Reduced Quantum Cost via Simultaneous Measurement

Transpilation and Circuit Synthesis

Pranav Gokhale Frederic T. Chong (University of Chicago)

Background

The originally envisioned quantum algorithms (Shor factoring, Grover search, etc.) required too many gates to comply with Noisy Intermediate-Scale Quantum (NISQ) hardware. To comply with NISQ constraints, a new paradigm of *variational* algorithms has emerged, for instance VQE (Variational Quantum Eigensolver, "killer app" for chemistry).

The catch: variational algorithms require **too many measurements**, e.g. $O(N^4)$. Naively, each measurement requires a fresh quantum execution. However, it was observed that certain measurements could be executed simultaneously.

We show that simultaneous measurement leads to a **30X+** reduction in measurement cost (in fact, an asymptotic reduction). Our work includes efficient transpilation, circuit synthesis, experimental validation, and statistical analysis.

Simultaneous Measurement

Quantum computing deals with N-character terms called Pauli strings, which have regex representation (I|X|Y|Z){N}, for example YIXIZ or ZXYZZ for N = 5. The I, X, Y, and Z strings represent the Pauli matrices in quantum mechanics.

Terms can be **measured simultaneously iff they commute**. For N = 1, the commutation rule is that I commutes with everything and (X, Y, Z) only commute with themselves.

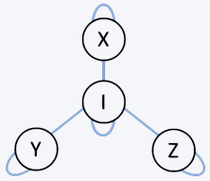

Figure 1: Commutation graph for N = 1.

For N > 1, the simplest rule is Qubit-Wise Commutativity (QWC): two terms QWC if they commute at each index. The more General Commutativity (GC) rule is that two terms GC if an even # of indices don't commute.

Graph Representation and Grouping

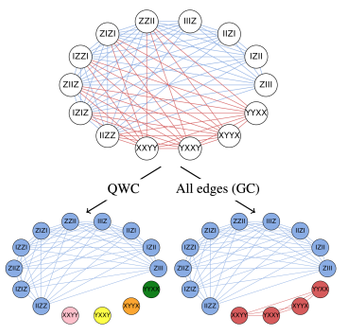

Figure 2: MIN-CLIQUE-COVER for H_2.

We seek to transpile VQE instances by grouping the terms into sets that can be simultaneously measured. In the commutation graph, we seek MIN-CLIQUE-COVER:

- Cliques, because all terms in a clique can be measured simultaneously.
- Cover, because we want to measure all of the terms.
- MIN, because we want to minimize total cost.

MIN-CLIQUE-COVER is NP-Hard, but we (a) use heuristics and (b) exploit underlying structure among the terms.

VQE Instance Transpilation

We transpile by grouping into min cliques using: Boppana-Halldórsson, Bron-Kerbosch, and OpenFermion. These achieve **30X+** lower measurement cost, but transpilation runtime is expensive, motivating us to study the structure of terms.

Linear-Time Transpilation

Molecular chem. terms have structure:
(1) Terms arise in 16-tuplets with MIN-CLIQUE-COVER of two.
(2) Terms arise in quadruplets that can be parallelized into disjoint schedules.

This structure yields N-sized cliques in linear time—reduces measurement cost from $O(N^4)$ to $O(N^3)$.

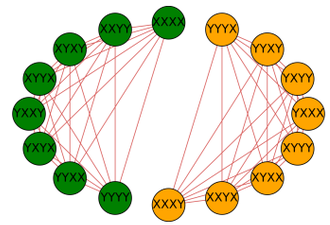

Figure 3: JW encoding yields 16 terms.

Results

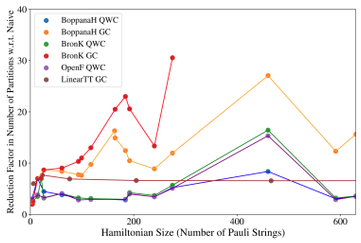

Figure 4: Measurement reduction (higher is better).

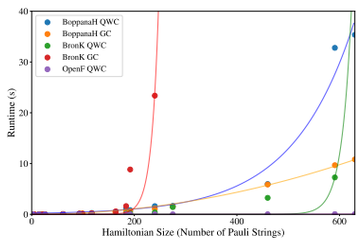

Figure 5: Transpilation runtime (lower is better).

Measurement Circuit Synthesis

Simultaneous measurement of QWC groups is easy. However, GC group measurement requires specialized circuits. We developed a circuit synthesis tool, based on the stabilizer formalism of quantum error correction. Critically, the synthesis is very fast, based on computations on 2N by N matrices.

Algorithm 2: Circuit synthesis for sim. measurement

input : $\{P_i\}$, a complete GC family of Pauli strings
output : Circuit for simultaneous measurement of $\{P_i\}$

$M \in F_2^{2N\times N} \leftarrow$ basis of $\{P_i\}$;
Full-rankify Z-matrix by applying H gates;
Gaussian eliminate X-matrix using CNOT & SWAP gates;
for *each diagonal element in Z-matrix* **do**
  if *element is 1* **then** apply S to corresponding qubit;
end
for *each element below diagonal of Z-matrix* **do**
  if *element is 1* **then** apply CZ to the row-col qubits;
end
Apply H to each qubit;
Measure each qubit;

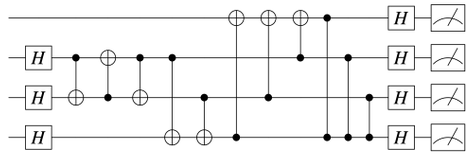

Figure 6: Simultaneous measurement circuit generated by our software for the green 8-clique in Figure 3.

Experimental Results

We experimentally validated our transpilation and circuit synthesis by executing on an IBM 20-qubit device. We measured the ground state energy of deuteron and found error of 835 keV when grouping for simultaneous measurement—11% lower than separate measurements.

These results will scale favorably for larger problem instances and better hardware, since we asymptotically improve reduction factor in measurement cost.

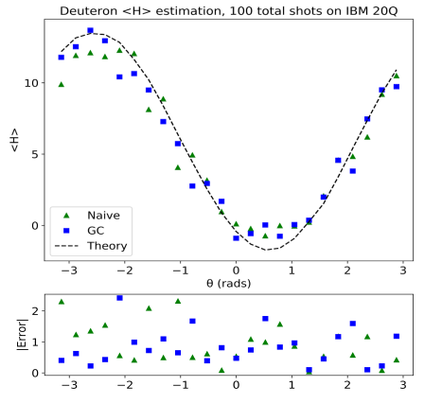

Figure 7: 11% lower average error with simultaneous measurement on IBM 20-qubit device with 100 shot budget.

Adaptive Covariance Mitigation

Simultaneous measurement introduces covariance terms. Under pathological conditions, these covariances can make simultaneous measurement worse (higher total variance) than separate measurement [McClean et al. 15]. We resolve this difficulty by showing:

(1) in expectation, simultaneous measurement is always better.

(2) the sample covariance matrices allow us to adaptively course-correct

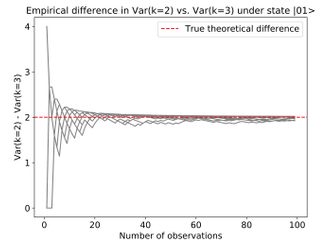

Figure 8: Sample covariances converge to their true value in just a few observations.

Future Work

Further experimental realizations of our methods will be promising, particularly as our results scale favorably. We also propose research towards optimization of simultaneous measurement circuits (e.g. gate cancellation).

Recently, others observed that our techniques could also apply to algorithms for machine learning and dynamics simulation [Sweke et al. 2019]. Understanding the commutativity structure in these algorithms would be fruitful.

We acknowledge helpful discussions and contributions from Olivia Angiuli, Yongshan Ding, Kaiwen Gui, Teague Tomesh, Martin Suchara, and Margaret Martonosi. For more information, see arXiv:1907.13623 and arXiv:1908.11857.