IQI Institute for Quantum Information

Accurate Quantum State Estimation

via

"Keeping the ~~Expert~~ Experimentalist Honest"

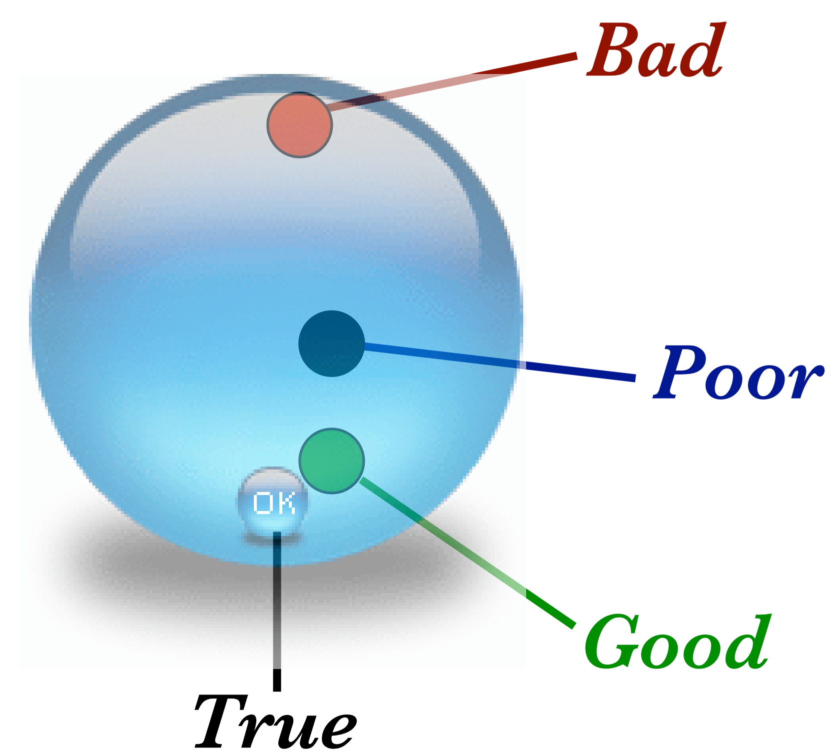
quant-ph/0603116

Robin Blume-Kohout (Caltech-IQI)

with Patrick Hayden (McGill)

and Karan Malhotra (IIT-Kanpur)

What is State Estimation?



- **Goal**: Characterize a *source* of quantum systems.
- You measure *N* identical copies.
- Measurement record:

$$\mathcal{M} = \left\{ \hat{E}_1, \hat{E}_2, \ldots \hat{E}_N \right\}$$

- Then report your best guess for ρ.
- **Key Application**: Verifying quantum hardware for fault tolerant quantum computation.
=> probabilities in $[10^{-3} \ldots 10^{-5}]$ are important!

What's the Problem?

- Current technology: **M**aximum **L**ikelihood **E**stimation

$$\mathcal{M} = \left\{ \hat{E}_1, \hat{E}_2, \ldots \hat{E}_N \right\} \longrightarrow \mathcal{L}(\rho) \equiv p(\mathcal{M}|\rho) \longrightarrow \rho_{\mathrm{MLE}}$$

- ρ_{MLE} does not honestly represent the experimentalist's knowledge about the observed ensemble.

- **Why?** ρ_{MLE} typically has zero eigenvalues:

$$\lambda_i = \langle \phi_i | \, \rho_{\mathrm{MLE}} \, | \phi_i \rangle = 0$$

- Zero *eigenvalue* = zero *probability*

↳ absolute certainty

↳ doesn't admit error bars!

Yes, it's a problem

PHYSICAL REVIEW A, VOLUME 64, 052312

Measurement of qubits

Daniel F. V. James,[1,*] Paul G. Kwiat,[2,3] William J. Munro,[4,5] and Andrew G. White[2,4]
[1]Theoretical Division T-4, Los Alamos National Laboratory, Los Alamos, New Mexico 87545
[2]Physics Division P-23, Los Alamos National Laboratory, Los Alamos, New Mexico 87545
[3]Department of Physics, University of Illinois, Urbana-Champaign, Illinois 61801
[4]Department of Physics, University of Queensland, Brisbane, Queensland 4072, Australia
[5]Hewlett-Packard Laboratories, Filton Road, Stoke Gifford, Bristol BS34 8QZ, United Kingdom
(Received 20 March 2001; published 16 October 2001)

$$\hat{\rho}=\begin{pmatrix} 0.5069 & -0.0239+i0.0106 & -0.0412-i0.0221 & 0.4833+i0.0329 \\ -0.0239-i0.0106 & 0.0048 & 0.0023+i0.0019 & -0.0296-i0.0077 \\ -0.0412+i0.0221 & 0.0023-i0.0019 & 0.0045 & -0.0425+i0.0192 \\ 0.4833-i0.0329 & -0.0296+i0.0077 & -0.0425-i0.0192 & 0.4839 \end{pmatrix}$$

This matrix is illustrated in Fig. 3 (right). In this case, the matrix has eigenvalues 0.986 022, 0.013 977 7, 0, and 0; and $\mathrm{Tr}\{\hat{\rho}^2\}=0.972\,435$, indicating that, while the linear reconstruction gave a nonphysical density matrix, the maximum likelihood reconstruction gives a legitimate density matrix.

1. D. F. V. James et. al., "Measurement of Qubits," *Phys. Rev. A*, 64:052312 (**2001**)

Yes, it's a problem

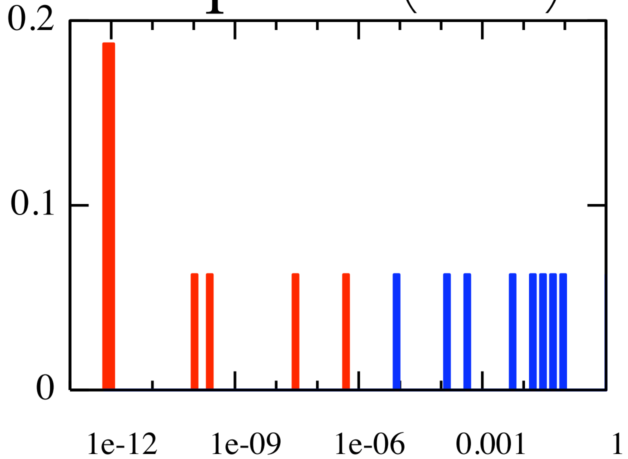
Vol 438|1 December 2005|doi:10.1038/nature04279 nature

LETTERS

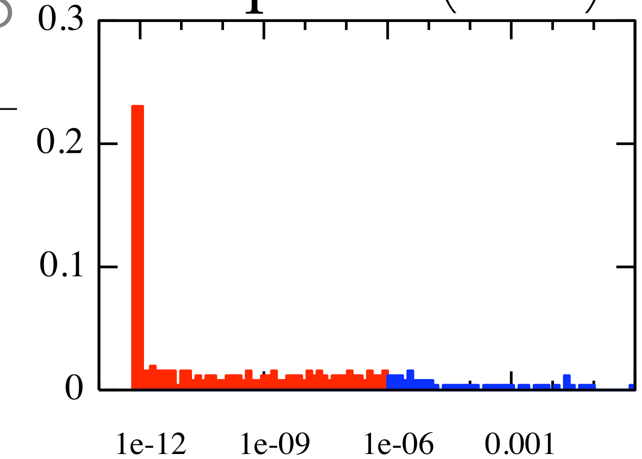
Scalable multiparticle entanglement of trapped ions

H. Häffner[1,3], W. Hänsel[1], C. F. Roos[1,3], J. Benhelm[1,3], D. Chek-al-kar[1], M. Chwalla[1], T. Körber[1,3], U. D. Rapol[1,3], M. Riebe[1], P. O. Schmidt[1], C. Becher[1]†, O. Gühne[3], W. Dür[2,3] & R. Blatt[1,3]

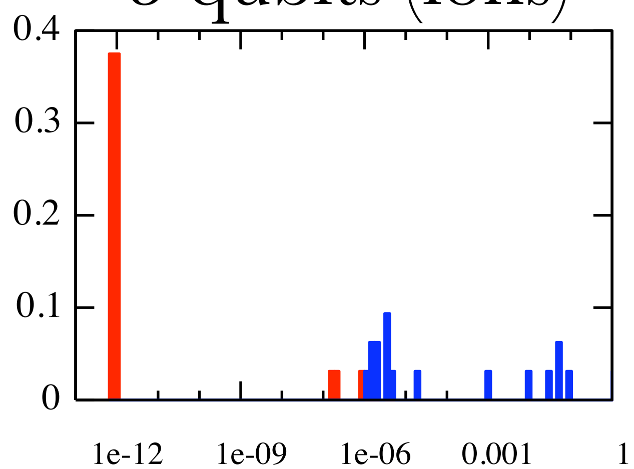
4 qubits (ions)



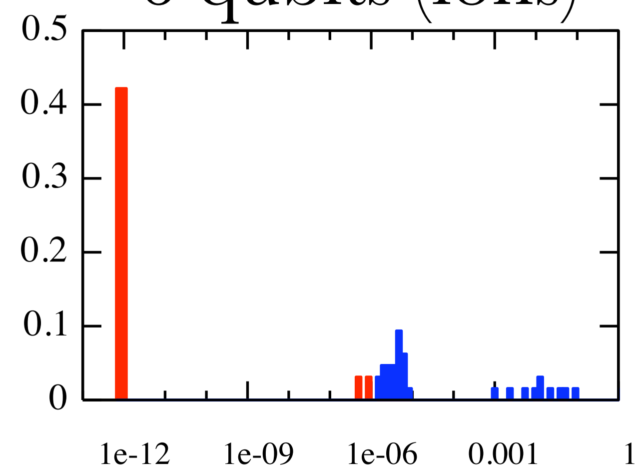
8 qubits (ions)



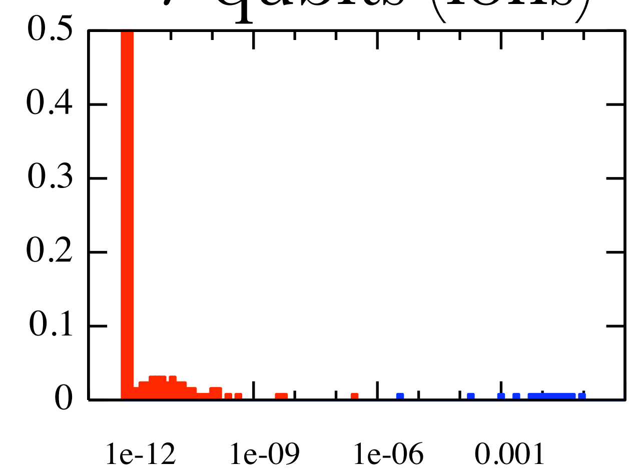
5 qubits (ions)



6 qubits (ions)



7 qubits (ions)

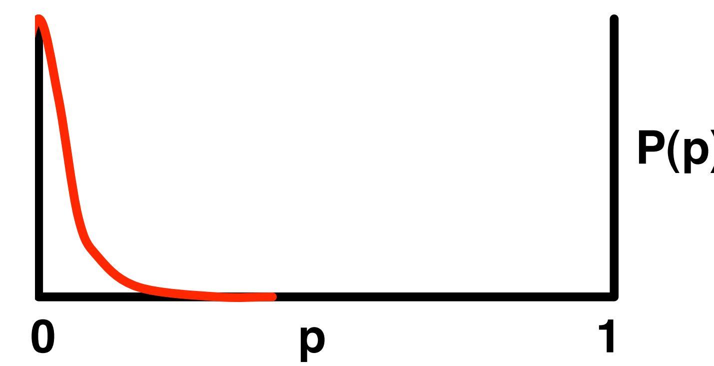


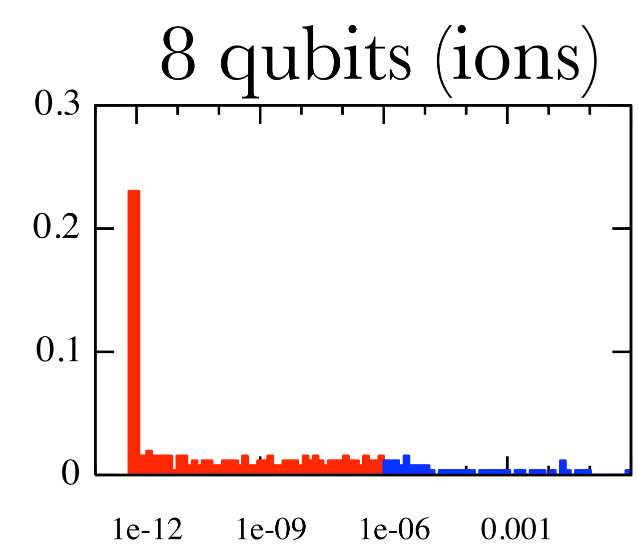
1. D. F. V. James et. al., "Measurement of Qubits," *Phys. Rev. A*, 64:052312 (**2001**)
2. Häffner et. al., "Scalable multiparticle entanglement of trapped ions," *Nature*, 438:643-6 (**2005**)

Error Bars?

- "But aren't there error bars in there?"
- Not always.
- What does ~~$p = 0 \pm 0.1$~~ mean?

$$\implies p = 0.05 \pm 0.05$$





States as Predictions

- Quantum states are like probability distributions: they predict the outcome of future measurements.
- Lets define a metric $f(\rho : \sigma)$ that measures how well σ predicts measurements on ρ.

1. **The best estimate of ρ is ρ itself**.
   – i.e., $f(\rho : \rho) > f(\rho : \sigma)$ for all $\sigma \neq \rho$.
2. **$f(\rho : \sigma)$ should correspond to an *operational* test**
   – i.e., someone's utility (reward or cost) for some practical procedure.

quant-ph/0603116

Quantum
Strictly Proper Scoring Rules

Victor the Verifier measures a copy of $\boldsymbol{\rho}$, and pays you $\boldsymbol{R_i(\sigma)}$ if he gets outcome $\boldsymbol{i}$.

1) $f(\rho:\sigma) = \sum_i p(i) R_i(\sigma)$

Measurement $= \{E_i\}$: $\sum E_i = \mathbb{1}$.

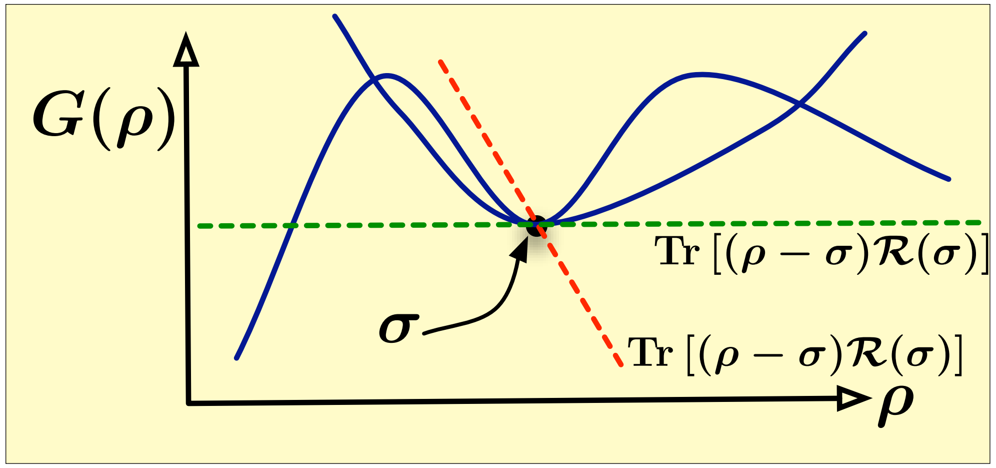
$$f(\rho:\sigma) = \sum_i \mathrm{Tr}[\rho E_i] R_i(\sigma)$$
$$= \mathrm{Tr}[\rho \mathcal{R}(\sigma)]$$

where $\mathcal{R}(\sigma) = \sum_i E_i R_i(\sigma)$.

Define "value": $G(\rho) \equiv f(\rho:\rho) = \mathrm{Tr}[\rho\mathcal{R}(\rho)]$.

2) $f(\rho:\rho) > f(\rho:\sigma)$ if $\sigma \neq \rho$.

$$f(\rho:\rho) > f(\sigma:\sigma) + f(\rho:\sigma) - f(\sigma:\sigma)$$
$$G(\rho) > G(\sigma) + \mathrm{Tr}\left[(\rho-\sigma)\mathcal{R}(\sigma)\right]$$



- $f(\rho:\sigma)$ is a subtangent to $G(\rho)$
- $G(\rho)$ must be strictly convex.

- The estimator's *expected loss* for lying is $\Delta(\rho:\sigma) \equiv G(\rho) - f(\rho:\sigma)$.
- $\Delta(\rho:\sigma)$ is an *operational divergence* = good measure of σ's predictive accuracy.
- Operational divergences are $1:1$ with strictly convex "entropies" $G(\rho)$.

A survey of metrics

- GOOD METRICS (OPERATIONAL DIVERGENCES)

1. L2 distance: $\mathbf{Tr}\left[(\boldsymbol{\rho} - \boldsymbol{\sigma})^{2}\right] \Leftrightarrow \boldsymbol{R_i} = \mathbf{2}\boldsymbol{s_i} - \mathbf{Tr}(\boldsymbol{\sigma}^{2}) - \mathbf{1}$

2. Relative entropy: $\boldsymbol{S}(\boldsymbol{\rho}||\boldsymbol{\sigma}) = \mathbf{Tr}\left[\boldsymbol{\rho}\log\frac{\boldsymbol{\sigma}}{\boldsymbol{\rho}}\right] \Leftrightarrow \boldsymbol{R_i} = \mathbf{ln}(\boldsymbol{s_i})$

- BAD METRICS

1. Overlap: $\mathbf{1} - \mathbf{Tr}[\boldsymbol{\rho\sigma}]$ — *Improper (motivates lying)*

2. Trace-norm: $||\boldsymbol{\rho} - \boldsymbol{\sigma}||_{\mathbf{1}}$

3. Fidelity: $\mathbf{1} - \left[\mathbf{Tr}\sqrt{\sqrt{\boldsymbol{\sigma}}\boldsymbol{\rho}\sqrt{\boldsymbol{\sigma}}}\right]^{\mathbf{2}}$

} *Non-operational* (2, 3)

Bayesian Mean Estimation

unconditionally optimizes every operational divergence

If you *know* $\boldsymbol{\rho}$, the optimal estimate is $\boldsymbol{\sigma} = \boldsymbol{\rho}$.

But what if your knowledge is uncertain (probabilistic)?

$\Longrightarrow$ the state could be $\boldsymbol{\rho_i}$ with probability $\boldsymbol{\pi_i}$.

KEY FACT: $\boldsymbol{f(\rho : \sigma)} = \mathrm{Tr}[\boldsymbol{\rho \mathcal{R}(\sigma)}]$ is *linear* in $\boldsymbol{\rho}$.

$\Longrightarrow$ *expected* utility is

$$\overline{\boldsymbol{f}} = \sum \boldsymbol{\pi_i f(\rho_i : \sigma)} = \boldsymbol{f}\left(\sum \boldsymbol{\pi_i \rho_i : \sigma}\right)$$

$\Longrightarrow$ optimal estimate is $\boldsymbol{\sigma} = \overline{\boldsymbol{\rho}} \equiv \sum \boldsymbol{\pi_i \rho_i}$

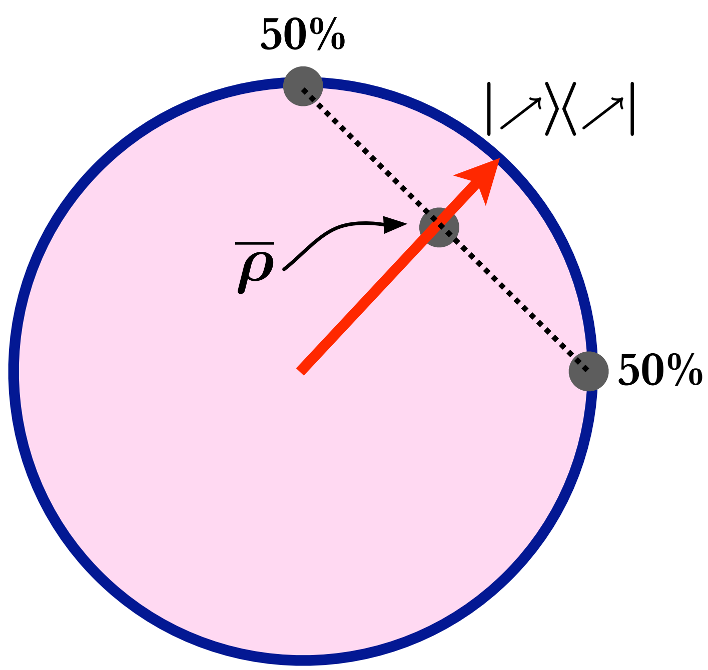
This is fairly straightforward, but tedious... see quant-ph/0603116

If unknown $\boldsymbol{\rho}$ was selected from distribution $\boldsymbol{\pi_0(\rho)}\mathrm{d}\boldsymbol{\rho}$, **and** measurements $\boldsymbol{\mathcal{M}} = \{\boldsymbol{E_1, E_2 \ldots}\}$ were made, **then**:

1. Your knowledge is $\boldsymbol{\pi(\rho)}\mathrm{d}\boldsymbol{\rho} = \frac{\boldsymbol{p(\mathcal{M}|\rho)\pi_0(\rho)}\mathrm{d}\boldsymbol{\rho}}{\int \boldsymbol{p(\mathcal{M}|\rho)\pi_0(\rho)}\mathrm{d}\boldsymbol{\rho}}$.
2. The optimal estimate is $\overline{\boldsymbol{\rho}} = \int \boldsymbol{\rho\pi(\rho)}\mathrm{d}\boldsymbol{\rho}$.

But isn't this obvious? *NO*

(a) Suppose we optimize fidelity.



$$\overline{F}\left(\rho_i, \overline{\rho}\right) = \frac{3}{4}$$
$$\overline{F}\left(\rho_i, |\nearrow\rangle\langle\nearrow|\right) = \frac{3+2\sqrt{2}}{4+2\sqrt{2}} \approx 0.85$$

MORAL: fidelity measures how well σ *simulates* ρ, not how well it *estimates* ρ.

(b) Suppose we assume the future will look (statistically) just like the past.

- We know what future datasets will look like
- We can add them to measurement record
- $\mathcal{M} \longrightarrow \mathcal{M} \cup \mathcal{M} \cup \mathcal{M} \cup \ldots$
- $p(\mathcal{M}|\rho) \longrightarrow p(\mathcal{M}|\rho)^{\infty}$
- So $\pi(\rho)d\rho \rightarrow \delta\left(\rho - \rho_{\mathrm{MLE}}\right)$

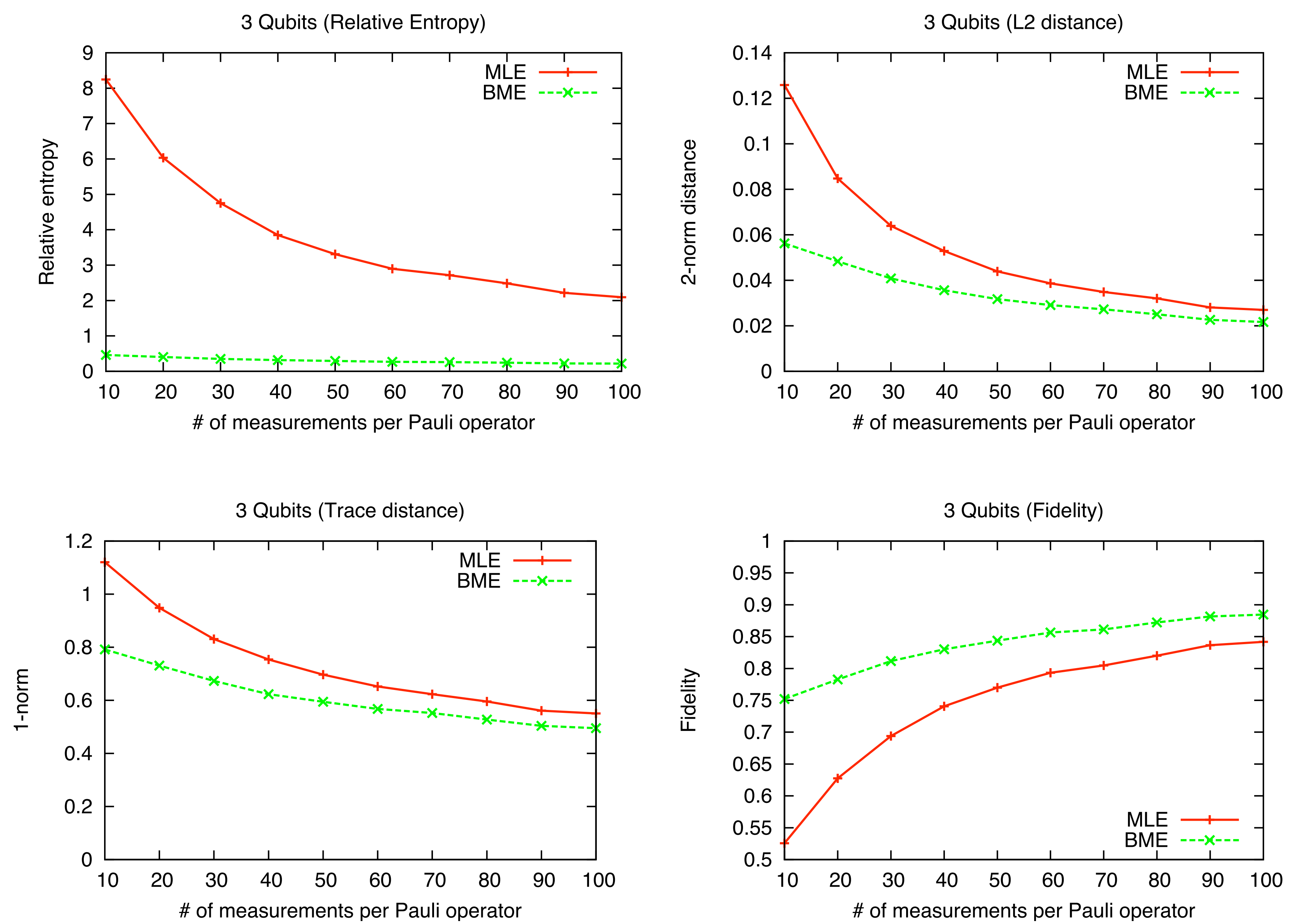
MORAL: MLE can be derived by assuming this "frequentist axiom".

(This explains a lot...)

How good is MLE?

NUMERICS:

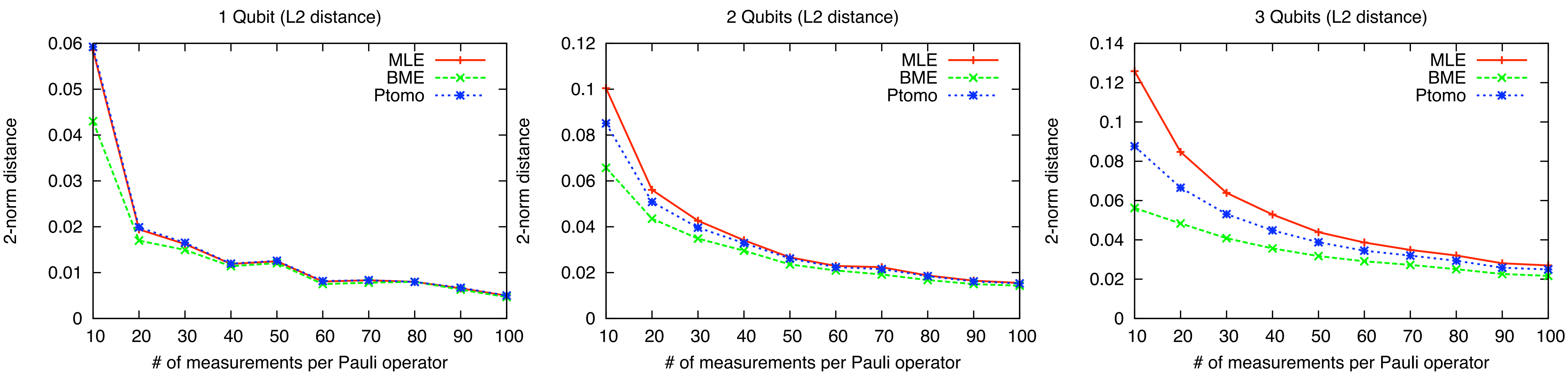
1. Generate random (Hilbert-Schmidt measure) mixed states $\boldsymbol{\rho}$ for $\boldsymbol{n}$ qubits,
2. Measure each of $\mathbf{4}^{\boldsymbol{n}-\mathbf{1}}$ Pauli observables $\boldsymbol{N}$ times,
3. Analyze measurement record to get $\boldsymbol{\sigma}$,
4. Compare $\boldsymbol{\sigma}$ to $\boldsymbol{\rho}$ using:
   (a) L2-norm, $\mathbf{Tr}\left[(\boldsymbol{\rho}-\boldsymbol{\sigma})^{\mathbf{2}}\right]$,
   (b) relative entropy, $\mathbf{Tr}\left[\boldsymbol{\rho}(\ln\boldsymbol{\rho}-\ln\boldsymbol{\sigma})\right]$,
   (c) fidelity, $\mathbf{Tr}\left[\sqrt{\sqrt{\boldsymbol{\sigma}}\boldsymbol{\rho}\sqrt{\boldsymbol{\sigma}}}\right]^{\mathbf{2}}$
   (d) L1-norm, $\mathbf{Tr}\left[|\boldsymbol{\rho}-\boldsymbol{\sigma}|\right]$.



Quick & Dirty Tomography

1. MLE gets computationally difficult for large systems.
2. BME is even harder!
3. $\frac{1}{\sqrt{3}}\left(|\mathbf{D.James}\rangle + |\mathbf{T.Havel}\rangle + |\mathbf{P.Jessen}\rangle\right)$ suggested "quick 'n' dirty tomography".

$$\sigma_{\text{tomo}} = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & -\lambda_3 & \\ & & & -\lambda_4 \end{pmatrix} \longrightarrow \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & 0 & \\ & & & 0 \end{pmatrix} \longrightarrow \frac{1}{\lambda_1 + \lambda_2} \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & 0 & \\ & & & 0 \end{pmatrix} = \sigma.$$



Conclusions

- It’s worth thinking carefully and deeply about state estimation.
- Operational divergences are a good way to evaluate the predictive accuracy of estimates.
- BME is optimal, and a baseline for evaluating other (more efficient) approaches.
- Quick & dirty methods can outperform MLE.