APOLLO QUANTUM

What Does "Chemically Useful" Actually Mean?

Nicole Bellonzi
Apollo Quantum
Head of Strategy & Operations

IPAM Workshop - Bridging the Gap Between NISQ and FTQC
Tuesday, February 17th, 2026

Two different conversations about “progress”

- Chemistry conversations:

  Are the results accurate enough to guide decisions?

- Quantum computing conversations:

  How many logical qubits do you need? How deep is the circuit?

These are **not the same question**

Performance metrics are means, not ends

- Accuracy
- Runtime
- Qubits / resources
- Scaling

These are **proxies** for something else:

Does the computation change a scientific or industrial decision?

“Useful” depends on the decision context

- Different decisions require different accuracy levels
- Reaction barrier prediction ≠ qualitative screening
- “Chemical accuracy” is not a universal requirement
- Accuracy, uncertainty, and cost are always coupled

Benchmarks measure tasks, not decisions

Many benchmarks evaluate:

- algorithmic subroutines
- synthetic Hamiltonians
- hardware-native tasks

These do **not automatically map** to chemical workflow value

Benchmark results depend on the problem distribution

- Problem instances are **not neutral**
- Dataset construction influences:
  - which solvers appear strong
  - which approaches appear scalable
- "Performance" often reflects **instance selection**

Performance targets are useful, but incomplete

- Performance targets help:
  - Track engineering progress
  - Compare algorithmic approaches
  - Define scaling milestones
- But alone they **do not answer**:

  What new decisions become possible?

Case study:

Ground-state energy estimation

Ground-state energy estimation as a benchmark domain

- Central problem in computational chemistry
- Clear accuracy targets
- Used to compare classical and quantum solvers
- Often used as a proxy for “chemical usefulness”

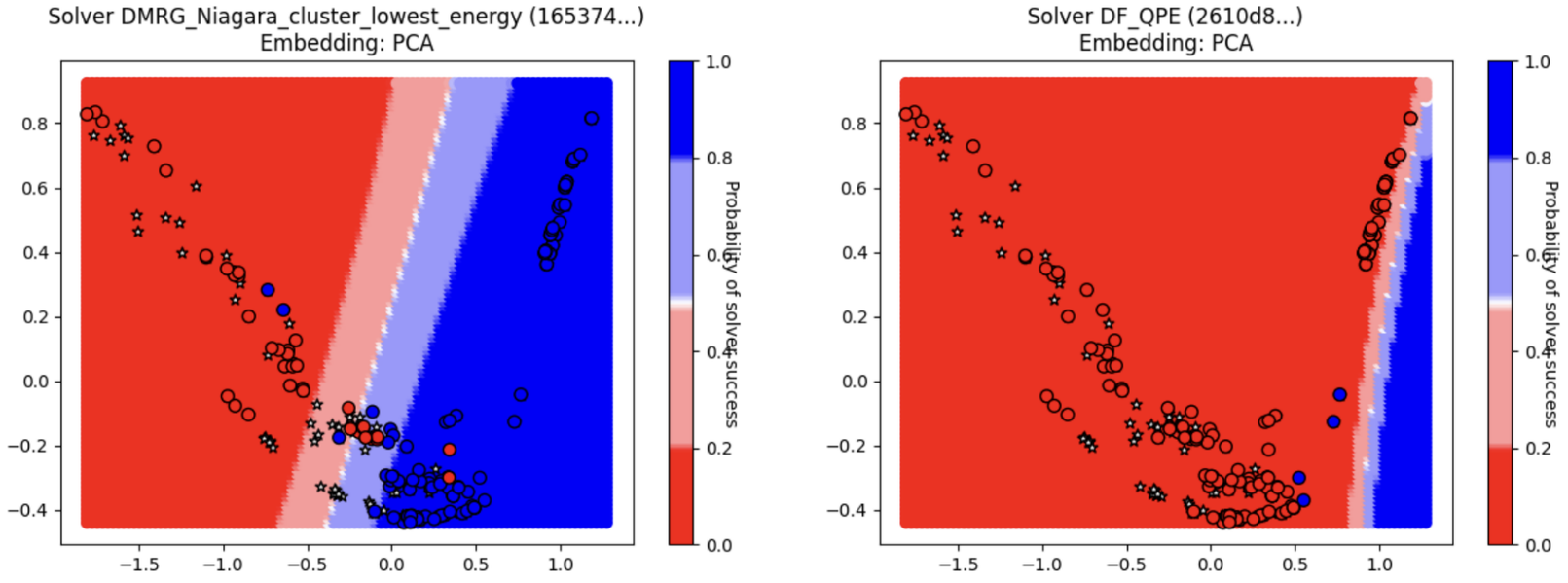
Modern benchmarking approaches

- Diverse Hamiltonian problem sets[1]
- Feature-based complexity characterization[2]
- Solver performance comparison[3]
- Solvability region analysis[4]

Representative References
[1] HamLib; Curtiss *et al.*, *J. Chem. Phys.* **106**, 1063 (1997); arXiv:2105.12767
[2] Berry *et al.*, *Commun. Math. Phys.* **356**, 1057 (2017); Motta *et al.*, *Nat. Phys.* **16**, 205 (2020)
[3] Holmes *et al.*, *JCTC* **12**, 3674 (2016); Reiher *et al.*, *PNAS* **114**, 7555 (2017)
[4] Bellonzi *et al.*, arXiv:2508.10873 (2025)

APOLLO QUANTUM

Solver performance depends on instance distribution



(a) First Run DMRG

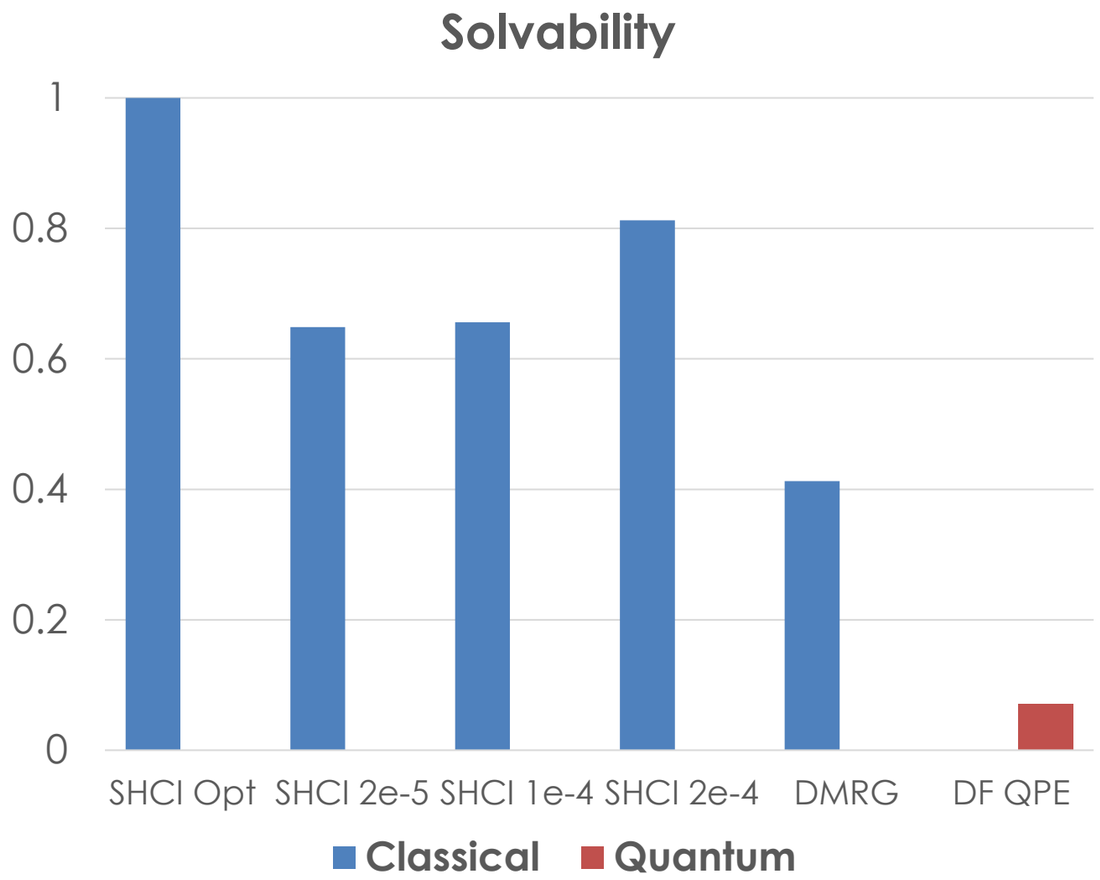
(b) DF QPE

Bellonzi *et al.*, arXiv:2508.10873 (2025)

- Solver success varies across problem space
- Feasibility depends on instance distribution
- Benchmark conclusions **are context-dependent**

Feasibility depends on assumptions and thresholds

- Classical solvers: broad solvability on current benchmark set
- Quantum DF-QPE: limited solvability under current runtime assumptions
- Interpretation depends on:
  - runtime thresholds
  - hardware assumptions
  - dataset composition



Bellonzi *et al.*, arXiv:2508.10873 (2025)

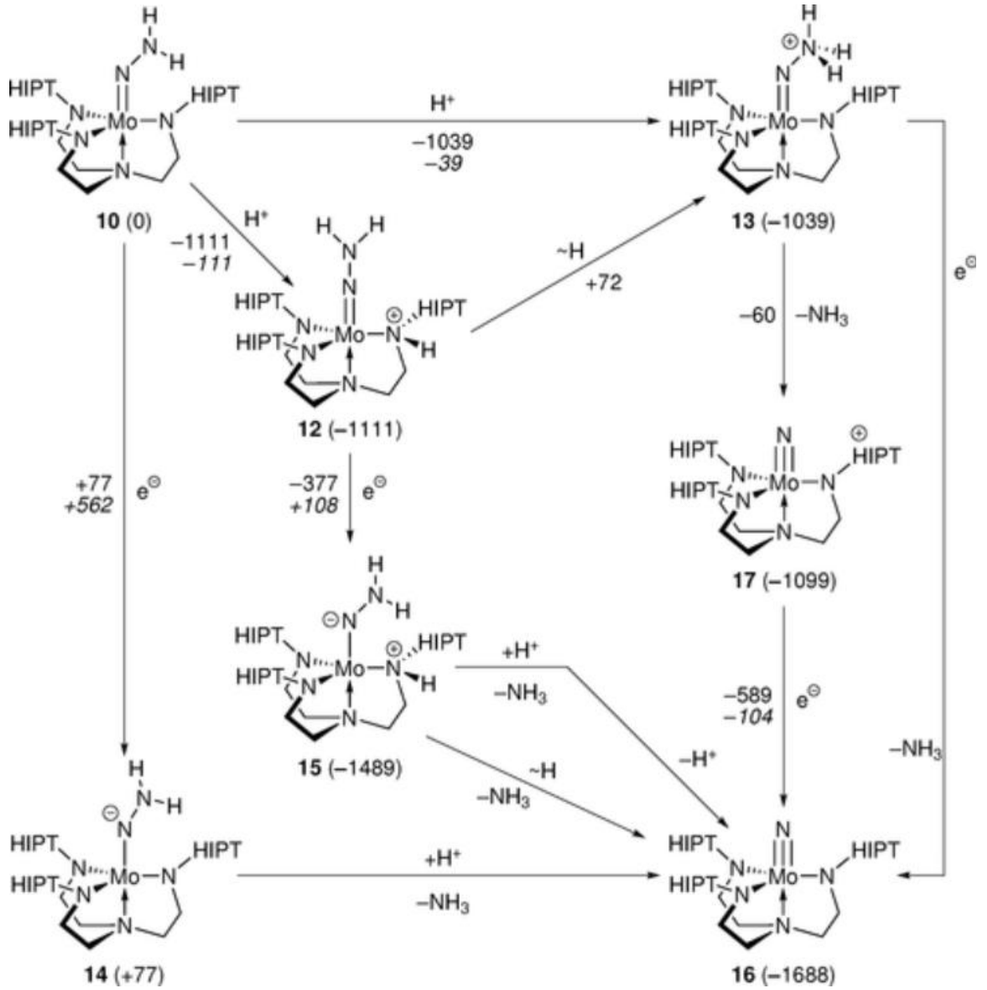
Case study:
Application-driven utility

When is a computation worth doing?

- Utility depends on:
  - Decision value
  - Accuracy required for the decision
  - Cost of computation
  - Availability of alternatives

Utility-driven modeling example

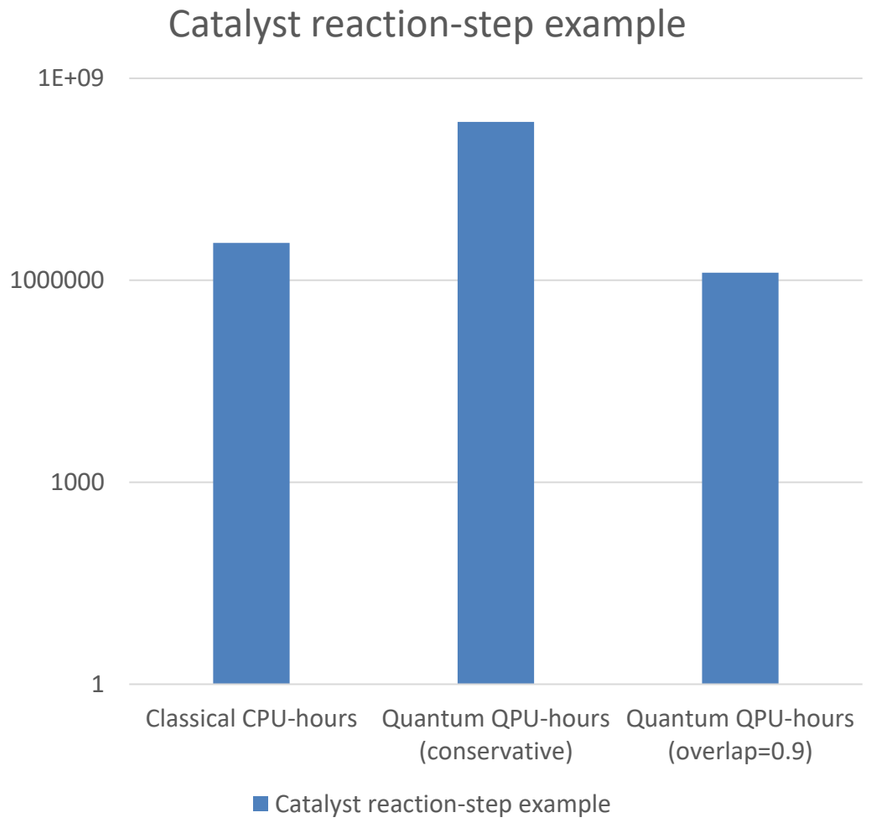
- Homogeneous catalyst discovery workflows
- Reaction energetics guide:
  - pathway selection
  - catalyst screening
  - experimental prioritization
- Some target calculations have **measurable economic value**



Schenk et al., Inorg. Chem. 47, 3634 (2008)

Expensive computations can still be useful

- Example: catalytic cycle steps with estimated economic utility (~$200k)
- Classical and quantum approaches both resource-intensive
- Key point:
  - Utility depends on **decision impact**, not only runtime
  - Feasibility thresholds shift when decision value is high



Bellonzi *et al.*, arXiv:2508.10873 (2025)

Many feasibility claims are assumption-sensitive

Feasibility depends on:

- initial state overlap assumptions
- error correction / hardware models
- active-space and model chemistry choices
- dataset composition
- workflow integration assumptions

Benchmarks should be read as conditional evidence

Benchmark results tell us:

- what is feasible **under specific assumptions**
- which improvements would change feasibility
- where conclusions are robust vs fragile

What does “chemically useful” mean?

- Chemical usefulness is **decision-dependent**
- Performance metrics are **proxies**, not endpoints
- Benchmark results must be interpreted in context:
  - Datasets
  - Assumptions
  - workflow impact
- Progress should be evaluated by:

Which decisions become newly possible?

Translating performance improvements into capability

- Identify application-relevant thresholds
- Connect benchmark milestones to decision capability
- Use decision-aware benchmarks to guide:
  - hardware targets
  - algorithm design
  - application selection


APOLLO
QUANTUM