

Title: Quantum Estimation via Convex Optimization

Date: Aug 27, 2008 09:30 AM

URL: <http://pirsa.org/08080045>

Abstract: A number of problems in quantum estimation can be formulated as a convex optimization [1]. Applications include: maximum likelihood estimation, optimal experiment design, quantum state detection, and quantum metrology under instrumentation constraints. This talk will draw on the work I have been involved with, e.g., [2], [3], [4]. Our work in optimal quantum error correction [5, 6] is also relevant. Great benefit is derived using an error model which is specific to the system. Obtaining the errors from tomography is a logical route. How to do this, however, is an open question. The constraint is the form required by the standard error-correction model upon which the optimization is constructed. I will present some ideas on how to do the tomography in this context. • Maximum Likelihood (ML) quantum estimation problems are easily formed as log-convex optimization problems [1]. These include estimation of the state (density), estimation of the distribution of known input states, estimation of the OSR elements for quantum process tomography, and estimation of the coefficients of a preselected basis set of OSR elements. Estimation of Hamiltonian parameters, unfortunately, is not a convex optimization. Associated with these estimation problems, including Hamiltonian parameter estimation, is an optimal experiment design (OED), which is convex, and which can determine the system configurations to maximize the estimation accuracy [2]. Experiments have been performed In Ian Walmsley's Group at Oxford using these methods [7, 8]. • Quantum state detection can be formulated as a convex optimization problem in the matrices of the POVM which characterize the measurement apparatus. Minimizing the error probability is a semidefinite program (SDP) [9]. Maximizing the posterior probability of detection is a quasiconvex optimization problem [3]. • Quantum metrology subject to instrumentation constraints can be cast as a convex optimization problem [4]. Focusing on the single parameter case, the optimization problem is a linear program (LP). The Fisher information from the LP solution for the constrained problem can be compared to what is possible with no constraints, the Quantum Fisher Information. This approach is easily extended to the multi-parameter case. • Quantum Error Correction (QEC) that is optimized with respect to the specific system at hand can reduce ancilla overhead while raising error thresholds for fault-tolerant operation [5, 6, 10, 11]. The problem is cast as a bi-convex optimization problem, iterating between encoding and recovery, each being an SDP. In [5] we introduced two new aspects of this approach: (i) we modified the objective functions to account for robustness, and (ii) posed the problem in an indirect form which can be solved via a sequence of constrained least-squares problems. This opens the way for solving extremely large problems in a reasonable time period both from offline models and online from measured data, i.e., tomography.

Quantum Estimation via Convex Optimization

Robert Kosut
SC Solutions, Sunnyvale, CA

collaborators

H. Rabitz, C. Brif, M. Grace
Princeton

I. Walmsley, M. Branderhorst
Oxford

D. Lidar, A. Shabani
USC

E. Yablonovitch, R. Caflisch
UCLA

Research supported by the DARPA QuIST Program
(Quantum Information Science & Technology)

Quantum Estimation via Convex Optimization

Robert Kosut
SC Solutions, Sunnyvale, CA

collaborators

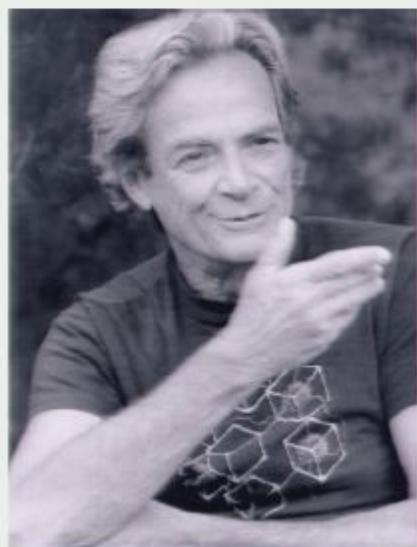
H. Rabitz, C. Brif, M. Grace
Princeton

I. Walmsley, M. Branderhorst
Oxford

D. Lidar, A. Shabani
USC

E. Yablonovitch, R. Caflisch
UCLA

Research supported by the DARPA QuIST Program
(Quantum Information Science & Technology)



“Quantum Mechanical Computers,”
Optics News, February 1985.

Richard P. Feynman

“In a machine such as this there are very many other problems due to imperfections. . . . there may be small terms in the Hamiltonian besides the ones we've written. . . . At least some of these problems can be remedied in the usual way by techniques such as error correcting codes . . . But until we find a specific implementation for this computer, I do not know how to proceed to analyze these effects. However, it appears that they would be very important in practice. This computer seems to be very delicate and these imperfections may produce considerable havoc.”

On Learning (a.k.a. “estimating”)

SEEKER “How do I gain **Good Judgement?**”

On Learning (a.k.a. “estimating”)

SEEKER “How do I gain **Good Judgement**? ”

SAGE “You must first acquire **Wisdom**. ”

On Learning (a.k.a. “estimating”)

SEEKER “How do I gain **Good Judgement?**”

SAGE “You must first acquire **Wisdom.**”

SEEKER “How do I gain **Wisdom?**”

On Learning (a.k.a. “estimating”)

SEEKER “How do I gain **Good Judgement?**”

SAGE “You must first acquire **Wisdom.**”

SEEKER “How do I gain **Wisdom?**”

SAGE “From **Bad Judgement.**”

On Learning (a.k.a. “estimating”)

SEEKER “How do I gain **Good Judgement**? ”

SAGE “You must first acquire **Wisdom**. ”

SEEKER “How do I gain **Wisdom**? ”

SAGE “From **Bad Judgement**. ”



First rule of learning: **Make Mistakes**

the following can be cast as convex optimization problems

- quantum state tomography
- quantum error correction
- quantum process tomography
- optimal experiment design
- quantum state detection

the following are generally *not* convex optimization problems

- Hamiltonian parameter estimation
 - optimal experiment design **is** a cvx opt
- control design
 - is there analogous optimal experiment design for control?

Convex Optimization

S. Boyd & L. Vandenberghe, *Convex Optimization*, Cambridge, 2004

Supporting Lecture Slides and Course Videos

<http://stanford.edu/boyd>

$$\begin{array}{ll}\text{minimize} & f(x) \\ \text{subject to} & x \in C\end{array}$$

if $f(x)$ is a convex function, and C is a convex set, then:

- any local solution is global.
- interior point method achieves the global optimum to within desired accuracy
- duality theory yields optimality conditions
- non-convex constraints can often be relaxed to convex constraints, thus always providing a lower bound, and sometimes an upper bound.
- often are unrecognized
- MATLAB based (free) software available, e.g.,
 - complilers: YALMIP (J. Lofberg, ETH), CVX (M. Grant, S. Boyd, Y. Ye, Stanford)
 - solvers: SEDUMI, SDPT3, ...

computation time (roughly)

least-squares

$$\begin{array}{ll} \text{minimize} & \|Ax - b\| \\ \text{cost} & n^2N \quad (A \in \mathbf{R}^{N \times n}) \end{array}$$

linear program

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \leq b \\ \text{cost} & n^2N \end{array}$$

convex optimization

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \dots, N \\ \text{cost} & \max \{n^3, n^2N, K\} \\ & K \text{ is cost of } f_i, \nabla f_i, \nabla^2 f_i \end{array}$$

- less cost with structure

Convexity & Quantum Mechanics

examples of convex sets

probability outcomes

$$\{p_i \in \mathbf{R}\}$$

$$\sum_i p_i = 1, \quad p_i \geq 0$$

density matrix

$$\{\rho \in \mathbf{C}^{n \times n}\}$$

$$\text{Tr } \rho = 1, \quad \rho \geq 0$$

positive operator
valued measure (POVM)

$$\{O_i \in \mathbf{C}^{n \times n}\}$$

$$\sum_i O_i = I_n, \quad O_i \geq 0$$

process matrix

system representation
in fixed basis

$$B_\alpha \in \mathbf{C}^{n \times n}, \quad \alpha = 1, \dots, n^2$$

$$\rho_{\text{out}} = \sum_{\alpha, \beta} X_{\alpha \beta} B_\alpha \rho_{\text{in}} B_\beta^\dagger$$

$$\{X \in \mathbf{C}^{n^2 \times n^2}\}$$

$$\sum_{\alpha, \beta} X_{\alpha \beta} B_\alpha^\dagger B_\beta = I_n, \quad X \geq 0$$

examples of non-convex sets

pure state $\left\{ \rho = |\psi\rangle\langle\psi| \mid \langle\psi|\psi\rangle = 1 \right\}$

Hamiltonian parameter estimation $\left\{ \rho(T, \theta) \mid H(t, \theta) = H_0(t) + \sum_k \theta_k H_k(t), \quad 0 \leq t \leq T \right\}$

Control design $\left\{ \rho(T \mid v(\cdot)) \mid H(t) = H_0 + \sum_k v_k(t) H_k, \quad 0 \leq t \leq T \right\}$

- not every non-convex problem cannot be efficiently solved
- these latter two, in general, can have many local optimal points, which are often quite good^{a b}

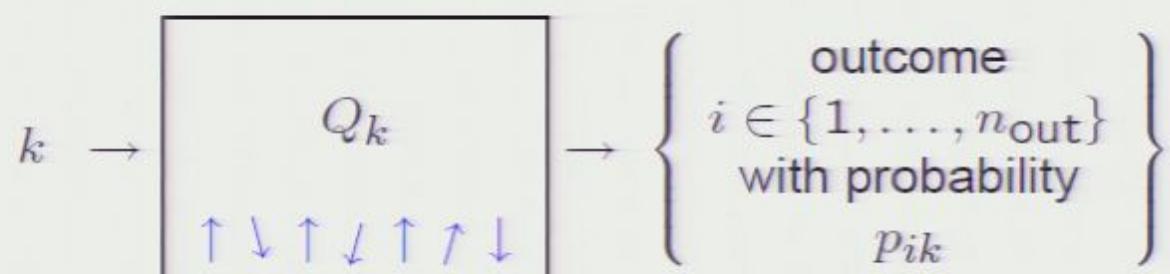
^aRabitz, Hsieh, Rosenthal. Quantum optimally controlled transition landscapes, *Science*, 2004

^bRabitz, Ho, Kosut, Demiralp. Topology of optimally controlled quantum mechanical transition probability landscapes, *Phys. Rev. A*, 2006
Pirsa: 08080045 Page 15/75

Quantum State Tomography

Collecting data from independent experiments

- with system in configuration k , repeat experiment ℓ_k times.
 - $k \in \{1, \dots, n_{\text{cfg}}\}$
 - configurations include: prepared initial states, sample times, frequencies, optical settings, applied control fields, ...
- record outcome counts n_{ik}
 - number of times outcome i occurred in configuration k
 - $i \in \{1, \dots, n_{\text{out}}\}$
- repeat ℓ_k times:



State tomography via Maximum Likelihood (ML) estimation

$$\text{minimize} \quad L(\rho) = - \sum_{i,k} n_{ik} \log p_{ik}(\rho)$$

$$\text{subject to} \quad p_{ik}(\rho) = \text{Tr } O_{ik} \rho, \quad \rho \geq 0, \quad \text{Tr } \rho = 1$$

- $L(\rho)$ – negative log-likelihood function
- n_{ik} – outcome counts per configuration (data)
- O_{ik} – system model (POVM)
- convex in ρ

State tomography via Least-Squares (LS) estimation

$$\text{minimize } V(\rho) = \sum_{i,k} (p_{ik}^{\text{emp}} - p_{ik}(\rho))^2$$

$$\text{subject to } p_{ik}(\rho) = \text{Tr } O_{ik} \rho, \quad \rho \geq 0, \quad \text{Tr } \rho = 1$$

- $V(\rho)$ – least-squares error
- $p_{ik}^{\text{emp}} = n_{ik}/\ell_k$ – empirical probability estimate (data)
- O_{ik} – system model (POVM)
- convex in ρ
 - not much less computational effort than ML
 - convex optimization for any norm of the error: $\|p^{\text{emp}} - p(\rho)\|$

noisy measurements

- replace POVM with

$$O_{ik} = \sum_{i'} p_{ii'}^{\text{noise}} O_{i'k}^{\text{noise-free}}$$

- $p_{ii'}^{\text{noise}}$ is probability of outcome i when noise-free outcome is i'
- rows and columns of p^{noise} both sum to 1
- if p^{noise} is incorrect, ML and LS remain convex optimizations, however, the estimate of ρ is biased.
- estimating ρ and p^{noise} separately are each convex, but not jointly
 - iterating often leads to good estimates

State tomography via Least-Squares (LS) estimation

$$\text{minimize } V(\rho) = \sum_{i,k} (p_{ik}^{\text{emp}} - p_{ik}(\rho))^2$$

$$\text{subject to } p_{ik}(\rho) = \text{Tr } O_{ik} \rho, \quad \rho \geq 0, \quad \text{Tr } \rho = 1$$

- $V(\rho)$ – least-squares error
- $p_{ik}^{\text{emp}} = n_{ik}/\ell_k$ – empirical probability estimate (data)
- O_{ik} – system model (POVM)
- convex in ρ
 - not much less computational effort than ML
 - convex optimization for any norm of the error: $\|p^{\text{emp}} - p(\rho)\|$

State tomography via Maximum Likelihood (ML) estimation

$$\text{minimize} \quad L(\rho) = - \sum_{i,k} n_{ik} \log p_{ik}(\rho)$$

$$\text{subject to} \quad p_{ik}(\rho) = \text{Tr } O_{ik} \rho, \quad \rho \geq 0, \quad \text{Tr } \rho = 1$$

- $L(\rho)$ – negative log-likelihood function
- n_{ik} – outcome counts per configuration (data)
- O_{ik} – system model (POVM)
- convex in ρ

State tomography via Least-Squares (LS) estimation

$$\text{minimize } V(\rho) = \sum_{i,k} (p_{ik}^{\text{emp}} - p_{ik}(\rho))^2$$

$$\text{subject to } p_{ik}(\rho) = \text{Tr } O_{ik} \rho, \quad \rho \geq 0, \quad \text{Tr } \rho = 1$$

- $V(\rho)$ – least-squares error
- $p_{ik}^{\text{emp}} = n_{ik}/\ell_k$ – empirical probability estimate (data)
- O_{ik} – system model (POVM)
- convex in ρ
 - not much less computational effort than ML
 - convex optimization for any norm of the error: $\|p^{\text{emp}} - p(\rho)\|$

noisy measurements

- replace POVM with

$$O_{ik} = \sum_{i'} p_{ii'}^{\text{noise}} O_{i'k}^{\text{noise-free}}$$

- $p_{ii'}^{\text{noise}}$ is probability of outcome i when noise-free outcome is i'
- rows and columns of p^{noise} both sum to 1
- if p^{noise} is incorrect, ML and LS remain convex optimizations, however, the estimate of ρ is biased.
- estimating ρ and p^{noise} separately are each convex, but not jointly
 - iterating often leads to good estimates

YALMIP code in MATLAB of (Idealized) Least-Squares Estimation

```
% load data: O - n x n x nout x ncfg, rho0 - n x n
% declare sdp variable
rho = sdpvar(n,n,'hermitian','complex');

% make true and model probabilities: p0(i,k) and p(i,k)
clear p0 p
for k=1:ncfg
    for i=1:nout
        p0(i,k) = trace(O(:,:,i,k)*rho0);
        p(i,k) = trace(O(:,:,i,k)*rho);
    end
end

obj = norm(p0-p,'fro');
clist = set( rho >= 0 ) + set( trace(rho) == 1 );

solvesdp(clist,obj);
```

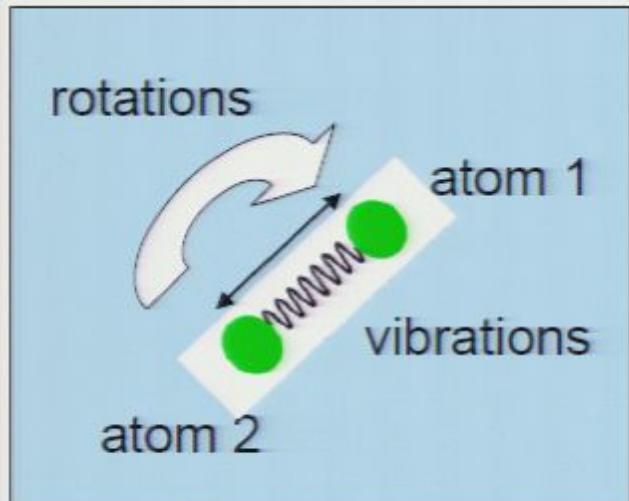
CPU times (2.4 Ghz, 2 GB RAM)

n_ρ	$n_{\text{real-par}}$	n_{out}	n_{cfg}	$t_{\text{cpu}} \text{ (sec)}$
2	3	2	500	0.35
			1000	0.38
4	15	2	500	0.40
			1000	0.53
8	63	2	500	1.00
			1000	1.75
8	63	4	500	1.40
			1000	2.4
16	255	4	500	10.10
			1000	15.5
20	399	4	500	20.1
			1000	33.2
32	1023	4	500	151.0

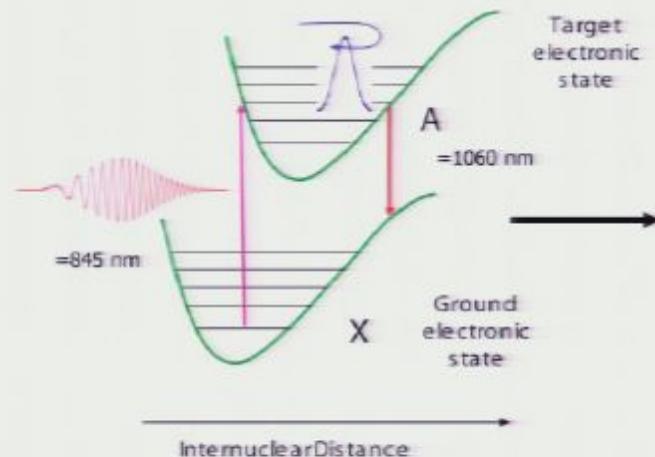
- randomly selected $\rho_{\text{true}}, O_{ik}$
- YALMIP calls the solver SDPT3
- t_{cpu} includes 10-20 SDP iterations (this is typical)

State tomography of vibrational wavepackets in diatomic molecules*

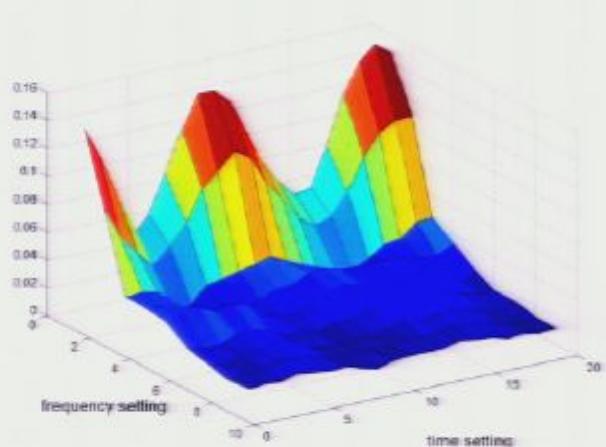
Diatomical molecular system



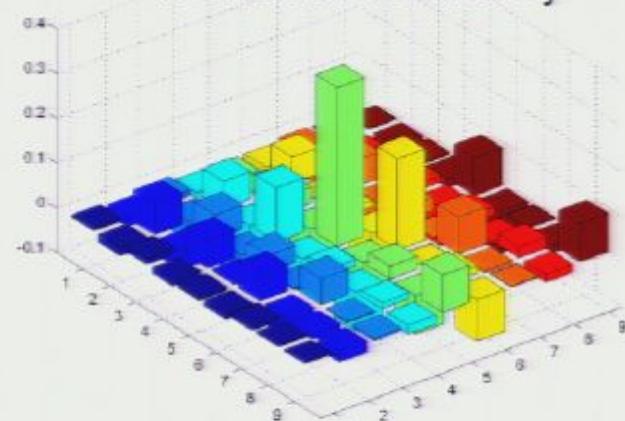
Electronic transition



Signal from fluorescence measurements of vibrations



Reconstructed density



$$P_{\alpha\gamma, \text{empirical}}$$

$$\alpha = 0, 1$$

$$\gamma = \Omega, T$$

$$\|p_{\alpha\gamma, \text{emp}} - p_{\alpha\gamma}(\rho_{\text{est}})\| \approx 0.01$$

adding prior information

- modify objective \Rightarrow maximize $\sum_{i,k} n_{ik} \log p_{ik}(\rho) + \log p_0(\rho)$
or add constraint $\Rightarrow \rho \in D$
- if each element of ρ is uniformly distributed, then add constraint
$$\left\{ \rho \mid \|W^{-1}\text{vec}(\rho - \rho_0)\|_\infty \leq 1 \right\}$$
- if each element of ρ is normally distributed, then add to objective,
$$\log p_0(\rho) \sim \|W^{-1/2}\text{vec}(\rho - \rho_0)\|_2^2$$

W is prior covariance matrix
- convex optimization in θ if $\rho(\theta)$ is affine in θ

Quantum Process Tomography

The characterization of dynamics of open quantum systems

- a fundamental problem in quantum information science and coherent control
 - for verifying the performance of an information-processing device
 - for design of decoherence prevention/correction methods

OSR → Process Matrix

- Expand $n \times n$ OSR elements $\{A_\mu, \mu = 1, \dots, \ell\}$ in a fixed basis

$$A_\mu = \sum_{\alpha=1}^{n^2} \mathbf{x}_{\mu\alpha} B_\alpha, \quad B_\alpha \in \mathbb{C}^{n \times n}$$

- Process matrix

$$\begin{aligned} (\mathbf{X})_{\alpha\beta} &= \sum_{\mu=1}^{\ell} \mathbf{x}_{\mu\alpha} \mathbf{x}_{\mu\beta}^* \\ \sum_{\alpha,\beta} \mathbf{X}_{\alpha\beta} B_\alpha^\dagger B_\beta &= I_n \end{aligned} \implies \left\{ \begin{array}{l} \mathbf{X} \in \mathbb{C}^{n^2 \times n^2} \\ \text{quadratic in } \mathbf{x}_\mu \in \mathbb{C}^{n^2} \\ \text{positive semidefinite} \\ n^4 - n^2 \text{ real parameters} \end{array} \right.$$

Process Matrix → (equivalent) OSR

$$\text{SVD of } \mathbf{X} = V S V^\dagger \implies A'_\mu = \sum_{\alpha=1}^{n^2} \sqrt{s_\mu} V_{\alpha\mu} B_\alpha, \quad \mu = 1, \dots, n^2$$

- number of numerically non-zero $s_\mu = \text{rank } \mathbf{X} \leq n^2$; often $\ll n^2$

Relax quadratic equality constraint to $\mathbf{X} \geq 0$

ML Process Tomography

$$\text{minimize } L(\mathbf{X}) = - \sum_{i,k} n_{ik} \log p_{ik}(\mathbf{X})$$

$$\text{subject to } \sum_{\alpha,\beta} \mathbf{X}_{\alpha\beta} B_\alpha^\dagger B_\beta = I_n, \quad \mathbf{X} \geq 0$$

$$p_{ik}(\mathbf{X}) = \text{Tr } \mathbf{X} R_{ik}, \quad (R_{ik})_{\alpha\beta} = \text{Tr } B_\alpha \rho_k B_\beta^\dagger O_{ik}$$

- convex optimization in $\mathbf{X} \in \mathbb{C}^{n^2 \times n^2}$
- $n^4 - n^2$ real parameters to estimate

Process Matrix to OSR

$$\mathbf{X} \rightarrow \mathbf{A} \text{ via SVD of } \mathbf{X} \Rightarrow \begin{cases} \text{relaxed optimal } \mathbf{X} \text{ is optimal ML estimate} \\ \text{number of OSR} \leq n^2 \end{cases}$$

OSR → Process Matrix

- Expand $n \times n$ OSR elements $\{A_\mu, \mu = 1, \dots, \ell\}$ in a fixed basis

$$A_\mu = \sum_{\alpha=1}^{n^2} \textcolor{red}{x}_{\mu\alpha} B_\alpha, \quad B_\alpha \in \mathbf{C}^{n \times n}$$

- Process matrix

$$\begin{aligned} (\textcolor{red}{X})_{\alpha\beta} &= \sum_{\mu=1}^{\ell} \textcolor{red}{x}_{\mu\alpha} \textcolor{red}{x}_{\mu\beta}^* \\ \sum_{\alpha,\beta} \textcolor{red}{X}_{\alpha\beta} B_\alpha^\dagger B_\beta &= I_n \end{aligned} \implies \left\{ \begin{array}{l} \textcolor{red}{X} \in \mathbf{C}^{n^2 \times n^2} \\ \text{quadratic in } \textcolor{red}{x}_\mu \in \mathbf{C}^{n^2} \\ \text{positive semidefinite} \\ n^4 - n^2 \text{ real parameters} \end{array} \right.$$

Process Matrix → (equivalent) OSR

$$\text{SVD of } \textcolor{red}{X} = V S V^\dagger \implies A'_\mu = \sum_{\alpha=1}^{n^2} \sqrt{s_\mu} V_{\alpha\mu} B_\alpha, \quad \mu = 1, \dots, n^2$$

- number of numerically non-zero $s_\mu = \text{rank } \textcolor{red}{X} \leq n^2$; often $\ll n^2$

Relax quadratic equality constraint to $\mathbf{X} \geq 0$

ML Process Tomography

$$\text{minimize } L(\mathbf{X}) = - \sum_{i,k} n_{ik} \log p_{ik}(\mathbf{X})$$

$$\text{subject to } \sum_{\alpha,\beta} \mathbf{X}_{\alpha\beta} B_\alpha^\dagger B_\beta = I_n, \quad \mathbf{X} \geq 0$$

$$p_{ik}(\mathbf{X}) = \text{Tr } \mathbf{X} R_{ik}, \quad (R_{ik})_{\alpha\beta} = \text{Tr } B_\alpha \rho_k B_\beta^\dagger O_{ik}$$

- convex optimization in $\mathbf{X} \in \mathbb{C}^{n^2 \times n^2}$
- $n^4 - n^2$ real parameters to estimate

Process Matrix to OSR

$$\mathbf{X} \rightarrow \mathbf{A} \text{ via SVD of } \mathbf{X} \Rightarrow \begin{cases} \text{relaxed optimal } \mathbf{X} \text{ is optimal ML estimate} \\ \text{number of OSR} \leq n^2 \end{cases}$$

OSR → Process Matrix

- Expand $n \times n$ OSR elements $\{A_\mu, \mu = 1, \dots, \ell\}$ in a fixed basis

$$A_\mu = \sum_{\alpha=1}^{n^2} \mathbf{x}_{\mu\alpha} B_\alpha, \quad B_\alpha \in \mathbb{C}^{n \times n}$$

- Process matrix

$$\begin{aligned} (\mathbf{X})_{\alpha\beta} &= \sum_{\mu=1}^{\ell} \mathbf{x}_{\mu\alpha} \mathbf{x}_{\mu\beta}^* \\ \sum_{\alpha,\beta} \mathbf{X}_{\alpha\beta} B_\alpha^\dagger B_\beta &= I_n \end{aligned} \implies \left\{ \begin{array}{l} \mathbf{X} \in \mathbb{C}^{n^2 \times n^2} \\ \text{quadratic in } \mathbf{x}_\mu \in \mathbb{C}^{n^2} \\ \text{positive semidefinite} \\ n^4 - n^2 \text{ real parameters} \end{array} \right.$$

Process Matrix → (equivalent) OSR

$$\text{SVD of } \mathbf{X} = V S V^\dagger \implies A'_\mu = \sum_{\alpha=1}^{n^2} \sqrt{s_\mu} V_{\alpha\mu} B_\alpha, \quad \mu = 1, \dots, n^2$$

- number of numerically non-zero $s_\mu = \text{rank } \mathbf{X} \leq n^2$; often $\ll n^2$

Relax quadratic equality constraint to $\mathbf{X} \geq 0$

ML Process Tomography

$$\text{minimize } L(\mathbf{X}) = - \sum_{i,k} n_{ik} \log p_{ik}(\mathbf{X})$$

$$\text{subject to } \sum_{\alpha,\beta} \mathbf{X}_{\alpha\beta} B_\alpha^\dagger B_\beta = I_n, \quad \mathbf{X} \geq 0$$

$$p_{ik}(\mathbf{X}) = \text{Tr } \mathbf{X} R_{ik}, \quad (R_{ik})_{\alpha\beta} = \text{Tr } B_\alpha \rho_k B_\beta^\dagger O_{ik}$$

- convex optimization in $\mathbf{X} \in \mathbb{C}^{n^2 \times n^2}$
- $n^4 - n^2$ real parameters to estimate

Process Matrix to OSR

$$\mathbf{X} \rightarrow \mathbf{A} \text{ via SVD of } \mathbf{X} \Rightarrow \begin{cases} \text{relaxed optimal } \mathbf{X} \text{ is optimal ML estimate} \\ \text{number of OSR} \leq n^2 \end{cases}$$

Parametric process tomography

OSR basis model

$$\overline{\mathcal{A}} = \left\{ \overline{A}_\mu \in \mathbb{C}^{n \times n} \mid \mu = 1, \dots, m \right\}$$

parametric OSR model

$$\mathcal{A}(\theta) = \left\{ \sqrt{\theta_\mu} \overline{A}_\mu \in \mathbb{C}^{n \times n}, \theta_\mu \in \mathbb{R}_+ \mid \mu = 1, \dots, m \right\}$$

ML parametric process tomography

$$\text{minimize } L(\theta) = - \sum_{i,k} n_{ik} \log(a_{ik}^T \theta), \quad (a_{ik})_\mu = \text{Tr } O_{ik} \overline{A}_\mu \rho_k \overline{A}_\mu^\dagger$$

$$\text{subject to } \theta \geq 0, \quad \sum_\mu \theta_\mu \overline{A}_\mu^\dagger \overline{A}_\mu = I_n$$

- a convex optimization in $\theta \in \mathbb{R}^m$
- $m \ll n^4 - n^2 \implies$ considerable reduction over standard QPT

Parametric process tomography

OSR basis model

$$\overline{\mathcal{A}} = \left\{ \overline{A}_\mu \in \mathbb{C}^{n \times n} \mid \mu = 1, \dots, m \right\}$$

parametric OSR model

$$\mathcal{A}(\theta) = \left\{ \sqrt{\theta_\mu} \overline{A}_\mu \in \mathbb{C}^{n \times n}, \theta_\mu \in \mathbb{R}_+ \mid \mu = 1, \dots, m \right\}$$

ML parametric process tomography

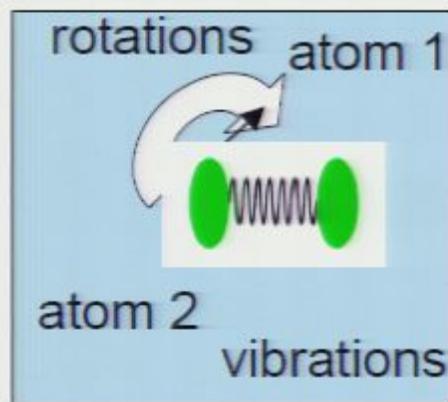
$$\text{minimize } L(\theta) = - \sum_{i,k} n_{ik} \log(a_{ik}^T \theta), \quad (a_{ik})_\mu = \text{Tr } O_{ik} \overline{A}_\mu \rho_k \overline{A}_\mu^\dagger$$

$$\text{subject to } \theta \geq 0, \quad \sum_\mu \theta_\mu \overline{A}_\mu^\dagger \overline{A}_\mu = I_n$$

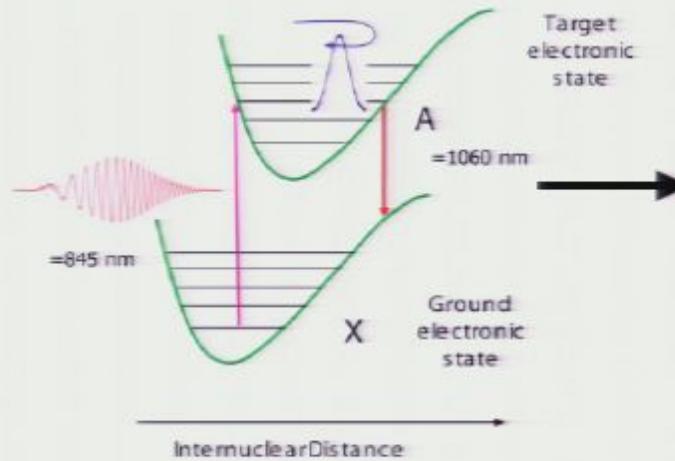
- a convex optimization in $\theta \in \mathbb{R}^m$
- $m \ll n^4 - n^2 \implies$ considerable reduction over standard QPT

Process tomography of vibrational wavepackets in diatomic molecules

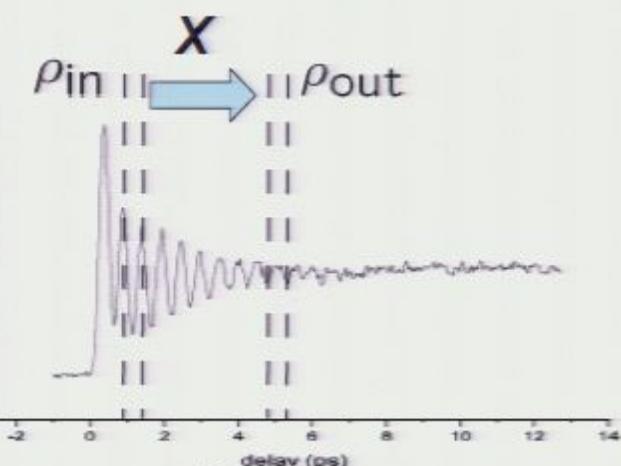
Diatomical molecular system



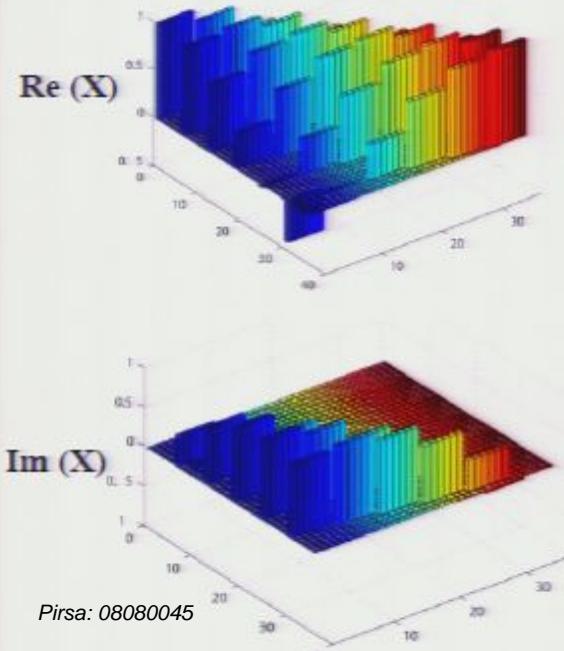
Electronic transition



Signal from fluorescence measurements of vibrations



Reconstructed Process Matrix



Reduction of variables from $6^4 - 6^2 = 1260$ in X to 150 in θ via model-based OSR basis set $\{A_\mu\}$

$$\text{minimize } L(\theta) = - \sum_{i,k} (p_{ik}^{\text{emp}} - \text{Tr } O_{ik} \rho_{\text{out}})^2$$

$$\text{subject to } \rho_{\text{out}} = \sum_{\mu} \theta_{\mu} A_{\mu} \rho_{\text{in}} A_{\mu}^{\dagger}$$

$$\sum_{\mu} \theta_{\mu} A_{\mu}^{\dagger} A_{\mu} = I_6, \quad \theta \geq 0$$

- how “good” is the parametric model?
- to test the limit of performance eliminate effect of selecting configurations, i.e., POVMs and input states

$$\begin{aligned}\|\rho_{\text{out}} - \rho_{\text{out}}(\theta)\|_{\text{fro}} &= \left\| \sum_{\mu} A_{\mu} \rho_{\text{in}} A_{\mu}^{\dagger} - \theta_{\mu} \overline{A}_{\mu} \rho_{\text{in}} \overline{A}_{\mu}^{\dagger} \right\|_{\text{fro}} \\ &\leq \left\| \sum_{\mu} A_{\mu}^* \otimes A_{\mu} - \theta_{\mu} \overline{A}_{\mu}^* \otimes \overline{A}_{\mu} \right\|_2 \|\text{vec}(\rho_{\text{in}})\|_2\end{aligned}$$

- configuration-free (convex) optimization

$$\begin{aligned}&\text{minimize} \quad \left\| \sum_{\mu} A_{\mu}^* \otimes A_{\mu} - \theta_{\mu} \overline{A}_{\mu}^* \otimes \overline{A}_{\mu} \right\|_2 \\ &\text{subject to} \quad \theta \geq 0, \quad \sum_{\mu} \theta_{\mu} \overline{A}_{\mu}^{\dagger} \overline{A}_{\mu} = I_n\end{aligned}$$

numerical example

- actual system – input state is $|0_E\rangle\langle 0_E| \otimes \rho_Q$

Hamiltonian $H = I_E \otimes H_Q + \Delta \in \mathbf{H}^{n_E n_Q}$

OSR $\mathcal{A} = \{A_1, \dots, A_{n_E}\}$

- basis system – input state is $|0_{\overline{E}}\rangle\langle 0_{\overline{E}}| \otimes \rho_Q$

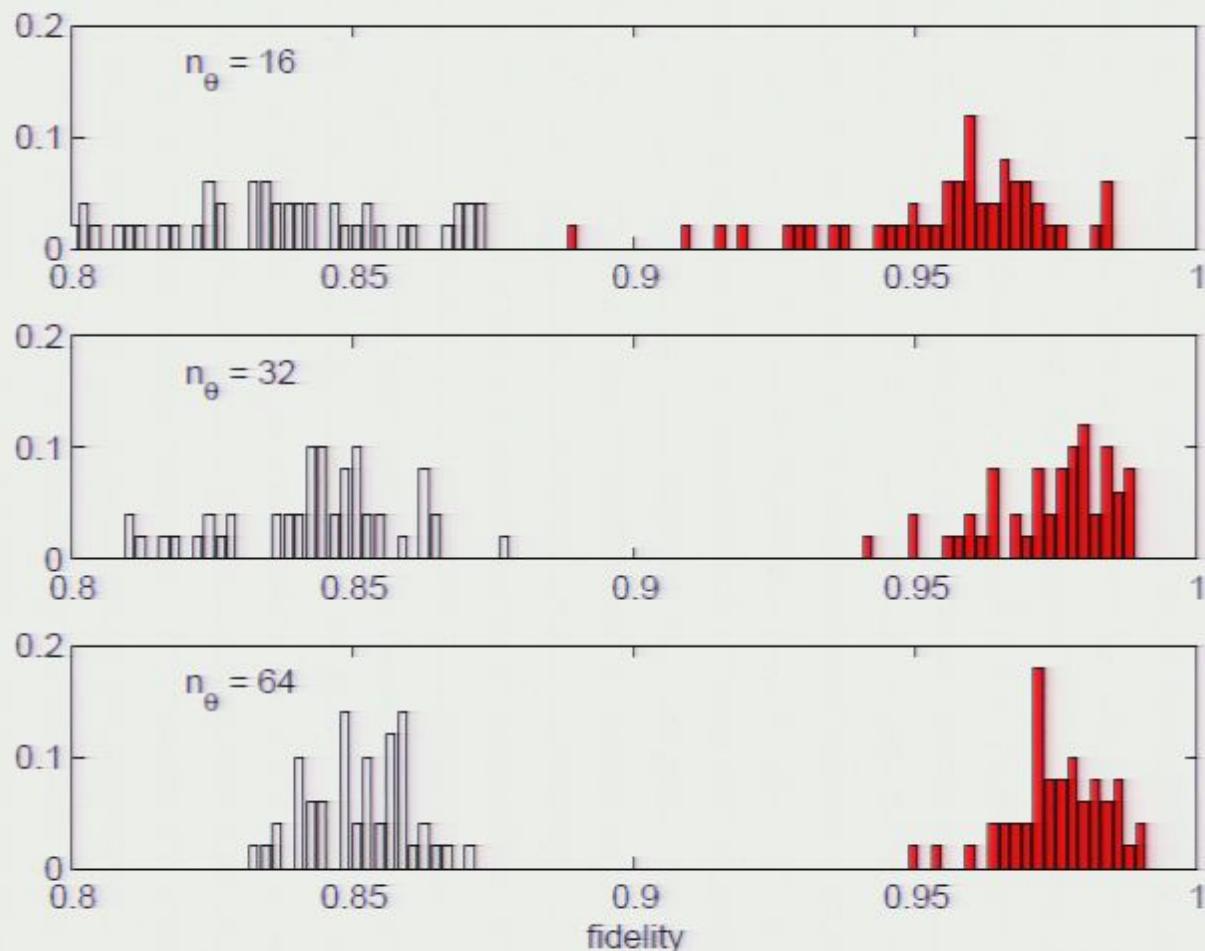
Hamiltonian $\overline{H} = I_{\overline{E}} \otimes H_Q + \overline{\Delta} \in \mathbf{H}^{\overline{n}_E n_Q}$

OSR $\overline{\mathcal{A}} = \{\overline{A}_1, \dots, \overline{A}_{n_\theta}\}, \quad n_\theta = \overline{n}_E$

- parametric OSR model

$$\mathcal{A}(\theta) = \left\{ \sqrt{\theta_1} \overline{A}_1, \dots, \sqrt{\theta_{n_\theta}} \overline{A}_{n_\theta} \right\}$$

Histograms of fidelity



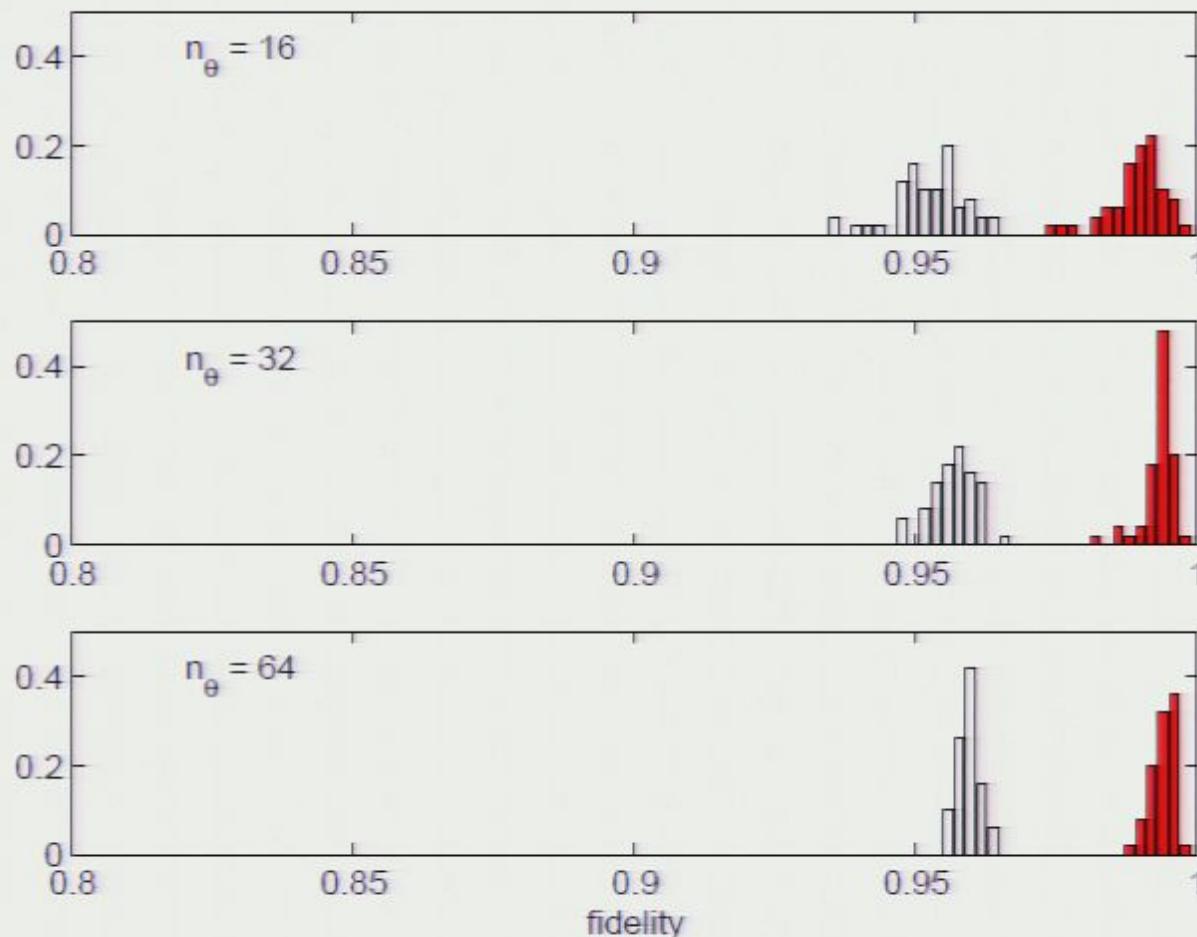
$$f(\bar{\mathcal{A}}, \mathcal{A})$$

$$f(\mathcal{A}(\theta), \mathcal{A})$$

- Δ random, 50 runs
- $\bar{\Delta}$ held fixed per run
- standard QPT: 240 parameters

- control: H_Q produces the Bell transform ($n_Q = 4$)
- system: $\|\Delta\| = 0.025$ with $n_E = n_Q^2 = 16 \implies \Delta \in \mathbb{H}^{64}$
- basis: $\|\bar{\Delta}\| = 1$ $\bar{n}_E \in \{16, 32, 64\} \implies \bar{\Delta} \in \mathbb{H}^{\{64, 128, 256\}}$

Histograms of fidelity



$f(\bar{\mathcal{A}}, \mathcal{A})$

$f(\mathcal{A}(\theta), \mathcal{A})$

- Δ random, 50 runs
- $\bar{\Delta}$ held fixed per run
- standard QPT: 240 parameters

- control: H_Q produces the Bell transform ($n_Q = 4$)
- system: $\|\Delta\| = 0.01$ with $n_E = n_Q^2 = 16 \implies \Delta \in \mathbb{H}^{64}$
- basis: $\|\bar{\Delta}\| = 0.5$ with $\bar{n}_E \in \{16, 32, 64\} \implies \bar{\Delta} \in \mathbb{H}^{\{64, 128, 256\}}$

numerical example

- actual system – input state is $|0_E\rangle\langle 0_E| \otimes \rho_Q$

Hamiltonian $H = I_E \otimes H_Q + \Delta \in \mathbf{H}^{n_E n_Q}$

OSR $\mathcal{A} = \{A_1, \dots, A_{n_E}\}$

- basis system – input state is $|0_{\overline{E}}\rangle\langle 0_{\overline{E}}| \otimes \rho_Q$

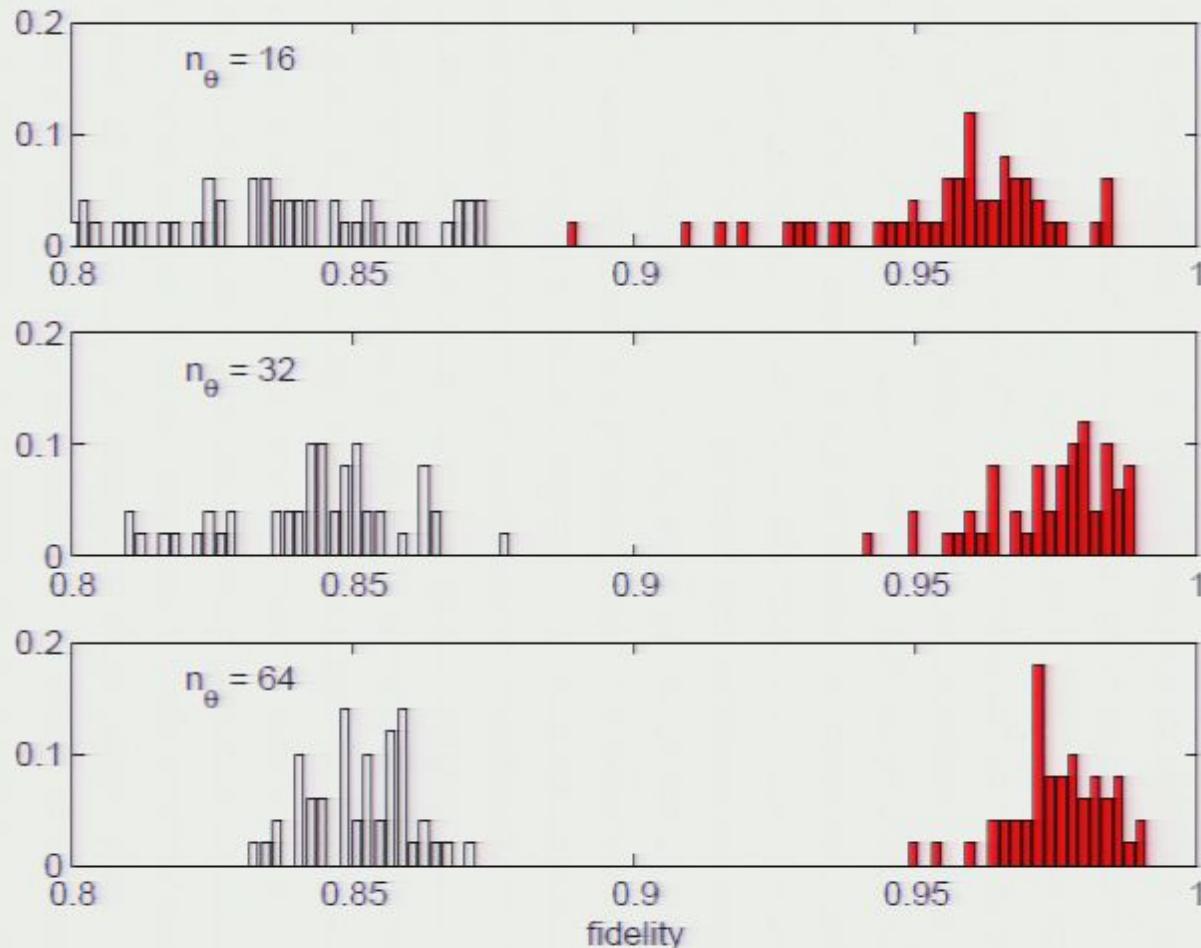
Hamiltonian $\overline{H} = I_{\overline{E}} \otimes H_Q + \overline{\Delta} \in \mathbf{H}^{\overline{n}_E n_Q}$

OSR $\overline{\mathcal{A}} = \{\overline{A}_1, \dots, \overline{A}_{n_\theta}\}, \quad n_\theta = \overline{n}_E$

- parametric OSR model

$$\mathcal{A}(\theta) = \left\{ \sqrt{\theta_1} \overline{A}_1, \dots, \sqrt{\theta_{n_\theta}} \overline{A}_{n_\theta} \right\}$$

Histograms of fidelity



$$f(\bar{\mathcal{A}}, \mathcal{A})$$

$$f(\mathcal{A}(\theta), \mathcal{A})$$

- Δ random, 50 runs
- $\bar{\Delta}$ held fixed per run
- standard QPT: 240 parameters

- control: H_Q produces the Bell transform ($n_Q = 4$)
- system: $\|\Delta\| = 0.025$ with $n_E = n_Q^2 = 16 \implies \Delta \in \mathbb{H}^{64}$
- basis: $\|\bar{\Delta}\| = 1$ $\bar{n}_E \in \{16, 32, 64\} \implies \bar{\Delta} \in \mathbb{H}^{\{64, 128, 256\}}$

Quantum State Detection

Optimal experiment design for state estimation

model $p_{ik}(\rho) = \text{Tr } O_{ik} \rho, \quad \text{Tr } \rho = 1, \quad \rho \geq 0$

data $n_{ik} = \#(i \mid k) \text{ from } \ell_k \text{ out of } N \text{ experiments}$

variables $\ell = \{\ell_k\}$

unbiased estimate $\hat{\rho}(\ell), \quad \sum_k \ell_k = N$

Variance lower bound from Cramér-Rao Inequality

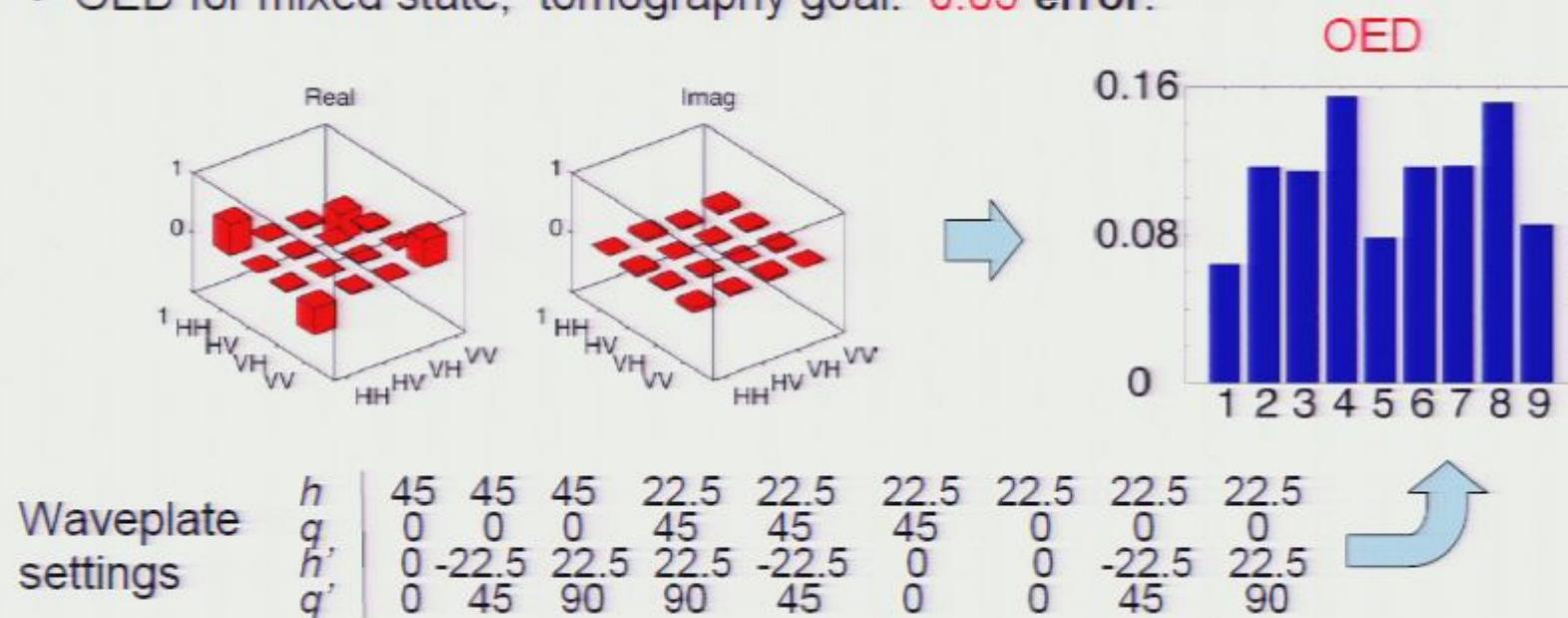
$$\mathbf{E} \|\hat{\rho}(\ell) - \rho_{\text{true}}\|^2 \geq V(\ell, \rho_{\text{true}}) = \text{Tr} \left(\sum_k \ell_k F_k(\rho_{\text{true}}) \right)^{-1}$$

$$F_k(\rho_{\text{true}}) = C^T \left(\sum_i \frac{(\text{vec } O_{ik})(\text{vec } O_{ik})^\dagger}{p_{ik}(\rho_{\text{true}})} \right) C \in \mathbb{R}^{n^2-1 \times n^2-1}$$

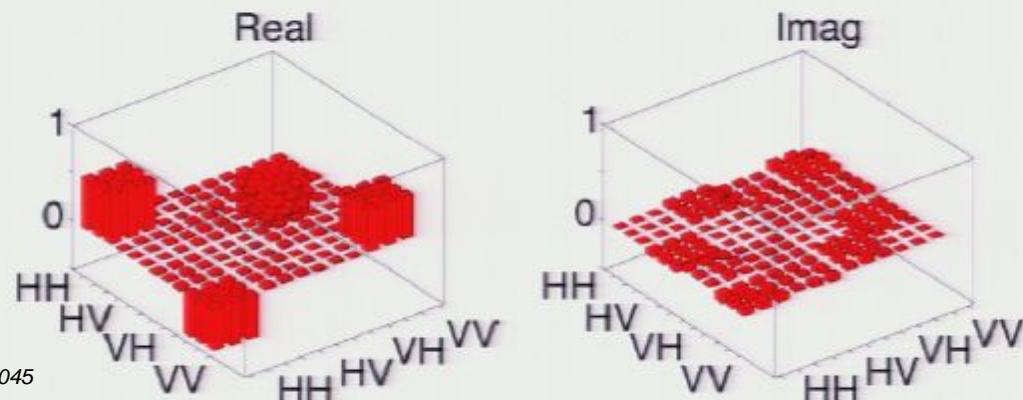
- $C \in \mathbb{R}^{n^2 \times n^2-1}$ accounts for constraint $\text{Tr } \rho = 1$
- a criterion for experiment design – $V(\ell, \hat{\rho})$ with surrogate $\hat{\rho}$ for ρ_{true}

Reconstructed two-photon mixed state via OED

- OED for mixed state; tomography goal: 0.05 error.



9 data sets, 5,200 experiments each for saturating CR bound.



- Mean CR from data set:

$$CR_{BOUND} = 0.051$$

Maximum likelihood Hamiltonian parameter estimation

$$\text{minimize } L(\theta) = - \sum_{i,k} n_{ik} \log p_{ik}(\theta)$$

$$\text{subject to } p_{ik}(\theta) = \text{Tr } O_{ik}(\theta) \rho_k, \quad \theta \in \Theta$$

- typically Θ is a convex set, e.g., $\|\theta - \theta_{\text{nom}}\| \leq \delta$
- example: Hamiltonian system model

$$\begin{aligned} H_k(t, \theta) &= H_{k,0}(t) + \sum_{\nu} \theta_{\nu} H_{k,\nu}(t), \quad 0 \leq t \leq T \\ O_{ik}(\theta) &= U_k(T, \theta)^{\dagger} M_{ik} U_k(T, \theta) \end{aligned}$$

- ML (or LS) **not**, in general, a convex optimization of $\theta \in \Theta$
- OED **is** a convex optimization in distribution of experiments

normalized Hamiltonian in rotating frame

$$H = H_1 + H_2 + H_{12}$$

$$H_1 = \frac{1}{2} [\varepsilon_{1z} \omega_0 (Z \otimes I_2) + \varepsilon_{1x} \omega_1 (X \otimes I_2)]$$

$$H_2 = \frac{1}{2} [\varepsilon_{2z} \omega_0 (I_2 \otimes Z) + \varepsilon_{2x} \omega_1 (I_2 \otimes X)]$$

$$H_{12} = \varepsilon_c \omega_c S$$

$$\omega_0 = \omega_{xy} = \gamma g_z^{\text{nom}} B_z, \quad \omega_1 = \gamma g_{xy}^{\text{nom}} B_{xy}, \quad \omega_c = g_c^{\text{nom}} / \hbar$$

controls – normalized gate voltages

$$\varepsilon_{1z}, \varepsilon_{1x}, \varepsilon_{2z}, \varepsilon_{2x}, \varepsilon_c$$

Semiconductor two-qubit gate – idealized g-factor control (UCLA/HRL)

$$H = H_1 + H_2 + H_{12} \quad \left\{ \begin{array}{l} H_1 = \overline{H}_1 \otimes I_2 \\ H_2 = I_2 \otimes \overline{H}_2 \end{array} \right.$$

$$\overline{H}_n = \frac{\hbar\gamma}{2} \left(g_{n,z} Z B_z + g_{n,xy} (X B_x + Y B_y) \right), \quad n = 1, 2$$

$$H_{12} = g_c S$$

- γ – gyromagnetic ratio
- X, Y, Z – 2×2 Pauli spin matrices
- S – 4×4 spin coupling matrix $S = X^{\otimes 2} + Y^{\otimes 2} + Z^{\otimes 2}$
- B_x, B_y, B_z – magnetic field $\left\{ \begin{array}{l} B_z = \text{constant} \\ B_x = B_{xy} \cos \omega_{xy} t \\ B_y = B_{xy} \sin \omega_{xy} t \end{array} \right.$
- controlled “g-factors” $g_c, \{g_{n,z}, g_{n,xy}, n = 1, 2\}$

normalized Hamiltonian in rotating frame

$$H = H_1 + H_2 + H_{12}$$

$$H_1 = \frac{1}{2} [\varepsilon_{1z} \omega_0 (Z \otimes I_2) + \varepsilon_{1x} \omega_1 (X \otimes I_2)]$$

$$H_2 = \frac{1}{2} [\varepsilon_{2z} \omega_0 (I_2 \otimes Z) + \varepsilon_{2x} \omega_1 (I_2 \otimes X)]$$

$$H_{12} = \varepsilon_c \omega_c S$$

$$\omega_0 = \omega_{xy} = \gamma g_z^{\text{nom}} B_z, \quad \omega_1 = \gamma g_{xy}^{\text{nom}} B_{xy}, \quad \omega_c = g_c^{\text{nom}} / \hbar$$

controls – normalized gate voltages

$$\varepsilon_{1z}, \varepsilon_{1x}, \varepsilon_{2z}, \varepsilon_{2x}, \varepsilon_c$$

“a” realization of the Bell transform

$$U_{\text{Bell}} = (U_{\text{had}} \otimes I_2) \sqrt{U_{\text{swap}}} (X^{-1/2} \otimes X^{1/2}) \sqrt{U_{\text{swap}}} (I_2 \otimes X)$$

pulse control sequence

ε_{1z}	ε_{1x}	ε_{2z}	ε_{2x}	ε_c	Δt	gate
0	0	0	1	0	$\frac{\pi}{\omega_1}$	$-iI_2 \otimes X$
0	0	0	0	1	$\frac{\pi}{8\omega_c}$	$e^{-i\frac{\pi}{8}}\sqrt{U_{\text{swap}}}$
0	0	0	1	0	$\frac{\pi}{2\omega_1}$	$e^{-i\frac{\pi}{4}}I_2 \otimes X^{1/2}$
0	1	0	0	0	$\frac{3\pi}{2\omega_1}$	$e^{-i\frac{3\pi}{4}}X^{-1/2} \otimes I_2$
0	0	0	0	1	$\frac{\pi}{8\omega_c}$	$e^{-i\frac{\pi}{8}}\sqrt{U_{\text{swap}}}$
$\frac{\omega_{\text{had}}}{\omega_0\sqrt{2}}$	$\frac{\omega_{\text{had}}}{\omega_1\sqrt{2}}$	0	0	0	$\frac{\pi}{\omega_{\text{had}}}$	$-iU_{\text{had}} \otimes I_2$

resulting gate $U(t_f) = e^{-i\frac{\pi}{4}} U_{\text{Bell}}, \quad t_f = \left(\frac{3}{\omega_1} + \frac{1}{4\omega_c} + \frac{1}{\omega_{\text{had}}} \right) \pi$

(iterative) adaptive control

initialize

- pulse sequence

repeat

- implement pulse sequence
- measure Z at
 - $t_f \Rightarrow n_{cfg} = 1 (n_{sa} = 1)$
 - $(t_f/2, t_f) \Rightarrow n_{cfg} = 2 (n_{sa} = 2)$
- ML estimate of ω_1

until

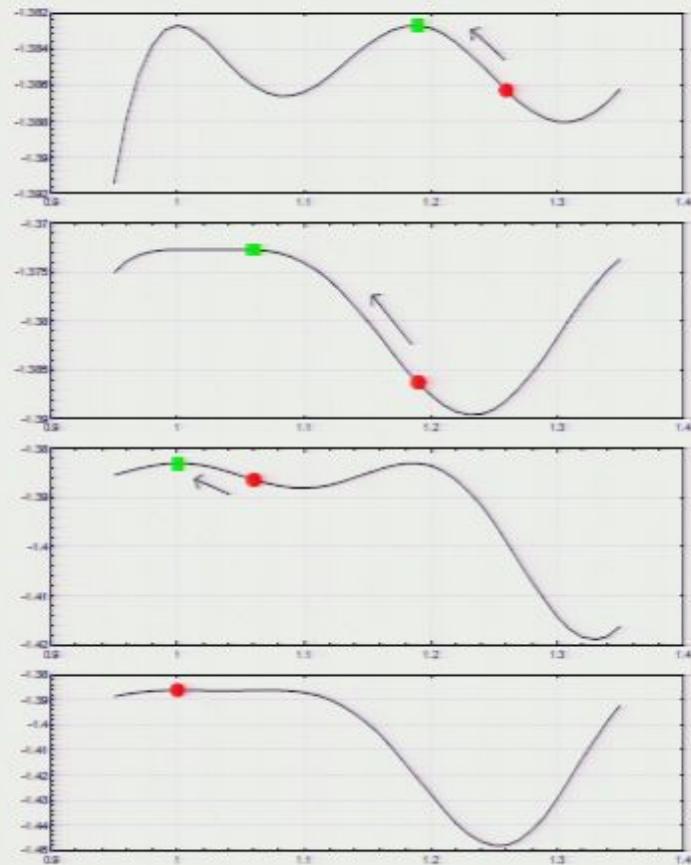
- performance criteriton satisfied (process tomography \rightarrow fidelity?)

example with infinite data

average likelihood function

$$\mathcal{L}(\omega_1) = \sum_{i,k} p_{ik}(\omega_1) \log \text{Tr } p_{ik}(\omega_1)$$

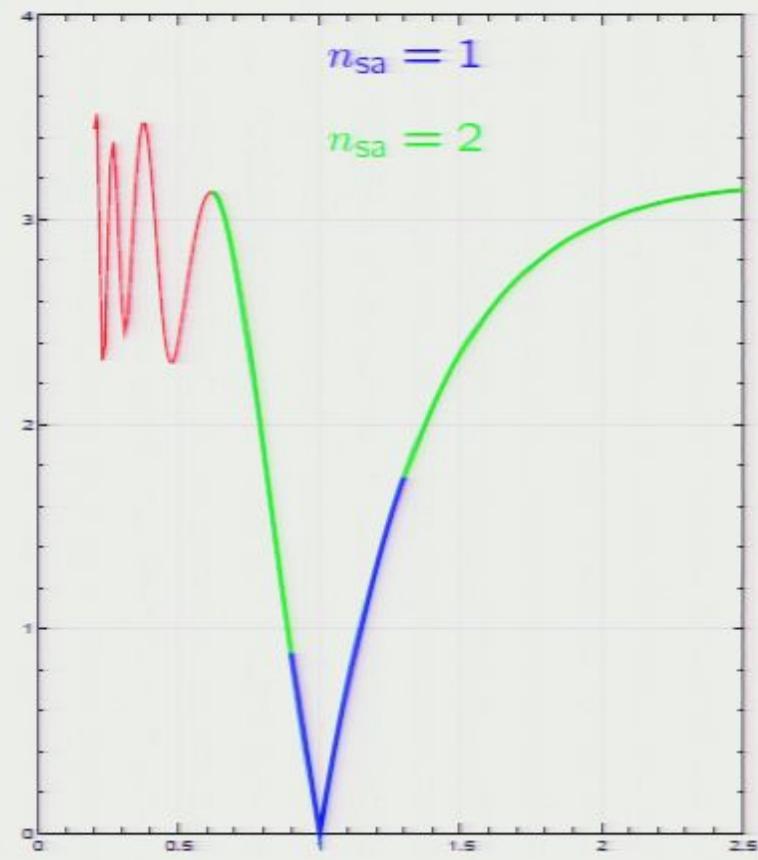
$n_{\text{sa}} = 1$



model parameter (ω_1/ω_1)

(in)fidelity

$$1 - f(U(\varepsilon(\omega_1)), U_{\text{Bell}})$$



model parameter (ω_1/ω_1)

Optimal experiment design for Hamiltonian parameter estimation

model

$$p_{ik}(\theta) = \text{Tr } O_{ik}(\theta) \rho_k$$

data

$$n_{ik} = \#(i \mid k) \text{ from } \ell_k \text{ out of } N \text{ experiments}$$

variables

$$\ell = \{\ell_k\}$$

unbiased estimate $\hat{\theta}(\ell)$

variance lower bound from Cramér-Rao Inequality

$$\mathbb{E} \|\hat{\theta}(\ell) - \theta\|^2 \geq V(\ell, \theta) = \text{Tr} \left(\sum_k \ell_k F_k(\theta) \right)^{-1}$$

$$F_k(\theta) = \sum_i \left(\nabla_\theta p_{ik}(\theta) \right) \left(\nabla_\theta p_{ik}(\theta) \right)^T / p_{ik}(\theta)$$

$$\sum_k \ell_k = N, \quad \ell_k \text{ is an integer}$$

State & Process Tomography: Summary

Convex Optimization? ^a

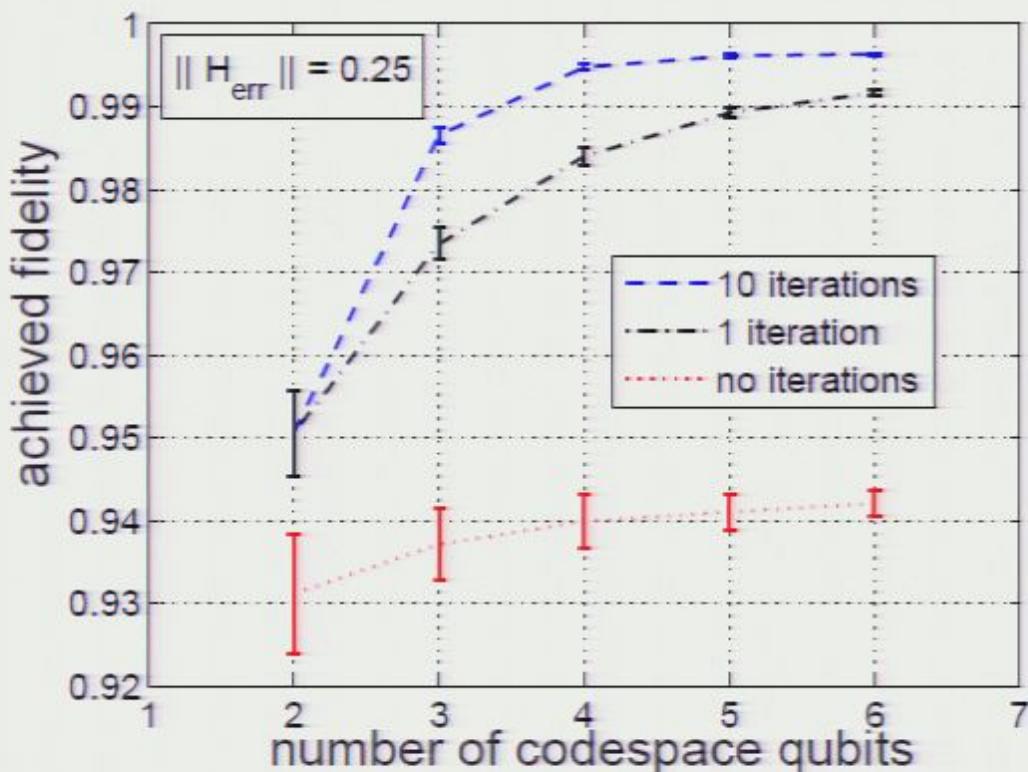
Tomography Variable	MLE	OED
State (density)	yes	yes
State distribution	yes	yes
Process matrix (fixed basis)	yes	yes
OSR distribution	yes	yes
Hamiltonian parameters	not always	yes

^aR. L. Kosut, I. A. Walmsley and H. Rabitz, Optimal experiment design for quantum state and process tomography and Hamiltonian parameter estimation, [quant-ph/0411093](https://arxiv.org/abs/quant-ph/0411093), Nov. 2004

Quantum Error Correction

Example: perfect tomography

$$H_{QE}(t) = I_E \otimes H_Q(t) + H_{\text{err}}(t), \quad n_{QE} = (2^{2+n_{CA}}) n_E$$



- the goal is to preserve 2 qubits ($n_S = 4$)
- $1-\sigma$ errorbars from 100 runs with H_{err} chosen randomly and then normalized to $\| H_{\text{err}} \| = 0.25$

- shows tradeoff: codespace qubits vs. tomographies, e.g.,
 $f_{\text{ach}} > 0.98$ costs (4 qubits, 1 tomog.) or (3 qubits, 10 tomog.)
 $f_{\text{ach}} > 0.99$ costs (6 qubits, 1 tomog.) or (4 qubits, 10 tomog.)

Concluding Remarks

On Learning (a.k.a. “estimating”)

First rule of learning: **Make Mistakes**

On Learning (a.k.a. “estimating”)

First rule of learning: **Make Mistakes**

Second rule of learning: **Correct Mistakes**

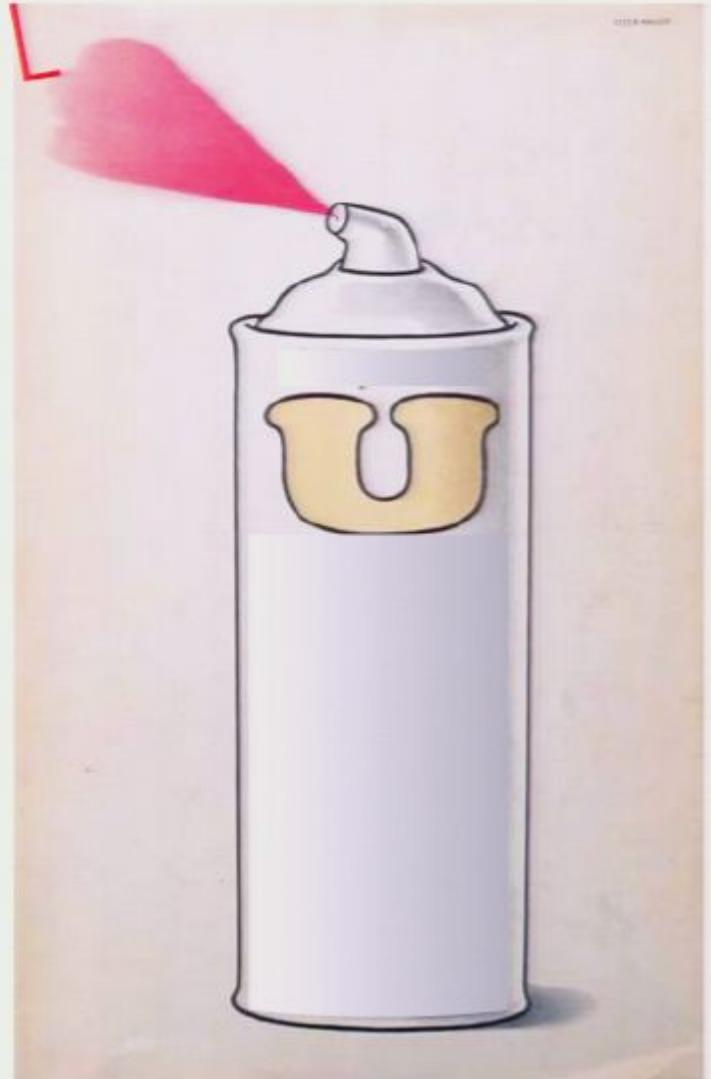


Feedback Control

C O
N T R O L

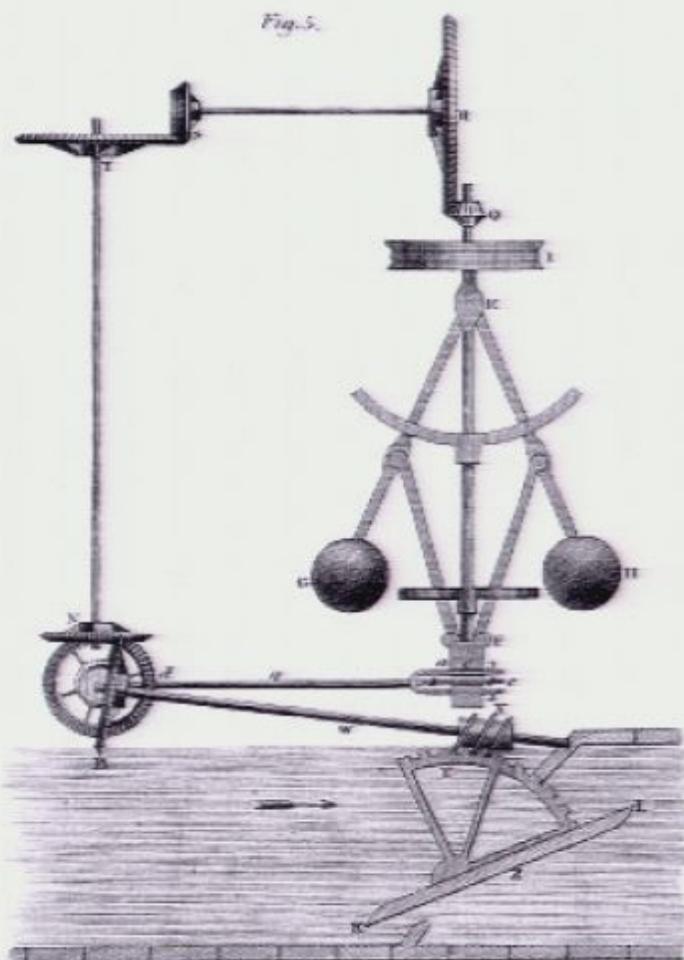
UBIQUITOUS FOR QIP/QEC

**CONTROL IS REQUIRED
EVERYWHERE
THERE IS A DESIRED UNITARY.**



Source: 1969 Doubleday Book Cover
UBIK by Philip K. Dick

The Feedback Control design paradigm



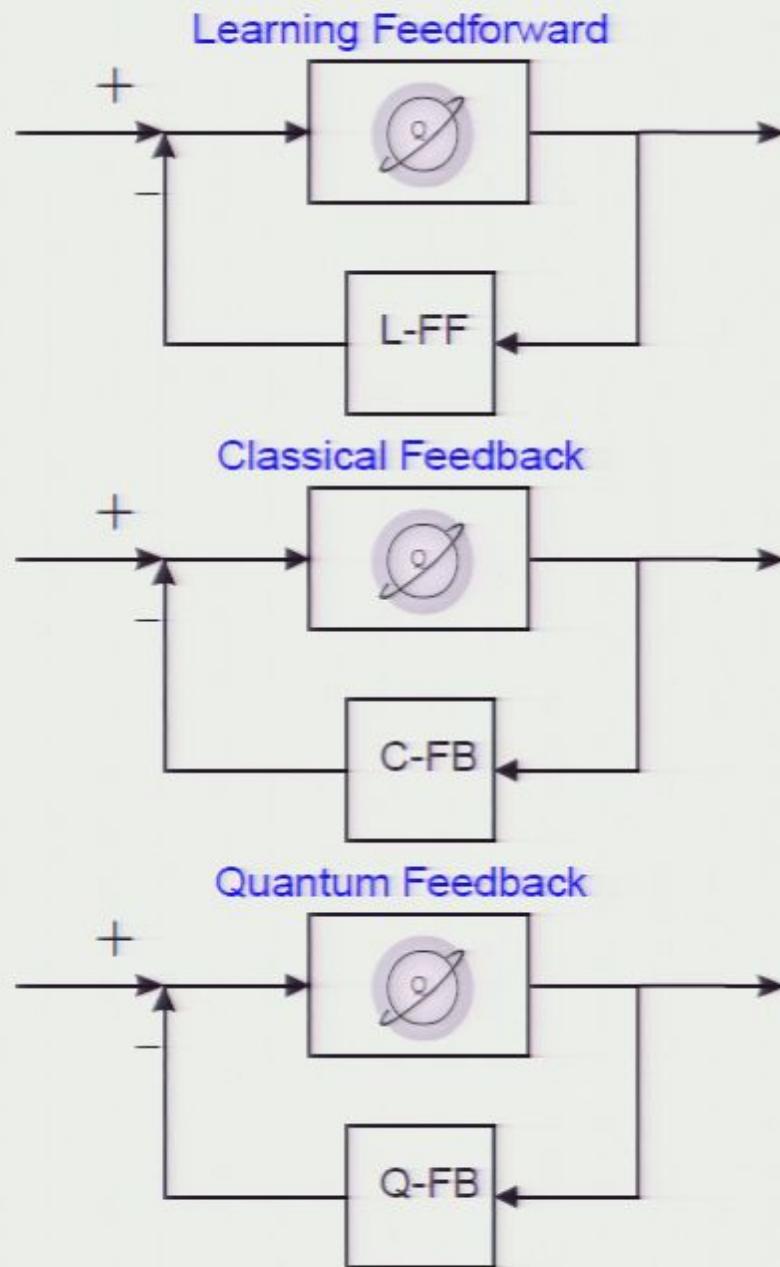
“Flyball Governor”
for steam engine control
James Watt (ca. 1788)

completely mechanical
“natural” speed regulation

Is there a
Quantum Flyball Governor?

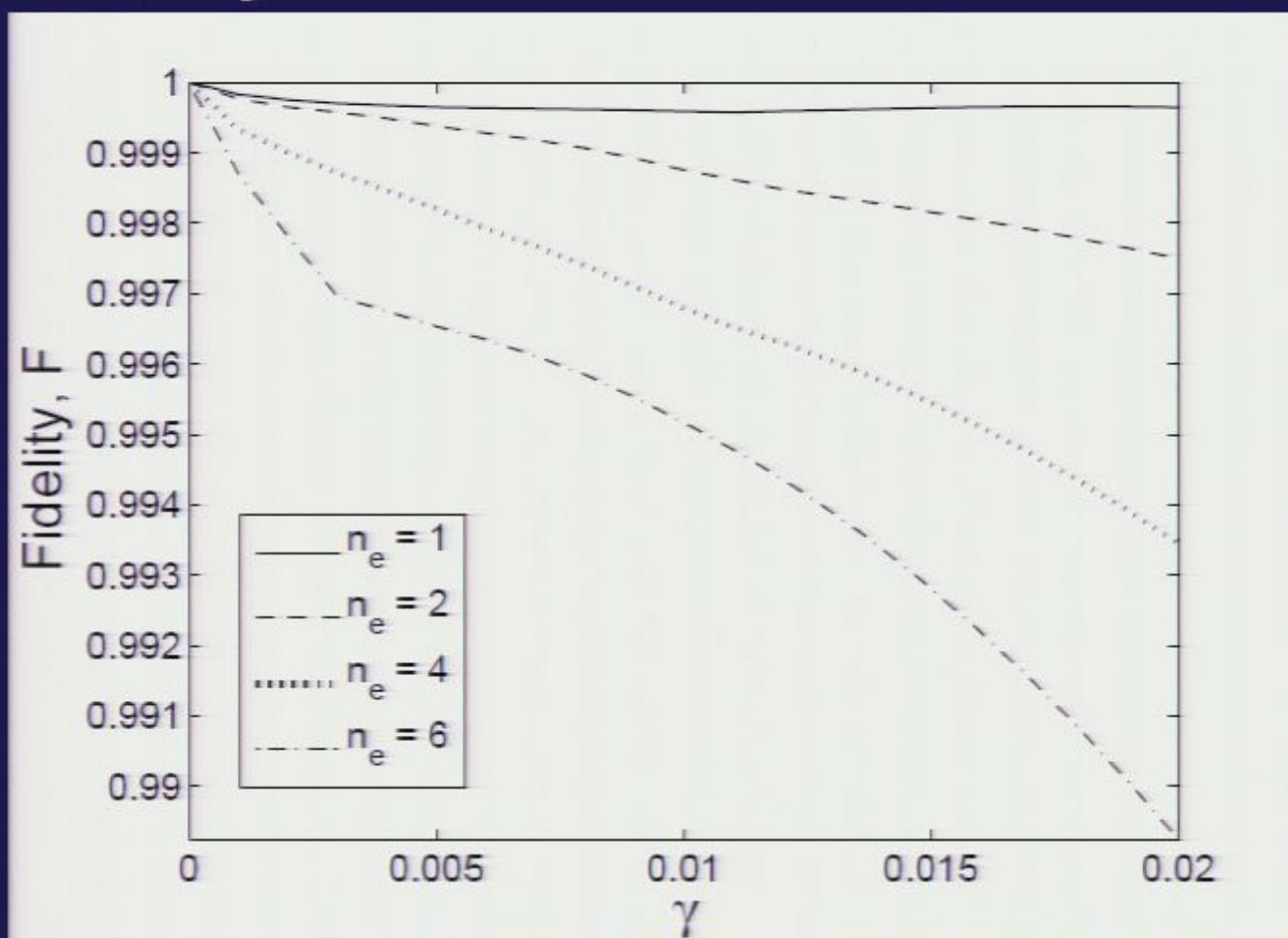
completely quantum mechanical
“natural” quantum error correction

Control of Quantum Systems



Optimal Control: single-qubit

Multiparticle environment: Hadamard





“ ... and these imperfections may produce considerable havoc.”

– Richard Feynman



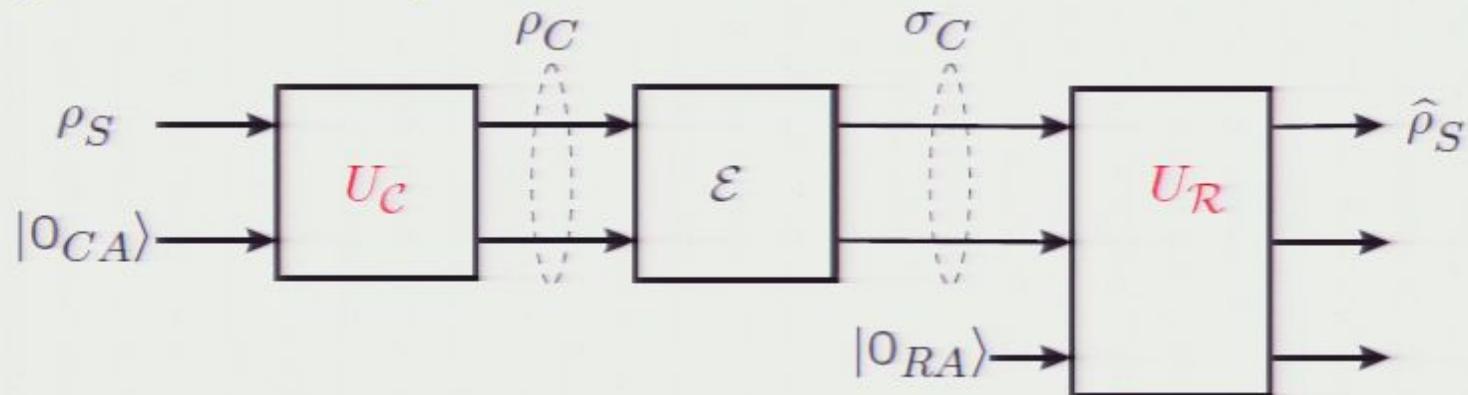
“It ain’t what you don’t know
that gets you into trouble.
It’s what you know for sure
that just ain’t so.”

– Mark Twain

Oxford, 1907

Quantum error correction model

System-ancilla representation



$$U_C = [C \cdots], \quad C \in \mathbb{C}^{n_C \times n_S}, \quad (n_C = n_S n_{CA}) \quad C^\dagger C = I_S$$

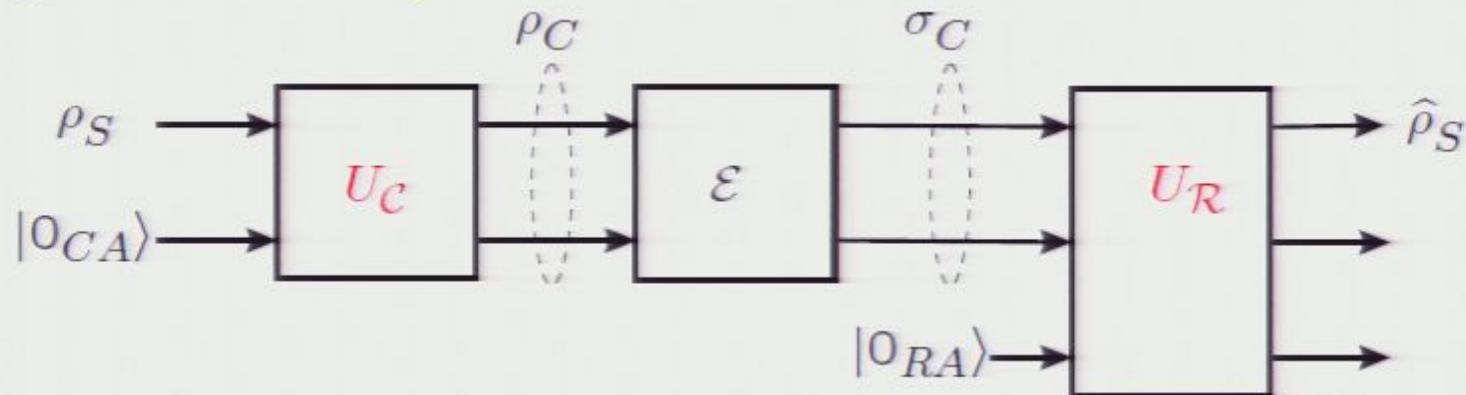
$$U_R = [R \cdots], \quad R = \begin{bmatrix} R_1 \\ \vdots \\ R_{n_{CA} n_{RA}} \end{bmatrix}, \quad R_i \in \mathbb{C}^{n_S \times n_C}, \quad R^\dagger R = I_C$$

Design goal

Determine encoding \mathcal{C} and recovery \mathcal{R} to so that the noisy channel, $\mathcal{R}\mathcal{E}\mathcal{C}$, is as close as possible to a desired unitary U_S .

Quantum error correction model

System-ancilla representation



$$U_C = [C \ \dots], \quad C \in \mathbb{C}^{n_C \times n_S}, \quad (n_C = n_S n_{CA}) \quad C^\dagger C = I_S$$

$$U_R = [R \ \dots], \quad R = \begin{bmatrix} R_1 \\ \vdots \\ R_{n_{CA} n_{RA}} \end{bmatrix}, \quad R_i \in \mathbb{C}^{n_S \times n_C}, \quad R^\dagger R = I_C$$

Design goal

Determine encoding \mathcal{C} and recovery \mathcal{R} so that the noisy channel, $\mathcal{R}\mathcal{E}\mathcal{C}$, is as close as possible to a desired unitary U_S .

Performance measures ^{a b c}

Channel fidelity between \mathcal{REC} and U_S

$$f = \frac{1}{n_S^2} \sum_{r,e} |\text{Tr } U_S^\dagger \mathbf{R}_r E_e \mathbf{C}|^2$$

Perfect error correction, $f = 1$, if and only if there are constants α_{re} such that

$$\mathbf{R}_r E_e \mathbf{C} = \alpha_{re} U_S, \quad \sum_{r,e} |\alpha_{re}|^2 = 1.$$

This suggests the **indirect** measure of fidelity, the “distance-like” error,

$$d = \min \left\{ \sum_{r,e} \|\mathbf{R}_r E_e \mathbf{C} - \alpha_{re} U_S\|_{\text{fro}}^2 \mid \sum_{r,e} |\alpha_{re}|^2 = 1 \right\}$$

^aB. Schumacher, *Phys. Rev. A*, (1996)

^bE. Knill, R. Laflamme, *Phys. Rev. A*, (1997)

Direct Fidelity Maximization ^{a b c d}

$$\begin{aligned} \text{maximize } f &= \frac{1}{n_S^2} \sum_{r,e} |\text{Tr } U_S^\dagger \mathbf{R}_r E_e \mathbf{C}|^2 \\ \text{subject to } \sum_r \mathbf{R}_r^\dagger \mathbf{R}_r &= I_C, \quad \mathbf{C}^\dagger \mathbf{C} = I_S \end{aligned}$$

Indirect Fidelity Maximization ^d

$$\begin{aligned} \text{minimize } d &= \sum_{r,e} \|\mathbf{R}_r E_e \mathbf{C} - \alpha_{re} U_S\|_{\text{fro}}^2 \\ \text{subject to } \sum_r \mathbf{R}_r^\dagger \mathbf{R}_r &= I_C, \quad \mathbf{C}^\dagger \mathbf{C} = I_S, \quad \sum_{r,e} |\alpha_{re}|^2 = 1 \end{aligned}$$

- iterating (either) between recovery and encoding \implies local optimum.
- relation between fidelity and distance

$$d = 2n_S(1 - \sqrt{f}), \quad f = (1 - d/2n_S)^2$$

^aM. Reimpell & R. F. Werner, *Phys. Rev. Lett.* (2005)

^bN. Yamamoto, S. Hara, & K. Tsumura, *Phys. Rev. A*, (2005)

^cA. S. Fletcher, P. W. Shor, & M. Z. Win, *Phys. Rev. A* (2007)

Error system uncertainty

- error system is one of a number of possible error systems:

$$\mathcal{E}_\beta = \left\{ E_{\beta e} \mid e = 1, \dots, m_E \right\}, \beta = 1, \dots, \ell$$

- sources of uncertainty
 - different runs of a tomography experiment can yield different error channels.
 - model Hamiltonian $H(\theta)$ dependent upon an uncertain set of parameters θ – take a sample $\{H(\theta_\beta)\}_{\beta=1}^\ell$ will result in a set of error systems
- errors can be mitigated by
 - fault-tolerant methods which rely on several levels of code concatenation
 - robust error correction – include specific uncertainty, e.g., use average error system:

$$\begin{aligned}\mathcal{E}_{\text{avg}} &= \left\{ \sqrt{p_\beta} E_{\beta e} \mid e = 1, \dots, m_E, \beta = 1, \dots, \ell \right\} \\ p_\beta &= 1/\ell \quad (\text{uniform probability})\end{aligned}$$

Quantum Estimation via Convex Optimization

Robert Kosut
SC Solutions, Sunnyvale, CA

collaborators

D. Lidar, A. Shabani E. Yablonovitch, R. Caflisch
USC UCLA

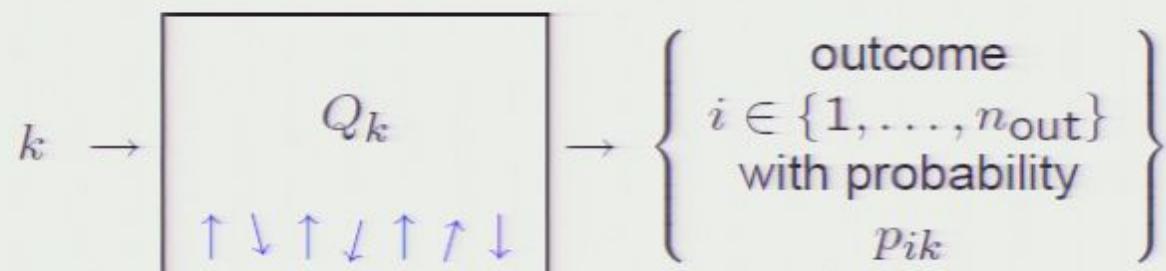
Research supported by the DARPA QuIST Program
(Quantum Information Science & Technology)

Convex Optimization

S. Boyd & L. Vandenberghe, *Convex Optimization*, Cambridge, 2004
Supporting Lecture Slides and Course Videos
<http://stanford.edu/boyd>

Collecting data from independent experiments

- with system in configuration k , repeat experiment ℓ_k times.
 - $k \in \{1, \dots, n_{\text{cfg}}\}$
 - configurations include: prepared initial states, sample times, frequencies, optical settings, applied control fields, ...
- record outcome counts n_{ik}
 - number of times outcome i occurred in configuration k
 - $i \in \{1, \dots, n_{\text{out}}\}$
- repeat ℓ_k times:



State tomography via Maximum Likelihood (ML) estimation

$$\text{minimize} \quad L(\rho) = - \sum_{i,k} n_{ik} \log p_{ik}(\rho)$$

$$\text{subject to} \quad p_{ik}(\rho) = \text{Tr } O_{ik} \rho, \quad \rho \geq 0, \quad \text{Tr } \rho = 1$$

- $L(\rho)$ – negative log-likelihood function
- n_{ik} – outcome counts per configuration (data)
- O_{ik} – system model (POVM)
- convex in ρ