Measurement Grouping and Shot Budget

Variational Quantum Algorithms · Module 2 — Parameterized Circuits

Node 2.8

Position in Knowledge Graph

System: Variational Quantum Algorithms **Structural Domain:** Parameterized Circuits **Node:** 2.8

The naive measurement protocol from node 2.7 — measure each Pauli term in the Hamiltonian separately, with N shots each — is operationally feasible but **wasteful**. For a Hamiltonian with M Pauli terms, it costs $M \times N$ circuit runs per cost evaluation. For chemistry-style Hamiltonians where M can be in the hundreds or thousands, this is prohibitive.

The remedy is **measurement grouping**: identify Pauli terms that can be measured *simultaneously* in the same circuit run, group them, and amortize the shot cost across the group. Doing this well can reduce the per-evaluation cost by an order of magnitude.

This node explains the grouping problem, the standard solutions, and how a finite shot budget gets allocated across groups in practice.

When Two Pauli Strings Can Be Measured Simultaneously

Two Pauli strings can be **measured in the same experiment** if and only if they commute and there exists a single basis change that diagonalizes both. The simplest sufficient condition is **qubit-wise commutativity (QWC)**: two Pauli strings are QWC if at every qubit position, either:

- they have the same Pauli letter (XX, YY, ZZ, II), or
- at least one of them is the identity (XI, IY, ZI, II, etc.).

If two strings are QWC, then the per-qubit basis change for one is compatible with the per-qubit basis change for the other, and a single measurement in the combined basis gives you the bitstring needed to compute eigenvalues for **both**.

Example: $P_1 = X_0 Z_1$ and $P_2 = X_0 I_1$. These are QWC — both have X on qubit 0, and P_2 has identity on qubit 1. A single measurement (rotate qubit 0 to Z basis via H, leave qubit 1 alone, read out both) gives bitstrings from which you compute both $\widehat{\langle P_1 \rangle}$ (using both bits) and $\widehat{\langle P_2 \rangle}$ (using only bit 0).

QWC is the **simplest** notion of joint measurability. There are stronger notions:

- **General commutativity**: two Pauli strings commute as operators, but the diagonalizing basis change is not local (it requires entangling gates in the basis-change layer).
- **Anticommuting partitions**: with extra ancilla qubits, you can measure even mutually anticommuting sets of Paulis jointly.

The thesis's regularization techniques don't depend on grouping strategy specifically, but the shot budget allocations they assume are based on standard QWC grouping.

The Grouping Problem

Given a Hamiltonian $H = \sum_j h_j P_j$, partition the Pauli strings into the **minimum number of QWC groups**. Each group can be measured by a single circuit (with a single basis-change layer); within a group, every shot contributes to estimating every term in the group.

This is exactly the **graph coloring problem** on a "compatibility graph": vertices are Pauli strings, edges connect non-QWC pairs. A valid grouping is a coloring; the minimum grouping is the chromatic number.

Graph coloring is NP-hard in general, but for chemistry Hamiltonians:

- A greedy coloring runs in $O(M^2)$ and produces near-optimal groupings in practice.
- Specialized algorithms (Sorted Insertion, Largest-First Degree-Sorted) consistently reduce M from "thousands" to "tens" for typical chemistry Hamiltonians.

For H_2 (15 Pauli terms after Jordan-Wigner): grouping reduces this to **2 groups**, meaning a 7.5× reduction in circuit count per evaluation. For larger molecules like BeH_2: groupings of ~10-20 from raw counts of 200-500. For very large Hamiltonians (FeMoco, biomolecules): hundreds of groups from tens of thousands of terms.

The reduction is not free — within a group, terms with different supports compete for the same shots, and the variance allocation has to be done carefully — but it is a major operational win.

Variance Within a Group

Suppose a group G contains Pauli terms $P_{j_1}, P_{j_2}, \ldots, P_{j_K}$. A single measurement in the group's basis produces a bitstring from which you compute one eigenvalue per term:

$$\widehat{e}_k = \prod_{i \in \text{support}(P_{j_k})} (-1)^{b_i}, \quad k = 1, \ldots, K.$$

After N_G shots:

$$\widehat{\langle P_{j_k} \rangle} = \frac{1}{N_G} \sum_{s=1}^{N_G} \widehat{e}_k^{(s)}.$$

The crucial point: **all K estimators come from the same shots**. Their variances are still each bounded by $1/N_G$, but they are **correlated** with each other (because they share the underlying measurement outcomes).

The total variance of the cost contribution from the group is:

$$\text{Var}\left[\sum_{k=1}^{K} h_{j_k} \widehat{\langle P_{j_k} \rangle}\right] = \frac{1}{N_G} \text{Var}_{\text{shot}}\left[\sum_k h_{j_k} \widehat{e}_k\right],$$

where the inner variance is over the joint distribution of one shot's worth of measurements in the group basis.

For QWC groups, the inner variance can be substantially smaller than $\sum_k h_{j_k}^2$ (the bound for independent measurements) when the Pauli terms are correlated under the group basis. So measurement grouping does not just save circuit runs — it can also **reduce the per-shot variance contribution**.

Allocating a Shot Budget

Given a total shot budget N_{total} per cost evaluation and a set of G groups, how should you allocate shots across groups?

Two reasonable strategies:

Equal allocation. Give each group $N_G = N_{\text{total}}/G$. Simple, and works fine when the per-group variance contributions are roughly equal.

Variance-weighted allocation. Give each group a number of shots proportional to its variance contribution:

$$N_G = N_{\text{total}} \cdot \frac{\sigma_G}{\sum_{G'} \sigma_{G'}},$$

where σ_G is an estimate of the per-shot standard deviation of the group's contribution. This minimizes the total variance for a fixed total shot count and is the optimal allocation in the sense of Lagrange multipliers. It is sometimes called the "Wu et al." or "Rubin" allocation.

In practice, optimal variance-weighted allocation requires knowing variances in advance, which you don't. A common compromise: use a few hundred shots per group as a "scout" to estimate variances, then allocate the remaining budget proportionally.

The Shot Budget As An Optimization Hyperparameter

The total shot budget N_{total} per evaluation is **part of the optimization design**, not a fixed number imposed by hardware. Tradeoffs:

- **More shots** → less estimator noise → cleaner gradient signal → faster per-iteration progress.
- **Fewer shots** → cheaper evaluations → more iterations per wall-clock budget → more total parameter updates.

Most VQA optimizer choices make implicit assumptions about this tradeoff:

- **SPSA** (Simultaneous Perturbation Stochastic Approximation) tolerates very low shot counts and many iterations.
- **COBYLA** assumes moderate shot counts and uses geometry-based updates.
- **Adam with parameter-shift gradients** wants high enough shot counts that the gradient is reliable.
- **HOPSO** (the thesis's preferred optimizer) tolerates noisy evaluations and benefits from more iterations.

The thesis runs experiments at fixed $N_{\text{total}} = 1024$ shots per evaluation to enable apples-to-apples comparison across optimizers. Real-hardware deployments often use **shot scheduling** — start with low shots (cheap, noisy) for early exploration and ramp up as the optimizer converges (expensive, accurate near the minimum). See node 5.7 for the thesis-level treatment.

What Goes Wrong If You Get This Wrong

Two common failure modes:

1. Insufficient shots per term. With $N = 100$ shots and a Pauli term whose variance is near 1, the estimator standard deviation is ~0.1. For an optimizer trying to detect a gradient component of magnitude 0.05, this is not enough — the noise drowns the signal, and the optimizer wanders. Symptom: cost trajectory looks like a random walk rather than a descent.

2. Imbalanced shot allocation. If you give the 14-Pauli-term H_2 Hamiltonian 100 shots per term but the dominant term has $|h_j| = 10$ and the rest have $|h_j| < 0.1$, then the dominant term contributes $100\times$ more uncertainty to the cost estimate than the rest combined — you have wasted 90% of your shot budget on

terms whose noise contribution is negligible. Symptom: variance-weighted reallocation produces a 5-10× efficiency gain.

Both failure modes are extremely common in beginner VQA work and very rarely diagnosed correctly. Watching the per-term variance contribution to the cost estimate is the single most informative debugging signal.

Structural Role

This is the **operational closure** node of Module 2. Module 2 has now described, end to end, what happens when the optimizer requests a cost evaluation: ansatz design (2.1-2.3) → compilation (2.4) → encoding (2.5) → state representation (2.6) → measurement protocol (2.7) → shot budget and grouping (this node).

Forwards: Module 5 (regularization) uses the shot budget as the resource against which improvements are measured. Module 9 (hardware noise) revisits the variance contributions when realistic noise models are layered in.

Relations to Other Concepts

- **Graph coloring** — the formal problem behind QWC grouping.
- **Importance sampling** — variance-weighted shot allocation is its discrete analogue.
- **Classical shadows** — an alternative to grouping that estimates many observables from random measurements.
- **Adaptive measurement protocols** — schemes that change basis based on previous outcomes.

Worked Example — Grouping for H_2

After Jordan-Wigner, the H_2 Hamiltonian (4 qubits) has 15 Pauli terms.

Let's enumerate a few: - $I_0I_1I_2I_3$ (identity, contributes a constant) - Z_0, Z_1, Z_2, Z_3 (single-qubit Z's) - Z_0Z_1, Z_0Z_2, Z_0Z_3, Z_1Z_2, Z_1Z_3, Z_2Z_3 (ZZ pairs) - $X_0X_1Y_2Y_3$, $X_0Y_1Y_2X_3$, $Y_0X_1X_2Y_3$, $Y_0Y_1X_2X_3$ (XXYY-type four-body)

The 11 Z-only terms are all mutually QWC (they all only have Z's and I's on every qubit) — they form **one group**, measured by reading out all 4 qubits in the Z basis.

The 4 XXYY terms are also mutually QWC: each has the same Pauli letter at each position (after checking carefully — they aren't quite, actually). In practice, they form 1-2 additional groups depending on the exact strings.

Total grouping: **2-3 groups**. From 15 separate measurements down to 2-3 — a roughly 5-7× speedup.

For a 100-shot budget per group (a generous allocation), the total cost per evaluation is 200-300 circuit runs instead of 1500. **That difference is the difference between a workable VQA inner loop and an unworkable one.**

Visualization

Pending. A graph diagram. Vertices: 15 Pauli strings labeled. Edges: connect non-QWC pairs. The vertices are colored to show a 2-coloring (the QWC grouping). Below: a bar chart showing "without grouping: 15 measurements" vs "with grouping: 2 measurements".

Exercises

1. For the Pauli strings $\{X_0Z_1, X_0Y_1, Z_0Z_1, X_0X_1\}$, determine which pairs are QWC. Find the minimum QWC grouping.
2. Show that for a Hamiltonian made entirely of Z-string terms (all Pauli operators are Z or I), every term is QWC with every other term, so the grouping reduces to a single group.
3. (VQA) The greedy QWC grouping algorithm has worst-case approximation ratio bounded by the maximum degree of the conflict graph. Sketch why this matters for chemistry Hamiltonians vs spin-model Hamiltonians.

Research Extensions

- **General-commutativity grouping** — beyond QWC, allowing entangling basis-change layers.
- **Anticommuting set measurements** — using ancilla qubits to measure mutually anticommuting Paulis.
- **Fragment-based methods** — partition the Hamiltonian into structurally meaningful fragments (one-body, two-body, etc.) and measure each fragment optimally.
- **Online shot allocation** — reallocate shots adaptively as the optimization proceeds.

Cross-links to Other Courses

- **Graph Theory** — coloring, chromatic number.
- **Operations Research** — resource allocation under uncertainty.
- **Statistics Module 9** — experimental design, optimal allocation of measurements.
- **Module 5 (Regularization)** — uses the shot budget as the headline cost metric.

Variational Quantum Algorithms · Module 2 — Parameterized Circuits · Node 2.8
Concepts: expectation-value, variance-and-covariance, graph-theory, concentration-of-measure
Prerequisites: 2.5, 2.7

hbar.university — own.your.cognition