Hierarchical Progressive Optimization for Multi-Qubit Pauli Noise Modeling

Xiangyu Ge(葛翔宇)[1], Jiafei Ge(葛佳菲)[2], Shengmei Zhao(赵生妹)[1], Le Wang(王乐)[1], and Anqi Zhang(张安琪)[1,*]

[1]Nanjing University of Posts and Telecommunications, Nanjing 210003, China

[2]School of Basic Medical Sciences & School of Public Health, Faculty of Medicine, Yangzhou University, Yangzhou 225009, PR China

August 21, 2026

Abstract

Quantum Noise Characterization (QNC) is indispensable for benchmarking and mitigating errors in Noisy Intermediate-Scale Quantum (NISQ) devices. However, traditional Quantum Process Tomography (QPT) suffers from an exponential parameter explosion, severely hindering its scalability. In this paper, we propose a Hierarchical Progressive Optimization (HPO) framework to efficiently extract high-order spatial crosstalk in multi-qubit systems. The complexity analysis shows that the combinatorial projection mask reduces the required number of Pauli transfer matrix (PTM) elements from $\mathcal{O}(16^N)$ to $\mathcal{O}(N^2 3^N)$. Numerical simulations on a 10-qubit HHL circuit achieve a fidelity of 0.9381 with the HPO method, compared to 0.7431 obtained using global depolarizing-noise mitigation.

Keywords: quantum noise characterization; Pauli noise modeling; hierarchical progressive optimization; quantum error mitigation; spatial crosstalk; quantum process tomography

PACS: 03.65.Wj, 03.67.Lx, 03.65.Yz, 03.67.-a

1. Introduction

In the Noisy Intermediate-Scale Quantum (NISQ) era, hardware imperfections and decoherence cause the execution outcomes of quantum algorithms on physical devices to deviate significantly from ideal simulations [1]. Such quantum noise not only degrades the fidelity of multi-qubit states in operations like quantum teleportation [2], but also poses a critical bottleneck for quantum metrology [3] and error mitigation tasks [4,5]. Recent studies on the fundamental limits of quantum error mitigation further emphasize that without highly accurate physical noise models, achieving practical quantum advantage remains extremely difficult [6]. Therefore, an accurate and scalable quantum noise model serves as the indispensable bridge connecting theoretical quantum algorithms with real-world physical hardware, fundamentally impacting the evaluation of quantum architectures and performance prediction.

*Corresponding author. E-mail: zhangaq@njupt.edu.cn

To characterize these complex quantum channels, various tomographic methods have been proposed. Full Quantum Process Tomography (QPT) and Pauli Transfer Matrix (PTM) reconstruction provide a mathematically exhaustive description of the noise maps [7,8]. Although advanced classical shadow techniques have been introduced to alleviate measurement overhead [9], these full-characterization methods still severely lack scalability due to exponential parameter explosion, requiring $\mathcal{O}(16^N)$ measurements for an N-qubit system. At the other extreme, simplified phenomenological models (e.g., independent depolarizing channels) are highly scalable but inherently fail to capture the ubiquitous spatial crosstalk in contemporary quantum processors. Early studies identified crosstalk as a primary source of multi-qubit errors [10], significantly manifesting in both readout processes [11,12] and microwave-driven gate operations [13]. Recent physical experiments explicitly demonstrate that simultaneous gate operations on superconducting qubits induce profound irreducible multi-body crosstalk ($\mathcal{E}_{ij} \neq \mathcal{E}_i \otimes \mathcal{E}_j$) [14], which can only be diagnosed via complex out-of-time-order correlators (OTOCs) [15] and suppressed by specialized hardware shielding structures [16]. Consequently, existing methods either suffer from parameter explosion or are overly simplified to express true physical spatial coupling. In particular, there is a distinct lack of a recursive extension mechanism that clearly answers "which parameters can be directly inherited from low-qubit systems, and which high-order parameters require additional optimization."

The core contradiction in current noise modeling lies in the exponentially growing dimensionality of full PTMs, in contrast to the severely constrained measurement budget on physical quantum hardware. Therefore, a structured modeling approach that can utilize low-qubit samples and recursively extend to high-qubit systems is critically required. To break through this bottleneck, recent studies have actively explored approximation schemes based on physical spatial structures. For instance, graph attention networks have been utilized to capture nearest-neighbor interactions in many-body systems [17], and local receptive fields have been applied to truncate redundant long-range correlations in quantum error correction codes [18,19]. These works reveal that discarding the global fully-connected assumption and leveraging locality in quantum systems has emerged as a critical strategy to mitigate the dimensionality curse. However, at the hardware level, mechanisms such as maintaining low-noise stability [20] and leveraging opposite-sign anharmonicity [21] or crosstalk-robust quantum control [22] reveal that quantum noise exhibits stable locality accompanied by fragile high-order correlations. Few studies have explicitly analyzed information retention capacity under Pauli weight and Hamming distance constraints. Specifically: (1) the dimension of the Pauli basis grows exponentially with the number of qubits; (2) quantum noise comprises not only single-body errors but also two-qubit crosstalk and multi-body correlations; and (3) hardware measurement costs are prohibitively high and sample sizes are strictly limited, precluding reliance on full parameter reconstruction. As the system scales, high-weight multi-qubit Pauli correlations are highly susceptible to decoherence and physically collapse toward zero—a phenomenon we define as *Pauli dilution*. Ultimately, quantum noise intrinsically presents a duality of robust "local coupling features" and fragile "high-order correlation effects." This physical constraint requires that a scalable, reliable noise model avoid both overly simplistic uncorrelated approximations and unrestricted full-dimensional expansion in the exponential Hilbert space.

To address these challenges, we propose a Hierarchical Progressive Optimization (HPO) framework for sparse Pauli noise modeling, tailored for characterizing multi-qubit quantum circuits. Unlike traditional approximation methods that optimize a truncated global space simultaneously, our framework physically decouples

the noise into a local baseline and high-order residuals. Specifically, we first establish a 2-qubit foundational model constrained by a Hamming distance mask ($d \leq 2$) to capture dominant local errors. When extending to N-qubit systems ($N > 2$), the framework strictly freezes the pre-trained low-order parameters inherited from the baseline.

The contributions of this paper are fourfold: 1) a Hamming distance mask is introduced for sparse Pauli noise modeling, which preserves dominant local errors and filters irrelevant Pauli correlations. 2) a hierarchical freezing mechanism is developed, in which the pre-trained low-order parameters are frozen when the system is extended to larger multi-qubit systems. 3) a topology-aware parameter sharing scheme is constructed, in which the optimized 2-qubit foundational model is reused on the global topology graph. 4) a recursive progressive optimization framework is established, in which the noise model is learned from the foundational baseline to the high-order residuals.

The proposed HPO framework provides a reliable, scalable paradigm for evaluating high-fidelity multi-qubit gate schemes [23], robust quantum gate optimizations [24], and reliable conversion schemes for hybrid entangled states in error-predicted devices [25].

2. Proposed Sparse Pauli Noise Modeling Methodology

To address the exponential parameter explosion inherent in full Quantum Process Tomography (QPT) while accurately capturing spatial crosstalk, we propose a Hierarchical Progressive Optimization (HPO) framework. The core philosophy is to transition from global unconstrained optimization to a strict physics-informed subspace search. As schematically illustrated in Fig. 1, the detailed formulation is presented in the following four subsections.

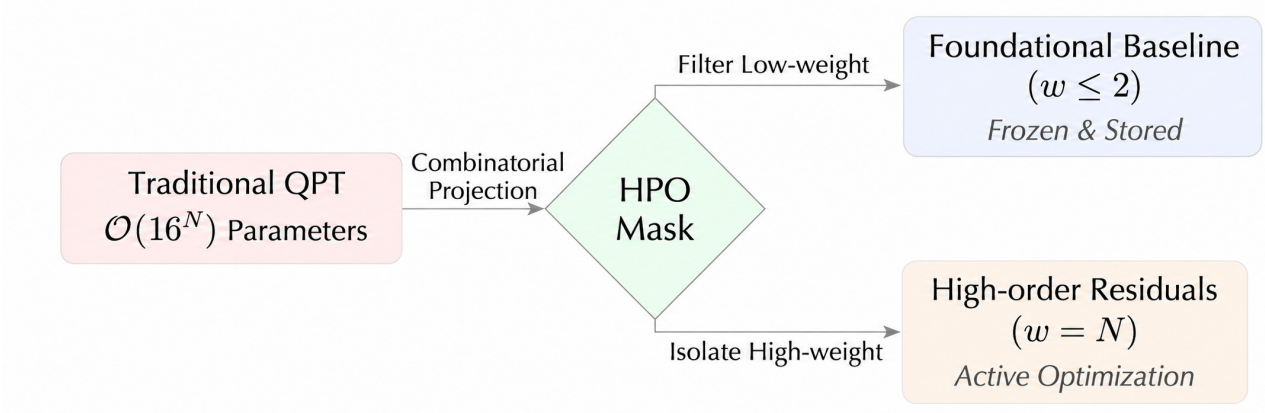


Figure 1: Schematic diagram of the proposed Hierarchical Progressive Optimization (HPO) framework. By introducing a deterministic combinatorial mask, the framework successfully decouples the intractable multi-body interactions into a frozen foundational baseline and an active optimization subspace, effectively avoiding the parameter explosion.

2.1. Mathematical Formulation of the 2-qubit Foundational Baseline

The HPO framework stems from systematic characterization of the elementary two-qubit coupling unit, the core building block for large-scale quantum systems. For an N-qubit system, the complete operator space is spanned by the Pauli basis $\mathcal{P}^{\otimes N} = \{I, X, Y, Z\}^{\otimes N}$. Under the Markovian approximation, any completely positive and trace-preserving (CPTP) quantum channel $\mathcal{E}$ can be uniquely mapped into a real-valued Pauli Transfer Matrix (PTM) $\mathcal{R}$, with its elements defined by the Hilbert-Schmidt inner product [8]:

$$\mathcal{R}_{ij} = \frac{1}{2^N} \mathrm{Tr}\left[P_i \mathcal{E}(P_j)\right]. \tag{1}$$

For a localized 2-qubit subsystem, the full PTM resides in a 16×16 dimensional continuous space, containing 256 matrix entries. However, in superconducting architectures, physical noise is predominantly highly localized, primarily driven by single-qubit relaxation (T_1), dephasing (T_2), and adjacent static ZZ-crosstalk. Simultaneously optimizing all 256 PTM entries with limited experimental measurements may cause severe overfitting to unphysical high-order entangled noise.

To enforce physical sparsity and isolate these dominant local errors, we introduce a Hamming distance mask $M^{(2)}$. Let $w(P)$ denote the Pauli weight (the number of non-identity operators in the string), and $D_H(P_i, P_j) = \sum_k \mathbb{I}(P_i^{(k)} \neq P_j^{(k)})$ denote the Hamming distance underlying the transition probability. This mask plays a central role in our framework: superconducting architectures demonstrate that simultaneous physical errors beyond two-body Markovian interactions are exponentially suppressed [26]. By restricting the Hamming distance, the mask effectively filters out unphysical noise artifacts and confines parameter dynamics to local physical processes. The foundational model is structurally constrained as:

$$\mathcal{R}_{\text{base}}^{(2)} = \mathbb{I} + \left(\Delta\mathcal{R} \odot M^{(2)}\right), \tag{2}$$

$$\text{s.t.} \quad M_{ij}^{(2)} = \begin{cases} 1, & \text{if } \left(w(P_i) \leq 2 \vee w(P_j) \leq 2\right) \wedge D_H(P_i, P_j) \leq 2 \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

where $\mathbb{I}$ represents the ideal identity channel, and $\odot$ denotes the Hadamard product. By applying this hard-truncation mask, this foundational phase precisely captures essential single-qubit decoherence and neighboring two-qubit crosstalk. The optimized parameters obtained in this stage form a stable, physically grounded baseline for subsequent high-order extensions.

2.2. Topology-Aware Recursive Extension via Kronecker Mapping

While the foundational model accurately captures local dynamics, unconstrained full-dimensional PTM reconstruction ($\mathcal{O}(16^N)$) for an N-qubit processor ($N > 2$) incurs an exponential parameter explosion, making it experimentally infeasible. To address this, we project the optimized foundational 2-qubit parameters onto the macroscopic system topology graph $G = (V, E)$, where V denotes the physical qubits and E denotes valid coupling edges.

We define a physical operator $\Lambda_e(\cdot)$ that maps a localized 2-qubit PTM on a specific edge $e = (u, v)$ into the global N-qubit Hilbert space. Under the assumption that spontaneous long-range crosstalk is absent without physical graph connectivity, the global sparse base PTM $\tilde{\mathcal{R}}_{\text{base}}^{(N)}$ is recursively constructed from a tensor product

sequence:

$$\tilde{\mathcal{R}}_{\text{base}}^{(N)} = \bigotimes_{e=(u,v)\in E} \Lambda_e\left(\mathcal{R}_{\text{base},e}^{(2)}\right), \quad \text{where } \Lambda_e(\mathcal{R}) = \mathcal{R} \otimes I_{V\setminus\{u,v\}}. \tag{4}$$

This recursive spatial extension projects the empirically verified local error characteristics onto the expansive N-qubit space. By strictly padding the non-interacting spectator qubits with the identity operator I, we ensure that no artificial, non-local crosstalk artifacts are mathematically generated during the scaling process. This step acts as a mean-field approximation of the global noise landscape.

2.3. Hierarchical Progressive Optimization with Gradient Projection

Although the recursively extended base model efficiently captures topological errors, it inherently underfits genuine high-order multi-body correlations that emerge during simultaneous multi-qubit gate operations. Our approach centers on the Hierarchical Progressive Optimization (HPO) framework, which characterizes weak spatial crosstalk via rigorous gradient projection.

We implement a topology-aware parameter sharing strategy to decouple the loss landscape. For any Pauli basis where the weight $w(P) < N$, the structural parameters are strictly frozen to their pre-trained values inherited from lower-order stages, denoted as $\mathcal{W}_{\text{frozen}}$. To capture the fragile N-body correlations without suffering from the barren plateau problem, the theoretical PTM residual space is confined to the N-th-order mask-selected subspace.

We mathematically define the dynamic progressive mask $M^{(N)}$ as a hard physical truncation filter:

$$M_{ij}^{(N)} = \begin{cases} 1, & \text{if } \big(w(P_i) = N \vee w(P_j) = N\big) \wedge D_H(P_i, P_j) \leq 2 \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

The effective global transformation matrix for the N-th order characterization is formulated as $\mathcal{R}_{\text{eff}}^{(N)} = \mathbb{I} + \mathcal{W}_{\text{frozen}} + \big(\Delta\mathcal{R}_{\text{res}} \odot M^{(N)}\big)$. Crucially, during the backpropagation process, the gradient of the tomographic loss function $\mathcal{L}$ with respect to the incremental residual tensor $\Delta\mathcal{R}_{\text{res}}$ is mathematically projected onto the defined subspace:

$$\theta_{t+1} = \theta_t - \eta\left(\nabla_{\Delta\mathcal{R}}\mathcal{L} \odot M^{(N)}\right). \tag{6}$$

The element-wise projection restricts the theoretical PTM residual entries to the N-th-order mask-selected space. This explicit orthogonal projection prevents the optimizer from wandering into the redundant 16^N flat manifold, thereby successfully avoiding complex, non-convex local minima and ensuring stable, machine-precision convergence for high-order residuals.

2.4. Combinatorial Complexity Analysis and Dimensionality Reduction

The primary bottleneck of full Quantum Process Tomography (QPT) is the curse of dimensionality. Let $\mathcal{P}^{\otimes N}$ be the N-qubit Pauli group. The required PTM-element count scales exponentially as $\mathcal{D}_{\text{full}} = 16^N$. The HPO framework systematically mitigates this issue through two-phase dimensionality reduction.

Phase 1: Linear Scaling of the Foundational Baseline. The localized foundational model restricts active parameters strictly to physical coupling edges $e \in E$. For a specific edge, the distance mask $d \leq 2$ bounds

the parameter count to a maximum of $16^2 = 256$. By mapping this onto the global topology $G = (V, E)$, the base network complexity is bounded by the linear sum over all physical edges:

$$\mathcal{D}_{\text{base}} \leq \sum_{e \in E} 256 = 256|E|. \tag{7}$$

For contemporary superconducting processors with bounded-degree connection topologies (e.g., a 1D chain where $|E| = N - 1$), the number of edges scales linearly with the number of qubits. Thus, the optimization complexity of the foundational baseline exhibits **linear growth for bounded-degree connection topologies**: $\mathcal{D}_{\text{base}} \sim \mathcal{O}(N)$.

Phase 2: Combinatorial Bound of Progressive Residuals. During the N-th stage of progressive optimization, attempting to capture full multi-body correlations would reintroduce the $\mathcal{O}(16^N)$ scaling. However, the progressive mask $M^{(N)}$ isolates only the critical N-th order subspace, defined by the logical constraint: $\big(w(P_i) = N \vee w(P_j) = N\big) \wedge D_H(P_i, P_j) \leq 2$.

We mathematically derive the exact active parameter count, $K_{\text{res}}^{(N)}$, using combinatorial analysis. Let $\mathcal{S}_N$ denote the set of Pauli strings with a full weight of $w = N$. Since these strings contain no identity operators (I), they are composed entirely of X, Y, Z, yielding $|\mathcal{S}_N| = 3^N$.

For each fixed full-weight Pauli string $P_i \in \mathcal{S}_N$, the number of target strings P_j satisfying $D_H(P_i, P_j) \leq 2$ is $1 + 3N + 9\binom{N}{2}$. The total number of such P_j over all 3^N choices of P_i is the following:

$$\begin{aligned} |A| &= |\mathcal{S}_N| \times \left[1 + \binom{N}{1} \times 3 + \binom{N}{2} \times 3^2\right] \\ &= 3^N \left(1 + 3N + 9\binom{N}{2}\right). \end{aligned} \tag{8}$$

By symmetry, the case where $P_j \in \mathcal{S}_N$ yields the same size $|B| = |A|$. To avoid double counting, we subtract the intersection $|A \cap B|$, where both $P_i, P_j \in \mathcal{S}_N$. In this intersection, any altered position must strictly change to another non-identity operator (only 2 choices per position, as I is excluded):

$$\begin{aligned} |A \cap B| &= 3^N \left[1 + \binom{N}{1} \times 2 + \binom{N}{2} \times 2^2\right] \\ &= 3^N \left(1 + 2N + 4\binom{N}{2}\right). \end{aligned} \tag{9}$$

By applying $K_{\text{res}}^{(N)} = 2|A| - |A \cap B|$, the expression for the high-order optimization complexity is obtained:

$$K_{\text{res}}^{(N)} = 3^N \left[1 + 4N + 14\binom{N}{2}\right], \tag{10}$$

thus, $K_{\text{res}}^{(N)}$ characterizes the size of the mask-selected PTM space for an N-qubit system. It should be noted that a general completely positive and trace-preserving (CPTP) channel acting on a d-dimensional Hilbert space contains $d^4 - d^2$ independent real degrees of freedom [27,28]. Therefore, for the two-qubit case, the foundational model contains $d^4 - d^2 = 256 - 16 = 240$ independent degrees of freedom. Hence, 240 serves as the starting point for the progressive optimization in the HPO framework.

Since directly optimizing this space becomes impractical as N increases, HPO freezes all parameters learned in previous stages and only optimizes the newly introduced parameters associated with the added qubit q_N. For the endpoint-grown topology, the corresponding trainable parameter count is denoted by

$$\Delta K_N = \sum_{w=1}^{N} 3^w, \qquad N > 2. \tag{11}$$

The resulting PTM-element counts are summarized in Table 1, where $C_{\text{mask}} = 1 - K_{\text{res}}^{(N)}/\mathcal{D}_{\text{full}}$ denotes the corresponding compression ratio.

Table 1: Comparison of the required number of PTM-entries under the CPTP constraint during 2-to-10-qubit progressive training of a 10-qubit system.

N	$\mathcal{D}_{\text{full}}$	$K_{\text{res}}^{(N)}$	C_{mask}	ΔK_N
2	240	240	–	240
3	4,032	1,485	63.17%	39
4	65,280	8,181	87.47%	120
5	1,047,552	39,123	96.27%	363
6	16,773,120	171,315	98.98%	1,092
7	268,419,072	706,401	99.74%	3,279
8	4,294,901,760	2,788,425	99.935%	9,840
9	68,719,214,592	10,648,503	99.9845%	29,523
10	1,099,510,579,200	39,621,879	99.9964%	88,572

3. Numerical Simulation and Results

Numerical simulations to evaluate the performance and scalability of the Hierarchical Progressive Optimization (HPO) framework are performed here.

3.1. Simulation Setup

To ensure stable and extremely precise optimization, an Adam optimizer is employed and initialized with a learning rate of $\eta = 0.002$, coupled with a Cosine Annealing learning rate scheduler ($\eta_{\min} = 10^{-5}$). To quantify the accuracy of the noise parameter extraction during the QPT process, we use the Mean Squared Error (MSE) between the predicted expectation values and the ground-truth:

$$\text{MSE} = \frac{1}{|\mathcal{S}|} \sum_{k=1}^{|\mathcal{S}|} \left(\langle O_k \rangle_{\text{pred}} - \langle O_k \rangle_{\text{true}} \right)^2 , \tag{12}$$

where $\mathcal{S}$ represents the selected combinatorial subset of high-weight Pauli observables. Furthermore, to evaluate the algorithmic impact of the noise model, the state fidelity $\mathcal{F}$ between the target density matrix ρ_{ideal} and the noise-injected density matrix σ_{noisy} is computed [29]:

$$\mathcal{F}(\rho_{\text{ideal}}, \sigma_{\text{noisy}}) = \left(\text{Tr} \sqrt{\sqrt{\rho_{\text{ideal}}} \sigma_{\text{noisy}} \sqrt{\rho_{\text{ideal}}}} \right)^2 . \tag{13}$$

To make HPO training computationally tractable, we further restrict the stage-wise trainable parameter space based on ΔK_N in Table 1. For $N \leq 5$, the trainable count equals ΔK_N. For $N > 5$, a Pauli-weight cutoff of $w \leq 3$ limits each stage to $3 + 9 + 27 = 39$ newly introduced parameters, while all parameters inherited from stage $N - 1$ remain fixed.

The simulations are executed on a GPU-accelerated quantum machine learning framework leveraging PennyLane [30], which enables high-efficiency backpropagation through complex quantum circuits.

3.2. Progressive Optimization of Pauli-Weight Noise Components

The learning dynamics of the HPO framework under varying qubit scales $N \in \{2, 3, 4, 5\}$ is validated firstly. The active parameter space is strictly bounded by the combinatorial progressive mask, preventing gradient dilution.

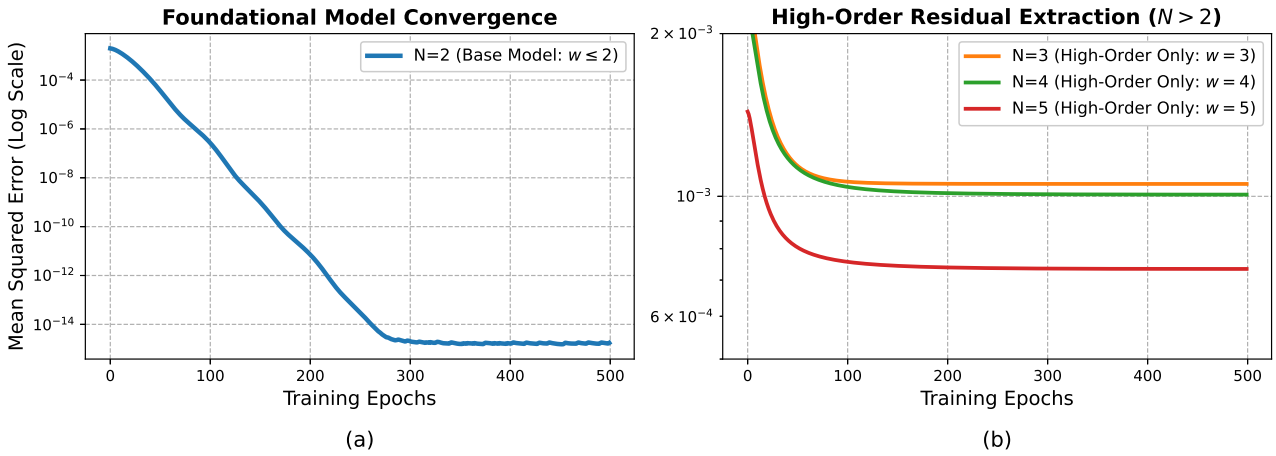


Figure 2: Convergence behavior of the HPO framework. (a) The convergence of the foundational baseline of the low-order terms; (b) The convergence of higher-order residual terms.

Figure 2 illustrates convergence behaviors induced by the hierarchical masking strategy of the HPO framework in different Pauli terms. As shown in Fig. 2(a), the HPO framework, which optimizes all Pauli terms with weight $w \leq 2$, exhibits a clear convergence trend, and eventually reaches an MSE of approximately 10^{-14}. This result indicates that the dominant low-order local noise components are sufficiently represented by the $w \leq 2$ Pauli interaction terms. In contrast, Fig. 2(b) reveals the learning behavior of the higher-order residual components with $N > 2$ under the HPO framework. After the previously optimized low-order baseline is fixed, the projection mask restricts subsequent optimization to the remaining higher-weight Pauli sectors. As a result, the residual learning trend remains smooth and stable, suggesting that the HPO framework can separate the well-captured local baseline from the weaker multi-body corrections. This behavior supports the hierarchical decomposition: the foundational model first absorbs the dominant local Pauli error structure, while the residual stages focus on higher-order correlations without destabilizing the already optimized baseline.

It should be emphasized that the larger MSEs in Fig. 2(b) do not imply reduced fidelity. In HPO, the low-order components are fixed before optimizing the weak high-order residuals. Although their small amplitudes make them more difficult to estimate, their limited contribution to the overall noise channel leaves the algorithm-level fidelity essentially unchanged.

3.3. Quantum Error Mitigation on 10-Qubit HHL Algorithm

To demonstrate the effectiveness of the HPO framework, the proposed method is evaluated through a circuit-level numerical simulation of the 10-qubit HarrowHassidimLloyd (HHL) algorithm [31], which requires accurate quantum phase estimation (QPE) and is highly sensitive to unmitigated phase accumulation and high-order crosstalk. As depicted in the circuit architecture of Fig. 3, the simulation uses 1 ancilla qubit, 8 clock qubits for precision QPE, and 1 target qubit. Four benchmarking settings are evaluated in terms of the target-state fidelity $\mathcal{F}$: an ideal noiseless state, the unmitigated noisy simulation, QEM via the traditional global depolarizing model, and QEM via the HPO framework.

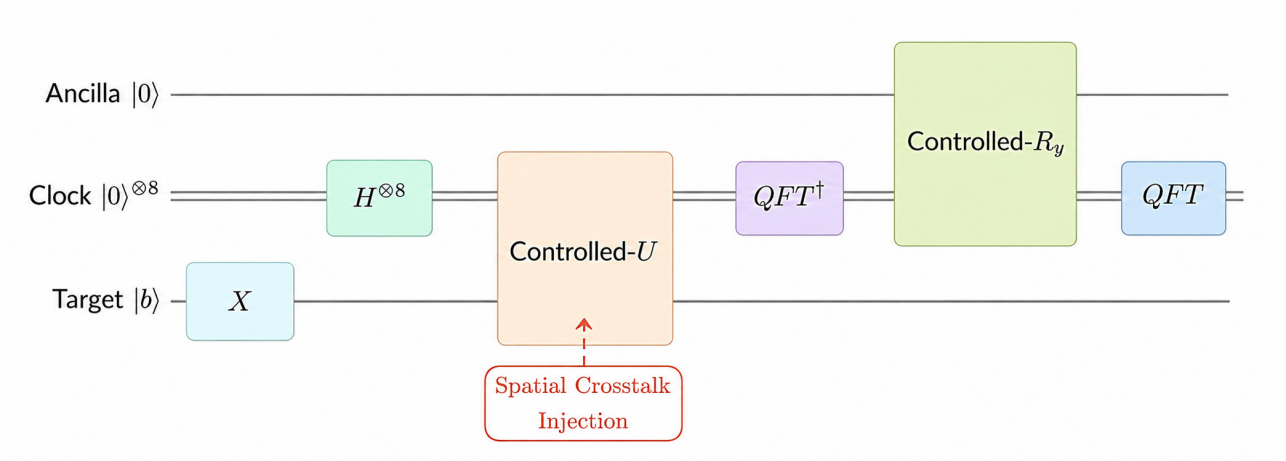


Figure 3: Quantum circuit architecture of the 10-qubit Harrow-Hassidim-Lloyd (HHL) algorithm used for spatial crosstalk mitigation. The circuit comprises 1 ancilla qubit, 8 clock qubits (denoted by the double-line bus) for high-precision quantum phase estimation, and 1 target qubit. The red dashed arrow indicates the injection points of unmitigated dynamic multi-body crosstalk during dense controlled-phase operations.

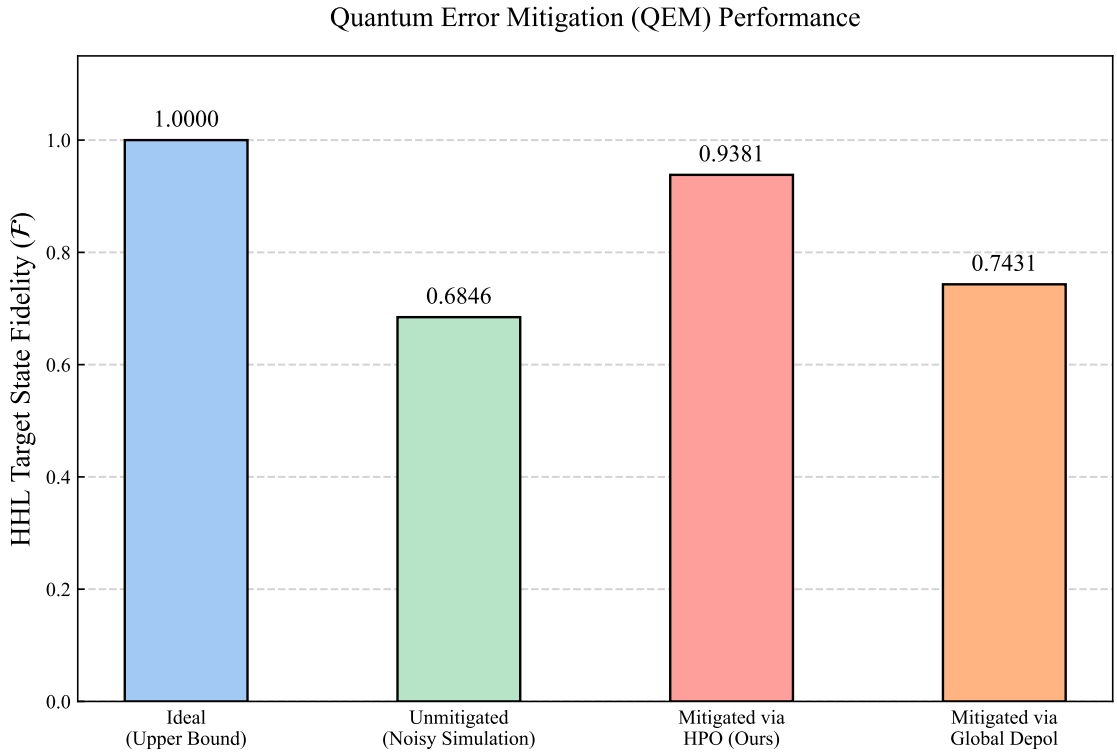


Figure 4: Algorithmic state fidelity evaluation of the 10-qubit HHL algorithm under various Quantum Error Mitigation (QEM) strategies.

As quantitatively illustrated in Fig. 4, the HPO-guided mitigation scheme targets and cancels these fragile high-weight errors. Consequently, HPO-guided mitigation recovers the target-state fidelity to 0.9381, corresponding to an absolute fidelity gain of 19.5 percentage points relative to the global depolarizing baseline. This result suggests that sparse high-order Pauli residuals can provide actionable information for algorithm-level error mitigation, especially when the dominant noise contains structured crosstalk components. In contrast, the unmitigated noisy simulation suffers from severe decoherence, yielding an unacceptable fidelity of 0.6846. Applying the global depolarizing mitigation baseline improves the target-state fidelity from 0.6846 to 0.7431, where the improvement remains limited, for the reason that the traditional model relies on the oversimplified assumption of independent single-qubit error channels, completely blinding itself to the non-local multi-body correlations (e.g., XX or ZZ crosstalk) that aggressively corrupt the deep QPE circuit.

3.4. Scalability and Parameter Overhead

To evaluate the scalability of the HPO framework, its PTM-element and physical-coefficient counts are compared with the required PTM-element count associated with Quantum Process Tomography (QPT).

Figure 5 distinguishes four counting quantities of *log* up to $N = 10$. As shown in it, the required PTM-element count $\mathcal{D}_{\text{full}}$ exhibits the fastest exponential growth as the number of qubits N increases. The theoretical mask-selected PTM-element count $K_{\text{res}}^{(N)}$ also grows rapidly, but remains smaller than $\mathcal{D}_{\text{full}}$. The untruncated incremental physical-coefficient count ΔK_N increases more moderately, although its training dimension still grows continuously with N.

In contrast, the stage-wise trainable coefficient count under the Pauli-weight cutoff $w \leq 3$ remains constant for $N \geq 6$, demonstrating that the truncation prevents the optimization dimension from increasing with the system size. This truncation strategy substantially reduces the stage-wise training dimension. As shown in Table 1, even under the ΔK_N scheme, $88{,}572 - 29{,}523 = 59{,}049$ newly introduced parameters must be optimized at $N = 10$. This scale still imposes a considerable optimization burden on classical computing platforms. Moreover, for $N > 10$, the fidelity improvement obtained by including additional high-order Pauli coefficients decreases rapidly. Therefore, optimizing all untruncated incremental parameters is both computationally expensive and inefficient, motivating the use of a Pauli-weight cutoff to limit the number of trainable parameters at each stage.

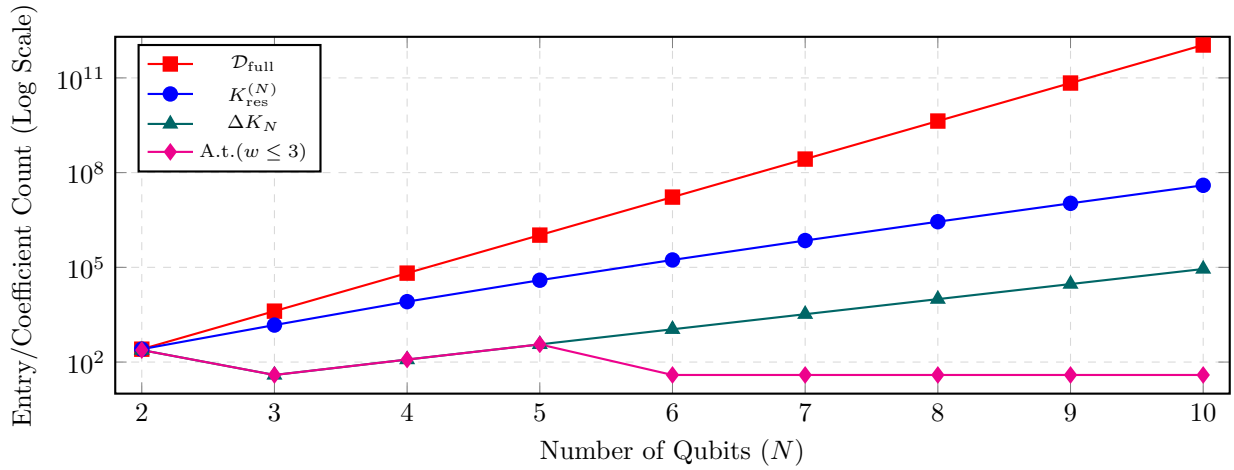


Figure 5: Log-scale comparison of the required PTM-element count, theoretical mask-selected PTM-element count, untruncated incremental physical-coefficient count, and actually trainable parameters count with the Pauli-weight cutoff $w \leq 3$ imposed for $N > 5$.

4. Conclusion

In summary, we have developed and numerically verified the Hierarchical Progressive Optimization (HPO) framework, addressing the exponential $\mathcal{O}(16^N)$ scaling of the required PTM-element count in quantum noise characterization. Our method introduces a deterministic combinatorial projection mask that strictly isolates N-th order spatial crosstalk residuals while freezing foundational low-weight topologies. This mathematical decoupling fundamentally prevents gradient dilution and avoids barren plateaus, significantly reducing the parameter space overhead.

The practical superiority of the HPO framework is demonstrated through an algorithmic-level Quantum Error Mitigation (QEM) simulation on a 10-qubit HHL algorithm. By precisely targeting fragile, high-weight Pauli correlations (such as XX and ZZ crosstalk) that are routinely overlooked by traditional simplified depolarizing models, the mitigation strategy achieves a definitive fidelity recovery to 0.9381, corresponding to an absolute fidelity gain of approximately 19.5 percentage points relative to the global depolarizing baseline.

Looking forward, the HPO framework's ability to efficiently map non-local multi-body correlations opens new avenues for characterizing dynamic noise in deeper variational quantum algorithms and large-scale su-

perconducting processors. Integrating this progressive masking technique with hardware-efficient topological error-correcting codes remains a promising direction for future investigation.

Data availability statement

The data that support the findings of this study are openly available. The source code for this work is available at: https://github.com/gexiangyu0608/HPO-QNC

Acknowledgment

This work was supported by the National Natural Science Foundation of China (No. 62375140) and the National Innovation and Entrepreneurship Training Program for College Students (No. 202610293069Z).

References

[1] Xue C, Chen Z Y, Wu Y C and Guo G P 2021 *Chin. Phys. Lett.* **38** 030302

[2] Wang A, Wei Y Z, Jiang M, Li Y C, Chen H and Huang X 2024 *Chin. Phys. B* **33** 080307

[3] Lei X, Fan J and Pang S 2024 *Chin. Phys. B* **33** 060304

[4] Li T, Zhang H, Zheng L and Li F 2025 *Chin. Phys. B* **34** 124703

[5] Li Y and Benjamin S C 2017 *Phys. Rev. X* **7** 021050

[6] Quek Y, Stilck França D, Khatri S, Meyer J J and Eisert J 2024 *Nat. Phys.* **20** 1648

[7] Stanchev S G and Vitanov N V 2024 *Sci. Rep.* **14** 18185

[8] Roncallo S, Maccone L and Macchiavello C 2023 *Quantum Sci. Technol.* **9** 015010

[9] Levy R, Luo D and Clark B K 2024 *Phys. Rev. Res.* **6** 013029

[10] Sarovar M, Proctor T, Rudinger K, Young K, Nielsen E and Blume-Kohout R 2020 *Quantum* **4** 321

[11] Maciejewski F B, Zimborás Z and Oszmaniec M 2020 *Quantum* **4** 257

[12] Maciejewski F B, Baccari F, Zimborás Z and Oszmaniec M 2021 *Quantum* **5** 464

[13] Wang R, Zhao P, Jin Y and Yu H 2022 *Appl. Phys. Lett.* **121** 152602

[14] Zhao P, Linghu K, Li Z, Xu P, Wang R, Xue G, Jin Y and Yu H 2022 *PRX Quantum* **3** 020301

[15] Zhang Y, Zhang Y, Li S, Zheng W and Yu Y 2024 *Chin. Phys. B* **33** 110306

[16] Tang Z, Lan D and Yu Y 2024 *J. Nanjing Univ. (Nat. Sci.)* **60** 707

[17] Yu Q H and Lin Z J 2024 *Chin. Phys. Lett.* **41** 030202

[18] Ji N H, Sun H Q, Xiao B, Song P L and Ma H Y 2025 *Chin. Phys. B* **34** 020309

[19] Li A Q, Tian C W, Xu X X, Ma H Y and Liang J Q 2025 *Chin. Phys. B* **34** 030306

[20] Sun D, Zhang J, Huang P, Zhang Y, Guo Z, Chen T, Wang R, Sun X, Zhang J, Huang W, Qiu J, Chu J, Tao Z, Guo W, Linpeng X, Jiang J, Niu J, Zhong Y and Yu D 2025 *arXiv:2505.00297 [quant-ph]*

[21] Zhao P, Xu P, Lan D, Chu J, Tan X, Yu H and Yu Y 2020 *Phys. Rev. Lett.* **125** 200503

[22] Zhou Z, Sitler R, Oda Y, Schultz K and Quiroz G 2023 *Phys. Rev. Lett.* **131** 210802

[23] Zhao K, Ma W G, Wang Z, Li H, Huang K, Shi Y H, Xu K and Fan H 2025 *Phys. Rev. Appl.* **24** 034064

[24] Wang J B 2025 *Chin. Phys. B* **34** 040302

[25] Du F F, Fan Z G, Ren X M, Ma M and Liu W Y 2025 *Chin. Phys. B* **34** 010303

[26] Sheldon S, Magesan E, Chow J M and Gambetta J M 2016 *Phys. Rev. A* **93** 060302

[27] Choi M D 1975 *Linear Algebra Appl.* **10** 285

[28] Lindblad G 1976 *Commun. Math. Phys.* **48** 119

[29] Jozsa R 1994 *J. Mod. Opt.* **41** 2315

[30] Bergholm V, Izaac J, Schuld M *et al.* 2018 *arXiv:1811.04968 [quant-ph]*

[31] Harrow A W, Hassidim A and Lloyd S 2009 *Phys. Rev. Lett.* **103** 150502