Efficiently estimating nonlinear properties of quantum states with very few measurement settings

Zihao Li*
(Dated: December 2, 2025)

I. ESTIMATION OF $\mathrm{Tr}(\rho^t)$ WITH GLOBAL HAAR RANDOM UNITARY

Consider that we are given multiple copies of an unknown quantum state ρ on the Hilbert space $\mathcal{H}$ of dimension d, and our goal is to estimate the t-th moment $\mathrm{Tr}(\rho^t)$ using single-copy measurements.

Theorem 1. *Suppose $0 < \epsilon < 1$ and $t = \mathcal{O}(1)$ is a positive integer. There exists an algorithm that can estimate $\mathrm{Tr}(\rho^t)$ within ϵ additive error by using single-copy measurements. The sample complexity and number of different random unitaries used by the algorithm are*

$$\mathcal{O}\left(\max\left\{\frac{d^{1-1/t}}{\epsilon^{2/t}}, \frac{1}{\epsilon^2}\right\}\right), \quad \text{and} \quad \max\left\{1, \frac{1}{d\epsilon^2}\right\}. \tag{1}$$

respectively.

Previously, [Chen and Wang, COLT 2025] and [arXiv:2505.16715] have shown that at least $\Omega(\epsilon^{-2})$ copies of ρ are necessary for the estimation task, even if collective operations are accessible.

A. Estimation protocol

Let S_t be the permutation group of order t. For any element $\pi \in \mathrm{S}_t$, we define its corresponding permutation operator to be the unitary operator R_π on $\mathcal{H}^{\otimes t}$ that satisfies

$$R_\pi\left(|\varphi_1\rangle \otimes \cdots \otimes |\varphi_t\rangle\right) = \left|\varphi_{\pi^{-1}(1)}\right\rangle \otimes \cdots \otimes \left|\varphi_{\pi^{-1}(t)}\right\rangle \qquad \forall\, |\varphi_1\rangle, \ldots, |\varphi_t\rangle \in \mathcal{H}. \tag{2}$$

Recall that t is a constant. Our key observation is that, if we can estimate $\sum_{\pi\in\mathrm{S}_s} \mathrm{Tr}(\rho^{\otimes s} R_\pi)$ within $\mathcal{O}(\epsilon)$ additive error for all $s = 1, \ldots, t$, then we can estimate $\mathrm{Tr}(\rho^t)$ accurately within $\mathcal{O}(\epsilon)$ error. Due to this reason, in the following we focus on the estimation of $\sum_{\pi\in\mathrm{S}_t} \mathrm{Tr}(\rho^{\otimes t} R_\pi)$. Our protocol runs as follows: For $q = 1, 2, \ldots, N_U$, do

1. Sample a random unitary $U_q \sim \mathbb{U}(d)$.
2. Measure N_M copies of ρ in the basis $\left\{U_q^\dagger |b\rangle\langle b| U_q\right\}_{b=0}^{d-1}$ and obtain outcomes $\hat{\mathbf{r}} = \{\hat{r}_1, \ldots, \hat{r}_{N_M}\}$.
3. Compute the estimator $\hat{M}_{U_q}$ defined in Eq. (3) using $\hat{\mathbf{r}}$.

Finally, output $\hat{M} := N_U^{-1} \sum_q \hat{M}_{U_q}$ as our estimate of $\sum_{\pi\in\mathrm{S}_t} \mathrm{Tr}(\rho^{\otimes t} R_\pi)$.

Define $\kappa_t = \binom{t+d-1}{t}$ and

$$\hat{M}_U := \binom{N_M}{t}^{-1} \sum_{i_1 < i_2 < \cdots < i_t} \frac{t!\,\kappa_t}{d}\, \mathbf{1}[\hat{r}_{i_1} = \hat{r}_{i_2} = \cdots = \hat{r}_{i_t}], \tag{3}$$

where $\mathbf{1}[Y] = 1$ when Y is true and 0 otherwise. Then

$$\begin{aligned}\mathbb{E}\left[\hat{M}_U\right] &= \mathbb{E}_U \mathbb{E}_{\hat{\mathbf{r}}}\left(\hat{M}_U \middle| U\right) = \mathbb{E}_U\left[\frac{t!\,\kappa_t}{d} \sum_{b=0}^{d-1} \Pr(\hat{r}_{i_1} = \cdots = \hat{r}_{i_t} = b\,|U)\right] = \frac{t!\,\kappa_t}{d}\, \mathbb{E}_U\left[\sum_{b=0}^{d-1} \langle b|U\rho U^\dagger|b\rangle^t\right] \\ &= \frac{t!\,\kappa_t}{d} \sum_{b=0}^{d-1} \mathrm{Tr}\left[\rho^{\otimes t}\, \mathbb{E}_{U\sim\mathbb{U}(d)} \left(U^\dagger |b\rangle\langle b| U\right)^{\otimes t}\right] = t!\, \mathrm{Tr}\left(\rho^{\otimes t} \Pi_{\mathrm{sym}}^{(t)}\right) = \sum_{\pi\in\mathrm{S}_t} \mathrm{Tr}\left(\rho^{\otimes t} R_\pi\right).\end{aligned} \tag{4}$$

Hence, $\hat{M}$ is an unbiased estimator for $\sum_{\pi\in\mathrm{S}_t} \mathrm{Tr}(\rho^{\otimes t} R_\pi)$.

* zihaoli@hku.hk

B. Variance analyses

The variance of $\hat{M}$ is $\mathrm{Var}[\hat{M}] = N_U^{-1}\,\mathrm{Var}[\hat{M}_U]$, where

$$\mathrm{Var}\Big[\hat{M}_U\Big] = \mathbb{E}\Big[\hat{M}_U^2\Big] - \mathbb{E}\Big[\hat{M}_U\Big]^2 = \mathbb{E}_U\mathbb{E}_{\hat{\mathbf{r}}}\Big(\hat{M}_U^2\Big|U\Big) - \left[\sum_{\pi\in\mathrm{S}_t}\mathrm{Tr}\big(\rho^{\otimes t}R_\pi\big)\right]^2. \tag{5}$$

In the following, we calculate the first term.

We have

$$\mathbb{E}_U\mathbb{E}_{\hat{\mathbf{r}}}\Big(\hat{M}_U^2\Big|U\Big) = \binom{N_M}{t}^{-2}\left(\frac{t!\,\kappa_t}{d}\right)^2 \sum_{\substack{i_1<i_2<\cdots<i_t\\ j_1<j_2<\cdots<j_t}} \mathbb{E}_U\mathbb{E}_{\hat{\mathbf{r}}}\Big(\mathbf{1}[\hat{r}_{i_1}=\cdots=\hat{r}_{i_t}]\mathbf{1}[\hat{r}_{j_1}=\cdots=\hat{r}_{j_t}]\Big|U\Big). \tag{6}$$

Based on the collision number (denoted as $\mathrm{Co}[(i_1,i_2,\ldots,i_t);(j_1,j_2,\ldots,j_t)]$), the expectation value will be different. When $\mathrm{Co}[(i_1,\ldots,i_t);(j_1,\ldots,j_t)] = t-k$, where $k\in\{0,1,\ldots,t-1\}$, we have

$$\mathbb{E}_U\mathbb{E}_{\hat{\mathbf{r}}}\Big(\mathbf{1}[\hat{r}_{i_1}=\cdots=\hat{r}_{i_t}]\mathbf{1}[\hat{r}_{j_1}=\cdots=\hat{r}_{j_t}]\Big|U\Big) = \mathbb{E}_U\sum_{b=0}^{d-1}\langle b|U\rho U^\dagger|b\rangle^{t+k} = \frac{d}{\kappa_{t+k}}\mathrm{Tr}\Big(\rho^{\otimes(t+k)}\Pi_{\mathrm{sym}}^{(t+k)}\Big) \le \frac{d}{\kappa_{t+k}}. \tag{7}$$

When $\mathrm{Co}[(i_1,\ldots,i_t);(j_1,\ldots,j_t)] = 0$, we have

$$\begin{aligned}
&\mathbb{E}_U\mathbb{E}_{\hat{\mathbf{r}}}\Big(\mathbf{1}[\hat{r}_{i_1}=\cdots=\hat{r}_{i_t}]\mathbf{1}[\hat{r}_{j_1}=\cdots=\hat{r}_{j_t}]\Big|U\Big)\\
&= \mathbb{E}_U\sum_{b,b'=0}^{d-1}\langle b|U\rho U^\dagger|b\rangle^t\langle b'|U\rho U^\dagger|b'\rangle^t = \mathbb{E}_U\sum_{b=0}^{d-1}\langle b|U\rho U^\dagger|b\rangle^{2t} + \mathbb{E}_U\sum_{b\neq b'}\langle\!\langle b^{\otimes t}\otimes b'^{\otimes t}|\mathcal{U}^{\otimes 2t}|\rho^{\otimes 2t}\rangle\!\rangle.
\end{aligned} \tag{8}$$

Here, the first term satisfies

$$\mathbb{E}_U\sum_{b=0}^{d-1}\langle b|U\rho U^\dagger|b\rangle^{2t} = \frac{d}{\kappa_{2t}}\mathrm{Tr}\Big(\rho^{\otimes 2t}\Pi_{\mathrm{sym}}^{(2t)}\Big) \le \frac{d}{\kappa_{2t}} = \mathcal{O}(d^{1-2t}); \tag{9}$$

the second term satisfies

$$\begin{aligned}
&\mathbb{E}_U\sum_{b\neq b'}\langle\!\langle b^{\otimes t}\otimes b'^{\otimes t}|\mathcal{U}^{\otimes 2t}|\rho^{\otimes 2t}\rangle\!\rangle\\
&= \sum_{\tau,\pi\in\mathrm{S}_{2t}} W_{\tau,\pi}^{\mathbb{U}(d),(2t)} \sum_{b\neq b'}\langle\!\langle b^{\otimes t}\otimes b'^{\otimes t}|R_\tau\rangle\!\rangle\langle\!\langle R_\pi|\rho^{\otimes 2t}\rangle\!\rangle\\
&= d(d-1)\sum_{\tau_1,\tau_2\in\mathrm{S}_t}\sum_{\pi\in\mathrm{S}_{2t}} W_{\tau_1\tau_2,\pi}^{\mathbb{U}(d),(2t)}\langle\!\langle R_\pi|\rho^{\otimes 2t}\rangle\!\rangle\\
&= d(d-1)\sum_{\tau_1,\tau_2\in\mathrm{S}_t} W_{\tau_1\tau_2,\tau_1\tau_2}^{\mathbb{U}(d),(2t)}\langle\!\langle R_{\tau_1\tau_2}|\rho^{\otimes 2t}\rangle\!\rangle + d(d-1)\sum_{\tau_1,\tau_2\in\mathrm{S}_t}\sum_{\substack{\pi\in\mathrm{S}_{2t}\\ \pi\neq\tau_1\tau_2}} W_{\tau_1\tau_2,\pi}^{\mathbb{U}(d),(2t)}\langle\!\langle R_\pi|\rho^{\otimes 2t}\rangle\!\rangle\\
&= d(d-1)\sum_{\tau_1,\tau_2\in\mathrm{S}_t} d^{-2t}\big(1+\mathcal{O}(d^{-1})\big)\langle\!\langle R_{\tau_1}|\rho^{\otimes t}\rangle\!\rangle\langle\!\langle R_{\tau_2}|\rho^{\otimes t}\rangle\!\rangle + d(d-1)d^{-2t}\mathcal{O}(d^{-1})\\
&= d^{2-2t}\left[\sum_{\tau\in\mathrm{S}_t}\mathrm{Tr}\big(\rho^{\otimes t}R_\tau\big)\right]^2 + \mathcal{O}(d^{1-2t}),
\end{aligned} \tag{10}$$

where the forth equality follows from the fact that $W_{\pi,\pi'}^{\mathbb{U}(d),(2t)} = d^{-2t}\big(\delta_{\pi,\pi'}+\mathcal{O}(d^{-1})\big)$.

We then group the estimators and reduce the summation in Eq. (6) to

$$
\begin{aligned}
&\mathbb{E}_U\mathbb{E}_{\hat{\mathbf{r}}}\left(\hat{M}_U^2\middle|U\right)\\
&=\binom{N_M}{t}^{-2}\left(\frac{t!\,\kappa_t}{d}\right)^2\left[\sum_{k=0}^{t}\sum_{\mathrm{Co}[(i_1,\ldots,i_t);(j_1,\ldots,j_t)]=t-k}\mathbb{E}_U\mathbb{E}_{\hat{\mathbf{r}}}\left(\mathbf{1}[\hat{r}_{i_1}=\cdots=\hat{r}_{i_t}]\mathbf{1}[\hat{r}_{j_1}=\cdots=\hat{r}_{j_t}]\middle|U\right)\right]\\
&=\binom{N_M}{t}^{-2}\left(\frac{t!\,\kappa_t}{d}\right)^2\left[\sum_{k=0}^{t}\binom{N_M}{t+k}\binom{t+k}{t}\binom{t}{k}\mathbb{E}_U\mathbb{E}_{\hat{\mathbf{r}}}\left(\mathbf{1}[\hat{r}_{i_1}=\cdots=\hat{r}_{i_t}]\mathbf{1}[\hat{r}_{j_1}=\cdots=\hat{r}_{j_t}]\middle|U\right)\right]\\
&\leq\binom{N_M}{t}^{-2}\left(\frac{t!\,\kappa_t}{d}\right)^2\left\{\sum_{k=0}^{t-1}\binom{N_M}{t+k}\binom{t+k}{t}\binom{t}{k}\frac{d}{\kappa_{t+k}}+\binom{N_M}{2t}\binom{2t}{t}\left[d^{2-2t}\left(\sum_{\tau\in\mathrm{S}_t}\mathrm{Tr}\big(\rho^{\otimes t}R_\tau\big)\right)^2+\mathcal{O}(d^{1-2t})\right]\right\}\\
&=\sum_{k=0}^{t-1}\binom{N_M}{t}^{-2}\binom{N_M}{t+k}\mathcal{O}(d^{t-1-k})+\binom{N_M}{t}^{-2}\binom{N_M}{2t}\binom{2t}{t}\left[1+\mathcal{O}\big(d^{-1}\big)\right]\left[\left(\sum_{\tau\in\mathrm{S}_t}\mathrm{Tr}\big(\rho^{\otimes t}R_\tau\big)\right)^2+\mathcal{O}(d^{-1})\right]\\
&\leq\sum_{k=0}^{t-1}\mathcal{O}\left(\frac{d^{t-1-k}}{N_M^{t-k}}\right)+\left(\sum_{\tau\in\mathrm{S}_t}\mathrm{Tr}\big(\rho^{\otimes t}R_\tau\big)\right)^2+\mathcal{O}(d^{-1})\\
&=\sum_{i=0}^{t}\mathcal{O}\left(\frac{d^{i-1}}{N_M^{i}}\right)+\left(\sum_{\tau\in\mathrm{S}_t}\mathrm{Tr}\big(\rho^{\otimes t}R_\tau\big)\right)^2,
\end{aligned}
\tag{11}
$$

where the first inequality follows from Eqs. (7), (9), (10), and the second inequality holds because

$$
\begin{aligned}
\binom{N_M}{t}^{-2}\binom{N_M}{2t}\binom{2t}{t}&=\left[\frac{N_M!}{t!(N_M-t)!}\right]^{-2}\frac{N_M!}{(2t)!(N_M-2t)!}\cdot\frac{(2t)!}{(t)!(t)!}=\frac{(N_M-t)!^2}{N_M!(N_M-2t)!}\\
&=\left(\frac{N_M-t}{N_M}\right)\left(\frac{N_M-t-1}{N_M-1}\right)\cdots\left(\frac{N_M-2t+1}{N_M-t+1}\right)\leq 1.
\end{aligned}
\tag{12}
$$

Equations (11) and (5) together imply that

$$
\mathrm{Var}[\hat{M}]=\frac{1}{N_U}\mathrm{Var}[\hat{M}_U]=\frac{1}{N_U}\left[\mathbb{E}_U\mathbb{E}_{\hat{\mathbf{r}}}\left(\hat{M}_U^2\middle|U\right)-\left(\sum_{\pi\in\mathrm{S}_t}\mathrm{Tr}\big(\rho^{\otimes t}R_\pi\big)\right)^2\right]\leq\frac{1}{N_U}\sum_{i=0}^{t}\mathcal{O}\left(\frac{d^{i-1}}{N_M^{i}}\right).
\tag{13}
$$

To ensure that the additive error is $\mathcal{O}(\epsilon)$, we need to bound this variance by $\mathcal{O}\big(\epsilon^2\big)$. Consider two cases below:

1. $d<1/\epsilon^2$. In this case, it suffices to let $N_U=\mathcal{O}\big(d^{-1}\epsilon^{-2}\big)$ and $N_M=\mathcal{O}(d)$. The sample cost is $N_MN_U=\mathcal{O}\big(\epsilon^{-2}\big)$.
2. $d\geq 1/\epsilon^2$. In this case, it suffices to let $N_U=\mathcal{O}(1)$ and $N_M=\mathcal{O}\big(d^{1-1/t}\epsilon^{-2/t}\big)$. So the sample cost is $\mathcal{O}(N_M)$.

To summarize, we have

$$
N_U=\max\left\{1,\frac{1}{d\epsilon^2}\right\},\qquad N_M=\min\left\{\frac{d^{1-1/t}}{\epsilon^{2/t}},d\right\},\qquad N_UN_M=\max\left\{\frac{d^{1-1/t}}{\epsilon^{2/t}},\frac{1}{\epsilon^2}\right\},
\tag{14}
$$

where we ignored constant coefficients. These results confirm the sample complexity illustrated in Theorem 1. Notably, in the usual case $d\geq 1/\epsilon^2$, we only need $N_U=\mathcal{O}(1)$ random unitaries to estimate the moment $\mathrm{Tr}(\rho^t)$ within additive error ϵ. This is the best result that one can expect.

II. ESTIMATION OF $\mathrm{Tr}(O\rho^t)$ WITH GLOBAL HAAR RANDOM UNITARY

Now we turn to estimate $\mathrm{Tr}(O\rho^t)$ using single-copy measurements, where O is an Hermitian observable on $\mathcal{H}$ that satisfies $\|O\|_\infty \le 1$. Let $B = \max\{\mathrm{Tr}(O_0^2), 1\}$, where $O_0 = O - \mathrm{Tr}(O)\mathbb{1}/d$ is the traceless part of O.

Theorem 2. *Suppose $t = \mathcal{O}(1)$ and M are positive integers, and $0 < \epsilon < 1$. There is an algorithm based on single-copy measurements that can estimate $\mathrm{Tr}(O\rho^t)$ to within ϵ additive error, with a high success probability. The sample complexity and number of different random unitaries used by the algorithm are*

$$\mathcal{O}\left(\max\left\{\frac{d^{1-1/t}B^{1/t}}{\epsilon^{2/t}}, \frac{B}{\epsilon^2}\right\}\right) \quad \text{and} \quad \max\left\{1, \frac{B}{d\epsilon^2}\right\}, \tag{15}$$

respectively.

A. Estimation protocol

Note that

$$\mathrm{Tr}(O\rho^t) = \mathrm{Tr}(O_0\rho^t) + d^{-1}\mathrm{Tr}(O)\mathrm{Tr}(\rho^t). \tag{16}$$

Thanks to Theorem 1 in the last section, here the second term can be estimated within ϵ error by using $\mathcal{O}(d^{1-1/t}\epsilon^{-2})$ copies. The remaining task is to estimate the traceless part of O on ρ^t. Hence, WLOG, in the following we assume that O is a traceless observable.

Recall that t is a constant. Similar as in Sec. I, if we can estimate $\sum_{\pi\in \mathrm{S}_{s+1}} \mathrm{Tr}[(\rho^{\otimes s}\otimes O)R_\pi]$ within $\mathcal{O}(\epsilon)$ additive error for all $s = 1, \ldots, t$, then we can estimate $\mathrm{Tr}(O\rho^t)$ accurately within $\mathcal{O}(\epsilon)$ error. Due to this reason, we can focus on the estimation of $\sum_{\pi\in \mathrm{S}_{t+1}} \mathrm{Tr}[(\rho^{\otimes t}\otimes O)R_\pi]$. Our protocol runs as follows: For $q = 1, 2, \ldots, N_U$, do

1. Sample a random unitary $U_q \sim \mathbb{U}(d)$.

2. Measure N_M copies of ρ in the basis $\{U_q^\dagger|b\rangle\langle b|U_q\}_{b=0}^{d-1}$ and obtain outcomes $\hat{\mathbf{r}} = \{\hat{r}_1, \ldots, \hat{r}_{N_M}\}$.

3. Compute the estimator $\hat{M}_{U_q}(O)$ defined in Eq. (17) using $\hat{\mathbf{r}}$.

Finally, output $\hat{M}(O) := N_U^{-1}\sum_q \hat{M}_{U_q}(O)$ as our estimate of $\sum_{\pi\in \mathrm{S}_{t+1}} \mathrm{Tr}[(\rho^{\otimes t}\otimes O)R_\pi]$.

Define

$$\hat{M}_U(O) := \binom{N_M}{t}^{-1} \sum_{i_1<i_2<\cdots<i_t} \frac{(t+1)!\kappa_{t+1}}{d}\, \mathbf{1}[\hat{r}_{i_1} = \hat{r}_{i_2} = \cdots = \hat{r}_{i_t}]\, \langle \hat{r}_{i_1}|UOU^\dagger|\hat{r}_{i_1}\rangle, \tag{17}$$

where $\mathbf{1}[Y] = 1$ when Y is true and 0 otherwise. Then

$$\begin{aligned}
\mathbb{E}\left[\hat{M}_U(O)\right] &= \mathbb{E}_U\mathbb{E}_{\hat{\mathbf{r}}}\left(\hat{M}_U(O)\middle|U\right) = \mathbb{E}_U\left[\frac{(t+1)!\kappa_{t+1}}{d}\sum_{b=0}^{d-1}\Pr(\hat{r}_{i_1} = \cdots = \hat{r}_{i_t} = b\,|U)\langle b|UOU^\dagger|b\rangle\right] \\
&= \frac{(t+1)!\kappa_{t+1}}{d}\,\mathbb{E}_U\left[\sum_{b=0}^{d-1}\langle b|U\rho U^\dagger|b\rangle^t\langle b|UOU^\dagger|b\rangle\right] \\
&= \frac{(t+1)!\kappa_{t+1}}{d}\sum_{b=0}^{d-1}\mathrm{Tr}\left[(\rho^{\otimes t}\otimes O)\mathbb{E}_U\left(U^\dagger|b\rangle\langle b|U\right)^{\otimes(t+1)}\right] \\
&= (t+1)!\,\mathrm{Tr}\left[(\rho^{\otimes t}\otimes O)\Pi_{\mathrm{sym}}^{(t+1)}\right] = \sum_{\pi\in \mathrm{S}_{t+1}}\mathrm{Tr}\left[(\rho^{\otimes t}\otimes O)R_\pi\right].
\end{aligned} \tag{18}$$

Hence, $\hat{M}(O)$ is an unbiased estimator for $\sum_{\pi\in \mathrm{S}_{t+1}} \mathrm{Tr}[(\rho^{\otimes t}\otimes O)R_\pi]$.

B. Variance analyses

The variance of $\hat{M}(O)$ is $\mathrm{Var}[\hat{M}(O)] = N_U^{-1}\,\mathrm{Var}[\hat{M}_U(O)]$, where

$$\mathrm{Var}\left[\hat{M}_U(O)\right] = \mathbb{E}\left[\hat{M}_U(O)^2\right] - \mathbb{E}\left[\hat{M}_U(O)\right]^2 = \mathbb{E}_U\mathbb{E}_{\hat{\mathbf{r}}}\left(\hat{M}_U(O)^2\middle|U\right) - \left(\sum_{\pi\in \mathrm{S}_{t+1}} \mathrm{Tr}\left[\left(\rho^{\otimes t}\otimes O\right)R_\pi\right]\right)^2. \tag{19}$$

In the following, we focus on the calculation of the first term.

We have

$$\mathbb{E}_U\mathbb{E}_{\hat{\mathbf{r}}}\left(\hat{M}_U(O)^2\middle|U\right) = \binom{N_M}{t}^{-2}\left(\frac{(t+1)!\kappa_{t+1}}{d}\right)^2 \sum_{\substack{i_1<i_2<\cdots<i_t\\ j_1<j_2<\cdots<j_t}} \mathbb{E}_U\mathbb{E}_{\hat{\mathbf{r}}}\left[f(\hat{r}_{i_1},\ldots,\hat{r}_{i_t},\hat{r}_{j_1},\ldots,\hat{r}_{j_t},U,O)\middle|U\right], \tag{20}$$

where

$$f(\hat{r}_{i_1},\ldots,\hat{r}_{i_t},\hat{r}_{j_1},\ldots,\hat{r}_{j_t},U,O) = \mathbf{1}[\hat{r}_{i_1}=\cdots=\hat{r}_{i_t}]\mathbf{1}[\hat{r}_{j_1}=\cdots=\hat{r}_{j_t}]\,\langle\hat{r}_{i_1}|UOU^\dagger|\hat{r}_{i_1}\rangle\langle\hat{r}_{j_1}|UOU^\dagger|\hat{r}_{j_1}\rangle. \tag{21}$$

Based on the collision number (denoted as $\mathrm{Co}[(i_1,i_2,\ldots,i_t);(j_1,j_2,\ldots,j_t)]$), the expectation value will be different. When $\mathrm{Co}[(i_1,\ldots,i_t);(j_1,\ldots,j_t)] = t-k$, where $k\in\{0,1,\ldots,t-1\}$, we have

$$\begin{aligned}
\mathbb{E}_U\mathbb{E}_{\hat{\mathbf{r}}}[f(\cdots)|U] &= \mathbb{E}_U\sum_{b=0}^{d-1}\langle b|U\rho U^\dagger|b\rangle^{t+k}\langle b|UOU^\dagger|b\rangle^2\\
&= \frac{d}{\kappa_{t+k+2}}\,\mathrm{Tr}\left[\left(\rho^{\otimes(t+k)}\otimes O^{\otimes 2}\right)\Pi_{\mathrm{sym}}^{(t+k+2)}\right]\\
&= \frac{d}{(t+k+2)!\,\kappa_{t+k+2}}\sum_{\pi\in\mathrm{S}_{t+k+2}}\mathrm{Tr}\left[\left(\rho^{\otimes(t+k)}\otimes O^{\otimes 2}\right)R_\pi\right] = \mathcal{O}\left(\frac{B}{d^{t+k+1}}\right),
\end{aligned} \tag{22}$$

where $B=\mathrm{Tr}(O^2)$.

When $\mathrm{Co}[(i_1,\ldots,i_t);(j_1,\ldots,j_t)] = 0$, we have

$$\begin{aligned}
\mathbb{E}_U\mathbb{E}_{\hat{\mathbf{r}}}[f(\cdots)|U] &= \mathbb{E}_U\sum_{b,b'=0}^{d-1}\langle b|U\rho U^\dagger|b\rangle^t\langle b'|U\rho U^\dagger|b'\rangle^t\langle b|UOU^\dagger|b\rangle\langle b'|UOU^\dagger|b'\rangle\\
&= \mathbb{E}_U\sum_{b=0}^{d-1}\langle b|U\rho U^\dagger|b\rangle^{2t}\langle b|UOU^\dagger|b\rangle^2 + \mathbb{E}_U\sum_{b\neq b'}\langle\!\langle b^{\otimes(t+1)}\otimes b'^{\otimes(t+1)}|\mathcal{U}^{\otimes(2t+2)}|(\rho^{\otimes t}\otimes O)^{\otimes 2}\rangle\!\rangle.
\end{aligned} \tag{23}$$

Here, the first term satisfies

$$\mathbb{E}_U\sum_{b=0}^{d-1}\langle b|U\rho U^\dagger|b\rangle^{2t}\langle b|UOU^\dagger|b\rangle^2 = \frac{d}{(2t+2)!\,\kappa_{2t+2}}\sum_{\pi\in\mathrm{S}_{2t+2}}\mathrm{Tr}\left[\left(\rho^{\otimes 2t}\otimes O^{\otimes 2}\right)R_\pi\right] = \mathcal{O}\left(\frac{B}{d^{2t+1}}\right); \tag{24}$$

the second term satisfies

$$
\begin{aligned}
&\mathbb{E}_U \sum_{b\neq b'} \langle\!\langle b^{\otimes(t+1)} \otimes b'^{\otimes(t+1)} | \mathcal{U}^{\otimes(2t+2)} | (\rho^{\otimes t}\otimes O)^{\otimes 2}\rangle\!\rangle \\
&= \sum_{\tau,\pi\in \mathrm{S}_{2t+2}} W^{\mathbb{U}(d),(2t+2)}_{\tau,\pi} \sum_{b\neq b'} \langle\!\langle b^{\otimes(t+1)} \otimes b'^{\otimes(t+1)} | R_\tau\rangle\!\rangle \langle\!\langle R_\pi | (\rho^{\otimes t}\otimes O)^{\otimes 2}\rangle\!\rangle \\
&= d(d-1) \sum_{\tau_1,\tau_2\in\mathrm{S}_{t+1}} \sum_{\pi\in\mathrm{S}_{2t+2}} W^{\mathbb{U}(d),(2t+2)}_{\tau_1\tau_2,\pi} \langle\!\langle R_\pi | (\rho^{\otimes t}\otimes O)^{\otimes 2}\rangle\!\rangle \\
&= d(d-1) \sum_{\tau_1,\tau_2\in\mathrm{S}_{t+1}} W^{\mathbb{U}(d),(2t+2)}_{\tau_1\tau_2,\tau_1\tau_2} \langle\!\langle R_{\tau_1\tau_2} | (\rho^{\otimes t}\otimes O)^{\otimes 2}\rangle\!\rangle + d(d-1) \sum_{\tau_1,\tau_2\in\mathrm{S}_{t+1}} \sum_{\substack{\pi\in\mathrm{S}_{2t+2}\\ \pi\neq\tau_1\tau_2}} W^{\mathbb{U}(d),(2t+2)}_{\tau_1\tau_2,\pi} \langle\!\langle R_\pi | (\rho^{\otimes t}\otimes O)^{\otimes 2}\rangle\!\rangle \\
&= d(d-1) \sum_{\tau_1,\tau_2\in\mathrm{S}_{t+1}} d^{-(2t+2)}\big(1+\mathcal{O}(d^{-1})\big) \langle\!\langle R_{\tau_1} | \rho^{\otimes t}\otimes O\rangle\!\rangle \langle\!\langle R_{\tau_2} | \rho^{\otimes t}\otimes O\rangle\!\rangle + d(d-1)d^{-(2t+2)}\mathcal{O}(d^{-1}B) \\
&= d^{-2t}\left(\sum_{\tau\in\mathrm{S}_{t+1}} \mathrm{Tr}\big[(\rho^{\otimes t}\otimes O)R_\tau\big]\right)^2 + \mathcal{O}\left(\frac{B}{d^{2t+1}}\right).
\end{aligned} \tag{25}
$$

where the forth equality follows from the fact that $W^{\mathbb{U}(d),(2t+2)}_{\pi,\pi'} = d^{-(2t+2)}\big(\delta_{\pi,\pi'} + \mathcal{O}(d^{-1})\big)$.

We then group the estimators and reduce the summation in Eq. (20) to

$$
\begin{aligned}
&\mathbb{E}_U \mathbb{E}_{\hat{\mathbf{r}}}\Big(\hat{M}_U(O)^2\Big|U\Big) \\
&= \binom{N_M}{t}^{-2}\left(\frac{(t+1)!\kappa_{t+1}}{d}\right)^2 \left[\sum_{k=0}^{t} \sum_{\mathrm{Co}[(i_1,\ldots,i_t);(j_1,\ldots,j_t)]=t-k} \mathbb{E}_U\mathbb{E}_{\hat{\mathbf{r}}}\Big[f(\hat{r}_{i_1},\ldots,\hat{r}_{i_t},\hat{r}_{j_1},\ldots,\hat{r}_{j_t},U,O)\Big|U\Big]\right] \\
&= \binom{N_M}{t}^{-2}\left(\frac{(t+1)!\kappa_{t+1}}{d}\right)^2 \left[\sum_{k=0}^{t} \binom{N_M}{t+k}\binom{t+k}{t}\binom{t}{k} \mathbb{E}_U\mathbb{E}_{\hat{\mathbf{r}}}\Big[f(\hat{r}_{i_1},\ldots,\hat{r}_{i_t},\hat{r}_{j_1},\ldots,\hat{r}_{j_t},U,O)\Big|U\Big]\right] \\
&\leq \binom{N_M}{t}^{-2}\left(\frac{(t+1)!\kappa_{t+1}}{d}\right)^2 \left\{\sum_{k=0}^{t-1} \binom{N_M}{t+k}\binom{t+k}{t}\binom{t}{k}\mathcal{O}\left(\frac{B}{d^{t+k+1}}\right)\right. \\
&\qquad\qquad \left. + \binom{N_M}{2t}\binom{2t}{t}\left[d^{-2t}\left(\sum_{\tau\in\mathrm{S}_{t+1}}\mathrm{Tr}\big[(\rho^{\otimes t}\otimes O)R_\tau\big]\right)^2 + \mathcal{O}\left(\frac{B}{d^{2t+1}}\right)\right]\right\} \\
&= \sum_{k=0}^{t-1}\binom{N_M}{t}^{-2}\binom{N_M}{t+k}\mathcal{O}(d^{t-1-k}B) + \binom{N_M}{t}^{-2}\binom{N_M}{2t}\binom{2t}{t}\big[1+\mathcal{O}(d^{-1})\big]\left[\left(\sum_{\tau\in\mathrm{S}_{t+1}}\mathrm{Tr}\big[(\rho^{\otimes t}\otimes O)R_\tau\big]\right)^2 + \mathcal{O}\left(\frac{B}{d}\right)\right] \\
&\leq \sum_{k=0}^{t-1}\mathcal{O}\left(\frac{d^{t-1-k}B}{N_M^{t-k}}\right) + \left(\sum_{\tau\in\mathrm{S}_{t+1}}\mathrm{Tr}\big[(\rho^{\otimes t}\otimes O)R_\tau\big]\right)^2 + \mathcal{O}\left(\frac{B}{d}\right) \\
&= \sum_{i=0}^{t}\mathcal{O}\left(\frac{d^{i-1}B}{N_M^{i}}\right) + \left(\sum_{\tau\in\mathrm{S}_{t+1}}\mathrm{Tr}\big[(\rho^{\otimes t}\otimes O)R_\tau\big]\right)^2,
\end{aligned} \tag{26}
$$

where the first equality follows from Eqs. (22), (24), (25), and the second equality holds because $\binom{N_M}{t}^{-2}\binom{N_M}{2t}\binom{2t}{t} \leq 1$.

Equations (11) and (5) together imply that

$$
\mathrm{Var}[\hat{M}(O)] = \frac{1}{N_U}\,\mathrm{Var}[\hat{M}_U(O)] = \frac{1}{N_U}\left[\mathbb{E}_U\mathbb{E}_{\hat{\mathbf{r}}}\Big(\hat{M}_U(O)^2\Big|U\Big) - \left(\sum_{\tau\in\mathrm{S}_{t+1}}\mathrm{Tr}\big[(\rho^{\otimes t}\otimes O)R_\tau\big]\right)^2\right] \leq \frac{1}{N_U}\sum_{i=0}^{t}\mathcal{O}\left(\frac{d^{i-1}B}{N_M^i}\right). \tag{27}
$$

To ensure that the additive error is $\mathcal{O}(\epsilon)$, we need to bound this variance by $\mathcal{O}(\epsilon^2)$. Consider two cases below:

1. $d < B/\epsilon^2$.
   In this case, it suffices to let $N_U = \mathcal{O}(Bd^{-1}\epsilon^{-2})$ and $N_M = \mathcal{O}(d)$. The sample cost is $N_M N_U = \mathcal{O}(B/\epsilon^2)$.

2. $d \geq B/\epsilon^2$.
   In this case, it suffices to let $N_U = \mathcal{O}(1)$ and $N_M = \mathcal{O}\big(d^{1-1/t}B^{1/t}\epsilon^{-2/t}\big)$. So the sample cost is $\mathcal{O}(N_M)$.

To summarize, we have

$$N_U = \max\left\{1, \frac{B}{d\epsilon^2}\right\}, \qquad N_M = \min\left\{\frac{d^{1-1/t}B^{1/t}}{\epsilon^{2/t}}, d\right\}, \qquad N_U N_M = \max\left\{\frac{d^{1-1/t}B^{1/t}}{\epsilon^{2/t}}, \frac{B}{\epsilon^2}\right\}, \tag{28}$$

where we ignored constant coefficients. These results confirm the complexity illustrated in Theorem 2. Notably, in the usual case $d \geq B/\epsilon^2$, we only need $N_U = 1$ random unitaries to estimate $\mathrm{Tr}(O\rho^t)$ within additive error ϵ.

III. ESTIMATING PROPERTIES OF PRINCIPAL EIGENSTATES

Suppose ρ is an unknown quantum state with spectral decomposition $\rho = \sum_{i=1}^{d} \lambda_i |\psi_i\rangle\langle\psi_i|$, where the eigenvalues are arranged in decreasing order $\lambda := \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$. Additionally, we assume that $\lambda := \lambda_1$ is larger than λ_2 by a constant value $0 < \mu \leq 1$, i.e., $\lambda - \lambda_2 \geq \mu$. Our goal is to predict properties of the principal eigenstate $|\psi\rangle := |\psi_1\rangle$, that is, to predict the value of $\langle\psi|O|\psi\rangle$, where O is an Hermitian observable with $\|O\|_\infty \leq 1$.

Theorem 3. *Suppose $0 < \epsilon < 1$ is a constant. By using single-copy measurements on $\mathcal{O}(d)$ copies of ρ, the quantity $\langle\psi|O|\psi\rangle$ can be estimated within ϵ additive error. In addition, the circuit complexity is only $\mathcal{O}(1)$.*

Lemma 1. *Suppose $0 < \epsilon < 1$ and integer $t \geq \lceil \log_{1-\mu/3} \frac{\mu\epsilon}{6} \rceil$. Then*

$$\left| \frac{\mathrm{Tr}(O\rho^t)}{\mathrm{Tr}(\rho^t)} - \langle\psi|O|\psi\rangle \right| \leq \epsilon. \tag{29}$$

Proof of Lemma 1. First, note that ρ can be expressed as

$$\rho = \lambda|\psi\rangle\langle\psi| + \sum_{j=1}^{m} p_j \tau_j, \tag{30}$$

where τ_j are quantum states on $\mathcal{H}$ with $\mathrm{Tr}(\tau_i\tau_j) = \delta_{i,j}$, $p_j \in [\mu/3, \lambda - \mu/3]$ for $j = 1, \ldots, m-1$, and $p_m \in (0, \lambda - \mu/3]$. Since $\mathrm{Tr}(\rho) = \lambda + \sum_{j=1}^{m} p_j = 1$, we have

$$m \leq \left\lceil \frac{1-\lambda}{\max_j p_j} \right\rceil \leq \left\lceil \frac{1-\lambda}{\mu/3} \right\rceil \leq \frac{3}{\mu}. \tag{31}$$

By Eq. (30), we have

$$\lambda^t \leq \mathrm{Tr}\big(\rho^t\big) = \lambda^t + \sum_{j=1}^{m} p_j^t \mathrm{Tr}\big(\tau_j^t\big) \leq \lambda^t + \sum_{j=1}^{m} p_j^t, \tag{32}$$

$$\mathrm{Tr}\big(O\rho^t\big) = \lambda^t \langle\psi|O|\psi\rangle + \sum_{j=1}^{m} p_j^t \mathrm{Tr}\big(O\tau_j^t\big). \tag{33}$$

It follows that

$$\begin{aligned}
\left| \frac{\mathrm{Tr}(O\rho^t)}{\mathrm{Tr}(\rho^t)} - \langle\psi|O|\psi\rangle \right| &\leq \max\left\{ \left| \frac{\sum_{j=1}^{m} p_j^t \mathrm{Tr}\big(O\tau_j^t\big)}{\lambda^t} \right|, \left| \frac{\lambda^t\langle\psi|O|\psi\rangle + \sum_{j=1}^{m} p_j^t \mathrm{Tr}\big(O\tau_j^t\big)}{\lambda^t + \sum_{j=1}^{m} p_j^t} - \langle\psi|O|\psi\rangle \right| \right\} \\
&\leq \max\left\{ \sum_{j=1}^{m} \Big(\frac{p_j}{\lambda}\Big)^t \big|\mathrm{Tr}\big(O\tau_j^t\big)\big|, \frac{\sum_{j=1}^{m} p_j^t \big|\mathrm{Tr}\big(O\tau_j^t\big) - \langle\psi|O|\psi\rangle\big|}{\lambda^t + \sum_{j=1}^{m} p_j^t} \right\} \\
&\leq \max\left\{ \sum_{j=1}^{m} \Big(\frac{p_j}{\lambda}\Big)^t \|O\|_\infty, \sum_{j=1}^{m} \Big(\frac{p_j}{\lambda}\Big)^t \big|\mathrm{Tr}\big(O\tau_j^t\big) - \langle\psi|O|\psi\rangle\big| \right\} \\
&\leq 2\sum_{j=1}^{m} \Big(\frac{p_j}{\lambda}\Big)^t \leq 2m\Big(\frac{\max_j p_j}{\lambda}\Big)^t \leq \frac{6}{\mu}\Big(\frac{\lambda - \mu/3}{\lambda}\Big)^t \leq \frac{6}{\mu}\Big(1 - \frac{\mu}{3}\Big)^t \leq \epsilon,
\end{aligned} \tag{34}$$

where the last inequality follows from the assumption $t \geq \lceil \log_{1-\mu/3} \frac{\mu\epsilon}{6} \rceil$. This completes the proof. □

Lemma 2. *Suppose $0 < \epsilon < 1$, integer $t \geq 1$, $a, b \in \mathbb{R}$, $|a - \mathrm{Tr}(O\rho^t)| \leq \mu^t\epsilon/3$, and $|b - \mathrm{Tr}(\rho^t)| \leq \mu^t\epsilon/3$. Then*

$$\left| \frac{\mathrm{Tr}(O\rho^t)}{\mathrm{Tr}(\rho^t)} - \frac{a}{b} \right| \leq \epsilon. \tag{35}$$

Proof of Lemma 2. Let $\tilde{\epsilon} := \mu^t \epsilon/3$. Then we have

$$\begin{aligned}
\left|\frac{\mathrm{Tr}(O\rho^t) \pm \tilde{\epsilon}}{\mathrm{Tr}(\rho^t) - \tilde{\epsilon}} - \frac{\mathrm{Tr}(O\rho^t)}{\mathrm{Tr}(\rho^t)}\right| &\leq \left|\frac{\mathrm{Tr}(O\rho^t) \pm \tilde{\epsilon}}{\mathrm{Tr}(\rho^t) - \tilde{\epsilon}} - \frac{\mathrm{Tr}(O\rho^t)}{\mathrm{Tr}(\rho^t) - \tilde{\epsilon}}\right| + \left|\frac{\mathrm{Tr}(O\rho^t)}{\mathrm{Tr}(\rho^t) - \tilde{\epsilon}} - \frac{\mathrm{Tr}(O\rho^t)}{\mathrm{Tr}(\rho^t)}\right| \\
&\leq \left|\frac{\tilde{\epsilon}}{\mathrm{Tr}(\rho^t) - \tilde{\epsilon}}\right| + \left|\frac{\mathrm{Tr}(O\rho^t)\tilde{\epsilon}}{\mathrm{Tr}(\rho^t)[\mathrm{Tr}(\rho^t) - \tilde{\epsilon}]}\right| \\
&\leq \left(\left|\frac{\mathrm{Tr}(O\rho^t)}{\mathrm{Tr}(\rho^t)}\right| + 1\right) \cdot \left|\frac{\tilde{\epsilon}}{\mathrm{Tr}(\rho^t) - \tilde{\epsilon}}\right| \\
&\leq \frac{2\tilde{\epsilon}}{|\mathrm{Tr}(\rho^t) - \tilde{\epsilon}|} \leq \frac{2\mu^t \epsilon/3}{\mu^t - \mu^t\epsilon/3} = \frac{2\epsilon}{3 - \epsilon} \leq \epsilon. \qquad (36)
\end{aligned}$$

In the last line we used the fact $\mathrm{Tr}(\rho^t) \geq \lambda^t \geq \mu^t$. Similarly, we have

$$\begin{aligned}
\left|\frac{\mathrm{Tr}(O\rho^t) \pm \tilde{\epsilon}}{\mathrm{Tr}(\rho^t) + \tilde{\epsilon}} - \frac{\mathrm{Tr}(O\rho^t)}{\mathrm{Tr}(\rho^t)}\right| &\leq \left|\frac{\mathrm{Tr}(O\rho^t) \pm \tilde{\epsilon}}{\mathrm{Tr}(\rho^t) + \tilde{\epsilon}} - \frac{\mathrm{Tr}(O\rho^t)}{\mathrm{Tr}(\rho^t) + \tilde{\epsilon}}\right| + \left|\frac{\mathrm{Tr}(O\rho^t)}{\mathrm{Tr}(\rho^t) + \tilde{\epsilon}} - \frac{\mathrm{Tr}(O\rho^t)}{\mathrm{Tr}(\rho^t)}\right| \\
&\leq \left|\frac{\tilde{\epsilon}}{\mathrm{Tr}(\rho^t) + \tilde{\epsilon}}\right| + \left|\frac{\mathrm{Tr}(O\rho^t)\tilde{\epsilon}}{\mathrm{Tr}(\rho^t)[\mathrm{Tr}(\rho^t) + \tilde{\epsilon}]}\right| \\
&\leq \left(\left|\frac{\mathrm{Tr}(O\rho^t)}{\mathrm{Tr}(\rho^t)}\right| + 1\right) \cdot \frac{\tilde{\epsilon}}{\mathrm{Tr}(\rho^t) + \tilde{\epsilon}} \leq \frac{2\tilde{\epsilon}}{\mathrm{Tr}(\rho^t)} \leq \frac{2\mu^t\epsilon/3}{\mu^t} \leq \epsilon. \qquad (37)
\end{aligned}$$

Therefore,

$$\left|\frac{a}{b} - \frac{\mathrm{Tr}(O\rho^t)}{\mathrm{Tr}(\rho^t)}\right| \leq \max\left\{\left|\frac{\mathrm{Tr}(O\rho^t) \pm \tilde{\epsilon}}{\mathrm{Tr}(\rho^t) - \tilde{\epsilon}} - \frac{\mathrm{Tr}(O\rho^t)}{\mathrm{Tr}(\rho^t)}\right|, \left|\frac{\mathrm{Tr}(O\rho^t) \pm \tilde{\epsilon}}{\mathrm{Tr}(\rho^t) + \tilde{\epsilon}} - \frac{\mathrm{Tr}(O\rho^t)}{\mathrm{Tr}(\rho^t)}\right|\right\} \leq \epsilon. \qquad (38)$$

This completes the proof. □

IV. ESTIMATION OF NEGATIVITY-MOMENTS

Consider that we are given multiple copies of an unknown quantum state ρ on the Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$. Let n_A and n_B be the qubit number of subsystems A and B, respectively. W.l.o.g., we assume that $n := n_A + n_B$ is even and $n_A \geq n_B$. Our goal is to estimate the t-th order negativity-moment $\mathrm{Tr}\Big[\big(\rho^{\top_B}\big)^t\Big]$, where t is a constant integer.

Theorem 4. *Suppose* $0 < \epsilon < 1$, $d \geq \Omega(\epsilon^{-2})$, *and* $t = \mathcal{O}(1)$ *is a positive integer. There exists an algorithm that can estimate* $\mathrm{Tr}\Big[\big(\rho^{\top_B}\big)^t\Big]$ *within* ϵ *error by using single-copy measurements. The sample complexity, classical computational complexity, and number of random unitaries used by the algorithm are* $\mathcal{O}\Big(d_{A_1}^{2-2/t} d_{A_2}^2 \epsilon^{-2}\Big)$, $\mathcal{O}\Big(d_{A_1}^{2-2/t} d_{A_2}^2 \epsilon^{-2}\Big)$, *and* $\mathcal{O}\big(\epsilon^{-2}\big)$, *respectively.*

A. Estimation protocol

Notice that if we can estimate $\sum_{\pi \in \mathrm{S}_s} \mathrm{Tr}\Big(\big(\rho^{\top_B}\big)^{\otimes s} R_\pi\Big)$ within $\mathcal{O}(\epsilon)$ additive error for all $s = 1, 2, \ldots, t$, then we can estimate $\mathrm{Tr}\Big[\big(\rho^{\top_B}\big)^t\Big]$ within $\mathcal{O}(\epsilon)$ error. To achieve this goal, our protocol runs as follows. First, we divide subsystem A into three components A_1, A_2, and A_3, whose qubit numbers satisfy $n_{A_1} = n_{A_3}$ and $n_{A_2} = n_B$. Then we do the following procedure for $q = 1, 2, \ldots, N_U$:

1. Sample a random unitary $U_q \sim \mathbb{U}(\mathcal{H}_{A_1 \cup A_2})$.

2. Prepare N_M copies of ρ, and apply U_q on the subsystem $A_1 \cup A_2$ of ρ.

3. For each of the rotated state $\tilde{\rho}_{U_q} = (U_q \otimes I_{A_3,B})\rho(U_q^\dagger \otimes I_{A_3,B})$, we

   (1) measure subsystems A_1 and A_3 in the Bell basis $\{|\Phi_b\rangle\langle\Phi_b|\}_{b=1}^{d_{A_1}^2}$, and obtain outcomes $\hat{\mathbf{b}} = \Big\{\hat{b}_1, \ldots, \hat{b}_{N_M}\Big\}$. Here $|\Phi_b\rangle = \frac{1}{\sqrt{d_{A_1}}} \sum_{i=1}^{d_{A_1}} |i\rangle_{A_1} \otimes W_b |i\rangle_{A_3}$, and W_b are unitary operators with $\mathrm{Tr}\Big(W_b^\dagger W_{b'}\Big) = d_{A_1}\delta_{b,b'}$

   (2) measure subsystems A_2 and B with the POVM $\Big\{\Pi_{\mathrm{sym}}^{(2)}, \Pi_{\mathrm{asym}}^{(2)}\Big\}$ (equivalently, one can perform the swap test on A_2 and B), and obtain outcomes $\hat{\mathbf{r}} = \{\hat{r}_1, \ldots, \hat{r}_{N_M}\}$, where $\hat{r}_j \in \{+1, -1\}$.

4. Compute the following estimator using $\hat{\mathbf{b}}$ and $\hat{\mathbf{r}}$:

$$\hat{M}_U^{\mathrm{neg}} := \binom{N_M}{t}^{-1} d_{A_1}^{2t-2} d_{A_2}^t \sum_{i_1 < i_2 < \cdots < i_t} (\hat{r}_{i_1} \hat{r}_{i_2} \cdots \hat{r}_{i_t}) \mathbf{1}\Big[\hat{b}_{i_1} = \hat{b}_{i_2} = \cdots = \hat{b}_{i_t}\Big]. \tag{39}$$

Finally, output $\hat{M}^{\mathrm{neg}} := N_U^{-1} \sum_q \hat{M}_{U_q}^{\mathrm{neg}}$ as our estimate of $\sum_{\pi \in \mathrm{S}_t} \mathrm{Tr}\Big(\big(\rho^{\top_B}\big)^{\otimes t} R_\pi\Big)$.

Next, we shall explain why $\hat{M}^{\text{neg}}$ is a valid estimator. Its expectation value is

$$
\begin{aligned}
\mathbb{E}\left[\hat{M}^{\text{neg}}\right] &= \mathbb{E}_U \mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\left(\hat{M}_U^{\text{neg}}\middle|U\right) = d_{A_1}^{2t-2} d_{A_2}^{t}\, \mathbb{E}_U \mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\left\{ (\hat{r}_{i_1}\hat{r}_{i_2}\cdots\hat{r}_{i_t})\mathbf{1}\left[\hat{b}_{i_1}=\cdots=\hat{b}_{i_t}\right]\middle|U\right\} \\
&= d_{A_1}^{2t-2} d_{A_2}^{t}\, \mathbb{E}_U\left\{\sum_{\ell=0}^{t}\sum_{b=1}^{d_{A_1}^2}(-1)^{\ell}\Pr\left(\#[-1 \text{ among } \{\hat{r}_{i_1},\hat{r}_{i_2},\ldots,\hat{r}_{i_t}\}]=\ell, \text{ and } \hat{b}_{i_1}=\cdots=\hat{b}_{i_t}=b\middle|U\right)\right\} \\
&= d_{A_1}^{2t-2} d_{A_2}^{t}\, \mathbb{E}_U\left\{\sum_{\ell=0}^{t}\sum_{b=1}^{d_{A_1}^2}\binom{t}{\ell}(-1)^{\ell}\Pr\left(\hat{r}=-1,\hat{b}=b\middle|U\right)^{\ell}\Pr\left(\hat{r}=+1,\hat{b}=b\middle|U\right)^{t-\ell}\right\} \\
&= d_{A_1}^{2t-2} d_{A_2}^{t}\, \mathbb{E}_U\left\{\sum_{b=1}^{d_{A_1}^2}\left[\Pr\left(\hat{r}=+1,\hat{b}=b\middle|U\right)-\Pr\left(\hat{r}=-1,\hat{b}=b\middle|U\right)\right]^{t}\right\} \\
&= d_{A_1}^{2t-2} d_{A_2}^{t}\sum_{b=1}^{d_{A_1}^2}\mathbb{E}_U\left\{\left(\operatorname{Tr}\left[\tilde{\rho}_U(|\Phi_b\rangle\langle\Phi_b|)_{A_1,A_3}\otimes(\Pi_{\text{sym}}^{(2)})_{A_2,B}\right]-\operatorname{Tr}\left[\tilde{\rho}_U(|\Phi_b\rangle\langle\Phi_b|)_{A_1,A_3}\otimes(\Pi_{\text{asym}}^{(2)})_{A_2,B}\right]\right)^{t}\right\} \\
&= d_{A_1}^{2t-2} d_{A_2}^{t}\sum_{b=1}^{d_{A_1}^2}\mathbb{E}_U\left\{\operatorname{Tr}[\tilde{\rho}_U(|\Phi_b\rangle\langle\Phi_b|)_{A_1,A_3}\otimes \mathrm{SWAP}_{A_2,B}]^{t}\right\} = d_{A_1}^{2t-2} d_{A_2}^{t}\sum_{b=1}^{d_{A_1}^2} d_{A_1}^{-t}\operatorname{Tr}\left(\rho^{\otimes t}\tilde{H}_{b,t}\right)
\end{aligned}
\tag{40}
$$

Here, $\tilde{H}_t$ in the last line is defined as

$$
\tilde{H}_{b,t} := \mathbb{E}_U\left\{\left[\left(U_{A_1,A_2}^{\dagger}\otimes I_{A_3,B}\right)\left(d_{A_1}(|\Phi_b\rangle\langle\Phi_b|)_{A_1,A_3}\otimes \mathrm{SWAP}_{A_2,B}\right)\left(U_{A_1,A_2}\otimes I_{A_3,B}\right)\right]^{\otimes t}\right\}. \tag{41}
$$

According to Lemma 3 below, we have

$$
\begin{aligned}
\tilde{H}_{b,t} &= \sum_{\pi,\tau\in S_t}\mathrm{Wg}(\pi^{-1}\tau, d_{A_1}d_{A_2})(R_\pi)_{A_1}\otimes(R_\pi)_{A_2}\otimes(R_\tau)_{A_3}\otimes(R_\tau^{\top})_B \\
&= (d_{A_1}d_{A_2})^{-t}\left[\sum_{\pi\in S_t}(R_\pi)_A\otimes(R_\pi^{\top})_B+\mathcal{O}\left(\frac{1}{d_{A_1}d_{A_2}}\right)\sum_{\pi,\tau\in S_t}(R_\pi)_{A_1}\otimes(R_\pi)_{A_2}\otimes(R_\tau)_{A_3}\otimes(R_\tau^{\top})_B\right],
\end{aligned}
\tag{42}
$$

where the second equality follows from the relation $\mathrm{Wg}(\pi^{-1}\tau,d)=d^{-t}\left(\delta_{\pi,\tau}+\mathcal{O}(d^{-1})\right)$.

By combining Eqs. (40) and (42), the expectation of $\hat{M}^{\text{neg}}$ can be further expressed as

$$
\begin{aligned}
\mathbb{E}\left[\hat{M}^{\text{neg}}\right] &= \sum_{\pi\in S_t}\operatorname{Tr}\left[\rho_{AB}^{\otimes t}(R_\pi)_A\otimes(R_\pi^{\top})_B\right]+\mathcal{O}\left(\frac{1}{d_{A_1}d_{A_2}}\right)\sum_{\pi,\tau\in S_t}\operatorname{Tr}\left[\rho_{AB}^{\otimes t}(R_\pi)_{A_1}\otimes(R_\pi)_{A_2}\otimes(R_\tau)_{A_3}\otimes(R_\tau^{\top})_B\right] \\
&= \sum_{\pi\in S_t}\operatorname{Tr}\left[\left(\rho_{AB}^{\top_B}\right)^{\otimes t}(R_\pi)_A\otimes(R_\pi)_B\right]+\mathcal{O}\left(\frac{1}{d_{A_1}d_{A_2}}\right) \\
&= \sum_{\pi\in \mathrm{S}_t}\operatorname{Tr}\left(\left(\rho^{\top_B}\right)^{\otimes t}R_\pi\right)+\mathcal{O}\left(\frac{1}{d_{A_1}d_{A_2}}\right).
\end{aligned}
\tag{43}
$$

Therefore, in the usual case with $2^{-n/2}=(d_{A_1}d_{A_2})^{-1}\le\mathcal{O}(\epsilon)$, $\hat{M}^{\text{neg}}$ is a valid estimator for $\sum_{\pi\in\mathrm{S}_t}\operatorname{Tr}\left(\left(\rho^{\top_B}\right)^{\otimes t}R_\pi\right)$. If we can estimate $\hat{M}^{\text{neg}}$ within error $\mathcal{O}(\epsilon)$, we can then estimate $\sum_{\pi\in\mathrm{S}_t}\operatorname{Tr}\left(\left(\rho^{\top_B}\right)^{\otimes t}R_\pi\right)$ within error $\mathcal{O}(\epsilon)$.

Lemma 3 (Eq. (119) in [1])**.** *Suppose $U\in\mathbb{U}(d)$ and k is a positive integer, then*

$$
\mathop{\mathbb{E}}_{U\sim\mu_H}\left[U^{\otimes k}\otimes U^{*\otimes k}\right]=\sum_{\pi,\tau\in S_k}\mathrm{Wg}(\pi^{-1}\sigma,d)|R_\pi\rangle\rangle\langle\langle R_\tau|, \tag{44}
$$

where $|R_\pi\rangle\rangle=\sum_{i_1,i_2,\ldots,i_k=1}^{d}R_\pi|i_1,i_2,\ldots,i_k\rangle\otimes|i_1,i_2,\ldots,i_k\rangle$ denotes the vectorization of R_π.

B. Variance analyses

The variance of $\hat{M}^{\mathrm{neg}}$ is $\mathrm{Var}[\hat{M}^{\mathrm{neg}}] = N_U^{-1}\,\mathrm{Var}[\hat{M}_U^{\mathrm{neg}}]$, where

$$\mathrm{Var}\left[\hat{M}_U^{\mathrm{neg}}\right] = \mathbb{E}\left[\left(\hat{M}_U^{\mathrm{neg}}\right)^2\right] - \mathbb{E}\left[\hat{M}_U^{\mathrm{neg}}\right]^2 \le \mathbb{E}_U\mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\left[\left(\hat{M}_U^{\mathrm{neg}}\right)^2\middle|U\right]. \tag{45}$$

Note that

$$\mathbb{E}_U\mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\left[\left(\hat{M}_U^{\mathrm{neg}}\right)^2\middle|U\right] = \binom{N_M}{t}^{-2} d_{A_1}^{4t-4} d_{A_2}^{2t} \sum_{\substack{i_1<i_2<\cdots<i_t\\ j_1<j_2<\cdots<j_t}} \mathbb{E}_U\mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\left[f\left(\hat{\mathbf{b}},\hat{\mathbf{r}},\{i_1,\dots,i_t,j_1,\dots,j_t\}\right)\middle|U\right]. \tag{46}$$

where

$$f\left(\hat{\mathbf{b}},\hat{\mathbf{r}},\{i_1,\dots,i_t,j_1,\dots,j_t\}\right) := (\hat{r}_{i_1}\hat{r}_{i_2}\cdots\hat{r}_{i_t}\hat{r}_{j_1}\hat{r}_{j_2}\cdots\hat{r}_{j_t})\mathbf{1}\left[\hat{b}_{i_1}=\cdots=\hat{b}_{i_t}\right]\mathbf{1}\left[\hat{b}_{j_1}=\cdots=\hat{b}_{j_t}\right]. \tag{47}$$

Based on the collision number (denoted as $\mathrm{Co}[(i_1,i_2,\dots,i_t);(j_1,j_2,\dots,j_t)]$), the expectation value in the last bracket will be different. When $\mathrm{Co}[(i_1,\dots,i_t);(j_1,\dots,j_t)] = t-k$, where $k\in\{0,1,\dots,t-1\}$, we have

$$\begin{aligned}
&\mathbb{E}_U\mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\left[f\left(\hat{\mathbf{b}},\hat{\mathbf{r}},\{i_1,\dots,i_t,j_1,\dots,j_t\}\right)\middle|U\right]\\
&\quad= \mathbb{E}_U\mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\left((\hat{r}_{i_1}\hat{r}_{i_2}\cdots\hat{r}_{i_t}\hat{r}_{j_1}\hat{r}_{j_2}\cdots\hat{r}_{j_t})\mathbf{1}\left[\hat{b}_{i_1}=\cdots=\hat{b}_{i_t}=\hat{b}_{j_1}=\cdots=\hat{b}_{j_t}\right]\middle|U\right).
\end{aligned} \tag{48}$$

Wlog, we assume that $(i_1,i_2,\dots,i_t)$ and $(j_1,j_2,\dots,j_t)$ coincide on the last $t-k$ pairs. Then the above quantity can be rewritten as

$$\begin{aligned}
&\mathbb{E}_U\mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\left\{(\hat{r}_{i_1}\hat{r}_{i_2}\cdots\hat{r}_{i_k}\hat{r}_{j_1}\hat{r}_{j_2}\cdots\hat{r}_{j_k})\mathbf{1}\left[\hat{b}_{i_1}=\cdots=\hat{b}_{i_t}=\hat{b}_{j_1}=\cdots=\hat{b}_{j_k}\right]\middle|U\right\}\\
&= \mathbb{E}_U\left\{\sum_{\ell=0}^{2k}\sum_{b=1}^{d_{A_1}^2}(-1)^\ell\,\Pr\left(\#[-1 \text{ among } \{\hat{r}_{i_1},\dots,\hat{r}_{i_t},\hat{r}_{j_1},\dots,\hat{r}_{j_k}\}]=\ell, \text{ and } \hat{b}_{i_1}=\cdots=\hat{b}_{i_t}=\hat{b}_{j_1}=\cdots=\hat{b}_{j_k}=b\middle|U\right)\right\}\\
&= \mathbb{E}_U\left\{\sum_{\ell=0}^{2k}\sum_{b=1}^{d_{A_1}^2}\binom{2k}{\ell}(-1)^\ell\,\Pr\left(\hat{r}=-1,\hat{b}=b\middle|U\right)^\ell\Pr\left(\hat{r}=+1,\hat{b}=b\middle|U\right)^{2k-\ell}\Pr\left(\hat{b}=b\middle|U\right)^{t-k}\right\}\\
&= \mathbb{E}_U\left\{\sum_{b=1}^{d_{A_1}^2}\Pr\left(\hat{b}=b\middle|U\right)^{t-k}\sum_{\ell=0}^{2k}\binom{2k}{\ell}(-1)^\ell\,\Pr\left(\hat{r}=-1,\hat{b}=b\middle|U\right)^\ell\Pr\left(\hat{r}=+1,\hat{b}=b\middle|U\right)^{2k-\ell}\right\}\\
&= \mathbb{E}_U\left\{\sum_{b=1}^{d_{A_1}^2}\Pr\left(\hat{b}=b\middle|U\right)^{t-k}\left[\Pr\left(\hat{r}=+1,\hat{b}=b\middle|U\right)-\Pr\left(\hat{r}=-1,\hat{b}=b\middle|U\right)\right]^{2k}\right\}\\
&= \sum_{b=1}^{d_{A_1}^2}\mathbb{E}_U\left\{\mathrm{Tr}\left[\tilde{\rho}_U(|\Phi_b\rangle\langle\Phi_b|)_{A_1,A_3}\otimes I_{A_2,B}\right]^{t-k}\mathrm{Tr}\left[\tilde{\rho}_U(|\Phi_b\rangle\langle\Phi_b|)_{A_1,A_3}\otimes \mathrm{SWAP}_{A_2,B}\right]^{2k}\right\}\\
&= \sum_{b=1}^{d_{A_1}^2} d_{A_1}^{-(t+k)}\,\mathrm{Tr}\left(\rho^{\otimes(t+k)}\mathcal{Q}_{b,t,k}\right)
\end{aligned} \tag{49}$$

Here, $\mathcal{Q}_{b,t,k}$ in the last line is defined as

$$\begin{aligned}
\mathcal{Q}_{b,t,k} := \mathbb{E}_U\Bigg\{&\left[\left(U_{A_1,A_2}^\dagger\otimes I_{A_3,B}\right)\left(d_{A_1}(|\Phi_b\rangle\langle\Phi_b|)_{A_1,A_3}\otimes I_{A_2,B}\right)\left(U_{A_1,A_2}\otimes I_{A_3,B}\right)\right]^{\otimes(t-k)}\\
&\otimes\left[\left(U_{A_1,A_2}^\dagger\otimes I_{A_3,B}\right)\left(d_{A_1}(|\Phi_b\rangle\langle\Phi_b|)_{A_1,A_3}\otimes \mathrm{SWAP}_{A_2,B}\right)\left(U_{A_1,A_2}\otimes I_{A_3,B}\right)\right]^{\otimes 2k}\Bigg\}.
\end{aligned} \tag{50}$$

According to Lemma 3, we have

$$\begin{aligned}
\mathcal{Q}_{b,t,k} &= \sum_{\pi,\tau\in S_{t+k}} \mathrm{Wg}(\pi^{-1}\tau, d_{A_1}d_{A_2})(R_\pi)_{A_1}\otimes(R_\pi)_{A_2}\otimes(R_\tau)_{A_3}\otimes\Big\{I_B^{\otimes(t-k)}\otimes \mathrm{Tr}_{1,2,\ldots,t-k}\big[\big(R_\tau^\top\big)_B\big]\Big\} \\
&= (d_{A_1}d_{A_2})^{-(t+k)}d_B^{t-k}\sum_{\pi,\tau\in S_{t+k}}\mathcal{O}(1)(R_\pi)_{A_1}\otimes(R_\pi)_{A_2}\otimes(R_\tau)_{A_3}\otimes(R_{f(\tau)})_B \\
&= \mathcal{O}\Big(d_{A_1}^{-(t+k)}d_{A_2}^{-2k}\Big)\sum_{\pi,\tau\in S_{t+k}}(R_\pi)_{A_1}\otimes(R_\pi)_{A_2}\otimes(R_\tau)_{A_3}\otimes(R_{f(\tau)})_B.
\end{aligned} \tag{51}$$

It follows that

$$\mathbb{E}_U\mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\Big[f\Big(\hat{\mathbf{b}},\hat{\mathbf{r}},\{i_1,\ldots,i_t,j_1,\ldots,j_t\}\Big)\Big|U\Big] = \sum_{b=1}^{d_{A_1}^2} d_{A_1}^{-(t+k)}\,\mathrm{Tr}\Big(\rho^{\otimes(t+k)}\mathcal{Q}_{b,t,k}\Big) = \mathcal{O}\big(d_{A_1}^{2-2k-2t}d_{A_2}^{-2k}\big). \tag{52}$$

When $\mathrm{Co}[(i_1,\ldots,i_t);(j_1,\ldots,j_t)]=0$, we have

$$\begin{aligned}
&\mathbb{E}_U\mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\Big[f\Big(\hat{\mathbf{b}},\hat{\mathbf{r}},\{i_1,\ldots,i_t,j_1,\ldots,j_t\}\Big)\Big|U\Big] \\
&= \mathbb{E}_U\bigg\{\Big[\mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\Big(\hat{r}_{i_1}\hat{r}_{i_2}\cdots\hat{r}_{i_t}\mathbf{1}\big[\hat{b}_{i_1}=\cdots=\hat{b}_{i_t}\big]\Big|U\Big)\Big]^2\bigg\} \\
&= \mathbb{E}_U\left\{\left[\sum_{\ell=0}^{t}\sum_{b=1}^{d_{A_1}^2}(-1)^\ell\,\mathrm{Pr}\left(\#[-1\text{ among }\{\hat{r}_{i_1},\ldots,\hat{r}_{i_t}\}]=\ell,\text{ and }\hat{b}_{i_1}=\cdots=\hat{b}_{i_t}=b\,\Big|U\right)\right]^2\right\} \\
&= \mathbb{E}_U\left\{\left[\sum_{b=1}^{d_{A_1}^2}\sum_{\ell=0}^{t}(-1)^\ell\binom{t}{\ell}\mathrm{Pr}\Big(\hat{r}=-1,\hat{b}=b\Big|U\Big)^\ell\,\mathrm{Pr}\Big(\hat{r}=+1,\hat{b}=b\Big|U\Big)^{t-\ell}\right]^2\right\} \\
&= \mathbb{E}_U\left\{\left(\sum_{b=1}^{d_{A_1}^2}\Big[\mathrm{Pr}\Big(\hat{r}=+1,\hat{b}=b\Big|U\Big)-\mathrm{Pr}\Big(\hat{r}=-1,\hat{b}=b\Big|U\Big)\Big]^t\right)^2\right\} \\
&= \mathbb{E}_U\left\{\left(\sum_{b=1}^{d_{A_1}^2}\mathrm{Tr}\Big[\tilde{\rho}_U(|\Phi_b\rangle\langle\Phi_b|)_{A_1,A_3}\otimes\mathrm{SWAP}_{A_2,B}\Big]^t\right)^2\right\} \\
&= \sum_{b_1,b_2=1}^{d_{A_1}^2}\mathbb{E}_U\Big\{\mathrm{Tr}\Big[\tilde{\rho}_U(|\Phi_{b_1}\rangle\langle\Phi_b|)_{A_1,A_3}\otimes\mathrm{SWAP}_{A_2,B}\Big]^t\,\mathrm{Tr}\Big[\tilde{\rho}_U(|\Phi_{b_2}\rangle\langle\Phi_b|)_{A_1,A_3}\otimes\mathrm{SWAP}_{A_2,B}\Big]^t\Big\} \\
&= \sum_{b_1,b_2=1}^{d_{A_1}^2} d_{A_1}^{-2t}\,\mathrm{Tr}\Big(\rho^{\otimes 2t}\tilde{\mathcal{Q}}_{b_1,b_2,t}\Big)
\end{aligned} \tag{53}$$

Here, $\tilde{\mathcal{Q}}_{b_1,b_2,t}$ in the last line is defined as

$$\begin{aligned}
\tilde{\mathcal{Q}}_{b_1,b_2,t} := \mathbb{E}_U\bigg\{&\Big[\Big(U^\dagger_{A_1,A_2}\otimes I_{A_3,B}\Big)\Big(d_{A_1}(|\Phi_{b_1}\rangle\langle\Phi_{b_1}|)_{A_1,A_3}\otimes\mathrm{SWAP}_{A_2,B}\Big)\Big(U_{A_1,A_2}\otimes I_{A_3,B}\Big)\Big]^{\otimes t} \\
&\otimes\Big[\Big(U^\dagger_{A_1,A_2}\otimes I_{A_3,B}\Big)\Big(d_{A_1}(|\Phi_{b_2}\rangle\langle\Phi_{b_2}|)_{A_1,A_3}\otimes\mathrm{SWAP}_{A_2,B}\Big)\Big(U_{A_1,A_2}\otimes I_{A_3,B}\Big)\Big]^{\otimes t}\bigg\}.
\end{aligned} \tag{54}$$

According to Lemma 3, when $b_1 = b_2$, we have

$$\begin{aligned}\tilde{\mathcal{Q}}_{b_1,b_2,t} = &\sum_{\pi,\tau\in S_{2t}} \mathrm{Wg}(\pi^{-1}\tau, d_{A_1}d_{A_2})(R_\pi)_{A_1}\otimes(R_\pi)_{A_2}\otimes\Big[\big(W_{b_1}^{\otimes t}\otimes W_{b_2}^{\otimes t}\big)_{A_3}(R_\tau)_{A_3}\big(W_{b_1}^{\otimes t}\otimes W_{b_2}^{\otimes t}\big)^\dagger_{A_3}\Big]\otimes(R_\tau^\top)_B.\\ &= (d_{A_1}d_{A_2})^{-2t}\Bigg[\sum_{\pi\in S_{2t}}\big(W_{b_1}^{\otimes t}\otimes W_{b_2}^{\otimes t}\big)_{A_3}(R_\pi)_A\big(W_{b_1}^{\otimes t}\otimes W_{b_2}^{\otimes t}\big)^\dagger_{A_3}\otimes(R_\pi^\top)_B\\ &\quad+\mathcal{O}\bigg(\frac{1}{d_{A_1}d_{A_2}}\bigg)\sum_{\pi,\tau\in S_{2t}}(R_\pi)_{A_1}\otimes(R_\pi)_{A_2}\otimes\Big[\big(W_{b_1}^{\otimes t}\otimes W_{b_2}^{\otimes t}\big)_{A_3}(R_\tau)_{A_3}\big(W_{b_1}^{\otimes t}\otimes W_{b_2}^{\otimes t}\big)^\dagger_{A_3}\Big]\otimes(R_\tau^\top)_B\Bigg].\end{aligned} \tag{55}$$

It follows that

$$\begin{aligned}&\mathbb{E}_U\mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\Big[f\Big(\hat{\mathbf{b}},\hat{\mathbf{r}},\{i_1,\ldots,i_t,j_1,\ldots,j_t\}\Big)\Big|U\Big]\\ &=\sum_{b_1,b_2=1}^{d_{A_1}^2} d_{A_1}^{-2t}\,\mathrm{Tr}\Big(\rho^{\otimes 2t}\tilde{\mathcal{Q}}_{b_1,b_2,t}\Big) = \sum_{b_1,b_2=1}^{d_{A_1}^2} d_{A_1}^{-2t}\mathcal{O}\Big((d_{A_1}d_{A_2})^{-2t}\Big) = \mathcal{O}\big(d_{A_1}^{4-4t}d_{A_2}^{-2t}\big).\end{aligned} \tag{56}$$

We then group the estimators and reduce the summation in Eq. (46) to

$$\begin{aligned}&\mathbb{E}_U\mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\bigg[\Big(\hat{M}_U^{\mathrm{neg}}\Big)^2\bigg|U\bigg]\\ &=\binom{N_M}{t}^{-2} d_{A_1}^{4t-4}d_{A_2}^{2t}\sum_{k=0}^{t}\sum_{\mathrm{Co}[(i_1,\ldots,i_t);(j_1,\ldots,j_t)]=t-k}\mathbb{E}_U\mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\Big[f\Big(\hat{\mathbf{b}},\hat{\mathbf{r}},\{i_1,\ldots,i_t,j_1,\ldots,j_t\}\Big)\Big|U\Big]\\ &=\binom{N_M}{t}^{-2} d_{A_1}^{4t-4}d_{A_2}^{2t}\sum_{k=0}^{t}\binom{N_M}{t+k}\binom{t+k}{t}\binom{t}{k}\mathbb{E}_U\mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\Big[f\Big(\hat{\mathbf{b}},\hat{\mathbf{r}},\{i_1,\ldots,i_t,j_1,\ldots,j_t\}\Big)\Big|U\Big]\\ &=\binom{N_M}{t}^{-2} d_{A_1}^{4t-4}d_{A_2}^{2t}\Bigg[\sum_{k=0}^{t-1}\binom{N_M}{t+k}\binom{t+k}{t}\binom{t}{k}\mathcal{O}\big(d_{A_1}^{2-2k-2t}d_{A_2}^{-2k}\big)+\binom{N_M}{2t}\binom{2t}{t}\mathcal{O}\big(d_{A_1}^{4-4t}d_{A_2}^{-2t}\big)\Bigg]\\ &=\sum_{k=0}^{t-1}\binom{N_M}{t}^{-2}\binom{N_M}{t+k}\mathcal{O}\big(d_{A_1}^{2t-2k-2}d_{A_2}^{2t-2k}\big)+\binom{N_M}{t}^{-2}\binom{N_M}{2t}\mathcal{O}(1)\\ &=\sum_{i=0}^{t}\mathcal{O}\Bigg(\frac{d_{A_1}^{2i-2}d_{A_2}^{2i}}{N_M^i}\Bigg)+\mathcal{O}(1).\end{aligned} \tag{57}$$

Therefore,

$$\mathrm{Var}\Big[\hat{M}^{\mathrm{neg}}\Big] = \frac{1}{N_U}\mathrm{Var}\Big[\hat{M}_U^{\mathrm{neg}}\Big] \le \frac{1}{N_U}\mathbb{E}_U\mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\bigg[\Big(\hat{M}_U^{\mathrm{neg}}\Big)^2\bigg|U\bigg] = \frac{1}{N_U}\sum_{i=0}^{t}\mathcal{O}\Bigg(\frac{d_{A_1}^{2i-2}d_{A_2}^{2i}}{N_M^i}\Bigg)+\mathcal{O}(1). \tag{58}$$

To ensure that the additive error of $\hat{M}^{\mathrm{neg}}$ is $\mathcal{O}(\epsilon)$, we need to bound this variance by $\mathcal{O}(\epsilon^2)$, which can be satisfied by choosing $N_U = \mathcal{O}(\epsilon^{-2})$ and $N_M = \mathcal{O}\Big(d_{A_1}^{2-2/t}d_{A_2}^2\Big)$. The sample complexity is $N_M N_U = \mathcal{O}\Big(d_{A_1}^{2-2/t}d_{A_2}^2\epsilon^{-2}\Big) \le \mathcal{O}\big(d\epsilon^{-2}\big)$.

V. ESTIMATION OF NEGATIVITY-MOMENTS

Consider that we are given multiple copies of an unknown quantum state ρ on the Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$. Let n_A and n_B be the qubit number of subsystems A and B, respectively. W.l.o.g., we assume that $d_A \geq d_B$. Our goal is to estimate the t-th order negativity-moment $\mathrm{Tr}\left[\left(\rho^{\top_B}\right)^t\right]$, where t is a constant integer.

Theorem 5. *Suppose* $0 < \epsilon < 1$, $d_A \geq \Omega(\epsilon^{-2})$, *and* $t = \mathcal{O}(1)$ *is a positive integer. There exists an algorithm that can estimate* $\mathrm{Tr}\left[\left(\rho^{\top_B}\right)^k\right]$ *within* ϵ *additive error for all* $k = 1, 2, \ldots, t$, *by using single-copy operations on* ρ. *The sample complexity and number of different random unitaries used by the algorithm are*

$$N_U N_M = \mathcal{O}\left(\max\left\{\frac{d}{(d_{A_1}\epsilon^2)^{1/t}}, \frac{d}{d_{A_1}\epsilon^2}\right\}\right) \quad \textit{and} \quad N_U = \max\left\{1, \frac{1}{d_{A_1}\epsilon^2}\right\}, \tag{59}$$

respectively, where $d_{A_1} = d_A/d_B$.

A. Estimation protocol

Notice that if we can estimate $\sum_{\pi \in \mathrm{S}_s} \mathrm{Tr}\left(\left(\rho^{\top_B}\right)^{\otimes s} R_\pi\right)$ within $\mathcal{O}(\epsilon)$ additive error for all $s = 1, 2, \ldots, t$, then we can estimate $\mathrm{Tr}\left[\left(\rho^{\top_B}\right)^t\right]$ within $\mathcal{O}(\epsilon)$ error. To achieve this goal, our protocol runs as follows. First, we divide subsystem A into two components A_1 and A_2, whose qubit numbers satisfy $n_{A_2} = n_B$ and $n_{A_1} + n_{A_2} = n_A$. Then we do the following procedure for $q = 1, 2, \ldots, N_U$:

1. Sample a random unitary $U_q \sim \mathbb{U}(\mathcal{H}_A)$.

2. Prepare N_M copies of ρ, and apply U_q on the subsystem A of ρ.

3. For each of the rotated state $\tilde{\rho}_{U_q} = (U_q \otimes I_B)\rho(U_q^\dagger \otimes I_B)$, we

   (1) measure subsystems A_1 in the computational basis $\{|b\rangle\}_{b=1}^{d_{A_1}}$, and obtain outcomes $\hat{\mathbf{b}} = \left\{\hat{b}_1, \ldots, \hat{b}_{N_M}\right\}$.

   (2) measure subsystems A_2 and B with the POVM $\left\{\Pi_{\mathrm{sym}}^{(2)}, \Pi_{\mathrm{asym}}^{(2)}\right\}$ (equivalently, one can perform the swap test on A_2 and B), and obtain outcomes $\hat{\mathbf{r}} = \{\hat{r}_1, \ldots, \hat{r}_{N_M}\}$, where $\hat{r}_j \in \{+1, -1\}$.

4. Compute the following estimator using $\hat{\mathbf{b}}$ and $\hat{\mathbf{r}}$:

$$\hat{M}_U^{\mathrm{neg}} := \binom{N_M}{t}^{-1} d_A^t d_{A_1}^{-1} \sum_{i_1 < i_2 < \cdots < i_t} (\hat{r}_{i_1}\hat{r}_{i_2}\cdots\hat{r}_{i_t})\mathbf{1}\left[\hat{b}_{i_1} = \hat{b}_{i_2} = \cdots = \hat{b}_{i_t}\right]. \tag{60}$$

Finally, output $\hat{M}^{\mathrm{neg}} := N_U^{-1}\sum_q \hat{M}_{U_q}^{\mathrm{neg}}$ as our estimate of $\sum_{\pi \in \mathrm{S}_t} \mathrm{Tr}\left(\left(\rho^{\top_B}\right)^{\otimes t} R_\pi\right)$.

Next, we shall explain why $\hat{M}^{\text{neg}}$ is a valid estimator. Its expectation value is

$$\begin{aligned}
\mathbb{E}\Big[\hat{M}^{\text{neg}}\Big] &= \mathbb{E}_U\mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\Big(\hat{M}_U^{\text{neg}}\Big|U\Big) = d_A^t\, d_{A_1}^{-1}\,\mathbb{E}_U\mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\Big\{(\hat{r}_{i_1}\hat{r}_{i_2}\cdots\hat{r}_{i_t})\mathbf{1}\big[\hat{b}_{i_1}=\cdots=\hat{b}_{i_t}\big]\Big|U\Big\}\\
&= d_A^t\, d_{A_1}^{-1}\,\mathbb{E}_U\left\{\sum_{\ell=0}^{t}\sum_{b=1}^{d_{A_1}}(-1)^\ell\Pr\Big(\#[-1 \text{ among } \{\hat{r}_{i_1},\hat{r}_{i_2},\ldots,\hat{r}_{i_t}\}]=\ell, \text{ and } \hat{b}_{i_1}=\cdots=\hat{b}_{i_t}=b\Big|U\Big)\right\}\\
&= d_A^t\, d_{A_1}^{-1}\,\mathbb{E}_U\left\{\sum_{\ell=0}^{t}\sum_{b=1}^{d_{A_1}}\binom{t}{\ell}(-1)^\ell\Pr\Big(\hat{r}=-1,\hat{b}=b\Big|U\Big)^\ell\Pr\Big(\hat{r}=+1,\hat{b}=b\Big|U\Big)^{t-\ell}\right\}\\
&= d_A^t\, d_{A_1}^{-1}\,\mathbb{E}_U\left\{\sum_{b=1}^{d_{A_1}}\Big[\Pr\Big(\hat{r}=+1,\hat{b}=b\Big|U\Big)-\Pr\Big(\hat{r}=-1,\hat{b}=b\Big|U\Big)\Big]^t\right\}\\
&= d_A^t\, d_{A_1}^{-1}\sum_{b=1}^{d_{A_1}}\mathbb{E}_U\left\{\bigg(\operatorname{Tr}\Big[\tilde{\rho}_U\Big(|b\rangle\langle b|_{A_1}\otimes(\Pi^{(2)}_{\text{sym}})_{A_2,B}\Big)\Big]-\operatorname{Tr}\Big[\tilde{\rho}_U\Big(|b\rangle\langle b|_{A_1}\otimes(\Pi^{(2)}_{\text{asym}})_{A_2,B}\Big)\Big]\bigg)^t\right\}\\
&= d_A^t\, d_{A_1}^{-1}\sum_{b=1}^{d_{A_1}}\mathbb{E}_U\Big\{\operatorname{Tr}\big[\tilde{\rho}_U\big(|b\rangle\langle b|_{A_1}\otimes\text{SWAP}_{A_2,B}\big)\big]^t\Big\} = d_A^t\, d_{A_1}^{-1}\sum_{b=1}^{d_{A_1}}\operatorname{Tr}\Big(\rho^{\otimes t}\tilde{H}_{b,t}\Big)
\end{aligned} \tag{61}$$

Here, $\tilde{H}_t$ in the last line is defined as

$$\tilde{H}_{b,t} := \mathbb{E}_U\left\{\Big[\Big(U_A^\dagger\otimes I_B\Big)\Big(|b\rangle\langle b|_{A_1}\otimes\text{SWAP}_{A_2,B}\Big)\Big(U_A\otimes I_B\Big)\Big]^{\otimes t}\right\}. \tag{62}$$

According to Lemma 3 below, we have

$$\begin{aligned}
\tilde{H}_{b,t} &= \sum_{\pi,\tau\in S_t}\text{Wg}(\pi^{-1}\tau,d_A)(R_\pi)_A\otimes(R_\tau^\top)_B\\
&= d_A^{-t}\left[\sum_{\pi\in S_t}(R_\pi)_A\otimes(R_\pi^\top)_B+\mathcal{O}\big(d_A^{-1}\big)\sum_{\pi,\tau\in S_t}(R_\tau)_A\otimes(R_\tau^\top)_B\right],
\end{aligned} \tag{63}$$

where the second equality follows from the relation $\text{Wg}(\pi^{-1}\tau,d)=d^{-t}\big(\delta_{\pi,\tau}+\mathcal{O}(d^{-1})\big)$.

By combining Eqs. (61) and (63), the expectation of $\hat{M}^{\text{neg}}$ can be further expressed as

$$\begin{aligned}
\mathbb{E}\Big[\hat{M}^{\text{neg}}\Big] &= \sum_{\pi\in S_t}\operatorname{Tr}\big[\rho_{AB}^{\otimes t}(R_\pi)_A\otimes(R_\pi^\top)_B\big]+\mathcal{O}\big(d_A^{-1}\big)\sum_{\pi,\tau\in S_t}\operatorname{Tr}\big[\rho_{AB}^{\otimes t}(R_\pi)_A\otimes(R_\tau^\top)_B\big]\\
&= \sum_{\pi\in S_t}\operatorname{Tr}\bigg[\Big(\rho_{AB}^{\top_B}\Big)^{\otimes t}(R_\pi)_A\otimes(R_\pi)_B\bigg]+\mathcal{O}\big(d_A^{-1}\big)\\
&= \sum_{\pi\in \mathrm{S}_t}\operatorname{Tr}\Big(\big(\rho^{\top_B}\big)^{\otimes t}R_\pi\Big)+\mathcal{O}\big(d_A^{-1}\big).
\end{aligned} \tag{64}$$

Therefore, in the usual case with $d_A^{-1}\le\mathcal{O}(\epsilon)$, $\hat{M}^{\text{neg}}$ is a valid estimator for $\sum_{\pi\in\mathrm{S}_t}\operatorname{Tr}\Big(\big(\rho^{\top_B}\big)^{\otimes t}R_\pi\Big)$. If we can estimate $\hat{M}^{\text{neg}}$ within error $\mathcal{O}(\epsilon)$, we can then estimate $\sum_{\pi\in\mathrm{S}_t}\operatorname{Tr}\Big(\big(\rho^{\top_B}\big)^{\otimes t}R_\pi\Big)$ within error $\mathcal{O}(\epsilon)$.

B. Variance analyses

The variance of $\hat{M}^{\text{neg}}$ is $\text{Var}[\hat{M}^{\text{neg}}]=N_U^{-1}\,\text{Var}[\hat{M}_U^{\text{neg}}]$, where

$$\text{Var}\Big[\hat{M}_U^{\text{neg}}\Big]=\mathbb{E}\bigg[\Big(\hat{M}_U^{\text{neg}}\Big)^2\bigg]-\mathbb{E}\Big[\hat{M}_U^{\text{neg}}\Big]^2=\mathbb{E}_U\mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\bigg[\Big(\hat{M}_U^{\text{neg}}\Big)^2\bigg|U\bigg]-\bigg[\sum_{\pi\in\mathrm{S}_t}\operatorname{Tr}\Big(\big(\rho^{\top_B}\big)^{\otimes t}R_\pi\Big)+\mathcal{O}\big(d_A^{-1}\big)\bigg]^2. \tag{65}$$

Note that

$$\mathbb{E}_U\mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\left[\left(\hat{M}_U^{\text{neg}}\right)^2\middle|U\right]=\binom{N_M}{t}^{-2}d_A^{2t}\,d_{A_1}^{-2}\sum_{\substack{i_1<i_2<\cdots<i_t\\ j_1<j_2<\cdots<j_t}}\mathbb{E}_U\mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\left[f\left(\hat{\mathbf{b}},\hat{\mathbf{r}},\{i_1,\ldots,i_t,j_1,\ldots,j_t\}\right)\middle|U\right]. \tag{66}$$

where

$$f\left(\hat{\mathbf{b}},\hat{\mathbf{r}},\{i_1,\ldots,i_t,j_1,\ldots,j_t\}\right):=(\hat{r}_{i_1}\hat{r}_{i_2}\cdots\hat{r}_{i_t}\hat{r}_{j_1}\hat{r}_{j_2}\cdots\hat{r}_{j_t})\mathbf{1}\left[\hat{b}_{i_1}=\cdots=\hat{b}_{i_t}\right]\mathbf{1}\left[\hat{b}_{j_1}=\cdots=\hat{b}_{j_t}\right]. \tag{67}$$

Based on the collision number (denoted as $\mathrm{Co}[(i_1,i_2,\ldots,i_t);(j_1,j_2,\ldots,j_t)]$), the expectation value in the last bracket will be different. When $\mathrm{Co}[(i_1,\ldots,i_t);(j_1,\ldots,j_t)]=t-k$, where $k\in\{0,1,\ldots,t-1\}$, we have

$$\begin{aligned}&\mathbb{E}_U\mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\left[f\left(\hat{\mathbf{b}},\hat{\mathbf{r}},\{i_1,\ldots,i_t,j_1,\ldots,j_t\}\right)\middle|U\right]\\&\quad=\mathbb{E}_U\mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\left((\hat{r}_{i_1}\hat{r}_{i_2}\cdots\hat{r}_{i_t}\hat{r}_{j_1}\hat{r}_{j_2}\cdots\hat{r}_{j_t})\mathbf{1}\left[\hat{b}_{i_1}=\cdots=\hat{b}_{i_t}=\hat{b}_{j_1}=\cdots=\hat{b}_{j_t}\right]\middle|U\right).\end{aligned} \tag{68}$$

Wlog, we assume that $(i_1,i_2,\ldots,i_t)$ and $(j_1,j_2,\ldots,j_t)$ coincide on the last $t-k$ pairs. Then the above quantity can be rewritten as

$$\begin{aligned}&\mathbb{E}_U\mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\left\{(\hat{r}_{i_1}\hat{r}_{i_2}\cdots\hat{r}_{i_k}\hat{r}_{j_1}\hat{r}_{j_2}\cdots\hat{r}_{j_k})\mathbf{1}\left[\hat{b}_{i_1}=\cdots=\hat{b}_{i_t}=\hat{b}_{j_1}=\cdots=\hat{b}_{j_k}\right]\middle|U\right\}\\
&=\mathbb{E}_U\left\{\sum_{\ell=0}^{2k}\sum_{b=1}^{d_{A_1}}(-1)^\ell\Pr\left(\#[-1\text{ among }\{\hat{r}_{i_1},\ldots,\hat{r}_{i_k},\hat{r}_{j_1},\ldots,\hat{r}_{j_k}\}]=\ell,\text{ and }\hat{b}_{i_1}=\cdots=\hat{b}_{i_t}=\hat{b}_{j_1}=\cdots=\hat{b}_{j_k}=b\middle|U\right)\right\}\\
&=\mathbb{E}_U\left\{\sum_{\ell=0}^{2k}\sum_{b=1}^{d_{A_1}}\binom{2k}{\ell}(-1)^\ell\Pr\left(\hat{r}=-1,\hat{b}=b\middle|U\right)^\ell\Pr\left(\hat{r}=+1,\hat{b}=b\middle|U\right)^{2k-\ell}\Pr\left(\hat{b}=b\middle|U\right)^{t-k}\right\}\\
&=\mathbb{E}_U\left\{\sum_{b=1}^{d_{A_1}}\Pr\left(\hat{b}=b\middle|U\right)^{t-k}\sum_{\ell=0}^{2k}\binom{2k}{\ell}(-1)^\ell\Pr\left(\hat{r}=-1,\hat{b}=b\middle|U\right)^\ell\Pr\left(\hat{r}=+1,\hat{b}=b\middle|U\right)^{2k-\ell}\right\}\\
&=\mathbb{E}_U\left\{\sum_{b=1}^{d_{A_1}}\Pr\left(\hat{b}=b\middle|U\right)^{t-k}\left[\Pr\left(\hat{r}=+1,\hat{b}=b\middle|U\right)-\Pr\left(\hat{r}=-1,\hat{b}=b\middle|U\right)\right]^{2k}\right\}\\
&=\sum_{b=1}^{d_{A_1}}\mathbb{E}_U\left\{\mathrm{Tr}\left[\tilde{\rho}_U\left(|b\rangle\langle b|_{A_1}\otimes I_{A_2,B}\right)\right]^{t-k}\mathrm{Tr}\left[\tilde{\rho}_U\left(|b\rangle\langle b|_{A_1}\otimes\mathrm{SWAP}_{A_2,B}\right)\right]^{2k}\right\}=\sum_{b=1}^{d_{A_1}}\mathrm{Tr}\left(\rho^{\otimes(t+k)}\mathcal{Q}_{b,t,k}\right)\end{aligned} \tag{69}$$

Here, $\mathcal{Q}_{b,t,k}$ in the last line is defined as

$$\begin{aligned}\mathcal{Q}_{b,t,k}:=\mathbb{E}_U\Bigg\{&\left[\left(U_A^\dagger\otimes I_B\right)\left(|b\rangle\langle b|_{A_1}\otimes I_{A_2,B}\right)\left(U_A\otimes I_B\right)\right]^{\otimes(t-k)}\\&\otimes\left[\left(U_A^\dagger\otimes I_B\right)\left(|b\rangle\langle b|_{A_1}\otimes\mathrm{SWAP}_{A_2,B}\right)\left(U_A\otimes I_B\right)\right]^{\otimes 2k}\Bigg\}.\end{aligned} \tag{70}$$

According to Lemma 3, we have

$$\begin{aligned}\mathcal{Q}_{b,t,k}&=\sum_{\pi,\tau\in S_{t+k}}\mathrm{Wg}(\pi^{-1}\tau,d_A)(R_\pi)_A\otimes\left\{I_B^{\otimes(t-k)}\otimes\mathrm{Tr}_{1,2,\ldots,t-k}\left[\left(R_\tau^\top\right)_B\right]\right\}\\&=d_A^{-(t+k)}d_B^{t-k}\sum_{\pi,\tau\in S_{t+k}}\mathcal{O}(1)(R_\pi)_A\otimes(R_{f(\tau)})_B\\&=\mathcal{O}\left(d_A^{-(t+k)}d_B^{t-k}\right)\sum_{\pi,\tau\in S_{t+k}}(R_\pi)_A\otimes(R_{f(\tau)})_B.\end{aligned} \tag{71}$$

It follows that

$$\mathbb{E}_U\mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\Big[f\Big(\hat{\mathbf{b}},\hat{\mathbf{r}},\{i_1,\ldots,i_t,j_1,\ldots,j_t\}\Big)\Big|U\Big]=\sum_{b=1}^{d_{A_1}}\mathrm{Tr}\Big(\rho^{\otimes(t+k)}\mathcal{Q}_{b,t,k}\Big)=\mathcal{O}\Big(d_{A_1}d_A^{-(t+k)}d_B^{t-k}\Big). \tag{72}$$

When $\mathrm{Co}[(i_1,\ldots,i_t);(j_1,\ldots,j_t)]=0$, we have

$$\begin{aligned}
&\mathbb{E}_U\mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\Big[f\Big(\hat{\mathbf{b}},\hat{\mathbf{r}},\{i_1,\ldots,i_t,j_1,\ldots,j_t\}\Big)\Big|U\Big]\\
&=\mathbb{E}_U\left\{\Big[\mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\Big(\hat{r}_{i_1}\hat{r}_{i_2}\cdots\hat{r}_{i_t}\mathbf{1}\big[\hat{b}_{i_1}=\cdots=\hat{b}_{i_t}\big]\Big|U\Big)\Big]^2\right\}\\
&=\mathbb{E}_U\left\{\left[\sum_{\ell=0}^{t}\sum_{b=1}^{d_{A_1}}(-1)^\ell\Pr\left(\#[-1\text{ among }\{\hat{r}_{i_1},\ldots,\hat{r}_{i_t}\}]=\ell,\text{ and }\hat{b}_{i_1}=\cdots=\hat{b}_{i_t}=b\Big|U\right)\right]^2\right\}\\
&=\mathbb{E}_U\left\{\left[\sum_{b=1}^{d_{A_1}}\sum_{\ell=0}^{t}(-1)^\ell\binom{t}{\ell}\Pr\Big(\hat{r}=-1,\hat{b}=b\Big|U\Big)^\ell\Pr\Big(\hat{r}=+1,\hat{b}=b\Big|U\Big)^{t-\ell}\right]^2\right\}\\
&=\mathbb{E}_U\left\{\left(\sum_{b=1}^{d_{A_1}}\Big[\Pr\Big(\hat{r}=+1,\hat{b}=b\Big|U\Big)-\Pr\Big(\hat{r}=-1,\hat{b}=b\Big|U\Big)\Big]^t\right)^2\right\}\\
&=\mathbb{E}_U\left\{\left(\sum_{b=1}^{d_{A_1}}\mathrm{Tr}\Big[\tilde{\rho}_U\big(|b\rangle\langle b|_{A_1}\otimes\mathrm{SWAP}_{A_2,B}\big)\Big]^t\right)^2\right\}\\
&=\sum_{b_1,b_2=1}^{d_{A_1}}\mathbb{E}_U\left\{\mathrm{Tr}\Big[\tilde{\rho}_U\big(|b_1\rangle\langle b_1|_{A_1}\otimes\mathrm{SWAP}_{A_2,B}\big)\Big]^t\,\mathrm{Tr}\Big[\tilde{\rho}_U\big(|b_2\rangle\langle b_2|_{A_1}\otimes\mathrm{SWAP}_{A_2,B}\big)\Big]^t\right\}\\
&=\sum_{b_1,b_2=1}^{d_{A_1}}\mathrm{Tr}\Big(\rho^{\otimes 2t}\tilde{\mathcal{Q}}_{b_1,b_2,t}\Big)
\end{aligned} \tag{73}$$

Here, $\tilde{\mathcal{Q}}_{b_1,b_2,t}$ in the last line is defined as

$$\begin{aligned}
\tilde{\mathcal{Q}}_{b_1,b_2,t}:=\mathbb{E}_U\Big\{&\Big[\Big(U_A^\dagger\otimes I_B\Big)\Big(|b_1\rangle\langle b_1|_{A_1}\otimes\mathrm{SWAP}_{A_2,B}\Big)\Big(U_A\otimes I_B\Big)\Big]^{\otimes t}\\
&\otimes\Big[\Big(U_A^\dagger\otimes I_B\Big)\Big(|b_2\rangle\langle b_2|_{A_1}\otimes\mathrm{SWAP}_{A_2,B}\Big)\Big(U_A\otimes I_B\Big)\Big]^{\otimes t}\Big\}.
\end{aligned} \tag{74}$$

When $b_1=b_2$, we have

$$\begin{aligned}
\tilde{\mathcal{Q}}_{b_1,b_2,t}=\tilde{H}_{b,2t}&=\sum_{\pi,\tau\in S_{2t}}\mathrm{Wg}(\pi^{-1}\tau,d_A)(R_\pi)_A\otimes(R_\tau^\top)_B\\
&=d_A^{-2t}\left[\sum_{\pi\in S_{2t}}(R_\pi)_A\otimes(R_\pi^\top)_B+\mathcal{O}\big(d_A^{-1}\big)\sum_{\pi,\tau\in S_{2t}}(R_\tau)_A\otimes(R_\tau^\top)_B\right];
\end{aligned} \tag{75}$$

when $b_1\neq b_2$, we have

$$\begin{aligned}
\tilde{\mathcal{Q}}_{b_1,b_2,t}&=\sum_{\pi,\tau\in S_{2t}}\mathrm{Wg}(\pi^{-1}\tau,d_A)\,(R_\pi)_A\otimes(R_\tau^\top)_B\,\langle\!\langle b_1^{\otimes t}\otimes b_2^{\otimes t}|R_\tau^\top\rangle\!\rangle\\
&=\sum_{\pi\in S_{2t}}\sum_{\tau_1,\tau_2\in S_t}\mathrm{Wg}(\pi^{-1}\tau_1\tau_2,d_A)\,(R_\pi)_A\otimes\big(R_{\tau_1}^\top\big)_B\otimes\big(R_{\tau_2}^\top\big)_B\\
&=d_A^{-2t}\left[\sum_{\tau_1,\tau_2\in S_t}(R_{\tau_1})_A\otimes(R_{\tau_2})_A\otimes\big(R_{\tau_1}^\top\big)_B\otimes\big(R_{\tau_2}^\top\big)_B+\mathcal{O}\big(d_A^{-1}\big)\sum_{\pi,\tau\in S_{2t}}(R_\tau)_A\otimes(R_\tau^\top)_B\right],
\end{aligned} \tag{76}$$

It follows that

$$
\begin{aligned}
&\mathbb{E}_U \mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\left[f\left(\hat{\mathbf{b}},\hat{\mathbf{r}},\{i_1,\ldots,i_t,j_1,\ldots,j_t\}\right)\middle|U\right] \\
&= \sum_{b=1}^{d_{A_1}} \operatorname{Tr}\left(\rho^{\otimes 2t}\tilde{\mathcal{Q}}_{b,b,t}\right) + \sum_{b_1\neq b_2} \operatorname{Tr}\left(\rho^{\otimes 2t}\tilde{\mathcal{Q}}_{b_1,b_2,t}\right) \\
&= d_{A_1} d_A^{-2t}\mathcal{O}(1) + d_{A_1}(d_{A_1}-1)d_A^{-2t}\left\{\operatorname{Tr}\left[\rho^{\otimes t}\sum_{\tau\in S_t}(R_\tau)_A\otimes(R_\tau^\top)_B\right]^2 + \mathcal{O}\big(d_A^{-1}\big)\right\} \\
&= \mathcal{O}\big(d_{A_1}d_A^{-2t}\big) + d_{A_1}^2 d_A^{-2t}\operatorname{Tr}\left[\rho^{\otimes t}\sum_{\tau\in S_t}(R_\tau)_A\otimes(R_\tau^\top)_B\right]^2 \\
&= \mathcal{O}\big(d_{A_1}d_A^{-2t}\big) + d_{A_1}^2 d_A^{-2t}\sum_{\pi\in \mathrm{S}_t}\operatorname{Tr}\left(\left(\rho^{\top_B}\right)^{\otimes t}R_\pi\right)
\end{aligned}
\tag{77}
$$

We then group the estimators and reduce the summation in Eq. (66) to

$$
\begin{aligned}
&\mathbb{E}_U \mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\left[\left(\hat{M}_U^{\text{neg}}\right)^2\middle|U\right] \\
&= \binom{N_M}{t}^{-2} d_A^{2t}\, d_{A_1}^{-2}\sum_{k=0}^{t}\sum_{\mathrm{Co}[(i_1,\ldots,i_t);(j_1,\ldots,j_t)]=t-k}\mathbb{E}_U \mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\left[f\left(\hat{\mathbf{b}},\hat{\mathbf{r}},\{i_1,\ldots,i_t,j_1,\ldots,j_t\}\right)\middle|U\right] \\
&= \binom{N_M}{t}^{-2} d_A^{2t}\, d_{A_1}^{-2}\sum_{k=0}^{t}\binom{N_M}{t+k}\binom{t+k}{t}\binom{t}{k}\mathbb{E}_U \mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\left[f\left(\hat{\mathbf{b}},\hat{\mathbf{r}},\{i_1,\ldots,i_t,j_1,\ldots,j_t\}\right)\middle|U\right] \\
&= \binom{N_M}{t}^{-2} d_A^{2t}\, d_{A_1}^{-2}\left\{\sum_{k=0}^{t-1}\binom{N_M}{t+k}\binom{t+k}{t}\binom{t}{k}\mathcal{O}\left(d_{A_1}d_A^{-(t+k)}d_B^{t-k}\right)\right. \\
&\qquad\qquad \left. + \binom{N_M}{2t}\binom{2t}{t}\left[\mathcal{O}\big(d_{A_1}d_A^{-2t}\big) + d_{A_1}^2 d_A^{-2t}\sum_{\pi\in\mathrm{S}_t}\operatorname{Tr}\left(\left(\rho^{\top_B}\right)^{\otimes t}R_\pi\right)\right]\right\} \\
&= \sum_{k=0}^{t-1}\binom{N_M}{t}^{-2}\binom{N_M}{t+k}\mathcal{O}\big(d_{A_1}^{-1}d^{t-k}\big) + \binom{N_M}{t}^{-2}\binom{N_M}{2t}\binom{2t}{t}\left[\mathcal{O}\big(d_{A_1}^{-1}\big) + \sum_{\pi\in\mathrm{S}_t}\operatorname{Tr}\left(\left(\rho^{\top_B}\right)^{\otimes t}R_\pi\right)\right] \\
&\le \sum_{k=0}^{t-1}\mathcal{O}\left(\frac{d_{A_1}^{-1}d^{t-k}}{N_M^{t-k}}\right) + \sum_{\pi\in\mathrm{S}_t}\operatorname{Tr}\left(\left(\rho^{\top_B}\right)^{\otimes t}R_\pi\right) + \mathcal{O}\big(d_{A_1}^{-1}\big).
\end{aligned}
\tag{78}
$$

Therefore,

$$
\begin{aligned}
\operatorname{Var}\left[\hat{M}^{\text{neg}}\right] &= \frac{1}{N_U}\operatorname{Var}\left[\hat{M}_U^{\text{neg}}\right] = \frac{1}{N_U}\left\{\mathbb{E}_U \mathbb{E}_{\hat{\mathbf{b}},\hat{\mathbf{r}}}\left[\left(\hat{M}_U^{\text{neg}}\right)^2\middle|U\right] - \mathbb{E}\left[\hat{M}_U^{\text{neg}}\right]^2\right\} \\
&= \frac{1}{N_U}\left[\sum_{k=0}^{t-1}\mathcal{O}\left(\frac{d_{A_1}^{-1}d^{t-k}}{N_M^{t-k}}\right) + \mathcal{O}\big(d_{A_1}^{-1}\big)\right] = \frac{1}{N_U}\sum_{i=0}^{t}\mathcal{O}\left(\frac{d^i}{d_{A_1}N_M^i}\right)
\end{aligned}
\tag{79}
$$

To ensure that the additive error is $\mathcal{O}(\epsilon)$, we need to bound this variance by $\mathcal{O}\big(\epsilon^2\big)$. Consider two cases below:

1. $d_{A_1} = o(\epsilon^{-2})$. In this case, it suffices to let $N_U = \mathcal{O}\big(d_{A_1}^{-1}\epsilon^{-2}\big)$ and $N_M = \mathcal{O}(d)$. Then $N_M N_U = \mathcal{O}\big(\epsilon^{-2}\big)$.
2. $d_{A_1} = \Omega(\epsilon^{-2})$. In this case, it suffices to let $N_U = 1$ and $N_M = \mathcal{O}\left(d\, d_{A_1}^{-1/t}\epsilon^{-2/t}\right)$. So the sample cost is $\mathcal{O}(N_M)$.

To summarize, we have

$$
N_U = \max\left\{1, \frac{1}{d_{A_1}\epsilon^2}\right\}, \qquad N_M = \min\left\{\frac{d}{d_{A_1}^{1/t}\epsilon^{2/t}}, d\right\}, \qquad N_U N_M = \max\left\{\frac{d}{d_{A_1}^{1/t}\epsilon^{2/t}}, \frac{d}{d_{A_1}\epsilon^2}\right\},
\tag{80}
$$

where we ignored constant coefficients.

Notably, in the usual case $d_{A_1} = \Omega(\epsilon^{-2})$, we only need $N_U = 1$ random unitary to estimate PT moments $\mathrm{Tr}\left[\left(\rho^{\top_B}\right)^k\right]$ within additive error ϵ for all $k = 1, 2, \ldots, t$. This is the best result that one can expect. However, when $d_{A_1} = o(\epsilon^{-2})$, the sample complexity is $\mathcal{O}\big(d\, d_{A_1}^{-1} \epsilon^{-2}\big)$, and one needs more than one random unitaries. To address this shortcoming, we can add $n_C = \ln\big[\Theta\big(d_{A_1}^{-1}\epsilon^{-2}\big)\big]$ ancilla qubits to the initial state ρ, and treat the joint subsystem A_1 and C as the new subsystem A_1'. Then we have $d_{A_1'} = d_{A_1} 2^{n_C} = \Theta(\epsilon^{-2})$. So now we still need only one random unitary (acting on the joint subsystem of A_1 and C), and the new sample complexity reads

$$N_M' = \mathcal{O}\Big(d\, d_C\, d_{A_1'}^{-1/t} \epsilon^{-2/t}\Big) = \mathcal{O}\big(d\, d_{A_1}^{-1} \epsilon^{-2}\big). \tag{81}$$

Therefore, we can reduce the number of random unitaries to $N_U = 1$ without increasing the sample cost.

[1] A. A. Mele, Introduction to Haar measure tools in quantum information: A beginner's tutorial, Quantum **8**, 1340 (2024).