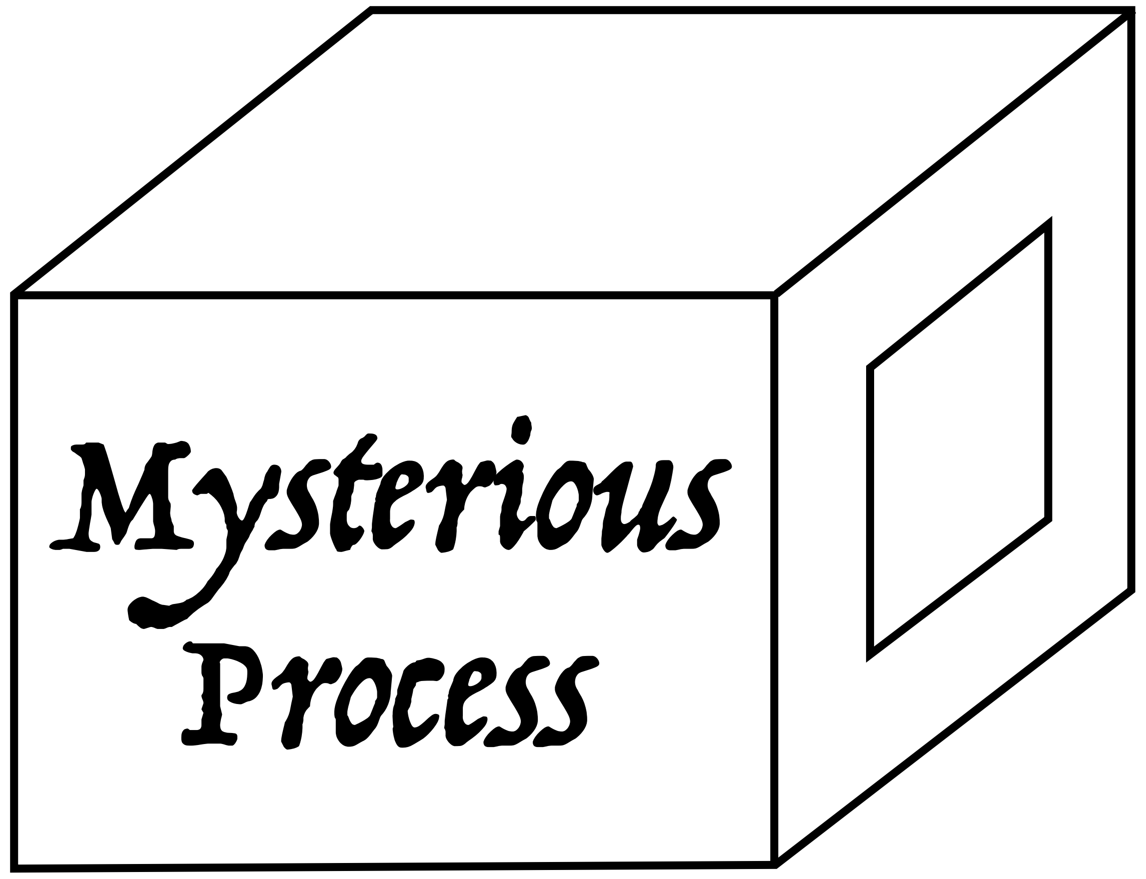
Sample-optimal classical shadows for pure states

Daniel Grier
UC San Diego

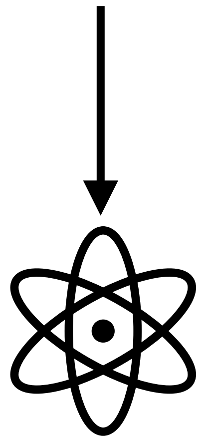
Hakop Pashayan
Freie Universität Berlin

Luke Schaeffer
University of Maryland

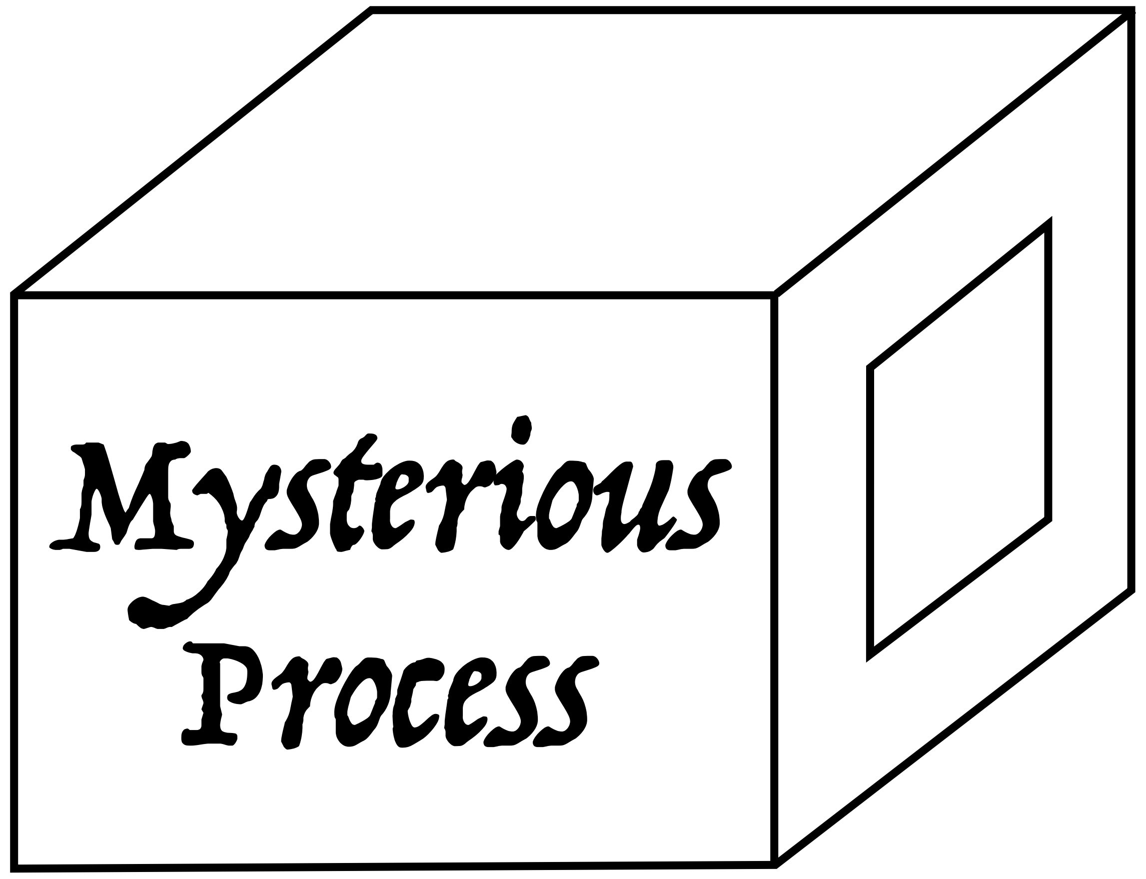
Classical shadows picture and review



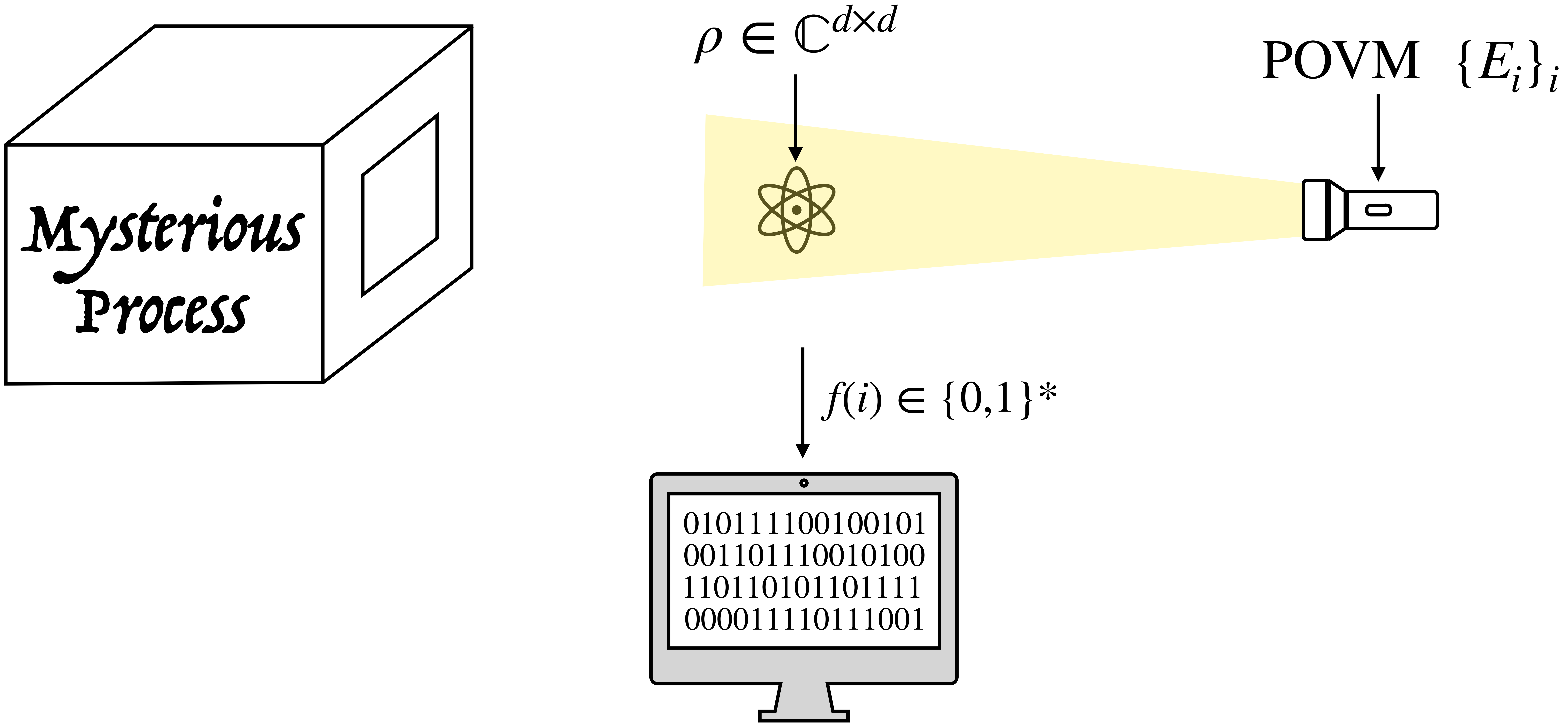
$\rho \in \mathbb{C}^{d\times d}$



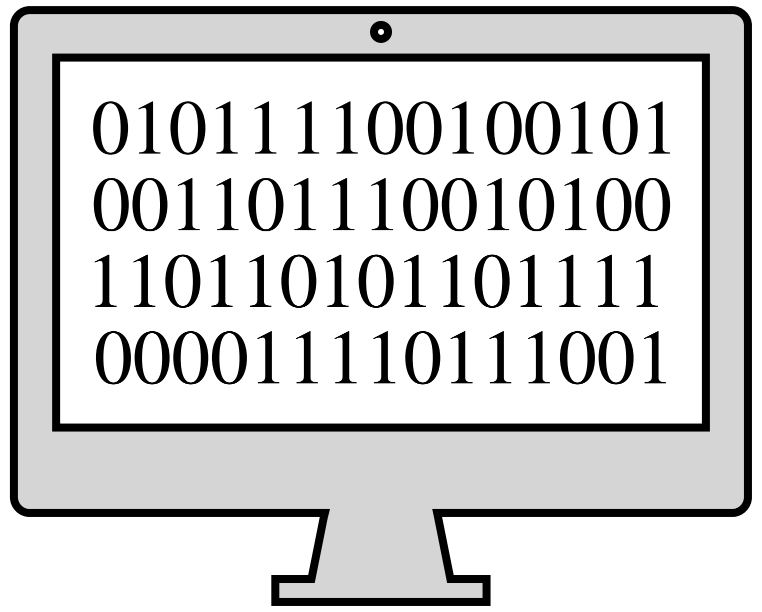
Classical shadows picture and review



Classical shadows picture and review



Classical shadows picture and review



Hermitian operator $H \in \mathbb{C}^{d\times d}$

Estimate $\hat{h}$ such that

$$|\hat{h} - \mathrm{Tr}(\mathrm{H}\rho)| \leq \epsilon$$

Question: How many copies of ρ needed to succeed w.h.p for any $H \in \mathscr{H}$?

Variants of classical shadows - Local Clifford

Question: How many copies of ρ needed to succeed w.h.p for any $H \in \mathscr{H}$?

Variant 1 (Local Clifford): Measure each qubit of ρ in a random Pauli basis.

→ [Huang, Kueng, Preskill 2020]: If $\mathscr{H} = \{k\text{-local Hamiltonians}\}$

Probability of success δ

Sample complexity: $$O\left(\frac{4^k}{\epsilon^2}\log(\delta^{-1})\right)$$

Accuracy parameter ϵ:
$|\hat{h} - \mathrm{Tr}(H\rho)| \leq \epsilon$

Variants of classical shadows - Global Clifford

Question: How many copies of ρ needed to succeed w.h.p for any $H \in \mathscr{H}$?

Variant 2 (Global Clifford): Apply random Clifford unitary to ρ and measure in the computational basis.

[Huang, Kueng, Preskill 2020]: $||\mathscr{H}||_F^2 = \max_{H \in \mathscr{H}} ||H||_F^2 = \max_{H \in \mathscr{H}} \mathrm{Tr}(H^2)$

Sample complexity: $$O\left(\frac{||\mathscr{H}||_F^2}{\epsilon^2} \log(\delta^{-1})\right)$$

Observation: If $\mathscr{H}$ has observables with operator norm 1, then $||\mathscr{H}||_F^2 = d$.

Variants of classical shadows - Global Clifford

Question: How many copies of ρ needed to succeed w.h.p for any $H \in \mathscr{H}$?

Variant 2 (Global Clifford): Apply random Clifford unitary to ρ and measure in the computational basis.

→ [Huang, Kueng, Preskill 2020]: $||\mathscr{H}||_F^2 = \max_{H\in\mathscr{H}} ||H||_F^2 = \max_{H\in\mathscr{H}} \mathrm{Tr}(H^2)$

Sample complexity: $O\left(\frac{d}{\epsilon^2}\log(\delta^{-1})\right)$

Compare to tomography: $O(d^3/\epsilon^2)$

Observation: If $\mathscr{H}$ has observables with operator norm 1, then $||\mathscr{H}||_F^2 = d$.

Variants of classical shadows - Joint measurement

Question: How many copies of ρ needed to succeed w.h.p for any $H \in \mathscr{H}$?

Variant 3 (Joint measurement): Get all copies of ρ at once (i.e., $\rho^{\otimes n}$) and can make an arbitrary measurement across all copies.

[G, Pashayan, Schaeffer]: Assuming ρ is *pure*

Sample complexity: $$O\left(\left(\frac{\|\mathscr{H}\|_F}{\epsilon} + \frac{1}{\epsilon^2}\right)\log(\delta^{-1})\right)$$

$$O\left(\frac{\|\mathscr{H}\|_F^2}{\epsilon^2}\log(\delta^{-1})\right)$$

Theorem: In constant δ regime, this is tight (up to a log factor).

Compression vs. classical shadows for pure states

Classical Shadows:
How many copies of $\psi \in \mathbb{C}^d$ do we need to *measure* to estimate the expected value of an unknown observable with high probability?

Compression: Given an *explicit description* of $\psi \in \mathbb{C}^d$ as a list of amplitudes, how many bits do we need to write down to estimate an unknown observable?

Theorem:

$$O\left(\left(\frac{\sqrt{d}}{\epsilon} + \frac{1}{\epsilon^2}\right)\log(\delta^{-1})\right)$$

sample are sufficient

Theorem [Gosset, Smolin 2018]:

$$\tilde{\Theta}\left(\left(\frac{\sqrt{d}}{\epsilon} + \frac{1}{\epsilon^2}\right)\log(\delta^{-1})\right)$$

bits are sufficient and necessary

Intuition: Measurements are giving the maximum possible information

Revisiting Variant 2 (Global Clifford) with pure states

Variant 2 (Independent measurement): Can only measure a single copy of ρ at once, but can apply arbitrary measurement.

Question: What happens when ρ is pure?

- Can still use Huang-Kueng-Preskill algorithm: $O(\|\mathscr{H}\|_F^2/\epsilon^2)$
- But their lower bound uses high rank states…

Sample complexity [G, Pashayan, Schaeffer]: $$O\left(\frac{\sqrt{d}\|\mathscr{H}\|_F}{\epsilon} + \frac{1}{\epsilon^2}\right)$$

- If $\|\mathscr{H}\|_F = d^{1/6}$ and $\epsilon = d^{-2/3}$, this is better than HKP bound

Outline for rest of talk

Sample complexity [G, Pashayan, Schaeffer]: $$O\left(\frac{\sqrt{d}\|\mathscr{H}\|_F}{\epsilon} + \frac{1}{\epsilon^2}\right)$$

Revisit the Huang-Kueng-Preskill algorithm

Define new estimator for the pure state case

Sketch analysis

Tensor networks

Representation theory

POVM that can be used for the Global Clifford analysis

HKP Global Clifford measurement: Apply random Clifford unitary to ρ and measure in the computational basis.

→ We use a continuous POVM instead

$$\left\{ d\, \mathbf{v}\mathbf{v}^\dagger \, \mathrm{d}\mu(\mathbf{v}) \right\}_{\mathbf{v}\in\mathbb{C}^d} \qquad d\int \mathbf{v}\mathbf{v}^\dagger \, \mathrm{d}\mu(\mathbf{v}) = I$$

↑ *Haar measure*

These measurements turn out to be equally good in the independent measurement setting

→ But not the joint measurement setting! $\propto \left\{ (\mathbf{v}\mathbf{v}^\dagger)^{\otimes n} \, \mathrm{d}\mu(\mathbf{v}) \right\}_{\mathbf{v}\in\mathbb{C}^d}$

Overview Huang-Kueng-Preskill algorithm

1) Choose measurement operators $\left\{ d\, \mathbf{v}\mathbf{v}^\dagger \; \mathrm{d}\mu(\mathbf{v}) \right\}_{\mathbf{v}\in\mathbb{C}^d}$
2) From measurement result, estimate state $\hat{\rho} = (d+1)\mathbf{v}\mathbf{v}^\dagger - I$
3) Compute expectation $\mathbb{E}[\hat{\rho}] = \rho \implies \mathbb{E}[\mathrm{Tr}(H\hat{\rho})] = \mathrm{Tr}(H\rho)$
4) Compute variance $\mathrm{Var}[\mathrm{Tr}(H\hat{\rho})] \approx \|H\|_F^2$

Repeat n times and average $\hat{\rho}_{\text{avg}} = \dfrac{\hat{\rho}_1 + \hat{\rho}_2 + \ldots + \hat{\rho}_n}{n}$

5) Variance of average $\mathrm{Var}[\mathrm{Tr}(H\hat{\rho}_{\text{avg}})] \approx \|H\|_F^2/n$
6) Chebyshev's inequality $\Pr[\,|\mathrm{Tr}(H\hat{\rho}_{\text{avg}}) - \mathrm{Tr}(H\rho)| \geq \epsilon] \leq \dfrac{\|H\|_F^2}{n\epsilon^2}$

Overview of our proof strategy for pure states

1) Choose measurement operators $\left\{d\, \mathbf{v}\mathbf{v}^\dagger \, \mathrm{d}\mu(\mathbf{v})\right\}_{\mathbf{v}\in\mathbb{C}^d}$

2) From measurement result, estimate state $\hat{\rho} = (d+1)\mathbf{v}\mathbf{v}^\dagger - I$

Repeat n times $\hat{\rho}_{\text{pairs}} = \dfrac{1}{n(n-1)} \sum_{i\neq j} \hat{\rho}_i \hat{\rho}_j$

3) Compute expectation $\mathbb{E}[\hat{\rho}_i\hat{\rho}_j] = \mathbb{E}[\hat{\rho}_i]\mathbb{E}[\hat{\rho}_j] = \rho\rho = \rho$

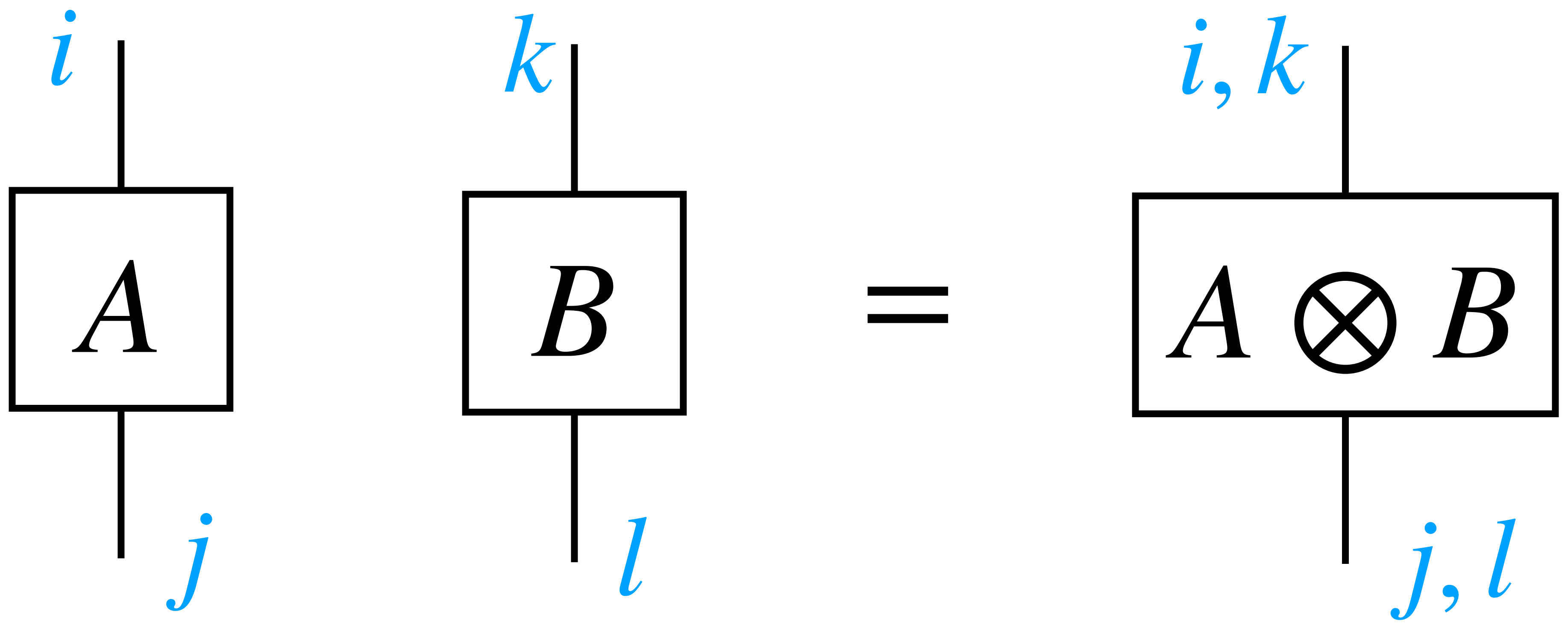
purity of ρ

4) Compute variance $\mathrm{Var}[\mathrm{Tr}(H\hat{\rho}_{\text{pairs}})] \approx d\|H\|_F^2/n^2 + 1/n$

Tensor Network Detour

Tensor networks

$$A = \begin{pmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,d} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ A_{d,1} & A_{d,2} & \cdots & A_{d,d} \end{pmatrix} = \{A_{i,j}\}_{i,j\in[d]}$$



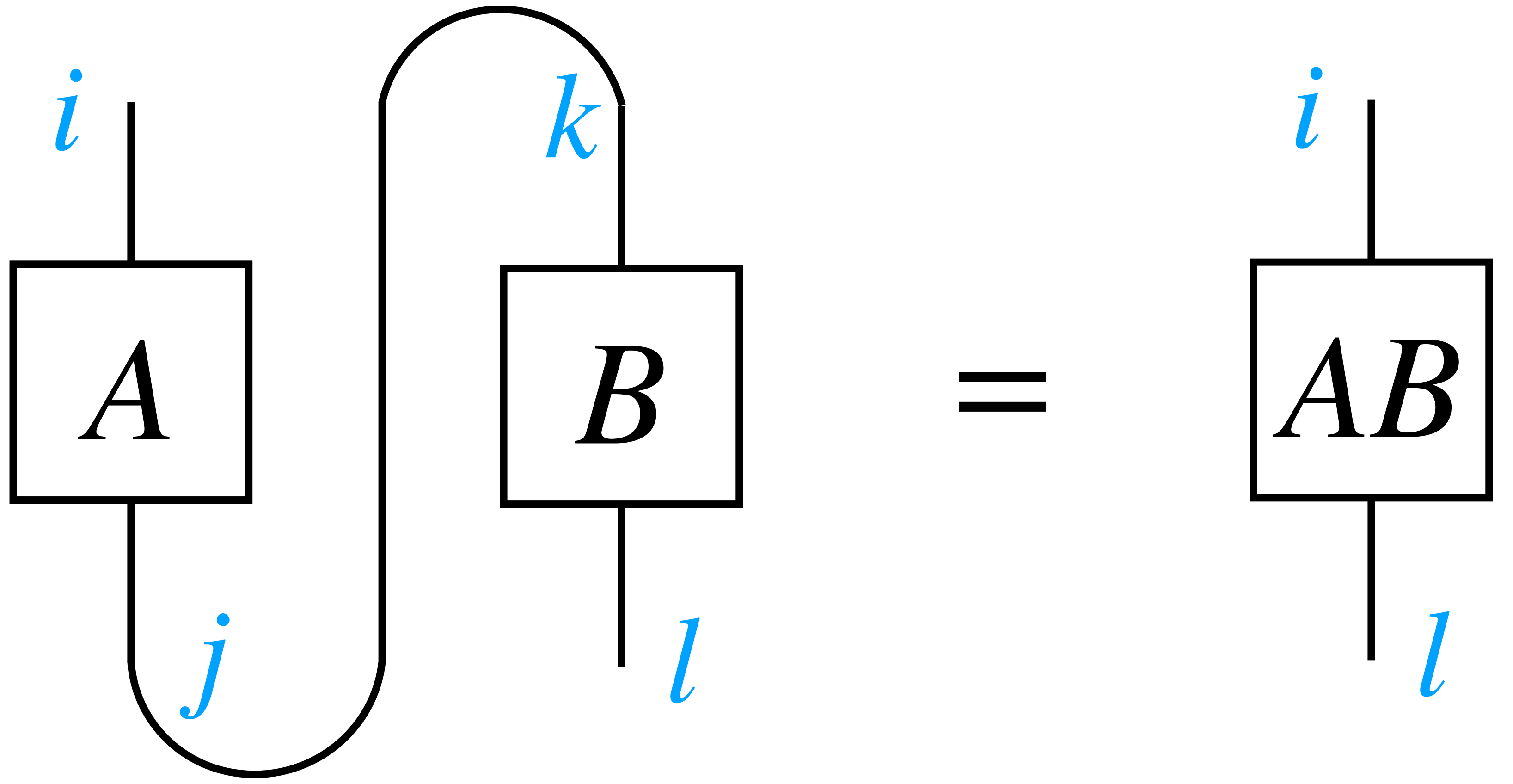
$$B = \begin{pmatrix} B_{1,1} & B_{1,2} & \cdots & B_{1,d} \\ B_{2,1} & B_{2,2} & \cdots & B_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ B_{d,1} & B_{d,2} & \cdots & B_{d,d} \end{pmatrix} = \{B_{k,l}\}_{k,l\in[d]}$$

Tensor Product:

$$(A \otimes B)_{i,k,j,l} = A_{i,j}B_{k,l}$$

Tensor networks

$$A = \begin{pmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,d} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ A_{d,1} & A_{d,2} & \cdots & A_{d,d} \end{pmatrix} = \{A_{i,j}\}_{i,j \in [d]}$$



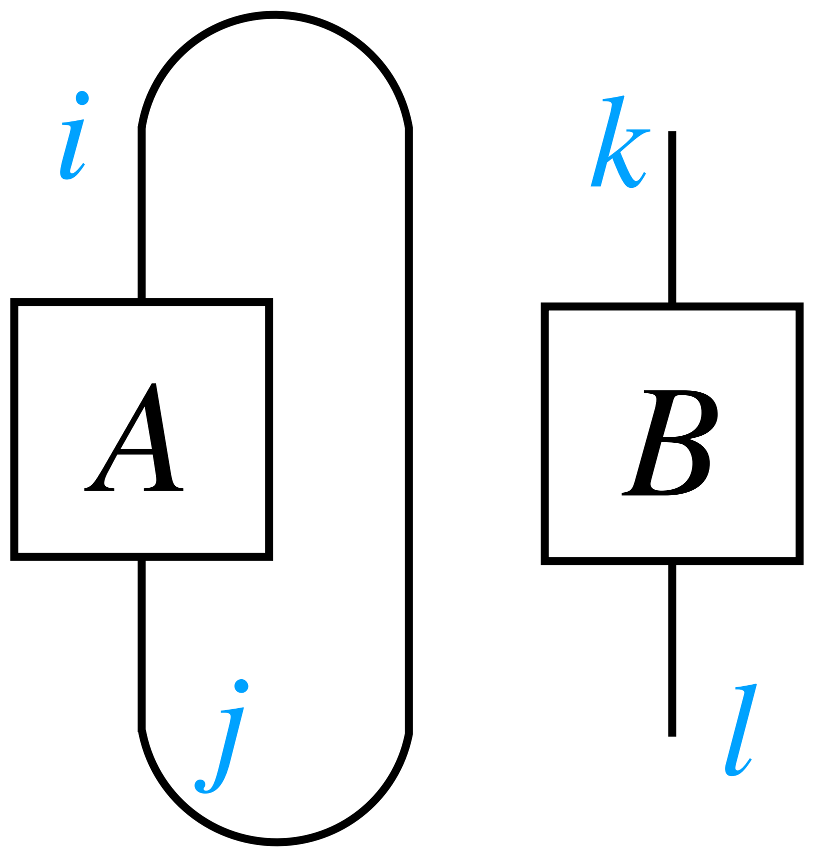
$$B = \begin{pmatrix} B_{1,1} & B_{1,2} & \cdots & B_{1,d} \\ B_{2,1} & B_{2,2} & \cdots & B_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ B_{d,1} & B_{d,2} & \cdots & B_{d,d} \end{pmatrix} = \{B_{k,l}\}_{k,l \in [d]}$$

Tensor Composition:

$$(AB)_{i,l} = \sum_k A_{i,k} B_{k,l}$$

Tensor networks

$$A = \begin{pmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,d} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ A_{d,1} & A_{d,2} & \cdots & A_{d,d} \end{pmatrix} = \{A_{i,j}\}_{i,j \in [d]}$$

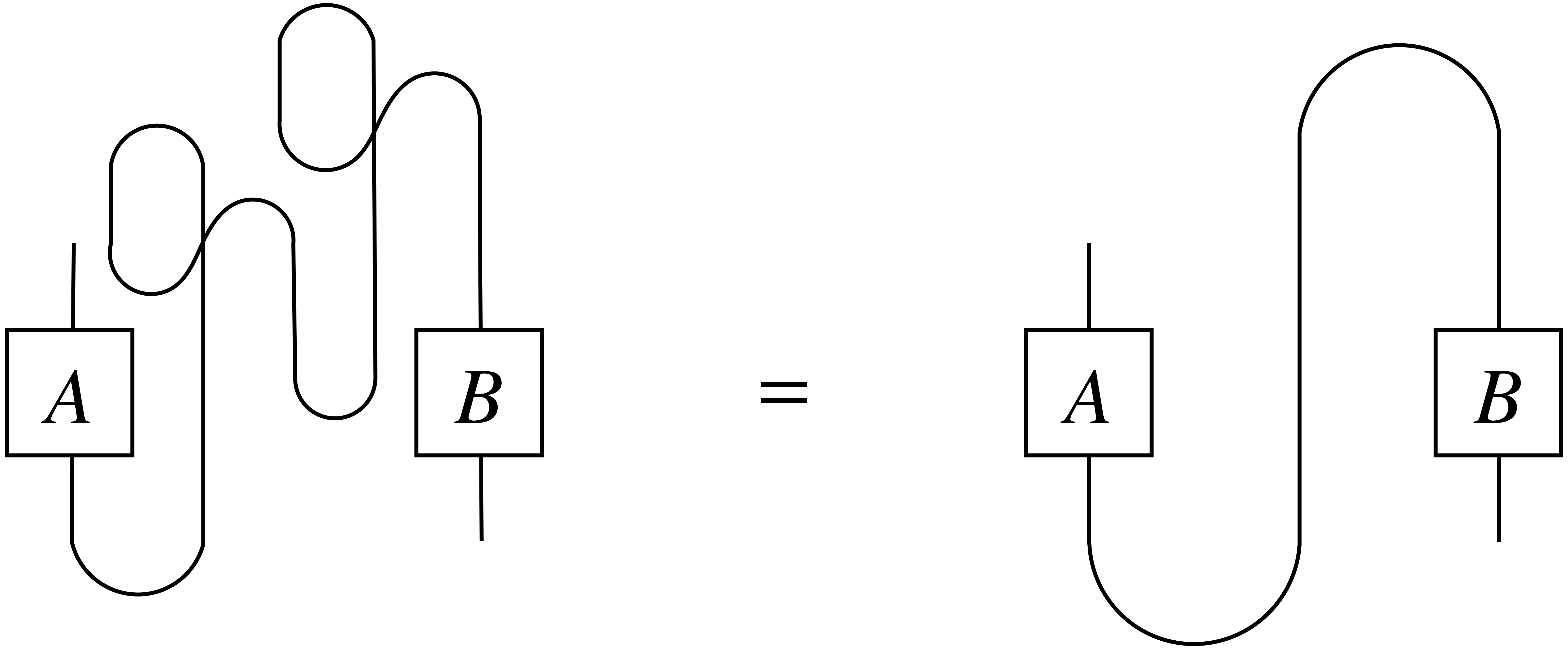


$$B = \begin{pmatrix} B_{1,1} & B_{1,2} & \cdots & B_{1,d} \\ B_{2,1} & B_{2,2} & \cdots & B_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ B_{d,1} & B_{d,2} & \cdots & B_{d,d} \end{pmatrix} = \{B_{k,l}\}_{k,l \in [d]}$$

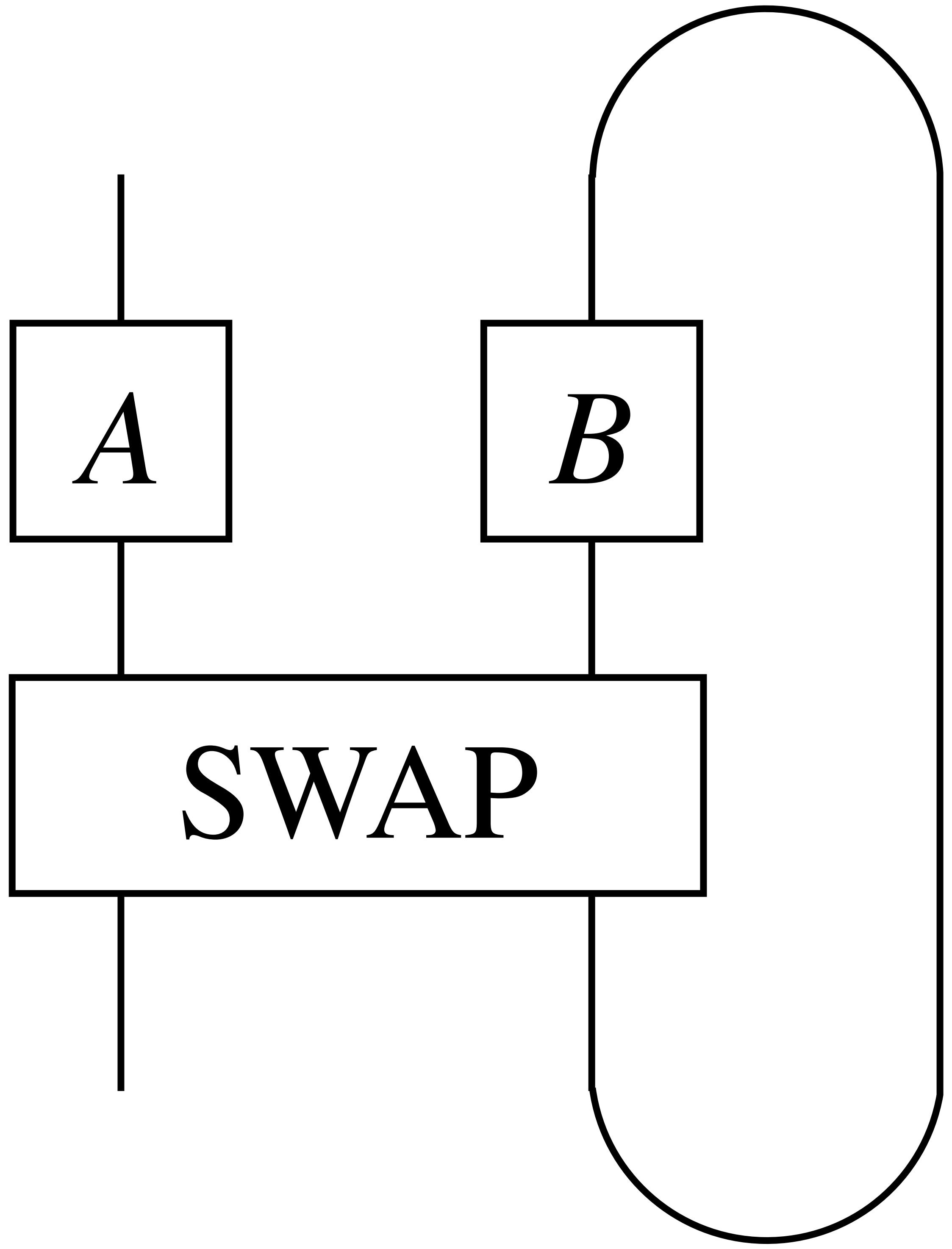
Self loops:

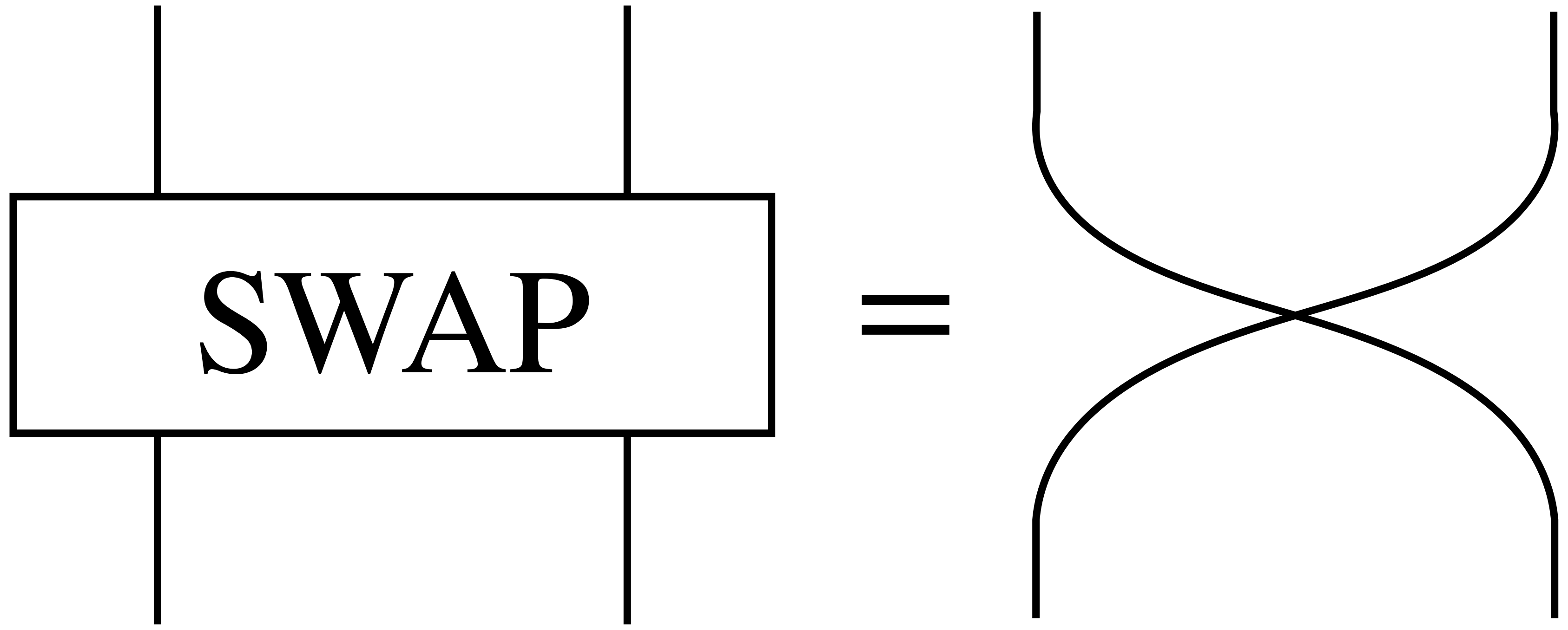
$$B_{k,l} \sum_i A_{i,i} = \mathrm{Tr}(A) B_{k,l}$$

Tensor networks

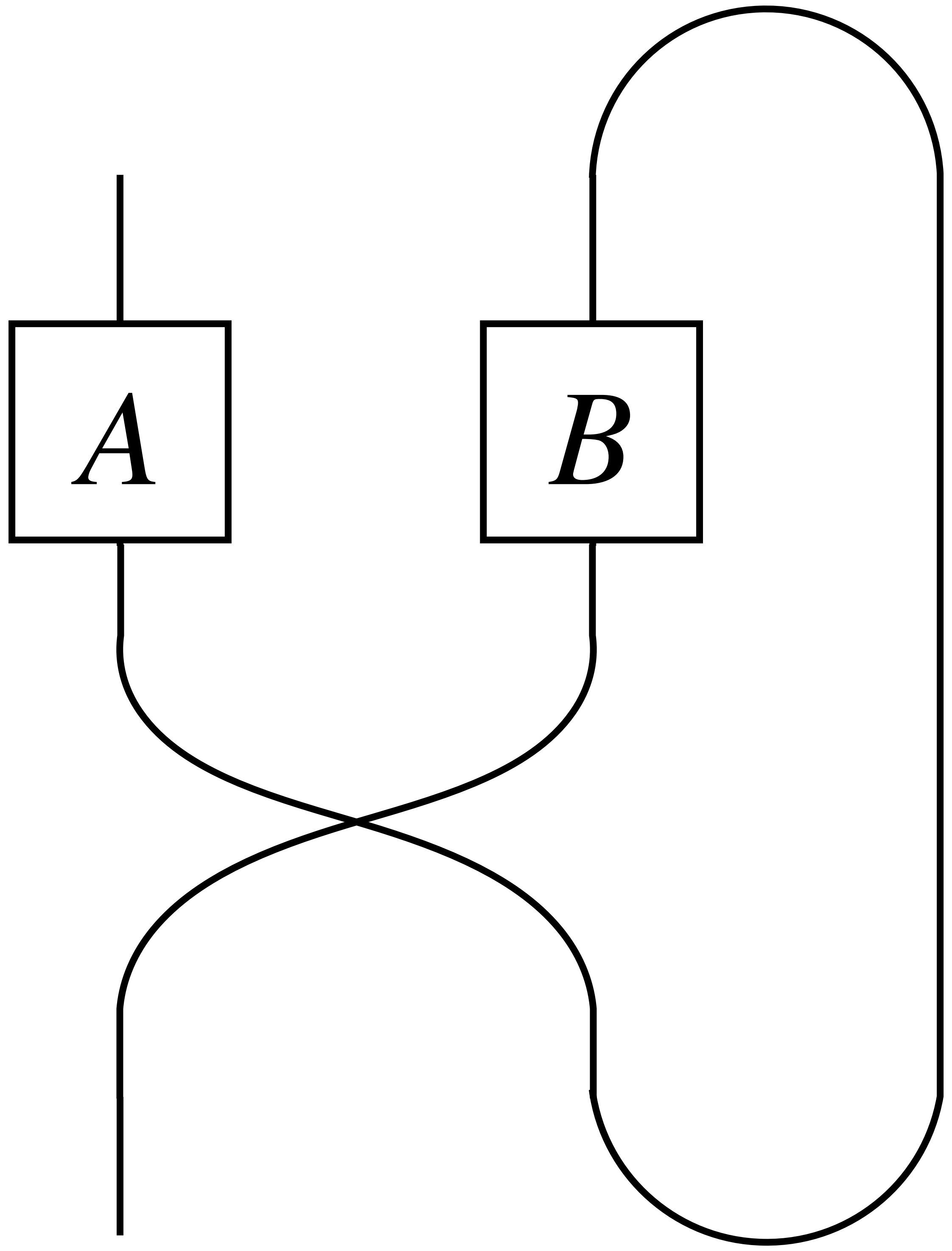


Tensor networks

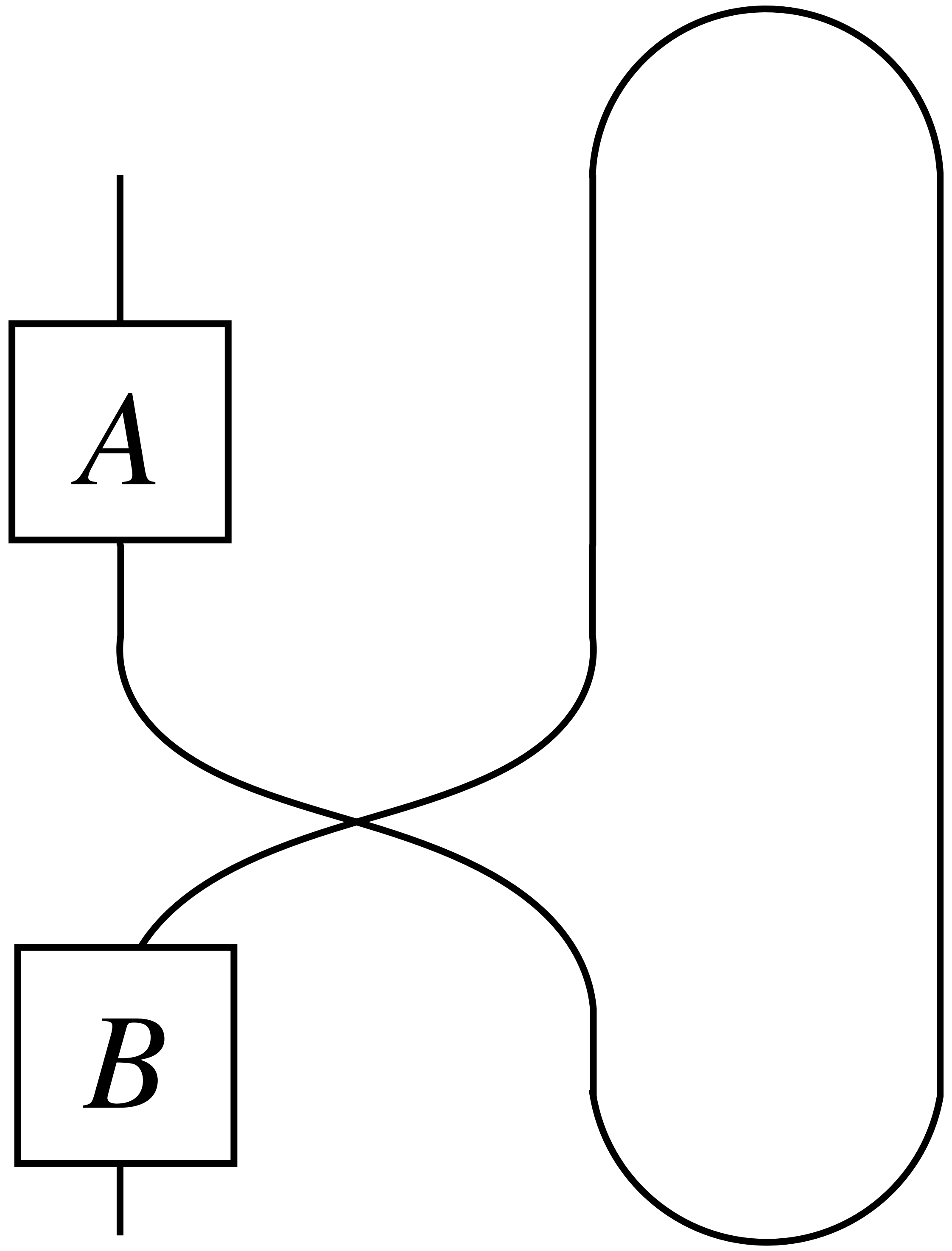




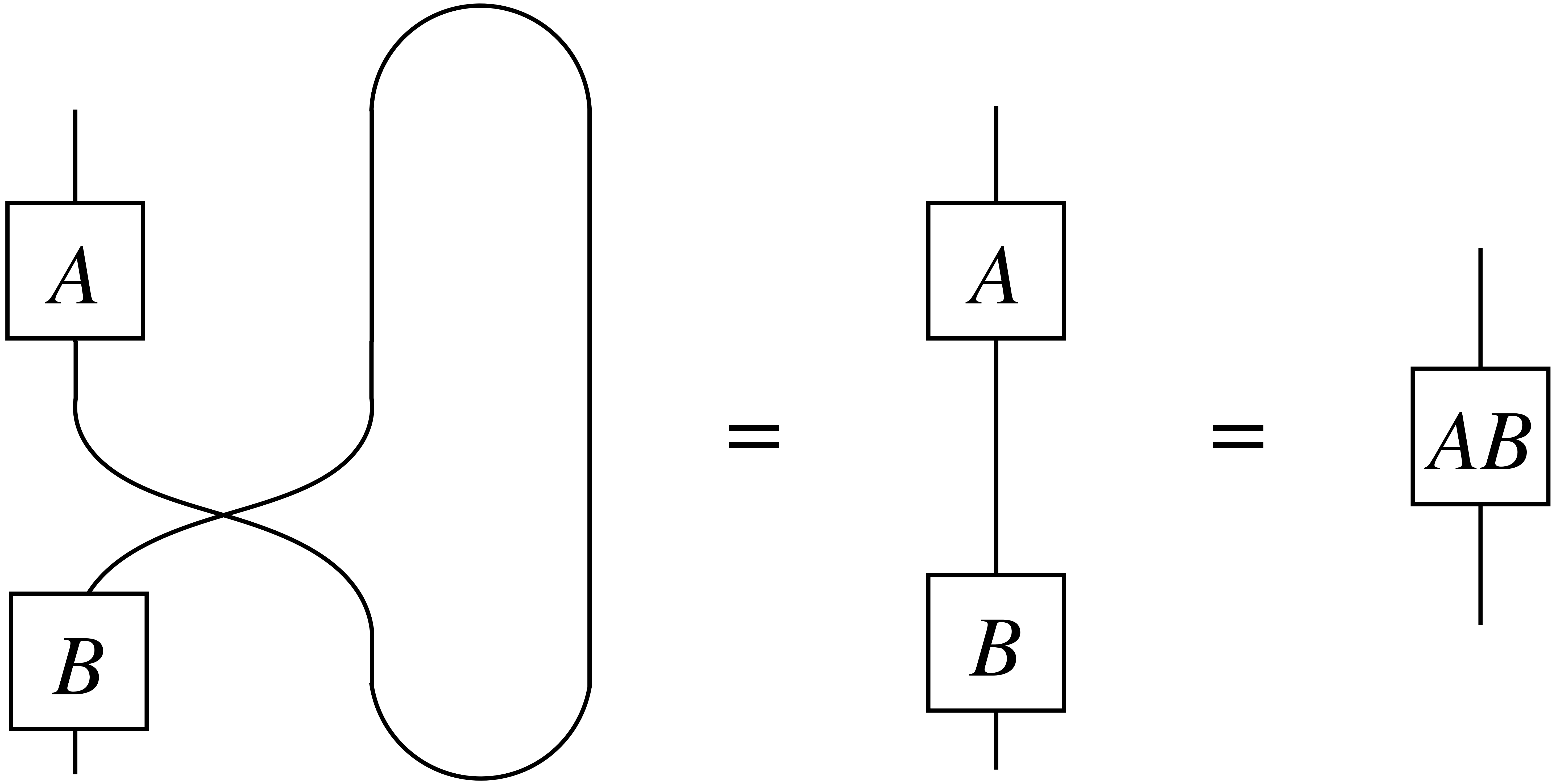
Tensor networks



Tensor networks



Tensor networks



Review of strategy

Measurement operators $\left\{ d\, \mathbf{v}\mathbf{v}^{\dagger}\, \mathrm{d}\mu(\mathbf{v}) \right\}_{\mathbf{v}\in\mathbb{C}^d}$

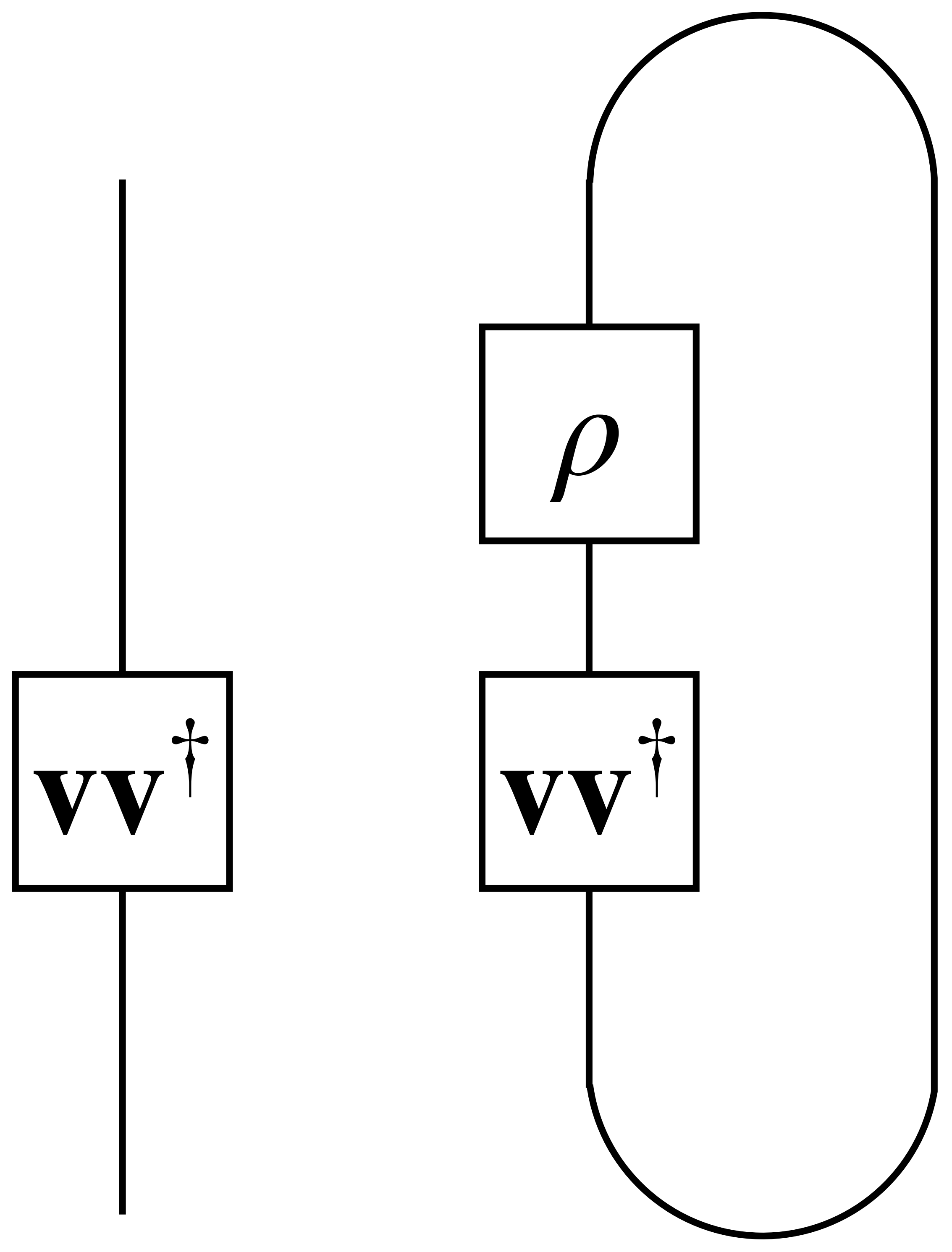
Estimate of state $\hat{\rho} = (d+1)\mathbf{v}\mathbf{v}^{\dagger} - I$

Goal: Show $\mathbb{E}[\hat{\rho}] = \rho$.

- V is random variable that is $\mathbf{v}\mathbf{v}^{\dagger}$ with probability $d\mathrm{Tr}(\mathbf{v}\mathbf{v}^{\dagger}\rho)\, \mathrm{d}\mu(\mathbf{v})$
- $\mathbb{E}[\hat{\rho}] = (d+1)\mathbb{E}[V] - I$

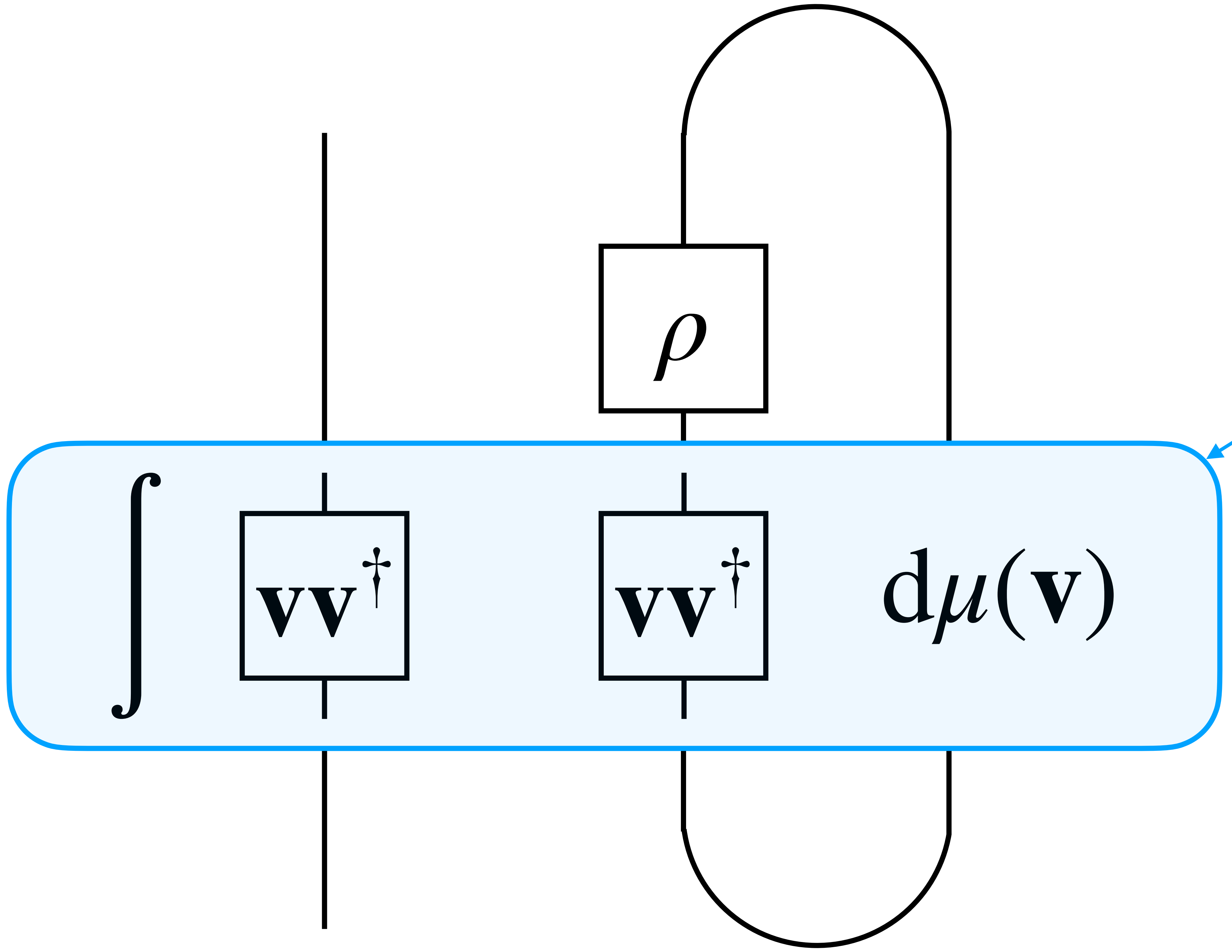
Computing the expectation

$$\mathbb{E}[V] = d \int \mathbf{v}\mathbf{v}^\dagger \mathrm{Tr}(\mathbf{v}\mathbf{v}^\dagger \rho)\ \mathrm{d}\mu(\mathbf{v})$$

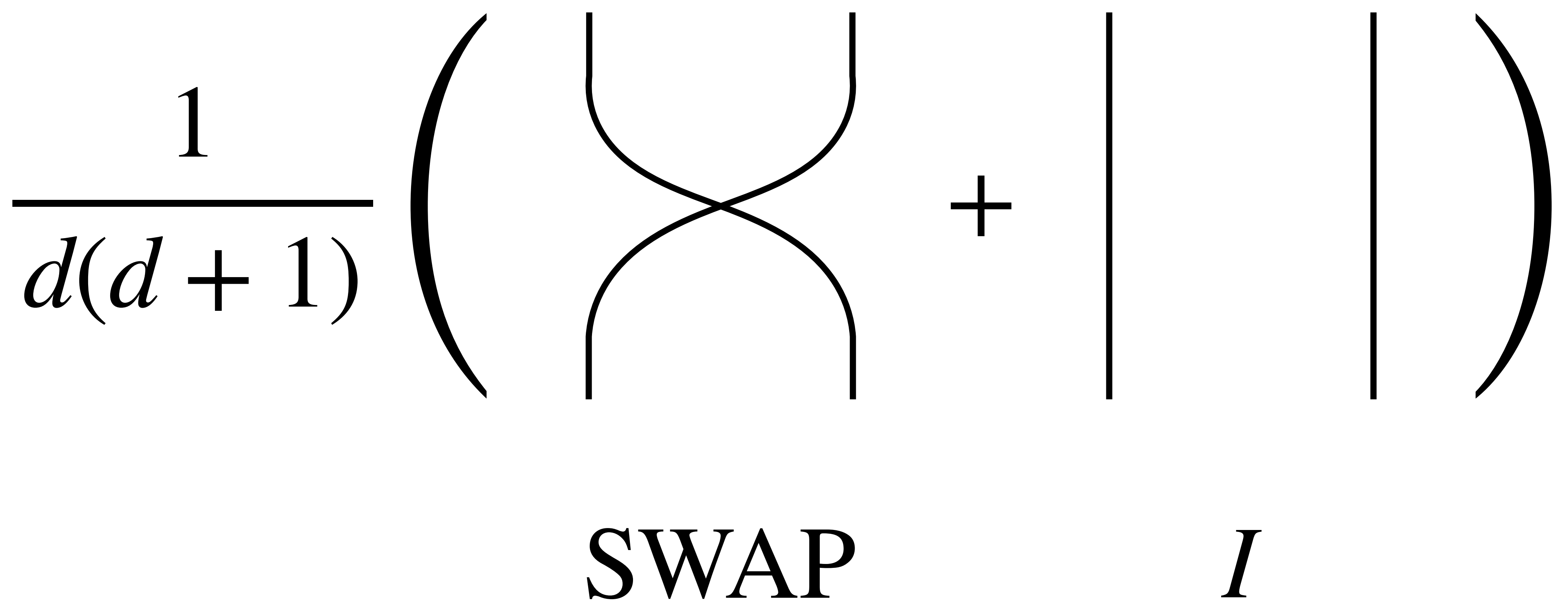


Computing the expectation

$$\mathbb{E}[V] = d \int \mathbf{v}\mathbf{v}^\dagger \mathrm{Tr}(\mathbf{v}\mathbf{v}^\dagger \rho)\ \mathrm{d}\mu(\mathbf{v})$$

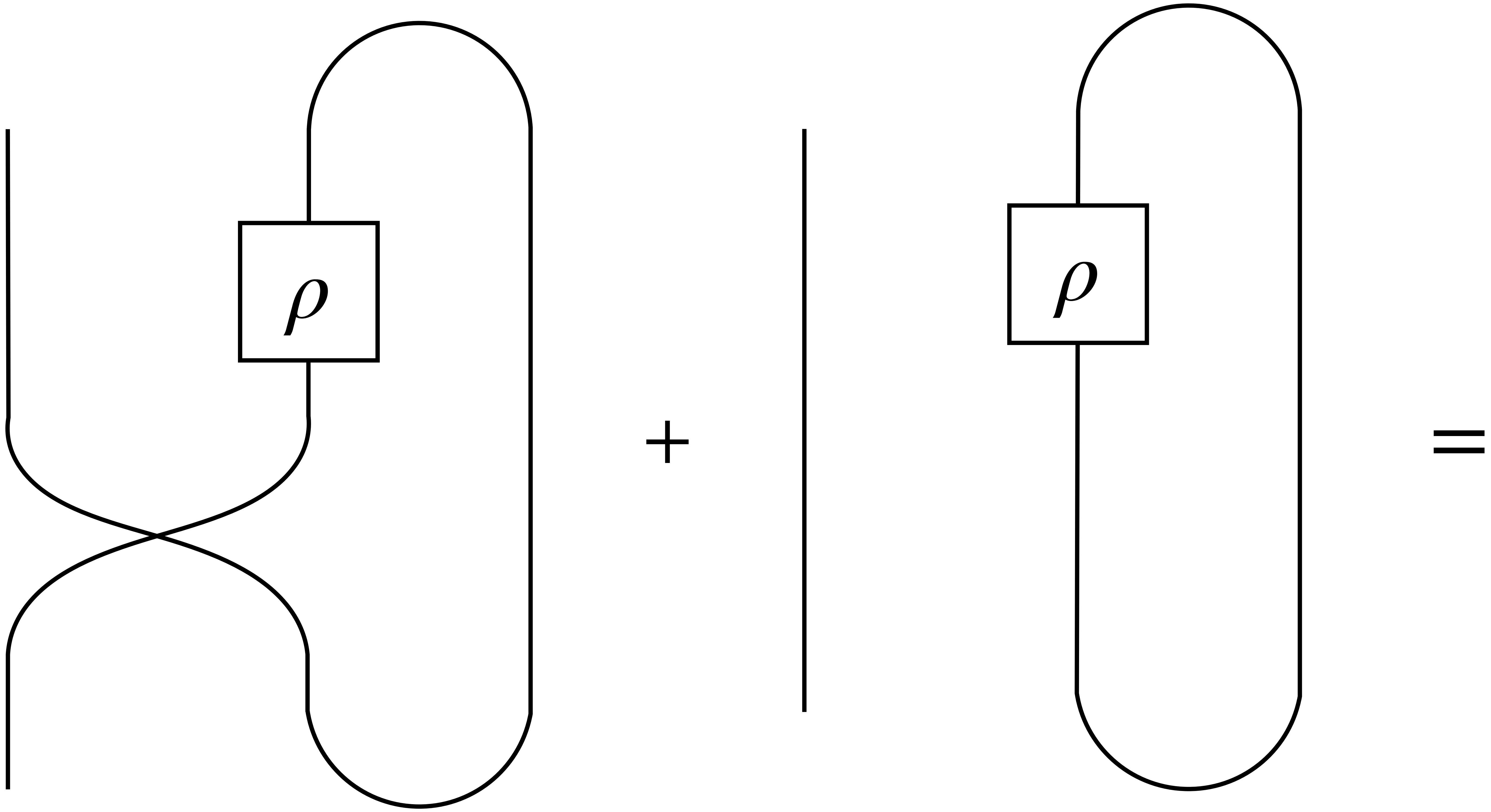


Schur's Lemma: This quantity is proportional to the sum over all permutations across the indices



Computing the expectation

$$\mathbb{E}[V] = d \int \mathbf{v}\mathbf{v}^{\dagger} \mathrm{Tr}(\mathbf{v}\mathbf{v}^{\dagger}\rho)\ \mathrm{d}\mu(\mathbf{v})$$



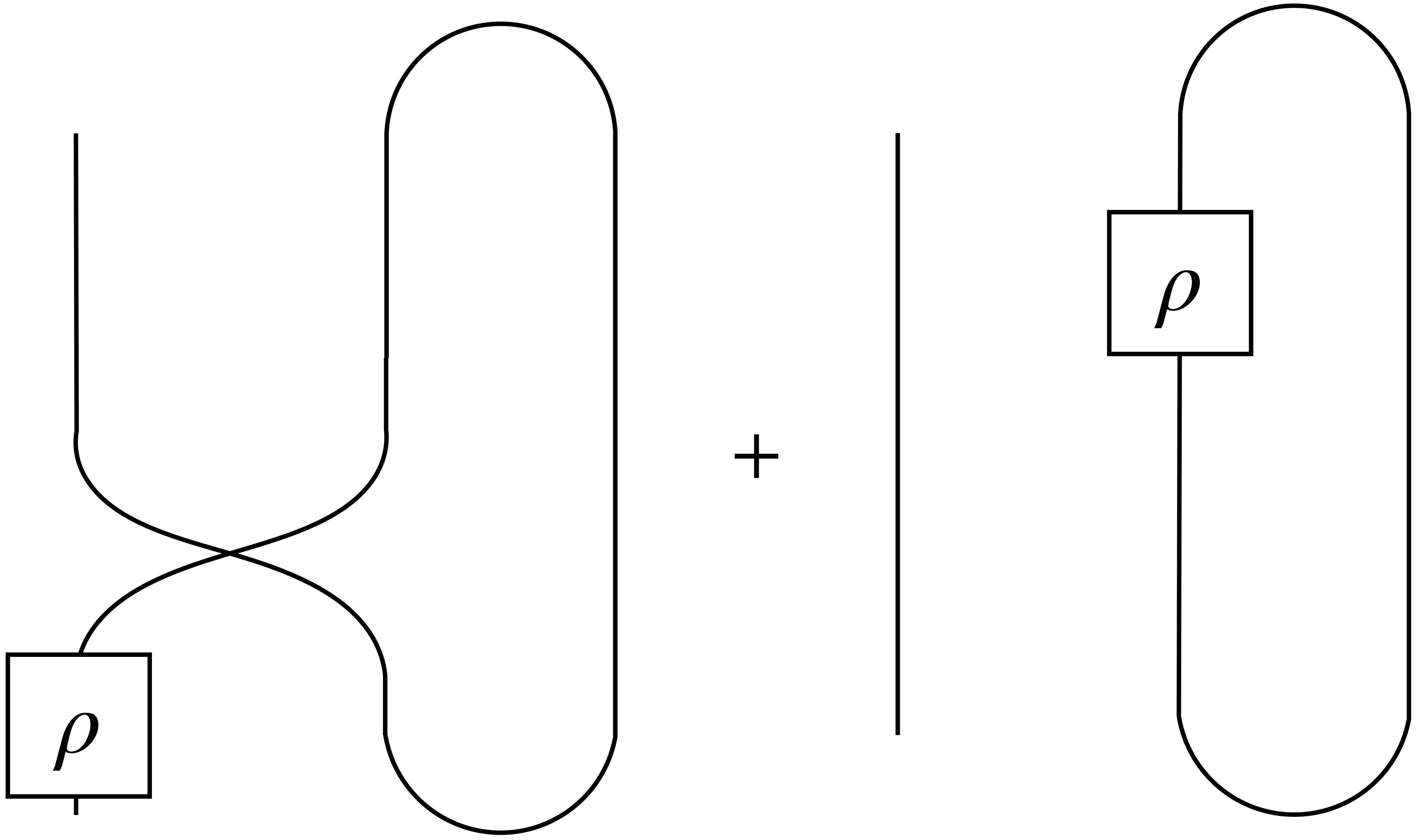
SWAP term

Identity term

Computing the expectation

$$\mathbb{E}[V] = d\int \mathbf{v}\mathbf{v}^\dagger \mathrm{Tr}(\mathbf{v}\mathbf{v}^\dagger \rho)\, \mathrm{d}\mu(\mathbf{v}) = d\frac{\rho + I}{d(d+1)} = \frac{\rho + I}{d+1}$$

Estimator:
$\hat{\rho} = (d+1)V - I$



SWAP term

Identity term

$$= \rho + \mathrm{Tr}(\rho)I = \rho + I$$

Overview of our proof strategy for pure states

1) Choose measurement operators $\left\{d\,\mathbf{v}\mathbf{v}^\dagger\ \mathrm{d}\mu(\mathbf{v})\right\}_{\mathbf{v}\in\mathbb{C}^d}$

2) From measurement result, estimate state $\hat{\rho} = (d+1)\mathbf{v}\mathbf{v}^\dagger - I$

Repeat n times $\quad \hat{\rho}_{\text{pairs}} = \dfrac{1}{n(n-1)}\sum_{i\neq j}\hat{\rho}_i\hat{\rho}_j$

3) Compute expectation $\quad \mathbb{E}[\hat{\rho}_i\hat{\rho}_j] = \mathbb{E}[\hat{\rho}_i]\mathbb{E}[\hat{\rho}_j] = \rho\rho = \rho$ ✓

4) Compute variance $\quad \mathrm{Var}[\mathrm{Tr}(H\hat{\rho}_{\text{pairs}})] \approx d\|H\|_F^2/n^2 + 1/n$

Computing the variance term by term

Estimator:

$$\hat{\rho}_{\text{pairs}} = \frac{1}{n(n-1)} \sum_{i \neq j} \hat{\rho}_i \hat{\rho}_j$$

$$\text{Var}[\text{Tr}(H\hat{\rho}_{\text{pairs}})] = \frac{1}{n^2(n-1)^2} \sum_{i \neq j} \sum_{k \neq \ell} \text{Cov}(\text{Tr}(H\hat{\rho}_i\hat{\rho}_j), \text{Tr}(H\hat{\rho}_k\hat{\rho}_\ell))$$

Case Analysis:

$\{i,j\} \cap \{k,\ell\} = \varnothing \implies \text{Tr}(H\hat{\rho}_i\hat{\rho}_j)$ and $\text{Tr}(H\hat{\rho}_k\hat{\rho}_\ell)$ are independent

$\implies \text{Cov}(\text{Tr}(H\hat{\rho}_i\hat{\rho}_j), \text{Tr}(H\hat{\rho}_k\hat{\rho}_\ell)) = 0$

Computing the variance term by term

Estimator:

$$\hat{\rho}_{\text{pairs}} = \frac{1}{n(n-1)} \sum_{i \neq j} \hat{\rho}_i \hat{\rho}_j$$

$$\text{Var}[\text{Tr}(H\hat{\rho}_{\text{pairs}})] = \frac{1}{n^2(n-1)^2} \sum_{i \neq j} \sum_{k \neq \ell} \text{Cov}(\text{Tr}(H\hat{\rho}_i\hat{\rho}_j), \text{Tr}(H\hat{\rho}_k\hat{\rho}_\ell))$$

Case Analysis:

One index matches ($|\{i,j\} \cap \{k,\ell\}| = 1$): $O(1)$

Both indices match ($|\{i,j\} \cap \{k,\ell\}| = 2$): $O(d\|H\|_F^2)$

$$\text{Var}[\text{Tr}(H\hat{\rho}_{\text{pairs}})] = O(d\|H\|_F^2/n^2 + 1/n)$$

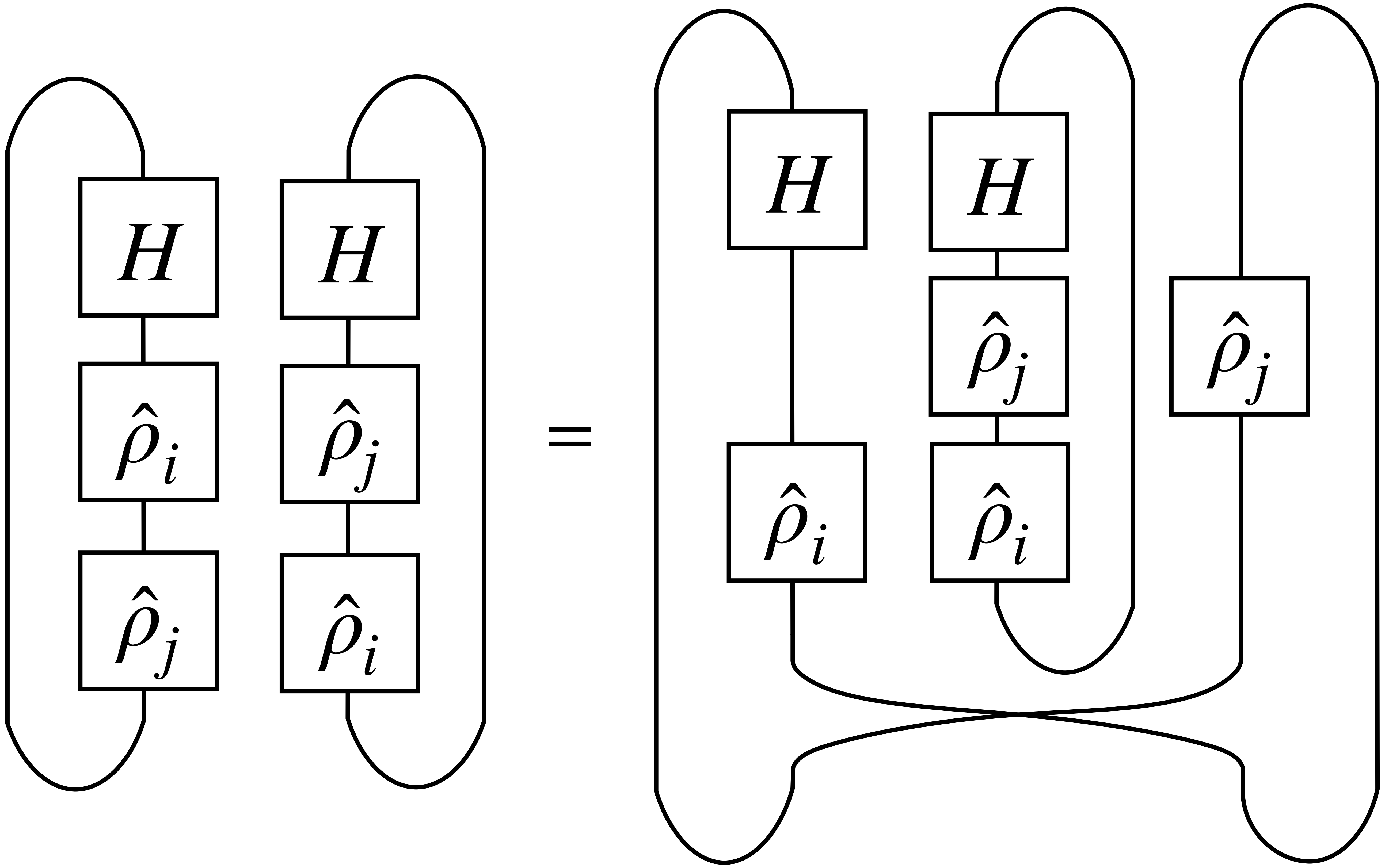
Isolating large covariance term

Big covariance term: $\mathrm{Cov}(\mathrm{Tr}(H\hat{\rho}_i\hat{\rho}_j), \mathrm{Tr}(H\hat{\rho}_i\hat{\rho}_j))$

$$= \mathbb{E}[\mathrm{Tr}(H\hat{\rho}_i\hat{\rho}_j)\overline{\mathrm{Tr}(H\hat{\rho}_i\hat{\rho}_j)}] - \mathbb{E}[\mathrm{Tr}(H\hat{\rho}_i\hat{\rho}_j)]\mathbb{E}[\overline{\mathrm{Tr}(H\hat{\rho}_i\hat{\rho}_j)}]$$

$$\leq \mathbb{E}[\mathrm{Tr}(H\hat{\rho}_i\hat{\rho}_j)\overline{\mathrm{Tr}(H\hat{\rho}_i\hat{\rho}_j)}]$$

$$= \mathbb{E}[\mathrm{Tr}(H\hat{\rho}_i\hat{\rho}_j)\mathrm{Tr}(H\hat{\rho}_j\hat{\rho}_i)]$$



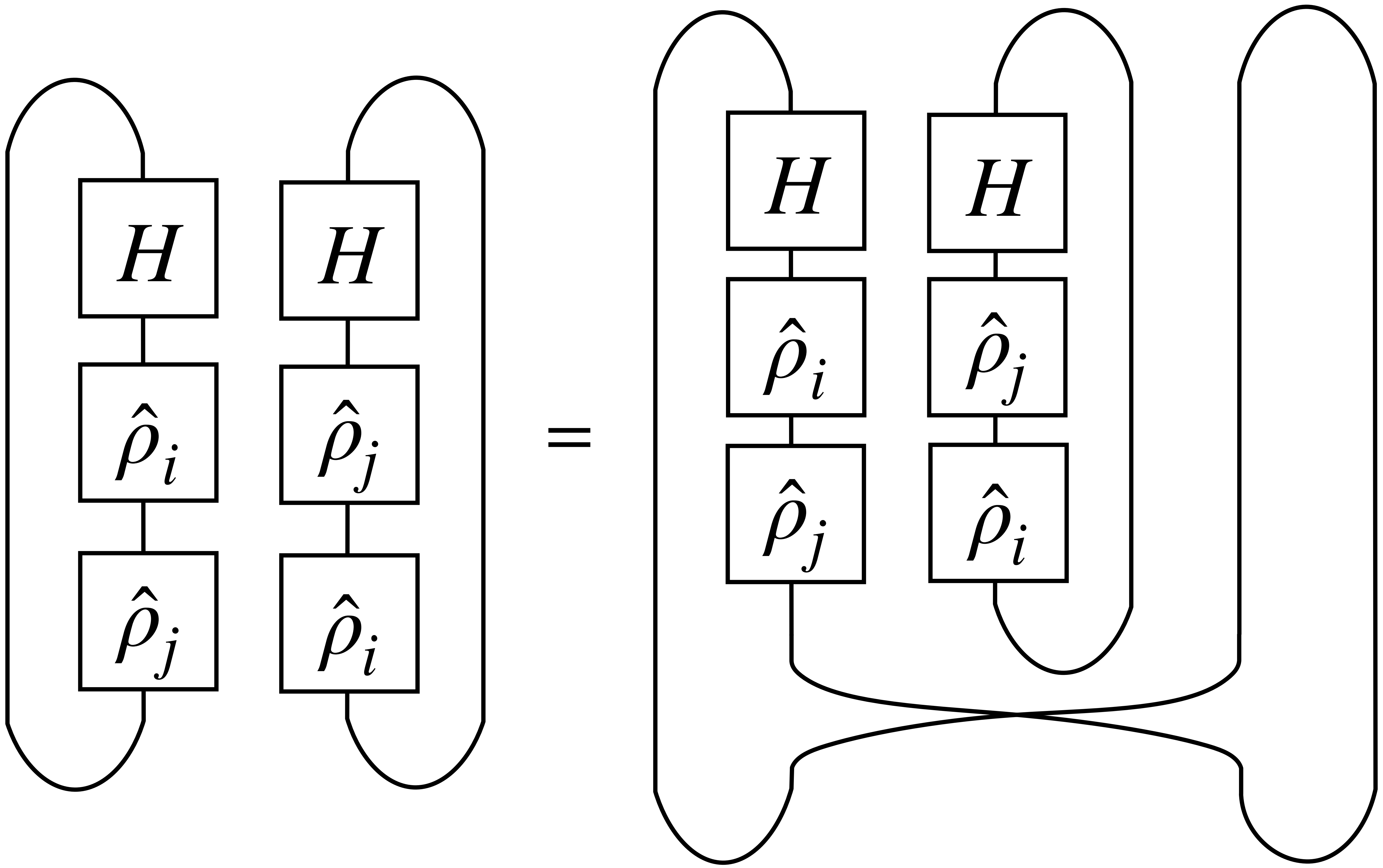
Isolating large covariance term

Big covariance term: $\mathrm{Cov}(\mathrm{Tr}(H\hat{\rho}_i\hat{\rho}_j), \mathrm{Tr}(H\hat{\rho}_i\hat{\rho}_j))$

$$= \mathbb{E}[\mathrm{Tr}(H\hat{\rho}_i\hat{\rho}_j)\overline{\mathrm{Tr}(H\hat{\rho}_i\hat{\rho}_j)}] - \mathbb{E}[\mathrm{Tr}(H\hat{\rho}_i\hat{\rho}_j)]\mathbb{E}[\overline{\mathrm{Tr}(H\hat{\rho}_i\hat{\rho}_j)}]$$

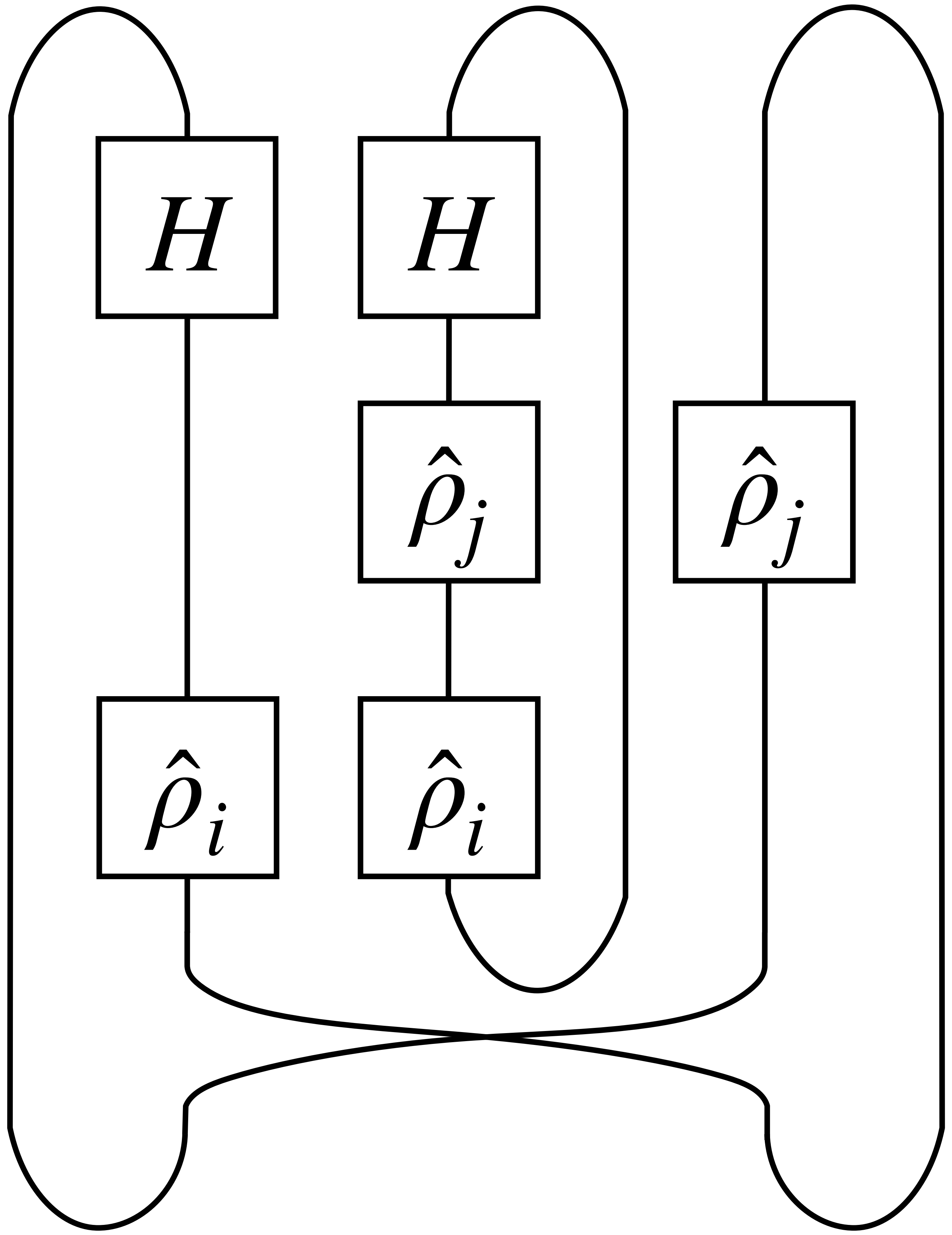
$$\leq \mathbb{E}[\mathrm{Tr}(H\hat{\rho}_i\hat{\rho}_j)\overline{\mathrm{Tr}(H\hat{\rho}_i\hat{\rho}_j)}]$$

$$= \mathbb{E}[\mathrm{Tr}(H\hat{\rho}_i\hat{\rho}_j)\mathrm{Tr}(H\hat{\rho}_j\hat{\rho}_i)]$$



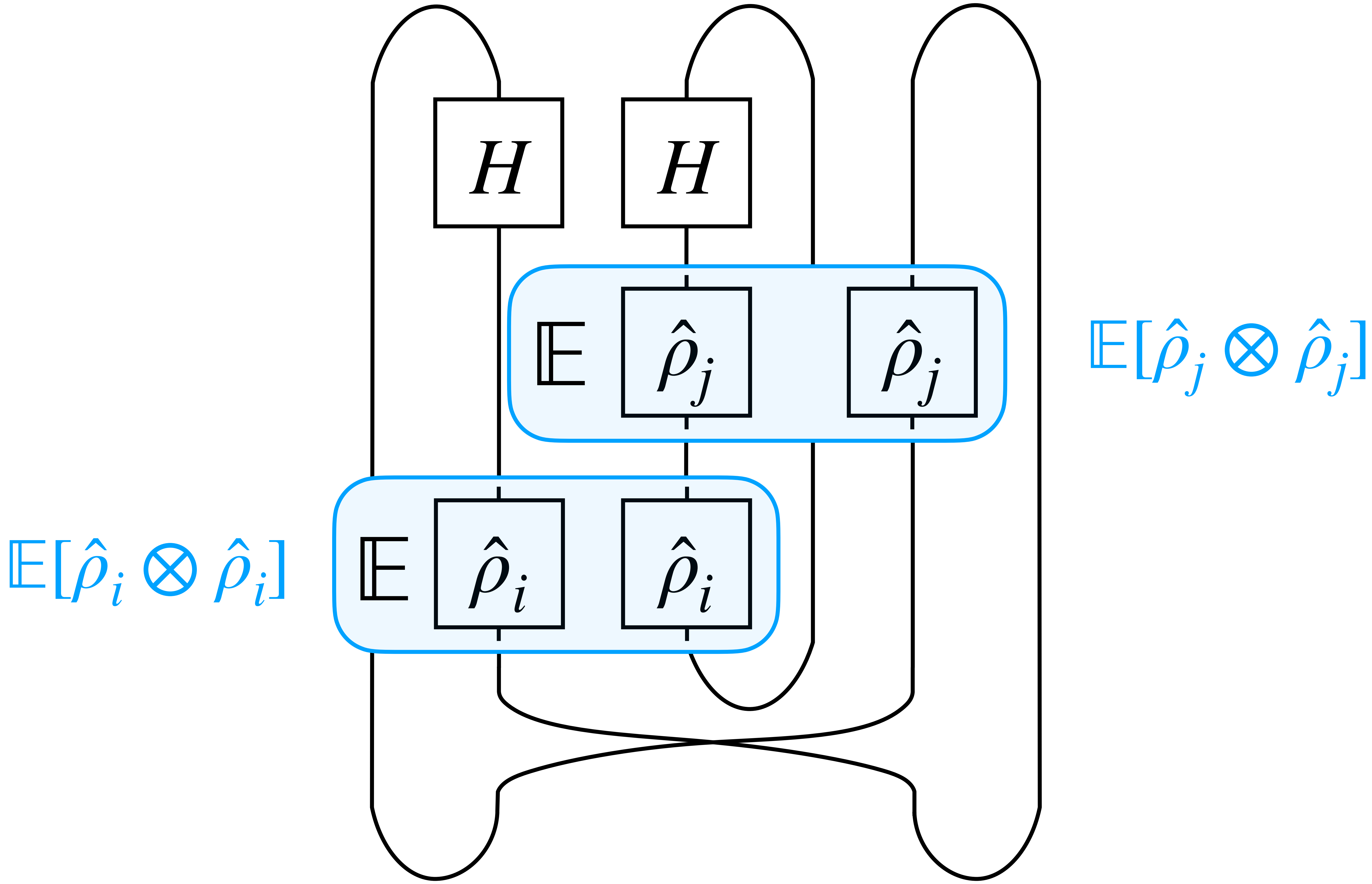
Covariance is broken into two expectations

$$\mathbb{E}[\mathrm{Tr}(H\hat{\rho}_i\hat{\rho}_j)\mathrm{Tr}(H\hat{\rho}_j\hat{\rho}_i)]$$



Covariance is broken into two expectations

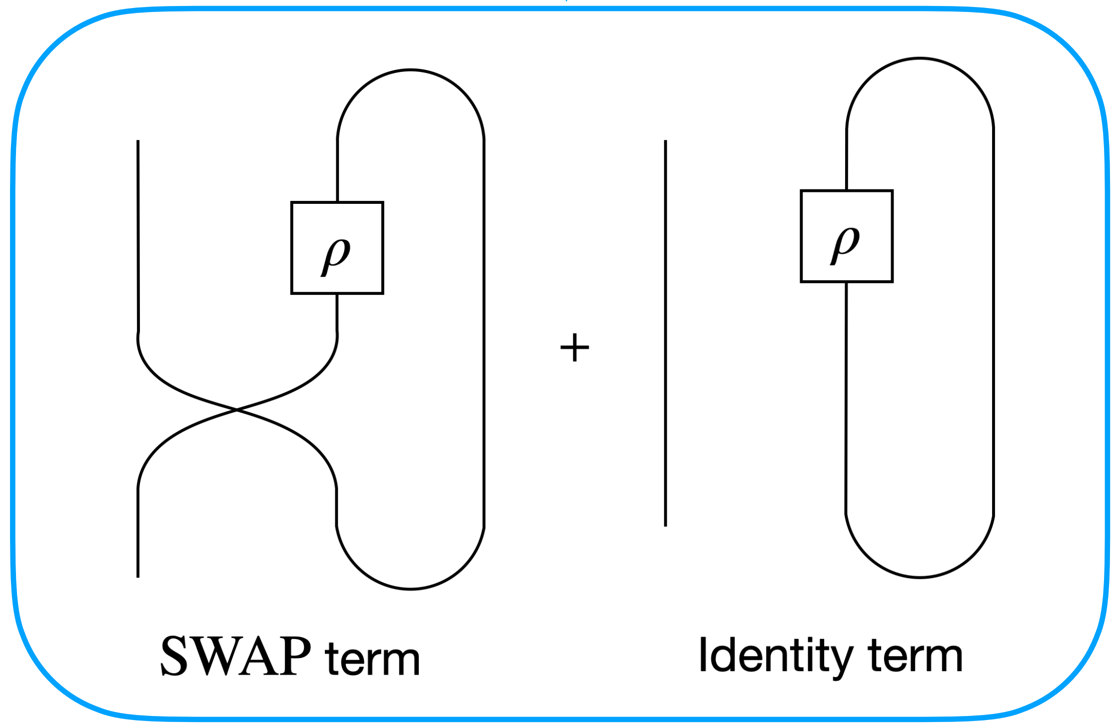
$$\mathbb{E}[\mathrm{Tr}(H\hat{\rho}_i\hat{\rho}_j)\mathrm{Tr}(H\hat{\rho}_j\hat{\rho}_i)]$$



Second moment calculation

Recall: $\hat{\rho} = (d+1)\mathbf{v}\mathbf{v}^\dagger - I$

- V is random variable that is $\mathbf{v}\mathbf{v}^\dagger$ with probability $d\mathrm{Tr}(\mathbf{v}\mathbf{v}^\dagger\rho)\ \mathrm{d}\mu(\mathbf{v})$
- $\mathbb{E}[\hat{\rho} \otimes \hat{\rho}] = (d+1)^2\mathbb{E}[V^{\otimes 2}] - (d+1)(\mathbb{E}[V] \otimes I + I \otimes \mathbb{E}[V]) + I \otimes I$



$$\frac{\rho + I}{d+1}$$

Second moment calculation

Recall: $\hat{\rho} = (d+1)\mathbf{v}\mathbf{v}^\dagger - I$

- V is random variable that is $\mathbf{v}\mathbf{v}^\dagger$ with probability $d\mathrm{Tr}(\mathbf{v}\mathbf{v}^\dagger\rho)\ \mathrm{d}\mu(\mathbf{v})$
- $\mathbb{E}[\hat{\rho} \otimes \hat{\rho}] = (d+1)^2\mathbb{E}[V^{\otimes 2}] - (d+1)(\mathbb{E}[V] \otimes I + I \otimes \mathbb{E}[V]) + I \otimes I$

$$= (d+1)^2\mathbb{E}[V^{\otimes 2}] - I \otimes I - \rho \otimes I - I \otimes \rho$$

Second moment calculation

Recall: $\hat{\rho} = (d+1)\mathbf{v}\mathbf{v}^\dagger - I$

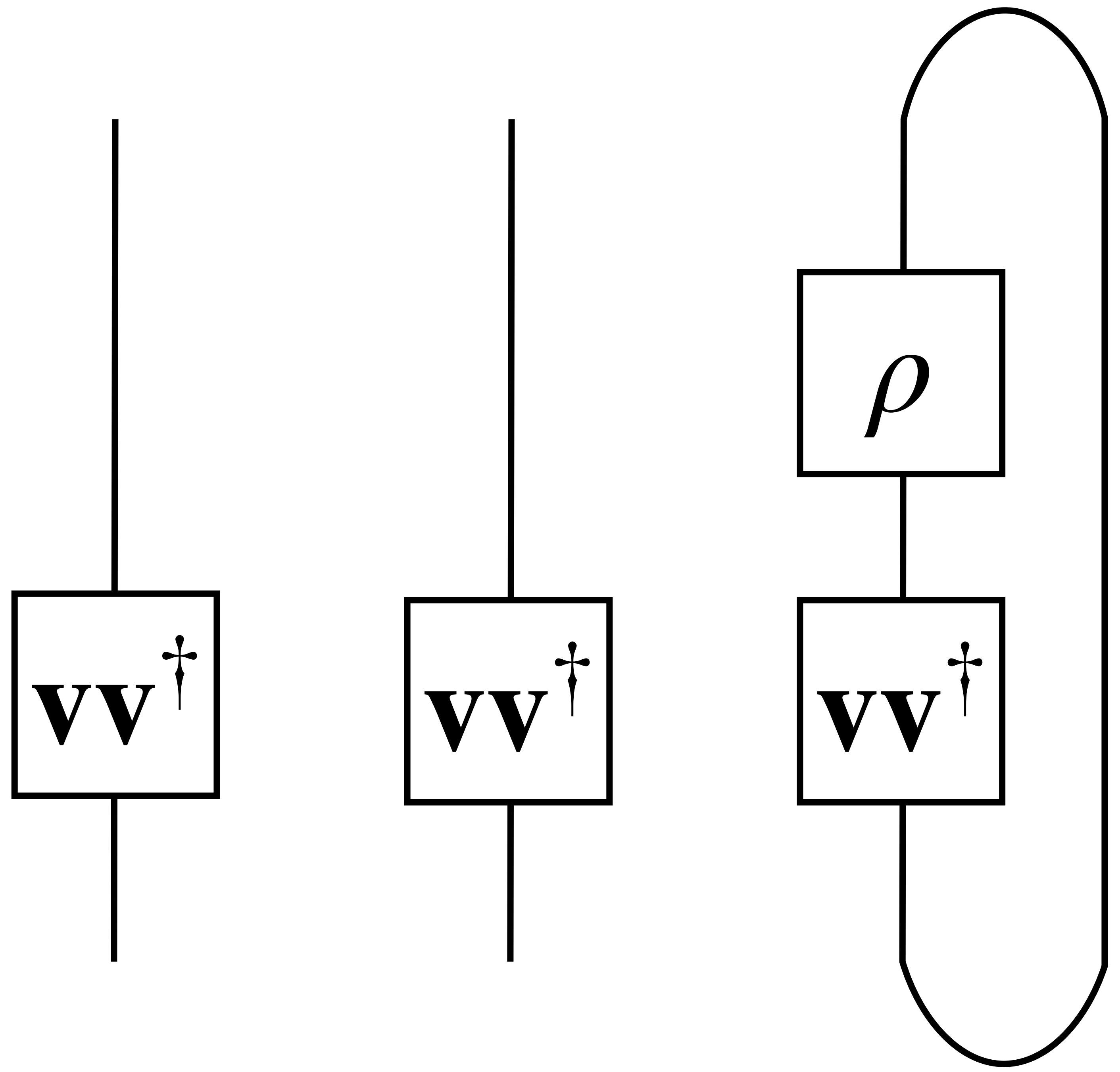
V is random variable that is $\mathbf{v}\mathbf{v}^\dagger$ with probability $d\mathrm{Tr}(\mathbf{v}\mathbf{v}^\dagger \rho)\ \mathrm{d}\mu(\mathbf{v})$

$$\mathbb{E}[\hat{\rho} \otimes \hat{\rho}] = (d+1)^2 \mathbb{E}[V^{\otimes 2}] - (d+1)(\mathbb{E}[V] \otimes I + I \otimes \mathbb{E}[V]) + I \otimes I$$

$$= (d+1)^2 \mathbb{E}[V^{\otimes 2}] - I \otimes I - \rho \otimes I - I \otimes \rho$$

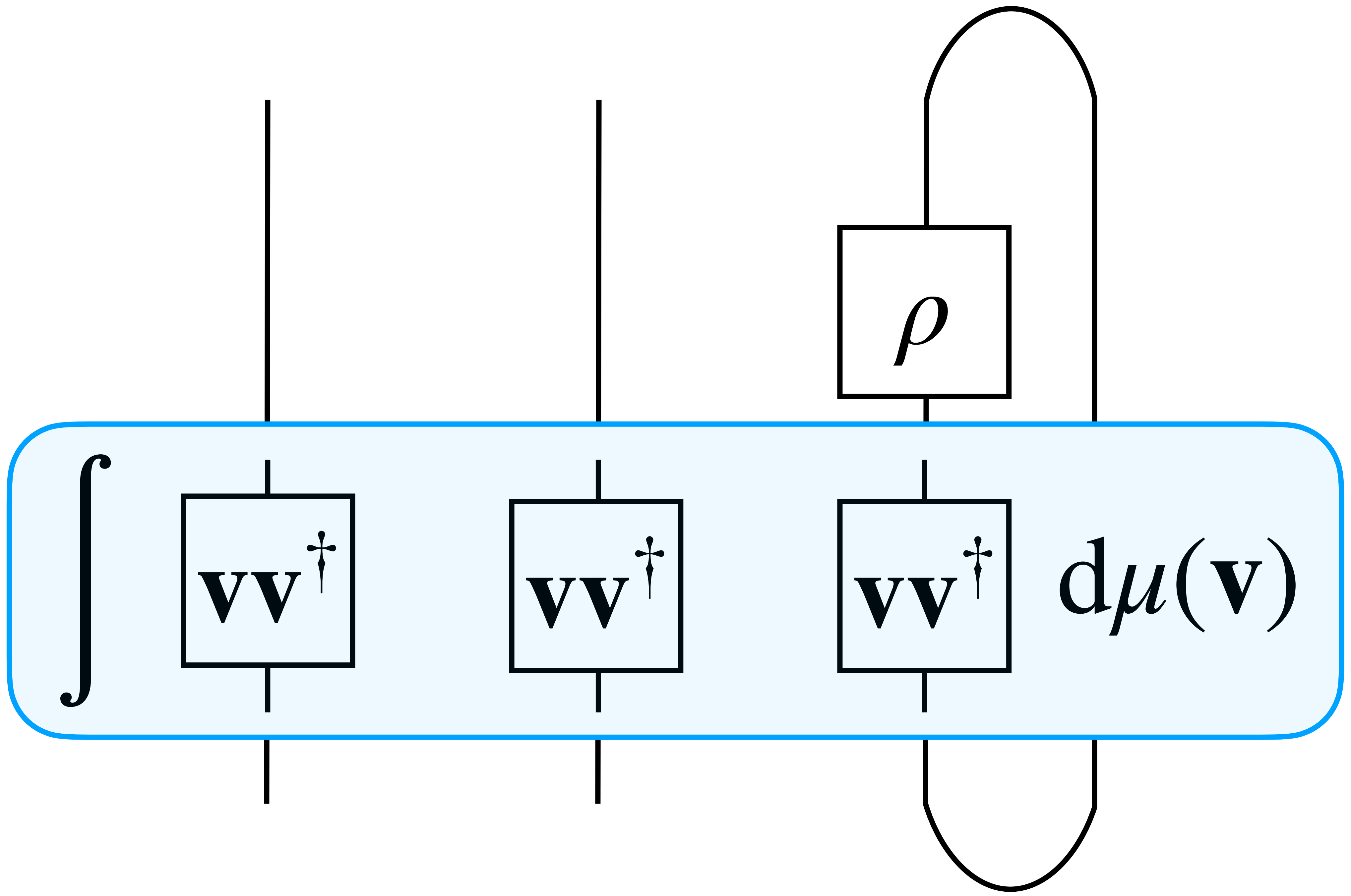
Second moment calculation

$$\mathbb{E}[V \otimes V] = d \int \mathbf{v}\mathbf{v}^\dagger \otimes \mathbf{v}\mathbf{v}^\dagger \mathrm{Tr}(\rho \mathbf{v}\mathbf{v}^\dagger)\, \mathrm{d}\mu(\mathbf{v})$$

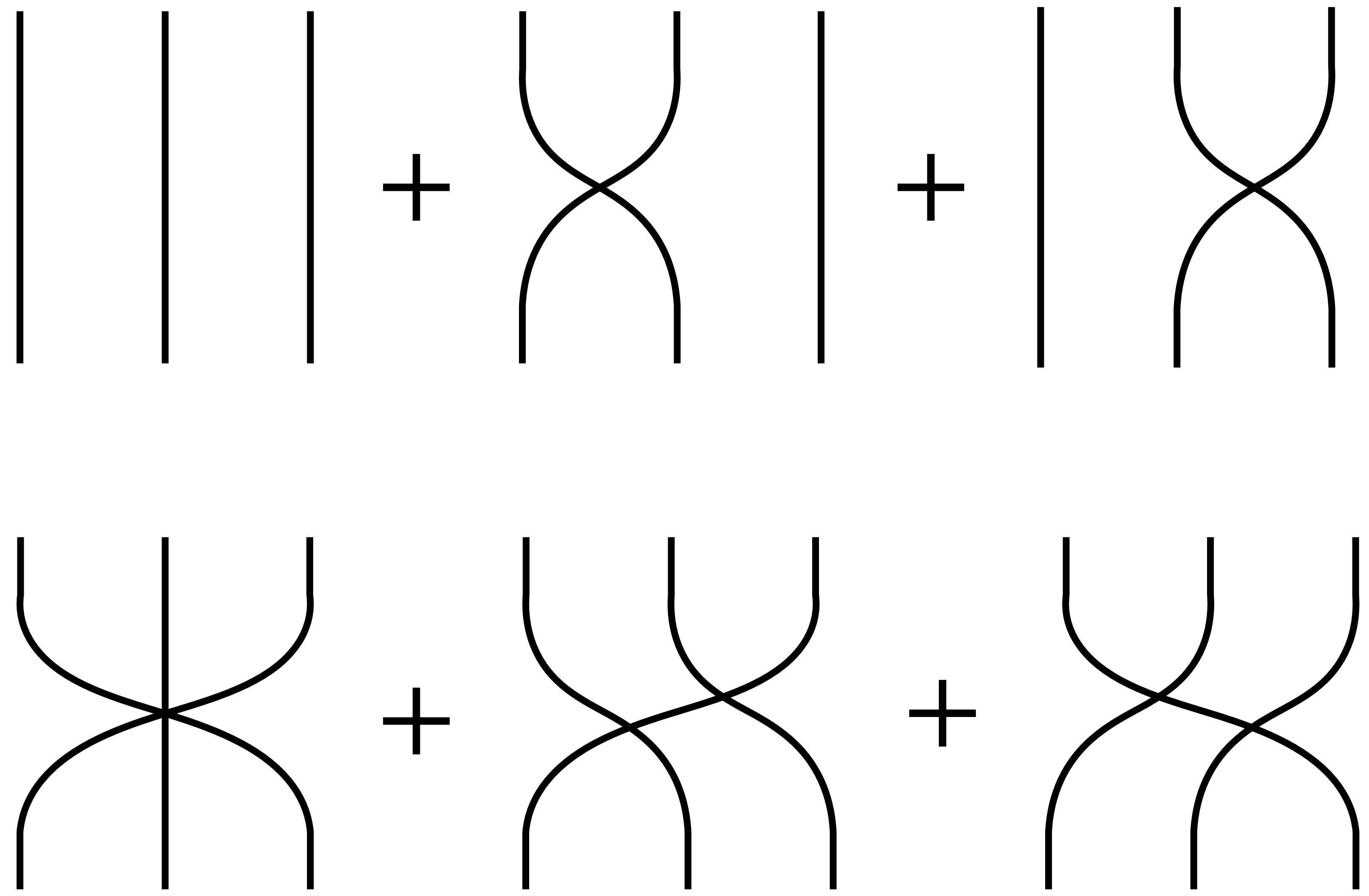


Second moment calculation

$$\mathbb{E}[V \otimes V] = d \int \mathbf{v}\mathbf{v}^\dagger \otimes \mathbf{v}\mathbf{v}^\dagger \mathrm{Tr}(\rho \mathbf{v}\mathbf{v}^\dagger)\, \mathrm{d}\mu(\mathbf{v}) = \frac{(I \otimes I + I \otimes \rho + \rho \otimes I)(I \otimes I + \mathrm{SWAP})}{(d+1)(d+2)}$$



Schur's Lemma:



Prefactor: $\frac{1}{d(d+1)(d+2)}$

Second moment calculation

Recall: $\hat{\rho} = (d+1)\mathbf{v}\mathbf{v}^\dagger - I$

- V is random variable that is $\mathbf{v}\mathbf{v}^\dagger$ with probability $d\,\mathrm{Tr}(\mathbf{v}\mathbf{v}^\dagger\rho)\ \mathrm{d}\mu(\mathbf{v})$
- $\mathbb{E}[\hat{\rho}\otimes\hat{\rho}] = (d+1)^2\mathbb{E}[V^{\otimes 2}] - (d+1)(\mathbb{E}[V]\otimes I + I\otimes\mathbb{E}[V]) + I\otimes I$

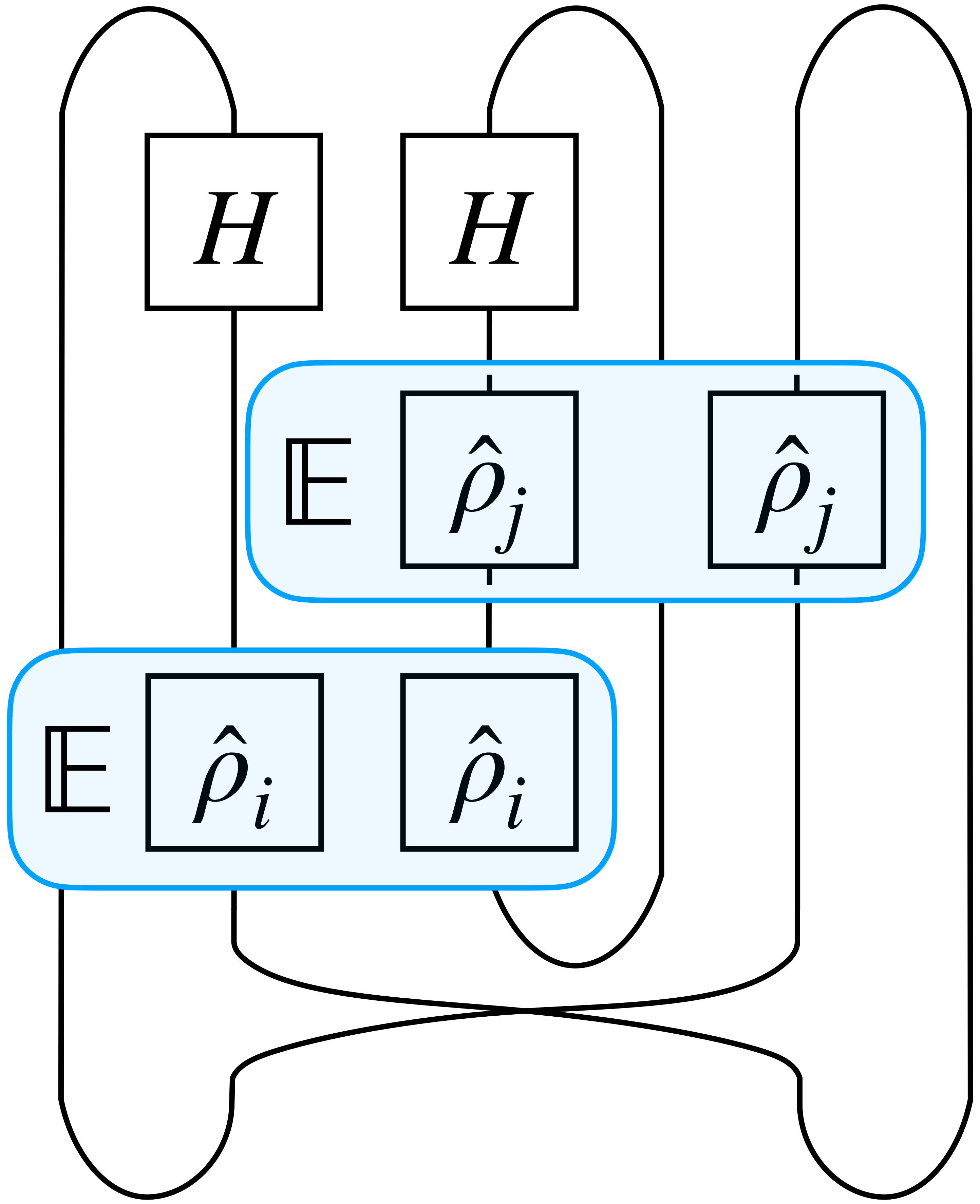
$$= (d+1)^2\mathbb{E}[V^{\otimes 2}] - I\otimes I - \rho\otimes I - I\otimes\rho$$

$$= (I\otimes I + \rho\otimes I + I\otimes\rho)\left(\cancelto{1}{\frac{d+1}{d+2}}\mathrm{SWAP} - \cancelto{0}{\frac{1}{d+2}}I\otimes I\right)$$

$$\approx (I\otimes I + \rho\otimes I + I\otimes\rho)\mathrm{SWAP}$$

Covariance is broken into two expectations

$$\mathbb{E}[\mathrm{Tr}(H\hat{\rho}_i\hat{\rho}_j)\mathrm{Tr}(H\hat{\rho}_j\hat{\rho}_i)]$$

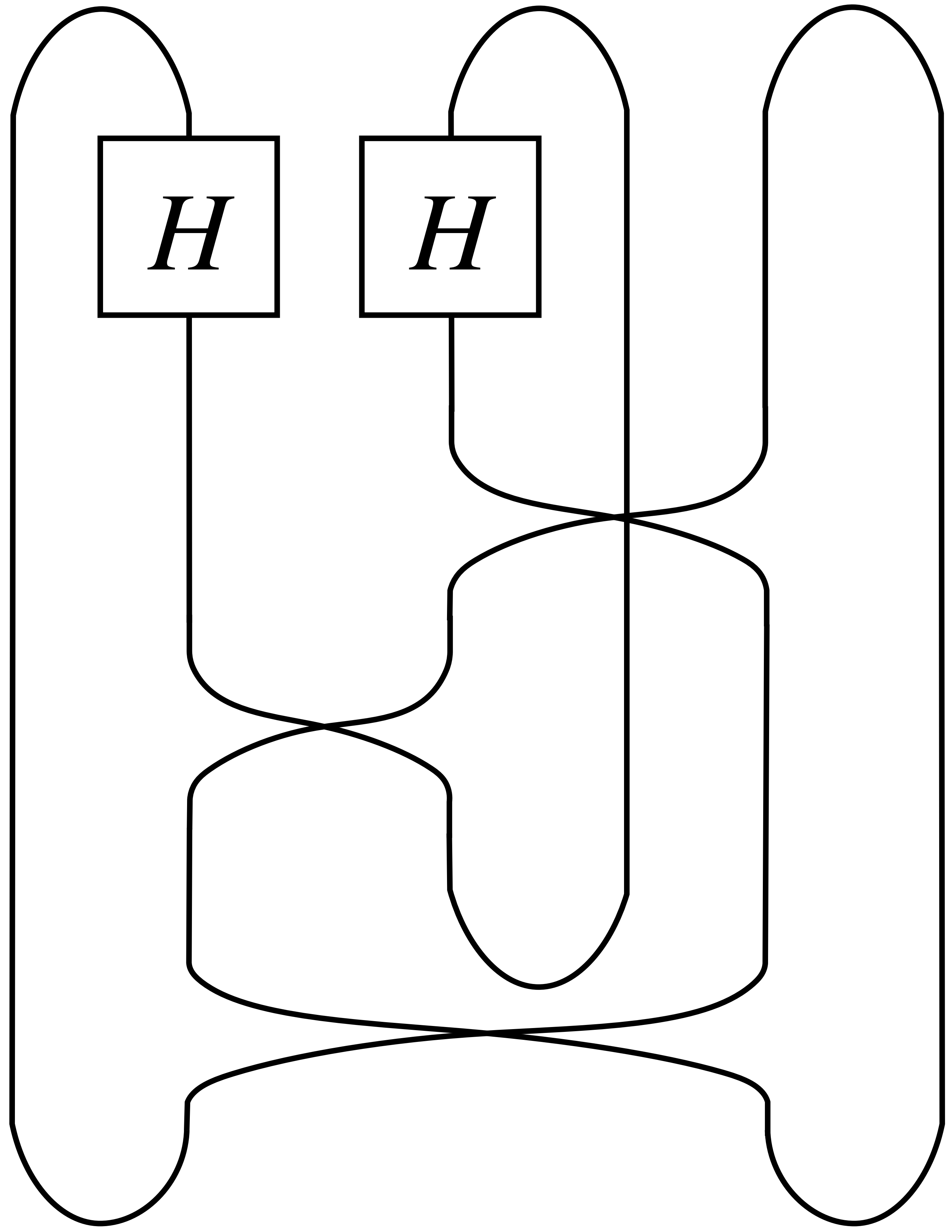


$(I \otimes I + \rho \otimes I + I \otimes \rho)\mathrm{SWAP}$

$(I \otimes I + \rho \otimes I + I \otimes \rho)\mathrm{SWAP}$

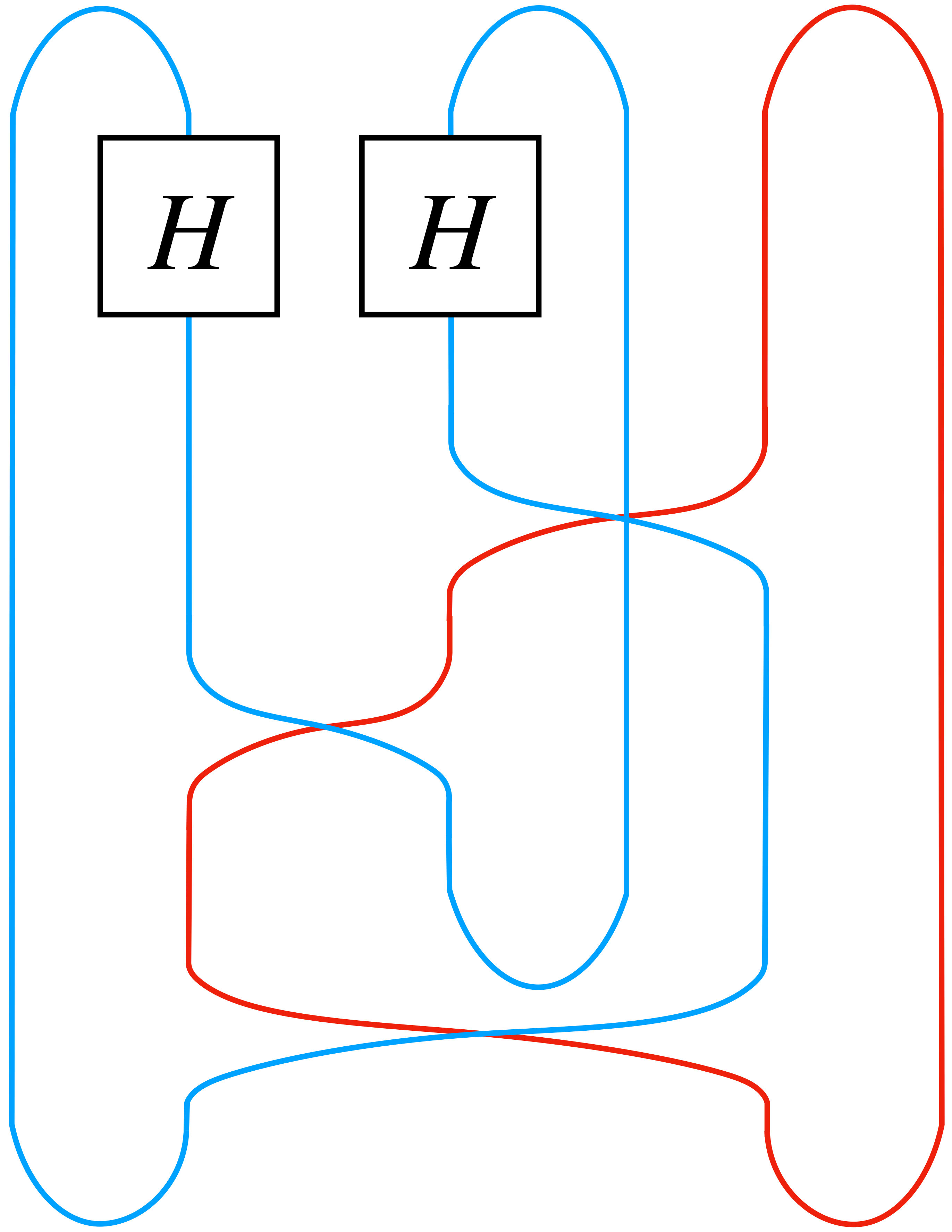
Covariance is broken into two expectations

$$\mathbb{E}[\mathrm{Tr}(H\hat{\rho}_i\hat{\rho}_j)\mathrm{Tr}(H\hat{\rho}_j\hat{\rho}_i)]$$



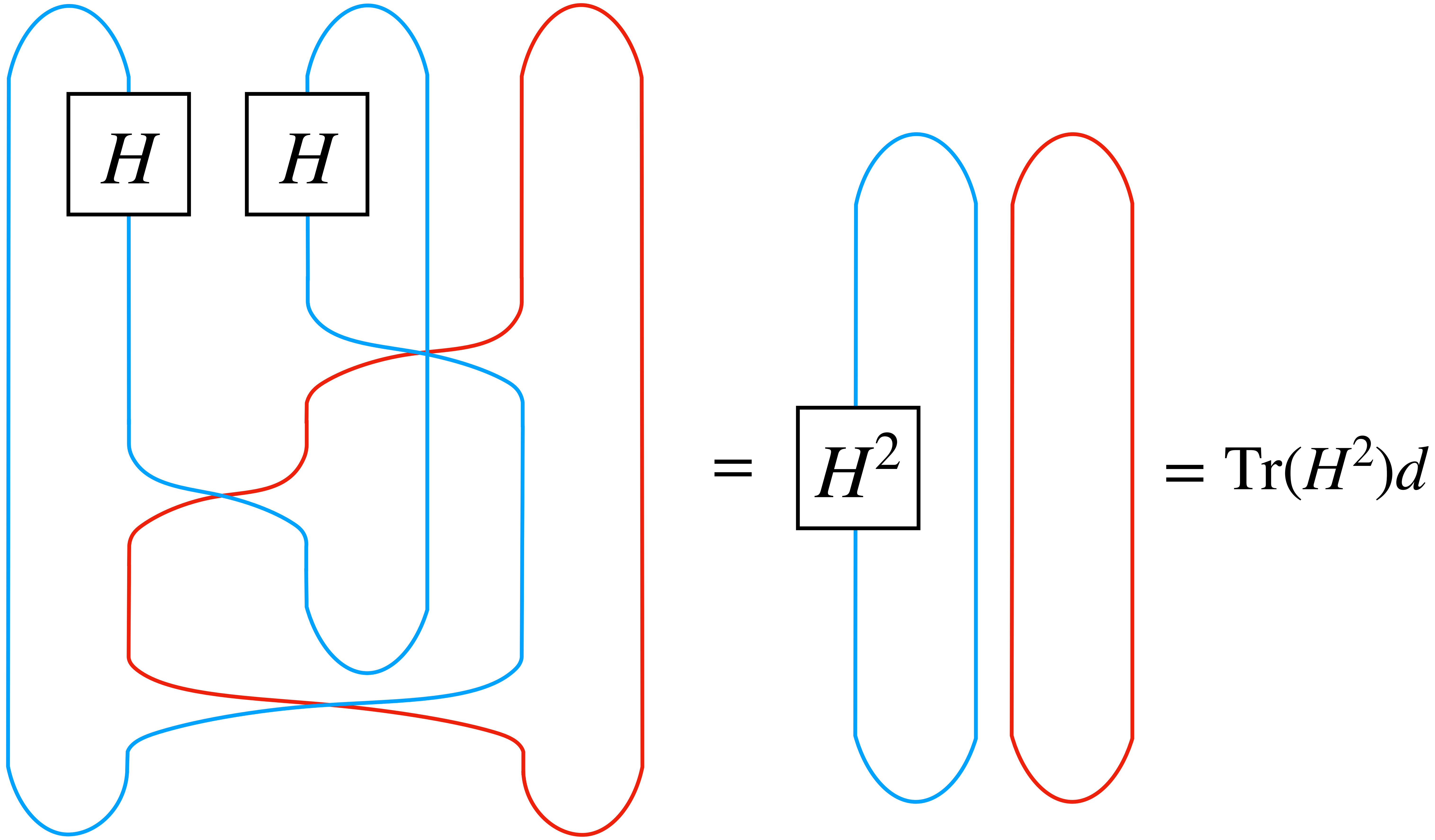
Covariance is broken into two expectations

$$\mathbb{E}[\mathrm{Tr}(H\hat{\rho}_i\hat{\rho}_j)\mathrm{Tr}(H\hat{\rho}_j\hat{\rho}_i)]$$



Covariance is broken into two expectations

$$\mathbb{E}[\mathrm{Tr}(H\hat{\rho}_i\hat{\rho}_j)\mathrm{Tr}(H\hat{\rho}_j\hat{\rho}_i)]$$



Computing the variance term by term

Estimator: $$\hat{\rho}_{\text{pairs}} = \frac{1}{n(n-1)} \sum_{i \neq j} \hat{\rho}_i \hat{\rho}_j$$

$$\text{Var}[\text{Tr}(H\hat{\rho}_{\text{pairs}})] = \frac{1}{n^2(n-1)^2} \sum_{i \neq j} \sum_{k \neq \ell} \text{Cov}(\text{Tr}(H\hat{\rho}_i\hat{\rho}_j), \text{Tr}(H\hat{\rho}_k\hat{\rho}_\ell))$$

Case Analysis:

Both indices match ($|\{i,j\} \cap \{k,\ell\}| = 2$):

$$\text{Cov}(\text{Tr}(H\hat{\rho}_i\hat{\rho}_j), \text{Tr}(H\hat{\rho}_i\hat{\rho}_j)) = O(d\|H\|_F^2)$$

$$\text{Var}[\text{Tr}(H\hat{\rho}_{\text{pairs}})] = O(d\|H\|_F^2/n^2 + 1/n)$$

Open questions

1) What's the right answer for the independent measurement setting?

2) Can we get a smooth scaling with the rank of the state?

3) What happens when the unknown states are not identical?