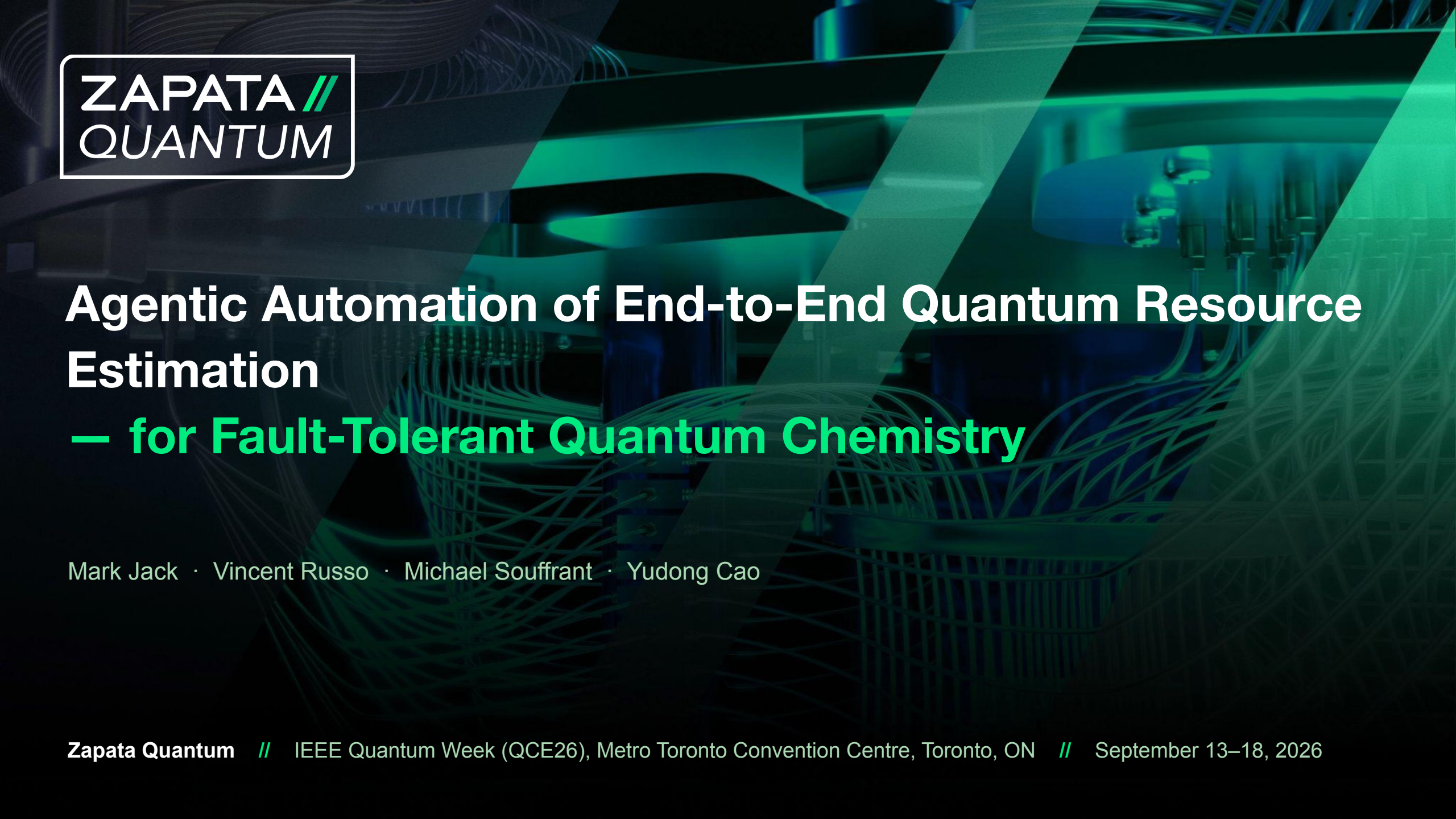


ZAPATA // QUANTUM

Agentic Automation of End-to-End Quantum Resource Estimation

— for Fault-Tolerant Quantum Chemistry

Mark Jack · Vincent Russo · Michael Souffrant · Yudong Cao

Zapata Quantum // IEEE Quantum Week (QCE26), Metro Toronto Convention Centre, Toronto, ON // September 13–18, 2026

1 Why quantum resource estimation matters

$2.23B

average R&D cost per drug, discovery to launch (2024)

10–15 yrs

typical timeline to an approved drug

90%

of drug candidates entering clinical trials fail

$400B

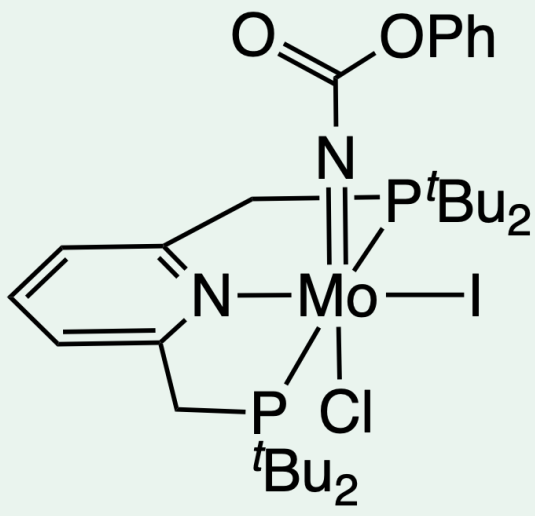
potential annual value of quantum computing to pharma by ~2035

- Molecular simulation to chemical accuracy — catalysts, drug–target binding — is the flagship quantum-computing use case.
- Chemicals & life sciences are projected among the first industries to capture value; quantum investment hit $12.6B in 2025.
- Every roadmap decision needs a price: qubits, runtime, error rates — on hardware that does not exist yet.
- **That pricing exercise is quantum resource estimation (QRE) — and today it is the bottleneck.**

Sources: Deloitte, Measuring the Return from Pharmaceutical Innovation (2025); Sun et al., Acta Pharm. Sin. B (2022); McKinsey Quantum Technology Monitor (2025–26).

2 The problem: one estimate is a months-to-years long study

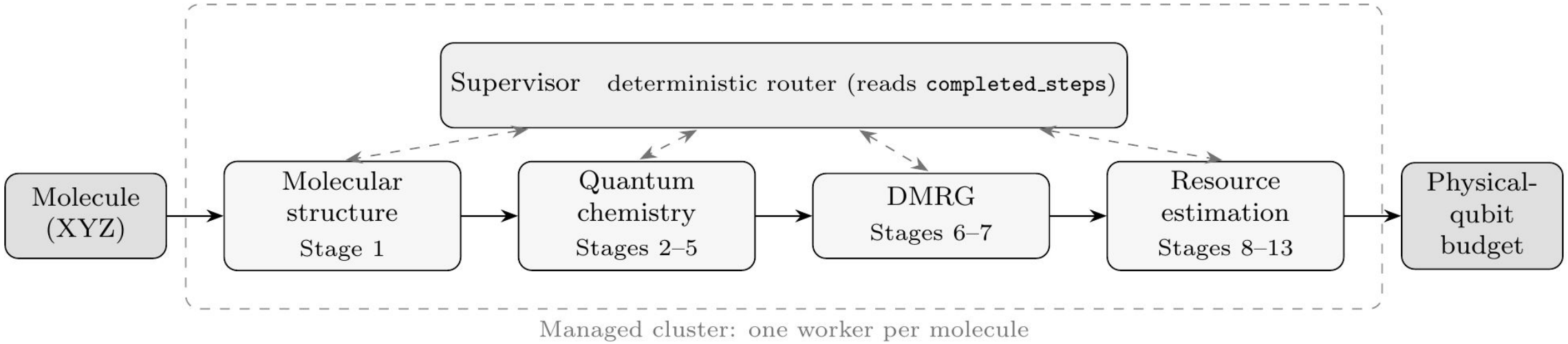
- **DARPA Quantum Benchmarking (QB):** the public reference — ten catalyst instances, ~3 years of analyst effort, ~4×10^5 CPU-hours of DMRG.
- The cost is **human coordination,** not computation — two kinds of toil:
  - **Operational —** environment failures, version mismatches, triaged re-runs.
  - **Algorithmic —** completed results that are physically wrong, and knowing the prescribed recourse of how to fix them.
- **Fixed, 13-stage, automated workflow:** molecule geometry → AVAS active space → Hamiltonian + double factorization + qubitization → DMRG energy + CSF overlap (logical qubits / T. gates) → QPE costing → physical qubits (surface code).
- **Every new molecule, method, or hardware target = a fresh manual study.**



2

Instance 2 — Mo-pincer catalyst, DARPA QB suite (31 orbitals, 44 electrons)

3 System architecture: a supervisor and four agents



- **Deterministic supervisor** — Python function, not an LLM: routes on typed state (completed_steps); auditable, cheap.
- Four **ReAct sub-agents** own contiguous stage blocks; a stronger LLM per agent, a lighter LLM per tool call.
- MCP servers wrap **PySCF, OpenFermion, Block2, QuantumMAMBO.jl,** and a **BenchQ**-style Stage-13 estimator.
- **Extractors** build the typed state; **validators** enforce domain checks (Pauli exclusion, energy sanity, CSF |γ| gate, odd code distance).
- Reported runs are **fully autonomous; complete auto-scaling** and **re-sizing of compute resources** enabled (CPU type, memory, threads etc.).
- Optional **human-in-the-loop** after Stage 1.

4 Automated workflow: absorbing both kinds of toil

Operational — repaired in place

- Infrastructure breaks identified as tool errors, bad dependency pins, coarse SCF tolerances.
- Surface as tool-error envelopes or validator violations.
- Sub-agent self-repairs: corrected parameters (e.g. pyscf_conv_tol), alternate tool path, flagged version mismatch.

Algorithmic — routed to a prescription

- All checks pass, yet the result is out of spec: low CSF overlap $|\gamma| \rightarrow$ intractable QPE shots. A retry is useless.
- Codified recourse, in order: raise bond dimension $\rightarrow$ revise active space $\rightarrow$ non-QPE estimator or flagged shot count.
- Disagreements with published references get an accountable cause, never silence.

Deployment

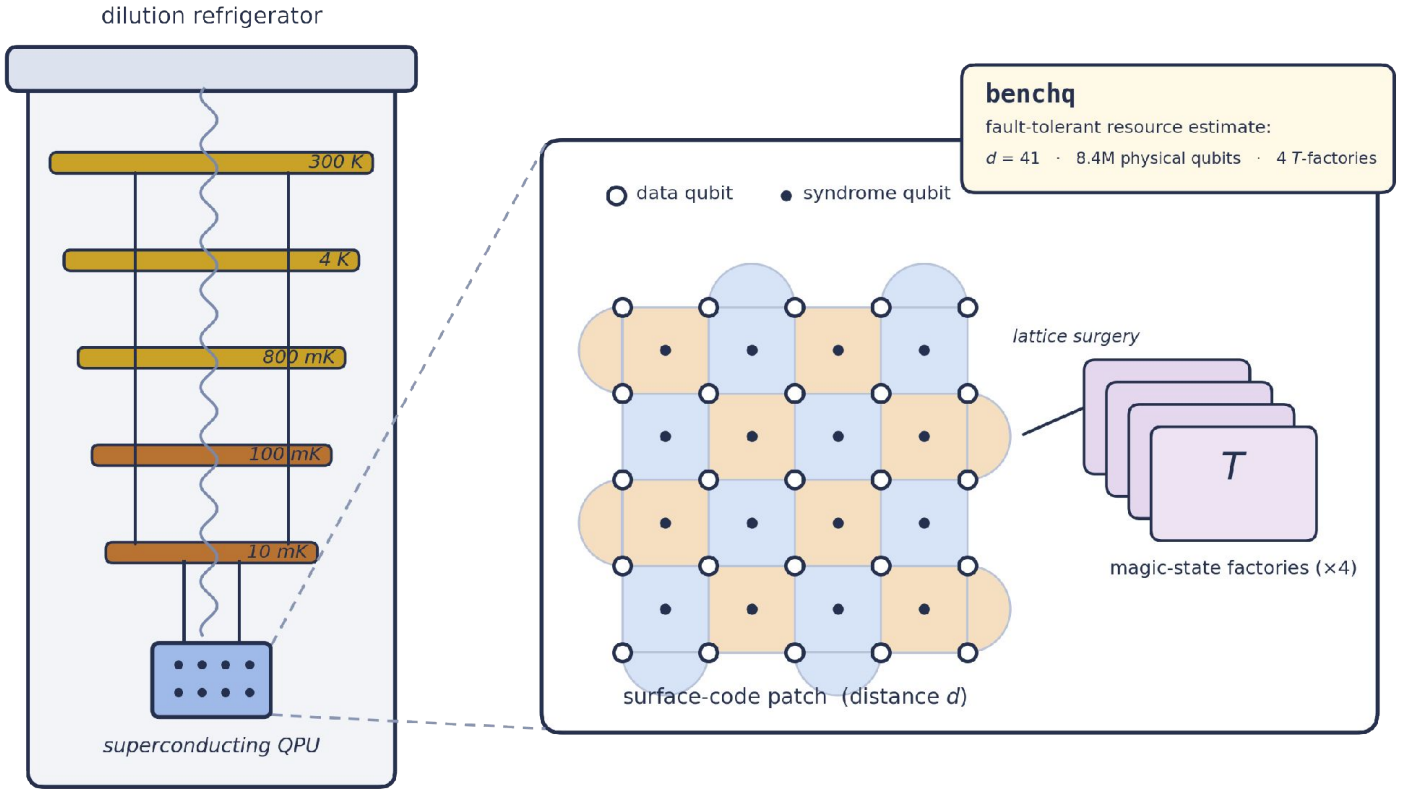
- Layered Docker container image (chemistry + Julia base, thin app layer), CI-built; batch jobs, one worker per molecule.
- GKE cluster auto-scales via LLM interface.
- Compute tier auto-derived from **active-space size dimension K** and **DMRG max bond dimension D**: compute time ~ K^3D^3; 16–80 vCPUs; up to 1200 GiB CPU memory/pod; up to 4100 GiB pod ephemeral storage.

5 Results: eight DARPA QB catalyst instances, end to end

Molecule	(orb, elec)	λ_tot	Toffoli	Logical	Physical	d	Wall (hr)	QPE (hr)
Single-reference instances (|γ| ≈ 0.92) — full 13-stage pipeline								
2	(31, 44)	218.3	2.50×10^{10}	994	8.41×10^{6}	41	15.23	967.0
PC	(31, 44)	216.5	1.32×10^{11}	1056	9.85×10^{6}	45	24.40	5115.9
RC	(31, 44)	210.4	1.25×10^{11}	1054	9.84×10^{6}	45	4.52	4849.5
$TS_{1/2}$	(31, 44)	216.8	1.34×10^{11}	1055	9.84×10^{6}	45	1.20	5175.9
Multi-reference instances (|γ| = 0.23–0.45) — gated at Stage 7, accepted shot-count penalty								
4a	(21, 31)	98.4	3.67×10^{10}	749	7.50×10^{6}	43	2.54	1423.3
PC^{-}	(49, 67)	410.4	2.43×10^{13}	2091	1.76×10^{7}	47	6.40	942944.7
I	(49, 67)	483.6	5.69×10^{11}	3171	2.51×10^{7}	47	8.02	22040.5
$TS_{1/4a}$	(49, 67)	462.3	4.04×10^{12}	3565	2.78×10^{7}	47	34.43	156471.0

- 1.2–34 hr per molecule, unattended, vs. a multi-year manual calendar; $TS_{1/2}$ (highlighted): 72 minutes on one node.
- Low CSF overlap |γ| in the lower block is genuine multi-reference character — shot counts inflated >10× are flagged, not silent.

6 What the budget buys: physical realization



- **$TS_{1/2}$ DMRG within 2 mHa of the DARPA reference (−2928.211 vs −2928.213 Ha).**
- Count differences trace to automated AVAS active spaces — flagged as an accountable choice.
- Suite: Toffoli gates: 2.5×10^{10}–2.4×10^{13}; logical qubits: 749–3565; physical qubits: 7.5–27.8 x 10^6 at code distance d = 41–47.
- Stage 8–13 closes in seconds; wall clock is integrals + the DMRG sweep.

Stage-13 estimate, instance 2: surface-code patches (d = 41), superconducting QPU, 8.4M physical qubits, four AutoCCZ magic-state factories.

7 Extensions in flight & where this sits

Shipped since submission ✓

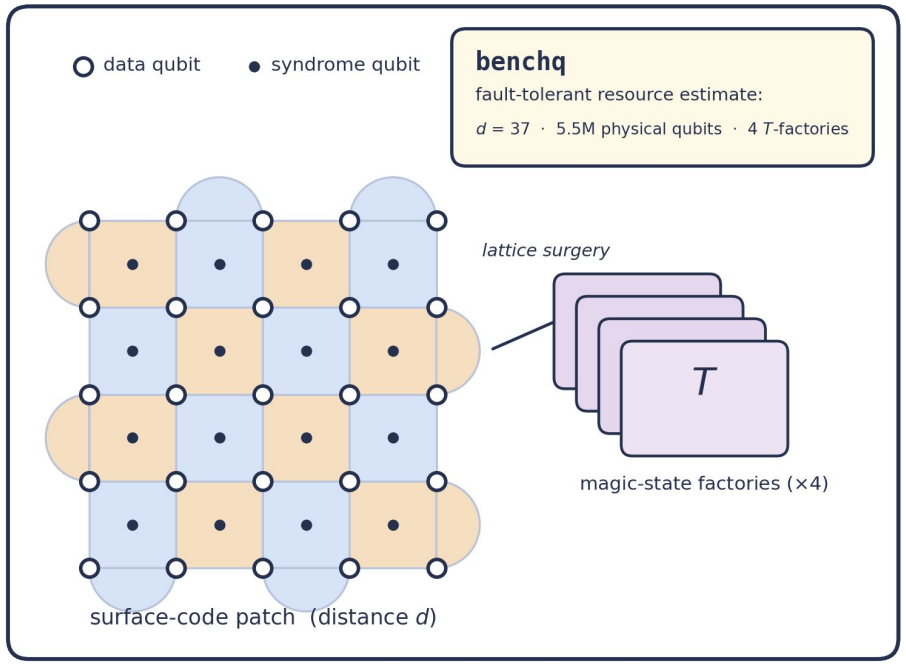
- Stage-as-a-service: per-stage API contracts + cloud results store (GCS, Supabase, queryable lookup API).
- Neutral-atom FTQC: pipeline extended to QuEra's surface-code architecture, cross-validated against Infleqtion's resource-superstaq as anchor.
- User interface: agentic geometry retrieval (PubChem / NIST CCCBDB / RDKit fallback) + metadata-inference (charge, spin, orbs).

In progress / in review

- GPU-accelerated DMRG via NVIDIA cuQuantum; GPU baselines on RunPod complete.
- Erasure-conversion error model: loss-centric Libra baseline, selectable erasure profiles, BenchQ cross-check (surface code).

Ready (queued)

- Bond-dimension / active-space escalation for the four multi-reference molecules (|γ| = 0.23–0.45).
- Resolve $TS_{1/2}$ active-space discrepancy vs. the reference study (31 vs 27 orbitals).



Alternate Stage-13 configuration (BenchQ model, d = 37, 5.5M qubits): the estimator layer is swappable (Azure QRE, Qualtran, superstaq).

Project board (Aug 2026): 55 done · 1 in review · 1 in progress · 6 ready.

Summary

Key findings

- 13-stage DARPA QB workflow fully automated: geometry → physical-qubit budget, no human in the loop.
- Specialist-months → 1–34 compute-hours per molecule; $TS_{1/2}$ in 72 minutes.
- DMRG reproduces the reference to 2 mHa; deviations attributed, not hidden.
- Operational toil self-repaired; algorithmic toil routed to codified prescriptions.
- Since submission: full 10-instance suite benchmarked on GKE + neutral-atom extension shipped.

Remaining issues & outlook

- Four multi-reference molecules (low $|\gamma|$) queued for bond-dimension / active-space escalation; 168-atom 1_lut_ts DMRG still under-converged.
- $TS_{1/2}$ active-space discrepancy vs. the reference (31 vs 27 orbitals) under investigation.
- Throughput & realism next: cuQuantum GPU DMRG (in progress) and erasure-conversion error models (in review).
- Agents absorb toil, not science — novel judgment stays with the specialist.

Thank you — questions welcome. // *Mark Jack · Zapata Quantum*