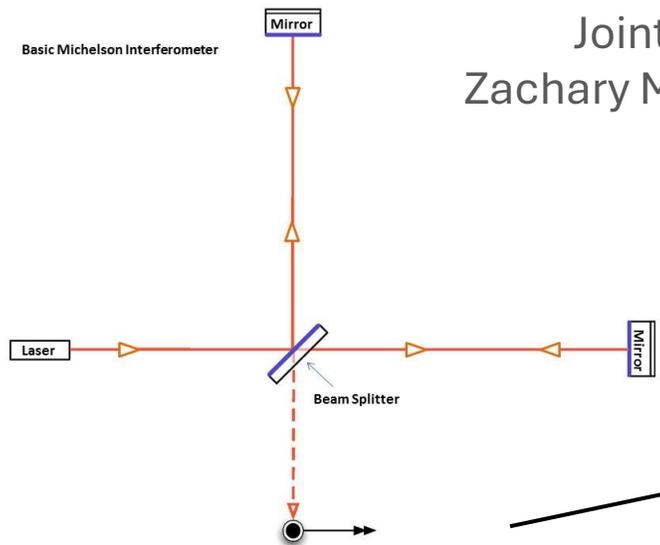


Towards sample-optimal learning of bosonic Gaussian quantum states

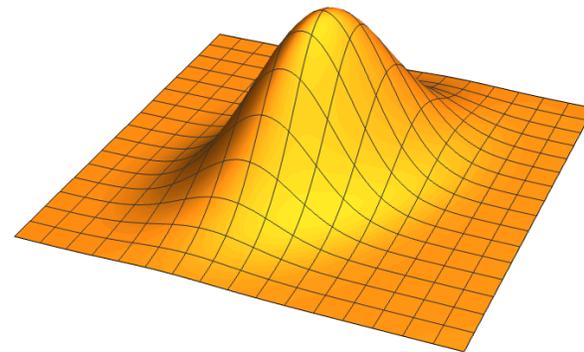
Senrui Chen (IQIM Postdoc)

IQIM Seminar, Feb 27, 2026

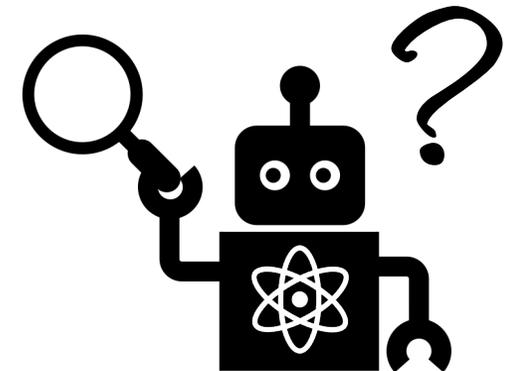
Joint work with Francesco Anna Mele, Marco Fanizza, Alfred Li, Zachary Mann, Hsin-Yuan (Robert) Huang, Yanbei Chen, and John Preskill
[\[arXiv 2603.18136\]](#)



[Credit: Caltech/MIT/LIGO Lab]



$\rho(\mu, \Sigma)$



Bosonic Quantum Information

- Quantum information processing with continuous-variable systems

[Caltech/MIT/LIGO Lab]



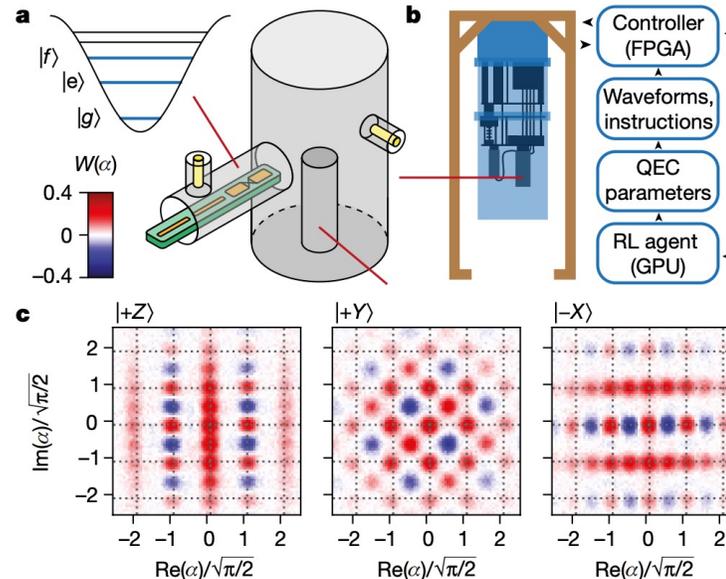
Quantum Sensing and Metrology

Gravitational wave (GW) detection

Dark matter search

Quantum imaging

[Sivak et al. Nature 616, 50–55 (2023)]

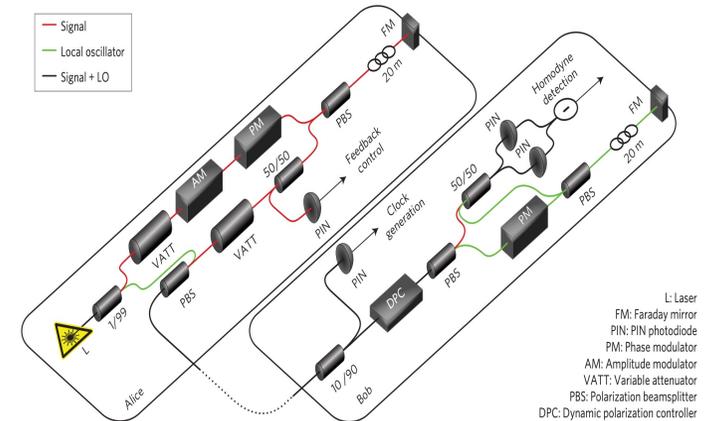


Quantum Computation

Bosonic quantum error correction

Boson Sampling

[Jouguet et al. Nat. Photon. 7, 378–381 (2013)]

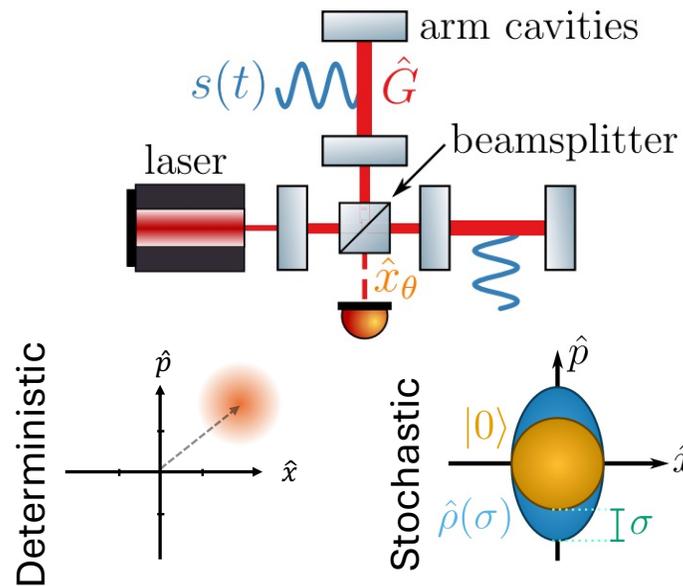
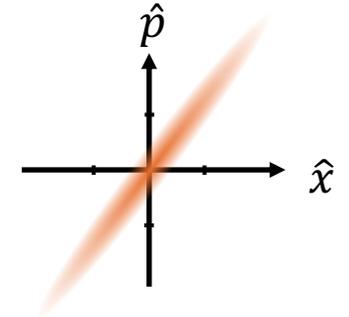


Quantum Communication

Continuous-Variable QKD

Gaussian states

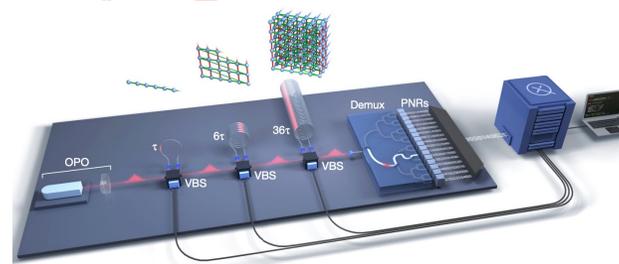
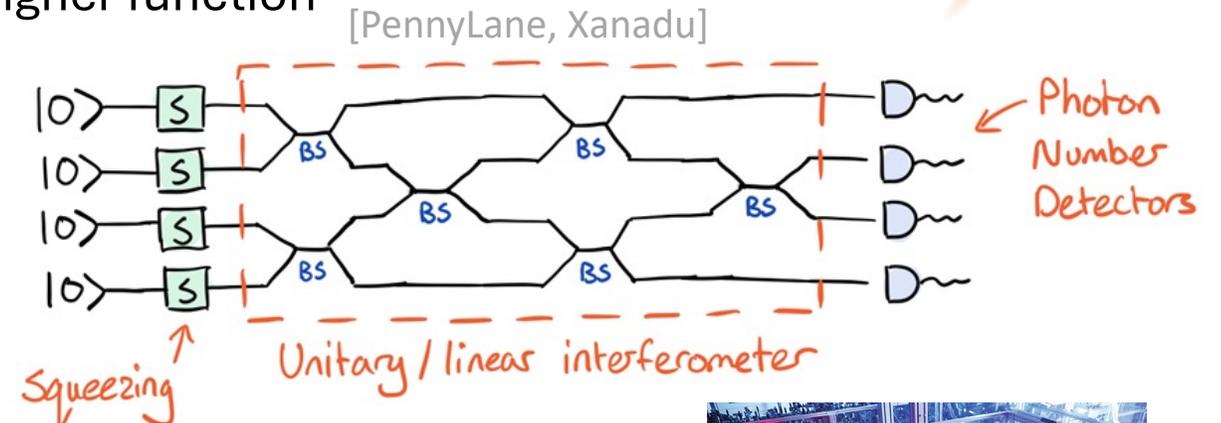
- Gaussian states – a natural class of bosonic quantum states
 - Gibbs states of quadratic Hamiltonian in quadrature operators (\hat{x}_i, \hat{p}_i)
 - Equivalently, States with a Gaussian Wigner function



[J Gardner et al. PRL **132**, 130801 (2024)]

[J Gardner et al. PRXQ **6**, 030311 (2025)]

Gravitational wave detection in LIGO



[Madsen et al. Nature **606**, 75–81 (2022)]



[Zhong et al. Science **370**, 6523 (2020)]

Gaussian Boson Sampling

Learning Gaussian states

- Learning Gaussian states is important for sensing, benchmarking, etc.
- How efficiently can we learn them?
- **Formally:** How many copies are needed to learn an unknown n -mode Gaussian state ρ , with energy no more than E , to trace distance ε with probability at least $2/3$?

$$p_{\text{succ}} = (1 + D_{\text{tr}})/2$$

- Trace distance: $D_{\text{tr}}(\rho, \sigma) = \frac{1}{2} \text{Tr}|\rho - \sigma|$, captures how good two states can be distinguished
- For sensing broadband or time-varying signals, or benchmarking large quantum devices, the scaling with **mode number** n is important
- We characterize energy E by the largest eigenvalues of ρ 's covariance matrix.
- **State-of-the-art:** $N = O(n^3 \varepsilon^{-2} + n \text{polyloglog} E)$ is sufficient^[1]
 - No meaningful lower bound is known.

Summary of Results

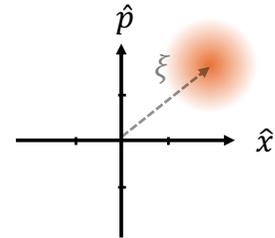
Task: Learn an n -mode $\rho(\mu, \Sigma)$ to ε trace distance with probability at least $2/3$. Promise: $\|\Sigma\|_{\text{op}} \leq \bar{E}$.

Class of states	Class of measurements	Bounds	
All Gaussian	Classical (including Gaussian)	$\tilde{O}(n^3/\varepsilon^2)$ [BMEM25]	$\Omega(n^3/\varepsilon^2)$ Thm. 2
	Any		$\Omega(n^2/\varepsilon^2)$ Thm. 3
Pure Gaussian	Any or Gaussian	$\tilde{\Theta}(n^2/\varepsilon^2)$ Thm. 3, 5	
Passive Gaussian	Any	$\tilde{O}(n^2/\varepsilon^2)$ Thm. 6	} Non-Gaussian Advantages
	Classical (including Gaussian)	$\Theta(n^3/\varepsilon^2)$ Thm. 2, 6	
1-mode Gaussian	Non-adaptive 1-copy Gaussian	$\tilde{\Theta}(\bar{E}/\varepsilon^2)$ Thm. 7	

Preliminaries

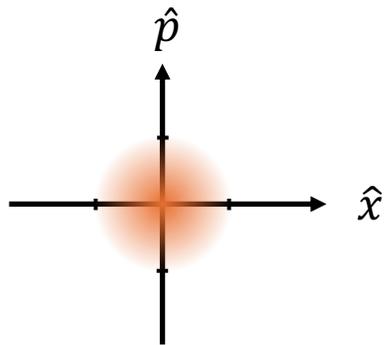
Basics of bosonic quantum information

- Hilbert space for n bosonic modes : $\mathcal{H}^n \cong L^2(\mathbb{R}^n; \mathbb{C})$
- Quadrature operators: $\hat{R} = [\hat{x}_1, \dots, \hat{x}_n, \hat{p}_1, \dots, \hat{p}_n]$
 - Canonical Commutation Relation (CCR): $[\hat{R}_k, \hat{R}_l] = i\Omega_{kl}$, where $\Omega = \begin{pmatrix} 0_n & -I_n \\ I_n & 0_n \end{pmatrix}$
- Mean μ and covariance Σ of a bosonic state ρ :
 - $\mu_i := \text{Tr}(\hat{R}_i \rho)$. $\Sigma_{kl} := \frac{1}{2} \text{Tr}(\{\hat{R}_k, \hat{R}_l\} \rho) - \mu_k \mu_l$
- Displacement operator: $\hat{D}_\xi := \exp(i\xi^T \Omega \hat{R})$, where $\xi \in \mathbb{R}^{2n}$
- **Wigner function:** $W_{\hat{O}}(r) := \frac{1}{(2\pi)^{2n}} \int d^{2n} \xi \exp(i\xi^T \Omega r) \text{Tr}(\hat{D}_\xi \hat{O})$, $r \in \mathbb{R}^{2n}$
 - For a quantum state ρ , $\int d^{2n} r W_\rho(r) = 1$, called **Wigner quasi-distribution**.
 - Inner product formula: $\text{Tr}(\rho \hat{O}) = (2\pi)^n \int d^{2n} r W_\rho(r) W_{\hat{O}}(r)$.

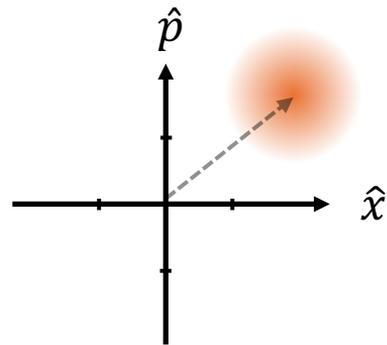


Basics of Gaussian states

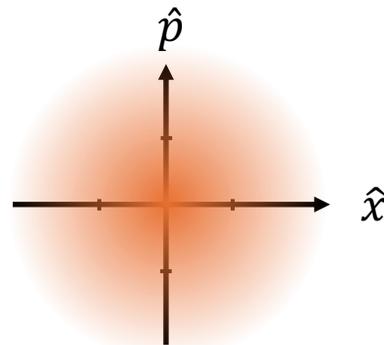
- **Def.** Gaussian states are states with Gaussian Wigner distribution $W_\rho = \mathcal{N}(\mu, \Sigma)$
 - Denoted by $\rho(\mu, \Sigma)$. Note the quantum and classical moments coincide.
- **Theorem** (Uncertainty relation)^[1]: $\rho(\mu, \Sigma)$ is a valid Gaussian state iff $\Sigma + i\Omega/2 \geq 0$
 - In 1-mode case, equivalent to $\lambda_1 \lambda_2 \geq \frac{1}{4}$ where λ 's are eigenvalues of Σ
- Examples of 1-mode Gaussian states (in Wigner distribution):



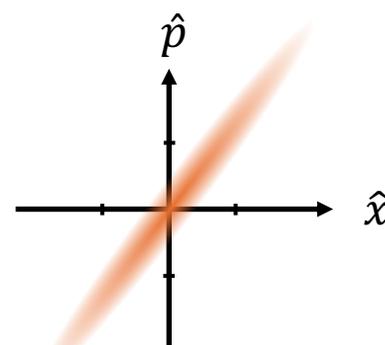
Vacuum state $|0\rangle$



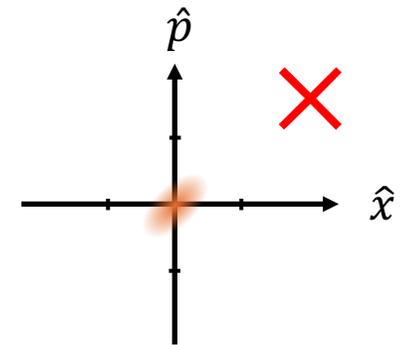
Coherent state



Thermal state



Squeezed state



NOT a valid state!!

[1] Originally by Robertson and Schrödinger. See Weedbrook et al., Rev. Mod. Phys. **84**, 621 (2012)

Measurements of bosonic systems

- Measurements are POVMs: $\{E_i\}$ where $E_i \geq 0$ and $\sum_i E_i = \mathbb{I}$. $\Pr(i|\rho) = \text{Tr}(E_i\rho)$

1. Classical meas.

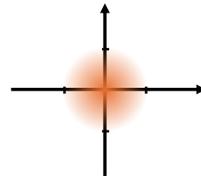
All POVM elements E_i have **non-negative** Wigner functions

- The most familiar instance is **Gaussian** measurements: $\left\{E_\xi := \frac{1}{(2\pi)^n} D_\xi \rho(0, V) D_\xi^\dagger\right\}$

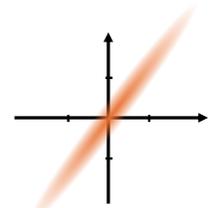
where $\rho(0, V)$ is a Gaussian state called the measurement **seed**. Needs $V + i\Omega/2 \geq 0$.

- Meas seed V on $\rho(\mu, \Sigma)$ yields $\xi \sim \mathcal{N}(\mu, \Sigma + V)$, i.e., blurred Wigner samples

- Examples: Heterodyne (vacuum seed)

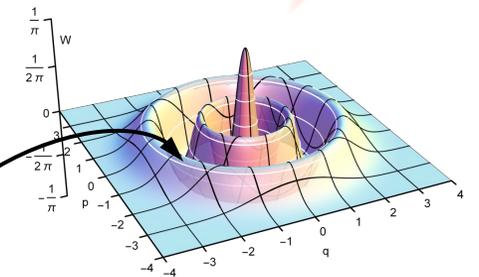


- Homodyne (squeezed seed)



2. Non-classical meas.

- Typical example: **photon number counting**: $\{E_n := |n\rangle\langle n|\}$
- Projection to the **Fock basis**, eigen basis of $\hat{x}_i^2 + \hat{p}_i^2$ at each mode
- Non-vacuum Fock states' Wigner functions contain negative values

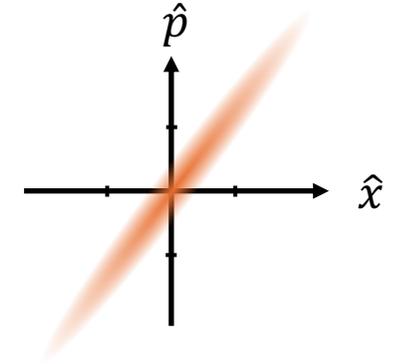


[Wigner function of $|4\rangle$, Wikipedia]

- Will Non-classical meas. help in learning Gaussian states? --- Yes, as we will see.

Learning Gaussian Wigner Distributions

Learning Gaussian Wigner distributions



- **Warm-up question:** How many copies are needed to learn the Wigner distribution of a Gaussian states to ε TV distance?
 - $\text{TV}(p, q) := \sum_x |p(x) - q(x)|/2.$ $\text{Pr}(\text{Succ in distinguishing}) = (1 + \text{TV})/2$
- **Known**^[1]: Learning classical $2n$ -dim Gaussians to ε -TVD requires $\Theta(n^2 \varepsilon^{-2})$ samples
- **Thm 1.** $N = \tilde{\Theta}(n^2 \varepsilon^{-2})$ are necessary and sufficient to learn Gaussian Wigner
 - Furthermore, this can be achieved using only **Gaussian measurements**.
 - Hide a doubly-logarithmic factor depending on energy (see next slide)

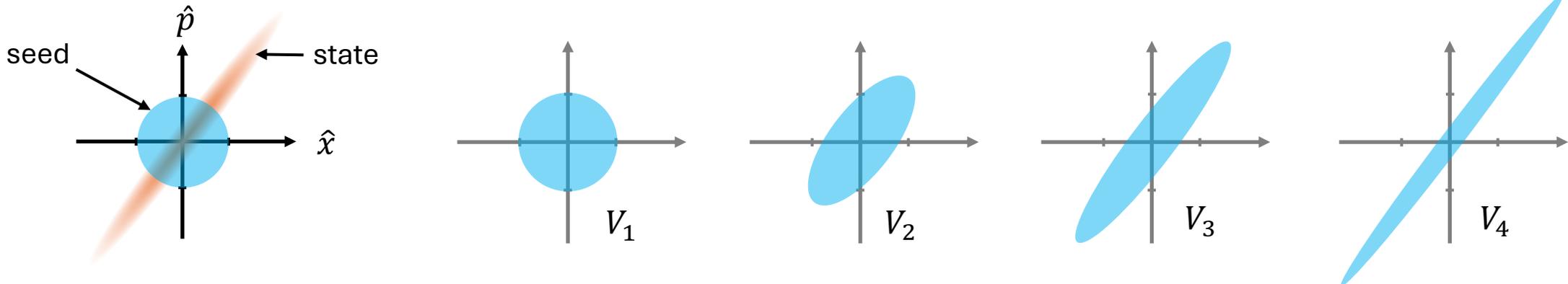
Learning Gaussian Wigner – upper bound

- For classical Gaussians, empirical mean & covariance estimators suffice to learn with $\Theta(n^2 \varepsilon^{-2})$ samples.

$$v_1, \dots, v_{2m} \in \mathbb{R}^{2n} \quad \hat{\mu} := \frac{1}{m} \sum_{i=1}^m v_i,$$

$$\hat{\Sigma} := \frac{1}{2m} \sum_{i=1}^m (v_{2i} - v_{2i-1})(v_{2i} - v_{2i-1})^T$$

- For bosonic Gaussian states, there are quantum noise!
 - An adaptive Gaussian protocol can learn with $N = \Theta(n^2 \varepsilon^{-2} + n \text{poly} \log \log E)$ copies given that $\|\Sigma\|_{\text{op}} \leq E$ [BMEM25]

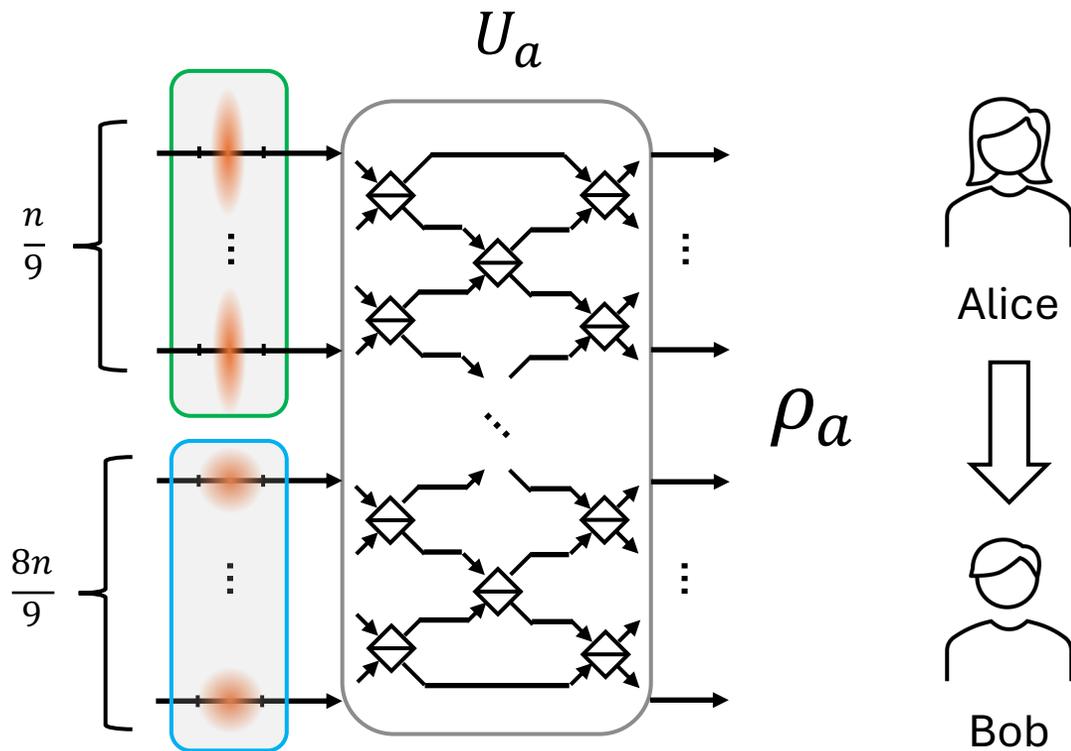


E.g. Vacuum seed overwhelms signals in highly-squeezed quadrature.

By adaptively refining seeds using $\Theta(n \cdot \text{poly} \log \log E)$ copies, one can find a good measurement seed V . Then empirical estimates suffice.

Learning Gaussian Wigner – lower bound

- Lower bound: $\Omega(n^2 \varepsilon^{-2})$ via an information-theoretic argument:



Gaussian state ensemble of size $2^{\Omega(n^2)}$

(i) Pick ρ_a at uniform random

(ii) Send N copies of ρ_a to Bob

(iii) Measure $\rho_a^{\otimes N}$ to guess a

Guarantee: $\forall a \neq b,$
 $\text{TV}(W_{\rho_a}, W_{\rho_b}) \geq 3\varepsilon$

↓
 If N suffices for learning,
 Bob can guess a w.h.p.

↓
 Fano's inequality:
 $I(A: B) \geq \Omega(n^2)$

↓
 Compute Holevo χ quantity
 $I(A: B) \leq N \cdot O(\varepsilon^2)$

↓
 $N \geq \Omega(n^2 \varepsilon^{-2})$ is necessary

Gaussian Wigner TVD \leftrightarrow Trace distance?

- We have seen $\tilde{\Theta}(n^2 \varepsilon^{-2})$ suffices to learn Gaussian Wigner to ε TVD.
 - And Gaussian measurements can achieve this.
- Is this the same as learning Gaussian states to ε trace distance?
 - $D_{\text{tr}}(\rho, \sigma) := \frac{1}{2} \text{Tr}|\rho - \sigma|$.
- **Lemma 1.** There exists constants $c_0, c_1, c_2 \geq 0$ such that for any n -mode Gaussian states ρ, σ satisfying $D_{\text{tr}}(\rho, \sigma) \leq c_0$, it holds that
$$c_1 \text{TV}(W_\rho, W_\sigma) \leq D_{\text{tr}}(\rho, \sigma) \leq \sqrt{n} \cdot c_2 \text{TV}(W_\rho, W_\sigma)$$
 - That is, Gaussian trace distance can be a factor of $\sim \sqrt{n}$ larger than Wigner TVD
 - Quantum samples can be more powerful than classical Wigner samples

An example of maximal separation

- Consider the following pair of n-mode Gaussian states:

$$\Sigma_0 = \frac{1}{2} I_{2n} \left(\begin{array}{c} \hat{p} \\ \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \\ \hat{x} \end{array} \right) \otimes n \quad \text{vs.} \quad \left(\begin{array}{c} \hat{p} \\ \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \\ \hat{x} \end{array} \right) \otimes n \quad \Sigma_1 = \frac{1}{2} \left(1 + \frac{\varepsilon}{n} \right) I_{2n}$$

- $\text{TV}(\mathcal{N}(0, \Sigma_0), \mathcal{N}(0, \Sigma_1)) \leq \frac{1}{\sqrt{2}} \|\Sigma_0^{-1/2} \Sigma_1 \Sigma_0^{-1/2} - I_{2n}\|_F = \varepsilon n^{-1/2}$ Wigner TV distance

- Apply **photon counting**,

- $\Pr(\text{Vacuum}|\rho_0) = 1,$

- $\Pr(\text{Vacuum}|\rho_1) = \langle 0|\rho_1|0\rangle = (1 + \varepsilon/2n)^{-n} \leq 1 - \varepsilon/4.$

- Thus $D_{\text{tr}}(\rho_0, \rho_1) \geq \varepsilon/4$, because D_{tr} is no smaller than any measured TVD

- This means $D_{\text{tr}}(\rho_0, \rho_1) \geq \sqrt{n}/4 \cdot \text{TV}(W_{\rho_0}, W_{\rho_1})$, maximal separation up to const.

Summary, so far

- We have proven that:
 - **Thm 1.** $N = \tilde{\Theta}(n^2 \varepsilon^{-2})$ is necessary and sufficient to learn Gaussian Wigner to ε TVD
 - **Lemma 1.** $\Omega\left(\text{TV}(W_\rho, W_\sigma)\right) = D_{\text{tr}}(\rho, \sigma) = \sqrt{n} \cdot O\left(\text{TV}(W_\rho, W_\sigma)\right)$
- This implies that, to learn a Gaussian state to ε trace distance:
 - $N = \tilde{O}(n^3 \varepsilon^{-2})$ copies are sufficient, using adaptive Gaussian measurements [BMEM25]
 - $N = \Omega(n^2 \varepsilon^{-2})$ copies are necessary, using any measurements.
- The rest of this talk
 - Can we close the n^2 vs. n^3 gap?
 - What are the roles of non-classical measurements and adaptivity?

Lower bound for classical measurements

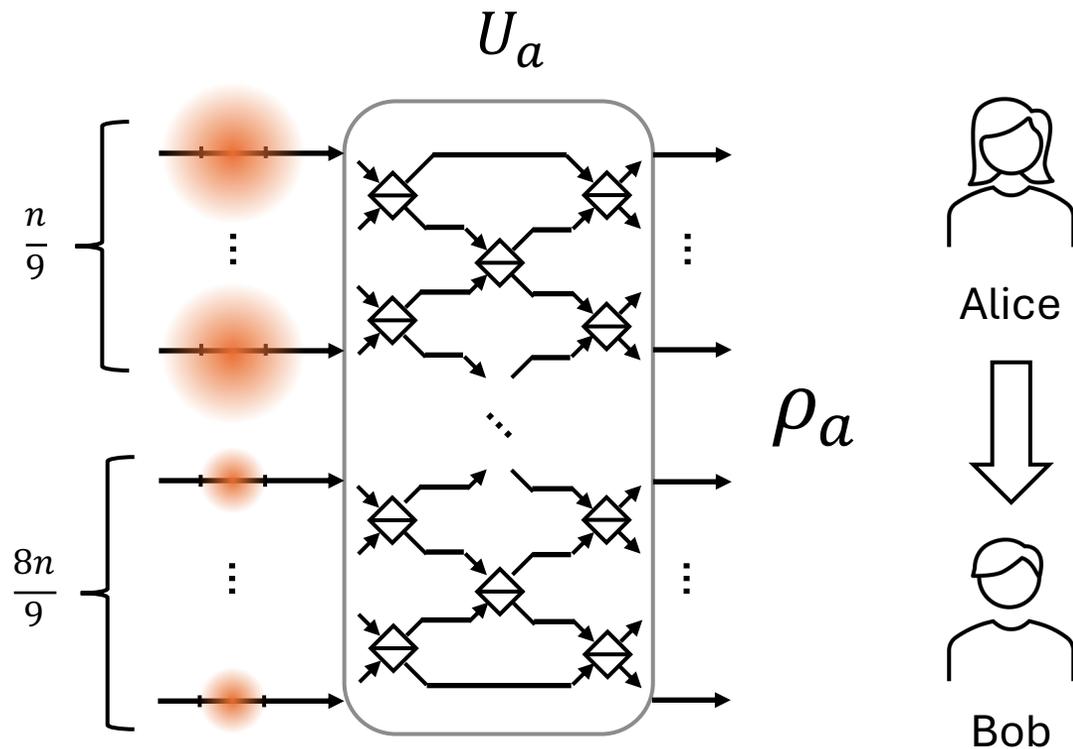
Lower bounds for classical measurements

- **Thm 2.** $N = \Omega(n^3 \varepsilon^{-2})$ are necessary for any **classical measurements**.
 - **Recall:** Classical POVMs have non-negative Wigner functions.
 - This implies [BMEM25] is the (near-)sample optimal Gaussian scheme.
 - Thm 2 holds even if ρ is promised to be **passive** –
 $\rho(\mu, \Sigma)$ is **passive** if it can be generated via passive linear optics on thermal states.
- Learning bosonic Gaussians via classical meas. is strictly harder than learning their Wigner distribution (n^3 vs. n^2)
- Key lemma [ME12]: Classical measurements on Gaussian states can be exactly simulated by samples from their Wigner distribution.

$$\Pr(\xi|\rho) = \text{Tr}[E_\xi \rho] \propto \int d^{2n}r \underbrace{W_{E_\xi}(r)}_{\text{Conditional distribution } \Pr(\xi|r)} \cdot \underbrace{W_\rho(r)}_{\text{Wigner distribution}}$$

Lower bounds for classical measurements

- $\Omega(n^3 \varepsilon^{-2})$ via an information-theoretical argument
 - Leverage the Wigner TVD – Trace distance separation in Lemma 1.



Passive Gaussian state ensemble of size $2^{\Omega(n^2)}$

(i) Pick ρ_a at uniform random

(ii) Send N **classical samples** from W_ρ to Bob.

(iii) Processing the classical samples to guess a

Guarantee: $\forall a \neq b,$
 $D_{\text{tr}}(\rho_a, \rho_b) \geq 3\varepsilon$



[ME12]

If N suffices for learning,
 Bob can guess a w.h.p.



Fano's inequality:
 $I(A: B) \geq \Omega(n^2)$



Obtain mutual info upper bound:
 $I(A: B) \leq N \cdot O(\varepsilon^2/n)$



$N \geq \Omega(n^3 \varepsilon^{-2})$ is necessary

Upper bounds

Learning Pure Gaussian states

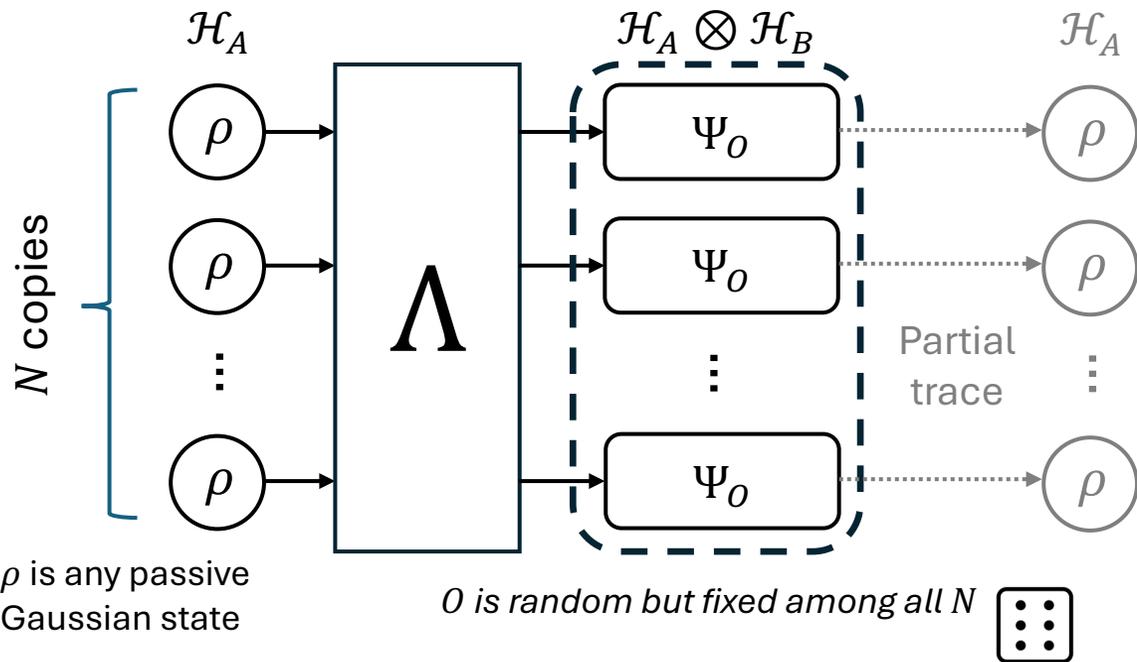
- When can we do better than $\tilde{O}(n^3 \varepsilon^{-2})$?
- **Thm 5.** With the promise that $\rho(\mu, \Sigma)$ is **pure** and $\|\Sigma\|_{\text{op}} \leq E$,
 $N = \tilde{O}(n^2 \varepsilon^{-2}) = O(n^2 \varepsilon^{-2} + n \text{ poly log log } E)$ suffices using Gaussian meas.
 - We also show a matching lower bound of $\tilde{\Omega}(n^2 \varepsilon^{-2})$ for pure states with **any** meas.
- **Lemma:** $D_{\text{tr}}(\rho, \sigma) = \Theta(\text{TV}(W_\rho, W_\sigma))$ for pure Gaussian states ρ and σ
 - Proof: Use $D_{\text{tr}}^2 = 1 - F$ for pure states and pure Gaussian fidelity formula.
- **Algorithm** using $\tilde{O}(n^2 \varepsilon^{-2})$ samples:
 1. Uses [BMEM25] to learn Wigner to ε TVD
 2. Project the estimator to a nearby pure state
 3. Convert from Wigner TV to trace distance.
- No non-Gaussian advantages for learning Pure Gaussian states!

Learning Passive Gaussian states

- **Thm 6.** With the promise that $\rho(\mu, \Sigma)$ is **passive** and $\|\Sigma\|_{\text{op}} \leq E$,
 1. $O(n^2 \varepsilon^{-2} + n \text{ poly log log } E)$ copies suffices using a non-classical scheme
 2. In contrast, any classical measurements require $\Omega(n^3 \varepsilon^{-2})$ copies.
- Thus, non-classical measurements provide advantages in learning!
- **Techniques:** Passive Random purification channels [MGC+25]

Learning Passive Gaussian states

- **Techniques:** Passive Random purification channels [MGC+25]



$$\Lambda^{(n,N)}(\rho_A^{\otimes N}) = \mathbb{E}_O[|\Psi_O\rangle\langle\Psi_O|^{\otimes N}] = \mathbb{E}_O\left[\left((I_A \otimes U_O) |\Psi\rangle\langle\Psi| (I_A \otimes U_O)^\dagger\right)^{\otimes N}\right]$$

- There exists a quantum channel Λ that maps N copies of any passive Gaussian states $\rho^{\otimes N}$ to N copies of a random purification $\Psi_O^{\otimes N}$
- Furthermore, Ψ_O are **pure** $2n$ -mode Gaussian states with $\|\Sigma_\Psi\|_{\text{op}} \leq 4\|\Sigma_\rho\|_{\text{op}}$
- **Algorithm:** First purify, then runs pure Gaussian state learning using $\tilde{O}(n^2 \varepsilon^{-2})$ copies (**Thm. 5**)

Facts:

- Λ is an **entangling** and **non-Gaussian** channel
- Experimental implementation of Λ is unclear
 - Open problem: Find an experimentally feasible scheme

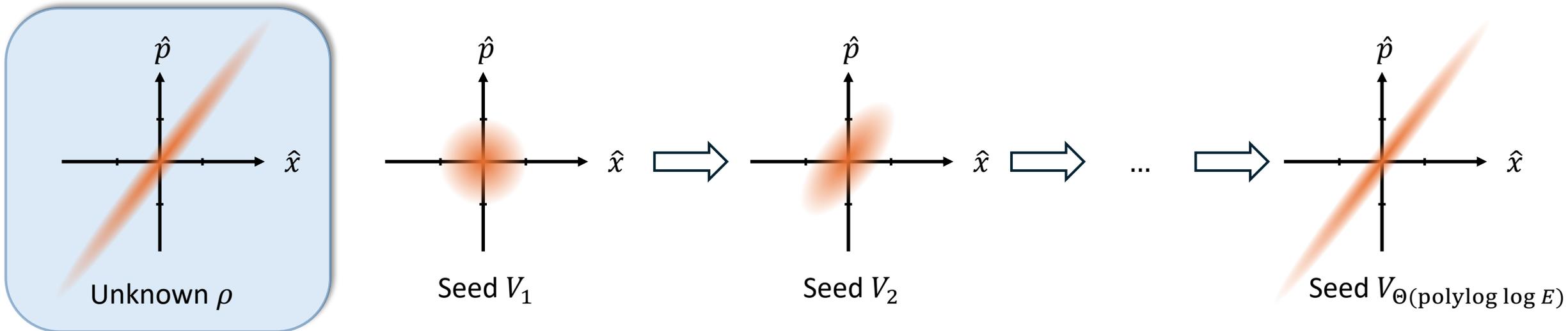
Conjectures for general Gaussian states

- **Proposition:** If there exists a random purification channel for any Gaussian states, such that the purified states are Gaussian whose covariance matrices satisfy $\|\Sigma_\Psi\|_{\text{op}} \leq \text{poly}(\|\Sigma\|_{\text{op}})$, then $N = \tilde{\Theta}(n^2 \varepsilon^{-2})$ suffices to learn any Gaussian states, and there are non-Gaussian advantages.
- **Conjecture:** There exist such random purification channels
 - Similar concepts of random purification for any n-qubit states are known [PSTW25]
- **Difficulty:** Symplectic group is not compact. Need careful analysis

On the role of adaptivity

Adaptive learning

- [BEME25]'s upper bounds has $O(\text{poly log log } E)$ energy dependence
 - $\Pr(x|\rho, V) = \mathcal{N}(\mu, \Sigma + V)$. We want V small in highly-squeezed directions of Σ
 - This is achieved by adaptively choosing a good seed V in Gaussian measurements
 - Use $\Theta(n \text{ polyloglog } E)$ copies to find a good V . Then measure $\Theta(n^2 \varepsilon^{-2})$ copies with V



- Is adaptivity necessary? How efficient can a non-adaptive scheme be?

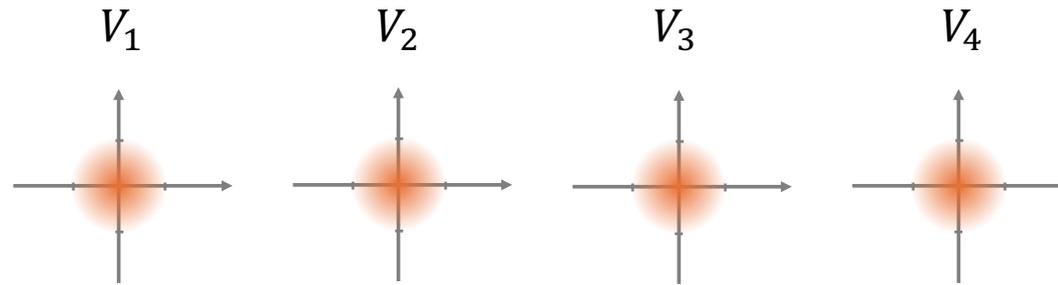
Non-adaptive lower bounds

- **Thm 7.** Consider learning 1-mode Gaussian state using single-copy Gaussian meas., for any **non-adaptive** schemes, $N = \Omega(E/\varepsilon^2)$ is needed to learn ρ such that $\|\Sigma\|_{op} \leq E$
- **Proof ideas:** Distinguishing $\rho_{0,\theta}$ with $\rho_{1,\theta}$ for random $\theta \sim \text{Unif}[0, \pi)$, where

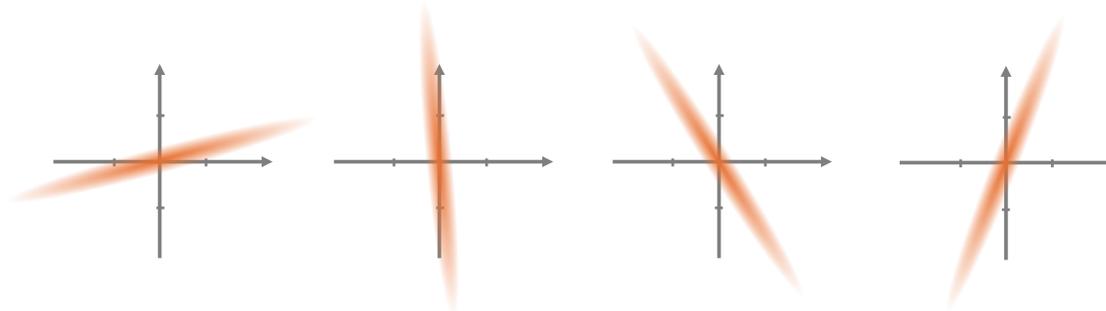
$$\Sigma_{0,\theta} = \frac{1}{2} R_\theta \begin{pmatrix} E & 0 \\ 0 & E^{-1} \end{pmatrix} R_\theta^T$$
$$\Sigma_{1,\theta} = \frac{1}{2} R_\theta \begin{pmatrix} E & 0 \\ 0 & E^{-1} + 2\varepsilon \end{pmatrix} R_\theta^T$$

- No fixed seed V can distinguish $\rho_{0,\theta}$ vs. $\rho_{1,\theta}$ well in average over θ
- Adaptivity is necessary for energy-independent scaling (for Gaussian meas.)

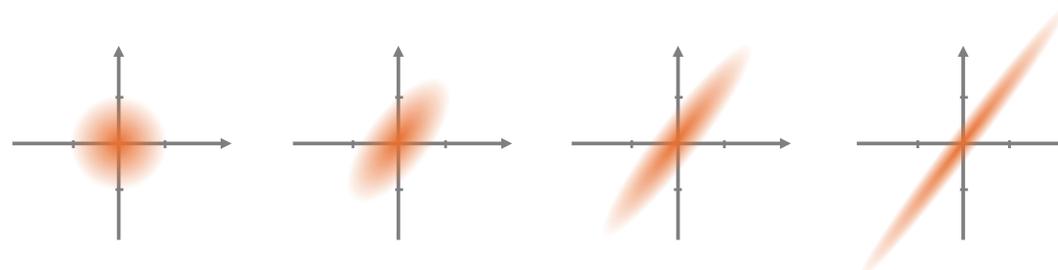
Learning 1-mode Gaussians - Comparison



(a) Heterodyne $N = \Theta(\bar{E}^2/\varepsilon^2)$



(b) Random Homodyne $N = \tilde{\Theta}(\bar{E}/\varepsilon^2)$



(c) Adaptive General-dyne $N = \tilde{\Theta}(1/\varepsilon^2)$

- We also obtain the following nearly-tight bounds for learning 1-mode Gaussian states with $\|\Sigma\|_{\text{op}} \leq E$:

- Heterodyne: $\Theta(E^2 \varepsilon^{-2})$

- Random-angle Homodyne $\tilde{\Theta}(E \varepsilon^{-2})$

- Near-optimal non-adaptive Gaussian schemes

- Upper bound not previously known

- Adaptive [BMEM25]

$$O(\text{poly log log } E \cdot \varepsilon^{-2}) = \tilde{\Theta}(\varepsilon^{-2})$$

Summary

Summary of Results

Task: Learn an n -mode $\rho(\mu, \Sigma)$ to ε trace distance with probability at least $2/3$. Promise: $\|\Sigma\|_{\text{op}} \leq \bar{E}$.

Class of states	Class of measurements	Bounds	
All Gaussian	Classical (including Gaussian)	$\tilde{O}(n^3/\varepsilon^2)$ [BMEM25]	$\Omega(n^3/\varepsilon^2)$ Thm. 2
	Any		$\Omega(n^2/\varepsilon^2)$ Thm. 3
Pure Gaussian	Any or Gaussian	$\tilde{\Theta}(n^2/\varepsilon^2)$ Thm. 3, 5	
Passive Gaussian	Any	$\tilde{O}(n^2/\varepsilon^2)$ Thm. 6	} Non-Gaussian Advantages
	Classical (including Gaussian)	$\Theta(n^3/\varepsilon^2)$ Thm. 2, 6	
1-mode Gaussian	Non-adaptive 1-copy Gaussian	$\tilde{\Theta}(\bar{E}/\varepsilon^2)$ Thm. 7	Adaptivity is helpful

Outlook

- Can we achieve $\tilde{O}(n^2 \varepsilon^{-2})$ for all Gaussian states?
- Experimentally-friendly non-classical learning scheme?
- Applications to concrete quantum sensing and benchmarking tasks?
- Learning other bosonic objects: Gaussian processes, non-Gaussian states?

Thank you!

Questions and comments are welcome

Based on joint work with Francesco Anna Mele, Marco Fanizza, Alfred Li, Zachary Mann, Hsin-Yuan (Robert) Huang, Yanbei Chen, and John Preskill.

[\[arXiv 2603.18136\]](https://arxiv.org/abs/2603.18136)