Innovations in Maximum Likelihood Quantum State Tomography

Theory: S. Glancy, E. Knill

Experiment: T. Gerrits, T. Clement, B. Calkins, A. Lita, A. Miller, A. Migdall, S. W. Nam, and R. Mirin

National Institute of Standards and Technology, USA

Experiences with Maximum Likelihood Quantum State Tomography

Theory: S. Glancy, E. Knill

Experiment: T. Gerrits, T. Clement, B. Calkins, A. Lita, A. Miller, A. Migdall, S. W. Nam, and R. Mirin

National Institute of Standards and Technology, USA

Struggles with Maximum Likelihood Quantum State Tomography

Theory: S. Glancy, E. Knill

Experiment: T. Gerrits, T. Clement, B. Calkins, A. Lita, A. Miller, A. Migdall, S. W. Nam, and R. Mirin

National Institute of Standards and Technology, USA

Plan

- "Schrödinger cat" making experiment
- State of quantum state tomography
- Our work
  - stopping criterion
  - improved maximum likelihood algorithm
  - approximate confidence intervals
- Preliminary cat state data

“Schrödinger Cat” States

- (I’m talking about the) state of a single harmonic oscillator.
- superposition of two coherent (or “displaced vacuum”) states

$$\left|\pm\right\rangle = \frac{1}{\sqrt{2 \pm 2e^{-2|\alpha|^2}}}\left(\left|-\alpha\right\rangle \pm \left|\alpha\right\rangle\right)$$

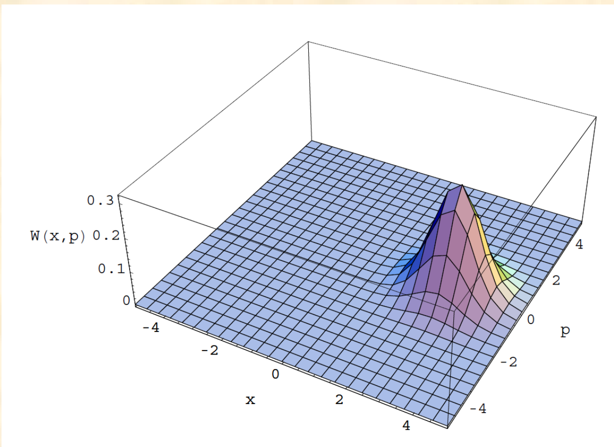
- $|+\rangle$ has only even numbers of photons.
- $|-\rangle$ has only odd numbers.
- $\langle +|-\rangle = 0$

- This type of Schrödinger cat states have been made in a light field trapped in a cavity, microwaves in a superconducting resonator, motion of a trapped ion, traveling light wave (others?)
- With photon (or phonon) numbers < 10.

“Schrödinger Cat” States

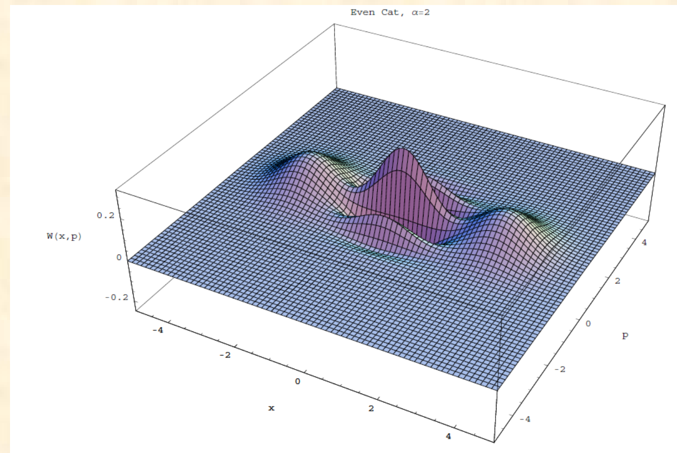
- (I’m talking about the) state of a single harmonic oscillator.
- superposition of two coherent (or “displaced vacuum”) states

coherent state



$\langle n \rangle = 4$

superposition of
coherent states



- This type of Schrödinger cat states have been made in a light field trapped in a cavity, microwaves in a superconducting resonator, motion of a trapped ion, traveling light wave (others?)
- With mean photon (or phonon) numbers < 10.

How to Make Cat States

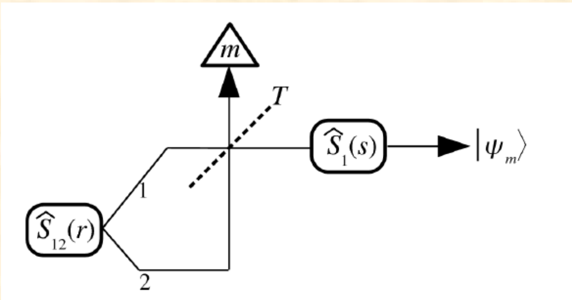
- Original cat making scheme:
  - Use Kerr effect Hamiltonian

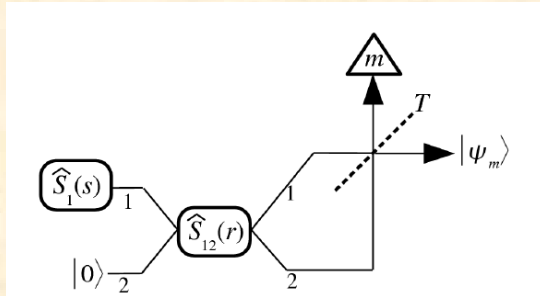
$$|\alpha\rangle \xrightarrow[t=\frac{\pi}{2\chi}]{\ \chi\hat{n}^2\ } |-\alpha\rangle + i|\alpha\rangle$$

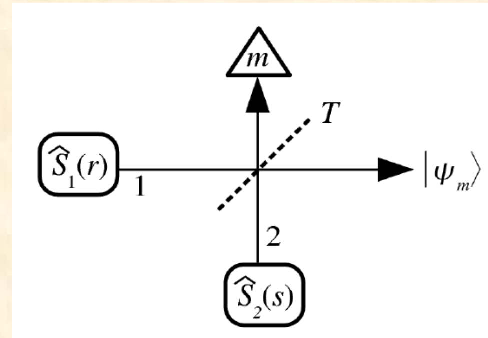
- Current materials have too much absorption and too small χ, but there is hope for EIT methods.
- We need to make cats with a specific optical mode shape, and Kerr effect interactions will disturb the mode.

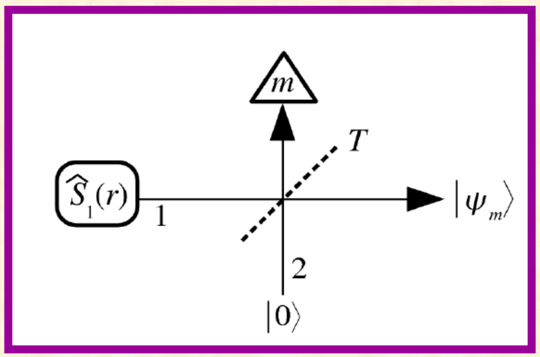
Yurke and Stoler
PRL **57**, 13

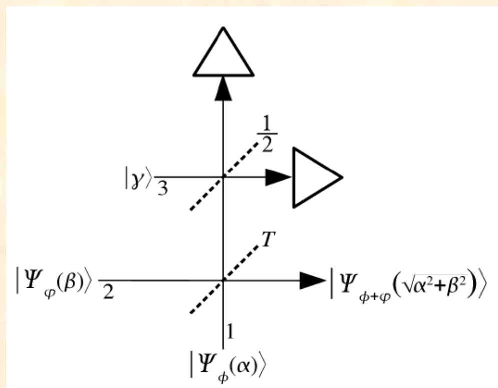
Lower Order Nonlinearity + Post Selection

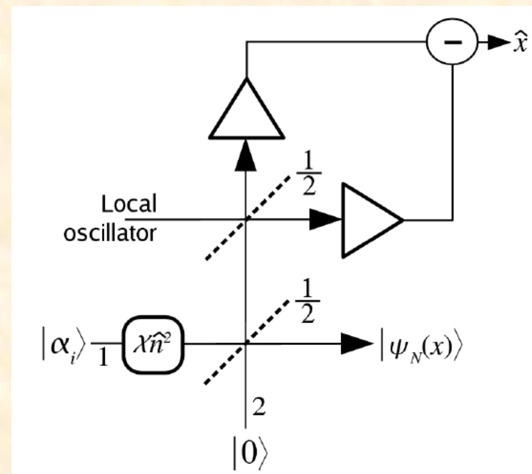


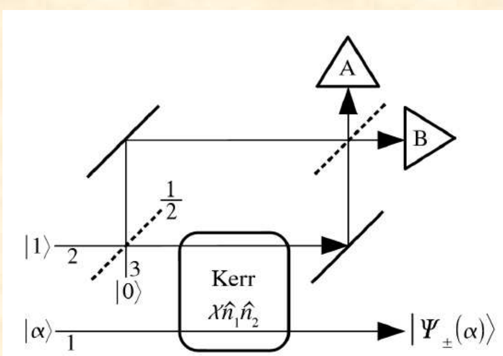








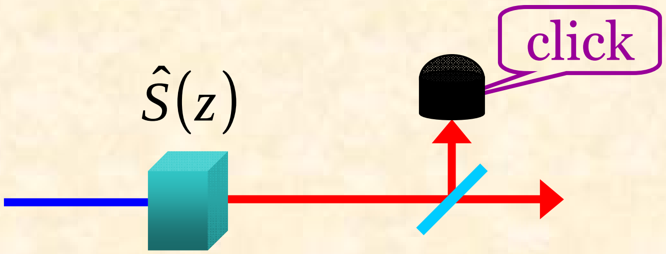




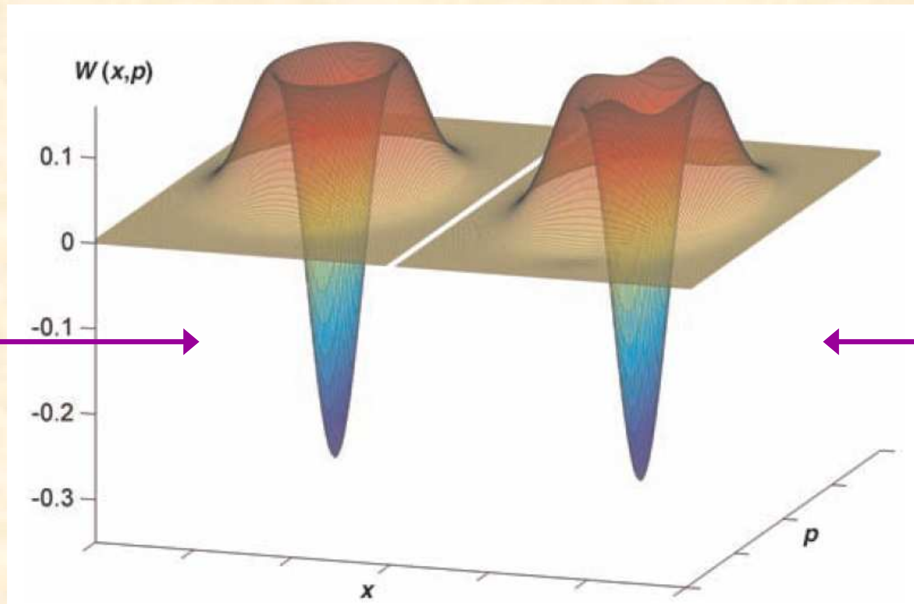
Glancy and Vasconcelos
arXiv:0705.2045

Photon Subtraction

- Make squeezed light by degenerate down conversion.
  $\omega_{pump} \rightarrow 2\omega_{squeezed}$
- Send through beam splitter.
- Trigger on observing a photon.
- Works like heralded single photon source, but with stronger squeezing ~3dB.



ideal photon subtracted state

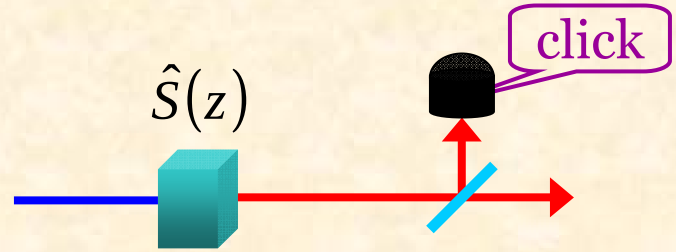


perfect cat state $|\alpha|^2=0.8$

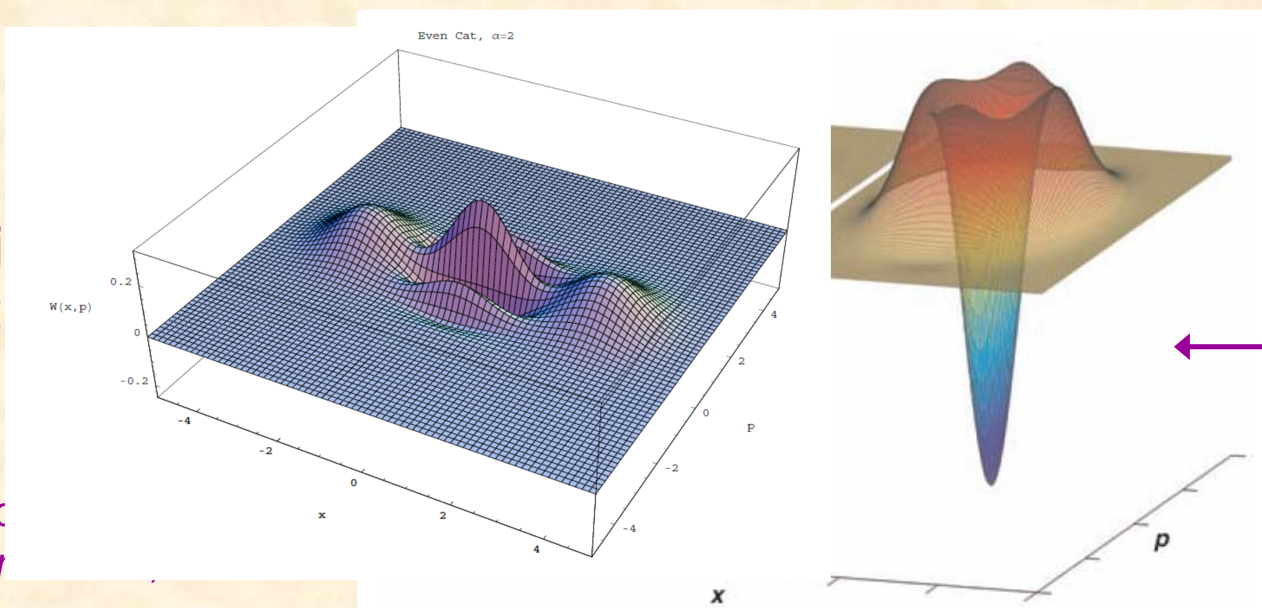
Ourjoumtsev et al. *Science* **312**, 83

Photon Subtraction

- Make squeezed light by degenerate down conversion.
  $\omega_{pump} \rightarrow 2\omega_{squeezed}$
- Send through beam splitter.
- Trigger on observing a photon.
- Works like heralded single photon source, but with stronger squeezing ~3dB.



i
subt

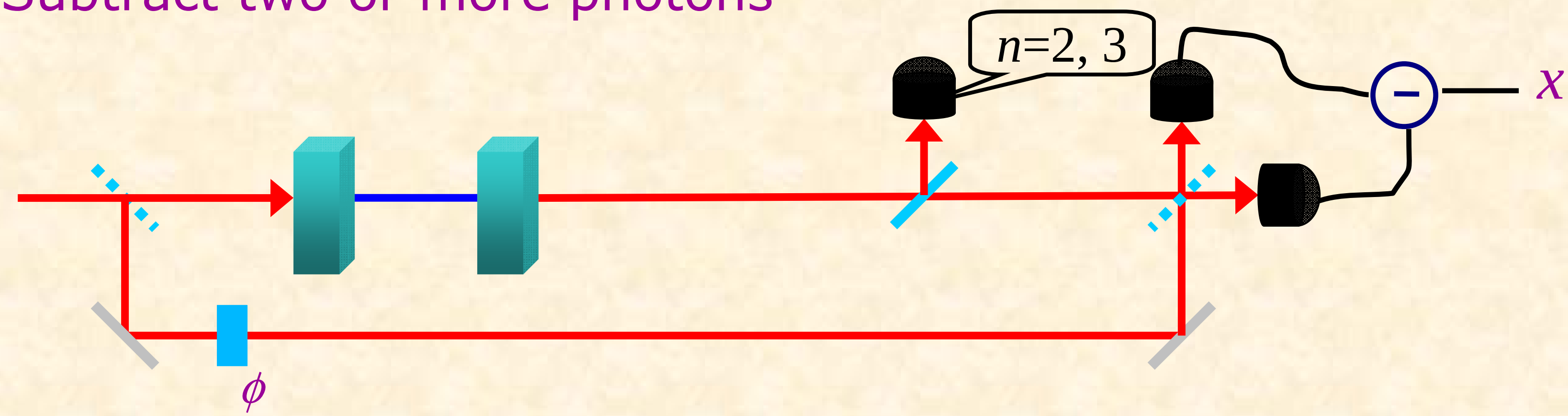


perfect cat state
$|\alpha|^2=0.8$

Ourj
Scien

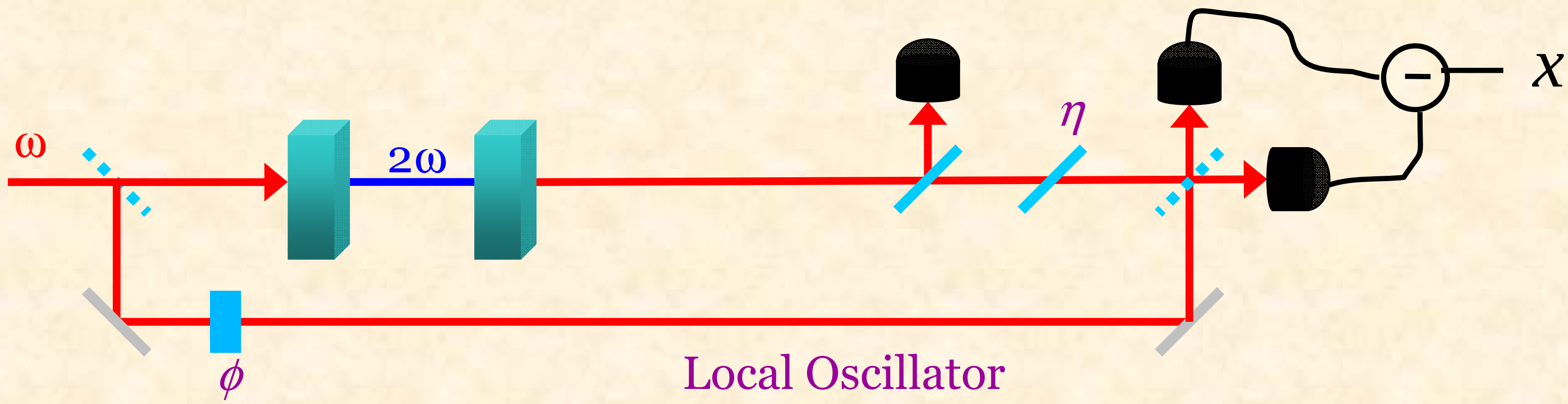
Our Photon Subtraction

- Subtract two or more photons



- Using superconducting transition edge sensor (TES) photon number resolving detectors.
  - efficiency ~ 90%
  - dark counts limited by black-body radiation
- Subtracting more photons makes a higher fidelity, larger cat, using less squeezing.
- Four Data Sets:
  - n=1 by avalanche photo diode (APD)
  - n=2 by APD
  - n=2 by TES
  - n=3 by TES

Measure by Homodyne Detection

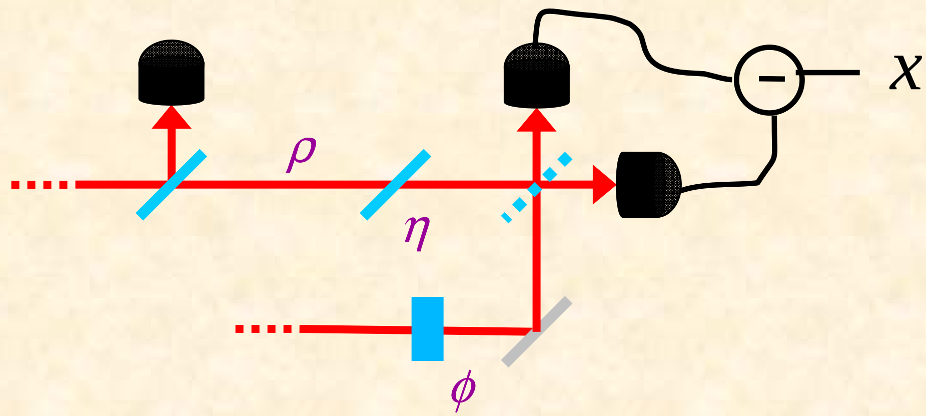


- Vary local oscillator phase ϕ, observe x.
- Record N pairs: $\{(x_m, \phi_m)|m=1\ldots N\}$.
- Calibrate system efficiency $\eta=\eta_{\text{opt}}\eta_{\text{pd}}\eta_{\text{mm}}\eta_{\text{dc}}$
  - η_{opt}=optical components 94.0% ± 0.5%
  - η_{pd}=photo-diodes 97.6% ± 2.2%
  - η_{mm}=mode-mismatch 95.0% ± 0.5%
  - η_{dc}=dark current 97.9% ± 0.1%
- η ~ 85% ± 3%

Lvovsky & Raymer
quant-ph/0511044

Forward Measurement Model

- Relate measurement probabilities to quantum state ρ:
  $$P(x|\phi) = \mathrm{Tr}[\Pi(x,\phi)\rho]$$



- $\Pi(x,\phi)$ is an element of a continuous POVM.
  $$\Pi(x,\theta) = \sum_n E_n(\eta) e^{-i\phi a^\dagger a} |x\rangle\langle x| e^{+i\phi a^\dagger a} E_n(\eta)^\dagger$$
  - $|x\rangle$ is the harmonic oscillator position eigenstate in photon number basis.
  - $e^{-i\phi a^\dagger a}$ is the phase evolution operator.
  - $E_n(\eta)$ are the Kraus operators for photon loss.

Do Quantum State Tomography

- Using set of observations $\{\Pi_m = \Pi(x_m,\phi_m)\,|\,m=1\ldots N\}$, and

  $P(x|\phi) = \mathrm{Tr}[\Pi(x,\phi)\rho]$

- infer state ρ.
- Choose tomography school:
  - linear inversion
  - maximum likelihood
  - Bayesian inference
  - maximum entropy

Paris & Řeháček (editors)
Quantum State Estimation

Maximum Likelihood

- Likelihood function:

$$L(\rho) = \prod_{m=1}^{N} \mathrm{Tr}(\Pi_m \rho)^{n_m}$$

- Loglikelihood function:

$$\mathcal{L}(\rho) = \sum_{m=1}^{N} n_m \log(\mathrm{Tr}(\Pi_m \rho))$$

- Maximize $\mathcal{L}(\rho)$ to find ρ.
- Respect ρ's constraints: hermitian, $\mathrm{Tr}(\rho)=1$.
- $\mathcal{L}$ is convex.

RρR Maximum Likelihood

- Iterative scheme:
  - begin with $\rho_o = \mathcal{N}(\mathrm{I})$=maximally mixed state
  - at each step i, compute

$$R(\rho_i) = \sum_{m=1}^{N} \frac{\Pi_m}{\mathrm{Tr}(\Pi_m \rho_i)}$$

  - find next $\rho_{i+1} = \mathcal{N}(R_i \rho_i R_i)$
  - at maximum likelihood point $\rho_{\mathrm{ML}} = \mathcal{N}(R_{\mathrm{ML}} \rho_{\mathrm{ML}} R_{\mathrm{ML}})$
- R is positive and hermitian, so each ρ_i is also hermitian and can be normalized to have trace 1.
- The "diluted algorithm", in which $\mathrm{R} \rightarrow \mathrm{I} + \varepsilon \mathrm{R}$, will increase $\mathcal{L}$ if, and ε is small enough.
- In practice $\varepsilon \rightarrow \infty$.

Řeháček et al.
quant-ph/0611244

RρR Virtues

- Always returns a density matrix.
- Has a clear method to incorporate measurement noise by adapting Π_m's.
- No need to parameterize ρ or use constraint equations.
- Simple implementation.

RρR Desiderata

- Stopping criterion
- Faster convergence
- Confidence region for ρ
  - or confidence intervals for observables of ρ.

$R\rho R$ Stopping Criterion

- Just do “a lot” of iterations?
- Compare likelihood found at each iteration?
- Compare fidelity (or trace distance?) between states at each iteration?
- We would like to bound the maximum likelihood using our knowledge of the current ρ.

$R\rho R$ Stopping Criterion

- Consider subsequent ρ's:

$$\rho' = \rho + \varepsilon(\sigma - \rho)$$

- Expand $\mathcal{L}$ to first order in ε:

$$\mathcal{L}(\rho') \approx \mathcal{L}(\rho) + \varepsilon \frac{\partial}{\partial \varepsilon} \mathcal{L}(\rho') \Big|_{\varepsilon=0}$$

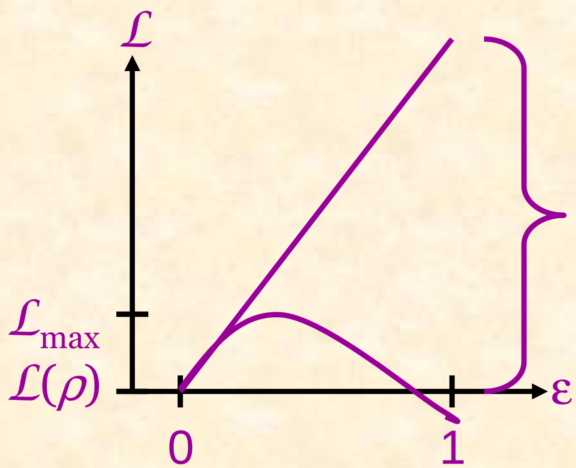
$$\mathcal{L}(\rho') \approx \mathcal{L}(\rho) + \varepsilon \mathrm{Tr}(R\sigma) - \varepsilon \mathrm{Tr}(R\rho)$$

$$\mathcal{L}(\rho') \approx \mathcal{L}(\rho) + \varepsilon \mathrm{Tr}(R\sigma) - \varepsilon N$$

- What σ will maximize $\mathcal{L}(\rho')$? $\sigma = |\psi\rangle\langle\psi|$, where $|\psi\rangle$ is the eigenvector of R with the largest eigenvalue.

$$\mathcal{L}(\rho') \approx \mathcal{L}(\rho) + \varepsilon\left[\max(\mathrm{eig}(R)) - N\right]$$

$R\rho R$ Stopping Criterion



$$\mathcal{L}(\rho') \approx \mathcal{L}(\rho) + \varepsilon\left[\max(\text{eig}(R(\rho))) - N\right]$$

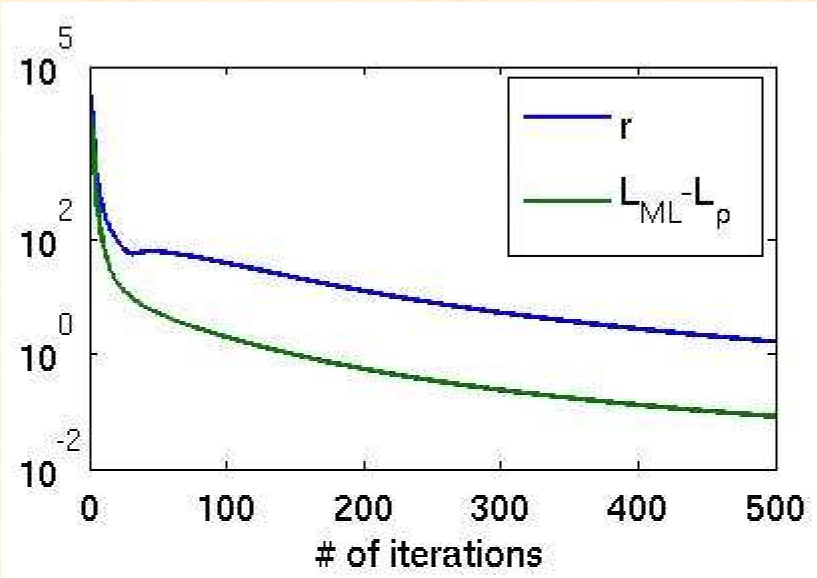
$$\mathcal{L}(\rho') \approx \mathcal{L}(\rho) + \varepsilon r$$

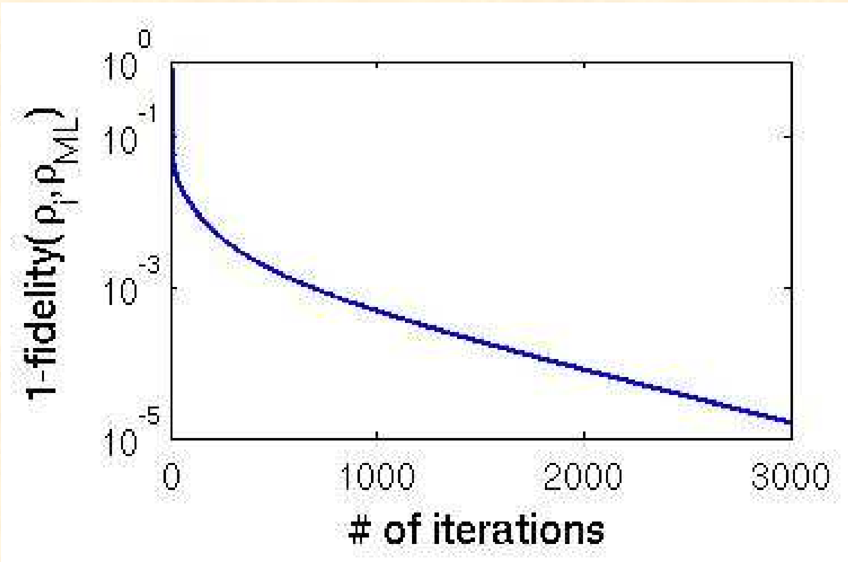
$$r = \max(\text{eig}(R(\rho))) - N$$

- $\mathcal{L}$ is convex, so $\mathcal{L}(\rho)+r \geq \mathcal{L}_{\max}$
- $\therefore$ stop iterations when r is small.

$R\rho R$ Stopping Criterion

$$r \geq \mathcal{L}_{\max} - \mathcal{L}(\rho)$$





≤10 photons, 40,000 measurements

- Bound is not very tight.
- Bounding $\mathcal{L}_{\mathrm{ML}}$ is good, but I wish we had a bound on the difference between ρ_i and ρ_{ML}.

Regularized Gradient Ascent

- Can we find an algorithm that converges faster?
- Strategy:
  - use traditional ideas of gradient ascent,
  - trust region / quadratic approximation of $\mathcal{L}$,
  - over-parameterize ρ to make optimization unconstrained.
  - To stay within trust region, restrict step size of each iteration.

Regularized Gradient Ascent

- Parameterization of ρ:

$$\rho_{i+1} = \mathcal{N}\left(\left(\rho_i^{1/2} + A\right)\left(\rho_i^{1/2} + A^\dagger\right)\right)$$

- A may be any matrix.
- Ensures ρ_{i+1} is a density matrix.
- Increases parameter space from d^2-1 to 2d^2, where d is Hilbert space dimension.

Regularized Gradient Ascent

- Quadratic approximation of $\mathcal{L}$:
  - $\rho_{i+1} = \rho_i + \Delta$, where Δ is 2nd order in A.

$$\mathcal{L}_{\mathrm{Q}}\left(\rho_{i+1}\right) \approx \mathcal{L}\left(\rho_i\right) + \mathrm{Tr}\left(R_i \Delta\right) - \frac{1}{2}\sum_{m=1}^{N} n_m \left(\frac{\mathrm{Tr}\left(\Pi_m \Delta\right)}{\mathrm{Tr}\left(\Pi_m \rho_i\right)}\right)^2$$

  - Write A as a $2d^2$ element real vector $\vec{A}$.

$$\mathcal{L}_{\mathrm{Q}}\left(\rho_{i+1}\right) \approx \mathcal{L}\left(\rho_i\right) + \vec{v}^T \vec{A} + \frac{1}{2}\vec{A}^T M \vec{A}$$

- Choose maximum step size: $s = \vec{A}^T \vec{A}$
- Maximize $\mathcal{L}_{\mathrm{Q}}(\rho_{i+1})$ subject to constraint $s \geq \vec{A}^T \vec{A}$:

$$\vec{A}(\lambda) = \left(2\lambda I - M\right)^{-1} \vec{v}$$

- λ is a Lagrange multiplier, which we set to $\lambda = \max(\mathrm{eig}(M))$ and increase if necessary.

Regularized Gradient Ascent

1. Choose step size s=1.

2. From ρ_i, calculate v, M.

3. λ=max(eig(M)), $\vec{A}(\lambda) = (2\lambda I - M)^{-1}\vec{v}$.

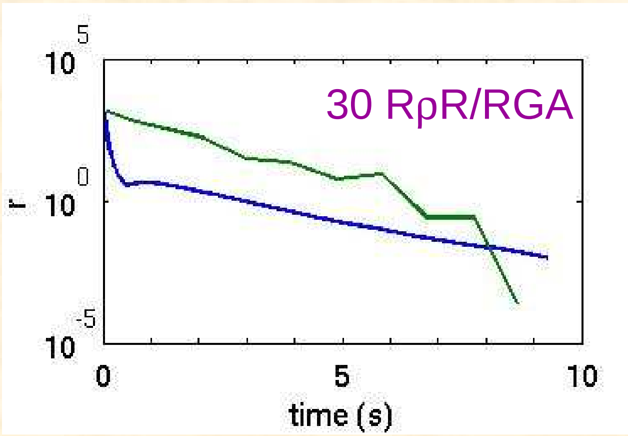
4. Check step size: if $\vec{A}(\lambda)^T \vec{A}(\lambda) \geq s$, increase lambda and goto 3.

5. Calculate new $\rho_{i+1} = \mathcal{N}\left(\left(\rho_i^{1/2} + A\right)\left(\rho_i^{1/2} + A^\dagger\right)\right)$.

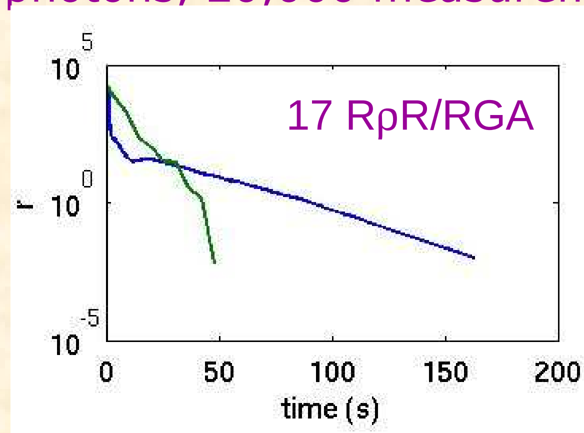
6. If (exact) $\mathcal{L}(\rho_{i+1}) \leq \mathcal{L}(\rho_i)$, reduce s and goto 4.

RGA vs. RρR Competition

10 photons, 2,000 measurements



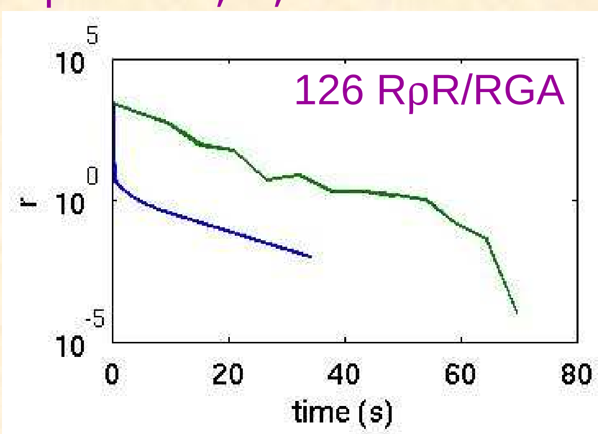
10 photons, 20,000 measurements



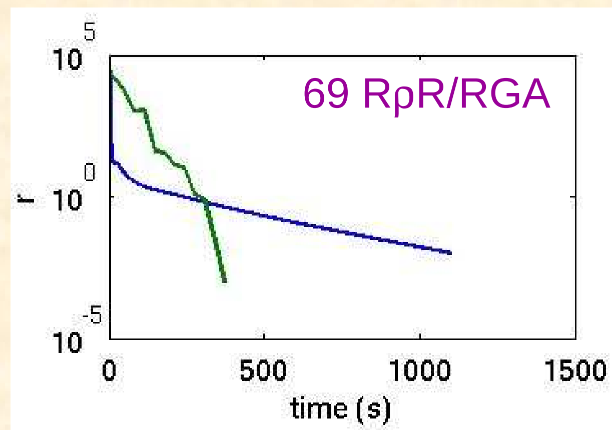
RρR —

RGA —

20 photons, 2,000 measurements



20 photons, 20,000 measurements

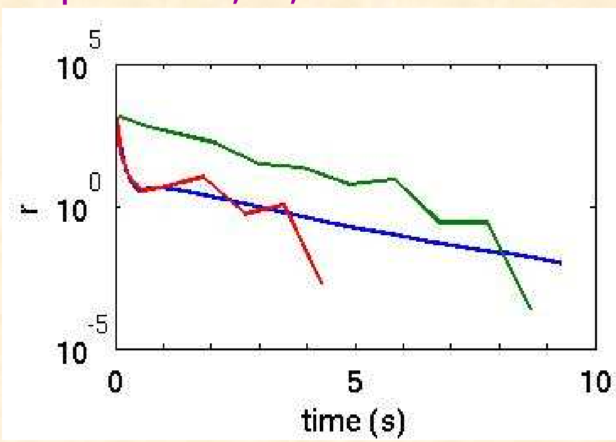


- If stopping r is small enough, RGA is faster.
- For high dimensions RρR can be faster for larger r.

RGA & RρR Cooperation

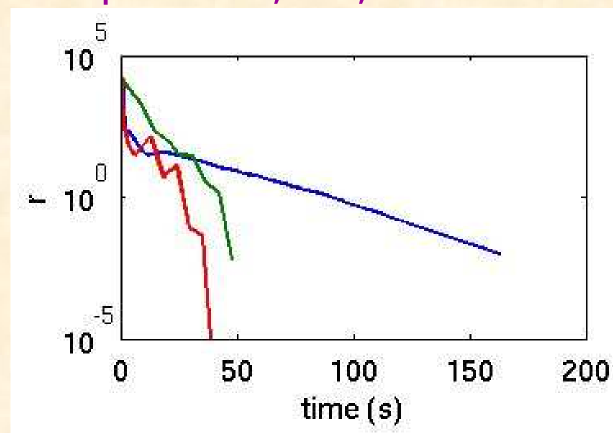
- Use RρR for time equal to one RGA iteration, then switch to RGA.

10 photons, 2,000 measurements

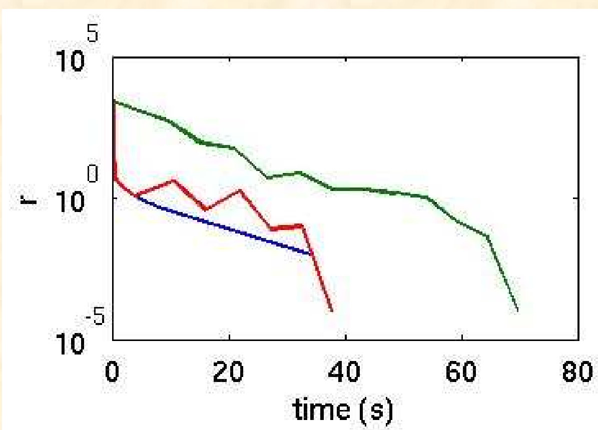


RρR —
RGA —
Coop. —

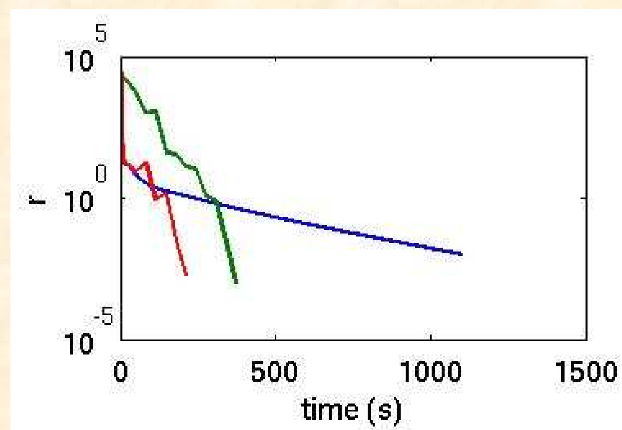
10 photons, 20,000 measurements



20 photons, 2,000 measurements



20 photons, 2,0000 measurements



Parametric Bootstrap

1. Use experimental data $\{(x_m, \phi_m)|m=1\ldots N\}$ to find ρ_{ML}.

2. Use ρ_{ML} to simulate B new data sets, each of which uses the same $\{\phi_m|m=1\ldots N\}$. For each ϕ_m, sample from

   $$P(x|\phi_m) = \mathrm{Tr}\left[\Pi(x,\phi_m)\rho_{\mathrm{ML}}\right]$$

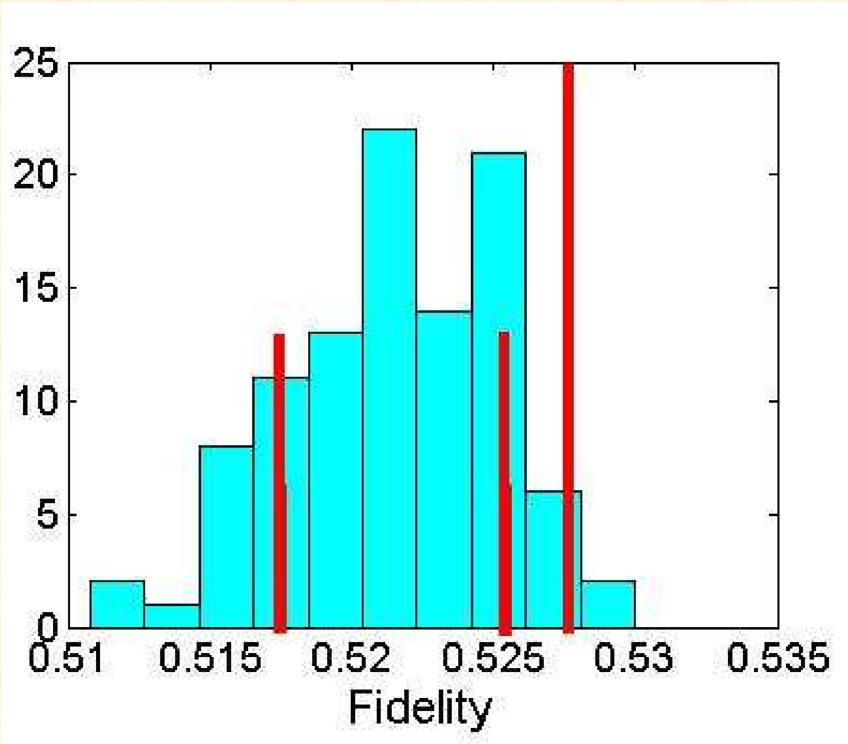
3. For each simulated data set, infer $\{\rho_{\mathrm{ML}}{}^{(b)}|b=1\ldots B\}$.

4. Use $\{\rho_{\mathrm{ML}}{}^{(b)}|b=1\ldots B\}$ to calculate parameter of interest $F(\rho)$.

5. Obtain distribution of $F^{(b)}=F(\rho_{\mathrm{ML}}{}^{(b)})$.

Parametric Bootstrap

- Resampling for fidelity with ideal cat state:



subtracting 1 photon

324,510 measurements

B=100 data sets

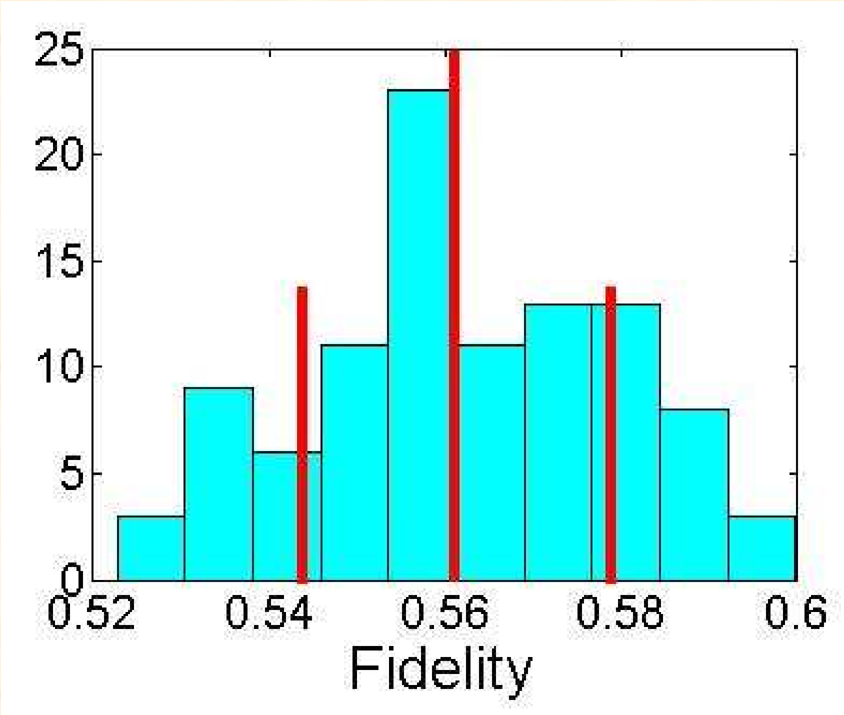
Long red line is maximum likelihood.

Shorter red lines mark central 68 percentile.

- Resampling is biased toward lower fidelity with pure state.
- We also see bias toward less negative Wigner function values.

Parametric Bootstrap

- Resampling for fidelity with ideal cat state:



subtracting 2 photons

41,223 measurements

B=100 data sets

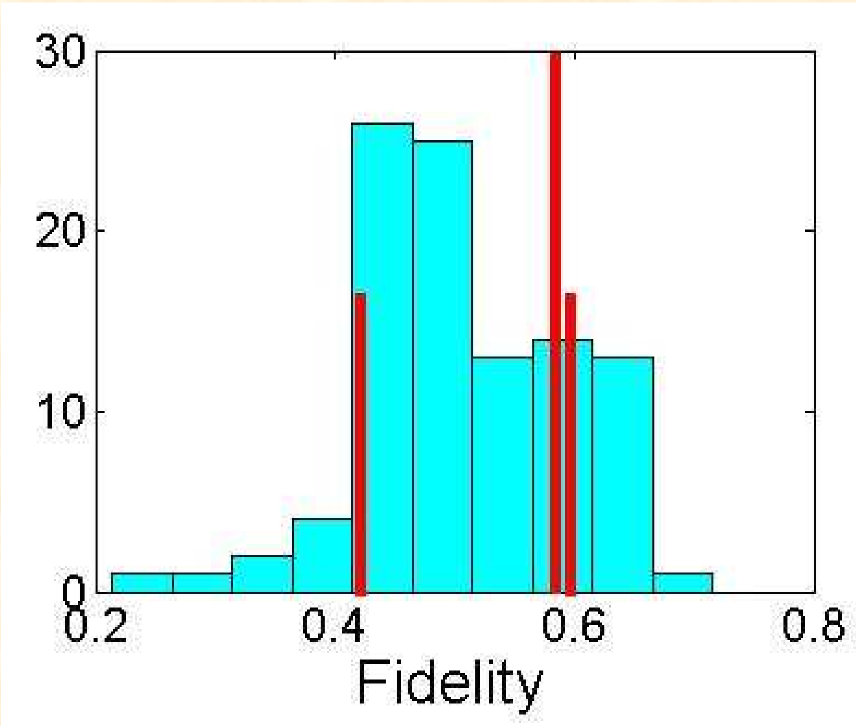
Long red line is maximum likelihood.

Shorter red lines mark central 68 percentile.

- Resampling is usually biased toward lower fidelity with pure state.
- We also see bias toward less negative Wigner function values.

Parametric Bootstrap

- Resampling for fidelity with ideal cat state:



subtracting 3 photons

1087 measurements

B = 100 data sets

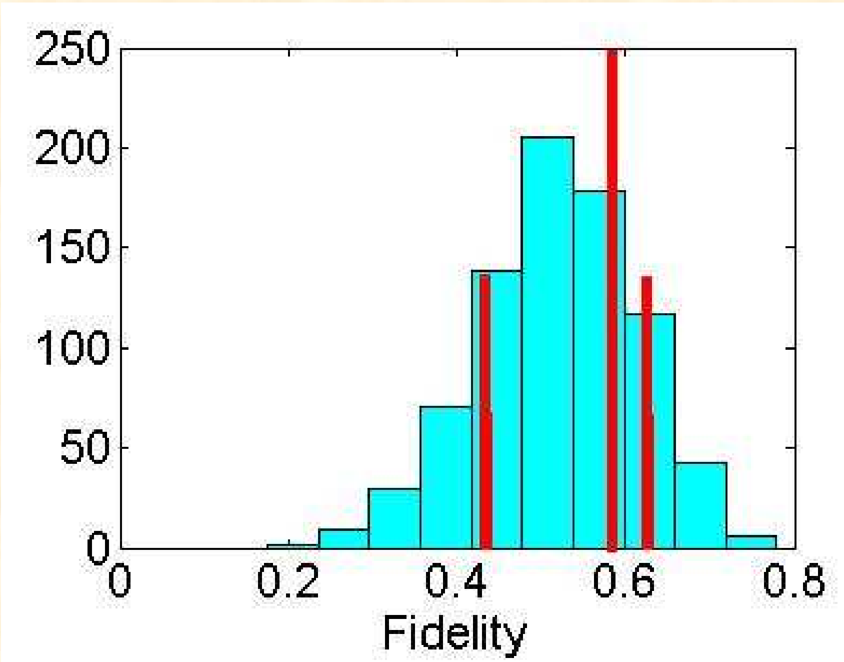
Central red line is maximum likelihood.

Outer red lines mark central 68 percentile.

- Resampling is biased toward lower fidelity with pure state.
- We also see bias toward less negative Wigner function values.

Parametric Bootstrap

- Resampling for fidelity with ideal cat state:



subtracting 3 photons

1087 measurements

B = 800 data sets

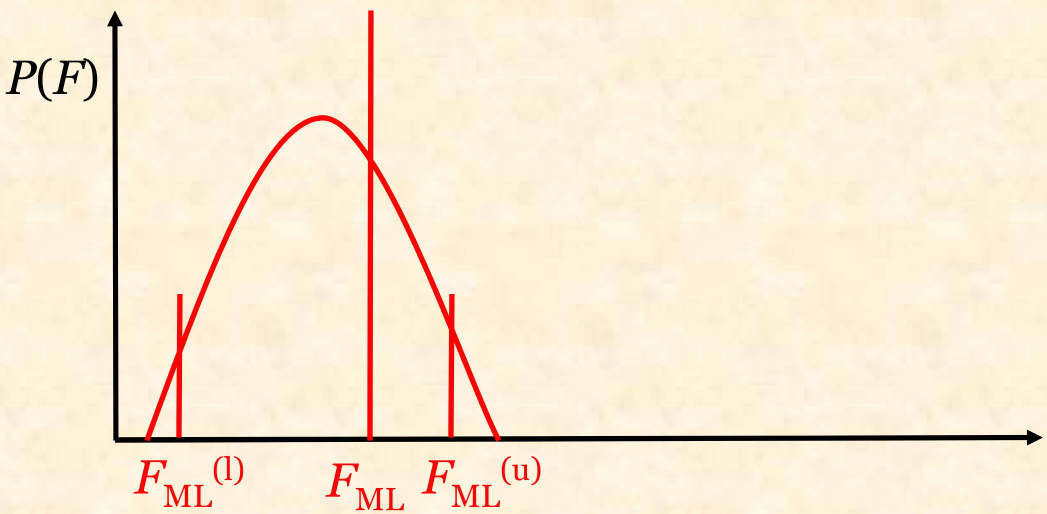
Central red line is maximum likelihood.

Outer red lines mark central 68 percentile.

- Resampling is biased toward lower fidelity with pure state.
- We also see bias toward less negative Wigner function values.

Bias Correcting Parametric Bootstrap

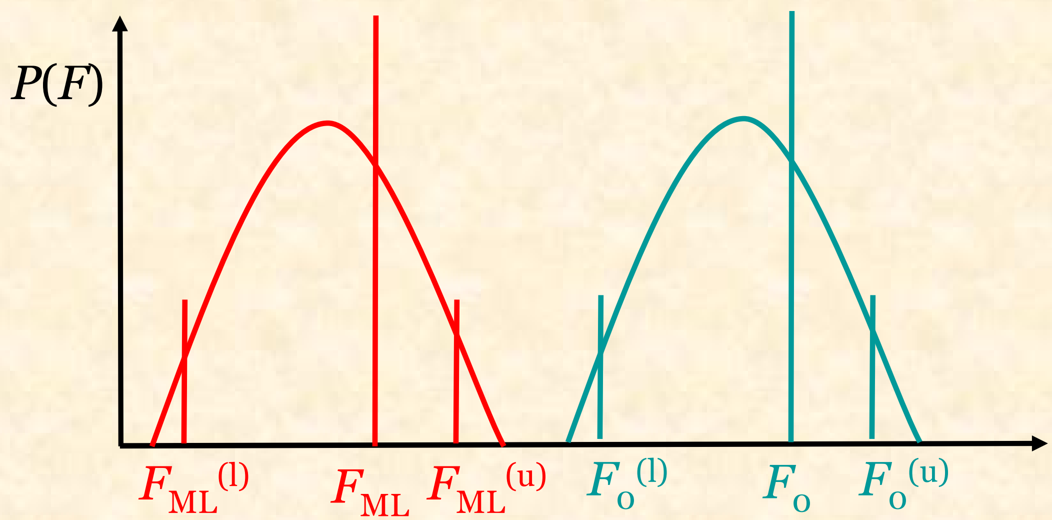
- Can we correct for the bias?
- Given: $F(\rho_{ML})=F_{ML}$, $P(F|\rho_{ML})$, $F_{ML}^{(l)}$, $F_{ML}^{(u)}$



Efron
Canadian J. Statistics ***9****, 139*
J. Amer. Statist. Assoc. ***82****, 171*

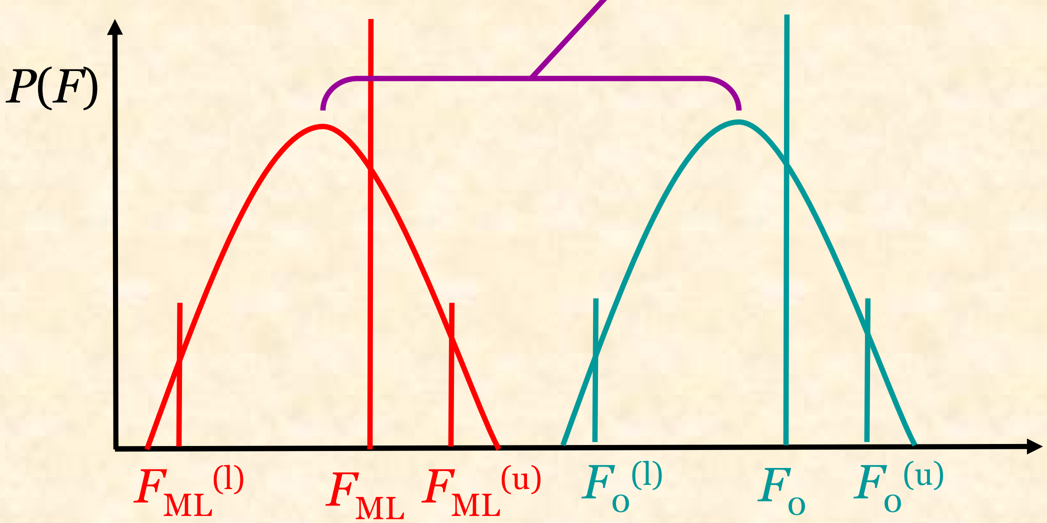
Bias Correcting Parametric Bootstrap

- Can we correct for the bias?
- Given: $F(\rho_{\mathrm{ML}})=F_{\mathrm{ML}}$, $P(F|\rho_{\mathrm{ML}})$, $F_{\mathrm{ML}}^{(l)}$, $F_{\mathrm{ML}}^{(u)}$
- Hypothesize ρ_{o}, a candidate for the true state ρ_{T}.
- Imagine $F(\rho_{\mathrm{o}})=F_{\mathrm{o}}$, $P(F|\rho_{\mathrm{o}})$, $F_{\mathrm{o}}^{(l)}$, $F_{\mathrm{o}}^{(u)}$.



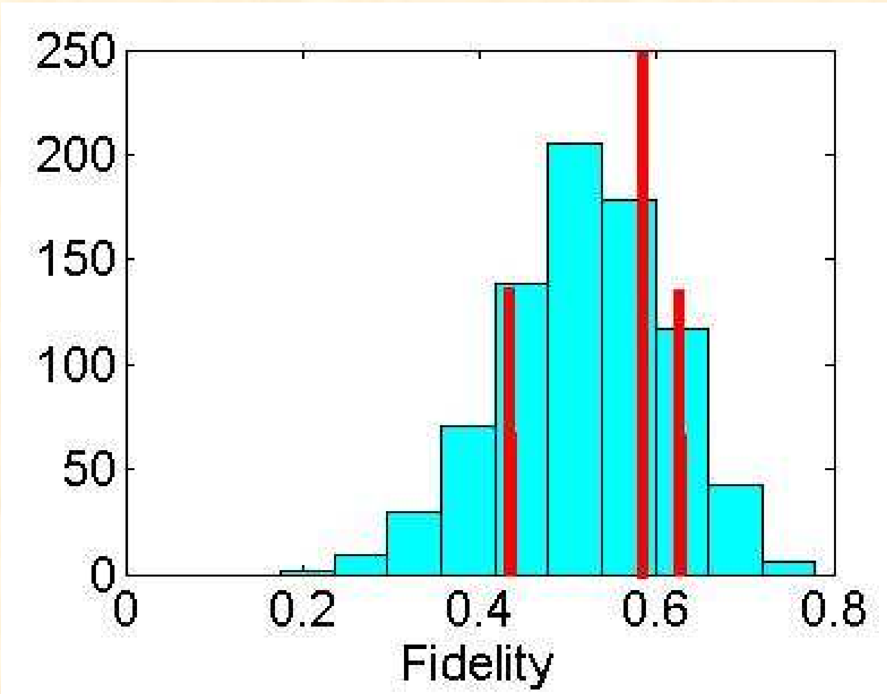
Bias Correcting Parametric Bootstrap

- Can we correct for the bias?
- Given: $F(\rho_{ML})=F_{ML}$, $P(F|\rho_{ML})$, $F_{ML}^{(l)}$, $F_{ML}^{(u)}$
- Hypothesize ρ_o, a candidate for the true state ρ_T.
- Imagine $F(\rho_o)=F_o$, $P(F|\rho_o)$, $F_o^{(l)}$, $F_o^{(u)}$.
- Assume $P(F|\rho_o) = P(F\text{-}f_o|\rho_{ML})$.



Parametric Bootstrap

- Resampling for fidelity with ideal cat state:



subtracting 3 photons

1087 measurements

B = 800 data sets

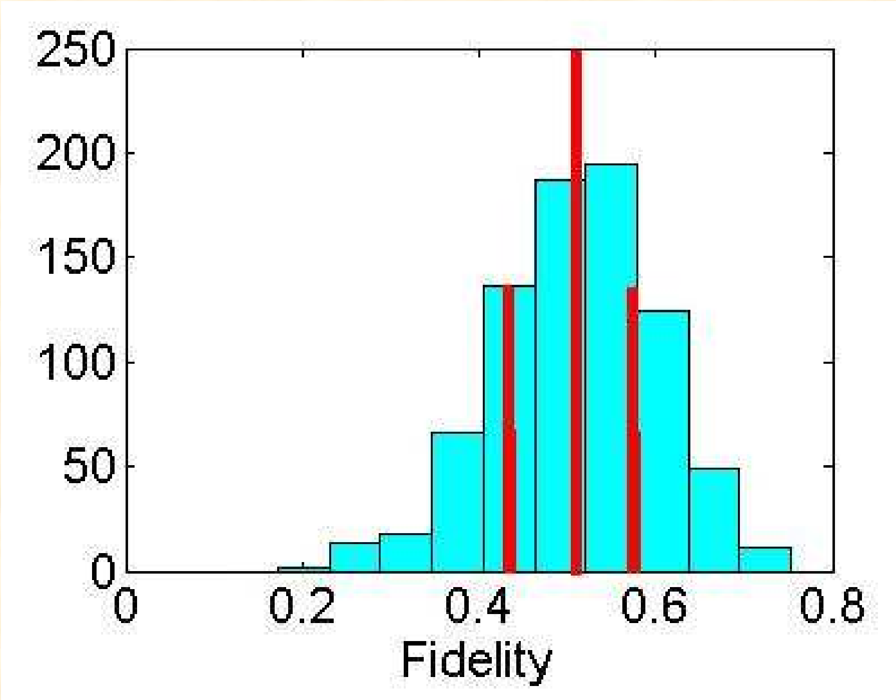
Central red line is maximum likelihood.

Outer red lines mark central 68 percentile.

- Resampling is biased toward lower fidelity with pure state.
- We also see bias toward less negative Wigner function values.

Parametric Bootstrap

- Resampling from state "close" to maximum likelihood



subtracting 3 photons

1087 measurements

B = 800 data sets

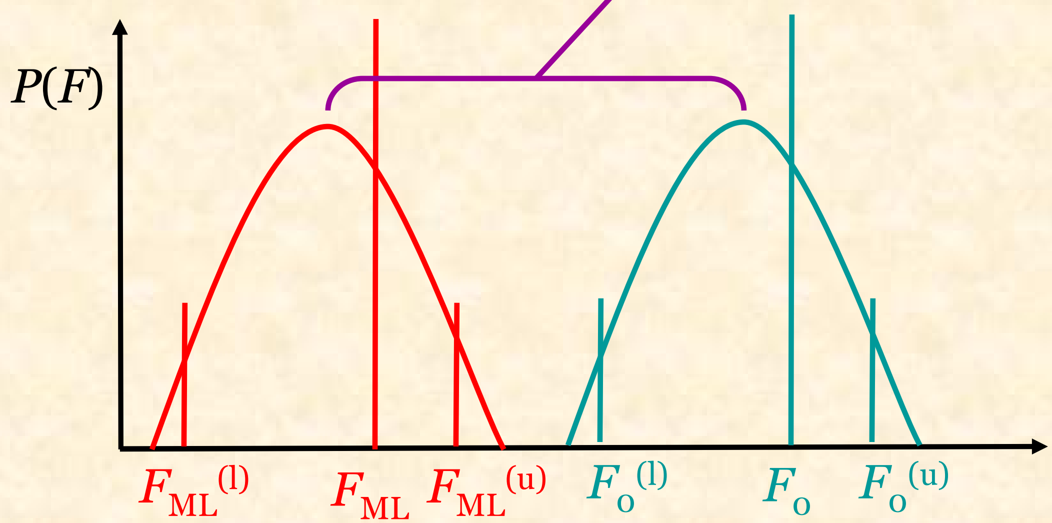
Central red line is fidelity of state used to generate data.

Outer red lines mark central 68 percentile.

- Histograms look similar, but clearly $P(F|\rho_o) = P(F\text{-}f_o|\rho_{ML})$ is not exactly true.

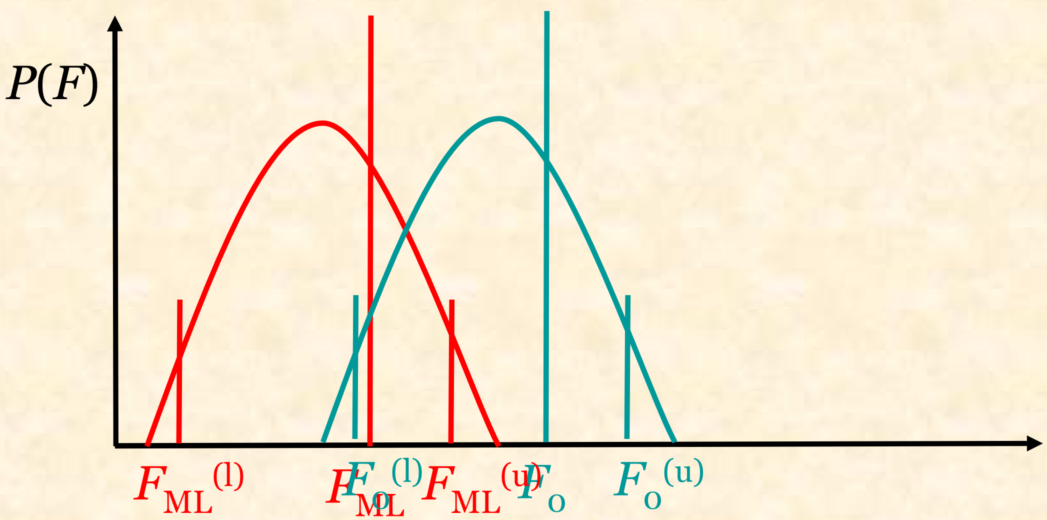
Bias Correcting Parametric Bootstrap

- Can we correct for the bias?
- Given: $F(\rho_{\mathrm{ML}})=F_{\mathrm{ML}}$, $P(F|\rho_{\mathrm{ML}})$, $F_{\mathrm{ML}}^{(l)}$, $F_{\mathrm{ML}}^{(u)}$
- Hypothesize ρ_{o}, a candidate for the true state ρ_{T}.
- Imagine $F(\rho_{\mathrm{o}})=F_{\mathrm{o}}$, $P(F|\rho_{\mathrm{o}})$, $F_{\mathrm{o}}^{(l)}$, $F_{\mathrm{o}}^{(u)}$.
- Assume $P(F|\rho_{\mathrm{o}}) = P(F\text{-}f_o|\rho_{\mathrm{ML}})$.



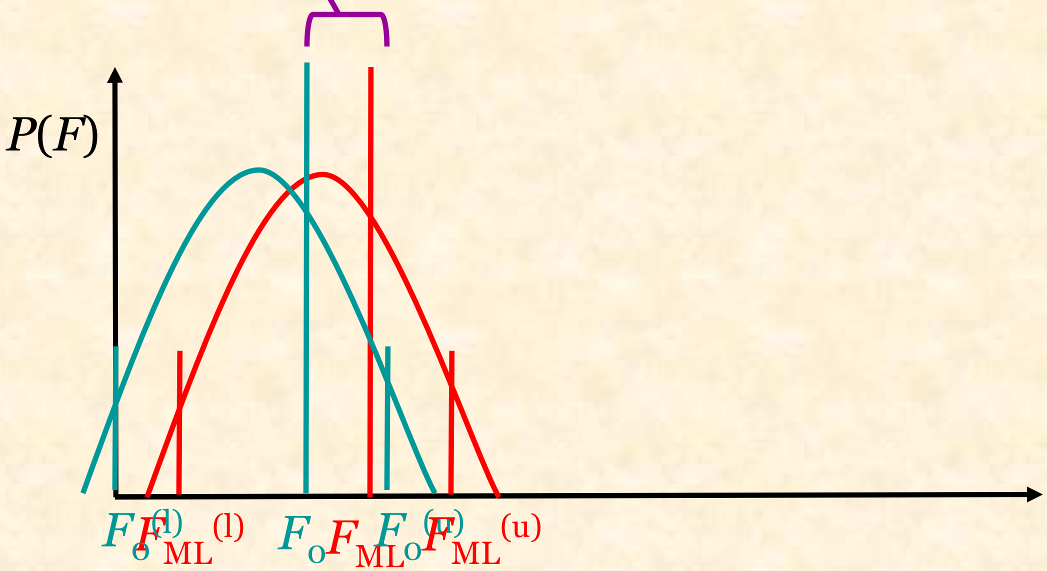
Bias Correcting Parametric Bootstrap

- Can we correct for the bias?
- Given: $F(\rho_{\mathrm{ML}})=F_{\mathrm{ML}}$, $P(F|\rho_{\mathrm{ML}})$, $F_{\mathrm{ML}}^{(l)}$, $F_{\mathrm{ML}}^{(u)}$
- Hypothesize ρ_{o}, a candidate for the true state ρ_{T}.
- Imagine $F(\rho_{\mathrm{o}})=F_{\mathrm{o}}$, $P(F|\rho_{\mathrm{o}})$, $F_{\mathrm{o}}^{(l)}$, $F_{\mathrm{o}}^{(u)}$.
- Assume $P(F|\rho_{\mathrm{o}}) = P(F\text{-}f_o|\rho_{\mathrm{ML}})$.
- If ρ_o is a good hypothesis, $F_{\mathrm{o}}^{(l)}<F_{\mathrm{ML}}<F_{\mathrm{o}}^{(u)}$



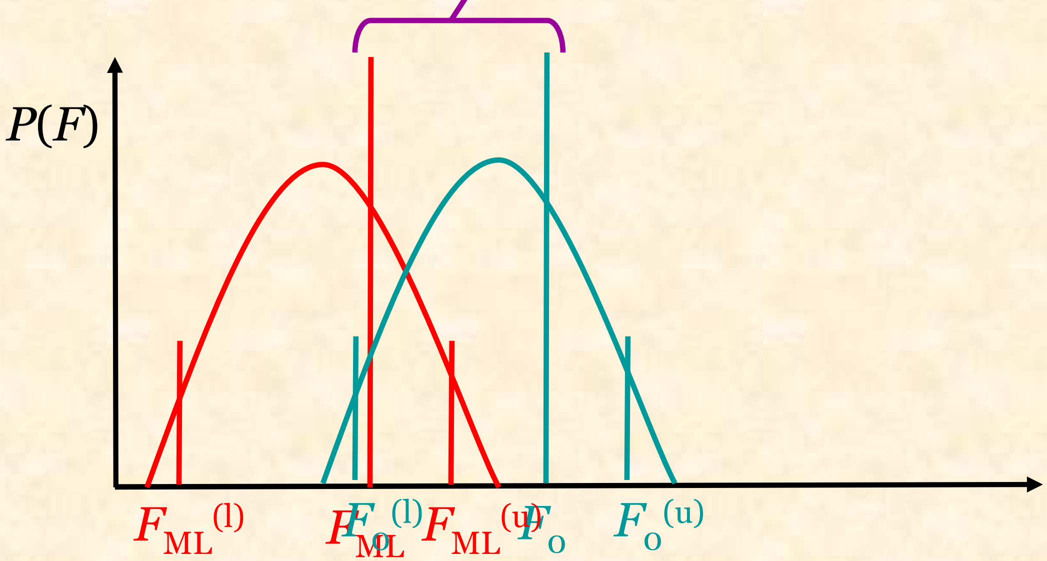
Bias Correcting Parametric Bootstrap

- What are the allowed locations for F_o, such that $F_o^{(l)}<F_{ML}<F_o^{(u)}$?
- $F_{ML}+F_o-F_o^{(u)}<F_o$



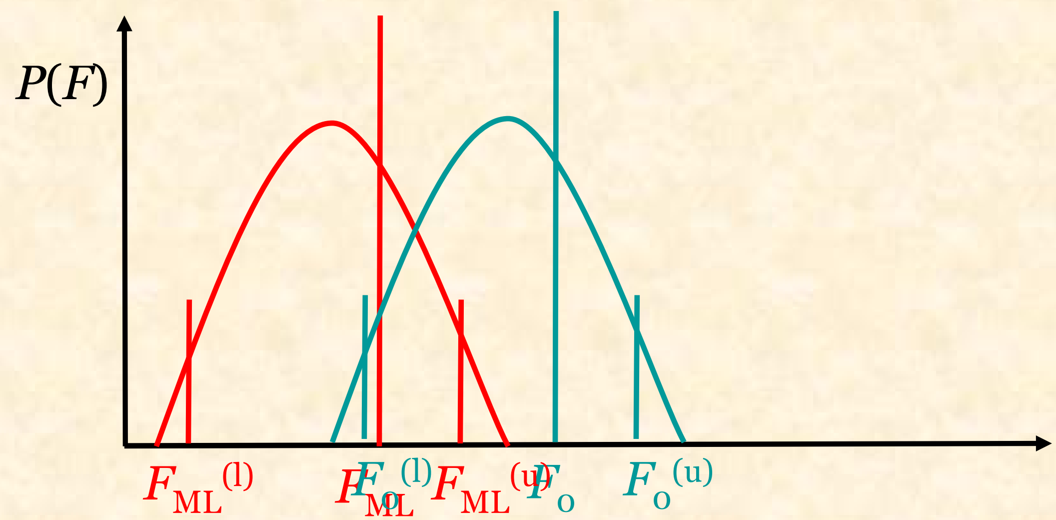
Bias Correcting Parametric Bootstrap

- What are the allowed locations for F_o, such that $F_o^{(l)}<F_{ML}<F_o^{(u)}$?
- $F_{ML}+F_o-F_o^{(u)}<F_o<F_{ML}+F_o-F_o^{(l)}$



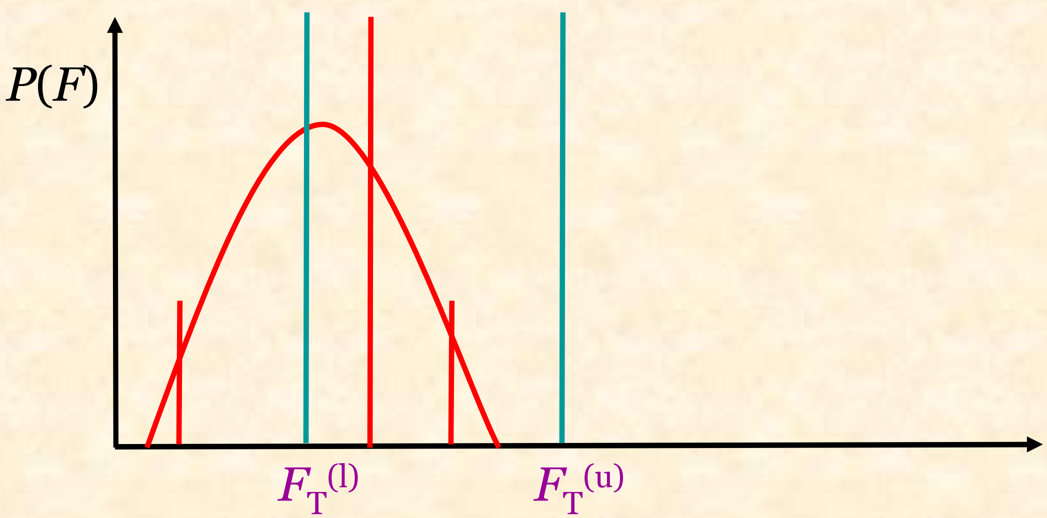
Bias Correcting Parametric Bootstrap

- What are the allowed locations for F_o, such that $F_o^{(l)}<F_{ML}<F_o^{(u)}$?
- $F_{ML}+F_o-F_o^{(u)}<F_o<F_{ML}+F_o-F_o^{(l)}$
- Because $P(F|\rho_o) = P(F\text{-}f_o|\rho_{ML})$,
  - $F_o-F_o^{(u)}=F_{ML}-F_{ML}^{(u)}$
  - $F_o-F_o^{(l)}=F_{ML}-F_{ML}^{(l)}$
- $\therefore\ 2F_{ML}-F_{ML}^{(u)}<F_o<2F_{ML}-F_{ML}^{(l)}$

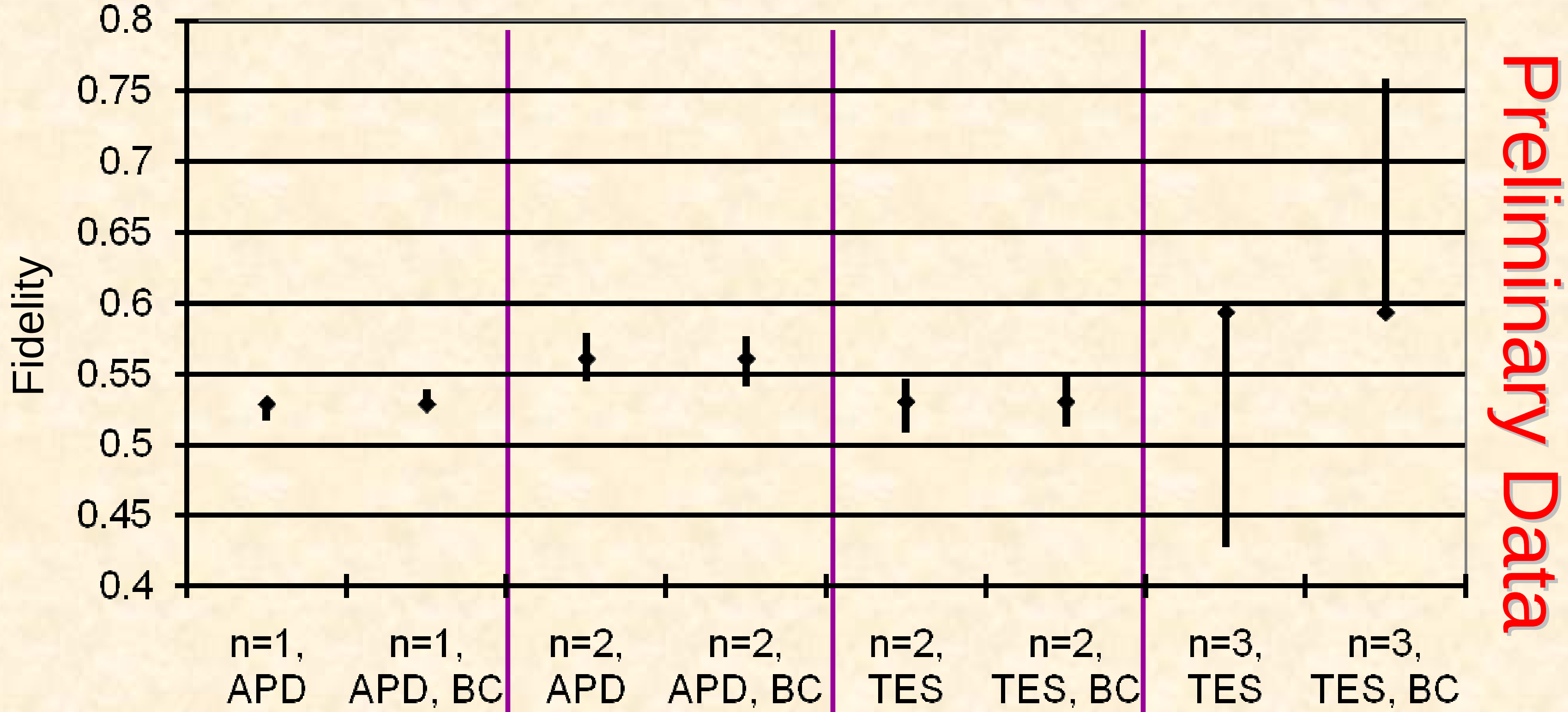


Bias Correcting Parametric Bootstrap

- What are the allowed locations for F_o, such that $F_o^{(l)}<F_{ML}<F_o^{(u)}$?
- $F_{ML}+F_o-F_o^{(u)}<F_o<F_{ML}+F_o-F_o^{(l)}$
- Because $P(F|\rho_o) = P(F\text{-}f_o|\rho_{ML})$,
  - $F_o-F_o^{(u)}=F_{ML}-F_{ML}^{(u)}$
  - $F_o-F_o^{(l)}=F_{ML}-F_{ML}^{(l)}$
- $\therefore\ 2F_{ML}-F_{ML}^{(u)}<F_o<2F_{ML}-F_{ML}^{(l)}$
- $F_T^{(l)} = 2F_{ML}-F_{ML}^{(u)}$, $F_T^{(u)} = 2F_{ML}-F_{ML}^{(l)}$

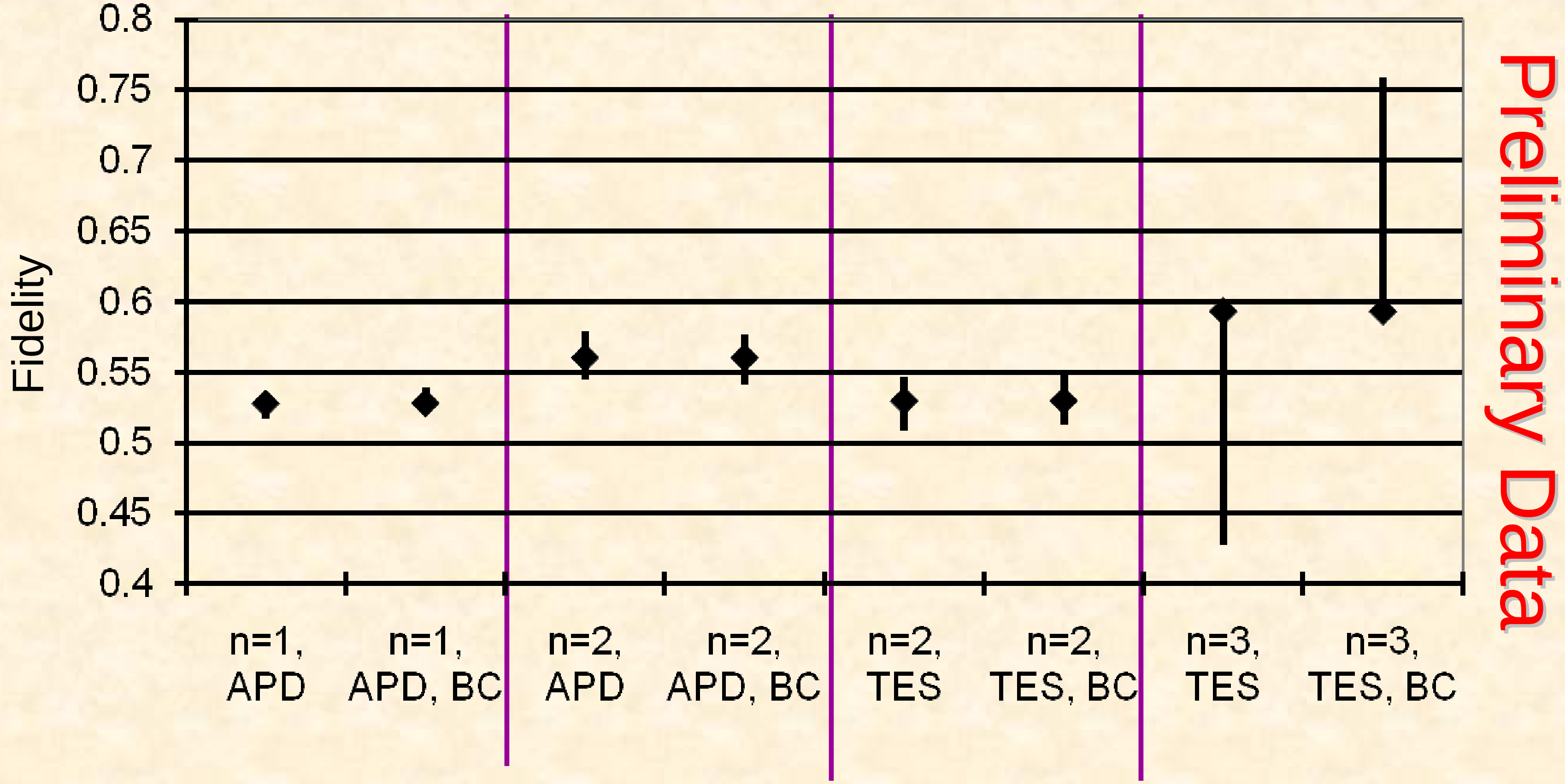


Bias Correcting Our Data



	n=1, APD	n=2, APD	n=2, TES	n=3, TES
N	324,510	41,223	24,790	1,087
Cat $\langle n \rangle$	1.88	1.96	1.24	3.10

Bias Correcting Our Data



- We must also include calibration uncertainty η~85%±3%.
- Choosing η=82% or 88% shifts F_{ML} by ~1.5%.
- So, I have increased the size of the data squares.

Conclusions

- MaxLikelihood stopping criterion:

  $$r = \max(\mathrm{eig}(R(\rho))) - N$$

  - bounds likelihood: $\mathcal{L}_{\max} \leq \mathcal{L}(\rho) + r$
- Regularized Gradient Ascent maximization algorithm.
  - faster convergence, but may not be practical in all cases
  - can optimize any convex function of ρ.
- Parametric bootstrap resampling with bias correction
  - correct low-purity bias of MaxLikelihood inference.
  - requires strong assumptions.
- Created approximate cat states by subtracting 3 photons.
  - $\langle n \rangle$ is fairly large, but fidelity needs improvement
  - requires higher purity squeezing

Open Problems

- More rigorous confidence intervals that don't require resampling.
  - Ask me about what we have tried that didn't work, and
  - ideas we have for achieving this.
- Test for number of photons required in density matrix
  - too many photons may cause "over fitting" problems.
- Fast method for Bayesian inference of ρ.
- How to make high purity, single mode, pulsed, squeezed light.