Reduction of Sampling Error in Quantum Krylov Subspace Diagonalization

Gwonhak Lee [1)], Artur Izmaylov [2,3)], Joonsuk Huh [1,4,5)] *

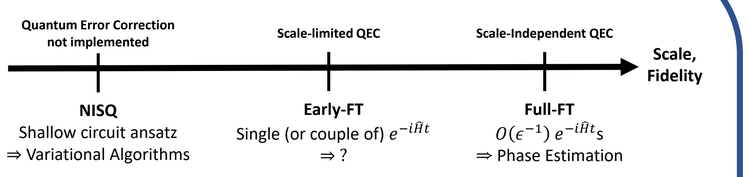
[1)] School of Advanced Institute of Nanotechnology, Sungkyunkwan University, Suwon, Republic of Korea
[2)] Chemical Physics Theory Group, Department of Chemistry, University of Toronto, Toronto, Canada
[3)] University of Toronto Scarborough, Department of Physical and Environmental Sciences, Toronto, Canada
[4)] Department of Chemistry, Sungkyunkwan University, Suwon, Republic of Korea
[5)] Institute of Quantum Biophysics, Sungkyunkwan University, Suwon, Republic of Korea

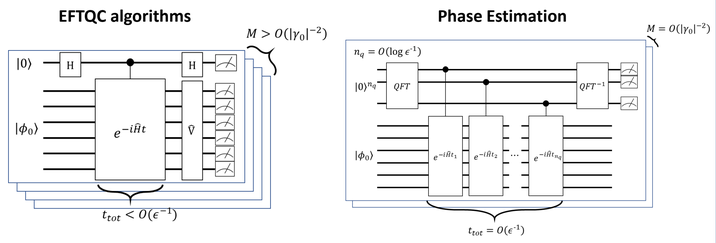
Abstract

An early fault-tolerant quantum approach to the Krylov subspace diagonalization (QKSD) or Lanczos method is emerging to approximately solve Hamiltonian diagonalization problems. Quantum computers are used to prepares a subspace projected by time evolution ansatz ($|\phi_k\rangle = e^{-i\hat{H}t_k}|\phi_0\rangle$) and then generalized eigenvalue problem (GEVP) in the subspace is solved classically. However, despite of its fast convergence to the ground state, the resulting GEVP usually suffers from large condition number, and thus becomes sensitive to the noise. Because we assume fault-tolerance, the finite sampling error (FSE) will be one of the major and fundamental noise sources. We analyze the error in QKSD method originated by FSE. Furthermore, techniques are presented to reduce the finite sampling noise for QKSD based on regularization, symmetry shift (BLISS), iterative coefficient splitting (ICS), and fermionic fluid fragment (F^3). Numerical experiments are demonstrated with small molecules(H_2, H_4, LiH, BeH_2, and H_2O) showing that the techniques reduce the effect of error by a factor of 20 - 500.

Results & Discussion

Early Fault-Tolerant Quantum Computing (EFTQC)





EFTQF algorithms for the estimation of eigenvalues E_i, where $\hat{H}|\psi_i\rangle = E_i|\psi_i\rangle$

- Due to the limited scalability (circuit depth, physical qubit counts, and # of non-Clifford operations), EFTQC algorithm establishes **a trade-off between circuit depth (t_{tot}) and repetition (M)** (see Table 1.).
- Relies on various **classical postprocessing** for eigenvalue estimation on the (noisy) sequences:

$$\left\{\tilde{a}(t_k) = \langle\phi_0|e^{-i\hat{H}t_k}|\phi_0\rangle + \delta_k^{(1)}\right\}_{k=1}^{N}, \qquad \left\{\tilde{a}_{\hat{O}}(t_k) = \langle\phi_0|\hat{O}e^{-i\hat{H}t_k}|\phi_0\rangle + \delta_k^{(2)}\right\}_{k=1}^{N}$$

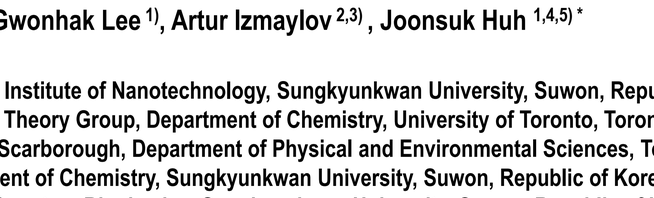
$\delta_k^{(1)}, \delta_k^{(2)}$: Error in the samples (Trotterization, finite sampling, decoherence)

- Examples:
  Time-series analysis [Somma, 2019] : discrete Fourier transform on $\{a(t_k)\}$
  Fourier Filtering [Lin & Tong, 2022] : Randomized time-series analysis for filtered CDF function
  Gaussian Filter [Wang et al. 2023] : Take a filter near the ground state energy
  QCELS [Ding & Lin, 2023] : Quantum Complex Exponential Least Square, find $\tilde{E}_0$ such that $\min_{\tilde{E}_0} \sum_k \left|a(t_k) - e^{-i\tilde{E}_0 t_k}\right|^2$

Gnd. State Energy Estimation Algorithms	$t_{\max}$ (Circuit Depth)	M (Circuit Repetitions)	Total Runtime
Phase Estimation	$O(\epsilon^{-1})$	$O(\lvert\gamma_0\rvert^{-2})$	$O(\epsilon^{-1}\lvert\gamma_0\rvert^{-2})$
[Somma, 2019]	$O(\epsilon^{-1}\text{polylog}(\lvert\gamma_0\rvert^2))$	$O(\epsilon^{-3}\lvert\gamma_0\rvert^{-4})$	$\tilde{O}(\epsilon^{-4}\lvert\gamma_0\rvert^{-4})$
[Lin & Tong, 2022]	$O(\epsilon^{-1}\text{polylog}(\lvert\gamma_0\rvert^2))$	$O(\lvert\gamma_0\rvert^{-4})$	$\tilde{O}(\epsilon^{-1}\lvert\gamma_0\rvert^{-4})$
[Wang et al. 2023]	$O(\Delta^{-1}\text{polylog}(\epsilon^{-1}\lvert\gamma_0\rvert^{-2}))$	$O(\epsilon^{-2}\Delta^2\lvert\gamma_0\rvert^{-4})$	$\tilde{O}(\epsilon^{-2}\Delta\lvert\gamma_0\rvert^{-4})$
[Ding & Lin, 2023]	$\tilde{O}\left(\epsilon^{-1}(1-\lvert\gamma_0\rvert^2)^{\frac{1}{2}}\right)$	$O\left((1-\lvert\gamma_0\rvert^2)^{-\frac{1}{2}(2+o(1))}\right)$	$\tilde{O}\left(\epsilon^{-1}(1-\lvert\gamma_0\rvert^2)^{-\frac{1}{2}(1+o(1))}\right)$

$\epsilon = |E_0 - \tilde{E}_0|$: eigenvalue accuracy
$\gamma_0 = \langle\phi_0|\psi_0\rangle$: initial overlap
$\Delta = E_1 - E_0$: Spectral gap

Quantum Krylov Subspace Diagonalization (QKSD)



- **Quantum Krylov Ansatz:**

$$|\psi(\boldsymbol{c})\rangle = \sum_{k=0}^{n-1} c_k e^{-i\pi k\hat{H}} |\phi_0\rangle$$

- **Project onto the Krylov subspace and apply variational principle, equivalently, solve generalized eigenvalue problem GEVP)**:

$$E_0^{(n)} = \min_{\boldsymbol{c}\in\mathbb{C}^n} \frac{\langle\psi(\boldsymbol{c})|\hat{H}|\psi(\boldsymbol{c})\rangle}{\langle\psi(\boldsymbol{c})|\psi(\boldsymbol{c})\rangle} \Longleftarrow \boldsymbol{Hc} = E^{(n)}\boldsymbol{Sc} \qquad [\boldsymbol{H}]_{kl} = \langle\phi_k|\hat{H}|\phi_l\rangle = \langle\phi_0|\hat{H}e^{-i\pi(l-k)\hat{H}}|\phi_0\rangle \quad [\boldsymbol{S}]_{kl} = \langle\phi_k|\phi_l\rangle = \langle\phi_0|e^{-i\pi(l-k)\hat{H}}|\phi_0\rangle$$

Error Analysis for QKSD gnd state and its energy

$$|\tilde{E}_0^{(n)} - E_0| \le |E_0^{(n)} - E_0| + |\tilde{E}_0^{(n)} - E_0^{(n)}|$$

Total Error | Projection Error | Matrix Perturbation

- **Projection Error (Noiseless QKSD)** – The gnd energy error decays exponentially to the QKSD dimension [Epperly *et al.*, 2022].

$$0 \le E_0^{(n)} - E_0 \le O\left((|\gamma_0|^{-2} - 1)e^{-nO(\Delta)}\right) \Longrightarrow n = \tilde{\Omega}(\Delta^{-1}\log(\epsilon_{\text{subspace}}^{-1}|\gamma_0|^{-2}))$$

- **Noiseless QKSD State Fidelity** – The fidelity between the QKSD gnd state and true state [In Preparation].

$$\mathcal{F}\left(|\lambda_0\rangle, \frac{|\psi(\boldsymbol{c}^*)\rangle}{\||\psi(\boldsymbol{c}^*)\rangle\|}\right) = \frac{|\langle\lambda_0|\psi(\boldsymbol{c}^*)\rangle|^2}{\langle\psi(\boldsymbol{c}^*)|\psi(\boldsymbol{c}^*)\rangle} \ge 1 - \frac{E_0^{(n)} - E_0}{\Delta}$$

Estimating GND State Property – After obtaining $\boldsymbol{c}$ with the standard QKSD method, an observable $\hat{O}$ can be estimated within a small error [In Preparation]:

$$\langle\lambda_0|\hat{O}|\lambda_0\rangle = \frac{\boldsymbol{c}^\dagger\boldsymbol{O}\boldsymbol{c}}{\boldsymbol{c}^\dagger\boldsymbol{S}\boldsymbol{c}} + O\left(\|\hat{O}\|\sqrt{\frac{(E_0^{(n)} - E_0)}{\Delta}}\right) \qquad [\boldsymbol{O}]_{kl} = \langle\phi_0|e^{i\pi k\hat{H}}\hat{O}e^{-i\pi l\hat{H}}|\phi_0\rangle$$

- **Matrix Perturbation (Noisy QKSD) –** Eigenvalue/state perturbation due to the erroneous matrices ($\tilde{\boldsymbol{H}} = \boldsymbol{H} + \boldsymbol{\Delta_H}, \tilde{\boldsymbol{S}} = \boldsymbol{S} + \boldsymbol{\Delta_S}$) :
  Eigenvalue perturbation [Mathias and Li, 2004]:

$$\left|\tilde{E}_0^{(n)} - E_0^{(n)}\right| \le O\left(d_0^{-1}\left(\|\boldsymbol{\Delta_H}\|^2 + \|\boldsymbol{\Delta_S}\|^2\right)^{1/2}\right)$$

Eigenstate perturbation [In Preparation + Davis and Kahan, 1970]:

$$\| |\psi(\boldsymbol{c})\rangle - |\tilde{\psi}(\tilde{\boldsymbol{c}})\rangle \| \le \frac{\|\boldsymbol{H}\|\|\boldsymbol{S}^{-1}\|}{\Delta(1 - \|\boldsymbol{S}^{-1}\boldsymbol{\Delta_S}\|)}\left(\text{cond}(\boldsymbol{S})\frac{\|\boldsymbol{\Delta_S}\|}{\|\boldsymbol{S}\|} + \frac{\|\boldsymbol{\Delta_H}\|}{\|\boldsymbol{H}\|}\right)$$

The condition number, $d_0^{-1} = \left(\min_{\boldsymbol{x}:\|\boldsymbol{x}\|=1} |\boldsymbol{x}^\dagger(\boldsymbol{H} + i\boldsymbol{S})\boldsymbol{x}|\right)^{-1}$ often depicts a large value, making the perturbation significant (ill-conditioning).

Finite Sampling Error Analysis

- **Measurement of Individual Matrix Elements**

$$a(t) = \langle\phi_0|e^{-i\hat{H}t}|\phi_0\rangle = \langle\Psi(t)|(\hat{\sigma}_x + i\hat{\sigma}_y)\otimes\hat{I}|\Psi(t)\rangle$$

$$a_{\hat{O}}(t) = \langle\phi_0|\hat{O}e^{-i\hat{H}t}|\phi_0\rangle = \langle\Psi(t)|(\hat{\sigma}_x + i\hat{\sigma}_y)\otimes\hat{O}|\Psi(t)\rangle$$

$$[\tilde{\boldsymbol{H}}]_{kl} = \sum_{j=1}^{N_\beta} a_{\hat{H}_j}(\pi(l-k)) + O\left(\|\hat{H}\|_\beta M_{kl}^{-1/2}\right) \qquad \hat{H} = \sum_{j=1}^{N_\beta}\hat{H}_j \text{ : Sum of diagonalizable fragments.}$$

$$[\tilde{\boldsymbol{S}}]_{kl} = a(\pi(l-k)) + O\left(\|\hat{H}\|_\beta M_{kl}^{-1/2}\right) \qquad M = \sum_{kl=0}^{n} M_{kl} \text{ : Number of shots to measure the (k,l) element.}$$

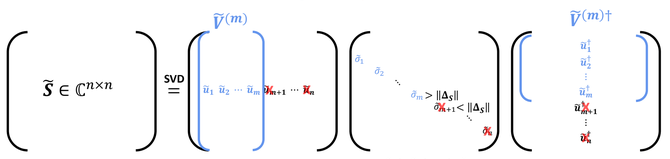
- **Error Matrix Norm** – Element errors are approximated to Gaussian errors, then apply random matrix theory([Roman Vershynin 2018]):

$$\|\boldsymbol{\Delta_H}\| \le \frac{2n\|\hat{H}\|_\beta\sqrt{2\log(2n)}}{\sqrt{M}}$$

$$\|\boldsymbol{\Delta_S}\| \le \frac{2n\sqrt{2\log(2n)}}{\sqrt{M}} \qquad \textbf{with a high probability.}$$

Reduction of Error: Regularized GEVP

Thresholding discards $(n - m)$ basis vectors, which have information as well as noise.
⇒ A trade-off between convergence and perturbation is established.



$$\tilde{\sigma}_1 \ge \tilde{\sigma}_2 \ge \cdots \ge \tilde{\sigma}_n$$

Projection to m basis vectors: $\tilde{\boldsymbol{A}} = \tilde{\boldsymbol{V}}^{(m)\dagger}\tilde{\boldsymbol{H}}\tilde{\boldsymbol{V}}^{(m)}$ and $\tilde{\boldsymbol{B}} = \tilde{\boldsymbol{V}}^{(m)\dagger}\tilde{\boldsymbol{S}}\tilde{\boldsymbol{V}}^{(m)}$ ⇒ Solve $(m \times m)$ GEVP, $\tilde{\boldsymbol{A}}\tilde{\boldsymbol{x}} = E^{(n,m)}\tilde{\boldsymbol{B}}\tilde{\boldsymbol{x}}$ instead.

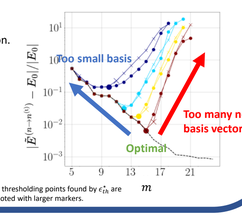
Based on the Weyl's inequality, $\tilde{\sigma}_k$'s less than $\|\boldsymbol{\Delta_S}\|$ are assumed to have very small information.

$$|\sigma_k - \tilde{\sigma}_k| \le \|\boldsymbol{\Delta_S}\|$$

Therefore, we heuristically assign ϵ_{th} with the upper bound of $\|\boldsymbol{\Delta_S}\|$,

$$\|\boldsymbol{\Delta_S}\| \le \frac{2n\sqrt{2\log 2n}}{\sqrt{M}} =: \epsilon_{th}^*,$$

This assignment found the optimal trade-off point.



The thresholding points found by ϵ_{th}^* are denoted with larger markers.

Reducing Norms by Shifting Technique

Minimize $\left\|\hat{H}\right\|_\beta$ to heuristically reduce the sampling variance ($\mathbb{E}[M(\epsilon^2)] \approx 4\left\|\hat{H}\right\|_\beta/\epsilon^2$)
Motivated by block invariant symmetry shift(**BLISS**) [I. Loaiza *et al.* arXiv:2208.08272]

Method

- Find **shift operator** $\hat{T}$ minimizing the norm of the shifted Hamiltonian

$$\min_{\hat{T}} \|\hat{H} - \hat{T}\|_\beta$$

where a Hermitian $\hat{T}$ satisfies the following to preserve the matrix element

$$\hat{T}|\phi_0\rangle = 0 \longrightarrow \langle\phi_0|(\hat{H} - \hat{T})|\phi_l\rangle = \langle\phi_0|\hat{H}|\phi_l\rangle$$

- How do we design $\hat{T}$?
  We consider the reference state $|\phi_0\rangle$ to be a Hartree Fock Slater determinant.

$$\hat{n}_p|\phi_0\rangle = n_p|\phi_0\rangle, \qquad n_p = \begin{cases} 0 & \text{if } p \in virt \\ 1 & \text{if } p \in occ \end{cases} \qquad \hat{n}_p = \hat{a}_p^\dagger\hat{a}_p$$

$$\hat{T}(\boldsymbol{w},\boldsymbol{v}) = \sum_p \left(v_p(\hat{n}_p - n_p) + \sum_{r<s} w_{prs}\hat{E}_r^s(\hat{n}_p - n_p)\right) \qquad \hat{E}_r^s = \hat{a}_r^\dagger\hat{a}_s + \hat{a}_s^\dagger\hat{a}_r$$

One-body | Two-body

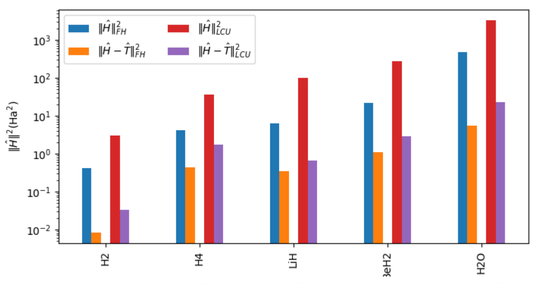
- The shifting technique utilizes the simplicity of bra (or ket) state to reduce the sampling cost

$$\langle\phi_0|\hat{T}|\phi_l\rangle = 0$$

Simple state | Complicated

Norm Reduction Results

- Numerical experiments with the electronic structure Hamiltonian of H_2, H_4, LiH, BeH2, and H_2O (STO-3G basis).



Reduction × 3~10

Fig 1. Norm before and after applying the shift.

Iterative Coefficient Splitting (ICS) & Fluid Fermion Fragment (F^3)

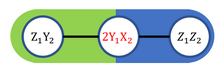
Reduction of Sampling Cost in Pauli Algebra - ICS

- Pauli Grouping

$$\hat{H} = \sum_p \alpha_p\hat{P}_p \quad \xrightarrow{\text{FH}} \quad \hat{H} = \sum_j \hat{H}_j = \sum_j \sum_{p\in\mathcal{C}_j} \alpha_{j,p}\hat{P}_p \qquad [\hat{P}_p, \hat{P}_q] = 0 \quad \forall p,q \in \mathcal{C}_m \text{ and } m.$$

$$\xrightarrow{\text{LCU}} \quad \hat{H} = \sum_j \beta_j\hat{U}_j = \sum_j \beta_j \sum_{p\in\mathcal{A}_j} \alpha'_{j,p}\hat{P}_p \qquad \{\hat{P}_p, \hat{P}_q\} = 2\delta_{p,q}\hat{I} \quad \forall p,q \in \mathcal{A}_m \text{ and } m.$$

Example: Splitting $\hat{H} = Z_1Y_2 + 2Y_1X_2 + Z_1Z_2$ by fragmented Hamiltonians



$$\hat{H}_1 = \hat{Z}_1\hat{Y}_2 + \alpha\hat{Y}_1\hat{X}_2$$
$$\hat{H}_2 = (2-\alpha)\hat{Y}_1\hat{X}_2 + \hat{Z}_1\hat{Z}_2 + 3\hat{X}_1\hat{Y}_2$$

Here, α doesn't change the measurement, but changes variance
⇒ Optimize variance by changing α.

$$\hat{H} = \sum_j \sum_{p\in\mathcal{G}_j} \alpha_p^{(j)}\hat{P}_p \qquad V_{j;0,l}^{(R)}(\boldsymbol{\alpha}^{(j)}) = \sum_{p,q\in\mathcal{G}_j} \alpha_p^{(j)}\alpha_q^{(j)}\text{Cov}_{0,l}^{(R)}(\hat{P}_p, \hat{P}_q)$$

Linear in α, The sum is fixed: $\sum_j \alpha_p^{(j)} = \alpha_p$ | Quadratic in α, the variance can be optimized

FH: $\text{Cov}_{0,l}^{(R)}(\hat{P},\hat{Q}) = \frac{1}{2}\left(\langle\phi_0|\hat{P}\hat{Q}|\phi_0\rangle + \langle\phi_l|\hat{P}\hat{Q}|\phi_l\rangle\right) - \text{Re}(\langle\phi_0|\hat{P}|\phi_l\rangle)\text{Re}(\langle\phi_0|\hat{Q}|\phi_l\rangle)$

LCU: $\text{Cov}_{0,l}^{(R)}(\hat{P},\hat{Q}) = \delta_{\hat{P},\hat{Q}} - \text{Re}(\langle\phi_0|\hat{P}|\phi_l\rangle)\text{Re}(\langle\phi_0|\hat{Q}|\phi_l\rangle)$

Reduction of Sampling Cost in Fermionic Algebra - F^3

- Low rank decomposition of electronic structure Hamiltonian:

$$\begin{aligned}\hat{H} &= \sum_{pq} g_{pq}\hat{E}_p^q + \frac{1}{2}\sum_{pqrs} h_{pq}^{rs}\hat{E}_p^q\hat{E}_r^s \\ &= \hat{U}^{(0)}\left(\sum_p \lambda_p^{(0)}\hat{n}_p\right)\hat{U}^{(0)\dagger} + \sum_{j>0}\hat{U}^{(j)}\left(\sum_{pq}\lambda_{pq}^{(j)}\hat{n}_p\hat{n}_q\right)\hat{U}^{(j)\dagger} \\ &= \sum_j \hat{H}_j\end{aligned}$$

- Then, since $\hat{n}_p^2 = \hat{n}_p$, we can transfer the fragments in two body terms with $p = q$ to one-body

$$\begin{aligned}\hat{H} &= \left[\hat{U}^{(0)}\left(\sum_p \lambda_p^{(0)}\hat{n}_p\right)\hat{U}^{(0)\dagger} + \sum_{j>0}\alpha_p^{(j)}\hat{U}^{(j)}\hat{n}_p\hat{U}^{(j)\dagger}\right] \\ &\quad + \sum_{j>0}\hat{U}^{(j)}\left(\sum_{pq}\lambda_{pq}^{(j)}\hat{n}_p\hat{n}_q - \sum_p \alpha_p^{(j)}\hat{n}_p\right)\hat{U}^{(j)\dagger} \\ &= \sum_j \hat{H}'_j\end{aligned}$$

- Perform the variance optimization as in ICS.

Main Result and Discussion

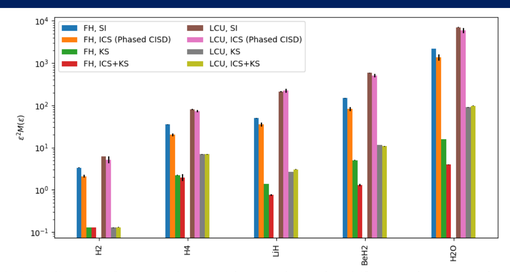


Fig. 2 Sampling cost for FH and LCU with or without the shifting technique and ICS.

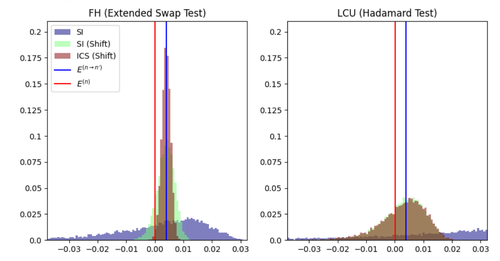


Fig. 3 Histogram of estimated eigenenergy error of the electronic structure Hamiltonian of H_2O.

Experiment Details

- The resulting data is based on 10,000 independent experiments for each setting.
- The number of shots(M) per transitional amplitude for each experiment is $10^8/n$, where n is Krylov order [$n = 3(H_2), 6(H_4), 8(LiH), 12(BeH_2), 14(H_2O)$]
- Fermion to qubit mapping is done by the Bravyi-Kitaev transform without qubit tapering.

Discussion

- ICS reduced the sampling cost by **30~40% in FH**, and **10~20% in LCU**(**Fig. 2**). The off-diagonal covariance in LCU was smaller than that in FH, thus, the reduction in LCU was smaller than FH.
- The **shifting** reduced the sampling cost by the factor of **10-100**. Applying **ICS with shifting**, the factor increased to **20-500**(**Fig. 2**).
- Despite the reduction in sampling noise and the corresponding energy error, there are still large errors between the estimated ground state energy and FCI (**Fig. 3**).
- Since the noiseless QKSD solution coincides with FCI, we can deduce that the error is amplified by the unstable GEVP (problem of large condition number).
- In general, the sampling cost of FH is less than that of LCU.

Acknowledgement

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education, Science and Technology (NRF-2022M3H3A106307411, NRF-2023M3K5A1094805, NRF-2023M3K5A1094813) and Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government(MSIP) (No. 2019-0-00003, Research and Development of Core technologies for Programming, Running, Implementing and Validating of Fault-Tolerant Quantum Computing System).