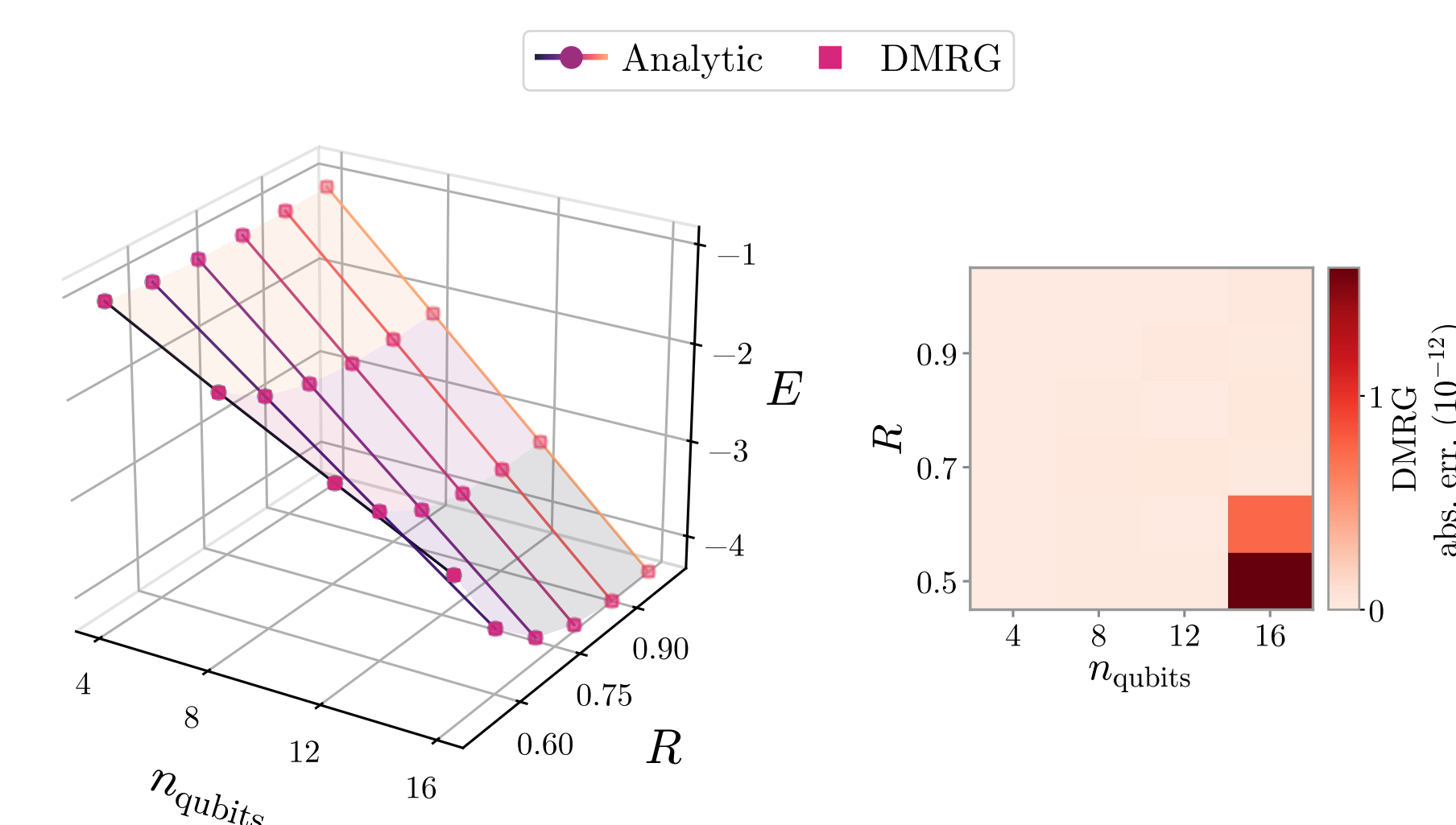


Methods and execution

Exact diagonalization provides tractable references. **VQE** (variational quantum eigensolver) and **IQPE** (iterative quantum phase estimation) provide quantum paths; **DMRG** (density matrix renormalization group) supplies a classical tensor-network comparison.

Backends span ideal and noisy simulation and real quantum hardware, with shared error-mitigation options.

3 Any Hamiltonian from HamLib

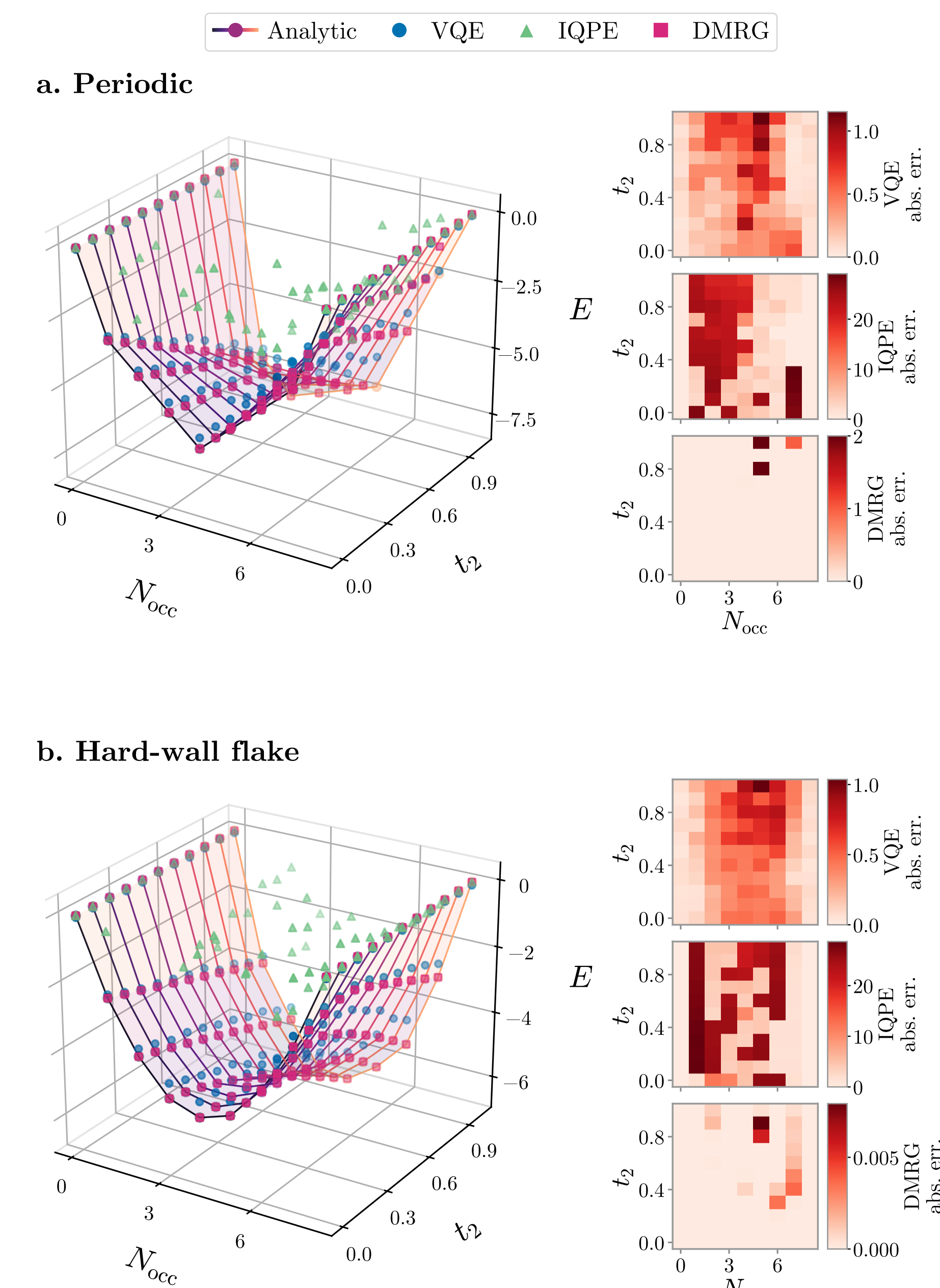


H_2 - H_8 chains in STO-6G: **24 Hamiltonians**, 4–16 mapped qubits and $R = 0.5$ – 1.0 Å. Exact and DMRG energies use the same workflow as the TFIM and 3-SAT examples.

4 Boundaries matter

Haldane: periodic vs hard-wall

2×2 unit cells; $t_1 = 1$; $M = 0$; $\phi = \pi/4$



Energies versus occupation and next-nearest hopping. Small panels show absolute errors on independent color scales.

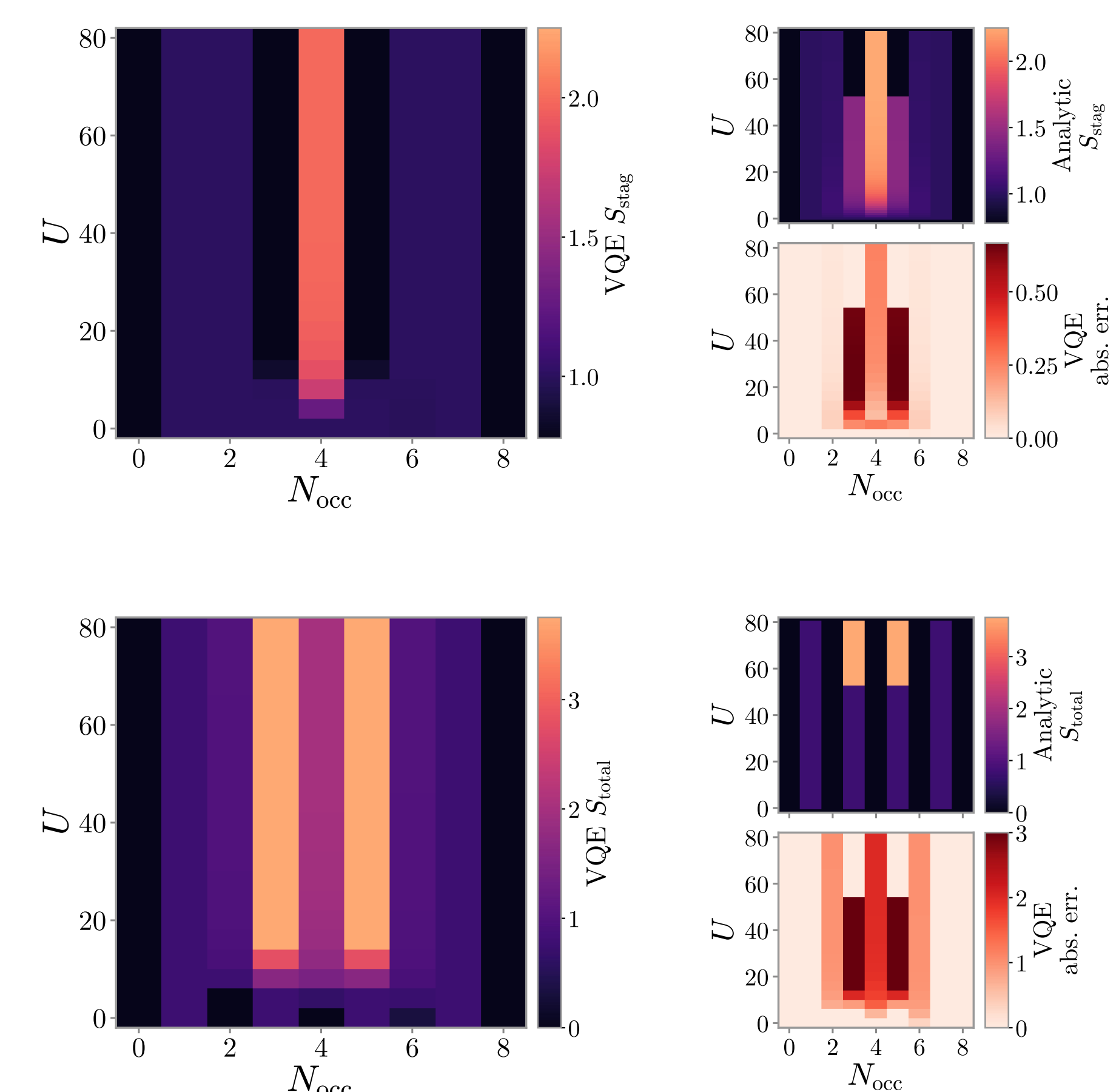
VQE tracks the exact surface in both geometries, with errors below about one energy unit. **IQPE** is less uniform. Isolated cells miss by about 20 units because one fixed evolution time, Trotter depth and iteration count does not suit every filling. **DMRG** is nearly exact on the hard-wall flake (errors below 5×10^{-3}).

Geometry becomes part of the benchmark.

5 Beyond energy: magnetism

Four-site cluster; periodic 2×1 lattice

Eight spin orbitals; $t = 1$; $U = 0$ – 80



Staggered (top) and total-spin (bottom) structure factors compare VQE, the exact reference and absolute error.

Exact crossing versus VQE onset

$U = 53$ vs $U \approx 16$

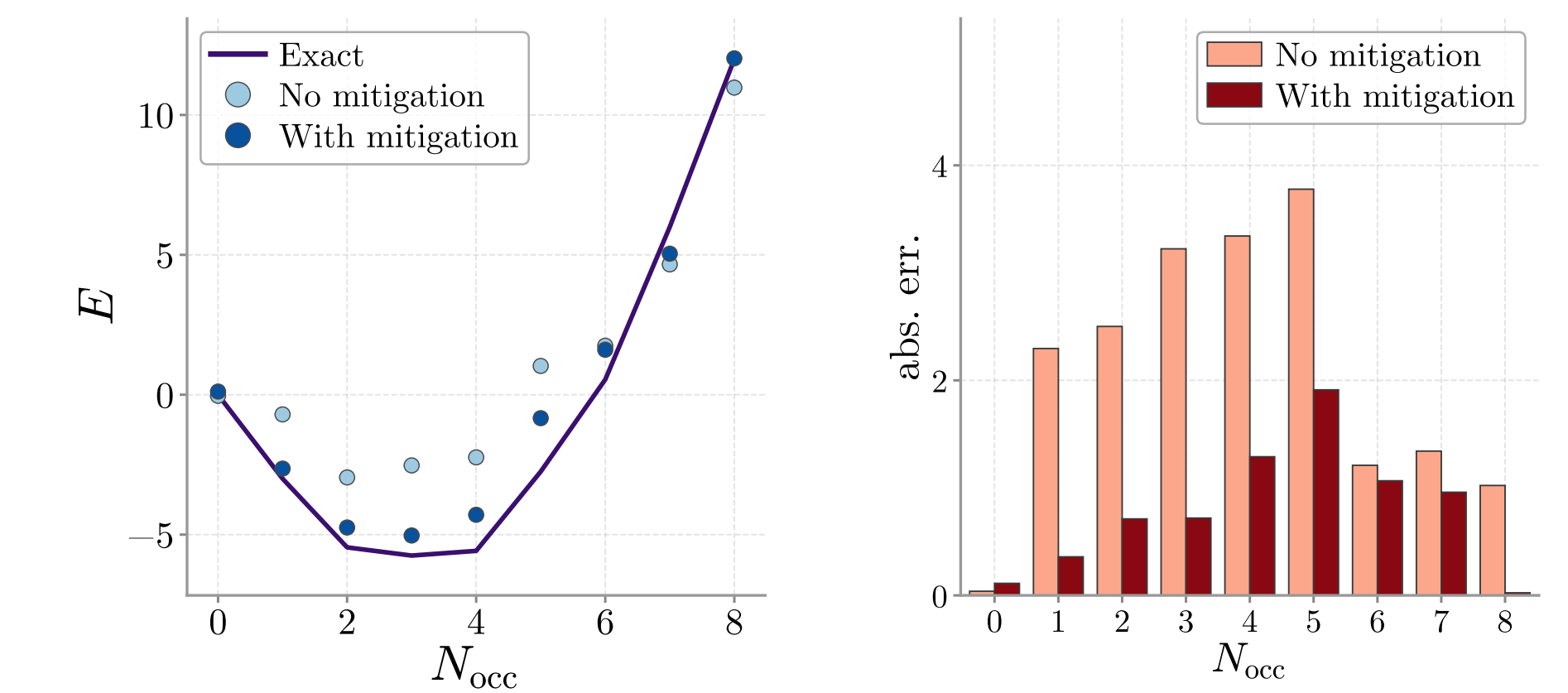
At occupations 3 and 5, VQE enters the **fully polarized regime too early**. The maps reveal a structured mismatch while capturing the half-filled antiferromagnetic trend more closely.

A small-cluster effect, not a true phase transition.

6 Running on real hardware

Hubbard energy versus occupation

8 qubits; $t = 1$, $U = 3$; 5 SPSA iterations



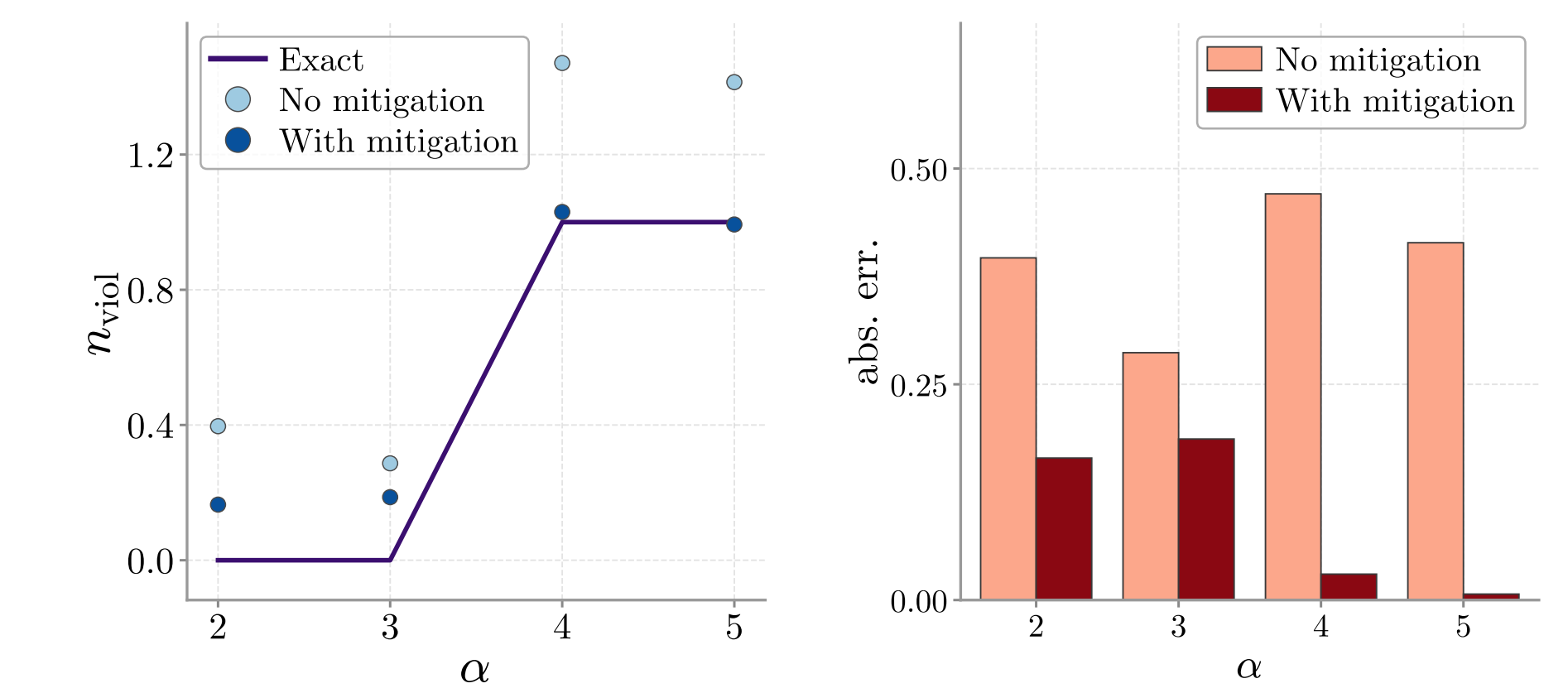
Observed mean $|\Delta E|$: unmitigated $\rightarrow$ mitigated

2.08 $\rightarrow$ 0.79

Both runs reproduce the broad filling dependence. The mitigated run has **lower mean error**, but correlated fillings retain substantial deviations from the exact reference.

Four-variable 3-SAT

4 qubits; $a = 2, 3, 4, 5$; 50 COBYLA iterations



Observed mean clause error: unmitigated $\rightarrow$ mitigated

0.39 $\rightarrow$ 0.10

Both runs round to the **exact violated-clause count** at all four sampled points. This compares a fixed instance family rather than probing the large-system random-3-SAT threshold.

Why compare both workloads?

One interface spans a particle-number-conserving lattice model and a diagonal Boolean cost model. The meaningful criterion is *accuracy on the chosen observable*; energy errors and clause errors use different units.

Hardware protocol and scope

All hardware runs use the **IQM Garnet** device. Each condition uses one independent VQE optimization per parameter cell and **4,096 shots** per circuit. The stack combines **M3** readout mitigation, **dynamical decoupling** and **two-point ZNE** at factors $\{1, 3\}$. Lower errors were observed in these runs; independent-run uncertainty and a causal mitigation benefit have not been established.

7 Takeaways and next steps

A landscape, not a single circuit

One sweep tests a method across physics, geometry and algorithm settings at once: smooth regions test its range while sharp boundaries test its resolution.

Scope and next steps

Our experiments used 4–16 qubits and were always compared against exact references. QBP will be published as **open source** on PyPI; new models, methods and mitigation strategies plug in through modular interfaces, and more **tensor-network baselines** are a natural next step.

Take-home message

One simulation pipeline turns any Hamiltonian family into a **phase-diagram benchmark** that is checked against exact and classical references on everything from ideal simulators to real hardware.