Generalized quantum subspace expansion method for error mitigation

Nobuyuki Yoshioka[1,2,a]   Hideaki Hakoshima[3]   Yuichiro Matsuzaki[3]
Yuuki Tokunaga[4]   Yasunari Suzuki[4,5]   Suguru Endo[4,b]

Abstract: The control over the effect of noise is crucial for extracting reliable results from noisy quantum computers. Since current available quantum resources are not fully fault-tolerant, it is essential to develop practical quantum error mitigation (QEM) methods to compensate for unwanted computation errors. Here, we propose a novel error mitigation scheme: the generalized quantum subspace expansion method that can mitigate dominant errors in quantum computers. Exploitation of the substantially extended subspace allows us to efficiently mitigate the noise present in the spectra of a given Hamiltonian in a noise-agnostic way. We found two practical subspaces for practical error mitigation: the subspaces spanned by powers of the noisy state ρ^m and a set of error-boosted states. We performed numerical simulations for both subspaces so that we verified suppression of errors by orders of magnitude. Out protocol inherits positive aspects of previous error-agnostic QEM techniques while significantly overcoming their drawbacks.

1. Introduction

Suppression of computation errors is one of crucial problems for implementing practical quantum computing algorithms with imperfect quantum devices [1,2]. On the road towards the goal of fully fault-tolerant computation, the number of required qubits was significantly reduced from theoretical studies while error rates of actual quantum devices were improved drastically in the recent years; however the realization of scalable fault-tolerant quantum computing is still decades away [3]. Therefore, it is vital to establish information processing techniques which makes the best of surging quantum resources *without* performing fully-functional error correction.

The quantum error mitigation (QEM) techniques post-process measurement data to compensate for undesirable bias from computation results, at the cost of additional measurement costs [4–15].The most prominent example is the probabilistic error cancellation method [5,7]. Once the characterization of the noise model is obtained, QEM operations are probabilistically inserted to invert each error map; then we can recover the correct computation result. However, the identification of the noise model is quite costly since we use gate-set tomography which is also easily affected by the drift of noise.

On the other hand, noise-agnostic QEM protocols without prior knowledge on error models have been developed: quantum subspace expansion (QSE) method [16–19] and the virtual distillation (VD) method, which is also called exponential error suppression (EES) method [20–23]. The QSE method enables us to realize a variational subspace spanned by a set of quantum states $\{|\psi_i\rangle\}_i$ as $|\psi\rangle = \sum_i c_i |\psi_i\rangle$, which can be effectively realized by performing additional measurements and post-processing. Although the QSE method was initially proposed to compute excited states from a ground state realized on a quantum device, it mitigates computation errors at the same time. The QSE method works well for mitigation of coherent errors which derive from insufficient variational optimization, lack of quantum circuit representability, and etc. Meanwhile, it cannot mitigate stochastic errors efficiently, because we generally need a linear combination of exponentially many Pauli operators with the system size to construct a projector to the error-free subspace [4, 16]. Remarkably, the VD/EES method, on the other hand, is complementary; it is suitable for suppression of stochastic errors. Entangling operations between M identical copies of noisy quantum states ρ allows us to obtain the error-mitigated computation results of a given observable O as $\langle O \rangle_{\text{VD}}^{(M)} = \text{Tr}[O\rho_{\text{VD}}^{(M)}]$ with $\rho_{\text{VD}}^{(M)} = \rho^M/\text{Tr}[\rho^M]$. The fidelity with a dominant eigenvector of ρ exponentially approaches unity, which indicates exponential suppression

[1] Department of Applied Physics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
[2] Theoretical Quantum Physics Laboratory, RIKEN Cluster for Pioneering Research (CPR), Wako-shi, Saitama 351-0198, Japan
[3] Research Center for Emerging Computing Technologies, National Institute of Advanced Industrial Science and Technology (AIST),1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan.
[4] NTT Computer and Data Science Laboratories, NTT Corporation, Musashino 180-8585, Japan
[5] JST, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama, 332-0012, Japan
[a] nyoshioka@ap.t.u-tokyo.ac.jp
[b] suguru.endou.uc@hco.ntt.co.jp

of errors. However, this method is entirely vulnerable to coherent errors which rotate the dominant eigenvector.

In this work, we propose a unified framework of error-agnostic QEM techniques: the generalized quantum subspace expansion (GSE) method. The central idea is to include general operators related to the target noisy quantum state in the expanded subspace, and we can more efficiently obtain an error-mitigated eigenstate of the target Hamiltonian, by inheriting the benefit of VD/EES methods, which is robustness to stochastic errors. We propose two types of practical subspaces for efficient noise suppression. The first example is the subspace consisting of powers of a noisy quantum state ρ^m, that achieves not only the exponential suppression of stochastic errors, but also drastic mitigation coherent errors. The second example is the subspace constructed by noise-boosted quantum states, which shares similarities with commonly used error-extrapolation method. In stark contrast with the error-extrapolation method, the GSE method is quite robust of fluctuations of noise-stretch factors, which is a crucial improvement for practical applications.

2. Framework of generalized quantum subspace expansion

Suppose a noisy approximation ρ of some desired state, e.g. the ground state of a given Hamiltonian H is obtained by using the variatioanal quantum eigensolver (VQE) [24–33]. The generalized subspace expansion method exploits the following ansatz in the extended subspace to express the eigenspectra of the Hamiltonian:

$$\rho_{\mathrm{EM}} = \frac{P^\dagger A P}{\mathrm{Tr}[P^\dagger A P]}, \qquad (1)$$

where $P = \sum_i \alpha_i \sigma_i$ $(\alpha_i \in \mathbb{C})$ is a general operator, σ_i is in general a non-Hermite operator, and A is a positive-semidefinite Hermite operator. Here, we refer to σ_i as a base of subspace. It is not difficult to verify that ρ_{EM} is a positive-semidefinite Hermite operator with unit norm, which makes ρ_{EM} correspond to a physical quantum state. Note that σ_i and A can be relevant to the noisy state ρ. For instance, we can set $\sigma_i = \rho$ and $A = \rho$; this contrasts the novel GSE method from the conventional QSE because it can also include general operators related to quantum states in the expanded subspace. We not that the most general subspace can be written as:

$$\sigma_i = \sum_k \beta_k^{(i)} \prod_{l=1}^{L_k} U_{lk}^{(i)} \rho_{lk}^{(i)} V_{lk}^{(i)}, \qquad (2)$$

where $\beta_k^{(i)} \in \mathbb{C}$, $\rho_{lk}^{(i)}$ is some quantum state, $U_{lk}^{(i)}$ and $V_{lk}^{(i)}$ are operators that allow for an efficient measurements on quantum computers (e.g. local Pauli operators or unitary operators) with L_k denoting the number of quantum state used in the subspace. See Appendix A.1 for more details.

In order to recover the noise-free spectra of the Hamiltonian, the following conditions need to be satisfied:
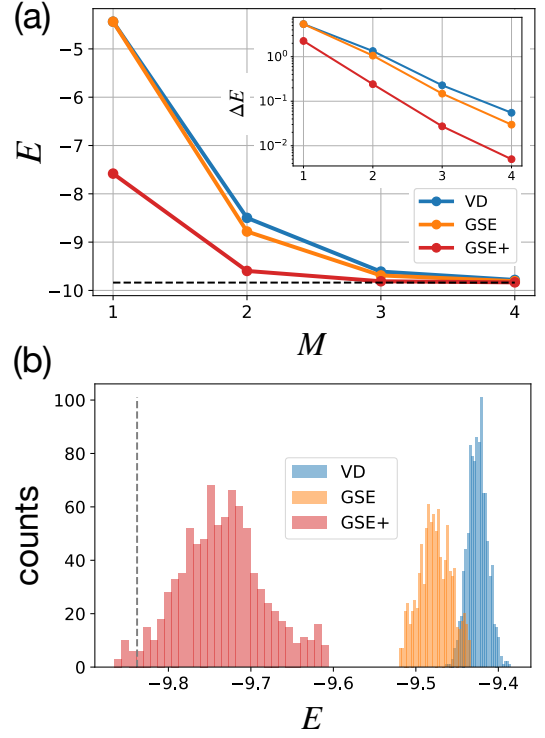
Fig. 1 (a) Mitigation of error in the ground-state energy by using the power subspace. For the GSE method, we take $\sigma_i = \rho_i$ $(i = 0, 1, ..., \frac{M}{2})$ and $A = I$ for even M's while we take $\sigma_i = \rho^i (i = 0, 1, ..., \frac{M-1}{2})$ and $A = \rho$ for odd M's. The blue and orange points correspond to data from the VD/EES and GSE method based on the power subspace, respectively, and red points from the GSE method with non-Hermite operators $\rho^m H$ $(m = 0, 1, ..., \lfloor M/2 \rfloor)$ additionally included in the subspace. The inset shows the log scale plot of the deviation ΔE from the true ground state energy. (b) Histograms of energy estimation for $M = 2$ copies. Here, the number of total measurements is set to 10^9. The gray dotted lines indicate the true ground state energy. We add depolarizing error after each gate of a variational quantum circuit, which is optimized by the VQE algorithm to represent the ground state of one-dimensional transverse-field Ising model with 8 qubits and $h = 1$. The depolarizing error rate p_{dep} is chosen so that the expected number of total errors in ρ is given as $N_{\mathrm{tot}} = N_{\mathrm{gate}} p_{\mathrm{dep}}$ with N_{gate} being the number of gates. Here, we set $N_{\mathrm{tot}} = 1.5$.

$$\min_{\vec{\alpha}} \mathrm{Tr}[\rho_{\mathrm{EM}} H] \quad \text{such that} \quad \mathrm{Tr}[\rho_{\mathrm{EM}}] = 1. \qquad (3)$$

Here we denote $\vec{\alpha} = (\alpha_0, \alpha_1, ...)$. The solution is given by solving the following generalized eigenvalue problem:

$$\mathcal{H}\vec{\alpha} = E\mathcal{S}\vec{\alpha}. \qquad (4)$$

Here, $\mathcal{H}_{ij} = \mathrm{Tr}[\sigma_i^\dagger A \sigma_j H]$ and $\mathcal{S}_{ij} = \mathrm{Tr}[\sigma_i^\dagger A \sigma_j]$ and E is the error-mitigated eigenenergy. The coefficients are normalized such that $\vec{\alpha}^\dagger \mathcal{S} \vec{\alpha} = 1$ which implies $\mathrm{Tr}[\rho_{\mathrm{EM}}] = 1$. We remark that $\mathcal{H}_{ij}$ and $\mathcal{S}_{ij}$ need to be efficiently evaluated on quantum computers. Once $\vec{\alpha}$ is found, the error-mitigated expectation value of any observable O can be computed as $\langle O \rangle = \sum_{ij} \alpha_i^* \alpha_j \mathrm{Tr}[\sigma_i^\dagger A \sigma_j O]$.

3. Demonstration of generalized quantum subspace expansion

We show that by implementing the generalized quantum
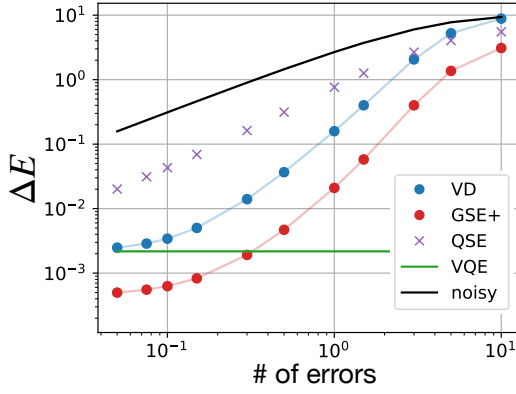
Fig. 2 The expected number of errors N_tot versus the ground-state energy deviation ΔE. Blue filled circles, red filled circles, and purple crosses correspond to the data from the VD/EES, GSE+, and the conventional QSE methods, respectively. Note that GSE+ represents the GSE method which additionally includes a term ρH in the bases of subspace $\{\sigma_i\}$. In addition, the result without QEM is expressed by the black line, and the noiseless outcome from the quantum circuit optimized by the VQE algorithm is shown by the green line. While the improvement of accuracy with the VD/EES methods is restricted by the insufficient representability of the variational quantum circuit, the GSE method can go beyond this limit by further exploring the expanded subspace. The numerical simulation is performed for the ground state of one-dimensional transverse-field Ising Hamiltonian for $N = 8$ qubits; and we have considered $M = 2$ copies of identical noisy quantum states.

subspaces spanned by Eq. (2), we can noise-agnostically suppress errors. To demonstrate the significance of our scheme, we will describe slightly more specific but highly practical two scenarios. Because of their distinct features, we refer to these subspaces as the *power subspace* and *fault subspace*, respectively.

3.1 Power subspace

Let us consider the bases of subspace described by powers of noisy quantum states as $\sigma_i = \rho^i$ $(i = 0, 1, ..., m)$ and set $A = I$:

$$\rho_\text{EM} = \sum_{i,j=0}^{m} \alpha_i^* \alpha_j \rho^{i+j}, \qquad (5)$$

which corresponds to the series expansion of the error-mitigated state ρ_EM with the noisy state ρ, i.e., $\rho_\text{EM} = \sum_{n=0}^{2m} f_n \rho^n$ where $f_n = \sum_{i+j=n} \alpha_i^* \alpha_j$. The choice of $m = 1$, for instance, results in $\rho_\text{EM} = f_0 I + f_1 \rho + f_2 \rho^2$. Note that we may alternatively choose $A = \rho$ to have $\rho_\text{EM} = f_1 \rho + f_2 \rho^2 + f_3 \rho^3$ for $m = 1$.

We remark that higher order states themselves are very useful [20, 21, 34]. By effectively evaluating the expectation value of an observable corresponding to the state $\rho_\text{VD}^{(M)} = \rho^M / \text{Tr}[\rho^M]$ $(M = 2, 3, ...)$, we can exponentially mitigate the contribution from the non-dominant eigenstates of ρ (See Appendix A.2). By using the power subspace, the contribution of non-dominant states will be suppressed even more efficiently by canceling with each other. In fact, it is straightforward to see that the power subspace for $A = I$ completely includes $\rho_\text{VD}^{(2m)}$, and therefore we can always sur-

pass the performance of the VD/EES method when the dominant vector approximates the ground state well (See Appendix A.3).

To clarify the expected advantages by our approach, we numerically demonstrate our algorithm. Figure 1 shows the result for the ground state energy of the one-dimensional transverse-field Ising model. The Hamiltonian is given as $H = -\sum_r Z_r Z_{r+1} + h \sum_r X_r$ where X_r and Z_r denote the x-and z-components of the Pauli matrix acting on the r-th site and h is the amplitude of the transverse magnetic field. We choose $h = 1$ in the following. It is obvious from Fig. 1(a) that both the VD/EES method and our GSE method yields exponential suppression of error with respect to the number of copies M. Moreover, the cancellation of non-dominant states in ρ yields quicker convergence of the expectation value $\text{Tr}[\rho_\text{EM} H]$ towards the ground state energy; this behavior is further highlighted by including additional operators such as $\rho^m H$ to the subspace, which is denoted as GSE+ method in the figures. The trade-off relationship between the accuracy and variance shown in Fig. 1(b) tells us that, even under the restriction of quantum resources (e.g. qubit counts and measurement shots), we benefit from the GSE method using the power subspace.

Now, we further investigate the effect of the crucial obstacle for the previous exponential error suppression techniques—the coherent errors. In Refs. [20, 21, 35], it has been shown that only in the presence of stochastic gate errors, a deviation of the dominant vector may be caused, which is referred to as the coherent mismatch. Furthermore, restrictions on the variational ansatz structure of quantum states due to experimental limitations may also cause coherent errors. In this regard, we stress that our method leads to a significant improvement over previous methods, because the expressibility of quantum states is enhanced effectively by the expansion of the subspace.

Figure 2 shows the result for numerics to support our observations. Here, we plot the energy deviation from the exact ground state energy at various noise levels where we add a term of ρH to bases of the subspace $\{\sigma_i\}$ in the GSE method to compensate for coherent errors. We assume that the error rate p is constant for the each gate, such that the expected number of total error satisfies $N_\text{tot} = N_\text{gate} p$ where N_gate is the number of gates. While the accuracy of the raw noisy state and the conventional QSE method scales only linearly with respect to N_tot, both the VD/EES and GSE methods using two copies of ρ provide quadratic error-suppression in the noisy regime. However, the difference of two methods is highlighted in the low-error regime, where the accuracy of the VD/EES method is restricted by the the performance of the original VQE simulation. Namely, when the noise-free ansatz quantum circuit is not powerful enough and involves algorithmic errors, we cannot remedy the shortage only by recovering the dominant vector. In sharp contrast, our method clearly mitigates such undesirable errors.

It is vital to remark that the required number of measure-
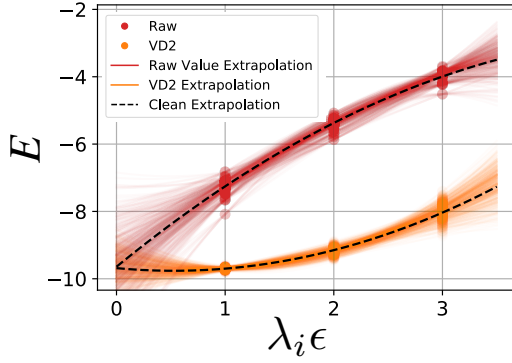
Fig. 3 Error-extrapolation method under imprecise noise level control, with *infinite* number of measurement shots. Red and orange dots indicate the energy estimated from raw quantum states and VD/EES method using $M = 2$ copies, respectively. The colored lines denote the zero-noise extrapolation under fluctuating noise level control, while the black dotted line is the extrapolation under perfect noise control. For each error unit ϵ, we generate 500 sets of noisy quantum states $\rho(\hat{\lambda}_i \epsilon)$ where $\hat{\lambda}_i = \lambda_i + \mathcal{N}(0, \lambda_i \sigma^2)$ for $\lambda_i \in \{1, 2, 3\}$ and $\sigma = 0.1$.

ments for the GSE method scales quadratically with respect to the desired accuracy, just as in the usual quantum measurements (See Appendices A.4 and A.5). When the dominant vector of the state ρ approximates the ground state well, this is mainly explained by the sampling cost deriving from higher powers ρ^M.

3.2 Fault subspace

Now we discuss another practical example of the GSE framework that utilizes non-identical quantum states to span the quantum subspace. Here, the practical QEM is realized by using quantum states from different noise levels, and hence refer to the subspace as the *fault subspace*; we take $\sigma_i = \rho(\lambda_i \epsilon)$ where ϵ denotes the unit of the error (e.g., infidelity per gate) and $\lambda_i \geq 1$ determines the actual noise level. Now, the error-mitigated state is represented as

$$\rho_{\mathrm{EM}} = \sum_{ij} \alpha_i^* \alpha_j \rho(\lambda_i \epsilon) \rho(\lambda_j \epsilon), \tag{6}$$

where we have set $A = I$ and $\sigma_i = \rho(\lambda_i \epsilon)$. We can also include high orders $\rho^m(\lambda_i \epsilon)$ ($m \geq 2$) or operators $U_l^{(i)}$ and $V_l^{(i)}$ in the extended subspace.

3.2.1 Relationship with error-extrapolation method

The fault subspace is closely related to the error-extrapolation method [5, 6]. In the error-extrapolation method, we estimate the error-free expectation value of a given observable O based on results at $n + 1$ noise levels $\langle O(\lambda_i \epsilon) \rangle = \mathrm{Tr}[\rho(\lambda_i \epsilon) O]$. The estimated noise-free result is given by

$$O^* = \sum_{i=0}^{n} \beta_i \langle O(\lambda_i \epsilon) \rangle + \mathcal{O}(\epsilon^{n+1}). \tag{7}$$

Here, $\beta_i = \prod_{j \neq i} \lambda_j (\lambda_j - \lambda_i)^{-1}$ are solutions of the linear equation

$$\sum_i \beta_i = 1, \sum_i \beta_i \lambda_i^k = 0. \tag{8}$$
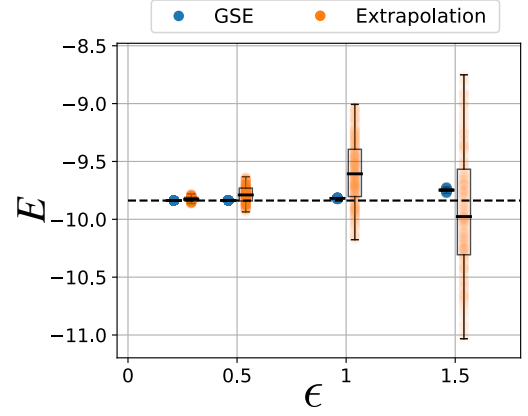
Fig. 4 Comparison between the GSE method and extrapolation under imprecise noise level control, with *infinite* number of measurement shots. The blue and orange points indicate the results from the GSE method using fault subspace and the zero-noise extrapolation for the VD/EES calculation using $M = 2$ copies, respectively. It evident that the extrapolation method under uncertain noise control results in both systematic deviation and increased variance. For each error unit ϵ, we generate 500 sets of noisy quantum states $\rho(\hat{\lambda}_i \epsilon)$ where $\hat{\lambda}_i = \lambda_i + \mathcal{N}(0, \lambda_i \sigma^2)$ for $\lambda_i \in \{1, 2, 3\}$ and $\sigma = 0.1$. As the Hamiltonian, we employ the transverse-field Ising model with $h = 1$ for $N = 8$ qubits.

Because $\{\beta_i\}$ is common among any physical observable, this is equivalent to estimate the noise free density matrix as $\rho_{\mathrm{ex}} = \sum_i \beta_i \rho(\lambda_i \epsilon)$. We remark that Eq. (7) can be utilized as long as $\langle O(\lambda \epsilon) \rangle$ is controlled by λ. In other words, the error-extrapolation method can be used not only for the calculation of the raw expectation value but also for the results obtained from the VD/EES method such as $\mathrm{Tr}[\rho^M(\lambda_i \epsilon) H]/\mathrm{Tr}[\rho^M(\lambda_i \epsilon)]$.

Since the error-extrapolation method is simple as well as practical, it has been studied widely both theoretically and experimentally. However, the extrapolation assumes that the noise level can be accurately controlled (e.g. by extending the gate execution duration), which may be difficult in pratice. Furthermore, since the extrapolation is a purely mathematical operation that does not take any physical constraint into account, it may yield unphysical results even when the measurement is error-free; as a result ρ_{ex} can be a unphysical state whose eigenvalues can be negative.

An instance reflecting the above problem is shown in Fig 3, where we assumed the stretch factor is imprecisely controlled as

$$\hat{\lambda}_i = \lambda_i + \mathcal{N}(0, \lambda_i \epsilon \sigma^2) \tag{9}$$

for $\lambda_i \in \{1, 2, 3\}$ and variance σ^2. We can see that, although we are taking the limit of infinite measurement shots, the true ground state energy cannot be recovered due to the imprecise knowledge on the stretch factor. Whereas the effect of the unwanted deviation in the noise control can be somehow mitigated by using data points calculated from the VD/EES method, it is still necessary to impose physical constraints on the underlying effective density matrix.

3.2.2 Demonstration of Fault subspace

The GSE method with the fault subspace provides the
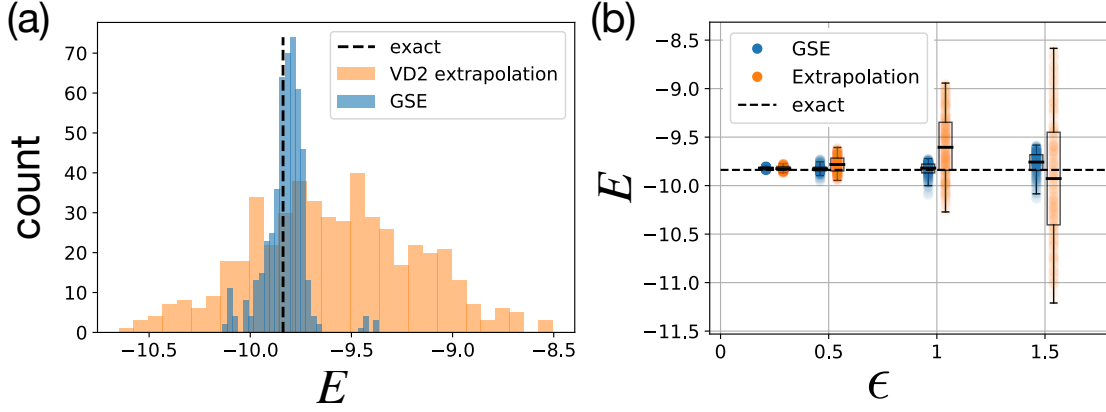
Fig. 5 Suppression of errors under imperfect noise control, with *finite* number of measurement shots. (a) Histogram of ground state energy obtained by error-agnostic mitigation techniques. Here, we compare the GSE method via fault subspace and the extrapolation of VD/EES method using $M = 2$ copies. The unit of controlled noise parameter is taken as $\epsilon = 1$. (b) Comparison of the effect of fluctuation in the stretch factor λ under various ϵ (in similar to Fig. 4). In all calculations, we consider the one-dimensional transverse-field Hamiltonian with $N = 8$ qubits with $h = 1$, and we estimate each Pauli term using 10^7 measurement shots.

solution to the above problems. First, the error-mitigated density matrix constructed by the GSE method corresponds to a physical density matrix. Second, the GSE method using the fault subspace does not rely on the accurate knowledge of noise levels. This is because the GSE method merely constructs a truncated Hilbert space so that the lowest eigenstate of the Hamiltonian is included. It is sufficient to employ bases that are not identical to each other to expand the subspace, while the suitable choice of error levels may result in a better performance.

We show the numerical simulation results in Fig. 4, where we assume the number of measurement shots n_s is infinite. We compare the ground state energy calculated using noisy states $\rho(\hat{\lambda}_i \epsilon)$ for various error unit ϵ, where the noise-stretch factor fluctuates as in Eq. (9). Concretely, the error-extrapolation method estimates the energy at the zero-noise limit by the Richardson extrapolation for each set of data $\hat{\mathcal{D}} = \{(\lambda_i, \text{Tr}[H\rho^2(\hat{\lambda}_i \epsilon)]/\text{Tr}[\rho^2(\hat{\lambda}_i \epsilon)])\}$, and the GSE method constructs effective density matrix corresponding to Eq. (6). While the extrapolated value fluctuates due to the random realization of $\hat{\lambda}_i$, the GSE method is quite robust to fluctuation of the noise-stretch parameter.

When n_s is finite, the result by the GSE method may not necessarily satisfy $E_{\text{GSE}} \geq E_{\text{GS}}$ where E_{GS} denotes the exact ground-state energy. Nevertheless, it is clear from Fig. 5 that the GSE method can estimate the error-free observable much more efficiently and accurately, because the imperfection in the noise control barely affects the construction of the error-mitigated state.

We remark that the GSE method using the fault subspace can surpass the conventional error-extrapolation method even when we assume the stretch factor λ_i can be precisely controlled. We consider distilling the error-extrapolated density matrix ρ_{ex} using $M = 2$ copies as $\rho_{\text{ex}}^2/\text{Tr}[\rho_{\text{ex}}^2]$. It is clear that this state is included in the ansatz for the GSE

method, i.e., $\rho_{\text{EM}} = \sum_{ij} \alpha_i^* \alpha_j \rho(\lambda_i \epsilon) \rho(\lambda_j \epsilon)$. When ρ_{ex} is a physical state, this also holds for higher orders since we can always construct a variational ansatz that includes $\rho_{\text{ex}}^M/\text{Tr}[\rho_{\text{ex}}^M]$. For example, we can simply take $\sigma_i = \rho(\lambda_i \epsilon)$ and $A = \rho_{\text{ex}}^{M-2}$ for $\rho_{\text{EM}} = \sum_{ij} \alpha_i^* \alpha_j \rho(\lambda_i \epsilon) \rho_{\text{ex}}^{M-2} \rho(\lambda_j \epsilon)$.

4. Summary and Outlook

In this work, we have proposed the generalized quantum subspace expansion (GSE) method which unifies the previously reported error-agnostic methods and inherits the advantages of them. As a practical demonstration, we have first discussed to include powers of the noisy quantum state ρ^m in the base of the subspace. This method does provide the exponential suppression of stochastic error which is even more efficient than the VD/EES method as well as eliminates the coherent errors of the dominant vector. In the second strategy, we propose the fault subspace which uses quantum states with various noise levels. Unlike the conventionally used error-extrapolation technique, the GSE method exhibits robust performance even when the noise-stretch factor fluctuates.

We envision several future directions. First, an efficient combination of the proposed scheme and other QEM methods is worth exploring. For example, we can combine the proposed method with the probabilistic error cancellation method to suppress unwanted bias of error-mitigated expectation values due to finite characterization errors. We also anticipate that exploitation of symmetry in the subspace [17, 34] will also help improve the computational accuracy. Second, our method is not restricted to near-term quantum computing, but may enhance computational accuracy even in the fault-tolerant quantum computing regimes in the case that problems of interest involve calculation of eigenspectra. Namely, we may apply the proposed method to suppress the effect of errors due to decoding of logical

qubits or insufficient number of T-gates without any characterization. This is in contrast with the previous works based on the probabilistic error cancellation method [36–39]. Third, although we concentrated on getting error-mitigated ground state or a specific eigenstate of a given Hamiltonian, the solution of Eq. (4) includes the excitation spectra if the subspace is properly chosen. The study of suitable subspace in our GSE framework will be also important in future works.

References

[1] M. A. Nielsen and I. Chuang, "Quantum computation and quantum information," 2002.

[2] D. A. Lidar and T. A. Brun, "*Quantum error correction*," *Cambridge university press*, 2013.

[3] J. Preskill, "Quantum Computing in the NISQ era and beyond," *Quantum*, vol. 2, p. 79, Aug. 2018.

[4] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, "Hybrid quantum-classical algorithms and quantum error mitigation," *Journal of the Physical Society of Japan*, vol. 90, no. 3, p. 032001, 2021.

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

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

[7] S. Endo, S. C. Benjamin, and Y. Li, "Practical quantum error mitigation for near-future applications," *Phys. Rev. X*, vol. 8, p. 031027, Jul 2018.

[8] S. McArdle, X. Yuan, and S. Benjamin, "Error-mitigated digital quantum simulation," *Phys. Rev. Lett.*, vol. 122, p. 180501, May 2019.

[9] X. Bonet-Monroig, R. Sagastizabal, M. Singh, and T. O'Brien, "Low-cost error mitigation by symmetry verification," *Physical Review A*, vol. 98, no. 6, p. 062339, 2018.

[10] Z. Cai, "Multi-exponential Error Extrapolation and Combining Error Mitigation Techniques for NISQ Applications," *arXiv:2007.01265 [quant-ph]*, July 2020.

[11] J. Sun, X. Yuan, T. Tsunoda, V. Vedral, S. C. Benjamin, and S. Endo, "Mitigating realistic noise in practical noisy intermediate-scale quantum devices," *Phys. Rev. Applied*, vol. 15, p. 034026, Mar 2021.

[12] 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, no. 7749, pp. 491–495, 2019.

[13] C. Song, J. Cui, H. Wang, J. Hao, H. Feng, and Y. Li, "Quantum computation with universal error mitigation on a superconducting quantum processor," *Science Advances*, vol. 5, no. 9, 2019.

[14] S. Zhang, Y. Lu, K. Zhang, W. Chen, Y. Li, J.-N. Zhang, and K. Kim, "Error-mitigated quantum gates exceeding physical fidelities in a trapped-ion system," *Nature Communications*, vol. 11, no. 1, p. 587, 2020.

[15] R. Sagastizabal, X. Bonet-Monroig, M. Singh, M. A. Rol, C. Bultink, X. Fu, C. Price, V. Ostroukh, N. Muthusubramanian, A. Bruno, *et al.*, "Experimental error mitigation via symmetry verification in a variational quantum eigensolver," *Physical Review A*, vol. 100, no. 1, p. 010302, 2019.

[16] J. R. McClean, M. E. Kimchi-Schwartz, J. Carter, and W. A. de Jong, "Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states," *Physical Review A*, vol. 95, Apr 2017.

[17] J. R. McClean, Z. Jiang, N. C. Rubin, R. Babbush, and H. Neven, "Decoding quantum errors with subspace expansions," *Nature Communications*, vol. 11, no. 1, p. 636, 2020.

[18] T. Takeshita, N. C. Rubin, Z. Jiang, E. Lee, R. Babbush, and J. R. McClean, "Increasing the representation accuracy of quantum simulations of chemistry without extra quantum resources," *Phys. Rev. X*, vol. 10, p. 011004, Jan 2020.

[19] N. Yoshioka, Y. O. Nakagawa, Y. ya Ohnishi, and W. Mizukami, "Variational quantum simulation for periodic materials," 2020, 2008.09492.

[20] W. J. Huggins, S. McArdle, T. E. O'Brien, J. Lee, N. C. Rubin, S. Boixo, K. B. Whaley, R. Babbush, and J. R. McClean, "Virtual distillation for quantum error mitigation," *arXiv preprint arXiv:2011.07064*, 2020.

[21] B. Koczor, "Exponential error suppression for near-term quantum devices," *arXiv preprint arXiv:2011.05942*, 2020.

[22] P. Czarnik, A. Arrasmith, L. Cincio, and P. J. Coles, "Qubit-efficient exponential suppression of errors," *arXiv preprint arXiv:2102.06056*, 2021.

[23] M. Huo and Y. Li, "Dual-state purification for practical quantum error mitigation," *arXiv preprint arXiv:2105.01239*, 2021.

[24] 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 Communications*, vol. 5, no. 1, p. 4213, 2014.

[25] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, "Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets," *Nature*, vol. 549, no. 7671, pp. 242–246, 2017.

[26] K. M. Nakanishi, K. Mitarai, and K. Fujii, "Subspace-search variational quantum eigensolver for excited states," *Physical Review Research*, vol. 1, no. 3, p. 033062, 2019.

[27] O. Higgott, D. Wang, and S. Brierley, "Variational Quantum Computation of Excited States," *Quantum*, vol. 3, p. 156, July 2019.

[28] T. Jones, S. Endo, S. McArdle, X. Yuan, and S. C. Benjamin, "Variational quantum algorithms for discovering hamiltonian spectra," *Phys. Rev. A*, vol. 99, no. 6, p. 062304, 2019.

[29] N. Yoshioka, Y. O. Nakagawa, K. Mitarai, and K. Fujii, "Variational quantum algorithm for nonequilibrium steady states," *Phys. Rev. Research*, vol. 2, p. 043289, Nov 2020.

[30] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, "Quantum computational chemistry," *Rev. Mod. Phys.*, vol. 92, p. 015003, Mar 2020.

[31] Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. Kieferová, I. D. Kivlichan, T. Menke, B. Peropadre, N. P. D. Sawaya, S. Sim, L. Veis, and A. Aspuru-Guzik, "Quantum chemistry in the age of quantum computing," *Chemical Reviews*, vol. 119, pp. 10856–10915, 10 2019.

[32] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, *et al.*, "Variational quantum algorithms," *arXiv preprint arXiv:2012.09265*, 2020.

[33] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, *et al.*, "Noisy intermediate-scale quantum (NISQ) algorithms," *arXiv preprint arXiv:2101.08448*, 2021.

[34] Z. Cai, "Quantum error mitigation using symmetry expansion," *arXiv preprint arXiv:2101.03151*, 2021.

[35] B. Koczor, "The dominant eigenvector of a noisy quantum state," *arXiv preprint arXiv:2104.00608*, 2021.

[36] Y. Suzuki, S. Endo, K. Fujii, and Y. Tokunaga, "Quantum error mitigation for fault-tolerant quantum computing," *arXiv preprint arXiv:2010.03887*, 2020.

[37] C. Piveteau, D. Sutter, S. Bravyi, J. M. Gambetta, and K. Temme, "Error mitigation for universal gates on encoded qubits," *arXiv preprint arXiv:2103.04915*, 2021.

[38] M. Lostaglio and A. Ciani, "Error mitigation and quantum-assisted simulation in the error corrected regime," *arXiv*

preprint arXiv:2103.07526, 2021.

[39] Y. Xiong, D. Chandra, S. X. Ng, and L. Hanzo, "Sampling overhead analysis of quantum error mitigation: Uncoded vs. coded systems," *IEEE Access*, 2020.

[40] Y. Suzuki, Y. Kawase, Y. Masumura, Y. Hiraga, M. Nakadai, J. Chen, K. M. Nakanishi, K. Mitarai, R. Imai, S. Tamiya, *et al.*, "Qulacs: a fast and versatile quantum circuit simulator for research purpose," 2011.13524.

[41] J. Johansson, P. Nation, and F. Nori, "Qutip 2: A python framework for the dynamics of open quantum systems," *Computer Physics Communications*, vol. 184, no. 4, pp. 1234–1240, 2013.

[42] L. N. Trefethen and D. Bau III, "*Numerical linear algebra*," *Siam*, vol. 50, 1997.

Appendix

A.1 Implementation of general subspace

The most general subspace can be implemented using the bases given as follows,

$$\sigma_i = \sum_k \beta_k^{(i)} \prod_{l=1}^{L_k} U_{lk}^{(i)} \rho_{lk}^{(i)} V_{lk}^{(i)}, \tag{A.1}$$

where $\beta_k^{(i)} \in \mathbb{C}$, $U_{lk}^{(i)}$ and $V_{lk}^{(i)}$ are general operators, L_k denotes the number of quantum state, and $\rho_{lk}^{(i)}$ is a quantum state. Suppose that the Hamiltonian H can be decomposed into a linear combination of products of Pauli operators as $H = \sum_a f_a P_a$. The elements $\mathcal{H}_{ij} := \mathrm{Tr}[\sigma_i^\dagger A \sigma_j H]$ and $\mathcal{S}_{ij} := \mathrm{Tr}[\sigma_i^\dagger A \sigma_j]$ (which we call a Hamiltonian in the expanded subspace and an overlap matrix, respectively) in the Eq. (4) can be described as:

$$\mathcal{H}_{ij} = \sum_{kk'} (\beta_k^{(i)})^* \beta_{k'}^{(j)} \gamma_{kk'}^{(ij)},$$

$$\gamma_{kk'}^{(ij)} = \sum_a f_a \mathrm{Tr}\Bigg[\left(\prod_{l=1}^{L_k} (V_{lk}^{(i)})^\dagger \rho_{lk}^{(i)} (U_{lk}^{(i)})^\dagger \right)$$
$$\times A \left(\prod_{l'=1}^{L_{k'}} U_{l'k'}^{(j)} \rho_{l'k'}^{(j)} V_{l'k'}^{(j)} \right) P_a \Bigg],$$

and

$$\mathcal{S}_{ij} = \sum_{kk'} (\beta_k^{(i)})^* \beta_{k'}^{(j)} \zeta_{kk'}^{(ij)},$$

$$\zeta_{kk'}^{(ij)} = \mathrm{Tr}\Bigg[\left(\prod_{l=1}^{L_k} (V_{lk}^{(i)})^\dagger \rho_{lk}^{(i)} (U_{lk}^{(i)})^\dagger \right)$$
$$\times A \left(\prod_{l'=1}^{L_{k'}} U_{l'k'}^{(j)} \rho_{l'k'}^{(j)} V_{l'k'}^{(j)} \right) \Bigg],$$

and these can be calculated on quantum computers by using a modified controlled derangement operator. Let us show an example when $A = \rho$ and $L_1 = 1$, assuming a single subspace given as $\sigma_i = U_1 \rho_1 V_1$. If both U_1 and V_1 are local Pauli operators, then the calculation of $\mathcal{H}_{ij}$ can be performed by the quantum circuit shown in Fig. A·1, while we can alternatively take them as unitary operators and modify the circuit.

A.2 Review of virtual distillation or exponential error suppression

In this section, we review the virtual distillation (VD)
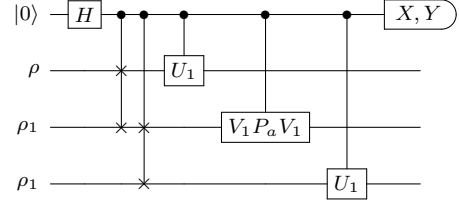
Fig. A·1 The quantum circuits for computing $\gamma_{kk'}^{(ij)}$ when U_1 and V_1 are local Pauli operators. Since $\gamma_{kk'}^{(ij)}$ is a complex number, we need to calculate the real and imaginary parts by measuring Pauli-X and -Y operators of the ancilla qubit, respectively. $\zeta_{kk'}^{(ij)}$ can be similarly evaluated by excluding the operator P_a from the circuit.

method or exponential error suppression (EES) method proposed in Refs. [20, 21]. We consider a noisy state ρ, which can be written in terms of the spectral decomposition:

$$\rho = p_0 |\psi_0\rangle \langle\psi_0| + \sum_{k=1} p_k |\psi_k\rangle \langle\psi_k|, \tag{A.2}$$

where we define $\langle\psi_i|\psi_j\rangle = \delta_{i,j}$ and $p_0 > p_1 \geq p_2 \geq \cdots \geq 0$, and we refer $|\psi_0\rangle$ as a dominant vector. In that method, we can effectively compute the expectation value of an observable from that dominant vector $|\psi_0\rangle$ with exponentially small error:

$$E_{\mathrm{VD}}^{(M)} = \frac{\mathrm{Tr}[\rho^M H]}{\mathrm{Tr}[\rho^M]}$$
$$= \frac{p_0^M \langle\psi_0| H |\psi_0\rangle + \sum_{k=1} (p_k)^M \langle\psi_k| H |\psi_k\rangle}{p_0^M + \sum_{k=1} (p_k)^M}$$
$$= E_{\mathrm{dom}} \left[1 + O\left((p_1/p_0)^M \right) \right], \tag{A.3}$$

by measuring the numerator $\mathrm{Tr}[\rho^M H]$ and the denominator $\mathrm{Tr}[\rho^M]$, respectively. Here, we have defined $E_{\mathrm{dom}} := \langle\psi_0| H |\psi_0\rangle$. These quantities can be measured by unitary diagonalization [20] or indirect (or non-destructive) measurement [20–22]. Fig. A·2 shows a controlled derangement quantum circuit for the case of $M = 3$ to calculate the numerator $\mathrm{Tr}[\rho^M H]$. We emphasize that when increasing the number of copies M, the virtual state $\rho_{\mathrm{VD}}^{(M)} = \rho^M / \mathrm{Tr}[\rho^M]$ exponentially gets closer to the dominant vector $|\psi_0\rangle$.

The VD/EES methods have the best performance when there are only stochastic (or orthogonal) errors that change an ideal state $|\psi_{\mathrm{id}}\rangle$ into its orthogonal states. In this case, we have $|\psi_0\rangle = |\psi_{\mathrm{id}}\rangle$ with sufficiently small error rates. However, this is not always the case. Some type of error causes a change in a state from $|\psi_{\mathrm{id}}\rangle$ to non-orthogonal states, and this leads to $|\psi_0\rangle \neq |\psi_{\mathrm{id}}\rangle$. The infidelity of these states is called the coherent mismatch $1 - |\langle\psi_0|\psi_{\mathrm{id}}\rangle|^2$ [35], and the VD/EES method cannot eliminate such coherent errors in general, although the effect of the coherent errors was investigated through numerical results [20, 21] and analytic results [35].

We also remark that even if the state is unphysical, i.e., the state has negative eigenvalues, as long as $|p_0| > |p_k|$ ($k = 1, 2, ...$), the contribution of $|p_k|/|p_0|$ exponentially vanishes as M increases, which implies virtual distillation still works.
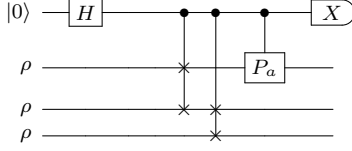
Fig. A·2 A derangement quantum circuit for evaluating $\mathrm{Tr}[\rho^M H] = \sum_a f_a \mathrm{Tr}[\rho^M P_a]$, where we set $M = 3$ and the Hamiltonian is decomposed into a linear combination of products of Pauli operators P_a as $H = \sum_a f_a P_a$.

This ensures that VD/EES method can be applied to error-mitigated unphysical states to further improve the computation accuracy.

A.3 Power subspace includes the virtual distillation

We compare the performance of the VD/EES method, $E_{\mathrm{VD}}^{(M)} = \frac{\mathrm{Tr}[\rho^M H]}{\mathrm{Tr}[\rho^M]}$, with that of the GSE method using power subspace, $E_{\mathrm{GSE}}^{(M)} = \min_{\vec{\alpha}} \mathrm{Tr}[\rho_{\mathrm{EM}} H]$. Since $E_{\mathrm{VD}}^{(M)}$ is not a monotonically decreasing function of M in general, we compare $\min_{1 \leq k \leq M} E_{\mathrm{VD}}^{(k)}$ with $E_{\mathrm{GSE}}^{(M)}$. Actually, we can show $\min_{1 \leq k \leq M} E_{\mathrm{VD}}^{(k)} \geq E_{\mathrm{GSE}}^{(M)}$. For this purpose, we can choose $\vec{\alpha} = \vec{\alpha}_k$ $(1 \leq k \leq m)$ such that

$$\mathrm{Tr}[\rho_{\mathrm{EM}} H] \mid_{\vec{\alpha}=\vec{\alpha}_k} = \begin{cases} E_{\mathrm{VD}}^{(2k)} & (A = I) \\ E_{\mathrm{VD}}^{(2k+1)} & (A = \rho) \end{cases} \qquad (A.4)$$

where we define a vector $\vec{\alpha}_k$:

$$\vec{\alpha}_k = \begin{cases} (\underbrace{0,0,\cdots,0}_{k}, \left(\mathrm{Tr}[\rho^{2k}]\right)^{-1/2}, \underbrace{0,\cdots,0}_{m-k}) & (A = I) \\ (\underbrace{0,0,\cdots,0}_{k}, \left(\mathrm{Tr}[\rho^{2k+1}]\right)^{-1/2}, \underbrace{0,\cdots,0}_{m-k}) & (A = \rho). \end{cases} \qquad (A.5)$$

Note that the $(k+1)$-th component of $\vec{\alpha}_k$ is a non-zero value and all the others are zero. Eq. (A.4) means that the power subspace expansion includes the expectation values obtained by the VD/EES methods. From the above result, we obtain

$$\min_{1 \leq k \leq M} E_{\mathrm{VD}}^{(k)} = \min_{1 \leq k \leq M} \mathrm{Tr}[\rho_{\mathrm{EM}} H] \mid_{\vec{\alpha}=\vec{\alpha}_k} \qquad (A.6)$$

$$\geq \min_{\vec{\alpha}} \mathrm{Tr}[\rho_{\mathrm{EM}} H] = E_{\mathrm{GSE}}^{(M)}. \qquad (A.7)$$

This implies that the convergence of the power subspace expansion is equal or greater than that of the VD/EES methods.

A.4 Effect of shot noise on the generalized subspace expansion

Here, we provide analytical results for the effect of shot noise to the solution of the generalized eigenvalue problem [42]. First we introduce noise-free matrices that describe the quantum subspace; let us denote the Hamiltonian in the expanded subspace and the overlap matrix denote as $\mathcal{H}_0$ and $\mathcal{S}_0$, respectively. Note that $\mathcal{S}_0 \geq 0$. The generalized eigenvalue problem without shot noise can be described as $\mathcal{H}_0 \vec{\alpha}_0 = E_{00} \mathcal{S}_0 \vec{\alpha}_0$, where $\vec{\alpha}_0$ is the error-free ideal solution.

In the following discussion, we assume the ideal eigenvalues are not degenerated. Denoting the effects of shot noise as $\delta\mathcal{H}$ and $\delta\mathcal{S}$, we represent $\mathcal{H} = \mathcal{H}_0 + \delta\mathcal{H}$ and $\mathcal{S} = \mathcal{S}_0 + \delta\mathcal{S}$. Now, we have

$$\mathcal{H}\vec{\alpha}_n = E_n \mathcal{S} \vec{\alpha}_n. \qquad (A.8)$$

Here, $\vec{\alpha}_n = \vec{\alpha}_{on} + \delta\vec{\alpha}_n$ and $E_n = E_{0n} + \delta E_n$ with E_{0n} and $\vec{\alpha}_{0n}$ representing the ideal n-th eigenvalue and solution of the generalized eigenvalue problem, respectively. Focusing on the first order of the deviation, we get

$$\mathcal{H}_0 \delta\vec{\alpha}_n + \delta\mathcal{H}\vec{\alpha}_{0n} = E_{0n}\mathcal{S}_0\delta\vec{\alpha}_n + E_{0n}\delta\mathcal{S}\vec{\alpha}_{0n} + \delta E_n\mathcal{S}_0\vec{\alpha}_{0n}. \qquad (A.9)$$

Now, by expanding $\delta\vec{\alpha}_n = \sum_m \epsilon_{nm}\vec{\alpha}_{0m}$, we have

$$\sum_m \epsilon_{nm}E_{0m}\mathcal{S}_0\vec{\alpha}_{0m} + \delta\mathcal{H}\vec{\alpha}_{0n} = E_{0n}\mathcal{S}_0 \sum_m \epsilon_{nm}\vec{\alpha}_{0m}$$
$$+ E_{0n}\delta\mathcal{S}\vec{\alpha}_{0n} + \delta E_n\mathcal{S}_0\vec{\alpha}_{0n}. \qquad (A.10)$$

By multiplying $\vec{\alpha}_{0n}^\dagger$ from the left in Eq. (A.10), we obtain

$$\delta E_n = \vec{\alpha}_{0n}^\dagger(\delta\mathcal{H} - E_{0n}\delta\mathcal{S})\vec{\alpha}_{0n}, \qquad (A.11)$$

where we used

$$\vec{\alpha}_{0m}^\dagger \mathcal{S}_0 \vec{\alpha}_{0n} = \delta_{nm}. \qquad (A.12)$$

Multiplying $\vec{\alpha}_{0l}^\dagger$ $(l \neq n)$ from the left in Eq. (A.10) leads to

$$\epsilon_{nl} = \frac{\vec{\alpha}_{0l}^\dagger(\delta\mathcal{H} - E_{0n}\delta\mathcal{S})\vec{\alpha}_{0n}}{E_{0n} - E_{0l}}. \qquad (A.13)$$

We also note that from the normalization condition $\vec{\alpha}_n^\dagger \mathcal{S} \vec{\alpha}_n = 1$, we can derive $\epsilon_{nn} = -\frac{1}{2}\vec{\alpha}_{0n}\delta\mathcal{S}\vec{\alpha}_{0n}$. To summarize, we obtain

$$E_n = E_{0n} + \vec{\alpha}_{0n}^\dagger(\delta\mathcal{H} - E_{0n}\delta\mathcal{S})\vec{\alpha}_{0n}$$
$$\vec{\alpha}_n = (1 - \frac{1}{2}\vec{\alpha}_{0n}\delta\mathcal{S}\vec{\alpha}_{0n})\vec{\alpha}_{0n} + \sum_{m \neq n} \frac{\vec{\alpha}_{0m}^\dagger(\delta\mathcal{H} - E_{0n}\delta\mathcal{S})\vec{\alpha}_{0n}}{E_{0n} - E_{0m}}\vec{\alpha}_{0m}. \qquad (A.14)$$

Let us denote the total number of measurements and the dimension of the subspace as N_s and D. Since each element of $\delta\mathcal{H}$ and $\delta\mathcal{S}$ is in the order of $O(DN_s^{-\frac{1}{2}})$, we have $\delta E_n = O(DN_s^{-\frac{1}{2}})$ and $\delta\vec{\alpha}_n = O(DN_s^{-\frac{1}{2}})$. We can see $\{S_0^{1/2}\vec{\alpha}_n\}_n$ are mutually orthogonal and normalized due to Eq. (A.12). Thus, for an arbitrary $D \times D$ matrix B, we have

$$|\vec{\alpha}_{0n}^\dagger B \alpha_{0n}| = |\vec{\alpha}_{0n}^\dagger \mathcal{S}_0^{1/2} \mathcal{S}_0^{-1/2} B \mathcal{S}_0^{-1/2} \mathcal{S}_0^{1/2} \vec{\alpha}_{0n}|$$
$$\leq \|\mathcal{S}_0^{-1/2} B \mathcal{S}_0^{-1/2}\|_{op}, \qquad (A.15)$$

where $\| \cdot \|_{op}$ denotes an operator norm. Substituting $B = \delta\mathcal{H} - E_{0n}\delta\mathcal{S}$ to Eq. (A.15) results in

$$|\delta E_n| \leq \|\mathcal{S}_0^{-1/2}(\delta\mathcal{H} - E_{0n}\delta\mathcal{S})\mathcal{S}_0^{-1/2}\|_{op}$$
$$\leq \|\mathcal{S}_0^{-1}\|_{op}(\|\delta\mathcal{H}\|_{op} + |E_{0n}|\|\delta\mathcal{S}\|_{op}) \qquad (A.16)$$
$$\lesssim \frac{2D\|\mathcal{S}_0^{-1}\|_{op}}{\sqrt{N_s}}(\|\mathcal{H}\|_F + |E_{0n}|\|\mathcal{S}\|_F), \qquad (A.17)$$

where $\| \cdot \|_F$ is the Frobenius norm and we used

$\|C\|_{op} \leq \|C\|_F$ for an arbitrary matrix C, and $\|\delta\mathcal{H}\|_F \lesssim \|\mathcal{H}\|_F / \sqrt{N_\mathrm{s} D^{-2}/2}$ and $\|\delta\mathcal{S}\|_F \lesssim \|\mathcal{S}\|_F / \sqrt{N_\mathrm{s} D^{-2}/2}$. Also, we assigned the same number of samples to the measurement of $\mathcal{H}$ and $\mathcal{S}$. Here, we assumed $A = \rho$ or $A = I/d$ with d being the system dimension. Note that rescaling the A matrix does not affect the overall accuracy.

When we decompose the Hamiltonian into the linear combination of Pauli operators as $H = \sum_a h_a P_a$, we get $|\mathcal{H}_{ij}| \leq \gamma$ with $\gamma = \sum_a |h_a|$. Thus, we obtain

$$|\delta E_n| \leq \frac{4\gamma D^2 \|\mathcal{S}_0^{-1}\|_{op}}{\sqrt{N_\mathrm{s}}}, \qquad (A.18)$$

where we used $E_{0n} \leq \|H\|_{op} \leq \gamma$, and also assumed $\mathcal{S}_{ij} \leq 1$, which is satisfied in highly practical cases such as the power subspace (without additional bases) and fault subspace. Thus, we can achieve the required accuracy ε when $N_\mathrm{s} \geq 16\gamma^2 D^4 \|\mathcal{S}_0^{-1}\|^2 / \varepsilon^2$.

We further discuss $\|\mathcal{S}_0^{-1}\|_{op}$ for the power subspace in the case of $D = 2$ and $D = 3$, which is highly practical for near-term devices. For $D = 2$, when we set $A = \rho$ for the ansatz given in Eq. (1), we have

$$\mathcal{S}_0 = \begin{pmatrix} 1 & \mathrm{Tr}[\rho^2] \\ \mathrm{Tr}[\rho^2] & \mathrm{Tr}[\rho^3] \end{pmatrix}, \qquad (A.19)$$

and we obtain $\|\mathcal{S}_0^{-1}\|_{op} \approx (\mathrm{Tr}[\rho^3] - (\mathrm{Tr}[\rho^2])^2)^{-1}$ under the assumption of $(\mathrm{Tr}[\rho^2])^2 \ll 1$. Similarly, we have $\|\mathcal{S}_0^{-1}\|_{op} \approx (\mathrm{Tr}[\rho^2])^{-1}$ for $A = I/d$. In the case of $D = 3$, we obtain $\|\mathcal{S}_0^{-1}\|_{op} \approx (\mathrm{Tr}[\rho^5] - (\mathrm{Tr}[\rho^4])^2 / \mathrm{Tr}[\rho^3])^{-1}$ $(A = \rho)$ and $\|\mathcal{S}_0^{-1}\|_{op} \approx (\mathrm{Tr}[\rho^4] - (\mathrm{Tr}[\rho^3])^2 / \mathrm{Tr}[\rho^2])^{-1}$ $(A = I/d)$, respectively.

A.5 Effect of shot noise on observable estimation

As is discussed in the previous section, the values of physical observable necessarily deviates from the ideal ones unless we take infinite number of measurements. Such fluctuations are often called the shot noise. For instance, results provided in Fig. 1 in the main text take the effect of shot noise into account. Each matrix element $\mathcal{H}_{ij}$ (or $\mathcal{S}_{ij}$) is given as $\mathcal{H}_{ij} = \mathcal{H}_{ij}^{(0)} + \delta\mathcal{H}_{ij}$, where $\mathcal{H}_{ij}^{(0)}$ is the expectation value and $\delta\mathcal{H}_{ij}$ is a Gaussian noise whose amplitude is determined from the variance that arise from projective measurements. In the following, we discuss the variance of estimators for physical observables under power and fault subspaces, where the generalization can also be done straightforwardly.

A.5.1 Power subspace

It is instructive to start from a case where we have M identical copies to compute the expectation value of an operator O. As was proposed in Ref. [20, 21], let us employ the cyclic shift operator $S^{(M)}$ which act on M systems as follows,

$$S^{(M)} |\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots \otimes |\psi_M\rangle = |\psi_2\rangle \otimes |\psi_3\rangle \otimes \cdots \otimes |\psi_M\rangle \otimes |\psi_1\rangle, \qquad (A.20)$$

where $|\psi_m\rangle$ denotes an arbitrary quantum state of m-th system. Practically, $S^{(M)}$ can be regarded as a product of shift operators between two systems as $S^{(M)} = \prod_m S_{m,m+1}^{(2)}$ that can be constructed as a tensor product of swap operators acting on all corresponding qubits where $S_{m,m+1}^{(2)} |\psi_m\rangle \otimes |\psi_{m+1}\rangle = |\psi_{m+1}\rangle \otimes |\psi_m\rangle$. After some computation, we can show that the expectation value for the power of quantum state ρ^M can be expressed as

$$\langle O \rangle = \frac{\mathrm{Tr}[\rho^M O]}{\mathrm{Tr}[\rho^M]} = \frac{\mathrm{Tr}[S^{(M)} O^{(M)} \bigotimes_{m=1}^M \rho]}{\mathrm{Tr}[S^{(M)} \bigotimes_{m=1}^M \rho]}, \qquad (A.21)$$

where we take $O^{(M)} = O \otimes I \otimes \cdots\cdots I$ to provide a unified view over various choice of M [20]. The variance for the single-shot estimation of the numerator can be given as

$$\begin{aligned} \mathrm{Var}(S^{(M)} O^{(M)}) &= \langle\langle (S^{(M)} O^{(M)})^2 \rangle\rangle - \langle\langle S^{(M)} O^{(M)} \rangle\rangle^2 \\ &= \langle\langle S^{(M)} O^{(M)} S^{(M)} O^{(M)} \rangle\rangle - \mathrm{Tr}[\rho^M O]^2 \\ &= \mathrm{Tr}[\rho O]^2 - \mathrm{Tr}[\rho^M O]^2, \qquad (A.22) \end{aligned}$$

where $\langle\langle \cdot \rangle\rangle := \mathrm{Tr}[(\cdot) \otimes_{m=1}^M \rho]$ is introduced to denote the expectation value computed in the extended Hilbert space, which consists of M systems. It is beneficial to remark on a case when the observable is a linear combination of Pauli operators; we decompose as $O = \sum_{k=1}^{N_O} c_k P_k$ where P_k is a product of Pauli operators and N_O denotes the number of terms. Then, the variance is given as

$$\mathrm{Var}(S^{(M)} O^{(M)}) = \sum_k |c_k|^2 \mathrm{Var}(S^{(M)}(P_k^{(M)})). \qquad (A.23)$$

Here, we assume that we measure $\mathrm{Tr}[\rho^M P_1]$, $\mathrm{Tr}[\rho^M P_2]$, $\cdots$, $\mathrm{Tr}[\rho^M P_{N_O}]$ in separate experiments, and therefore the total variance (A.23) is estimated as a sum of each term. The variance for the single-shot estimation of the denominator can be obtained from a parallel argument as

$$\begin{aligned} \mathrm{Var}(S^{(M)}) &= \langle\langle (S^{(M)})^2 \rangle\rangle - \langle\langle S^{(M)} \rangle\rangle^2 \\ &= 1 - \mathrm{Tr}[\rho^M]^2. \qquad (A.24) \end{aligned}$$

Finally, we can calculate the variance of the estimator of the observable itself by substituting the above expressions into the standard formula. If we take n_s measurement shots for every term, the overall variance can be calculated as

$$\begin{aligned} \frac{\mathrm{Var}(O)}{n_\mathrm{s}} \simeq \frac{\langle O \rangle^2}{n_\mathrm{s}} &\left(\frac{\mathrm{Var}(S^{(M)} O^{(M)})}{\langle\langle S^{(M)} O^{(M)} \rangle\rangle^2} + \frac{\mathrm{Var}(S^{(M)})}{\langle\langle S^{(M)} \rangle\rangle^2} \right. \\ &\left. -2 \frac{\mathrm{Covar}(S^{(M)} O^{(M)}, S^{(M)})}{\langle\langle S^{(M)} O^{(M)} \rangle\rangle \langle\langle S^{(M)} \rangle\rangle} \right), \qquad (A.25) \end{aligned}$$

where the covariance between $S^{(M)} O^{(M)}$ and $S^{(M)}$ is set to zero in our numerical simulation. This is because we have chosen $O^{(M)}$ to be $[S^{(M)} O^{(M)}, S^{(M)}] \neq 0$. In this case, the denominator and numerator of Eq. (A.21) cannot be measured simultaneously, and hence we can ignore the covariance by measuring them separately. Let us also remark that overall variance can be improved, e.g., by weighing the number of measurements according to the coefficient c_k.

We have used Eq. (A.25) to estimate the variance of

the matrix elements $\mathcal{H}_{ij}$ and $\mathcal{S}_{ij}$ for the power subspace. Note that bases of subspaces $\{\sigma_i\}$, including the additional terms $\rho^m H$, are chosen so that all elements are given as $\text{Tr}[\rho^m H^k]$ ($k \in \mathbb{Z}^+$).

A.5.2   Fault subspace

Next, we discuss the case when the error-mitigated quantum state is expressed by non-identical quantum states. It is practical to first set $M = 2$ so that we can describe the case for the fault subspace. Given a quantum state $\rho_{\text{EM}} = \sum_{i,j=0}^{m} \alpha_i^* \alpha_j \rho_i \rho_j$ with $\rho_i = \rho(\lambda_i \epsilon)$ and $M = 2m$, the expectation value of a physical observable O is expressed as

$$\langle O \rangle = \frac{\text{Tr}[\rho_{\text{EM}} O]}{\text{Tr}[\rho_{\text{EM}}]} = \frac{\sum_{ij} \alpha_i^* \alpha_j \text{Tr}[\rho_i \rho_j O]}{\sum_{ij} \alpha_i^* \alpha_j \text{Tr}[\rho_i \rho_j]}$$

$$= \frac{\sum_{ij} \alpha_i^* \alpha_j \text{Tr}[S^{(2)} O^{(2)} (\rho_i \otimes \rho_j)]}{\sum_{ij} \alpha_i^* \alpha_j \text{Tr}[S^{(2)} (\rho_i \otimes \rho_j)]}$$

$$= \frac{\sum_{ij} \alpha_i^* \mathcal{O}_{ij} \alpha_j}{\sum_{ij} \alpha_i^* \mathcal{S}_{ij} \alpha_j}, \qquad (\text{A.26})$$

where we have defined $\mathcal{O}_{ij} = \text{Tr}[\rho_i \rho_j O]$. Based on the discussion of Eq. (A.22), we obtain the variance for the single-shot estimation of an element $\mathcal{O}_{ij}$ as

$$\text{Var}(S^{(2)} O^{(2)})_{ij} \equiv \langle\langle (S^{(2)} O^{(2)})^2 \rangle\rangle_{ij} - \langle\langle S^{(2)} O^{(2)} \rangle\rangle_{ij}^2$$

$$= \langle\langle (I \otimes O)(O \otimes I) \rangle\rangle_{ij} - \text{Tr}[\rho_i \rho_j O]^2$$

$$= \text{Tr}[\rho_i O]\text{Tr}[\rho_j O] - \text{Tr}[\rho_i \rho_j O]^2 (\text{A.27})$$

where we have defined $\langle\langle (\cdot) \rangle\rangle_{ij} = \text{Tr}[(\cdot)(\rho_i \otimes \rho_j)]$. Also, in a similar way to Eq. (A.24), the variance for single-shot estimation of $\mathcal{S}_{ij}$ is given as

$$\text{Var}(S^{(2)})_{ij} \equiv \langle\langle (S^{(2)})^2 \rangle\rangle_{ij} - \langle\langle S^{(2)} \rangle\rangle_{ij}^2$$

$$= 1 - \text{Tr}[\rho_i \rho_j]^2. \qquad (\text{A.28})$$

Using the above expressions, we find that the variance under n_s measurement shots for each term can be formally given as

$$\frac{\text{Var}(O)}{n_s} = \frac{\langle O \rangle^2}{n_s} \left(\frac{\sum_{ij} (\alpha_i^*)^2 (\alpha_j)^2 \text{Var}(S^{(2)} O^{(2)})_{ij}}{\text{Tr}[\rho_{\text{EM}} O]^2} \right.$$

$$\left. + \frac{\sum_{ij} (\alpha_i^*)^2 (\alpha_j)^2 \text{Var}(S^{(2)})_{ij}}{\text{Tr}[\rho_{\text{EM}}]^2} \right). \qquad (\text{A.29})$$

In the main text, we calculate the variance of the physical observable (in particular the energy $\langle H \rangle$) in a stochastic way rather than calculating from Eq. (A.29). This procedure can be summarized as follows: (1) add a Gaussian noise to $\mathcal{H}_{ij}$ and $\mathcal{S}_{ij}$ whose variance is determined from Eqs. (A.27) and (A.28), (2) regularize $\mathcal{S}$ to omit small/negative eigenvalues, (3) solve the generalized linear equation, and (4) repeat (1)-(3) until the variance can be estimated with sufficient accuracy.

A.6   Details regarding the structure of variational quantum circuits

In the main text, we have simulated the ground state of the one-dimensional transverse-field Ising model $H = -\sum_r Z_r Z_{r+1} + h \sum_r X_r$ where X_r and Z_r denote the x- and z-component of the Pauli matrices acting on the r-th site under open boundary condition. Throughout this paper, we have employed the hardware-efficient ansatz structure as shown in Fig. A·3, where the repetition number of units has been taken as $d = 12$ for results in Figs. 1 and 4, and $d = 6$ for that in Fig. 2. Noisy quantum states are generated by adding local depolarizing noise after each quantum gate.
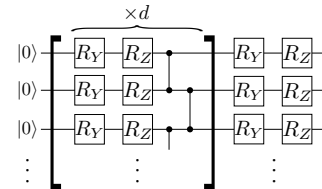
Fig. A·3   Structure of hardware-efficient ansatz employed in this work.