Department of Computer Science

NEURAL INFORMATION PROCESSING SYSTEMS

Robust Integrated Learning and Pauli Noise Mitigation for Parametrized Quantum Circuits

Md Mobasshir Arshed Naved Wenbo Xie Wojciech Szpankowski Ananth Grama

Department of Computer Science
Purdue University

NeurIPS 2025

Outline

- **Motivation**
- **Contribution**
- **Framework**
- **Conclusions**

The NISQ Era Challenge and Motivation

- **Noise Is Dynamic**
  - NISQ noise changes with temperature and environment
  - Pre-calibrated models quickly become inaccurate
- **Static Calibrations**
  - Periodic re-calibration adds latency
  - Fails to capture real-time fluctuations during training

Training Noisy PQC using Noise Mitigation

- **Zero-Noise Extrapolation (ZNE)** [Kandala et al., 2019]
  - Relies on precise hardware calibration
  - Requires multiple runs at different noise levels $\rightarrow$ large runtime cost
- **Probabilistic Error Cancellation (PEC)** [Temme et al., 2017]
  - Requires full noise characterization
  - Requires an exponential number of measurement shots to achieve high precision
  - Hard to implement for large systems and highly variable noise profiles
- **Van den Berg et al.'s sparse Pauli–Lindblad Formulation** [Van Den Berg et al., 2023]
  - Depend on pre-calibration
  - Assume fixed noise profiles — not true on real hardware

Limitation

High overhead and static noise assumptions make these approaches impractical for real-time learning.

Contribution: Integrated Learning Framework

We propose a novel gradient-based framework for learning parameterized quantum circuits (PQCs) in the presence of Pauli noise

- **Simultaneous Optimization**
    - PQC model parameters ($\overrightarrow{\theta}$)
    - Inverse noise parameters ($\overrightarrow{\sigma}$)
- **Adaptive Mitigation**
    - Dynamically update inverse noise parameters with model parameters
    - Provide *task-specific*, *data-driven* noise mitigation
    - Eliminates the need for noise pre-characterization and calibration, streamlining the training process.
- **Bounded Optimization**
    - Employ constrained updates to ensure *bounded*, *stable*, and *convergent* updates even under varying Pauli noise

Objective Function and Assumptions

Goal: jointly optimize ($\overrightarrow{\theta}$) and ($\overrightarrow{\sigma}$) on noisy quantum device

- *Optimization:*

$$\min_{\overrightarrow{\sigma},\overrightarrow{\theta}\in\mathbb{R}^{|(\overrightarrow{\sigma},\overrightarrow{\theta})|}} \mathcal{L}(\overrightarrow{\sigma},\overrightarrow{\theta})+\mathcal{G}(\overrightarrow{\sigma},\overrightarrow{\theta})$$

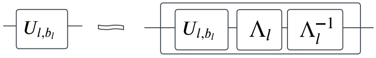
- *Loss Function:*

$$\mathcal{L}(\overrightarrow{\sigma},\overrightarrow{\theta})=\frac{1}{|\mathcal{D}|}\sum_{t=1}^{|\mathcal{D}|}\left(y_t-\mathrm{tr}\left(\mathcal{M}\mathcal{U}_R(\rho_t)\right)\right)^2$$

- *Constraint Function:*

$$\mathcal{G}(\overrightarrow{\sigma},\overrightarrow{\theta})=\begin{cases}0 & \overrightarrow{\sigma},\overrightarrow{\theta}\in\mathcal{X}_{sigma}\times\mathcal{X}_{theta}\\ +\infty & o.w.\end{cases}$$

Noise Mitigated PQC (nmPQC)



l-th layer of PQC   l-th layer of Noise Mitigated PQC

Objective Function and Assumptions

Goal: jointly optimize ($\overrightarrow{\theta}$) and ($\overrightarrow{\sigma}$) on noisy quantum device

- *Optimization:*

$$\min_{\overrightarrow{\sigma},\overrightarrow{\theta}\in\mathbb{R}^{\left|(\overrightarrow{\sigma},\overrightarrow{\theta})\right|}} \mathcal{L}(\overrightarrow{\sigma},\overrightarrow{\theta}) + \mathcal{G}(\overrightarrow{\sigma},\overrightarrow{\theta})$$

- *Loss Function:*

$$\mathcal{L}(\overrightarrow{\sigma},\overrightarrow{\theta}) = \frac{1}{|\mathcal{D}|}\sum_{t=1}^{|\mathcal{D}|}\left(y_t - \mathrm{tr}\left(\mathcal{M}\mathcal{U}_R(\rho_t)\right)\right)^2$$

- *Constraint Function:*

$$\mathcal{G}(\overrightarrow{\sigma},\overrightarrow{\theta}) = \begin{cases} 0 & \overrightarrow{\sigma},\overrightarrow{\theta} \in \mathcal{X}_{sigma}\times\mathcal{X}_{theta} \\ +\infty & o.w. \end{cases}$$

Assumptions

1. Bounded Noise constraints
   - $\lambda_{l,k} \in [0, \mathfrak{B}^{(l,k)}] \subseteq [0,1]$
2. Noiseless single qubit Paulis
   - I, X, Y, Z are considered noiseless
3. Absence of SPAM noise
4. Binary Observable
   - $\mathcal{M}$ is Hermitian with eigenvalue ± 1

Analytic Gradient Expressions

- Gradient w.r.t. PQC parameters

$$\frac{\partial \mathcal{L}_t}{\partial \theta_j} = -\left(y_t - \mathrm{tr}\left(\mathcal{M}\mathcal{U}_R(\rho_t)\right)\right) \cdot \Big[\mathrm{tr}\left(\mathcal{M}\mathcal{U}_{t,>j}\left(\Lambda_j^{-1} \circ \Lambda_j \circ \mathrm{Ad}_{U_j(\theta_j + \frac{\pi}{2})}\left(\rho_t^{(j-1)}\right)\right)\right)$$
$$-\mathrm{tr}\left(\mathcal{M}\mathcal{U}_{t,>j}\left(\Lambda_j^{-1} \circ \Lambda_j \circ \mathrm{Ad}_{U_j(\theta_j - \frac{\pi}{2})}\left(\rho_t^{(j-1)}\right)\right)\right)\Big]$$

- Gradient w.r.t. inverse noise parameters

$$\frac{\partial \mathcal{L}_t}{\partial \sigma_{j,q}} = -4\left(y_t - \mathrm{tr}\left(\mathcal{M}\mathcal{U}_R(\rho_t)\right)\right) \cdot$$
$$\Big[\mathrm{tr}\left(\mathcal{M}\mathcal{U}_{t,>j}\left(\Lambda_j^{-1}\left(\rho_{t,j}\right)\right)\right) - \mathrm{tr}\left(\mathcal{M}\mathcal{U}_{t,>j}\Lambda_j^{-1}\left(P_q^{(j)}\rho_{t,j}(P_q^{(j)})^\dagger\right)\right)\Big]$$

Key Contribution

Efficient, unbiased, parameter-shift-like gradient estimation for *inverse* noise parameters enables scalable joint optimization.

Universal Estimation Algorithm (UEA) and Sample Complexity

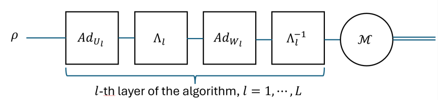
Common structural trace form:

$$\mathrm{tr}\left(\mathcal{M} \bigcirc_{l=1}^{L} \Lambda_l^{-1} \circ \mathrm{Ad}_{W_l} \circ \Lambda_l \circ \mathrm{Ad}_{U_{l,b_l}}(\rho_t)\right)$$

Key Insights

- Single unbiased estimator for both gradients
- Avoids explicit noise process tomography
- Uses a lightweight classical preprocessing step

Quantum Circuit Representation

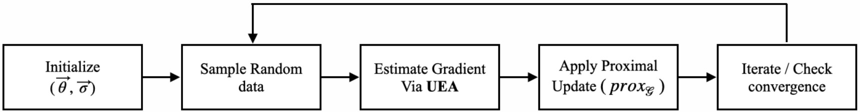


Sample Complexity

- Achieves an additive error of ε with high probability using $\mathcal{O}\left(\frac{|\text{params}|}{\varepsilon^2} \log \frac{|\text{params}|}{\delta}\right)$ measurements.
- Uses random parameter selection to reduce sample cost

Algorithm and Convergence

- **Algorithm:**



- **Convergence Summary**

Key Conditions

- $\mathcal{L}$ locally Lipschitz on compact $\mathcal{X}$; lower bounded
- Estimated gradient has bounded variance
- Step size η_t decreases over time

Intuitive Statement

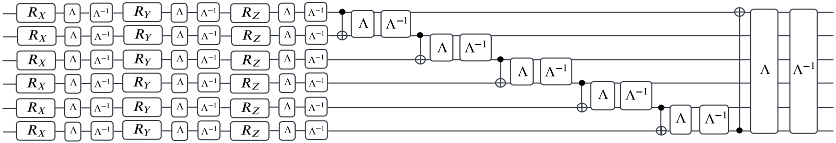
- Proximal SGD converges to a stationary point of $\mathcal{L} + \mathcal{G}$
- Subsampling factor p trades *convergence speed* vs *sample cost*

Takeaway

- Provably *stable*
- *Scales efficiently* with coordinate subsampling
- Preserves *convergence guarantees* under stochastic noise

Numerical Results: Binary Classification on MNIST & Fashion-MNIST

6 qubit Hardware-Efficient Ansatz(HEA)



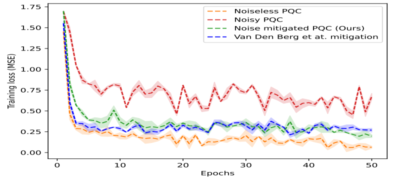
Method	MNIST (3 vs 6)	Fashion-MNIST (Pullover vs Shirt)
Noiseless PQC (oracle)	*94.13 ± 0.31*	*78.87 ± 1.32*
Noisy PQC (baseline)	79.69 ± 3.70	69.57 ± 3.35
Mitigated PQC (Van den Berg)	85.37 ± 3.44	72.37 ± 2.53
Ours (Integrated Learning)	**86.43 ± 3.14**	**74.23 ± 2.84**

All numbers are final test accuracies under *static* noise; mean ± stdev over runs.

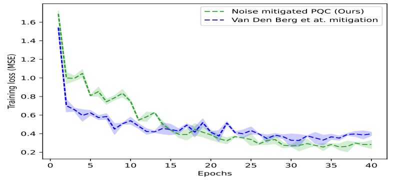
Experimental Wins Under Static and Drifting Noise

MNIST dataset

- Static Noise Scenario

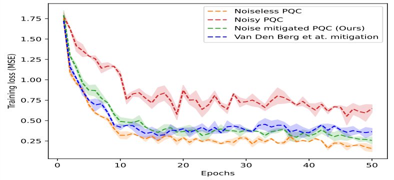


- Dynamic Noise Scenario

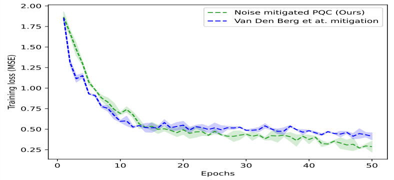


Fashion-MNIST dataset

- Static Noise Scenario



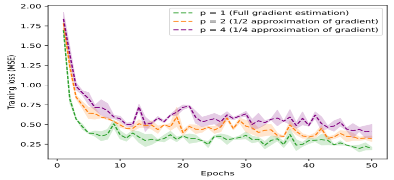
- Dynamic Noise Scenario



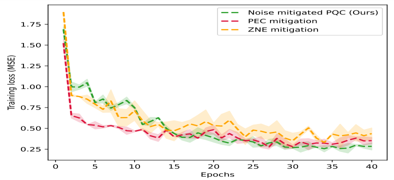
Additional Experiments

MNIST dataset

- Effect of Probabilistic subsampling

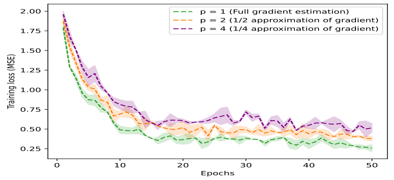


- Comparison with PEC and ZNE under Dynamic noise Scenario

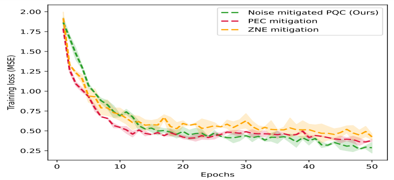


Fashion-MNIST dataset

- Effect of Probabilistic subsampling



- Comparison with PEC and ZNE under Dynamic noise Scenario



Conclusions and Future Work

Key Achievements

- Joint optimization of model and inverse noise parameters
- Parameter-shift-like rule for scalable gradient estimation
- Improved accuracy and adaptability under static and dynamic noise

Future Directions

- Reduce complexity of inverse noise operators
- Extend applicability to general measurement schemes

Thank you!

References

Kandala, A., Temme, K., Córcoles, A. D., Mezzacapo, A., Chow, J. M., and Gambetta, J. M. (2019).
Error mitigation extends the computational reach of a noisy quantum processor.
Nature, 567(7749):491–495.

Temme, K., Bravyi, S., and Gambetta, J. M. (2017).
Error mitigation for short-depth quantum circuits.
Physical review letters, 119(18):180509.

Van Den Berg, E., Minev, Z. K., Kandala, A., and Temme, K. (2023).
Probabilistic error cancellation with sparse pauli–lindblad models on noisy quantum processors.
Nature physics, 19(8):1116–1121.