Resource-aware error mitigation for variational quantum eigensolvers: coupling zero-noise extrapolation, shot allocation and adaptive circuit growth on noisy hardware models

[1]Saswat Swain, [2]Chandrasekhar Panda, [3]Pravat Satpathy
[1]Ph.D Scholar, [2]Researcher, [3]Researcher
[1]KSOM — KIIT Deemed to be University,
Bhubaneswar, Odisha, India
[2,3]Research and Development Division,
Tekstein Solutions Private Limited, Bhubaneswar, Odisha, India
[1]saswat.mck@gmail.com, [2]contact@chandrasekharpanda.com, [3]analytics.pravat@gmail.com

***Abstract*—Variational quantum eigensolvers (VQE) on noisy intermediate-scale quantum (NISQ) hardware are limited by three coupled resources: gate noise that biases every energy evaluation, the finite shot budget that sets the statistical precision, and the circuit depth that both the ansatz and any noise-amplifying mitigation scheme consume. Zero-noise extrapolation (ZNE), variance-aware shot allocation and adaptive ansatz construction are each established in isolation, yet they interact: extrapolation amplifies variance, allocation changes the noise floor against which an adaptive method decides whether to grow, and growth changes the depth that extrapolation must fold. This paper presents RAM-VQE, a resource-aware mitigation framework in which the three mechanisms share one shot ledger and one noise model: adaptive operator selection is screened on the mitigated energy landscape with a significance test and geometric precision escalation, shots are allocated jointly across noise-scale factors and commuting measurement groups by a Neyman rule, and extrapolation uses a depolarizing-channel (exponential) model inside the loop and for the final estimate. Two propositions give the allocation gain and the extrapolation variance amplification in closed form. The framework is evaluated on a gate-level density-matrix model with per-gate depolarizing and readout errors, validated against PennyLane to 10^{-12} Ha, for H_2 (4 qubits), stretched LiH (6 qubits) and a 4-site XXZ chain, against seven ablation arms over six seeds, with noise-rate and budget sweeps. At a two-qubit error rate of 5×10^{-3} the root-mean-square error relative to readout-mitigated VQE falls from 88.5 to 3.12 mHa (H_2) and from 82.9 to 5.22 mHa (LiH, with 48.0 rather than 64.0 two-qubit gates), but at 8.65 times the shot cost of applying the same mitigation to a fixed ansatz, whose error is 6.83 mHa; on the spin chain the adaptive pool is ansatz-limited and the fixed ansatz is better. Three negative findings are reported: adaptive growth screened on raw noisy gains stalls at the reference state, in-loop extrapolation without shot allocation degrades the optimisation, and data-driven selection among extrapolation models from three scale factors is unreliable. Code, seeds and result logs are released.**

***Index Terms*—Variational quantum eigensolver, zero-noise extrapolation, quantum error mitigation, shot allocation, adaptive ansatz, NISQ computing, depolarizing noise, Neyman allocation, quantum chemistry simulation, spin-chain Hamiltonians.**

I. Introduction

The variational quantum eigensolver [1], [2] prepares a parameterised trial state on a quantum processor and delegates the minimisation of the measured energy to a classical optimiser. Because the circuits are shallow and the only quantum output is a set of Pauli expectation values, VQE has become the reference workload for noisy intermediate-scale quantum (NISQ) devices [3], [4]. The same shallowness is also its weakness: every two-qubit gate contributes a systematic bias to the measured energy, every expectation value carries shot noise whose cost scales with the squared precision, and the optimiser must navigate a landscape that is both biased and stochastic.

Three families of remedies have matured independently. Error-mitigation techniques such as zero-noise extrapolation (ZNE) [5]–[7] and probabilistic error cancellation [8] trade an increased number of circuit executions for a reduced bias. Measurement-frugal optimisation [9]–[11] distributes a finite shot budget across Hamiltonian terms or across optimisation steps so that a target precision is reached with the fewest samples. Adaptive ansatz construction, exemplified by ADAPT-VQE and its qubit-pool variant [12], [13], grows the circuit one operator at a time and therefore spends two-qubit gates only where the energy gradient justifies them.

These remedies are usually benchmarked separately, but on hardware they are coupled through the resources they consume: ZNE folds the circuit, so its benefit and its variance grow with the depth the ansatz has accumulated; shot allocation sets the

statistical floor against which an adaptive scheme decides whether a candidate operator helps; and growth screened on the raw noisy energy rejects any operator whose depolarizing penalty exceeds its correlation-energy gain, which, as Section VIII shows, is every 14-gate double excitation at realistic error rates. A framework that accounts for these interactions is needed before any of the three can be relied on in a practical NISQ workflow.

This paper proposes RAM-VQE, a resource-aware mitigation framework that couples the three mechanisms through a shared shot ledger and a shared noise model. The main contributions are as follows.

1) A noise-screened adaptive growth rule that selects operators from the mitigated energy landscape and accepts them only when the predicted gain passes a sequential significance test against the estimated shot-noise floor, with geometric precision escalation when no candidate passes, a patience mechanism for gradient troughs and a revert step for operators that never pay off.
2) A joint Neyman allocation of shots over the product of noise-scale factors and qubit-wise-commuting measurement groups, with a closed-form variance bound (Proposition 1) and an explicit expression for the variance amplification of any linear extrapolation estimator (Proposition 2).
3) A budgeted ZNE design in which a cheap two-point exponential extrapolation is used inside every optimisation step and a three-point extrapolation is reserved for the final estimate, with the exponential form justified by the depolarizing-channel bias model derived in Section V.
4) A validated gate-level density-matrix simulator (cross-checked against PennyLane statevector and mixed-state devices to 10^{-12} Ha) and an eight-arm ablation study on H_2, stretched LiH and a 4-site XXZ chain, with noise and budget sweeps, reporting the improvements, their resource cost and three negative results.

II. Related Work

Table I summarises the works most relevant to the present study. Early VQE demonstrations [1], [14] established that chemical accuracy for H_2 is attainable with a single-parameter ansatz, but did so by fitting the measured energy curve rather than by mitigating gate noise. Richardson-type extrapolation was introduced independently in [5] and [6]; the digital folding form used here follows [7], which also observed that the choice between linear, polynomial and exponential fits is data dependent. The review in [15] catalogues the bias–variance trade-off of every mitigation family and states the sampling overhead of ZNE as the sum of squared extrapolation coefficients, the quantity made explicit in Proposition 2.

On the measurement side, [9] allocated shots in proportion to Hamiltonian coefficients, [10] replaced the constant by the measured variance inside a stochastic-gradient loop, and [11] adapted the shot count per iteration to the gradient norm; commuting-group measurement schemes [16], [17] reduce the number of settings but do not decide how many shots each receives. ADAPT-VQE [12] and qubit-ADAPT [13] select operators by gradient magnitude on a noiseless simulator, whereas [18] showed that gate noise flattens the landscape, which is the regime in which gradient screening fails. Variable-noise Clifford data regression [19], [20] combines ZNE with learned corrections at a training cost that is not budgeted against the optimisation.

None of these works treats the shot budget, the extrapolation overhead and the ansatz depth as one resource problem, and none reports what happens to an adaptive method when its screening signal is itself noisy; those are the gaps addressed here.

TABLE I
Summary of related work and the gap addressed by this study

Ref.	Year	Method	Limitation relative to this work
[5]	2017	Richardson ZNE (pulse stretching)	Variance amplification not budgeted; fixed ansatz
[6]	2017	Extrapolation + quasi-probability	No allocation across scale factors
[7]	2020	Digital (unitary-folding) ZNE	Fit model left to the user
[9]	2015	Shots ∝ |coefficient|	Ignores variance and commuting groups
[10]	2020	Variance-weighted sampling (Rosalin)	No mitigation; bias untreated
[11]	2020	Adaptive shots per step (iCANS)	Fixed ansatz; no extrapolation
[12]	2019	ADAPT-VQE	Screening assumes noiseless device
[13]	2021	qubit-ADAPT-VQE	Same noiseless screening
[18]	2021	Noise-induced barren plateaus	Diagnostic only
[19]	2021	vnCDR (Clifford regression + ZNE)	Training circuits not budgeted
This work	2026	RAM-VQE: coupled growth, allocation, ZNE	Simulation only; depolarizing model

III. Problem Statement

Let $H = \Sigma_k w_k P_k$ be an n-qubit Hamiltonian written as a sum of M Pauli strings, with ground-state energy E_0. A NISQ device prepares a trial state $\rho(\boldsymbol{\theta})$ whose circuit contains N_{CX} two-qubit gates, each followed by a depolarizing channel of strength p_2, and whose readout is corrupted by a symmetric bit-flip of probability p_{ro}. Three deficiencies of the conventional workflow are addressed.

Biased landscape. The device returns $E_{noisy}(\boldsymbol{\theta}) = \mathrm{Tr}[\rho(\boldsymbol{\theta})H]$ rather than the ideal energy. Under depolarizing noise the bias is a contraction towards the maximally mixed energy $E_{mix} = \mathrm{Tr}\, H/2^n$; it moves the location of the minimum, so an optimiser run on the raw device converges to wrong parameters even if the final value is later corrected. *Unbudgeted mitigation.* ZNE with $\lambda \in \{1, 2, 3\}$ multiplies the two-qubit gate executions by six for a Richardson fit and amplifies the shot variance by $\Sigma\gamma_\lambda^2$, which reaches 19 for the quadratic estimator; unweighed against the budget, mitigation can raise the mean-square error. *Blind growth.* When an adaptive scheme takes its gain estimate on the noisy device, the depolarizing penalty of the candidate is

included in the measured gain, which is negative for any operator costing more than a few two-qubit gates.

The engineering objective is to obtain an estimate $\hat{E}_0$ with minimal mean-square error $|\hat{E}_0 - E_0|^2$ for a given total shot budget B and a given device error model, where the budget includes every pilot, screening, folding and final-evaluation circuit, and the depth of the circuit is an outcome of the method rather than an input.

IV. Proposed Framework

Fig. 1 shows the architecture. A resource controller owns the shot ledger and the noise parameters and feeds three modules that exchange information at every step: Module I grows the ansatz, Module II allocates shots and Module III extrapolates. A measurement layer groups the Hamiltonian into K qubit-wise-commuting (QWC) settings by greedy first-fit colouring and applies readout correction. The loop is driven by a Rotosolve-type coordinate minimiser [21], which needs no gradient and whose per-parameter minimum follows in closed form from a fixed number of evaluations.

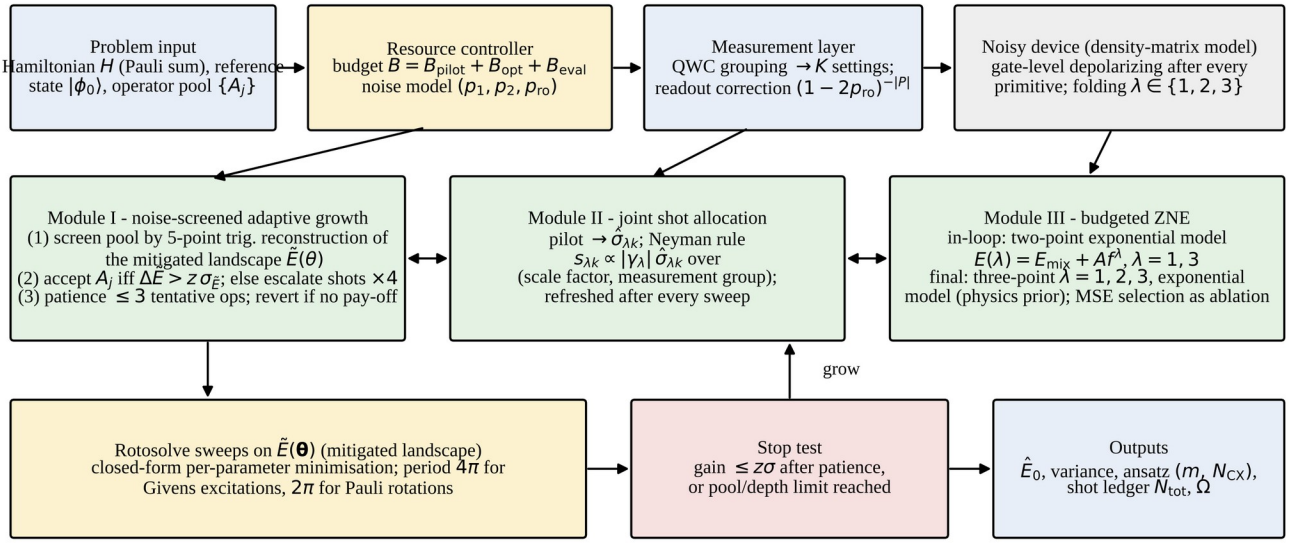


Fig. 1. Architecture of RAM-VQE. The resource controller (centre top) issues the shot budget and noise model to three coupled modules: noise-screened adaptive growth (Module I), joint shot allocation over scale factors and measurement groups (Module II) and budgeted zero-noise extrapolation (Module III). All circuit executions, including pilots and screening, are charged to one ledger.

A. Module I: noise-screened adaptive growth

The operator pool is problem specific: spin-conserving single and double fermionic excitations (Givens rotations) for the molecules, and Givens rotations on all qubit pairs plus nearest-neighbour ZZ phase rotations for the spin chain. At each growth step every pool operator A_j is appended with a free angle and the mitigated energy along that angle is reconstructed from $2R + 1$ evaluations (Section V-C), giving the exact minimum of the one-dimensional landscape rather than only its slope. The pool is ranked at a base precision of s_{base} shots per evaluation; the top two candidates are then re-screened at the current working precision $s_{\text{opt}} \geq s_{\text{base}}$, and the better of them is accepted if its predicted gain satisfies $\Delta\tilde{E} > z\sigma_{\tilde{E}}$, where $\sigma_{\tilde{E}}$ is the predicted standard deviation of the mitigated estimator at the current allocation and $z = 3$. If no candidate passes, the working precision is escalated by a factor of four, up to a cap s_{cap}, and the screen is repeated; the shot count per evaluation thus grows geometrically as the optimisation approaches the noise floor, which is the same rule applied to the sweep convergence of the fixed-ansatz arms. Only at the cap may operators be added tentatively, up to three in a row, to escape gradient troughs; the cumulative gain of a tentative block is re-measured with four times the working shots against the energy recorded at the last significant operator, and the block is removed before the final sweeps unless that paired difference exceeds one standard deviation; the asymmetry between the acceptance test (3σ) and the retention test (1σ) reflects that retaining a marginal block costs only depth, whereas accepting one costs further screening.

B. Module II: joint shot allocation

A pilot of uniform shots at the current parameters yields sample standard deviations $\hat{\sigma}_{\lambda k}$ for every pair of scale factor λ and measurement group k; the shots of every subsequent evaluation are distributed as $s_{\lambda k} \propto |\gamma_\lambda|\,\hat{\sigma}_{\lambda k}$, where γ_λ are the intercept sensitivities of the extrapolation in use (Proposition 2). The pilot is refreshed after every sweep, is charged to the ledger, and a floor of 20 shots per setting is kept.

C. Module III: budgeted zero-noise extrapolation

Noise is scaled by gate-level unitary folding $G \rightarrow GG^\dagger G$ [7], [22]; $\lambda = 3$ folds every primitive and $\lambda = 2$ folds the first half of the two-qubit gates. Inside the loop only $\lambda \in \{1, 3\}$ are executed and the intercept is taken from the two-point exponential model of Section V-B, whose contraction factor is tracked by an exponential moving average so that the allocation weights of Module II are available before the next evaluation. The final estimate executes $\lambda \in \{1, 2, 3\}$ with the full evaluation budget, reports the linear, quadratic-Richardson and exponential intercepts together with their delta-method variances, and returns the exponential intercept by default. A data-driven

alternative that selects the model by estimated mean-square error is retained as an ablation.

V. Mathematical Model

A. Noise model and estimator

After every primitive gate on wires w the state is passed through a k-qubit depolarizing channel,

$$\mathcal{D}_w(\rho) = (1-q)\rho + q\,(I_w/2^k) \otimes \mathrm{Tr}_w\rho, \quad q = p\cdot 4^k/(4^k-1), \tag{1}$$

with $p = p_1$ for $k = 1$ and $p = p_2$ for $k = 2$, so that p is the probability that a uniformly random non-identity Pauli error is applied. Readout applies a symmetric bit-flip with probability p_{ro} to every qubit, under which the measured expectation of a Pauli string of weight $|P|$ is attenuated to $(1-2p_{ro})^{|P|}\langle P\rangle$. Dividing the sample mean by this factor removes the readout bias exactly for the symmetric model, at the price of a variance amplification $(1-2p_{ro})^{-2|P|}$; this correction is applied in every arm so that the comparison isolates gate-noise mitigation.

The energy estimator from K QWC groups with s_k shots each is

$$\hat{E} = w_0 + \Sigma_{k=1}^{K}\hat{G}_k, \quad \mathrm{Var}(\hat{E}) = \Sigma_k\,\sigma_k^2/s_k, \tag{2}$$

where $\hat{G}_k$ is the sample mean of the weighted parity sum of group k and σ_k^2 its single-shot variance, which includes the covariance between strings measured in the same setting.

Proposition 1 (allocation gain). For a total of S shots, the uniform allocation $s_k = S/K$ gives $\mathrm{Var}_u = K\Sigma_k\sigma_k^2/S$, whereas the Neyman allocation $s_k = S\sigma_k/\Sigma_j\sigma_j$ gives $\mathrm{Var}_N = (\Sigma_k\sigma_k)^2/S$, and

$$\mathrm{Var}_u/\mathrm{Var}_N = K\,\Sigma_k\sigma_k^2\,/\,(\Sigma_k\sigma_k)^2 \geq 1, \tag{3}$$

with equality only when all σ_k are equal. *Proof.* Minimising $\Sigma\sigma_k^2/s_k$ subject to $\Sigma s_k = S$ by a Lagrange multiplier gives $s_k \propto \sigma_k$; the inequality is Cauchy–Schwarz applied to the vectors (σ_k) and $(1, \ldots, 1)$. ∎ The ratio is measured in Section VIII-A for the three Hamiltonians.

B. Depolarizing bias and extrapolation

If each of the d two-qubit gates were followed by a global depolarizing channel of strength ε, the noisy energy at scale factor λ would obey

$$E(\lambda) = E_{mix} + (E_{ideal} - E_{mix})\,f^{\lambda}, \quad f = (1-\varepsilon)^d, \tag{4}$$

an exact exponential in λ. The local channels of (1) do not reduce to a single global channel, but the fitted curves in Fig. 2 show that (4) describes the folded energies of all three benchmarks to within a few milli-units out to $\lambda = 7$. Two consequences follow. First, a polynomial extrapolation is biased by the neglected curvature of f^{λ}: for the linear estimator the residual bias is $(E_{ideal} - E_{mix})(1-f)^2 f(3-f)/2$ to leading order, which grows with the depth through f. Second, two measurements suffice to identify f and the intercept,

$$\hat{f} = [(\hat{E}(3) - E_{mix})/(\hat{E}(1) - E_{mix})]^{1/2}, \tag{5}$$

$$\hat{E}(0) = E_{mix} + (\hat{E}(1) - E_{mix})/\hat{f}, \tag{6}$$

since E_{mix} is known analytically from the Hamiltonian coefficients. For the molecules E_{mix} is the identity coefficient; for the traceless spin chain it is zero.

Proposition 2 (variance amplification). Any extrapolation whose intercept is a differentiable function $\hat{E}(0) = g(\hat{E}(\lambda_1), \ldots, \hat{E}(\lambda_L))$ has, to first order, $\mathrm{Var}[\hat{E}(0)] = \Sigma_l\,\gamma_l^2\,\mathrm{Var}[\hat{E}(\lambda_l)]$ with $\gamma_l = \partial g/\partial\hat{E}(\lambda_l)$. For the exact-fit polynomial of degree $L-1$ the γ_l are the Lagrange basis values at zero; for $\lambda = (1, 2, 3)$ they are $(3, -3, 1)$ with $\Sigma\gamma^2 = 19$, for the least-squares line through the same points $(4/3, 1/3, -2/3)$ with $\Sigma\gamma^2 = 7/3$, and for the two-point model (5)–(6) they are $(1.5/f, -0.5/f^3)$ with $\Sigma\gamma^2 \to 2.5$ as $f \to 1$. Minimising $\Sigma_l\gamma_l^2\Sigma_k\sigma_{\lambda_{\{l}k\}}^2/s_{lk}$ subject to a total budget gives the joint rule $s_{lk} \propto |\gamma_l|\;\sigma_{\lambda_{\{l}k\}}$ used by Module II, with minimal variance $(\Sigma_l\Sigma_k|\gamma_l|\sigma_{\lambda_{\{l}k\}})^2/S$. ∎

C. Trigonometric reconstruction

For a Pauli rotation $\exp(-i\theta P/2)$ the energy is $E(\theta) = a_0 + a_1\cos\theta + b_1\sin\theta$ and three evaluations determine it. A Givens single or double excitation has generator eigenvalues $\{0, \pm 1/2\}$ in the convention used by PennyLane, so the energy contains the frequencies 1/2 and 1 in θ and is 4π-periodic; it is therefore reconstructed as a second-order trigonometric polynomial in $u = \theta/2$ from five evaluations at $u_j = u_0 + 2\pi j/5$. The per-parameter minimum is located on a grid and refined by golden-section search. Because the fit is exact for the noiseless function, every evaluation is spent on the mitigated energy and the only error is statistical, which is why $\sigma_{\hat{E}}$ is the natural threshold for Module I.

VI. Algorithm and Complexity

Algorithm 1: RAM-VQE

1: **Input:** H (groups $k = 1\ldots K$), reference $|\varphi_0\rangle$, pool $\{A_j\}$, noise (p_1, p_2, p_{ro}), z, patience T, s_{base}, s_{cap}

2: ansatz ← ∅; **θ** ← ∅; $s_{opt} \leftarrow s_{base}$; pilot $\hat{\sigma}_{\lambda k}$ at $|\varphi_0\rangle$; tentative ← 0; anchor ← 0

3: **while** |ansatz| < m_{max} **do**

4: rank pool: for each A_j reconstruct $\tilde{E}(\theta_j)$ from $2R+1$ evaluations at s_{base}; $\Delta\tilde{E}_j \leftarrow \tilde{E}_{cur} - \min\tilde{E}(\theta_j)$

5: re-screen the top-2 at s_{opt}; $j \leftarrow$ ***argmax*** $\Delta\tilde{\mathbf{E}}_j$; $\tau \leftarrow z\sigma_{\tilde{E}}(s_{opt})$

6: **if** $\Delta\tilde{E}_j \leq \tau$ **and** $s_{opt} < s_{cap}$ **then** $s_{opt} \leftarrow 4s_{opt}$; refresh pilot; **continue**

7: **if** $\Delta\tilde{E}_j \leq \tau$ **then** tentative ← tentative + 1; **if** tentative > T **then break**

8: append A_j at its reconstructed minimum; two Rotosolve sweeps on $\tilde{E}$ at s_{opt}; refresh pilot

9: measure $\tilde{E}_{new}$ with $4s_{opt}$ shots; **if** $\Delta\tilde{E}_j > \tau$ **or** $\tilde{E}_{anchor} - \tilde{E}_{new} > \sigma_{diff}$ **then** tentative ← 0, anchor ← |ansatz|, $\tilde{E}_{anchor} \leftarrow \tilde{E}_{new}$

10: truncate ansatz to anchor if tentative > 0; two final sweeps

11: pilot at $\lambda \in \{1,2,3\}$; allocate $s_{\lambda k} \propto |\gamma_\lambda|\hat{\sigma}_{\lambda k}$; measure; return $\hat{E}(0)$, variance, ledger

Time complexity. One growth step costs $(2R+1)|\text{pool}|$ evaluations at s_{base} for ranking, $2(2R+1)$ at s_{opt} for the re-screen and $2(2R+1)m$ at s_{opt} for the sweeps when the ansatz holds m operators, each evaluation being two circuit executions ($\lambda = 1, 3$). With m_{max} operators the number of evaluations is $O(m_{max}(|\text{pool}| + m_{max}))$ and the two-qubit-gate executions are $O(m_{max}^2 c)$ for a mean operator cost c, plus the fixed final evaluation; the shot count is dominated by the sweeps at the escalated precision, $O(m_{max}^2 s_{cap})$. The fixed-ansatz baseline costs $O(m_{full}\cdot\text{sweeps})$ evaluations on a circuit that is already m_{full} operators deep, which is why its gate-execution count (Ω in Section VII) remains comparable despite far fewer shots. The classical reconstruction and allocation are $O(K)$ per evaluation and negligible. *Space complexity* is dominated by the density

matrix of the simulator, $O(4^n)$; on hardware the classical memory is $O(K + m)$.

VII. Simulation Setup

Hamiltonians were generated with PennyLane 0.45 (qchem module, STO-3G, differentiable Hartree–Fock) [23]: H_2 at 0.7414 Å (4 qubits, 15 Pauli terms, 5 QWC groups) and LiH at 3.0 Å with a 2-electron, 3-orbital active space (6 qubits, 62 terms, 21 groups). The stretched LiH geometry was chosen deliberately: at the equilibrium length the active-space correlation energy is only 1.05 mHa, below the noise floor of every arm, and the adaptive method correctly terminates at a single operator, which makes the benchmark uninformative. At 3.0 Å the correlation energy is 16.3 mHa. The spin benchmark is an open 4-site XXZ chain, $H = J\Sigma(X_iX_{i+1} + Y_iY_{i+1} + \Delta Z_iZ_{i+1})$ with $J = 1$, $\Delta = 0.5$ (3 groups, 21 terms), prepared from the domain-wall product state $|0011\rangle$, which lies in the $S_z = 0$ sector of the ground state. Reference energies are from exact diagonalisation of the qubit Hamiltonians: −1.13727 Ha, −7.72709 Ha and −5.42434 J respectively.

The noisy device is a purpose-built density-matrix simulator (NumPy/SciPy) that decomposes every PennyLane operation to {CNOT, H, RY, RZ, RX} and applies (1) after each primitive. It agrees with PennyLane's default.qubit and default.mixed devices on the H_2 circuit to 10^{-12} Ha in both the noiseless and noisy energies, and its sampling layer reproduces the predicted shot variance to within 3%. Operator costs are 14 CNOTs per double excitation, 2 per single excitation and 2 per two-qubit Pauli rotation. Table II lists the parameters. All runs use a fixed global seed (2026) offset by the seed index, six seeds per arm in the main study and three in the sweeps, and ran on two CPU cores; a LiH run of the full framework takes 20–80 s depending on how far the precision escalates.

TABLE II
Simulation parameters (defaults; sweeps vary one parameter at a time)

Parameter	Symbol	Value
Two-qubit depolarizing probability	p_2	5×10^{-3} (sweep: 2.5×10^{-3} – 2×10^{-2})
Single-qubit depolarizing probability	p_1	$p_2/10$
Readout bit-flip probability	p_{ro}	10^{-2}
Base / cap shots per in-loop evaluation	s_{base}, s_{cap}	8 000 / 128 000 (split over λ and groups)
Precision escalation factor, re-screened candidates	–	4, top-2
Final-evaluation budget	B_{eval}	2×10^5 (sweep: 2×10^4 – 5×10^5)
Pilot fraction of B_{eval}	–	0.1
Noise-scale factors (loop / final)	λ	{1, 3} / {1, 2, 3}
Acceptance multiplier, pay-off shot multiplier	z, –	3, 4×
Patience, maximum operators	T, m_{max}	3, 12 (14 for XXZ)
Rotosolve sweeps (growth / final / fixed)	–	2 / 2 / ≤ 10 (stop when sweep gain < $z\sigma$ at s_{cap})
Seeds (main study / sweeps)	–	6 / 3

Eight arms are compared (Table IV), labelled A–H in Tables IV, VI and VII and in Fig. 3: A, readout mitigation only; B, A with ZNE on the final estimate; C, A with two-point ZNE inside every evaluation; D, A with Neyman allocation; E, adaptive growth screened on raw noisy gains; F, adaptive growth screened on mitigated gains with uniform shots; G, the fixed ansatz with allocation and in-loop and final ZNE; H, RAM-VQE. Arms A–D use the fixed full-pool ansatz (one layer of all singles and doubles for the molecules, two layers of the pool for XXZ). Arm E screens growth on the raw noisy gain, as a noiseless-trained ADAPT loop would when deployed on hardware; arm G isolates the contribution of growth. Three metrics are reported: the signed energy error of the final estimate, the total shots N_{tot} over the whole run, and the execution overhead $\Omega = \Sigma$(shots × two-qubit gates) over every executed circuit, a hardware-time proxy that charges folded circuits at their folded depth. The ansatz error, the ideal energy of the final parameters minus E_0, is reported separately so that optimisation drift can be distinguished from residual bias.

VIII. Results and Discussion

A. Noise-free characterisation

Fig. 2 and Table III characterise the mitigation problem at the ideal optimum of the fixed ansatz without shot noise. The folded energies follow the exponential form (4) closely: the two-point fit from λ = 1 and 3 predicts the λ = 5 and 7 energies to within 1% of the raw bias for every benchmark and error rate. The consequence for extrapolation is in Table III: at $p_2 = 5 \times 10^{-3}$ the raw bias of 81.5 mHa for LiH is reduced to 16.2 mHa by the linear fit, to 43.9 mHa by quadratic Richardson and to -6.55 mHa by the exponential model. The quadratic estimator is worse than the linear one at every setting, because the curvature of f^{λ} is not captured by three points and its coefficients (3, −3, 1) amplify the model error. The exponential intercept slightly overshoots (negative residual) since the local channels contract the state faster than a single global channel at the same gate count. The Neyman ratio of Proposition 1, measured at the same states with readout error, is 2.36 for H_2, 4.12 for LiH and 1.08 for XXZ: allocation is worth a factor of four in shots for the 21-group LiH Hamiltonian and almost nothing for the three balanced groups of the spin chain.

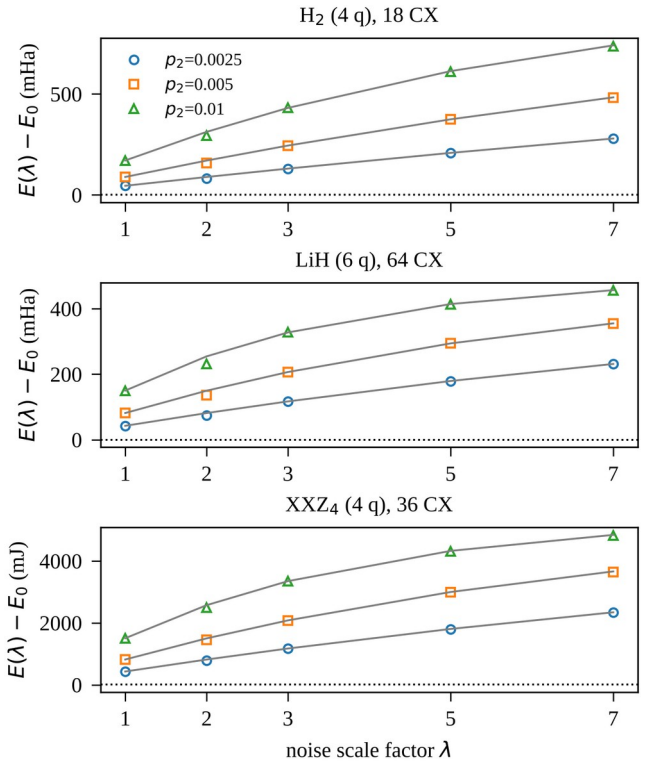


Fig. 2. Noisy energy versus noise-scale factor at the ideal optimum of the fixed ansatz (no shot noise). Markers: density-matrix simulation at three two-qubit error rates; grey lines: exponential model (4) fitted from λ = 1 and 3 only; dotted line: ideal energy of the ansatz.

TABLE III

SHOT-NOISE-FREE RESIDUAL BIAS OF THE EXTRAPOLATION MODELS AT THE FIXED-ANSATZ OPTIMUM (MHA FOR MOLECULES, MJ FOR XXZ; CONTRACTION FACTOR *F* FITTED FROM Λ = 1, 3)

Case / p_2	Raw λ=1	Linear	Rich.	Exp.	*f*
H2 / 0.0025	45.5	1.26	19.0	-2.19	0.956
H2 / 0.005	89.1	8.75	35.7	-4.19	0.914
H2 / 0.01	171	38.1	64.7	-7.64	0.837
LiH / 0.0025	42.5	3.31	23.0	-3.03	0.914
LiH / 0.005	81.5	16.2	43.9	-6.55	0.836
LiH / 0.01	150	58.8	82.6	-14.6	0.699
XXZ / 0.0025	420	44.6	87.9	-9.23	0.923
XXZ / 0.005	807	176	173	-16.1	0.852
XXZ / 0.01	1493	593	373	-22.8	0.728

B. Main comparison

Table IV and Fig. 3 give the main result over six seeds. Four observations hold. First, readout mitigation alone (A) leaves a bias equal to the depolarizing contraction of the full ansatz, 88.5 mHa for H_2 and 82.9 mHa for LiH, while the ansatz error of the same runs is only 0.01 and 3.12 mHa; for these shallow circuits the error is in the measurement, not in the parameters, and shot allocation alone (D) therefore changes nothing. Second, extrapolating the final estimate (B) removes most of the bias (5.57, 7.09 mHa, 50.8 mJ). Third, moving the extrapolation inside the loop with uniform shots (C) is worse than B on every benchmark, most clearly for LiH (14.8 mHa, with an ansatz error of 14.3 mHa): the two-point mitigated landscape has roughly 2.5 times the variance of the raw one, and the coordinate minimiser drifts on it. Adding Neyman allocation to the same loop (G) restores the ansatz error to 1.51 mHa and the RMSE to 6.83 mHa, so allocation is what makes in-loop mitigation usable rather than an independent saving. Fourth, the full framework (H) gives the lowest RMSE for the molecules, 3.12 mHa and 5.22 mHa, with an ansatz error of 1.01 mHa and 48.0 two-qubit gates for LiH instead of 64.0; the improvement over G is 23.5% for LiH and within the seed scatter for H_2, where the fixed ansatz has only three operators and nothing to prune.

TABLE IV

MAIN COMPARISON AT $p_2 = 5 \times 10^{-3}$ OVER SIX SEEDS (ARMS A–H, SECTION VII). ERRORS IN MHA (H_2, LIH) OR MJ (XXZ); ANSATZ = IDEAL ENERGY OF THE FINAL PARAMETERS MINUS E_0; N_{TOT} IN MILLIONS OF SHOTS; Ω IN MILLIONS OF CNOT-SHOTS; N_{CX} = TWO-QUBIT GATES OF THE FINAL ANSATZ

Arm	Error (mean ± s.d.)	RMSE	Ansatz	N_{tot}	Ω	N_{CX}
H_2						
A	88.5 ± 2.12	88.5	0.01	3.37	60.6	18
B	2.69 ± 5.35	5.57	0.01	3.37	64.2	18
C	-4.12 ± 5.00	6.15	0.10	3.78	136	18
D	89.1 ± 1.34	89.1	0.01	3.37	60.6	18
E	21.5 ± 2.39	21.6	20.6	18.3	111	0
F	0.01 ± 8.12	7.41	0.02	30.1	1743	17
G	-1.40 ± 3.45	3.44	0.03	4.14	127	18
H	2.41 ± 2.17	3.12	0.03	34.4	1623	18
LiH						
A	82.9 ± 2.61	82.9	3.12	10.4	666	64
B	-2.62 ± 7.22	7.09	3.12	10.4	678	64
C	3.66 ± 15.7	14.8	14.3	9.73	1245	64
D	80.7 ± 1.09	80.7	1.07	9.73	623	64
E	16.6 ± 1.29	16.7	16.3	21.9	142	1
F	2.54 ± 6.81	6.72	3.95	68.6	7875	50
G	-5.74 ± 4.05	6.83	1.51	10.1	1142	64
H	1.37 ± 5.52	5.22	1.01	87.6	7524	48
XXZ						
A	839 ± 16.2	839	60.3	18.4	662	36
B	36.9 ± 38.3	50.8	60.3	18.4	670	36
C	26.2 ± 76.0	74.1	58.6	18.4	1325	36
D	835 ± 15.4	835	52.1	19.3	693	36
E	236 ± 8.94	236	36.8	43.2	464	8
F	47.0 ± 23.0	51.5	57.0	61.0	1577	11
G	39.2 ± 42.1	54.9	46.9	23.0	1403	36
H	60.0 ± 43.4	71.9	66.5	45.2	800	8

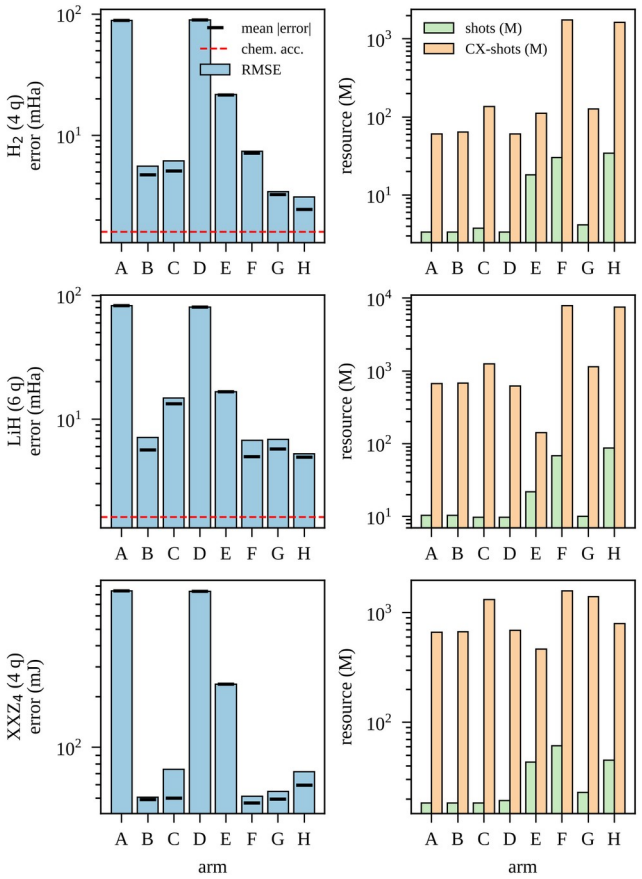


Fig. 3. Main comparison over six seeds, arms A–H as in Table IV. Left: root-mean-square energy error (bars) and mean absolute error (ticks) on a logarithmic scale; the dashed line marks chemical accuracy (1.6 mHa). Right: total shots and CNOT-shot overhead Ω.

The resource picture is the other side of this result. The full framework spends 87.6 M shots on LiH against 10.1 M for G and 10.4 M for A, because every screening evaluation is doubled by the $\lambda = 3$ circuit and the precision escalation raises the per-evaluation shots to 128 000 for the sweeps of the growing circuit. The gate-execution overhead Ω is 7.52 G CNOT-shots for H against 1.14 G for G, so the shorter circuit does not compensate for the larger number of executions. A 23.5% reduction in RMSE and a 25.0% reduction in depth are therefore bought with roughly nine times the shots. The spin chain shows the limit of the adaptive module from a different side: its greedy pool reaches an ansatz floor of 66.5 mJ (1.2% of E_0) with 4.17 operators, whereas the two-layer fixed ansatz has a noiseless floor of 12.2 mJ, and the fixed arms B and G (50.8 and 54.9 mJ) are better than H (71.9 mJ). Where the pool cannot express the ground state compactly, depth control has nothing to offer and mitigation of the fixed circuit is the better use of the budget.

C. Negative results

Arm E, adaptive growth screened on raw noisy gains, is what a practitioner obtains by running a noiseless-validated ADAPT loop on hardware. For H_2 it accepts the double excitation in 1 of six seeds and ends at 21.5 ± 2.39 mHa, which is the Hartree–Fock correlation energy: the measured gain of the 14-gate operator is negative because its depolarizing penalty (about 69.3 mHa for 14 gates at this error rate) exceeds its ideal gain of 20.6 mHa. For LiH the same arm accepts 0.5 operators on average, all of them two-gate single excitations, and reaches 16.6 ± 1.29 mHa; for XXZ it stops after 4.00 operators at 236 mJ. Noise-aware screening is therefore a prerequisite for adaptive methods on hardware, not a refinement. The second negative result is arm C above: mitigation inside the loop without allocation is harmful. The third concerns model selection (Table V). The rule that picks the extrapolation model by estimated mean-square error (RAM-auto) has an RMSE of 22.0 mHa on LiH and 17.5 mHa on H_2, against 5.22 and 3.12 mHa for the fixed exponential model; the quadratic Richardson estimator (RAM-rich) and the linear fit (RAM-lin) are between three and six times worse than the exponential model on every benchmark. With three scale factors the quadratic estimate that the selection rule uses as its bias reference is itself the noisiest and the most model-biased of the candidates, so the rule cannot separate bias from its own variance. The extrapolation model should be fixed from a noise characterisation of the device, as in (4), rather than chosen per run from the extrapolation data; this agrees with hardware practice, where the exponential form has been preferred for depolarizing-dominated processors [24], [25].

TABLE V

ABLATION OF THE FINAL EXTRAPOLATION MODEL WITHIN RAM-VQE (SIX SEEDS; ERROR AND RMSE IN MHA OR MJ; Σ = MEAN PREDICTED STANDARD DEVIATION OF THE ESTIMATOR)

Model	Error (mean ± s.d.)	RMSE	σ
H_2			
exponential (prior)	2.41 ± 2.17	3.12	4.32
linear least squares	16.8 ± 12.4	20.3	3.56
quadratic Richardson	-0.83 ± 19.9	18.2	10.8
MSE-selected	-5.28 ± 18.2	17.5	9.15
LiH			
exponential (prior)	1.37 ± 5.52	5.22	3.37
linear least squares	13.4 ± 9.18	15.8	2.57
quadratic Richardson	2.29 ± 19.3	17.8	7.68
MSE-selected	-3.47 ± 23.8	22.0	8.67
XXZ			
exponential (prior)	60.0 ± 43.4	71.9	21.0
linear least squares	73.4 ± 41.1	82.4	19.5
quadratic Richardson	105 ± 121	152	58.4
MSE-selected	116 ± 105	151	45.5

D. Optimisation drift

The per-sweep traces stored with every run (reproducibility bundle) show the mechanism behind Table IV. The fixed arms converge within four sweeps and about 10 M shots; on the raw landscape (A, B) the parameters settle a few milli-hartree above the optimum, and on the unallocated mitigated landscape (C) they wander to tens of milli-hartree. The adaptive arm (H) spends most of its budget after the escalation to 128 000 shots per evaluation, where each growth step costs several million shots; its ideal energy decreases through the accepted operators and the final sweeps leave it within about one milli-hartree of

the optimum, while its raw noisy energy rises as operators are added, which is exactly the signal that defeats arm E.

E. Noise and budget sweeps

Fig. 4 and Table VI vary p_2 from 2.5×10^{-3} to 2×10^{-2}. At the lowest rate the full framework reaches 0.98 mHa on H_2, inside chemical accuracy, and 2.67 mHa on LiH, against 2.96 mHa for the fixed mitigated ansatz. At $p_2 = 2 \times 10^{-2}$ the ordering changes. The fixed ansatz with final extrapolation stays at 4.69 mHa for H_2 but degrades to 34.0 mHa for LiH, where the 64-gate circuit folded to $\lambda = 3$ is almost fully depolarized; in-loop extrapolation of the same circuit (G) degrades to 20.0 mHa for H_2 and 14.8 mHa for LiH. The adaptive arm reaches 8.46 mHa for H_2 but fails outright for LiH (258 mHa): two of three runs hit the operator cap and one diverged to an ansatz error of 0.41 Ha, because the minimum of a trigonometric fit to noisy points is biased downward by an amount that grows with the noise, so at high error rates the screening becomes optimistic, the ansatz over-grows, and the coordinate minimiser is then driven by noise on a landscape it can no longer resolve. The regime in which the coupled framework is preferable is therefore the moderate one, $p_2 \lesssim 5 \times 10^{-3}$, which is where present superconducting and trapped-ion processors operate; at an order of magnitude worse no arm is reliable, and the fixed-ansatz final extrapolation is the safest choice.

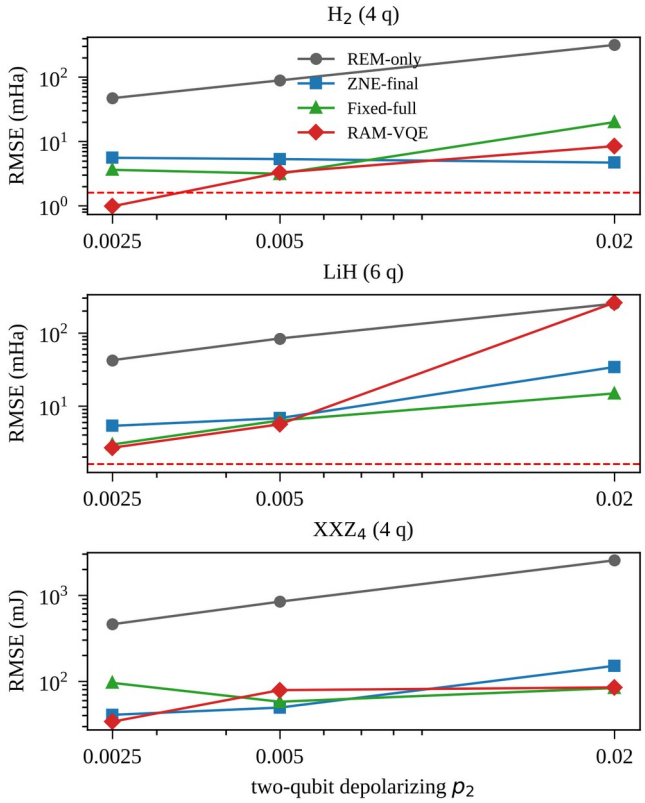


Fig. 4. Root-mean-square energy error versus two-qubit depolarizing probability (three seeds per point; six at p2 = 5 × 10−3). Dashed line: chemical accuracy.

TABLE VI

RMSE VERSUS TWO-QUBIT ERROR RATE P_2 (MHA FOR MOLECULES, MJ FOR XXZ; THREE SEEDS, SIX AT 5×10^{-3})

Case / p_2	A	B	G	H
H2 / 0.0025	47.0	5.58	3.63	0.98
H2 / 0.005	88.8	5.33	3.15	3.29
H2 / 0.02	316	4.69	20.0	8.46
LiH / 0.0025	42.1	5.32	2.96	2.67
LiH / 0.005	83.4	6.78	6.31	5.59
LiH / 0.02	250	34.0	14.8	258
XXZ / 0.0025	458	40.7	95.7	33.9
XXZ / 0.005	838	49.4	57.8	78.7
XXZ / 0.02	2529	151	83.3	84.7

Table VII varies the final-evaluation budget for LiH with the optimisation budget held fixed. The predicted standard deviation of the full framework falls from 11.5 to 2.23 mHa as the budget rises from 20 k to 500 k shots, but its RMSE only falls from 6.32 to 4.79 mHa: beyond about 100 k shots the error is dominated by a systematic residual of 4–5 mHa, made of the ansatz floor (about 1 mHa) and the remaining extrapolation bias. The fixed mitigated ansatz (G) improves from 16.4 to 4.96 mHa over the same range and overtakes H at the largest budget, since its deeper circuit yields a larger raw bias but a smaller model error once the shot noise is gone. The readout-only arm is flat at its bias, and the final-only extrapolation (B) is erratic across budgets with three seeds, which reflects its large seed-to-seed scatter in Table IV rather than a trend. The practical reading is that the final budget should be set where the predicted standard deviation reaches the expected residual bias, a few milli-hartree here, and the remainder of the shots spent on the optimisation.

TABLE VII

LIH RMSE IN MHA (PREDICTED Σ IN PARENTHESES) VERSUS FINAL-EVALUATION BUDGET B_{EVAL}; THREE SEEDS, SIX AT 200 K

B_{eval}	A	B	G	H
20k	76.0 (6.34)	4.85 (29.0)	16.4 (11.7)	6.32 (11.5)
50k	82.6 (3.90)	15.5 (16.8)	7.64 (7.60)	5.85 (6.83)
100k	80.9 (3.14)	16.8 (11.8)	7.59 (5.39)	6.78 (4.99)
200k	82.9 (2.06)	7.09 (8.71)	6.83 (3.86)	5.22 (3.31)
500k	81.8 (1.35)	8.37 (5.48)	4.96 (2.45)	4.79 (2.23)

IX. COMPARATIVE ANALYSIS

Table VIII places the framework against the methods it combines, on the criteria a hardware user faces. A numerical comparison with published hardware results is not attempted because the noise models differ; the comparison is structural, and Section VIII shows that the structural advantage is conditional on the noise regime and the pool. What distinguishes RAM-VQE is that each decision is taken with the cost and noise consequences of the other two modules in view.

TABLE VIII
STRUCTURAL COMPARISON WITH EXISTING APPROACHES

Approach	Bias / shots	Depth / budget	Screening
Raw VQE + readout mitigation	Readout only / uniform	Fixed / none	n/a
ZNE on final estimate [5]–[7]	Final only / uniform	Fixed / final circuits	n/a
Rosalin, iCANS [10], [11]	None / adaptive	Fixed / optimisation only	n/a
ADAPT-VQE [12], [13]	None / uniform	Adaptive / none	Noiseless gradient
vnCDR [19]	Learned + ZNE / uniform	Fixed / excludes training	n/a
RAM-VQE (this work)	In-loop + final ZNE / joint Neyman	Adaptive / one ledger	Mitigated gain, escalation

X. KEY OUTCOMES

- A unified treatment of extrapolation, allocation and growth under one shot ledger, with two propositions that give the allocation gain and the variance amplification in closed form, reproduced numerically (Table III, Section VIII-A).
- On a gate-level depolarizing model at $p_2 = 5 \times 10^{-3}$, an RMSE of 5.22 mHa for stretched LiH and 3.12 mHa for H_2 (from 82.9 and 88.5 mHa with readout mitigation alone), with 25% fewer two-qubit gates for LiH, at a shot cost roughly nine times that of mitigating a fixed ansatz, which itself reaches 6.83 mHa.
- Evidence that the exponential, depolarizing-motivated extrapolation model should be fixed a priori: it is three to six times more accurate than linear and quadratic fits on every benchmark, and per-run model selection from three scale factors fails.
- Three interaction effects that single-component studies cannot see: adaptive growth must be screened on mitigated energies or it stalls; in-loop mitigation needs variance-proportional allocation or it degrades the optimiser; and at high error rates the optimistic bias of noisy screening makes adaptive growth over-grow.
- For practice: the coupled framework is the right choice when the operator pool is compact and the device is in the moderate-noise regime; when the pool is ansatz-limited (the spin chain here) or the error rate is an order of magnitude worse, mitigating a fixed circuit is both cheaper and more accurate.

XI. CONCLUSION AND FUTURE SCOPE

RAM-VQE couples zero-noise extrapolation, variance-proportional shot allocation and noise-screened adaptive circuit growth so that each is operated with knowledge of the other two. On two molecular benchmarks under a validated gate-level noise model it gives the lowest energy error of eight configurations at the default error rate and reduces the two-qubit depth of the LiH ansatz by a quarter, at a substantially higher shot cost; on the spin chain, where the operator pool is the limit, the fixed mitigated ansatz is better. The study's more durable contributions are the interaction effects it isolates, the closed-form allocation and variance results, and the evidence that the extrapolation model should be fixed from the noise physics rather than selected from the data.

The limitations are explicit. All results are simulated; the noise model is depolarizing with symmetric readout error, which favours the exponential extrapolation and omits coherent errors, crosstalk and drift that real devices exhibit; the sampling overhead of any mitigation scheme is also bounded below by the exponential cost established in [26], so the gains reported here cannot be extrapolated to large depth. The noise resilience of the optimiser itself has been studied in [27] and is consistent with the drift reported in Table IV. The benchmarks are at most six qubits, and the greedy operator pool used for the spin chain leaves an ansatz floor that the fixed two-layer circuit does not. The shot cost of the framework is several times that of mitigating a fixed ansatz, and the precision escalation makes the late growth steps expensive. The optimisation layer is implemented in Python (SciPy) rather than MATLAB; a MATLAB port of the allocation and extrapolation routines is straightforward but has not been produced or verified here.

Future work follows from these limits. The exponential model should be replaced by a fitted noise model learned from a sparse Pauli–Lindblad characterisation [8], which would also allow probabilistic error cancellation to be budgeted inside the same ledger. The screening threshold z, the patience and the escalation schedule should be set from a formal sequential-testing bound rather than fixed, and the downward bias of the reconstructed minimum at high noise should be corrected analytically before screening. Larger active spaces and tapered Hamiltonians would test whether the depth advantage persists when double excitations dominate the pool. Finally, the framework should be executed on hardware through Qiskit [28] or PennyLane device plugins, with the simulator of this paper serving as the digital twin that predicts the shot and gate budgets before a job is submitted.

ACKNOWLEDGMENT

The authors thank the Research and Development Division of Tekstein Solutions Private Limited for computing resources. No external funding was received.

REFERENCES

[1] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, "A variational eigenvalue solver on a photonic quantum processor," *Nature Commun.*, vol. 5, p. 4213, 2014.

[2] J. Tilly, H. Chen, S. Cao, D. Picozzi, K. Setia, Y. Li, E. Grant, L. Wossnig, I. Rungger, G. H. Booth, and J. Tennyson, "The variational quantum eigensolver: A review of methods and best practices," *Phys. Rep.*, vol. 986, pp. 1–128, 2022.

[3] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, "Variational quantum algorithms," *Nature Rev. Phys.*, vol. 3, pp. 625–644, 2021.

[4] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K.

Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik, "Noisy intermediate-scale quantum algorithms," *Rev. Mod. Phys.*, vol. 94, p. 015004, 2022.

[5] K. Temme, S. Bravyi, and J. M. Gambetta, "Error mitigation for short-depth quantum circuits," *Phys. Rev. Lett.*, vol. 119, p. 180509, 2017.

[6] Y. Li and S. C. Benjamin, "Efficient variational quantum simulator incorporating active error minimization," *Phys. Rev. X*, vol. 7, p. 021050, 2017.

[7] T. Giurgica-Tiron, Y. Hindy, R. LaRose, A. Mari, and W. J. Zeng, "Digital zero noise extrapolation for quantum error mitigation," in *Proc. IEEE Int. Conf. Quantum Comput. Eng. (QCE)*, 2020, pp. 306–316.

[8] E. van den Berg, Z. K. Minev, A. Kandala, and K. Temme, "Probabilistic error cancellation with sparse Pauli–Lindblad models on noisy quantum processors," *Nature Phys.*, vol. 19, pp. 1116–1121, 2023.

[9] D. Wecker, M. B. Hastings, and M. Troyer, "Progress towards practical quantum variational algorithms," *Phys. Rev. A*, vol. 92, p. 042303, 2015.

[10] A. Arrasmith, L. Cincio, R. D. Somma, and P. J. Coles, "Operator sampling for shot-frugal optimization in variational algorithms," arXiv:2004.06252, 2020.

[11] J. M. Kübler, A. Arrasmith, L. Cincio, and P. J. Coles, "An adaptive optimizer for measurement-frugal variational algorithms," *Quantum*, vol. 4, p. 263, 2020.

[12] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, "An adaptive variational algorithm for exact molecular simulations on a quantum computer," *Nature Commun.*, vol. 10, p. 3007, 2019.

[13] H. L. Tang, V. O. Shkolnikov, G. S. Barron, H. R. Grimsley, N. J. Mayhall, E. Barnes, and S. E. Economou, "qubit-ADAPT-VQE: An adaptive algorithm for constructing hardware-efficient ansätze on a quantum processor," *PRX Quantum*, vol. 2, p. 020310, 2021.

[14] P. J. J. O'Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, et al., "Scalable quantum simulation of molecular energies," *Phys. Rev. X*, vol. 6, p. 031007, 2016.

[15] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O'Brien, "Quantum error mitigation," *Rev. Mod. Phys.*, vol. 95, p. 045005, 2023.

[16] V. Verteletskyi, T.-C. Yen, and A. F. Izmaylov, "Measurement optimization in the variational quantum eigensolver using a minimum clique cover," *J. Chem. Phys.*, vol. 152, p. 124114, 2020.

[17] W. J. Huggins, J. R. McClean, N. C. Rubin, Z. Jiang, N. Wiebe, K. B. Whaley, and R. Babbush, "Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers," *npj Quantum Inf.*, vol. 7, p. 23, 2021.

[18] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, "Noise-induced barren plateaus in variational quantum algorithms," *Nature Commun.*, vol. 12, p. 6961, 2021.

[19] A. Lowe, M. H. Gordon, P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, "Unified approach to data-driven quantum error mitigation," *Phys. Rev. Res.*, vol. 3, p. 033098, 2021.

[20] P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, "Error mitigation with Clifford quantum-circuit data," *Quantum*, vol. 5, p. 592, 2021.

[21] M. Ostaszewski, E. Grant, and M. Benedetti, "Structure optimization for parameterized quantum circuits," *Quantum*, vol. 5, p. 391, 2021.

[22] R. LaRose, A. Mari, S. Kaiser, P. J. Karalekas, A. A. Alves, P. Czarnik, M. El Mandouh, M. H. Gordon, Y. Hindy, A. Robertson, et al., "Mitiq: A software package for error mitigation on noisy quantum computers," *Quantum*, vol. 6, p. 774, 2022.

[23] V. Bergholm, J. Izaac, M. Schuld, C. Gogolin, S. Ahmed, V. Ajith, M. S. Alam, G. Alonso-Linaje, B. AkashNarayanan, A. Asadi, et al., "PennyLane: Automatic differentiation of hybrid quantum-classical computations," arXiv:1811.04968, 2022.

[24] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, "Error mitigation extends the computational reach of a noisy quantum processor," *Nature*, vol. 567, pp. 491–495, 2019.

[25] Y. Kim, A. Eddins, S. Anand, K. X. Wei, E. van den Berg, S. Rosenblatt, H. Nayfeh, Y. Wu, M. Zaletel, K. Temme, and A. Kandala, "Evidence for the utility of quantum computing before fault tolerance," *Nature*, vol. 618, pp. 500–505, 2023.

[26] R. Takagi, S. Endo, S. Minagawa, and M. Gu, "Fundamental limits of quantum error mitigation," *npj Quantum Inf.*, vol. 8, p. 114, 2022.

[27] E. Fontana, N. Fitzpatrick, D. Muñoz Ramo, R. Duncan, and I. Rungger, "Evaluating the noise resilience of variational quantum algorithms," *Phys. Rev. A*, vol. 104, p. 022403, 2021.

[28] A. Javadi-Abhari, M. Treinish, K. Krsulich, C. J. Wood, J. Lishman, J. Gacon, S. Martiel, P. D. Nation, L. S. Bishop, A. W. Cross, B. R. Johnson, and J. M. Gambetta, "Quantum computing with Qiskit," arXiv:2405.08810, 2024.