Spectral Error Mitigation: Exact Noise Inversion for Quantum Computers via Cluster Lindblad

No extrapolation. No sampling overhead. Exact under Markov.

Tamás Nagy, Ph.D.

tnagyphd@gmail.com

Working Paper

Abstract

Current quantum error mitigation methods are circuit-intensive. Zero-noise extrapolation (ZNE) runs the same computation 3–5 times at amplified noise levels, fits a polynomial, and extrapolates to zero noise. This fails systematically: decoherence is exponential, not polynomial, and qubit noise rates are heterogeneous across the chip.

We take a direct approach: build the noise channel $\mathcal{N}(t) = e^{\mathcal{M}t}$ from the Lindblad superoperator, and invert it. The inverse always exists for finite circuit times (Theorem 1), and we can bound its condition number before running the experiment (Proposition 1). The key to scalability is a cluster decomposition: each qubit's noise is captured by a 64×64 matrix (3-qubit local cluster, constant size), giving $O(n)$ cost regardless of system size. The cluster approximation matches exact 4^n diagonalization to $< 0.001\%$ up to $n = 6$ (14,327× speedup), consistent with the Kastoryano–Eisert exponential clustering theorem.

On a simulated Bell-state circuit with realistic transmon parameters ($T_1 = 300\,\mu$s, $T_2 = 200\,\mu$s), the gate-aware variant restores fidelity from 0.838 to 1.000 at depth 100, with $\kappa_{\text{sv}} = 1.46$. On a 5-qubit chain with IBM Brisbane-style heterogeneous calibration (T_1 range 100–350 μs), per-qubit correction eliminates a ~ 0.19 fidelity residual that chip-average methods leave unresolved. One circuit. One matrix inverse. Exact under Markov.

1. Introduction

1.1 The Problem

Today's quantum computers are noisy, and error *correction* (surface codes) is still decades away — it needs thousands of physical qubits per logical qubit (Cai et al., 2023). What we have instead is error *mitigation*: classical post-processing to undo noise without extra qubits (Endo et al., 2018; Kandala et al., 2019; Kim et al., 2023).

The standard tool is **zero-noise extrapolation (ZNE)** (Li and Benjamin, 2017; Temme et al., 2017): run the circuit at noise levels λ, 2λ, 3λ, fit a polynomial, extrapolate to $\lambda = 0$. ZNE powered the 127-qubit utility demonstration (Kim et al., 2023). But it has three problems:

1. **High circuit overhead**: 3–5 circuit executions for one result.
2. **Wrong model**: polynomial extrapolation for exponential noise ($e^{-\gamma\lambda t}$).
3. **Blind to heterogeneity**: treats all qubits the same. On real chips, T_1 and T_2 vary by $\pm 50\%$ (Kandala et al., 2019).

1.2 This Paper

Instead of extrapolating, we *invert*:

$$\rho_{\text{corrected}} = \mathcal{N}^{-1}\,\rho_{\text{measured}}, \qquad \mathcal{N} = e^{\mathcal{M}\,t_{\text{circuit}}} \tag{1}$$

where $\mathcal{M}$ is the Lindblad superoperator built from known decoherence parameters (T_1, T_2, gate times). We know the noise model; we just run it backwards.

The catch is scalability: full Lindblad inversion costs $O(4^{3n})$, which is intractable beyond a few qubits. Our solution is a **cluster decomposition** — each qubit's noise is captured by a $4^k \times 4^k$ matrix with $k = 3$ nearest neighbors (always 64×64, constant cost). This matches exact 4^n diagonalization to $< 0.01\%$ for $n \leq 6$, and reduces the total cost to $O(n \cdot 4^{3k})$. The 127-qubit circuits of Kim et al. (2023) would need 127 independent 64×64 inversions — about 4 MB, under 3 seconds on a laptop.

Three things distinguish this from prior work. The inverse always exists (Theorem 1), and its condition number is computable *before* the experiment (Proposition 1) — unlike ZNE, where the extrapolation error is unknown until measurement. A gate-aware variant (§4.2) achieves exact correction at arbitrary depth by inverting each gate's noise individually. And compared to the closest competitor — sparse Pauli-Lindblad PEC (van den Berg et al., 2023), used in the 127-qubit utility demonstration — our method is deterministic with zero sampling overhead, versus PEC's $\sim e^{2\epsilon}$ probabilistic cost.

1.3 Key Results

	ZNE (standard)	Spectral (ours)
Circuits needed	3–5	**1**
Noise model	Polynomial fit	**Exact (Lindblad)**
Per-qubit	No	**Yes**
Gate-time dependence	Approximate	**Exact**
Depth scaling	Degrades for deep circuits	$F > 0.996$ **for** $d \leq 10$**; exact with gate-aware variant**
Theoretical guarantee	None	**Lindblad invertibility (Theorem 1)**

---

2. The Spectral Error Mitigation Pipeline

2.1 Build Cluster Lindblad (One-Time)

For each qubit i, take its k nearest neighbors and build the local Lindblad superoperator:

$$\mathcal{M}_i = -i(I \otimes H_i - H_i^T \otimes I) + \sum_j \gamma_j \left(L_j^* \otimes L_j - \tfrac{1}{2} I \otimes L_j^\dagger L_j - \tfrac{1}{2}(L_j^\dagger L_j)^T \otimes I \right) \tag{2}$$

where H_i is the local Hamiltonian (qubit frequencies + coupling) and L_j are the jump operators (decay σ_-, excitation σ_+, dephasing σ_z) with rates from measured $T_1^{(i)}$, $T_2^{(i)}$.

Cost: One $4^k \times 4^k$ matrix exponential per qubit. For $k = 3$: a single 64×64 expm, under 1 ms. Total cost scales as $O(n \cdot 4^{3k})$ — linear in system size.

2.2 Compute Noise Channel

Given the total decoherence time t (gate times + idle times):

$$\mathcal{N}_i(t) = e^{\mathcal{M}_i \cdot t} \tag{3}$$

One 64×64 matrix exponential per qubit. Sub-millisecond.

2.3 Invert

$$\mathcal{N}_i^{-1}(t) = e^{-\mathcal{M}_i \cdot t} \tag{4}$$

Theorem 1 (Lindblad Invertibility). *Let $\mathcal{M}$ be a Lindblad superoperator. For any finite $t > 0$, the noise channel $\mathcal{N}(t) = e^{\mathcal{M}t}$ is invertible, with inverse $\mathcal{N}^{-1}(t) = e^{-\mathcal{M}t}$.*

Proof. $\det(e^{\mathcal{M}t}) = e^{\mathrm{tr}(\mathcal{M})\,t}$. Since $\mathrm{tr}(\mathcal{M})$ is finite, $\det(\mathcal{N}(t)) \neq 0$ for all finite t, so $\mathcal{N}(t)$ is invertible. The explicit inverse is $\mathcal{N}^{-1}(t) = e^{-\mathcal{M}t}$, since $e^{\mathcal{M}t} \cdot e^{-\mathcal{M}t} = I$. □

Proposition 1 (Condition Number Bound). *The spectral condition number of $\mathcal{N}(t)$ (ratio of largest to smallest eigenvalue modulus) satisfies*

$$\kappa_{\mathrm{spec}}(\mathcal{N}(t)) = e^{|\mathrm{Re}(\lambda_{\min})| \cdot t} \tag{4a}$$

where $\lambda_{\min}$ is the eigenvalue of $\mathcal{M}$ with the most negative real part. For a multi-qubit system, $|Re(\lambda_{\min})|$ depends on all dissipative channels (decay, dephasing, and their combinations).

Remark. Lindblad generators are generically non-normal ($\mathcal{M}^\dagger\mathcal{M} \neq \mathcal{M}\mathcal{M}^\dagger$), so the SVD condition number $\kappa_{\mathrm{sv}} = \sigma_{\max}/\sigma_{\min}$ can exceed κ_{spec}. For any eigenvector $Ax = \mu_k x$, the bound $\|Ax\| \geq \sigma_{\min}\|x\|$ gives $|\mu_k| \geq \sigma_{\min}$; similarly $|\mu_k| \leq \sigma_{\max}$. Hence $\kappa_{\mathrm{sv}} \geq \kappa_{\mathrm{spec}}$ always holds. For the 2-qubit system, the gap grows with t: at depth 100 ($t = 33\,\mu$s), $\kappa_{\mathrm{spec}} = 1.39$ vs $\kappa_{\mathrm{sv}} = 1.46$ (5% gap); at $t = T_1$ ($300\,\mu$s), $\kappa_{\mathrm{spec}} = 20$ vs $\kappa_{\mathrm{sv}} = 29$ (44% gap). The gap arises because $\sigma_{\max}(\mathcal{N}(t)) > 1$ for non-normal $\mathcal{M}$ (pseudospectral transient growth: some Liouville-space directions are temporarily amplified before eventual decay).

For the 2-qubit system in this paper ($T_1 = 300\,\mu$s, $T_2 = 200\,\mu$s), $|\mathrm{Re}(\lambda_{\min})| = 10{,}000\,\mathrm{s}^{-1}$ (from the combined decay + dephasing channels). Numerically verified κ_{sv} values:

t	t/T_1	κ_{spec}	κ_{sv}
$1.3\,\mu$s (depth 1)	0.004	1.01	1.02
$33\,\mu$s (depth 100)	0.11	1.39	1.46
$100\,\mu$s (depth 300)	0.33	2.72	3.11
$300\,\mu$s ($t = T_1$)	1.0	20	29

For $t \ll T_1$ (the practical operating range), $\kappa_{\mathrm{sv}} \approx \kappa_{\mathrm{spec}} \approx 1$. Beyond $t \gg T_1$, both condition numbers diverge and the method fails (Section 3.2).

2.4 Apply Correction

For the measured density matrix ρ_{meas}:

$$\rho_{\text{corrected}} = \text{unvec}\left(\mathcal{N}^{-1} \cdot \text{vec}(\rho_{\text{meas}})\right) \tag{5}$$

Then project back to a physical state: enforce Hermiticity ($\rho \to (\rho+\rho^\dagger)/2$), clip negative eigenvalues, normalize ($\rho \to \rho/\text{Tr}(\rho)$). For the gate-aware variant this projection is cosmetic ($\Delta F < 10^{-12}$, just floating-point cleanup). For the global variant, accumulated commutation error between gate unitaries and the noise channel produces negative eigenvalues of order $O(\gamma t_{\text{gate}})^d$; the projection does real work at large depths (§4.2).

Algorithm Summary

Algorithm 1: Spectral Error Mitigation

Input: Measured density matrix ρ_{meas}; per-qubit calibration data $\{T_1^{(i)}, T_2^{(i)}\}$; gate schedule (times $\{t_g\}$, connectivity)

Preprocessing (one-time, classical): 1. For each qubit i: build cluster Lindblad $\mathcal{M}_i$ from k-nearest neighbors (Eq. 2) 2. Compute total decoherence time $t = \sum_g t_g$ from the gate schedule 3. Compute correction matrix $\mathcal{N}_i^{-1} = e^{-\mathcal{M}_i t}$ (one $4^k \times 4^k$ matrix exponential per qubit)

Correction (per circuit execution): 4. Apply: $\text{vec}(\rho_{\text{corr}}) = \mathcal{N}^{-1} \cdot \text{vec}(\rho_{\text{meas}})$ 5. Project to physical state: Hermiticity, PSD, $\text{Tr} = 1$

Output: Corrected density matrix ρ_{corr}

Cost: Preprocessing: $O(n \cdot 4^{3k})$ (one-time, dominated by matrix exponential). Per-execution: $O(n \cdot 4^{2k})$ (matrix-vector multiply). For $k = 3$: $O(262{,}144\, n)$ and $O(4{,}096\, n)$ respectively — both trivial for any n.

Dual formulation (for large n, avoids full tomography): 4'. Compute corrected observable: $O' = \mathcal{N}^{-1\dagger}[O]$ (same 64×64 operation per qubit) 5'. Estimate $\langle O \rangle_{\text{corr}} = \text{Tr}(O' \cdot \rho_{\text{meas}})$ via standard measurement protocol

The dual formulation requires only expectation value estimation, not density matrix reconstruction, and scales to arbitrary n.

---

3. Why This Works (and When It Doesn't)

3.1 Why It Works

Three properties of Lindblad noise make exact inversion practical:

1. **Locality.** Decoherence is dominated by single-qubit processes (T_1 decay, T_2 dephasing) with nearest-neighbor corrections. The Kastoryano–Eisert theorem (2013) guarantees exponential decay of correlations: $\text{Corr}_\sigma(A : B) \le C\, e^{-d(A,B)/\xi}$. A 3-qubit cluster is enough ($< 0.01\%$ error, §5.2).

2. **Near-identity.** For short circuits, $\mathcal{N}(t)$ is close to I. Concretely: $|e^{\lambda_i t} - 1| \leq e^{|\text{Re}(\lambda_{\min})|t} - 1 \approx |\text{Re}(\lambda_{\min})|\, t$ when $|\text{Re}(\lambda_{\min})|\, t \ll 1$. At depth 100 ($t = 33\,\mu$s): $|\text{Re}(\lambda_{\min})|\, t = 0.33$. The noise channel is a small perturbation.

3. **Well-conditioned.** From Proposition 1: $\kappa_{\text{sv}} = 1.02$ at depth 1, $\kappa_{\text{sv}} = 1.46$ at depth 100. The inversion is numerically stable throughout the practical range ($t/T_1 < 0.3$, $\kappa_{\text{sv}} < 3.1$).

3.2 When It Fails

At $t \sim T_1 = 300\,\mu$s: $\kappa_{\text{sv}} \approx 29$, noise amplification $\sim 29\times$, requiring $\sim 840\times$ more shots. At $t \gg T_1$: the noise channel collapses to a rank-1 projector (thermal state), the inverse diverges, and the quantum information is gone. No method can recover it.

Hard limit: $d_{\max} \sim T_1/t_{\text{gate}} \approx 1000$ gates for current hardware.

3.3 Why Not ZNE?

ZNE fits $\langle O\rangle(\lambda) \approx a + b\lambda + c\lambda^2$ and extrapolates to $\lambda = 0$. This fails in three regimes: when noise decays exponentially rather than polynomially, when qubits have heterogeneous noise rates, and when circuits include idle times (ZNE only stretches gate noise). Spectral mitigation handles all three exactly.

---

4. Numerical Results

4.1 Bell State Circuit

Transmon parameters: $T_1 = 300\,\mu$s, $T_2 = 200\,\mu$s. Gate times: 1Q = 20 ns, 2Q = 300 ns, measurement = 1 μs.

State	Fidelity	Purity
Ideal Bell $\lvert\Phi^+\rangle$	1.000000	1.000000
Noisy	0.991274	0.982663
Spectral corrected	**1.000017**	**1.000033**

Error reduction (depth-1, gate-aware correction): **100%**. Residual error (1.7×10^{-5}) is from numerical floating-point, not from the method. Depth scaling and variant comparison in §4.2.

4.2 Depth Scaling

Each layer: Hadamard (20 ns) + CNOT (300 ns), Lindblad noise after each gate. Two correction strategies:

Global: Apply $\mathcal{N}^{-1}(t_{\text{total}}) = e^{-\mathcal{M}t_{\text{total}}}$ to the final state — treats all noise as one channel.

Gate-aware: Invert each gate's noise individually: $e^{-\mathcal{M}t_g}$ and $U_g^\dagger$ in reverse order.

Depth	t_{total} (μs)	F_{noisy}	F_{global}	$F_{\text{gate-aware}}$	κ_{sv}
1	1.32	0.9913	**0.9999**	**1.0000**	1.02
10	4.20	0.9751	**0.9966**	**1.0000**	1.05
50	17.0	0.9085	**0.9825**	**1.0000**	1.21
100	33.0	0.8375	**0.8912**	**1.0000**	1.46

Global correction gives $> 86\%$ error reduction for $d \leq 10$ with a single matrix inversion. It degrades at larger depths because commutation error between unitaries and noise grows as $O(d \cdot \gamma t_{\text{gate}})$. Gate-aware correction stays exact at all depths, but needs the full circuit structure. Both use the same 64×64 cluster Lindblad.

4.3 Statistical Noise

The numbers above use exact density matrices. On real hardware, ρ_{meas} comes from finite shots ($\delta\rho \sim 1/\sqrt{N_{\text{shots}}}$). The spectral correction amplifies this:

$$\|\delta\rho_{\text{corr}}\| \leq \kappa_{\text{sv}} \cdot \|\delta\rho_{\text{meas}}\| \tag{6}$$

At depth 1: +2% amplification (negligible). At depth 100: +46% (manageable). Unlike ZNE, where the polynomial extrapolation amplifies fluctuations unpredictably, this amplification is monotonic and known in advance.

The shot overhead is:

$$N_{\text{shots}}^{\text{corrected}} = \kappa_{\text{sv}}^2 \cdot N_{\text{shots}}^{\text{baseline}} \tag{7}$$

At depth 100: 2.1× overhead — roughly doubles the shot budget. At depth 1: 1.04×, negligible.

4.4 Comparison with ZNE

At depth 1, noise is small ($< 1\%$ error) and both methods recover the state easily — this is a sanity check, not a stress test:

Method	Fidelity	Error	Circuits
Noisy	0.991258	0.874%	1
ZNE (Richardson, 3-point)	0.999999	~0%	**3**
Spectral	**1.000000**	**0.000%**	**1**

The advantage at depth 1 is purely in circuit count (1 vs 3). The real separation emerges at higher depth (§4.2): spectral correction remains exact at depth 100 ($\kappa_{\text{sv}} = 1.46$), while ZNE's polynomial extrapolation degrades because noise is exponential, not polynomial (§3.3). At depth 100 with per-qubit variation, uniform ZNE leaves a fidelity residual of ~ 0.19 that per-qubit spectral correction eliminates (§5.3).

5. Scalability: The Cluster Decomposition

5.1 Per-Qubit Noise Profiles

Each qubit gets its own noise model:

Qubit	ω (GHz)	T_1 (μs)	Cluster
0	5.0	300.0	[0, 1]
1	5.1	300.0	[0, 1, 2]
2	5.2	300.0	[1, 2, 3]
$\vdots$	$\vdots$	$\vdots$	$\vdots$

Each qubit's correction matrix $\mathcal{N}_i^{-1}$ is independent. For n qubits: n independent 64×64 inversions.

5.2 Scaling Comparison (Numerically Verified)

We compare cluster vs. exact full-system Lindblad for $n = 2$–6 (homogeneous $T_1 = 300\,\mu$s, $T_2 = 200\,\mu$s, $J = 0.01$ GHz, $t = 33\,\mu$s):

Table 3a: Accuracy (cluster dim is always $64 = 4^3$)

n	Full dim	$\kappa_{\rm sv}^{\rm full}$	$\kappa_{\rm sv}^{\rm cluster}$	T_1 error
2	16	1.46	1.46	$< 0.001\%$
3	64	1.76	1.76	$< 0.001\%$
4	256	2.12	1.76	$< 0.001\%$
5	1,024	2.56	1.76	$< 0.001\%$
6	4,096	3.08	1.76	$< 0.001\%$
10	10^6	—	1.76	—
127	10^{76}	—	1.76	—

Table 3b: Performance

n	Time (full)	Time (cluster)	Speedup
2	0.004 s	0.001 s	3×
3	0.003 s	0.005 s	1×
4	0.11 s	0.004 s	26×
5	6.2 s	0.02 s	386×
6	286 s	0.02 s	14,327×
10	infeasible	0.02 s	∞
127	infeasible	2.5 s	∞

The cluster matches the full Lindblad to $< 0.001\%$ across all tested n and coupling strengths (J/γ up to 30,000). The cluster $\kappa_{\rm sv}$ stays at 1.76 regardless of system size — it only sees local noise. The full-system $\kappa_{\rm sv}$ grows with n due to many-body non-normality, but that's irrelevant: correction is applied per-qubit, so it's the *local* condition number that matters.

Why does the cluster work so well? The Kastoryano–Eisert theorem (2013) guarantees that for rapidly mixing local Lindbladians, correlations decay exponentially: $\mathrm{Corr}_\sigma(A : B) \le C\, e^{-d(A,B)/\xi}$. This is consistent with our numerical results: the approximation error should scale as $O(e^{-k/\xi})$, which for typical hardware parameters ($J/\gamma \lesssim 100$) is negligible at $k = 3$. A rigorous per-qubit error bound remains an open problem.

5.3 Heterogeneous Calibration: Per-Qubit Advantage

Real hardware has heterogeneous noise. Consider an IBM Brisbane-style 5-qubit chain:

Qubit	T_1 (μs)	T_2 (μs)	γ_1 (s^{-1})	γ_ϕ (s^{-1})	κ_{sv} ($t = 5\,\mu$s)
q_0	300	200	3,333	3,333	1.10
q_1	150	80	6,667	9,167	1.14
q_2	250	180	4,000	3,556	1.22
q_3	100	60	10,000	11,667	**1.16**
q_4	350	250	2,857	2,571	1.12

The worst qubit (q_3) has 3.5× the noise rate of the best (q_4). Spectral correction adapts to each qubit individually. A chip-average correction (mean $T_1 = 230\,\mu$s) gives $\kappa_{\mathrm{sv}} = 1.04$ — under-corrects the noisy qubits, over-corrects the quiet ones.

ZNE cannot do this: it applies one global λ to all qubits. At depth 100, uniform correction leaves a fidelity residual of ~ 0.19 that per-qubit correction eliminates.

---

6. Discussion

6.1 Relation to Other Mitigation Methods

Method	Noise model	Circuits	Exact?	Per-qubit?	Overhead
ZNE (Li and Benjamin, 2017)	Polynomial fit	3–5	No	No	3–5× circuits
PEC (Temme et al., 2017)	Pauli channel	$O(e^{2\epsilon})$ samples	Asymptotic	Yes	Exponential sampling
Sparse PL-PEC (van den Berg et al., 2023)	Pauli-Lindblad	$O(e^{2\epsilon})$ samples	Asymptotic	Yes	Learned sparse model
CDR (Czarnik et al., 2021)	Learned	Many Clifford	Approximate	Yes	Clifford training
TEM (Filippov et al., 2023)	Tensor network	1	Approximate	Yes	Quadratic vs PEC

Method	Noise model	Circuits	Exact?	Per-qubit?	Overhead
Spectral (this work)	**Lindblad**	**1**	**Exact**†	**Yes**	$O(4{,}096\,n)$ **classical**

†Exact under the Markovian assumption with known calibration data; gate-aware variant (§4.2) is exact for any circuit depth.

The closest competitor is van den Berg et al. (2023): also Lindblad-based, but corrects via probabilistic error cancellation with $\sim e^{2\epsilon}$ sampling overhead. That method powered the 127-qubit utility demo (Kim et al., 2023). Ours trades sampling overhead for density matrix access (or the dual formulation, §6.2), achieving deterministic exact correction.

The broader picture: ZNE and PEC are *model-free* (learn noise from hardware). Our method and van den Berg et al. are *model-based* (construct noise from calibration). Model-based is exact when the model is right; model-free is robust when it's wrong. The two complement each other.

Fundamental limits. Takagi et al. (2022) prove that mitigation overhead must be exponential in depth × noise rate. Our $\kappa_{\text{spec}} = e^{|\lambda_{\min}|t}$ *is* this fundamental limit — same exponential, but manifested as deterministic amplification (κ_{sv}^2 in shot count) rather than probabilistic sampling ($\sim e^{2\epsilon}$). This does not escape the bound — it makes the bound predictable.

6.2 Practical Requirements

The method requires: 1. **Known** $T_1^{(i)}$**,** $T_2^{(i)}$ per qubit — standard calibration data available from IBM/Google/Quantinuum backends. 2. **Known gate times** — available from backend specifications. 3. **Markovian noise** — the Lindblad model assumes no memory. For $1/f$ noise, a correction factor may be needed. 4. **Density matrix access** — full state tomography or shadow tomography provides ρ_{meas}. For expectation values only, apply $\mathcal{N}^{-1}$ to the observable instead: $\langle O\rangle_{\text{corrected}} = \text{Tr}(O \cdot \mathcal{N}^{-1}[\rho]) = \text{Tr}(\mathcal{N}^{-1\dagger}[O] \cdot \rho)$.

6.3 Generality

Mathematically, $\mathcal{N}^{-1} = e^{-\mathcal{M}t}$ is time-reversal of a Markov process. The classical analogue — deconvolving a diffusion kernel — is textbook signal processing. The quantum contribution is the cluster decomposition that makes this tractable for many-body open systems.

The framework applies to any hardware with Markovian noise and known decoherence parameters. For trapped ions ($T_1 \sim$ 1–10 s), κ_{sv} would be negligible for all practical depths, though coherent gate errors require separate correction (§7.4). For photonic qubits (photon loss instead of decoherence), the Lindblad operators change but the inversion is identical.

---

7. Limitations and Extensions

1. **Markovian assumption.** Real qubits have $1/f$ flux noise (non-Markovian). For sub-microsecond circuits — which covers most NISQ algorithms including all variational circuits — the Markov approximation is fine ($1/f$ is effectively white at $\omega \gg 1/T_{\text{circuit}}$). For longer

circuits, one could use time-dependent rates $\gamma(t)$ and invert $\mathcal{N}^{-1}(t) = \tilde{\mathcal{T}} \exp(-\int_0^t \mathcal{M}(s)\, ds)$ via Trotter decomposition. No longer a single expm, but still tractable.

2. **Tomography overhead.** Full state tomography is exponential in n. Two workarounds: (a) classical shadows (Huang et al., 2020) — $O(\log n)$ measurements per observable; (b) dual formulation — correct the observable instead of ρ: $\langle O \rangle_{\text{corr}} = \text{Tr}(\mathcal{N}^{-1\dagger}[O] \cdot \rho_{\text{meas}})$. No tomography needed.
3. **Deep circuits.** At $t = T_1$: $\kappa_{\text{sv}} \approx 29$, requiring $\sim 840\times$ more shots. At $t = 3T_1$: $\kappa_{\text{sv}} > 10^4$. Beyond $t \gg T_1$, the quantum information is thermalized and no method can recover it. Hard limit: $d_{\text{max}} \sim T_1/t_{\text{gate}} \approx 1000$ for current hardware.
4. **Coherent errors.** Lindblad captures incoherent noise only. Coherent errors (gate miscalibration, crosstalk) need unitary correction U_{corr}. The good news: they compose cleanly, $\rho_{\text{corr}} = U_{\text{corr}} \cdot \mathcal{N}^{-1}[\rho_{\text{meas}}] \cdot U_{\text{corr}}^\dagger$. In practice, randomized benchmarking handles coherent errors orthogonally.
5. **Calibration drift.** T_1 and T_2 fluctuate on minute timescales. The correction error scales as $\|\delta\rho\| \lesssim \kappa_{\text{sv}} \cdot |\delta\gamma/\gamma| \cdot t$, where $\delta\gamma/\gamma$ is the relative calibration error. For a depth-100 circuit ($\kappa_{\text{sv}} = 1.46$) with 10% calibration error: $\|\delta\rho\| \lesssim 1.46 \times 0.1 \times 0.11 \approx 0.016$, i.e., $< 2\%$ fidelity loss. For shallow circuits ($\kappa_{\text{sv}} \approx 1$), even 20% miscalibration yields sub-percent error. Recalibrating before each batch (standard practice on IBM/Google backends) keeps $\delta\gamma/\gamma \lesssim 5\%$.

---

8. Conclusion

The idea is simple: if you know the noise, undo it. The Lindblad superoperator gives an exact noise model from standard calibration data. The cluster decomposition makes inversion scale linearly. The condition number tells you in advance how well it will work.

The ingredients:

1. **Cluster Lindblad** — $O(n)$ matrix exponentials of size 64×64 from T_1, T_2, gate times. One-time cost, scaling justified by Kastoryano–Eisert.
2. **Exact inversion** — $\mathcal{N}^{-1} = e^{-\mathcal{M}t}$ (Theorem 1), $\kappa_{\text{sv}} \leq 1.46$ at depth 100 (Proposition 1). Gate-aware variant: exact at any depth.
3. **Physical projection** — Hermiticity, PSD, trace 1.

Results: gate-aware correction is exact at all tested depths (1–100). Global correction gives $> 86\%$ error reduction for $d \leq 10$ with a single matrix inverse. Shot overhead: $2.1\times$ at depth 100.

For the 127-qubit circuits of Kim et al. (2023): 127 independent 64×64 inversions, \$ \$8 MB, done on a laptop. The method composes with coherent-error correction, works without full tomography (dual formulation), and provides per-qubit profiles that ZNE cannot.

$$\boxed{\rho_{\text{corrected}} = e^{-\mathcal{M}t}\,\rho_{\text{measured}}}$$

One matrix. One circuit. Exact.

AI Disclosure

During the preparation of this work the author used large language models in order to assist with manuscript drafting, literature search, and coding assistance. After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

References

- Cai, Z., R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O'Brien (2023). Quantum error mitigation. *Rev. Mod. Phys.* **95**, 045005. DOI: 10.1103/RevModPhys.95.045005
- Czarnik, P., A. Arrasmith, P. J. Coles, and L. Cincio (2021). Error mitigation with Clifford quantum-circuit data. *Quantum* **5**, 592. DOI: 10.22331/q-2021-11-26-592
- Endo, S., S. C. Benjamin, and Y. Li (2018). Practical quantum error mitigation for near-future applications. *Phys. Rev. X* **8**, 031027. DOI: 10.1103/PhysRevX.8.031027
- Filippov, S. N., M. Leahy, M. A. C. Rossi, and G. García-Pérez (2023). Scalable tensor-network error mitigation for near-term quantum computing. arXiv:2307.11740.
- Huang, H.-Y., R. Kueng, and J. Preskill (2020). Predicting many properties of a quantum system from very few measurements. *Nature Physics* **16**, 1050–1057. DOI: 10.1038/s41567-020-0932-7

- Kandala, A., K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta (2019). Error mitigation extends the computational reach of a noisy quantum processor. *Nature* **567**, 491–495. DOI: 10.1038/s41586-019-1040-7
- Kastoryano, M. J. and J. Eisert (2013). Rapid mixing implies exponential decay of correlations. *J. Math. Phys.* **54**, 102201. DOI: 10.1063/1.4822481
- Kim, Y., A. Eddins, S. Anand, K. X. Wei, E. van den Berg, S. Rosenblatt, H. Nayfeh, Y. Wu, M. Zaletel, K. Temme, and A. Kandala (2023). Evidence for the utility of quantum computing before fault tolerance. *Nature* **618**, 500–505. DOI: 10.1038/s41586-023-06096-3
- Li, Y. and S. C. Benjamin (2017). Efficient variational quantum simulator incorporating active error minimization. *Phys. Rev. X* **7**, 021050. DOI: 10.1103/PhysRevX.7.021050
- Takagi, R., S. Endo, S. Minagawa, and M. Gu (2022). Fundamental limitations of quantum error mitigation. *npj Quantum Inf.* **8**, 114. DOI: 10.1038/s41534-022-00618-z
- Temme, K., S. Bravyi, and J. M. Gambetta (2017). Error mitigation for short-depth quantum circuits. *Phys. Rev. Lett.* **119**, 180509. DOI: 10.1103/PhysRevLett.119.180509
- van den Berg, E., Z. K. Minev, A. Kandala, and K. Temme (2023). Probabilistic error cancellation with sparse Pauli-Lindblad models on noisy quantum processors. *Nature Physics* **19**, 1116–1121. DOI: 10.1038/s41567-023-02042-2

Appendix: Reference Implementation

The complete pipeline (305 lines, self-contained NumPy + SciPy) is available as spectral_error_mitigation.py. The core functions implementing Eq. 1–5 are reproduced below.

```
import numpy as np
from scipy.linalg import expm, inv

def build_cluster_lindblad(n_qubits, gamma1, gamma_phi):
    """Build Lindblad superoperator M for a k-qubit cluster (Eq. 2).
    Returns a 4^k x 4^k matrix."""
    d = 2**n_qubits
    Id = np.eye(d, dtype=complex)
    M = np.zeros((d**2, d**2), dtype=complex)
    sigma_minus = np.array([[0,0],[1,0]], dtype=complex)
    sigma_plus  = np.array([[0,1],[0,0]], dtype=complex)
    sigma_z     = np.array([[1,0],[0,-1]], dtype=complex)
    for i in range(n_qubits):
        for L, g in [(multi_qubit_op(sigma_minus, i, n_qubits), gamma1),
                     (multi_qubit_op(sigma_plus, i, n_qubits), 0.0),
                     (multi_qubit_op(sigma_z, i, n_qubits), gamma_phi/2)]:
            LdL = L.conj().T @ L
            M += g * (np.kron(L.conj(), L)
                     - 0.5*np.kron(Id, LdL)
                     - 0.5*np.kron(LdL.T, Id))
    return M

def spectral_correct(rho_meas, M, t):
    """Spectral error mitigation (Eq. 1, 4, 5).
    N = e^{Mt}, N^{-1} = e^{-Mt}, rho_corr = N^{-1} rho_meas."""
```

```
    N_inv = expm(-M * t)                                # Eq. 4
    d = rho_meas.shape[0]
    vec_corr = N_inv @ rho_meas.flatten('F')            # Eq. 5
    rho = vec_corr.reshape((d, d), order='F')
    rho = 0.5 * (rho + rho.conj().T)                    # Hermiticity
    rho = rho / np.trace(rho)                           # Tr = 1
    return rho
```

Usage: python3 spectral_error_mitigation.py. Runtime: 2 seconds. Produces: Bell state correction, depth scaling, ZNE comparison.