Quantum Noise Channels

Variational Quantum Algorithms · Module 9 — Hardware Noise Models

Node 9.1

Position in Knowledge Graph

System: Variational Quantum Algorithms **Structural Domain:** Hardware Noise Models **Node:** 9.1

This is the **opening node** of Module 9. Modules 1-8 have mostly treated VQAs as if they ran on ideal quantum computers: unitary evolution, exact measurements, no decoherence. That picture is useful for theory but wrong in every practical sense. **Real VQAs run on noisy hardware**, and noise changes the analysis of essentially every module:

- It shifts the cost landscape (Modules 5, 6).
- It creates new barren plateaus (Module 4.5).
- It modifies the expressivity-trainability tradeoff (Module 8.5).
- It introduces drift that adaptive controllers must reject (Module 7.6, Module 10).

Module 9 is the module that takes noise seriously. It builds up the noise framework from first principles — quantum channels, specific noise models, gate error budgets, mitigation techniques — and by the end of the module, you can read any experimental VQA paper and understand what noise assumptions are in play.

This node introduces the mathematical framework: **quantum channels**, the operator-sum (Kraus) representation, and the Stinespring dilation. Noise in VQAs is modeled as a quantum channel applied after each gate (or after the whole circuit), and everything downstream builds on this foundation.

What A Quantum Channel Is

A **quantum channel** $\mathcal{E}$ is a linear map from density matrices to density matrices:

$$\mathcal{E} : \rho \mapsto \mathcal{E}(\rho),$$

satisfying three properties:

1. **Linear:** $\mathcal{E}(a\rho_1 + b\rho_2) = a\mathcal{E}(\rho_1) + b\mathcal{E}(\rho_2)$.
2. **Trace-preserving:** $\mathrm{Tr}(\mathcal{E}(\rho)) = \mathrm{Tr}(\rho)$ for all ρ.
3. **Completely positive:** $(\mathcal{E} \otimes I_n)(\rho)$ is positive-semidefinite for all ρ and all $n \geq 0$.

Together these are called **CPTP maps** (completely positive trace-preserving). A quantum channel is the most general physically allowed transformation of a quantum state, and every noise process can be modeled as one.

Examples:

- **Unitary evolution:** $\mathcal{E}(\rho) = U\rho U^\dagger$. A noise-free channel.
- **Measurement + discard outcome:** $\mathcal{E}(\rho) = \sum_k P_k \rho P_k$ where $\{P_k\}$ are projectors. The post-measurement state with the outcome forgotten.
- **Depolarizing channel:** $\mathcal{E}(\rho) = (1-p)\rho + p \cdot I/d$. Section 9.2.
- **Amplitude damping:** models energy loss $|1\rangle \to |0\rangle$. Section 9.3.

The Kraus Representation

Every quantum channel can be written as

$$\mathcal{E}(\rho) = \sum_k K_k \rho K_k^\dagger, \quad \sum_k K_k^\dagger K_k = I,$$

where the $\{K_k\}$ are **Kraus operators** (or "operation elements"). The Kraus representation is not unique — a channel has many equivalent Kraus decompositions — but any one of them fully specifies the channel.

Properties:

1. **A channel with one Kraus operator is a unitary** (up to global phase): $K = U$, so $\mathcal{E}(\rho) = U\rho U^\dagger$.
2. **A channel with two Kraus operators is a "one-error" channel**: probability p of applying an error, probability $1-p$ of the identity. The depolarizing channel can be written with 4 Kraus operators (I, X, Y, Z).
3. **The number of Kraus operators is at most** d^2 for a d-dimensional system. So a single-qubit channel has at most 4 Kraus operators; a two-qubit channel has at most 16.
4. **The Kraus representation is a kind of "unraveling"** — it decomposes the channel into probabilistic branches, each branch being a distinct way the noise could act. This is the operational intuition.

Canonical Noise Channels

A handful of channels appear everywhere in quantum hardware modeling:

Bit flip channel

$$\mathcal{E}(\rho) = (1-p)\rho + p \cdot X\rho X.$$

Flips $|0\rangle \leftrightarrow |1\rangle$ with probability p. Classical bit error analogue.

Phase flip channel

$$\mathcal{E}(\rho) = (1-p)\rho + p \cdot Z\rho Z.$$

Flips the relative phase $|+\rangle \leftrightarrow |-\rangle$ with probability p. Purely quantum — no classical analogue.

Depolarizing channel

$$\mathcal{E}(\rho) = (1-p)\rho + p/3(X\rho X + Y\rho Y + Z\rho Z).$$

Applies a uniformly random Pauli error with probability p. The "worst-case" unbiased error model. Section 9.2.

Amplitude damping channel

$$\mathcal{E}(\rho) = E_0\rho E_0^\dagger + E_1\rho E_1^\dagger,$$

with $E_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix}, E_1 = \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix}$.

Models energy relaxation: $|1\rangle \to |0\rangle$ with probability γ. The dominant error in superconducting qubits. Section 9.3.

Dephasing channel

$$\mathcal{E}(\rho) = \begin{pmatrix} \rho_{00} & \sqrt{1-\lambda}\rho_{01} \\ \sqrt{1-\lambda}\rho_{10} & \rho_{11} \end{pmatrix}.$$

Destroys coherence between $|0\rangle$ and $|1\rangle$ without changing populations. Closely related to the phase flip but continuous rather than probabilistic.

These five channels cover 95% of the noise models you'll encounter in practice. Real hardware is some weighted combination of these, with parameters (p, γ, λ) determined by calibration.

Stinespring Dilation

A deeper view: every quantum channel can be written as a unitary on a larger system.

Theorem (Stinespring 1955). For any CPTP channel $\mathcal{E}$ on a d-dimensional system, there exists an ancilla system of dimension at most d^2, a pure initial state $|0\rangle_{\text{anc}}$, and a unitary U on the joint system such that

$$\mathcal{E}(\rho) = \text{Tr}_{\text{anc}} \left[U(\rho \otimes |0\rangle\langle 0|_{\text{anc}}) U^\dagger \right].$$

Interpretation. Every noise channel is equivalent to: add an ancilla in a known state, apply a joint unitary, then trace out (i.e., forget) the ancilla. The noise comes from the information lost when you discard the ancilla.

Why this matters. It means noise is not fundamentally mysterious — it is always "unitary evolution with part of the system hidden." This is important because:

1. It explains why quantum error correction works: if you can track the ancilla, the noise is reversible.
2. It unifies different noise models: they are all different choices of ancilla and joint unitary.
3. It connects to the fluctuation-dissipation theorem of statistical mechanics.

The Pauli Transfer Matrix

A useful computational tool: the **Pauli transfer matrix (PTM)** of a channel.

For a single-qubit channel, write $\rho = (I + r_x X + r_y Y + r_z Z)/2$ (Bloch vector representation). The channel acts linearly on $\mathbf{r} = (r_x, r_y, r_z)$:

$$\mathbf{r}' = T\mathbf{r} + \mathbf{t},$$

where T is a 3×3 matrix and $\mathbf{t}$ is a translation vector. Together $(T, \mathbf{t})$ form the **affine map** representation of the channel.

For the standard channels:

- **Identity:** $T = I$, $\mathbf{t} = 0$.
- **Bit flip with prob p:** $T = \text{diag}(1, 1-2p, 1-2p)$, $\mathbf{t} = 0$.
- **Depolarizing with prob p:** $T = (1-p)I$, $\mathbf{t} = 0$.
- **Amplitude damping with prob γ:** $T = \text{diag}(\sqrt{1-\gamma}, \sqrt{1-\gamma}, 1-\gamma)$, $\mathbf{t} = (0, 0, \gamma)$.

The PTM is convenient for computations — composition of channels is just matrix multiplication of PTMs, and expectation values of Pauli observables can be read off directly.

Composing Channels

Real circuits have many noise events. How do they compose?

Sequential composition. If you apply channels $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_L$ in sequence, the total channel is $\mathcal{E}_{\text{total}} = \mathcal{E}_L \circ \cdots \circ \mathcal{E}_2 \circ \mathcal{E}_1$.

In Kraus form: if $\mathcal{E}_i(\rho) = \sum_k K_{i,k} \rho K_{i,k}^\dagger$, then $\mathcal{E}_{\text{total}}$ has Kraus operators $K_{L,k_L} \cdots K_{2,k_2} K_{1,k_1}$ over all index sequences $(k_1, \ldots, k_L)$. The number of Kraus operators can grow exponentially — for an L-gate circuit with 2-Kraus-operator channels, up to 2^L terms.

In PTM form: $T_{\text{total}} = T_L \cdots T_2 T_1$. A single matrix product, regardless of circuit depth. This is much cheaper computationally.

For a depth-L circuit with uniform depolarizing channel $\mathcal{E}_p$ at each gate, the composed channel is $\mathcal{E}_{\text{total}}(\rho) = (1-p)^L \rho + (1-(1-p)^L) \cdot I/d$. The **fidelity decays exponentially in** L — this is the root of all hardware-imposed depth limits.

Noise Rate Vs Fidelity

Two common ways of parameterizing noise:

1. Error probability p**.** The probability that a gate application produces an error. Used for discrete noise models (Pauli channels).

2. Process fidelity F**.** The fidelity of the noisy channel to the ideal channel. Related to p by (for depolarizing) $F = 1 - p(d^2-1)/d^2 \approx 1 - p$ for small p.

3. Average gate fidelity $\bar{F}$**.** The fidelity averaged over input states (Haar average). Related to process fidelity by $\bar{F} = (dF+1)/(d+1)$.

For practical purposes, **gate fidelity** is the number you'll see on hardware spec sheets: "single-qubit fidelity 99.9%, two-qubit fidelity 99.5%". The corresponding error probabilities are 10^{-3} and 5×10^{-3} respectively.

In 2026, superconducting qubits typically achieve $\sim 99.9\%$ single-qubit and $\sim 99.5 - 99.8\%$ two-qubit fidelity. Trapped ions reach $\sim 99.99\%$ single-qubit and $\sim 99.9\%$ two-qubit. Neutral atoms are somewhere in between.

For a VQA circuit with hundreds of gates, these small per-gate errors compound: a 500-gate circuit at 99.5% two-qubit fidelity has total circuit fidelity $0.995^{500} \approx 0.08$ — only 8% of shots give correct results, 92% are pure noise.

Noise In The Heisenberg Picture

An alternative and often cleaner view: push the noise to the **observables**, not the states.

In the Heisenberg picture, a channel $\mathcal{E}$ acts on observables via its **adjoint** $\mathcal{E}^\dagger$:

$$\text{Tr}[O\mathcal{E}(\rho)] = \text{Tr}[\mathcal{E}^\dagger(O)\rho].$$

Implication for VQAs: instead of computing $\mathcal{E}(\rho)$ and then measuring O, you can compute $\mathcal{E}^\dagger(O)$ once and measure it against the noiseless state.

For depolarizing noise: $\mathcal{E}_p^\dagger(O) = (1-p)O + p \cdot \text{Tr}(O)/d$. The observable's coefficient shrinks by $(1-p)$ while the identity component is unaffected. For Pauli observables (traceless), this is just $O \to (1-p)O$ — the expectation value is scaled down by $(1-p)$.

Compound effect: after L noisy gates with depolarizing rate p, the effective observable is $(1-p)^L O$. Measuring this is the same as measuring O on the noiseless state but with the result scaled by $(1-p)^L$ — which is exactly the expected fidelity decay.

This Heisenberg-picture view is the foundation of **zero-noise extrapolation** (node 9.5) and other mitigation techniques — you're essentially rescaling the measurement to compensate for the noise's effect on observables.

How VQAs See Noise

When running a VQA on noisy hardware, noise affects several things:

1. The cost function becomes biased. $C_{\text{noisy}}(\boldsymbol{\theta}) = \text{Tr}[H\mathcal{E}(|\psi(\boldsymbol{\theta})\rangle\langle\psi(\boldsymbol{\theta})|)] \neq C_{\text{ideal}}(\boldsymbol{\theta})$. The optimizer is optimizing the wrong function.

2. The bias depends on $\boldsymbol{\theta}$. Different parameter values give different noisy costs, not all shifted by the same amount. So the optimizer can be actively misled.

3. The gradient becomes biased. $\nabla C_{\text{noisy}} \neq \nabla C_{\text{ideal}}$. Parameter-shift gradients are still *exact for the noisy cost*, but that's not the cost you want.

4. The landscape changes shape. Local minima, saddle points, and plateaus can all shift under noise. Module 4.5 covered the noise-induced plateau as an extreme example.

5. Shot noise adds on top. Even the noisy cost has to be estimated from finite shots, so there's an additional stochastic layer.

Practical upshot: the optimizer trained on a noisy VQA converges to the wrong minimum, and the wrongness grows with circuit depth and noise rate. This is the core practical problem that Modules 9-10 address.

Structural Role

This node is the **mathematical foundation** of Module 9. It introduces quantum channels, the Kraus and Stinespring representations, the Pauli transfer matrix, the composition rules, and the impact of noise on VQA cost functions.

It is the prerequisite for every other node in Module 9 — subsequent nodes (9.2 depolarizing, 9.3 amplitude damping, 9.4 mitigation, 9.5 ZNE, etc.) all build on the quantum-channel framework established here.

Relations to Other Concepts

- **Kraus (1971)** — the operator-sum representation.
- **Stinespring (1955)** — the dilation theorem.
- **Nielsen-Chuang** — *Quantum Computation and Quantum Information*, canonical reference.
- **Module 4.5 (Noise-induced plateaus)** — the qualitative consequence of these models.
- **Module 7.6 (Feedback control)** — treating noise as a disturbance to be rejected.

Worked Example — Depolarizing Channel On A 1-Qubit State

Take $\rho = |+\rangle\langle+| = (I+X)/2$ and apply the single-qubit depolarizing channel at rate $p = 0.1$:

$$\mathcal{E}_{0.1}(\rho) = 0.9 \cdot \rho + 0.1 \cdot I/2 = 0.9 \cdot (I+X)/2 + 0.1 \cdot I/2 = (I + 0.9X)/2.$$

The Bloch vector has shrunk from $(1, 0, 0)$ to $(0.9, 0, 0)$ — the state has moved toward the maximally mixed state.

Expectation value of X: $\mathrm{Tr}[X\rho] = 1$ (ideal) vs $\mathrm{Tr}[X\mathcal{E}(\rho)] = 0.9$ (noisy). The noise has decreased the expectation by a factor of $0.9 = 1 - p$.

Effect on a VQA cost function. If the cost is $C = \langle X \rangle$, the VQA sees $C_{\text{noisy}} = 0.9C_{\text{ideal}}$. The optimizer still finds the right minimum (scaling doesn't change the location of extrema), just with a weaker signal. Signal-to-noise ratio is decreased, but convergence location is preserved.

Multi-layer circuit. Apply L layers of depolarizing noise. The expectation becomes $0.9^L\langle X \rangle$. For $L = 20$, this is $0.9^{20} \approx 0.12$ — only 12% of the signal remains. Below some threshold, the signal is indistinguishable from shot noise and the optimizer stalls.

This is the quantitative mechanism behind the depth limit on noisy hardware.

Visualization

Pending. A Bloch sphere showing a pure state and its trajectory under depolarizing noise as p increases from 0 to 1. The Bloch vector shrinks toward the center (maximally mixed). Caption: "Noise contracts the Bloch sphere. Everything collapses to $I/2$."

Exercises

1. Write down the Kraus operators for a bit flip channel with probability p. Verify they satisfy $\sum K^\dagger K = I$.
2. For a two-qubit system with single-qubit depolarizing noise (rate p) on each qubit independently, write down the joint channel as a Kraus decomposition. How many Kraus operators does it have?
3. (VQA) A VQA circuit has 200 gates with average two-qubit error $p = 0.005$. Estimate the total circuit fidelity. Is this sufficient for a useful result?

Research Extensions

- **Non-Markovian noise models** — correlations between errors that violate the CPTP assumption.
- **Learned noise models** — data-driven noise characterization from hardware calibration.
- **Structured noise for VQA design** — ansätze designed to exploit specific noise channels.
- **Coherent noise** — over-rotations and miscalibrations that don't fit the standard Pauli framework.

Cross-links to Other Courses

- **Open Quantum Systems** — Lindblad equation, master equations, decoherence theory.
- **Quantum Error Correction** — stabilizer codes, syndromes, logical operations.
- **Module 4.5 (Noise-induced plateaus)** — the qualitative effect on trainability.
- **Module 9.2 (Depolarizing Noise)** — the canonical example.

Variational Quantum Algorithms · Module 9 — Hardware Noise Models · Node 9.1
Concepts: quantum-channels, density-matrix, decoherence, linear-operators
Prerequisites: 1.1
Enables: 9.2, 9.3, 9.4, 9.8

hbar.university — own.your.cognition