The second step uses the fact that the uniform distribution over the Clifford group has the same first (in fact three) moments as the Haar measure.

Combining this with Theorem 231 yields the following lower bound for Pauli shadow tomography.

Corollary 233 (Shadow tomography lower bound for Pauli observables)**.** *Any learning protocol which only performs single-copy measurements on copies of unknown state ρ and outputs ϵ-accurate estimates for $\{\mathrm{tr}(P\rho)\}$ for all Pauli operators P with high probability requires $\Omega(2^n/\epsilon^2)$ copies of ρ.*

3. Lower bound for non-adaptive two-copy measurements

Note that the learning protocol in Section 1 crucially relies on adaptivity in order to determine the appropriate measurement to perform in the second stage of the protocol where one resolves the signs. In this section, we show that this adaptivity is necessary: any protocol that performs *non-adaptive* two-copy measurements requires an exponential amount of copies of ρ:

Theorem 234. *Any protocol that estimates $\{\mathrm{tr}(P\rho)\}$ for all Pauli operators P to additive error less than 1 using* non-adaptive *two-copy measurements requires $\Omega(2^{n/2})$ copies of ρ.*

As in the previous lower bounds, we assume that the learning protocol uses rank-1 POVMs of the form $\{w_s |\psi_s\rangle\langle\psi_s|\}_s$ for for $2n$-qubit pure states $\{|\psi_s\rangle\}$ and nonnegative weights w_s with $\sum_s w_s = 2^{2n}$.

To prove the nonadaptive lower bound in Theorem 234, we consider a distinguishing task in which we are given access to copies of an unknown n-qubit quantum state ρ and want to distinguish between two cases:

- ρ is sampled from $\{\rho_a^+ = \frac{I+P_a}{2^n}\}_{a\in[4^n-1]}$ uniformly over non-identity Paulis P_a
- ρ is sampled from $\{\rho_a^- = \frac{I-P_a}{2^n}\}_{a\in[4^n-1]}$ uniformly over non-identity Paulis P_a.

If we have a protocol that can estimate all Pauli observables to additive error less than 1 with high probability, then we can solve this distinguishing task using the same protocol. It thus suffices to prove a lower bound on the number of copies of ρ needed to solve the distinguishing problem with high probability. Note that this distinguishing problem is different from the distinguishing problem we considered for single-copy measurements; there we wished to distinguish between maximally mixed versus states of the form $(I \pm P_a)/2^n$, whereas here we wish to distinguish between *two different* ensembles over nontrivial projectors.

Given an $a \in \{1, ..., 4^n - 1\}$, and given the t-th POVM $\mathcal{M}_t = \{w_s^t |\psi_s^t\rangle\langle\psi_s^t|\}$, we denote by $p_{a,t}^+$ and $p_{a,t}^-$ the probability distributions over outcomes s_t from measuring ρ_a^+ and ρ_a^- respectively. We also denote by p_a^+ and p_a^- the probability distribution over the T measurement outcomes $s_1, ..., s_T$ for measuring ρ_a^+ and ρ_a^-.

Our goal is to show a bound on the total variation distance between $\mathbb{E}_a[p_a^+]$ and $\mathbb{E}_a[p_a^-]$ for any nonadaptive sequence of two-copy POVM measurements $\{\mathcal{M}_t\}_{t=1}^T$, i.e.,

$$\mathrm{TV}(\mathbb{E}_a[p_a^+], \mathbb{E}_a[p_a^-]) \leq o(1)\,,$$

as Le Cam's lemma would then imply that any nonadaptive protocols using at most T two-copy measurements cannot distinguish between these two cases with high probability.

Note that for a sequence of nonadaptive measurements, the probability distribution p_a^+ and p_a^- can be written as a tensor product of the probability distributions for each individual measurement $p_{a,t}^+$ and $p_{a,t}^-$. We thus have

$$\begin{aligned}\mathrm{TV}(\mathbb{E}_a[p_a^+], \mathbb{E}_a[p_a^-]) &\leq \mathbb{E}_a[\mathrm{TV}(p_a^+, p_a^-)] \\ &= \mathbb{E}_a\Big[\mathrm{TV}\Big(\bigotimes_{t=1}^T p_{a,t}^+, \bigotimes_{t=1}^T p_{a,t}^-\Big)\Big] \\ &\leq \sum_{t=1}^T \mathbb{E}_a[\mathrm{TV}(p_{a,t}^+, p_{a,t}^-)] \\ &\leq T \max_{\text{two copy } \mathcal{M}_t} \mathbb{E}_a[\mathrm{TV}(p_{a,t}^+, p_{a,t}^-)]\,.\end{aligned}$$

Note that this sequence of inequalities crucially uses the fact that the measurements are non-adaptive.

It thus suffices to bound the average total variation distance for a *single* two-copy measurement. Define

$$\mathbf{N} \triangleq \left(2^n(\mathrm{SWAP}_{1,3} + \mathrm{SWAP}_{1,4} + \mathrm{SWAP}_{2,3} + \mathrm{SWAP}_{2,4}) - 4I^{\otimes 4}\right)\,.$$

For any such $\mathcal{M}_t$, we have

$$\begin{aligned}&\mathbb{E}_a[\mathrm{TV}(p_{a,t}^+, p_{a,t}^-)] \\ &\quad= \mathbb{E}_a\Big[\frac{1}{2}\sum_s \Big|\mathrm{tr}\Big[\Big(\frac{I+P_a}{2^n}\Big)^{\otimes 2} w_s^t \left|\psi_s^t\right\rangle\left\langle\psi_s^t\right|\Big] - \mathrm{tr}\Big[\Big(\frac{I-P_a}{2^n}\Big)^{\otimes 2} w_s^t \left|\psi_s^t\right\rangle\left\langle\psi_s^t\right|\Big]\Big|\Big] \\ &\quad= \mathbb{E}_a\Big[\frac{1}{2^{2n}}\sum_s w_s^t \left|\mathrm{tr}[(I\otimes P_a + P_a\otimes I)\left|\psi_s^t\right\rangle\left\langle\psi_s^t\right|]\right|\Big] \\ &\quad\leq \frac{1}{2^{2n}}\mathbb{E}_a\Big[\sqrt{\sum_s w_s^t \,\mathrm{tr}^2[(I\otimes P_a + P_a\otimes I)\left|\psi_s^t\right\rangle\left\langle\psi_s^t\right|]\cdot\sum_s w_s^t}\Big] \\ &\quad\leq \frac{1}{2^n}\sqrt{\mathbb{E}_a\Big[\sum_s w_s^t \,\mathrm{tr}^2\left[(I\otimes P_a + P_a\otimes I)\left|\psi_s^t\right\rangle\left\langle\psi_s^t\right|\right]\Big]} \\ &\quad= \frac{1}{2^n}\sqrt{\mathbb{E}_a\Big[\sum_s w_s^t \left\langle\psi_s^t\right|\left\langle\psi_s^t\right|(I\otimes P_a + P_a\otimes I)^{\otimes 2}\left|\psi_s^t\right\rangle\left|\psi_s^t\right\rangle\Big]} \\ &\quad= \frac{1}{2^n}\sqrt{\frac{1}{2^{2n}-1}\sum_s w_s^t \left\langle\psi_s^t\right|\left\langle\psi_s^t\right|\mathbf{N}\left|\psi_s^t\right\rangle\left|\psi_s^t\right\rangle} \\ &\quad\leq \frac{1}{2^n}\sqrt{\frac{1}{2^{2n}-1}\sum_s w_s^t(4\cdot 2^n - 4)} \\ &\quad= \frac{1}{2^n}\sqrt{\frac{4\cdot 2^{2n}}{2^n+1}} = O\Big(\frac{1}{2^{n/2}}\Big)\end{aligned}$$

where the third step follows by Cauchy-Schwarz inequality, the fourth step follows from the fact that $\sum_s w_s^t = 2^{2n}$ and Jensen's inequality, and the sixth step follows from the two-design property of the Clifford group (here $\mathrm{SWAP}_{i,j}$ denotes the SWAP operator on the i-copy and the j-th copy). Therefore, we have

$$\mathrm{TV}(\mathbb{E}_a[p_a^+], \mathbb{E}_a[p_a^-]) \leq O(T\cdot 2^{-n/2}).$$

This indicates that any nonadaptive protocols with $T \leq o(2^{n/2})$ two-copy measurement can not solve this distinguishing task with high probability, which yields the $\Omega(2^{n/2})$ lower bound in Theorem 234 for two-copy nonadaptive protocols for solving Pauli shadow tomography.

4. Lower bound for protocols with limited quantum memory

In this section, we consider an extension of the setting of Section 2.1 to one in which the learner has access to a nonzero but small amount of additional quantum memory - more than is needed to perform single-copy measurements, but not enough to perform two-copy measurements. We prove that the exponential lower bound of Section 2.1 persists in this setting.

The main result of this section is the following:

Theorem 235 (Shadow tomography with bounded quantum memory)**.** *Any learning algorithm with k qubits of quantum memory requires $T \geq \Omega\left(2^{(n-k)/3}\right)$ copies of ρ to predict $|\mathrm{tr}(P\rho)|$ for all n-qubit Pauli observables P with at least probability 2/3.*

We first formalize our notion of *quantum memory* by generalizing the notion of learning tree from before.

Definition 236 (Tree representation for learning states with bounded quantum memory)**.** *Fix an unknown n-qubit quantum state ρ. A learning algorithm with size-k quantum memory can be expressed as a rooted tree $\mathcal{T}$ of depth T, where each node encodes the current state of the quantum memory in addition to the transcript of measurement outcomes the algorithm has seen so far. Specifically, the tree satisfies the following properties:*

(1) Each node u is associated with a k-qubit unnormalized mixed state $\mathbf{\Sigma}^\rho(u)$ corresponding to the current state of the quantum memory.

(2) For the root r of the tree, $\mathbf{\Sigma}^\rho(r)$ is an initial state denoted by $\mathbf{\Sigma}_0$.

(3) At each node u, we apply a POVM measurement $\{F_s^u\}_s$ on $\mathbf{\Sigma}^\rho(u) \otimes \rho$ to obtain a classical outcome s. Each child node v of u is connected through the edge $e_{u,s}$.

(4) For POVM element $F = M^\dagger M$ and $\mathbf{\Sigma} \in \mathbb{H}^{2^k \times 2^k}$, define

$$A_M^\rho(\mathbf{\Sigma}) \triangleq \mathrm{tr}_{>k}\left(M(\mathbf{\Sigma} \otimes \rho)M^\dagger\right).$$

If v is the child node of u connected through the edge $e_{u,s}$, then

$$\mathbf{\Sigma}^\rho(v) \triangleq A_{M_s^u}^\rho(\mathbf{\Sigma}^\rho(u)). \tag{72}$$

$A_M^\rho(\mathbf{\Sigma})$ is the first k-qubit of the unnormalized post-measurement state.

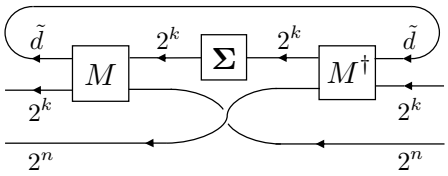
(5) Note that for any node u, $p^\rho(u) \triangleq \mathrm{tr}(\mathbf{\Sigma}^\rho(u))$ is the probability that the transcript of measurement outcomes observed by the learning algorithm after t measurements is u. And $\mathbf{\Sigma}^\rho(u)/p^\rho(u)$ is the state of the k-qubit memory at the node u.

Given ρ, we will abuse notation and let p^ρ denote the distribution on leaves of $\mathcal{T}$ given by the probabilities $\{p^\rho(\ell)\}_\ell$. Let $P_1, \ldots, P_{4^n-1}$ denote the collection of non-identity n-qubit Pauli operators. We will use $\mathbb{E}_P[\cdot]$ to denote expectation with respect to a uniformly random such P.

As before, we consider a many-versus-one distinguishing task where we want to distinguish between the completely mixed state ρ_{mm} versus the set of n-qubit states $\{\rho_P\}$ where P ranges over all n-qubit Pauli observables not equal to identity, where $\rho_P \triangleq (I+P)/2^n$. And as before, a lower bound for this task immediately translates to one for shadow tomography. For the former, it suffices to show that for $T = o\left(2^{(n-k)/3}\right)$, $\mathsf{TV}(p^{\rho_{\mathrm{mm}}}, \mathbb{E}_P[p^{\rho_P}]) = o(1)$.

As in the proof of the single-copy lower bound, the primary technical ingredient will be a second moment bound. To formulate this, we will need the following object:

Definition 237. *Given a POVM element $F = M^\dagger M$ and an unnormalized mixed state $\mathbf{\Sigma} \in \mathbb{H}^{2^k \times 2^k}$, $\tau_{M,\mathbf{\Sigma}} \in \mathbb{H}^{2^{n+k} \times 2^{n+k}}$ is given by*

$\tau_{M,\mathbf{\Sigma}} =$ 

.

(The dimensions of the Hilbert spaces corresponding to the edges have been labeled.)

One can think of the following lemma as bounding a matrix-valued analogue of the quantity defined in the single-copy lower bound. Indeed, when $k = 0$, the following specializes up to constant factors to the second moment bound for the single-copy case.

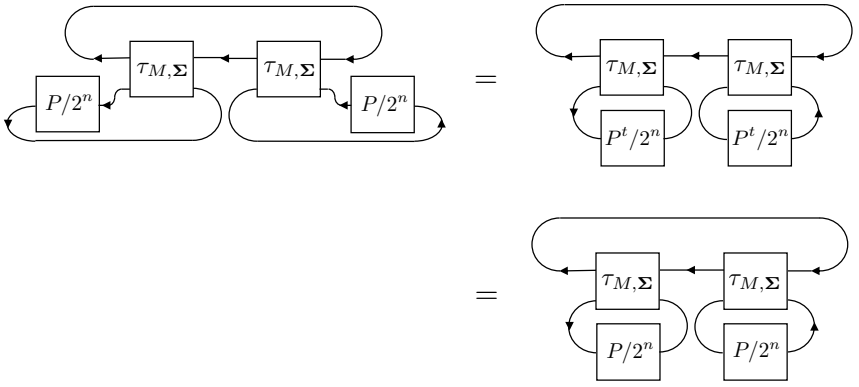
Lemma 238. *For any POVM element M and unnormalized mixed state $\mathbf{\Sigma} \in \mathbb{H}^{2^k \times 2^k}$,*

$$\mathbb{E}_P[\|A_M^{\rho_{\mathrm{mm}}}(\mathbf{\Sigma}) - A_M^{\rho_P}(\mathbf{\Sigma})\|_{\mathsf{tr}}^2] \leq \frac{1}{2^{n-k}} \cdot \frac{1}{2^{2n}-1} \cdot \mathrm{tr}\left(\tau_{M,\mathbf{\Sigma}}\right)^2. \tag{73}$$

Proof. Note that $A_M^{\rho_{\mathrm{mm}}}(\mathbf{\Sigma}) - A_M^{\rho_P}(\mathbf{\Sigma}) = -\mathrm{tr}_{>k}\left(M\left(\frac{P}{2^n} \otimes \mathbf{\Sigma}\right) M^\dagger\right)$, so by the fact that $\|X\|_{\mathrm{tr}}^2 \leq 2^k \mathrm{tr}(X^2), \forall X \in \mathbb{H}^{2^k \times 2^k}$, the left-hand side of (73) is upper bounded by

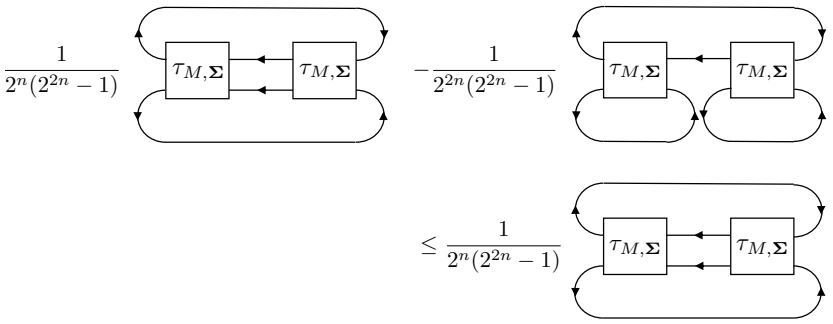
$$2^k \mathbb{E}_P\left[\mathrm{tr}\left(\left[\mathrm{tr}_{>k}\left(M\left(\frac{P}{2^n} \otimes \mathbf{\Sigma}\right) M^\dagger\right)\right]^2\right)\right]. \tag{74}$$

For fixed P, we express the expression inside the expectation diagrammatically as



(75)
.

By the 2-design property of the Clifford group, averaging (75) with respect to P yields



where the inequality follows from the fact that the second term is equal to $\frac{1}{2^{2n}(2^{2n}-1)}\mathrm{tr}(\mathrm{tr}_{>k}(\tau)^2)$ which is non-negative. The claim follows from the fact that $\mathrm{tr}(\tau^2) \leq \mathrm{tr}(\tau)^2$ (as τ is positive-semidefinite) and utilizing Eqn. (74). □

We will not be able to make use of the convexity trick from the proof of the single-copy lower bound. Instead, we make use of a careful pruning argument; intuitively, for any leaf ℓ, we will essentially ignore the contribution to $\mathsf{TV}(p^{\rho_{\mathrm{mm}}}, \mathbb{E}_P[p^{\rho_P}])$ coming from Paulis P for which $A^{\rho_P}_{M_s^u}$ behaves too differently from $A^{\rho_{\mathrm{mm}}}_{M_s^u}$ for some edge $e_{u,s}$ on the path from root to ℓ.

Definition 239. *A Pauli P is* bad *for an edge $e_{u,s}$ if*

$$\|A^{\rho_{\mathrm{mm}}}_{M_s^u}(\mathbf{\Sigma}^{\rho_{\mathrm{mm}}}(u)) - A^{\rho_P}_{M_s^u}(\mathbf{\Sigma}^{\rho_{\mathrm{mm}}}(u))\|_{\mathsf{tr}} \geq \frac{1}{2^{(n-k)/3}} \cdot \sqrt{\frac{1}{2^{2n}-1}} \cdot \mathrm{tr}\left(\tau_{M_s^u, \mathbf{\Sigma}^{\rho_{\mathrm{mm}}}(u)}\right).$$

Otherwise we say P is good *for $e_{u,s}$. Given node u, let $P[u]$ denote the set of all Paulis which are good for all edges on the path from root to u.*

The following is an immediate consequence of Lemma 238 and Markov's:

Fact 240. *For any edge $e_{u,s}$, there are at most $2^{-(n-k)/3} \cdot (4^n - 1)$ bad Paulis $P \in \mathbb{H}^{2^n \times 2^n}$. In particular, along any given root-to-leaf path of the learning tree, there are at most $T \cdot 2^{-(n-k)/3} \cdot (4^n - 1)$ Paulis which are bad for some edge along the path.*

Lemma 240 allows us to bound the $\mathsf{TV}(p^{\rho_{\mathrm{mm}}}, \mathbb{E}_P[p^{\rho_P}])$ by a small term coming from bad Paulis and a term coming from good ones:

Lemma 241.

$$\mathsf{TV}\left(p^{\rho_{\mathrm{mm}}}, \mathbb{E}_P[p^{\rho_P}]\right) \leq T \cdot 2^{-(n-k)/3} + \frac{1}{4^n - 1} \sum_{\ell \in \mathrm{leaf}(\mathcal{T})} \sum_{P \in P[\ell]} \|\mathbf{\Sigma}^{\rho_{\mathrm{mm}}}(\ell) - \mathbf{\Sigma}^{\rho_P}(\ell)\|_{\mathsf{tr}} \tag{76}$$

PROOF. Let $\mathcal{L}$ denote the set of leaves ℓ for which $p^{\rho_{\mathrm{mm}}}(\ell) \geq \mathbb{E}_P[p^{\rho_P}(\ell)]$. Then

$$\begin{aligned}
\mathsf{TV}\left(p^{\rho_{\mathrm{mm}}}, \mathbb{E}_P[p^{\rho_P}]\right) &= \sum_{\ell \in \mathcal{L}} p^{\rho_{\mathrm{mm}}}(\ell) - \mathbb{E}_P[p^{\rho_P}(\ell)] \\
&\leq \sum_{\ell \in \mathcal{L}} \mathbb{E}_P[\min(p^{\rho_{\mathrm{mm}}}(\ell), |p^{\rho_{\mathrm{mm}}}(\ell) - p^{\rho_P}(\ell)|)] \\
&\leq \sum_{\ell \in \mathcal{L}} \mathbb{E}_P[\min(p^{\rho_{\mathrm{mm}}}(\ell), \|\boldsymbol{\Sigma}^{\rho_{\mathrm{mm}}}(\ell) - \boldsymbol{\Sigma}^{\rho_P}(\ell)\|_{\mathsf{tr}})] \qquad (77) \\
&\leq \sum_{\ell \in \mathcal{L}} [\![*]\!] \Pr P \notin P[\ell] \cdot p^{\rho_{\mathrm{mm}}}(\ell) + \frac{1}{4^n - 1} \sum_{P \in P[\ell]} \|\boldsymbol{\Sigma}^{\rho_{\mathrm{mm}}}(\ell) - \boldsymbol{\Sigma}^{\rho_P}(\ell)\|_{\mathsf{tr}},
\end{aligned}$$

The first equality uses the fact that $\mathsf{TV}(p, q) = \frac{1}{2}\sum_i |p_i - q_i| = \sum_{i: p_i \geq q_i}(p_i - q_i)$. Inequality (77) uses the fact that $\|\boldsymbol{\Sigma}^{\rho_{\mathrm{mm}}}(\ell) - \boldsymbol{\Sigma}^{\rho_P}(\ell)\|_{\mathsf{tr}} \geq \mathrm{tr}(\boldsymbol{\Sigma}^{\rho_{\mathrm{mm}}}(\ell) - \boldsymbol{\Sigma}^{\rho_P}(\ell)) = p^{\rho_{\mathrm{mm}}}(\ell) - p^{\rho_P}(\ell)$. The lemma follows from Fact 240 and the fact that $\sum_\ell p^{\rho_{\mathrm{mm}}}(\ell) \leq 1$. □

We are now ready to prove Theorem 235.

PROOF OF THEOREM 235. By Lemma 241, it suffices to control the latter term on the right-hand side of Eq. (76). We do so via a hybrid argument. For any leaf ℓ with parent u and incoming edge $e_{u,s}$, and any $P \in P[\ell]$, we can upper bound $\|\boldsymbol{\Sigma}^{\rho_{\mathrm{mm}}}(\ell) - \boldsymbol{\Sigma}^{\rho_P}(\ell)\|_{\mathsf{tr}}$ by

$$\|A^{\rho_{\mathrm{mm}}}_{M^u_s}(\boldsymbol{\Sigma}^{\rho_{\mathrm{mm}}}(u)) - A^{\rho_P}_{M^u_s}(\boldsymbol{\Sigma}^{\rho_{\mathrm{mm}}}(u))\|_{\mathsf{tr}} + \|A^{\rho_P}_{M^u_s}\left(\boldsymbol{\Sigma}^{\rho_{\mathrm{mm}}}(u) - \boldsymbol{\Sigma}^{\rho_P}(u)\right)\|_{\mathsf{tr}}, \qquad (78)$$

where we have used the transition formula in Eq. (72) and the triangle inequality. We can upper bound the former term by

$$\frac{1}{2^{(n-k)/3}} \cdot \sqrt{\frac{2^{2n}}{2^{2n} - 1}} \cdot \mathrm{tr}\left(\frac{1}{2^n} \tau_{M^u_s, \boldsymbol{\Sigma}^{\rho_{\mathrm{mm}}}(u)}\right) \qquad (79)$$

because P is good for edge $e_{u,s}$; see Definition 239 and note that $\sqrt{2^{2n}} \cdot \frac{1}{2^n} = 1$. Using Definition 237 and the fact that $\{(M^u_s)^\dagger M^u_s\}_s$ is a POVM hence $\sum_s (M^u_s)^\dagger M^u_s = \mathbb{1}$, we have the following identity.

$$\sum_s \mathbb{E}_P[\mathrm{tr}\left(\frac{1}{2^n} \tau_{M^u_s, \boldsymbol{\Sigma}^{\rho_{\mathrm{mm}}}(u)}\right)] = \mathbb{E}_P[\mathrm{tr}(\boldsymbol{\Sigma}^{\rho_{\mathrm{mm}}}(u))], \ \forall u : \text{node on } \mathcal{T}.$$

Therefore, we have

$$\sum_{\substack{u,s:\ e_{u,s} \text{ connected} \\ \text{to a leaf}}} \mathbb{E}_P\left[\mathrm{tr}\left(\frac{1}{2^n} \tau_{M^u_s, \boldsymbol{\Sigma}^{\rho_{\mathrm{mm}}}(u)}\right)\right] = \mathbb{E}_P\left[\sum_{\substack{u:\ \text{connected} \\ \text{to a leaf}}} \mathrm{tr}(\boldsymbol{\Sigma}^{\rho_{\mathrm{mm}}}(u))\right] = 1.$$

As for the latter term in Eq. (78), note that for any leaf ℓ with parent u, $P[u] \subseteq P[\ell]$, so

$$\sum_{\ell, P \in P[\ell]} \| A_{M_s^u}^{\rho_P} \left(\mathbf{\Sigma}^{\rho_{\mathrm{mm}}}(u) - \mathbf{\Sigma}^{\rho_P}(u) \right) \|_{\mathsf{tr}} \leq \sum_{\substack{u,s:\ e_{u,s} \text{ connected} \\ \text{to a leaf}, P \in P[u]}} \| A_{M_s^u}^{\rho_P} \left(\mathbf{\Sigma}^{\rho_{\mathrm{mm}}}(u) - \mathbf{\Sigma}^{\rho_P}(u) \right) \|_{\mathsf{tr}}$$

$$= \sum_{\substack{u:\ u \text{ connected} \\ \text{to a leaf}, P \in P[u]}} \sum_s \| A_{M_s^u}^{\rho_P} \left(\mathbf{\Sigma}^{\rho_{\mathrm{mm}}}(u) - \mathbf{\Sigma}^{\rho_P}(u) \right) \|_{\mathsf{tr}} \tag{80}$$

Note that for any node u, the map

$$\mathcal{E} : X \mapsto \sum_s |s\rangle \langle s| \, s \otimes A_{M_s^u}^{\rho_P}(X)$$

is a quantum channel, so in particular $\| \mathcal{E}(X) \|_{\mathsf{tr}} \leq \| X \|_{\mathsf{tr}}$. In particular, we can see that

$$\| \mathcal{E}(\mathbf{\Sigma}^{\rho_{\mathrm{mm}}}(u) - \mathbf{\Sigma}^{\rho_P}(u)) \|_{\mathsf{tr}} = \sum_s \| A_{M_s^u}^{\rho_P} \left(\mathbf{\Sigma}^{\rho_{\mathrm{mm}}}(u) - \mathbf{\Sigma}^{\rho_P}(u) \right) \|_{\mathsf{tr}}.$$

We may thus upper bound Eq. (80) by

$$\sum_{\substack{u:\ \text{connected to} \\ \text{a leaf}, P \in P[u]}} \| \mathbf{\Sigma}^{\rho_{\mathrm{mm}}}(u) - \mathbf{\Sigma}^{\rho_P}(u) \|_{\mathsf{tr}} = \sum_{\substack{u:\ \text{at depth } T-1 \\ P \in P[u]}} \| \mathbf{\Sigma}^{\rho_{\mathrm{mm}}}(u) - \mathbf{\Sigma}^{\rho_P}(u) \|_{\mathsf{tr}}. \tag{81}$$

Combining Eq. (78), (79), and (81), we conclude that

$$\sum_{\substack{\ell \in \mathrm{leaf}(\mathcal{T}) \\ P \in P[\ell]}} \| \mathbf{\Sigma}^{\rho_{\mathrm{mm}}}(\ell) - \mathbf{\Sigma}^{\rho_P}(\ell) \|_{\mathsf{tr}} \leq \frac{1}{2^{(n-k)/3}} \cdot \sqrt{\frac{2^{2n}}{2^{2n}-1}} + \sum_{\substack{u \text{ at depth } T-1 \\ P \in P[u]}} \| \mathbf{\Sigma}^{\rho_{\mathrm{mm}}}(u) - \mathbf{\Sigma}^{\rho_P}(u) \|_{\mathsf{tr}}$$

We can thus conclude by induction and by Lemma 241,

$$\mathsf{TV}\left(p^{\rho_{\mathrm{mm}}}, \mathbb{E}_P[p^{\rho_P}] \right) \leq T \cdot 2^{-(n-k)/3} + T \cdot 2^{-(n-k)/3} \cdot \sqrt{\frac{2^{2n}}{2^{2n}-1}}.$$

In order to achieve the many-versus-one distinguishing task with probability at least $2/3$, we must have $2/3 \leq \mathsf{TV}(p^{\rho_{\mathrm{mm}}}, \mathbb{E}_P[p^{\rho_P}])$ from Le Cam's two point method. However for $T = o(2^{(n-k)/3})$, $\mathsf{TV}(p^{\rho_{\mathrm{mm}}}, \mathbb{E}_P[p^{\rho_P}]) = o(1)$. This concludes the lower bound that $T \geq \Omega(2^{(n-k)/3})$. □