Quantum Tomography and Information

likelihood and entropy in physics

Zdeněk Hradil

Department of Optics, Palacký University

Olomouc, Czech Rep.

Work done in collaboration with J. Řeháček, D. Mogilevtsev. L. Sanchez-Soto, B.Englert, Y-S Teo, B. Stoklasa

Full Program

Lecture 1: Motivation for tomogrphy and MaxLik estimation

•Introduction: inverse problems, Radon and Inverse Radon transformation
•basics of quantum mechanics and estimation theory
•MaxLik estimation and implementation in QM

Lecture 2: General Concepts and Advanced Problems

•Fisher information
•Quantum diagnostics
•Information principles and MLME estimation
•Full scheme for MaxLik tomography
•Resource analysis for tomography of 5 qbits

Lecture 3: Examples of MaxLik tomography

Many, many Examples:
operational phase, absorption tomography, reconstruction of CP maps, **homodyne tomography**, loop detectors, optical scanning devices, Fisher info in quantum interferometry, …

Lesson 1: Motivation and Basic MaxLik Tomo

Linear inverse problems

ML estimation is excellent tool for solving linear inverse problems with constraints (= tomography)

$$I_j = \sum_k c_{jk}\,\mu_k$$

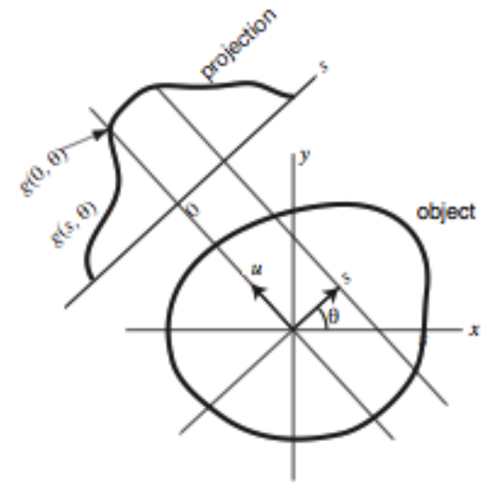
detected mean values I_j, $j = 1,2,\ldots M$
reconstructed signal μ_k $k = 1,2,\ldots N$

Over-determined problems $M > N$
Well defined problems $M = N$
Under-determined problems $M < N$

Tomography and Inverse Radon Transformation

Radon transformation

$$g(s,\theta) = \int dxdy f(x,y)\delta(x\cos\theta + y\sin\theta - s)$$

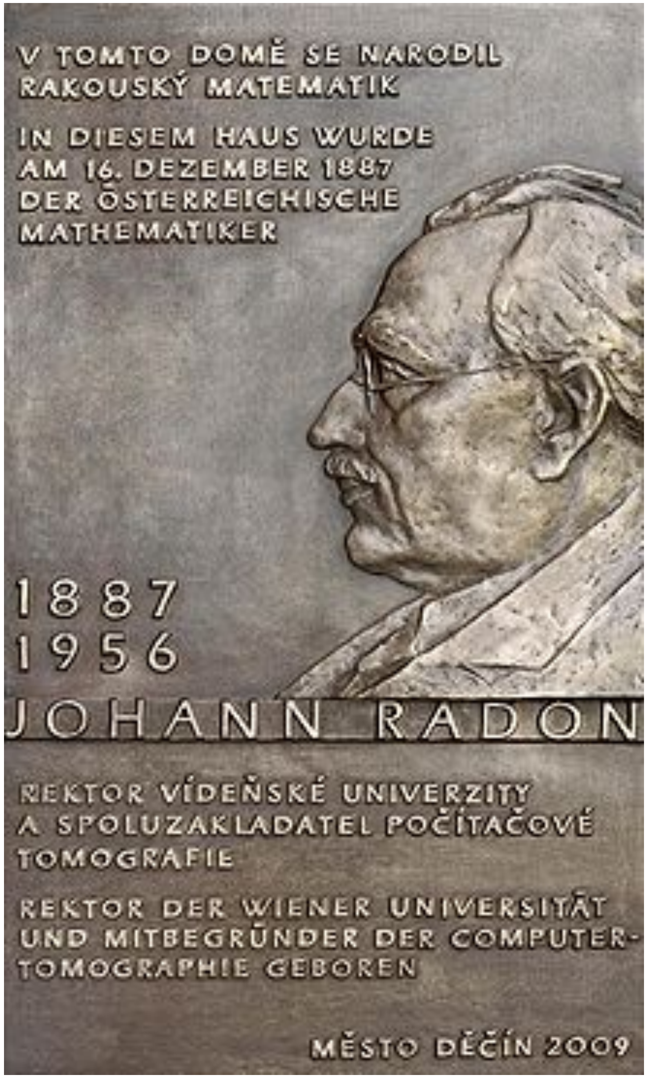


Projection theorem
(ray sum)

$$g(s,\theta) = \int_{-\infty}^{\infty} f(s\cos\theta - u\sin\theta, s\sin\theta + u\cos\theta)\, du$$

Inverse Radon transformation-
Fourier transformation method

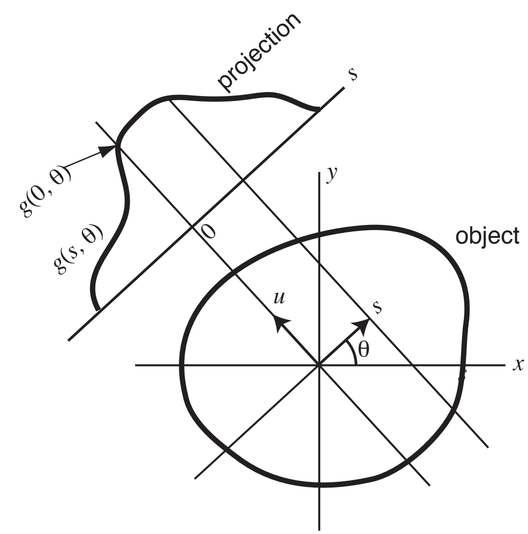
$$G_\theta(\xi) = F(\xi\cos\theta, \xi\sin\theta)) \qquad f(x,y) = F^{-1}G_\theta$$



Johann Radon,
16.12. 1887 Děčín - 25. 5. 1956,Vienna

1. Radon and inverse Radon transformation

- Derive the analytical expression for Radon and Inverse Radon Transformations.



Geometry: Projection along the line

$$x \cos \theta + y \sin \theta - s = 0$$

Radon transformation

$$g(s, \theta) = \int\int_{-\infty}^{\infty} f(x, y)\ \delta(x \cos \theta + y \sin \theta - s) dx dy$$

For obtaining Radon transformation as a Ray sum, define rotated system of coordinates

$$\begin{pmatrix} s \\ u \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

Hence

$$x = s\cos\theta - u\sin\theta$$
$$y = s\sin\theta + u\cos\theta$$

$$\int \delta(s)ds = 1$$

Radon transformation as a Ray sum

$$g(s,\theta) = \int_{-\infty}^{\infty} f(s\cos\theta - u\sin\theta,\, s\sin\theta + u\cos\theta)\, du$$

Inverse Radon transformation follows directly from the definition

$$G_\theta(\xi) = \int_{-\infty}^{\infty} g(s,\theta)\exp(-i2\pi\xi s)ds$$

$$G_\theta(\xi) = F(\xi\cos\theta, \xi\sin\theta)$$

$$f(x,y) = \int_0^\infty \int_0^{2\pi} \xi d\xi d\theta F(\xi\cos\theta, \xi\sin\theta) e^{2\pi i(x\xi\cos\theta + y\xi\sin\theta)}$$
$$= \int_0^\infty \int_0^{2\pi} \xi d\xi d\theta G_\theta(\xi) e^{2\pi i(x\xi\cos\theta + y\xi\sin\theta)}$$

Elements of quantum theory

Probability in Quantum Mechanics:

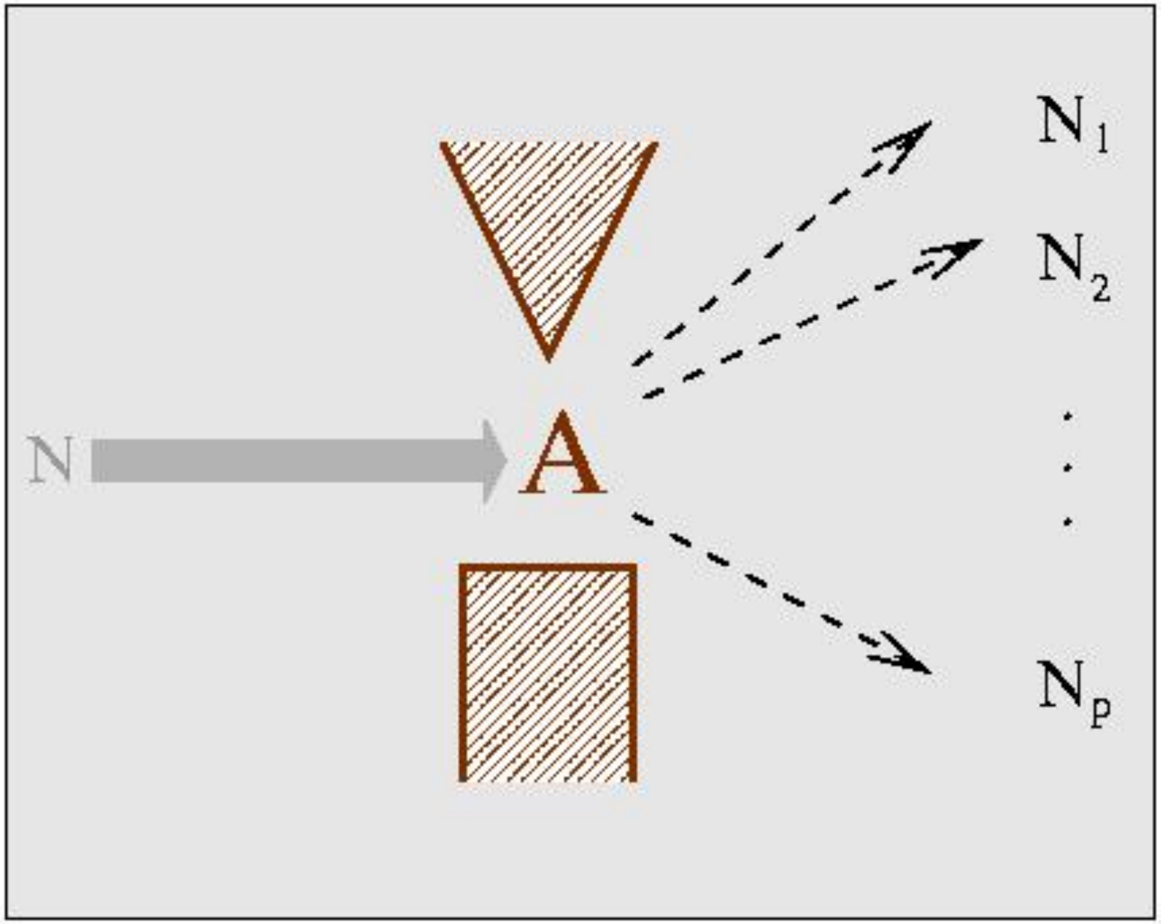
$$p_j = Tr(\rho A_j)$$

Measurement: elements of positive-valued operator measure (POVM) $A_j \geq 0$

Relation of completeness $\Sigma_j A_j = 1$

Signal: density matrix $\rho \geq 0$

Von Neumann Measurement



Estimation Theory in Words

- Variable of interest is a **c-number** θ_{true}
- This variable cannot addressed directly
- Only some variable-dependent **data D** can be detected
- Presence of variable θ_{true} is manifested by **conditional probability distribution** $p(D|\theta_{true})$
- **Estimator** $\theta = \theta(D)$ relates the data to the variable of interest
- Due to the stochastic nature of data there is no unique and deterministic mapping between D and θ.
- The inversion can be formulated just in statistical sense by Bayes theorem

$$p(\theta|D) = p(D|\theta)\, p(\theta)\, p(D)^{-1},$$

prior distribution $p(\theta)$

normalization $p(D) = \int d\theta\, p(D|\theta)\, p(\theta)$

Estimation Theory in Words...

•The quality of estimation should be assessed by **the cost function $C(\theta, \theta_{true})$**

-least square fit $C(\theta, \theta_{true}) = (\theta - \theta_{true})^2$

-maximum likelihood fit $C(\theta, \theta_{true}) = -\delta(\theta - \theta_{true})$

•**The risk function**

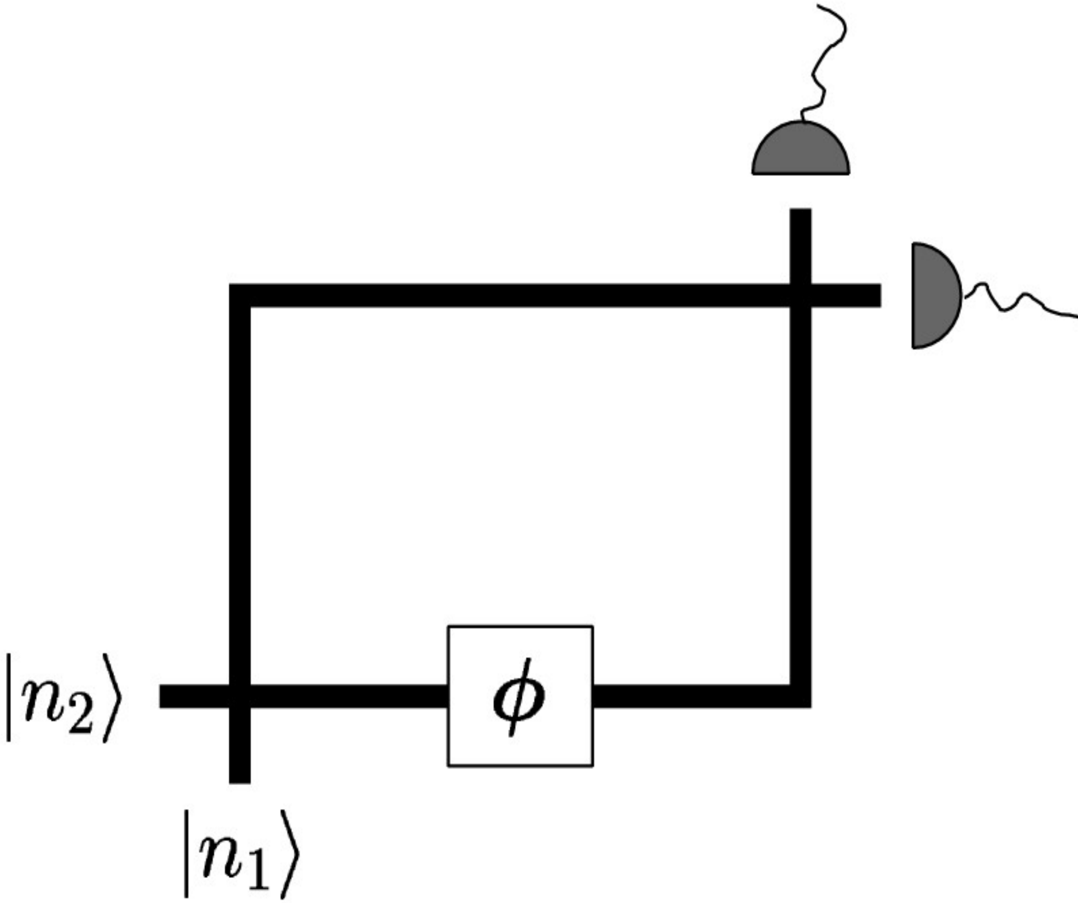
$$R(\theta|D) = \int d\theta_{true} \; C(\theta, \theta_{true}) \; p(\theta_{true}|D)$$

•Optimal strategy minimizes the risk taking into account all prior probabilities and costs

•**Conclusion: for the choice of no prior and delta peaked cost function to minimize risk means to maximize the likelihood**

$$L \sim p(D|\theta) \sim p(\theta|D)$$

Estimation Theory in Drawings



Necessary ingredients:

- Input signal
- Controllable transformation
- Feasible detection

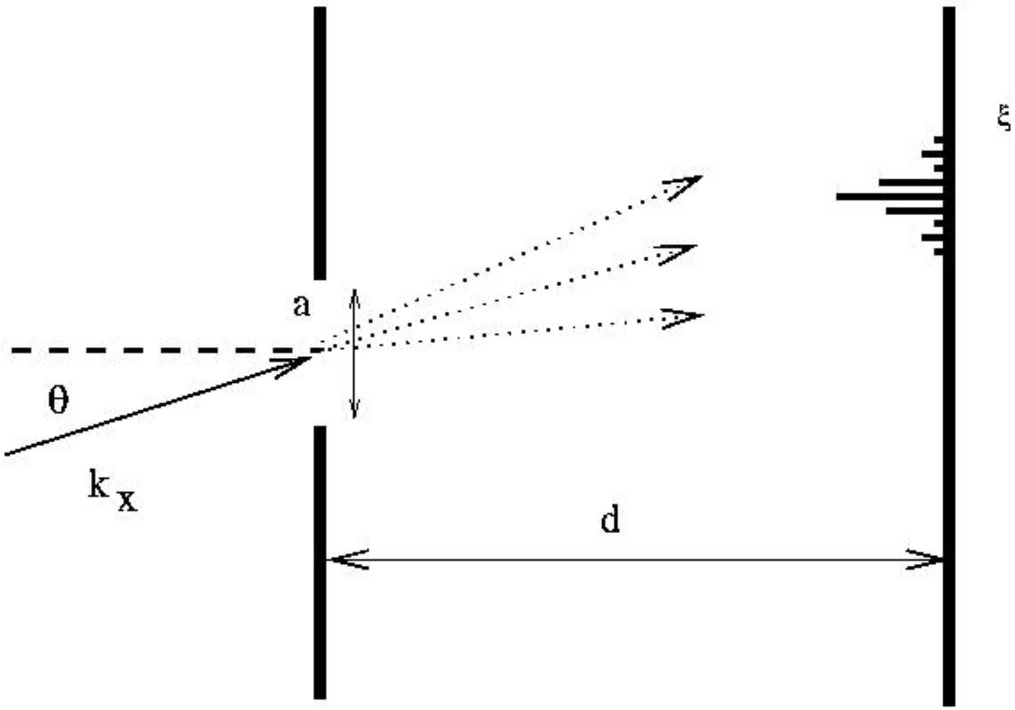
Quantum Estimation Theory

Quantum Estimation Theory
= Quantum Theory + Estimation Theory

Some peculiarities:

- Quantum state ρ plays the role of c-number (matrix) with special constraints ($\rho \geq 0$)
- Quantum measurement must obey uncertainty principle

Motivation: Diffraction on the slit



Detection on the screen may be used as geometrical estimate for impulse since $\theta = \xi/d$ and $p_x = h \sin\theta/\lambda$

Diffraction continues 1 …

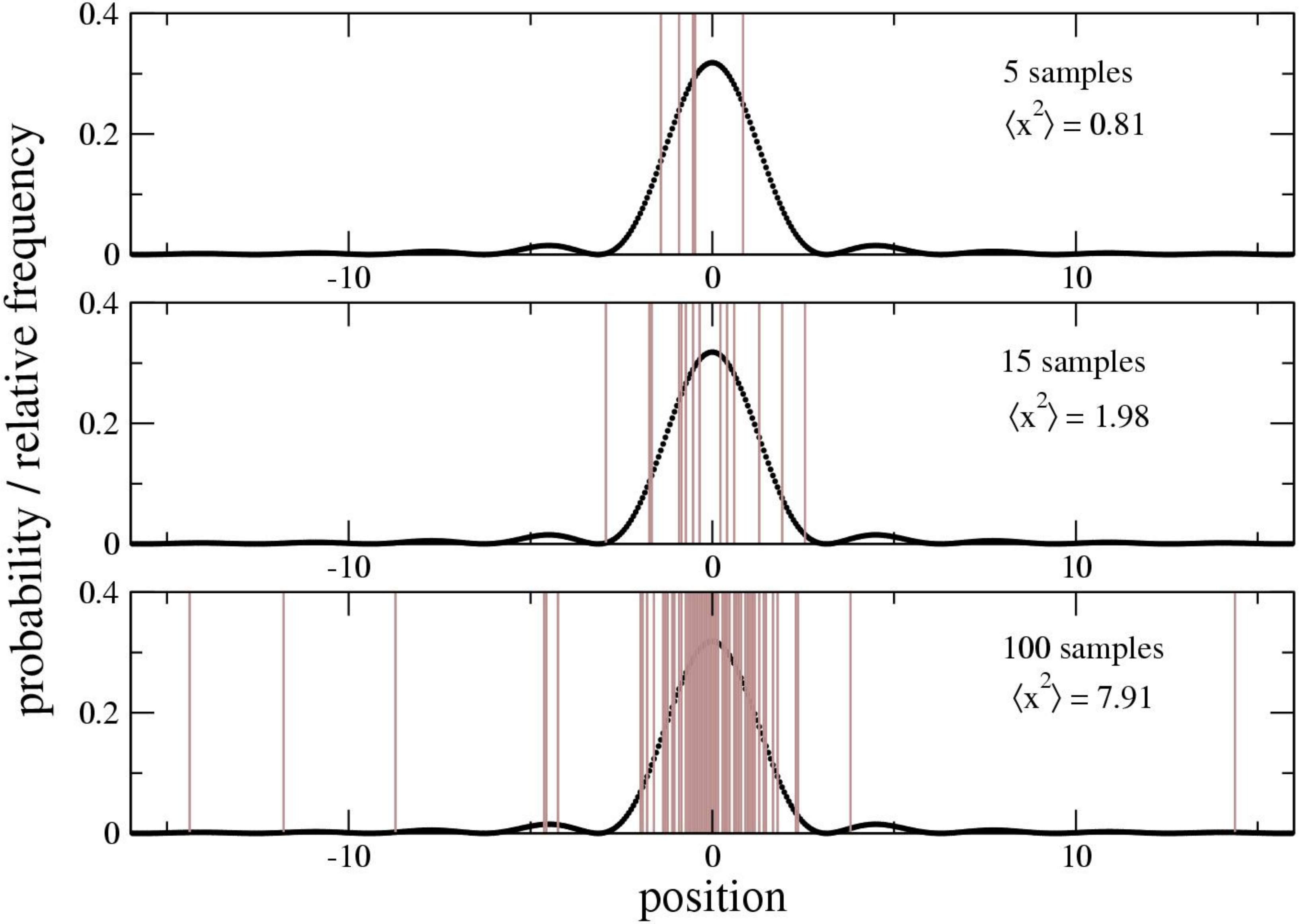
- **The uncertainty is given by wave theory**

  $P(\mu|\nu) = \pi^{-1}\,\mathrm{sinc}^2(\mu - \nu)$; $\mu = \xi\,(\pi a/\lambda d)$, $\nu = p_x\, a/2\hbar$

- **Straightforward but wrong argumentation based on the first minimum of sinc function gives**

$\Delta x = a/2$, $\Delta p_x = (h/a)$ and therefore $\Delta x\,\Delta p_x \sim h/2$!

- But the correctly calculated variance of sinc^2 function gives the infinite width !!
- The estimate of p_x based on single event will be very uncertain !!!
- The remedy is to accumulate the events and relate the estimate to some collective variable (=centre of mass of the interference pattern)
- Proper estimation theory should be formulated with the mathematical statistics.



probability / relative frequency
5 samples
$\langle x^2 \rangle = 0.81$
15 samples
$\langle x^2 \rangle = 1.98$
100 samples
$\langle x^2 \rangle = 7.91$
0.4
0.2
0
-10
0
10
position

Diffraction continues 2 …

- The prediction should be based on some posterior distribution

$$P(v)_{post} = \Pi_\mu \; p(\mu|v)^{N_\mu} = \exp[\Sigma_\mu N_\mu \log p(\mu|v)].$$

Here v is our estimate of some true value v_{true}, which is hidden in detected data μ

- Note: product of detected probabilities is denoted as likelihood L and its logarithm in exponential is called log-likelihood log L
- Significant sampling (N large) $N_\mu = N\, p(\mu|v_{true})$
- Gaussian approximation of log L as the expansion near v_{true} :

$\Sigma_\mu N_\mu \log p(\mu|v) \sim N\, \Sigma_\mu p(\mu|v_{true}) \log p(\mu|v) \sim$

(1st term) $N\, \Sigma_\mu p(\mu|v_{true}) \log p(\mu|v_{true})$

(2nd term) $+ N\, \Sigma_\mu p(\mu|v_{true})\; \partial_v \log p(\mu|v)|_{true}\; (v-v_{true})$

(3rd term) $+ \frac{1}{2} N\, \Sigma_\mu p(\mu|v_{true})\, \partial^2{}_v \log p(\mu|v)|_{true}\; (v-v_{true})^2$

Diffraction continues 3 …

1st term is entropy $S = \sum_\mu p(\mu|v_{true}) \log p(\mu|\, v_{true})$

2nd term is zero since $\sum_\mu p(\mu|v_{true})\ \partial_v \log p(\mu|v)|_{true} =$
$\sum_\mu \partial_v\, p(\mu|v)|_{true}\ (v-v_{true}) = (v-v_{true})\ \partial_v 1 = 0$

3rd term similarly gives the only nonzero contribution

$$F = N \sum_\mu p(\mu|v_{true})\, [\, \partial_v \log p(\mu|v)|_{true}\,]^2$$
$$= N \sum_\mu p(\mu|v_{true})^{-1}\, [\, \partial_v\, p(\mu|v)|_{true}\,]^2$$

F = Fisher information

$$L \sim \exp(S) \exp[\, - \tfrac{1}{2} F\, (v-v_{true})^2\,]$$

This means that parameter estimation is done with the precision 1/F !

Diffraction continues 4 …

Believe or not Fisher information is remedy for uncertainty relations on the slit!

$(\Delta x)^2 = a^2/12$

$(\Delta v)^2 = (a/2\hbar)^2 (\Delta p_x)^2$ and $F = 4\pi^{-1} \int d\mu\, [\partial_\mu \text{sinc}\, \mu]^2 = 4/3$

and therefore $\Delta x\, \Delta p_x = \hbar/2$!

This is not an accident but a consequence of Cramer-Rao inequalities (N=1):

Unbiased estimator : $\Sigma_\mu\, p(\mu|v_{true})\, (v - v_{true}) = 0 \quad / \partial v_{true}$

$$\Sigma_\mu\, \partial v_{true}\, p(\mu|v_{true})\, (v - v_{true}) = 1 \quad \text{/Cauchy-Schwarz inequality}$$

$$\Sigma_\mu\, [p(\mu|v_{true})]^{-1/2}\, \partial v_{true}\, p(\mu|v_{true})\, [p(\mu|v_{true})]^{1/2}\, (v - v_{true}) = 1$$

$$(\Delta v)^2\, F \geq 1$$

Some pedagogical remarks …

$$\Delta A \, \Delta B \geq \tfrac{1}{2} \, |[A,B]|$$

- The meaning of Heisenberg uncertainty principle is pedagogically confusing. Does it mean the constraints on measurement? Which one? Both?
- No, this is the constraint on possible quantum states (see the derivation or see the condition for covariance matrix).
- Heisenberg uncertainty is weaker than Cramer-Rao inequality

$$(\Delta v)^2 \, F \geq 1$$

- Cramer-Rao can be formulated even for simultaneous estimation (measurement) of several parameters.

Maximum Likelihood Estimation (1922)

Sir Ronald Aylmer Fisher, FRS (17 February 1890 – 29 July 1962)
http://digital.library.adelaide.edu.au/coll/special/fisher/papers.html

- Maximum Likelihood (MaxLik) principle is not a rule that requires justification: Bet Always On the Highest Chance!
- Numerous applications in signal analysis, optics, geophysics, nuclear physics,…
- A. Witten, The application of ML estimator to tunnel detection, Inverse Problems 7(1991), 49.
- MaxLik analysis= pea plant experiment of G. Mendel was contrived (too good to be true, statistically ☺)

Statistics matters !

Think of Jan Hendrik Schön

The world's first organic electrical laser!
The smallest ever transistor!

Fisher information

- B. Roy Frieden, *Physics from Fisher information: A Unification*, Cambridge University Press, 1999
- Fisher information for shift invariant distributions $p(x)$

$$p(x|\theta) = p(x-\theta)$$

Amplitude $q(x)$ as generalized coordinate $p(x) = q(x)^2$

$$F= \int dx\ (dp/dx)^2/\ p(x) = \int dx\ (dq/dx)^2$$

Fisher information measures the gradient content of the field $q(x)$ and the "square gradient term" is a part of all Lagrangians, see the second order Lagrange-Euler equations, e.g. classical mechanics $L = \frac{1}{2}\ m\ (dq/dt)^2 - V(q)$

Maximum Likelihood Tomography

- Likelihood **L** quantifies the degree of belief in certain hypothesis under the condition of the given data.
- MaxLik principle selects the most likely configuration
- Information is updated according to the Bayes rule

prior probability ➔ posterior probability

$$P(\rho|D) = P(D|\rho)\, p(\rho)\, [p(D)]^{-1}$$

ML reconstruction: Complete measurement

Log-likelihood for generic measurement $p_i = Tr(\rho A_i)$

$$L(\rho) = \Pi_i \, p_j^{Ni}$$

Normalization $Tr(\rho) = 1$

Constraint $\rho \geq 0$

Maximize the likelihood !!!

Jensen inequality (inequality between geometric and arithmetic means) $\Pi_i \, (x_i/a_i)^{fi} \leq \sum_i f_i \, x_i/a_i$

$$L(\rho)^{1/N} = \Pi_i \, p_j^{fi} \leq (\Pi_i \, a_i^{fi}) \, Tr(R \, \rho)$$

$$R = \sum_i (f_i/a_i) \, A_i$$

Let us chose for extreme $a_i = Tr(\rho A_i)$

Extremal equation $R \, \rho = \rho$

Easy derivation

Differentiate formally the Log-likelihood with the constraint

$$\log L(\rho) = \sum_i N_i \log p_j(\rho) - \lambda \, Tr(\rho) \qquad / \partial \rho_{kl}$$

$$\sum_i N_i / p_j(\rho) \quad (A_i)_{kl} \; |k\rangle\langle l| - \lambda \, \delta_{kl} \, |k\rangle\langle l| = 0 \qquad / \rho$$

$$\sum_i N_i / p_j(\rho) \, A_i \, \rho = \lambda \, \rho \qquad / Tr\rho = 1$$

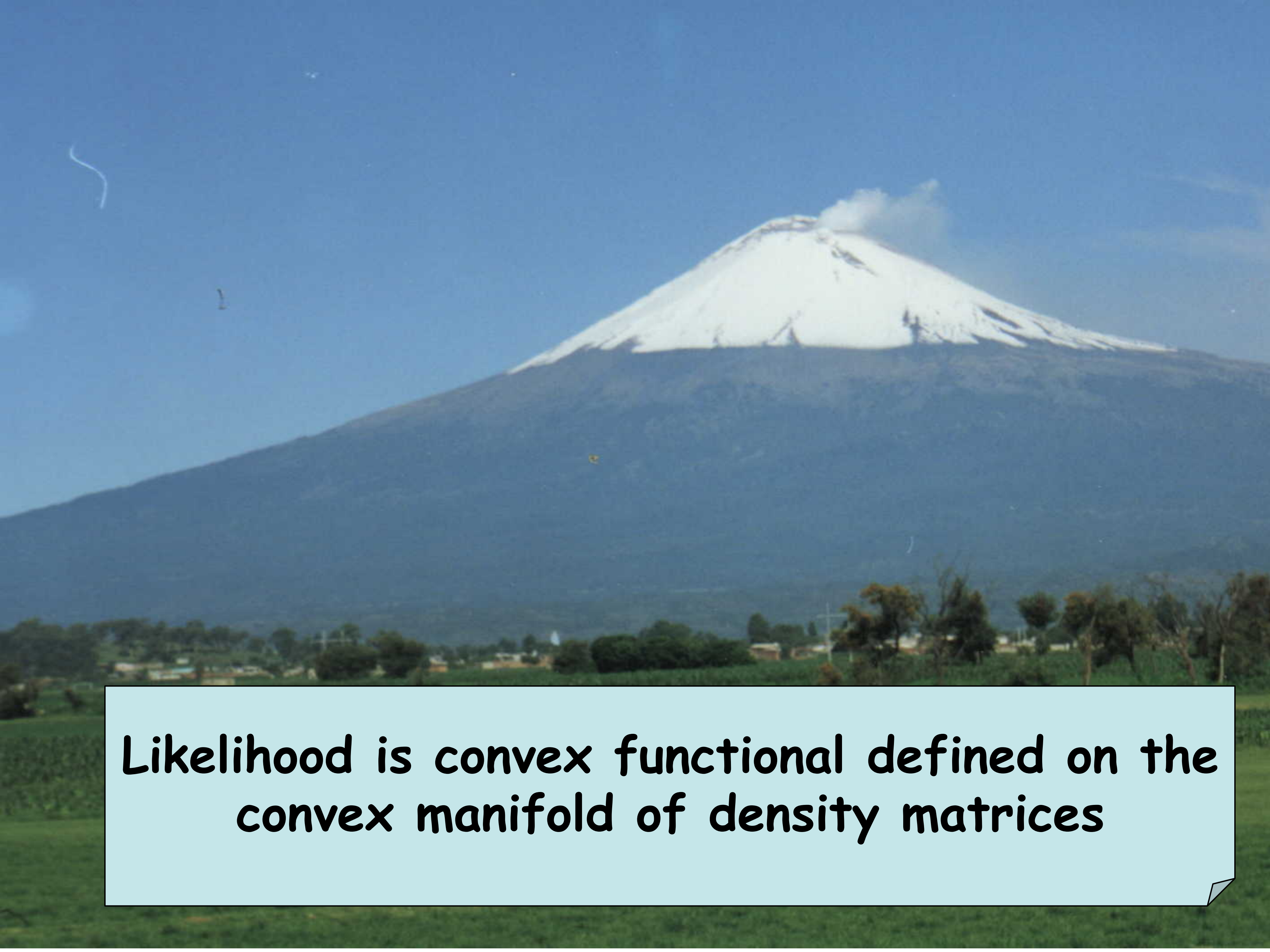
$$R \, \rho = \rho$$

Other hints:

$\rho = \sum_i \lambda_i \, |\varphi_i\rangle\langle\varphi_i|$, $\partial\langle\varphi_i| \; [\langle\varphi_i| A_j \, |\varphi_i\rangle] = A_j \, |\varphi_i\rangle$;

$\rho = \Omega\Omega^\dagger$ $\qquad \partial \, \Omega^\dagger \; Tr(A_j \, \Omega\Omega^\dagger) = A_j \, \Omega$

(Log)-likelihood is convex functional over the convex manifold of density matrices = convex optimization


Likelihood is convex functional defined on the convex manifold of density matrices

Lesson 2: Advanced MaxLik

MaxLik interpretation

Linear inversion

$\Sigma_k A_k \equiv 1$ $\qquad$ $Tr(\rho A_k) = f_k$

MaxLik inversion

$\Sigma_k A'_k = 1_G$ $\qquad$ $Tr(\rho A'_k) \equiv f_k$

where $A'_k = (f_k/p_k) A_k$

MaxLik in terms of Quantum Mechanics

Fluctuations in k-th channel

$$(\Delta\varepsilon_k)^2 = Tr(\rho A_k)\,[1 - Tr(\rho A_k)]$$

All the observations cannot be equally trusted!

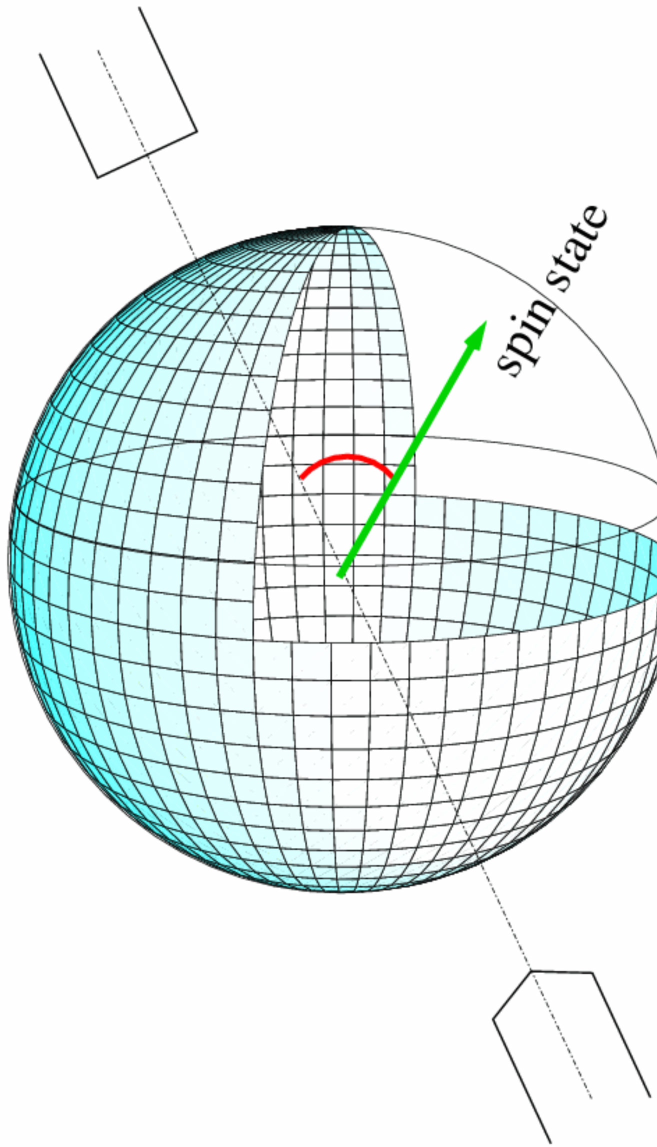
MaxLik estimation in 3 steps:

1. Re-define POVM elements $A_k \Rightarrow \mu_k A_k$
2. Postulate mean values $\mu_k Tr(\rho A_k) = f_k$
3. Postulate the closure relation

$$\Sigma_k \mu_k A_k = 1$$

Why the optimal estimation must be nonlinear:

- Various projections are counted with different accuracy.
- Accuracy depends on the unknown quantum state.
- Optimal estimation strategy must re-interpret the registered data and estimate the state simultaneously.
- Optimal estimation should be nonlinear. MaxLik is doing this.



Objective (Biased) Tomography

•Reconstruction is not equally good in the full Hilbert space: Field of view defines the visible part of the Hilbert space
•How to reconstruct and where to reconstruct are NOT independent tasks in generic tomography schemes

Hradil, Mogilevtsev, Rehacek, Biased tomography schemes: an objective approach, PRL 96, 230401 (2006).

Generic over-complete / un-complete measurements

$$\sum_j A_j = G \geq 0$$

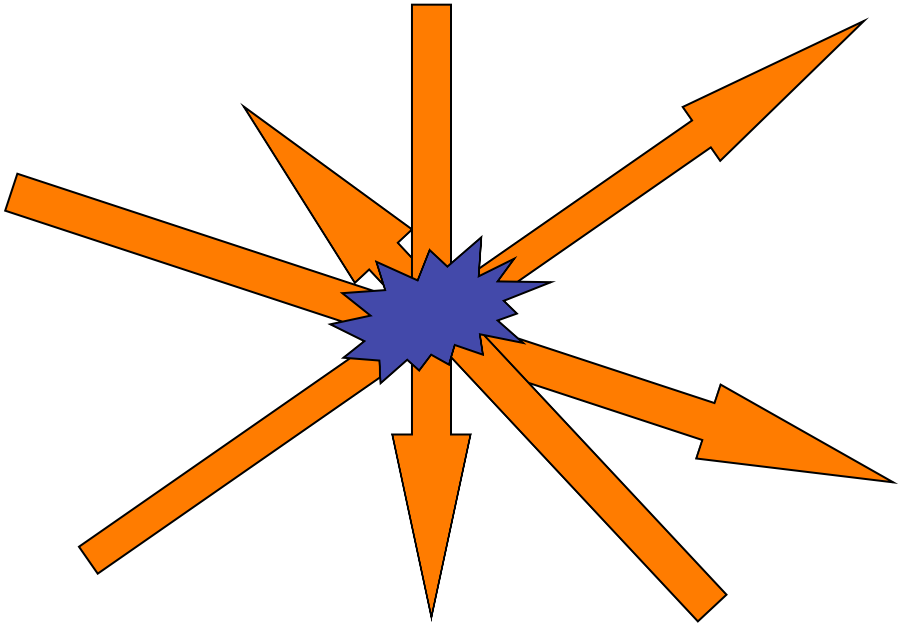
may always be cast in the form of POVM

$$G^{-1/2} / \quad \ldots\ldots\ldots \quad / G^{-1/2}$$

$$\sum_j G^{-1/2} A_j G^{-1/2} = 1_G$$

Every measurement is complete…somewhere !!!

Geometry: overlap of states



Projector
$A_i = |y_i\rangle\langle y_i|$

Overlap of all projectors
$\Sigma_i \; |\langle y_i|\varphi\rangle|^2$

Maximum overlap
$\{\Sigma_i \; |y_i\rangle\langle y_i|\}|\varphi\rangle = \lambda|\varphi\rangle$

Generic reconstruction scheme

Log-likelihood for generic measurement

$$\log L = \sum_i N_j \log p_j / (\sum_k p_k)$$

(probabilities are mutually normalized)

Equivalent formulation: estimation of parameters with Poissonian probabilities and unknown mean λ

(constrained MaxLik by Fermi)

$$\log L = \sum_j N_j \log (\lambda\, p_j) - \lambda \sum_j p_j$$

Extremal equation

$$\mathbf{R}\,\rho = \mathbf{G}\,\rho$$

$$\mathbf{R} = (\Sigma_j p_i)\,/(\Sigma_j N_i)\,\Sigma_k\,(N_k/p_k)\,A_k$$

$$G = \Sigma_i\,A_i$$

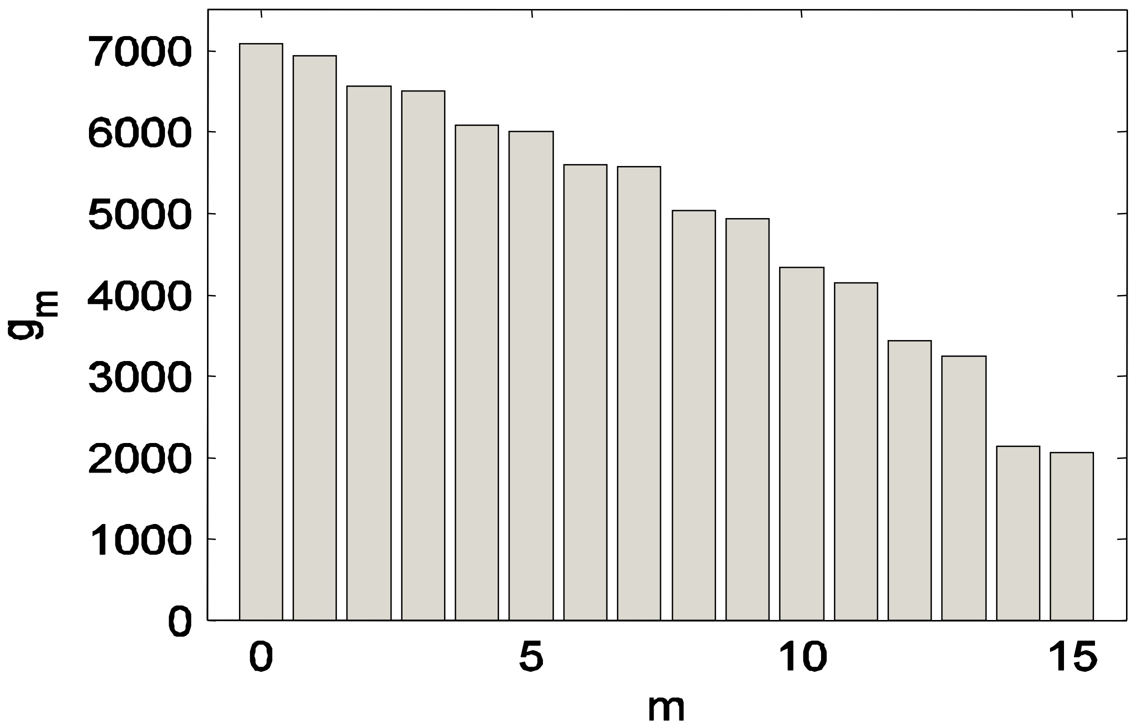
$$\mathbf{R}_G\,\rho_G = \rho_G$$

$$R_G = G^{-1/2} R G^{-1/2},\ \rho_G = G^{1/2}\,\rho\,G^{1/2}$$

Solution in the iterative form

$$\rho_G = R_G \rho_G R_G$$

Field of view in tomography

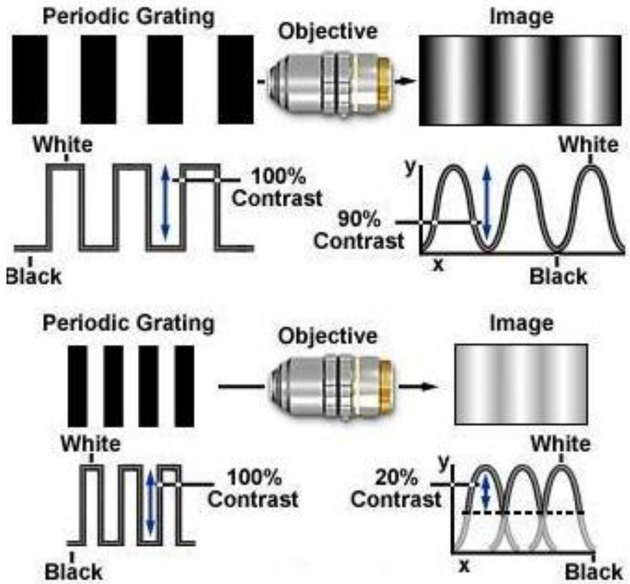


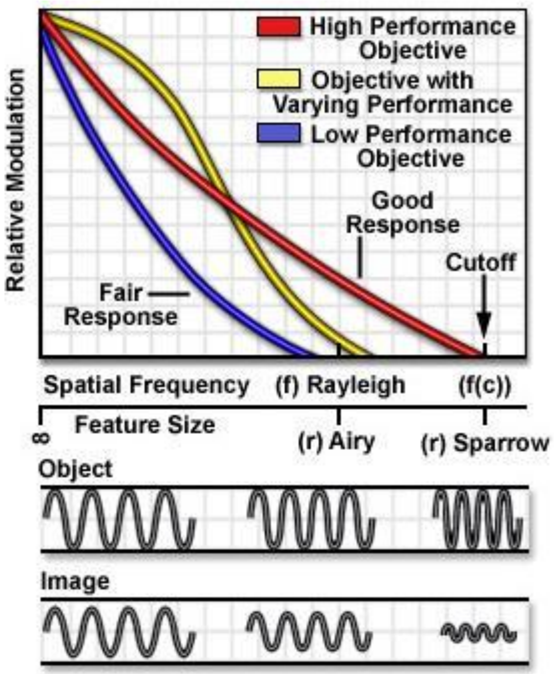
Since $\rho = G^{-1/2}\,\rho_G\,G^{-1/2}$ the solution must be searched in the subspace spanned by nonzero eigenvalues of G !

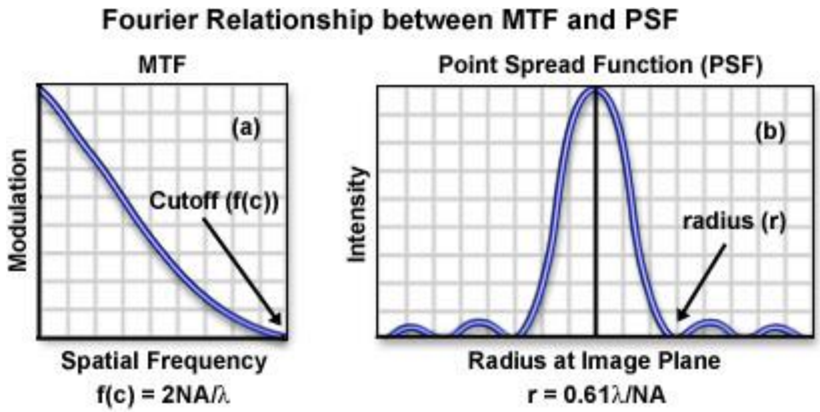
Analogies in optics:

People with better optics should see better!

Optical transfer function







Even classical observations differ in the observed extent and details

Warning !

- Gaussian approximation of the likelihood is not recommended, in general!

$$\log L = - \sum_j (p_j - f_j)^2 / (2\, \sigma^2)$$

- Solution is equally complicated as the original one but there are no links to quantum mechanics!
- Provided that probabilities p_i are not mutually normalized, the solution is even incorrect !

Tomography for quantum diagnostics

- The most likely state does not surely tell everything.
- The result of MaxLik reconstruction is not a single state but a family of states with some posterior distribution.
- MaxLik reconstruction characterizes the estimated state as random variable.
- Any prediction based on tomography e.g. fidelity, Wigner function at origin, etc. is uncertain

$$Q = \langle Q \rangle_{ML} \pm \Delta Q$$

- Quantum state = set of $M = d^2-1$ parameters
- Ω_i … generator basis

$$\rho = \Omega_0/d + \sum_I \rho_i \Omega_i , \quad \rho^{ML} = \Omega_0/d + \sum_I \rho_i^{ML} \Omega_i ,$$

- Relative coordinate $r_i = \rho - \rho^{ML}$, $\mathbf{r} = (r_0, r_1, \ldots r_{M-1})$

- Posterior (multi-normal) distribution

$$P^{\rho}(r) = (2\pi)^{-M/2} (\det F)^{1/2} \exp(-\tfrac{1}{2} \mathbf{rFr})$$

Fisher information matrix, $P = \sum_i p_i$

$$F_{jk} = N^2 \sum_i 1/N_i \, \partial r_j [p_i/P] \, \partial r_k [p_j/P]$$

- Performance measure linear in quantum state

$$z = \mathrm{Tr}(\mathbf{Z}\rho)$$

- Wigner function at origin $\quad Z = \sum_n (-1)^n |n\rangle\langle n|$
- Fidelity $\quad Z = |\psi_{true}\rangle\langle \psi_{true}|$

- Expansion in fixed operator basis

$$Z = \sum_i z_i \Omega_i ; \ |\mathbf{z}\rangle = (z_0, z_1, \dots z_{M-1})$$

- Experimental uncertainty

$$(\Delta z)^2 = \langle \mathbf{z} | \mathbf{F}^{-1} | \mathbf{z} \rangle$$

- Experimental uncertainty relations

$$(\Delta a)^2 (\Delta b)^2 >= |\langle a | F^{-1} | b \rangle|^2$$

- Self-consistency check:

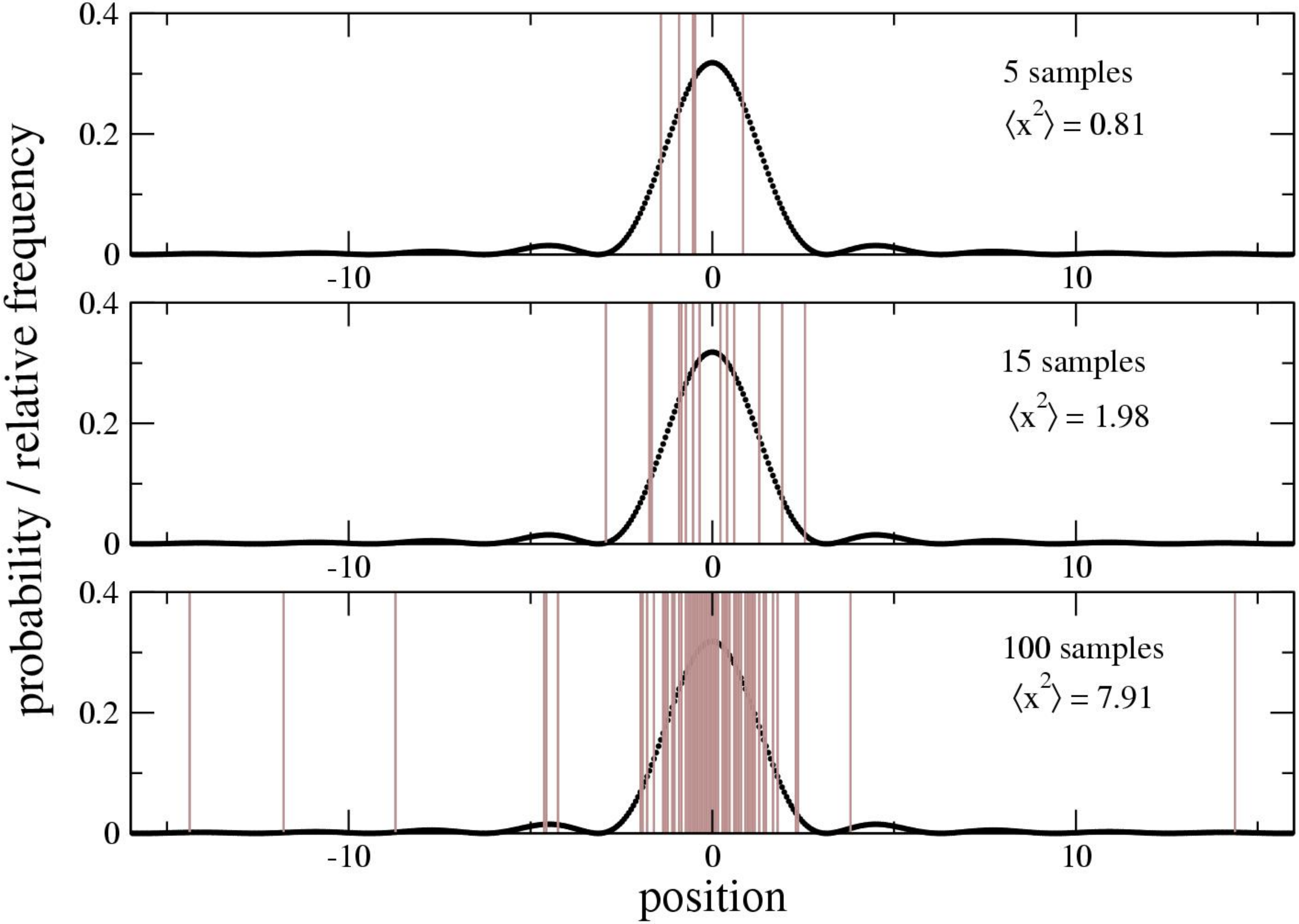
measured data f_k should be compared with the mean values $Tr(\rho A_k)$ within the error $\langle \mathbf{a} | \mathbf{F}^{-1} | \mathbf{a} \rangle$

- **Diagnostics inferred from quantum tomography should be always related to statistical prediction**

$$z = \mathrm{Tr}(Z\rho_{ML}) \pm \{\langle \mathbf{z}|\mathbf{F}^{-1}|\mathbf{z}\rangle\}^{1/2}$$

F … Fisher information matrix
|**z**>… vector with components of Z in the fixed operator basis

- Any tomography scheme should be tailored to a particular purpose, it cannot be universally optimal !!!
- The mean value of the effect $\mathbf{Tr(Z\rho_{ML})}$ and its variance $\langle \mathbf{z}|\mathbf{F}^{-1}|\mathbf{z}\rangle$ are equally important for diagnostic purposes
- The variance term scales with the dimension and depends strongly on the measurement! Indeed, one cannot do any prediction about quantities which have not been measured!



probability / relative frequency
position
5 samples
$\langle x^2 \rangle = 0.81$
15 samples
$\langle x^2 \rangle = 1.98$
100 samples
$\langle x^2 \rangle = 7.91$
0.4
0.2
0
-10
0
10

Information criteria and MaxLik tomography

"The most valuable commodity I know of is information, wouldn't you agree?" (M. Douglas as tycoon Gordon Gekko in the movie Wall Street)

Good statistical models

Many random phenomena, such as those arising in biological and ecological applications, are extremely complex, potentially involving an endless assortment of variables and interactions, „good" models are needed.An optimal statistical model is characterized by three fundamental attributes:

1. Parsimony (model simplicity)
2. Goodness-of-fit (conformity of the fitted model to the data at hand)
3. Generalizability (applicability of the fitted model to describe or predict new data)

Parsimony

- **Law of Parsimony: No more causes should be assumed than those that will account for the effect.**

More philosophy behind:

- Occam's Razor: "Plurality should not be posited without necessity." (Franciscan monk William of Ockham 1285–1349)
- "Everything should be made as simple as possible, but not simpler." (Albert Einstein, 1879–1955).
- "When you hear hoofbeats, think horses, not zebras." (popular adage from medical schools and residency programs)
- "Simplicity is the ultimate sophistication." (Leonardo da Vinci, 1452–1519).

- Laplace's Principle of Insufficient Reasoning: If there is no reason to prefer among several possibilities, than the best strategy is to consider them as equally likely and pick up the average.

All models are wrong, some are useful (George E. P. Box)

Akaike´s information criterion (AIC)

Akaike,IEEE Trans. Auto Control 19, 716 (1974)

Rationale behind AIC

- Could MaxLik be used for comparing various models?
- No, MaxLik favours overfitting: More complex model means better fit!!
- Akaike´s suggestion: Use Mean LogLik instead of LogLik itself !!!

- Akaike´s Information Criterion- remove the bias from MaxLik

$$AIC = \log L(x|\theta_{ML}) - M$$

M..... dimension of parameter space θ

Heuristic sketch of AIC

Ingredients needed for the „proof"

$I(\theta_0, \theta + \Delta\theta) = D(p = f(x|\theta_0) || q = f(x|\theta_0 + \Delta\theta)) = \frac{1}{2} ||\Delta\theta||_F$

The variance $||\Delta\theta||_F = \langle \Delta\theta | \mathbf{F} | \Delta\theta \rangle$

Fisher information matrix

$$F_{ij} = \langle \partial/\partial\theta_i \, p(x|\theta) \, \partial/\partial\theta_j \, p(x|\theta) \rangle$$

Variance fluctuates with chi-square distribution with the M degrees of freedom

$$\langle \log L(x|\theta_{ML}) \rangle = \log L(\theta_{ML}) - M$$

Schwarz and Bayesian Information Criterion (BIC)

Schwarz, Annals of Stat. 6, 461 (1978)

Konishi, Ando, Imoto, Biometrica 91, 27 (2004)

Penalized MaxLik estimation

Hint: Consider the averaged Likelihood. The normalization term is state independent but dimension dependent!

Modified Schwarz information

$$I_{MS} = \log L(\rho) - \tfrac{1}{2} M \log N + \tfrac{1}{2} M \log(2\pi) - \tfrac{1}{2} \log \det F$$

M … dimension of estimated variable (density matrix)
N … dimension of data set

Rationale behind BIC

MATHEMATICS:

$$\int d\theta^M \exp(-N/2\, \boldsymbol{\theta}^T F\, \boldsymbol{\theta}) = (2\pi/N)^{M/2} (\det F)^{-\frac{1}{2}}$$

$$(I_{MS} = \log L(\rho) - \tfrac{1}{2} M \log N + \tfrac{1}{2} M \log(2\pi) - \tfrac{1}{2} \log \det F)$$

PHYSICS:

MaxLik always improves with growing dimension of parameter space. But there is a difference if just noise or some missing component is fitted.

Entropy and quantification of ignorance

Yong Siah Teo, Huangjun Zhu, B-G Englert, J. Řeháček, Z. Hradil,
Quantum-State Reconstruction by Maximizing Likelihood and Entropy,**Phys. Rev. Lett. 107, 020404 (2011)**

MLME estimation

Likelihood $L(\rho)$ quantifies the knowledge

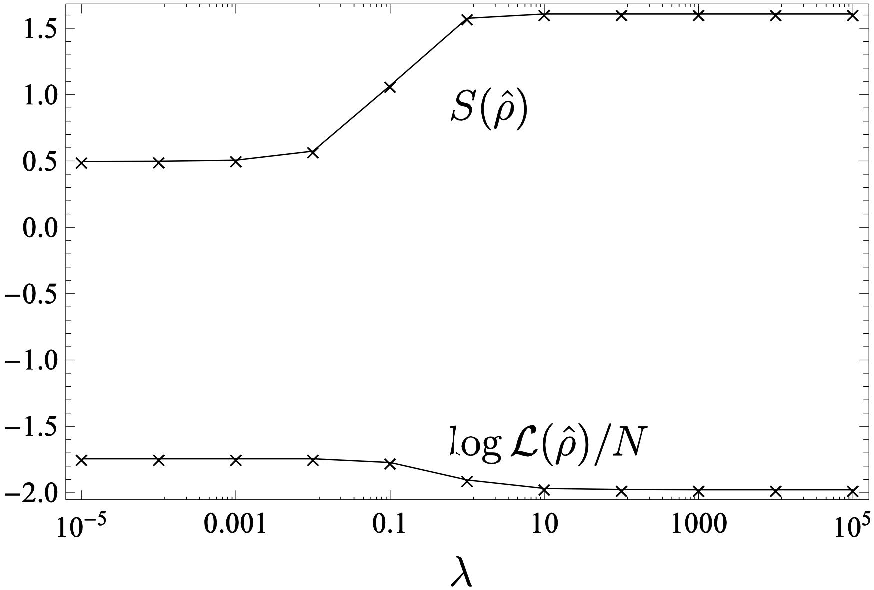
Entropy $S = -Tr(\rho \log \rho)$ quantifies the ignorance

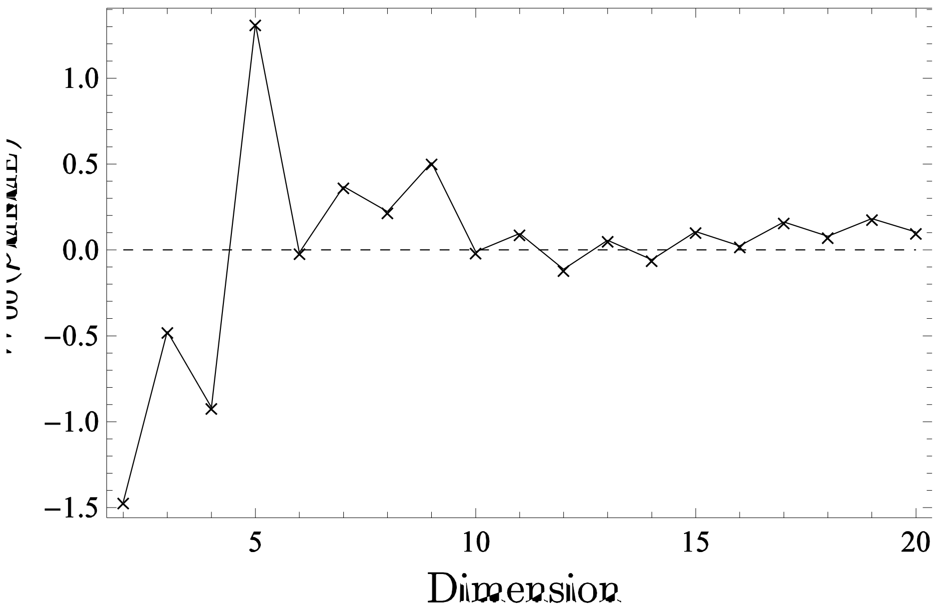
$$I(\lambda,\rho) = \lambda\, S(\rho) + 1/N \log L(\rho)$$

In the limit $\lambda = 0$ we are searching for the most likely states with the highest entropy.

MLME is robust and always selects the single solution.

Some MLME results





Left panel: As lambda decreases entropy and likelihood sets their optimal values;
Right panel: State with positive value of Wigner function in 20 dim Hilbert space is estimated as non classical with mild negativity in low dimensional spaces.

Resource analysis

- To control the quantum system means to control all relevant errors….

- Pure state in dimension d: 2d -1 real parameters

Estimation is not a convex problem…

- Density matrix $d^2 - 1$ real parameters

Fisher info matrix: $\frac{1}{2}(d^2-1)(d^2-2)$ real parameters

- CP maps: $d^2(d^2-1)$ real parameters

Fisher info matrix for CP maps: $\frac{1}{2}d^2(d^2-1)(d^4-d^2-1)$ real parameters

Quantum computation with 5 qbits: $d = 2^5 = 32$
Quantum state: ~ 10^3 parameters
Fisher info: ~ 10^6 parameters
CP maps: ~ 10^6 parameters
Fisher info of CP maps: ~ 10^{12} parameters

Lesson 3: Examples of MaxLik Tomo

Several examples

- Phase estimation
- Transmission tomography
- Tomography of CP maps
- Reconstruction of photocount statistics
- Image reconstruction
- Vortex beam analysis
- Quantification of entanglement
- Reconstruction of neutron wave packet
- Reconstruction based on homodyne detection
- Full reconstruction based on on/off detection
- Reconstruction of coherent matrix

The recommended MaxLik strategy

- Proper description of detected signal (quantum measurement)
- Specification of the field of view (domain of reconstructed signal)
- Solution of equation for MaxLik extremal state (for example iterative solution)
- Error analysis of Fisher information matrix

- Diagnostics of the noise of desired variable

Operational phase concepts

Interference of classical waves

$$I_{3,4} = I(1 \pm V\cos\vartheta) \qquad I_{5,6} = I(1 \pm V\sin\vartheta)$$

Operational phase Noh, Fougeres, Mandel

$$e^{i\vartheta} = \frac{I_3 - I_4 + i(I_5 - I_6)}{\sqrt{(I_3 - I_4)^2 + (I_5 - I_6)^2}}$$

Phase estimator maximizes the Gaussian likelihood

$$\mathcal{L}(\vartheta) \propto \prod_{i=3,4,5,6} \exp\left(-\frac{(I_{est,i} - I_i)^2}{2\sigma^2}\right)$$

More phase concepts ...

For interference of particles with Poissonian statistics

$$\mathcal{L} = \Pi_i (I_i)^{N_i}$$

MaxLik phase estimates is given as

$$exp(i\theta) \propto \frac{N_3 - N_4}{N_3 + N_4} + i\frac{N_5 - N6}{N_5 + N_6}$$

Suggest your own estimator for another statistics…

J. Řeháček, et al.:Testing of quantum phase in matter-wave optics, Phys. Rev. A 60, (1999) 473—479; J. Řeháček, et al.Testing operational phase concepts in quantum optics, J. Opt. B Quant and Semiclassical Opt 3, (2000), 237-244.

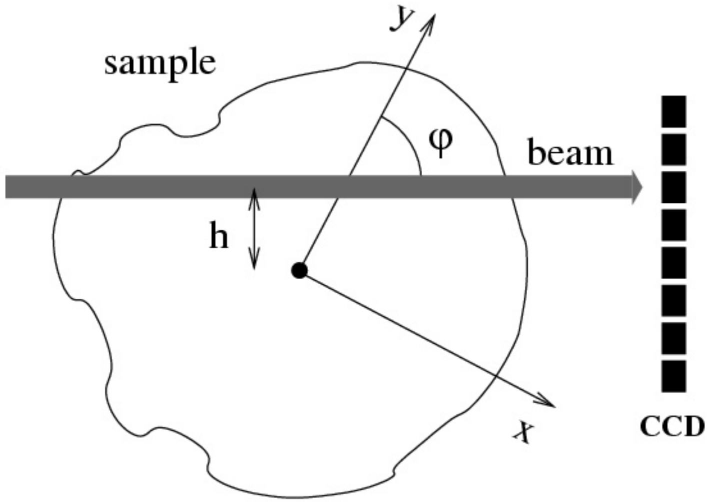
Reconstruction of q-bit states

- **Spin (qbit) :** *Z. Hradil, et al., Phys. Rev. A 62, (2000) 014101-1.*
- **Entangled qbits : J***. Řeháček, et al., Phys. Rev. A 63, (2001) 040303 (R).*

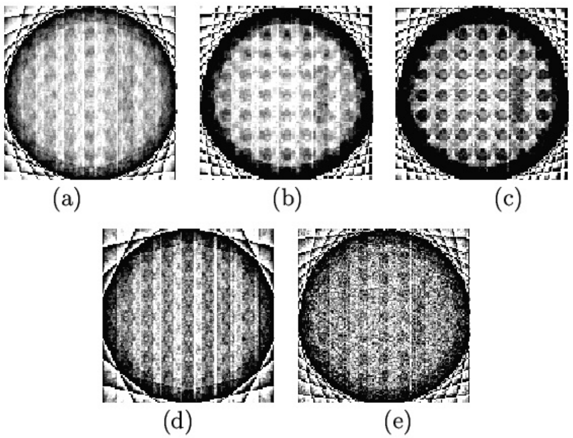
(Neutron) Transmission tomography

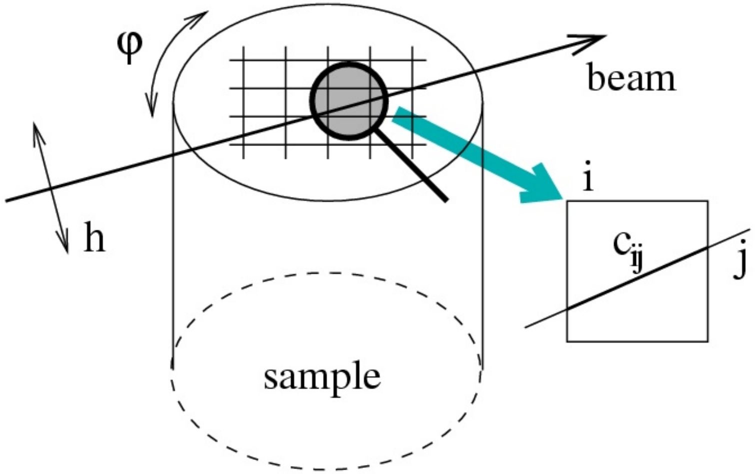
- Exponential attenuation

$$I(h,\theta) = I_0 \, e^{-\int_{\sigma(h,\theta)} \mu(x,y)\, d\sigma}$$

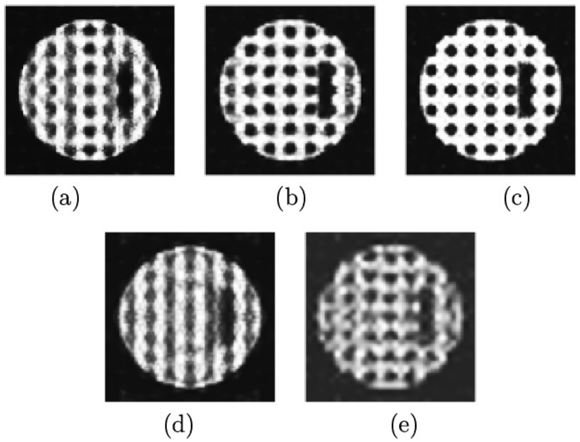


Filtered back projection





Maximum likelihood

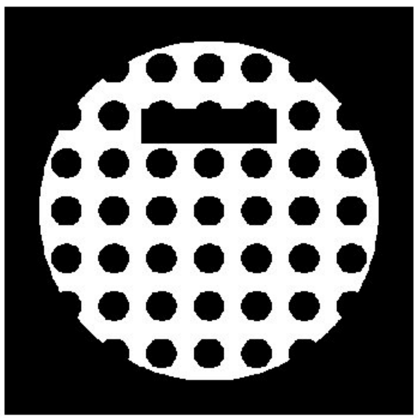


J. Řeháček, Z. Hradil, M. Zawisky, W. Treimer, M. Strobl: Maximum Likelihood absorption tomography, Europhys. Lett. 59 694- 700 (2002).

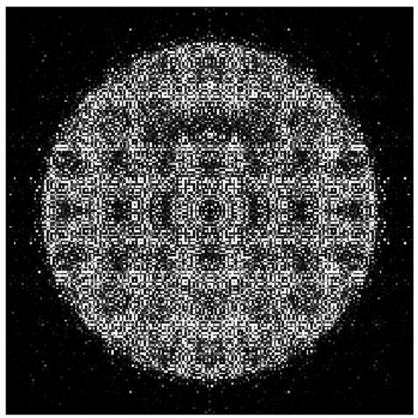
MLME=MaxEnt assisted MaxLik

Numerical simulations using 19 phase scans, 101 pixels each (M=1919)

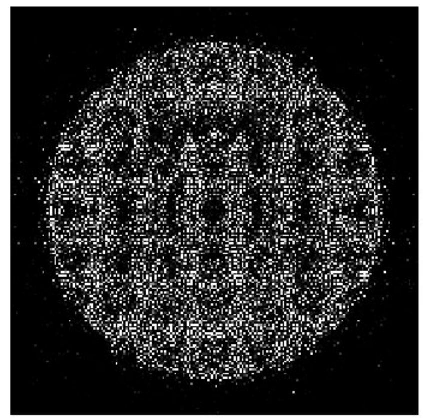
Reconstruction on the grid 201x 201 bins (N= 40401)



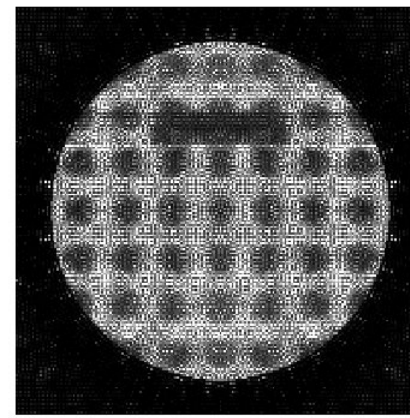
Object



MaxLik1

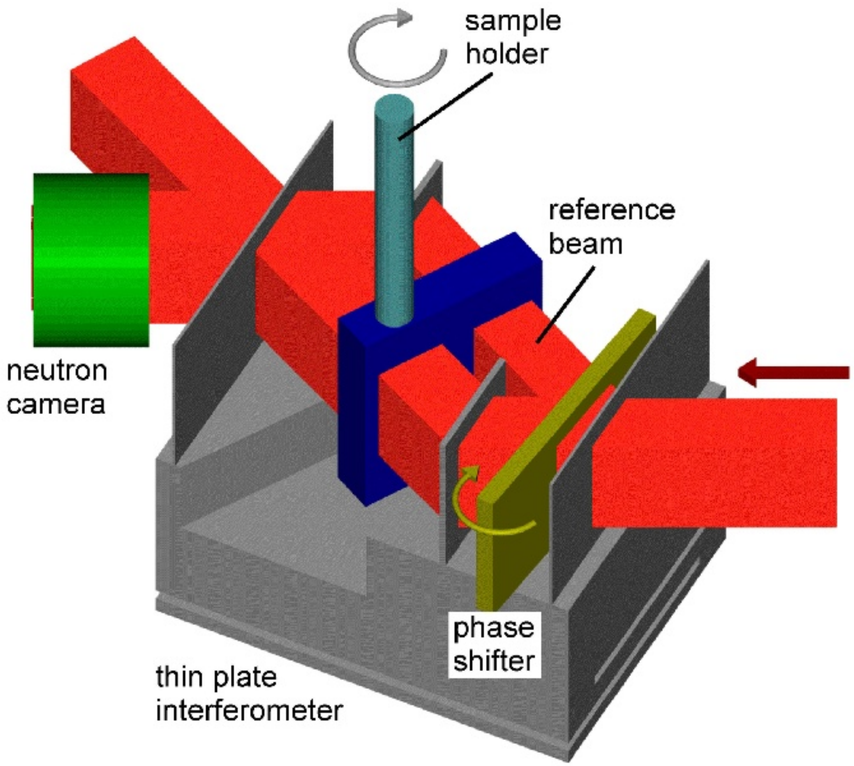


MaxLik2

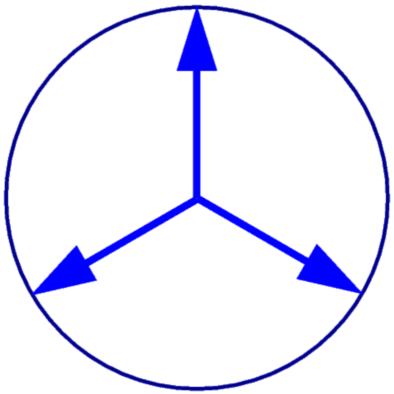


MaxEnt+Lik

Neutron Phase Tomography

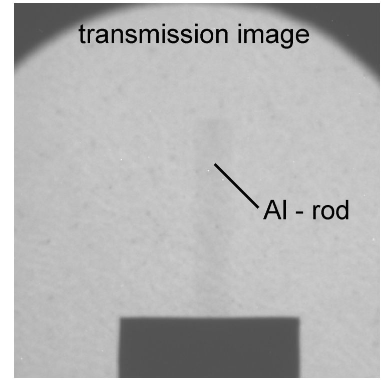


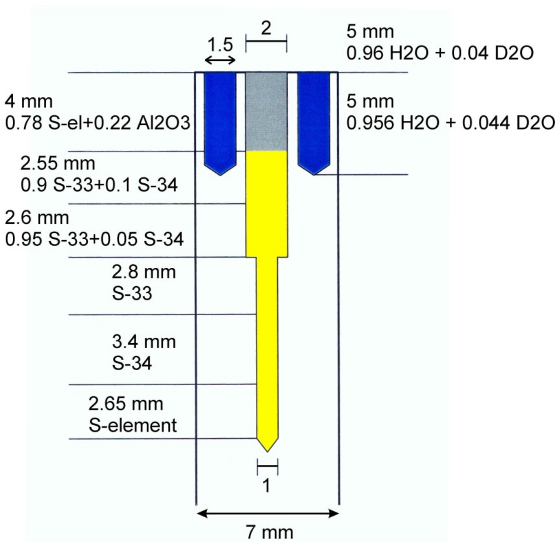
Perfect crystal interferometer

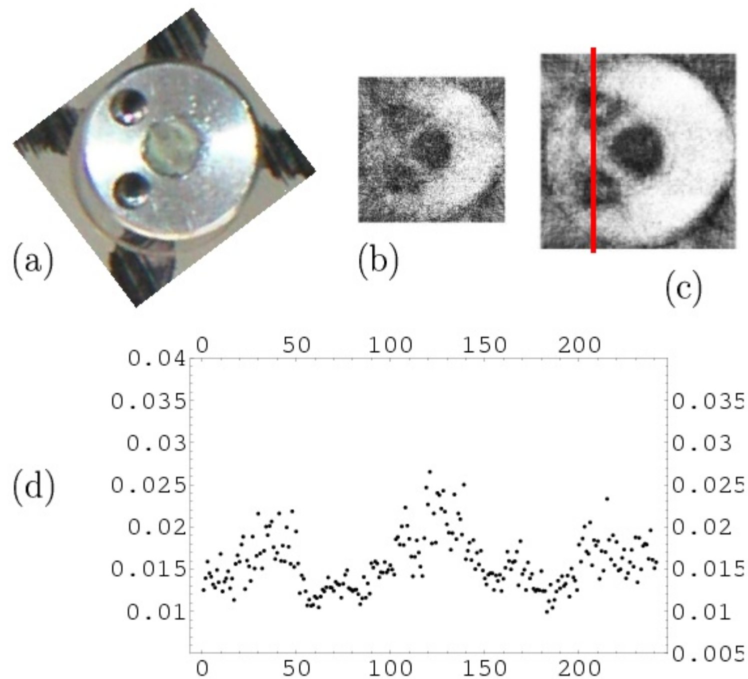


3 reference phases

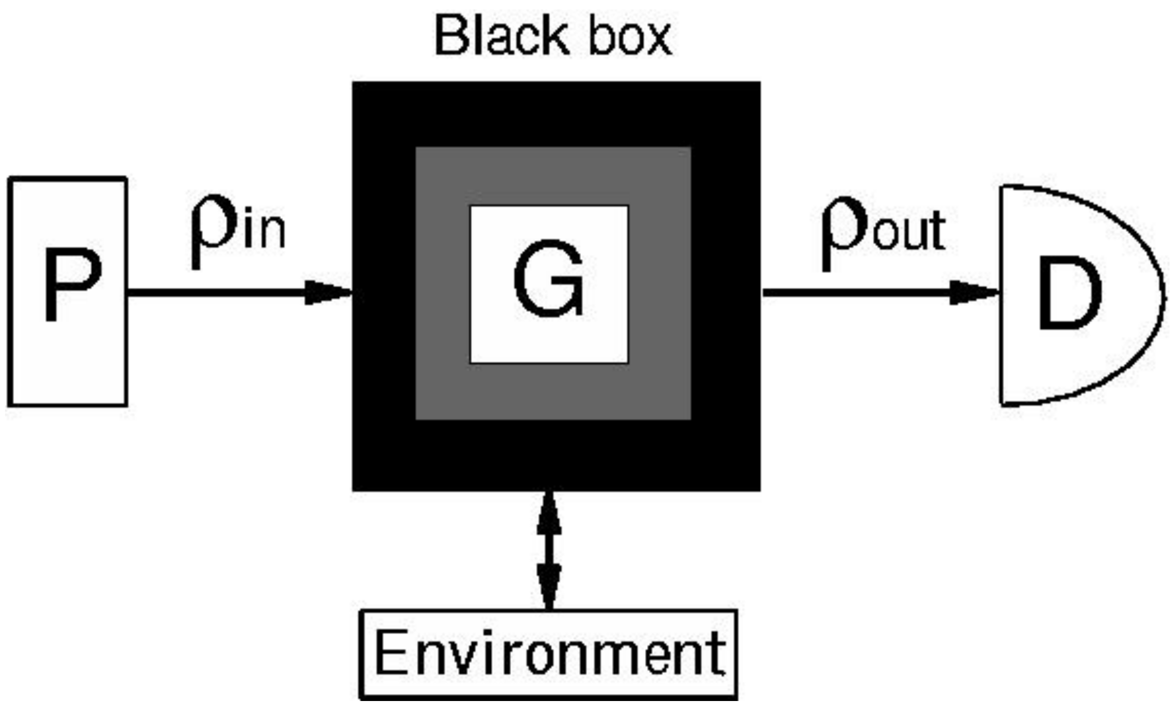
Phase Contrast Tomography with neutrons







Tomography of CP maps



J.Fiurášek, Z. Hradil, Phys. Rev. A 63 , (2001) 020101(R); M. Ježek et al., Phys. Rev.A68 (2003) 012305.

Complete Positive map K ➔ H

Kraus decomposition of

$\rho_{out} = \sum_i A_i \rho_{in} A_i^\dagger$, $\sum_i A_i^\dagger A_i = 1$

Jamiolkowski isomorphism : E ≥ 0 is positive semidefinite operator on HxK : $\rho_{out} = Tr_H[E \ \rho_{in}^T \times I_K]$

Constraints $Tr_K[E] = I_H$

CP tomography: POVM Π_{kl} measured for several density matrices

Probability: $p_{kl} = Tr(E \rho_k^T \times \Pi_{kl})$

Log-likelihood for CP maps

$$\log L(E) = \sum_{kl} f_{kl} \log p_{kl}(E) - Tr(\Lambda E)$$

Coupled nonlinear operator equations:

$$E = \Lambda^{-1} D E D \Lambda^{-1}$$

$D = \sum_{kl} (f_{kl}/p_{kl}) \rho_k^T \times \Pi_{kl}$ and $\Lambda = \{Tr_K [DED]\}^{1/2}$

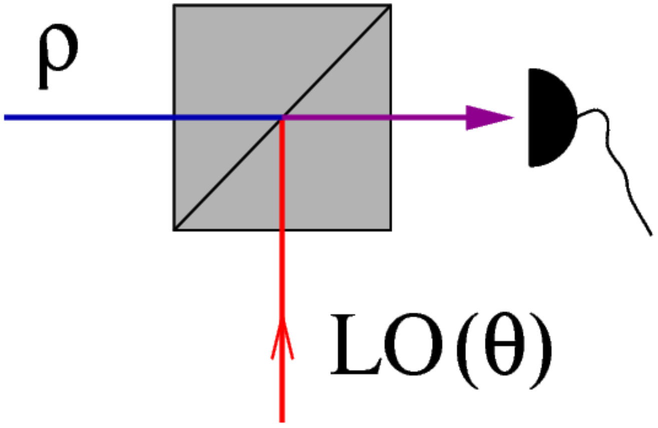
Quantum tomography based on homodyne detection

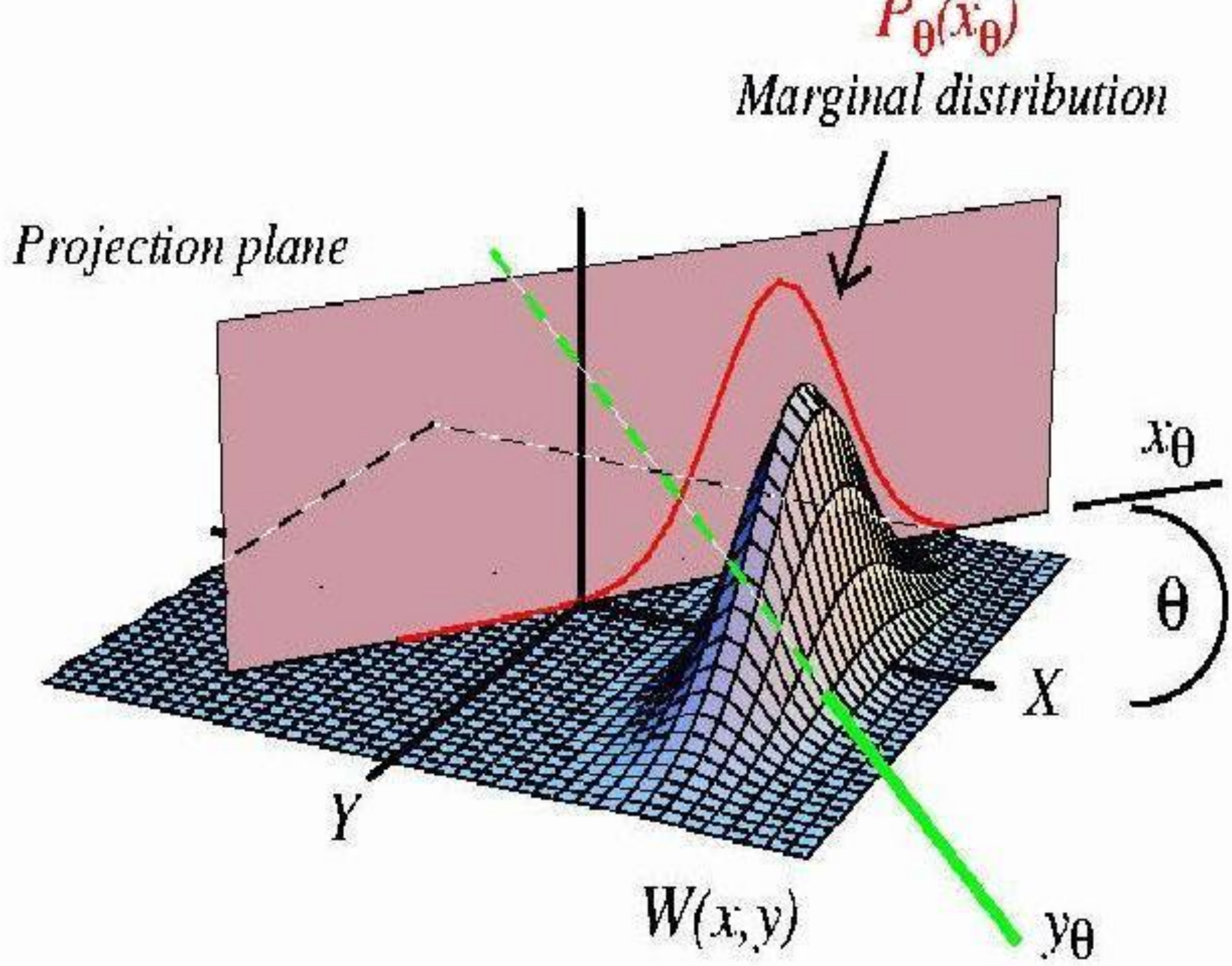
- rotated quadratures

$$\hat{x}_\theta = \sqrt{\frac{\hbar}{2}}\left[\hat{a}e^{-i\theta} + \hat{a}^\dagger e^{i\theta}\right], \qquad \hat{x}_{\theta+\pi/2} = \frac{\sqrt{\hbar}}{i\sqrt{2}}\left[\hat{a}e^{-i\theta} - \hat{a}^\dagger e^{i\theta}\right]$$

- marginal distributions for $P_\theta(x_\theta)$

Homodyne measurement





$P_\theta(x_\theta)$
Marginal distribution
Projection plane
x_θ
θ
X
Y
$W(x,y)$
y_θ

Parameters used in realistic experiments and simulations

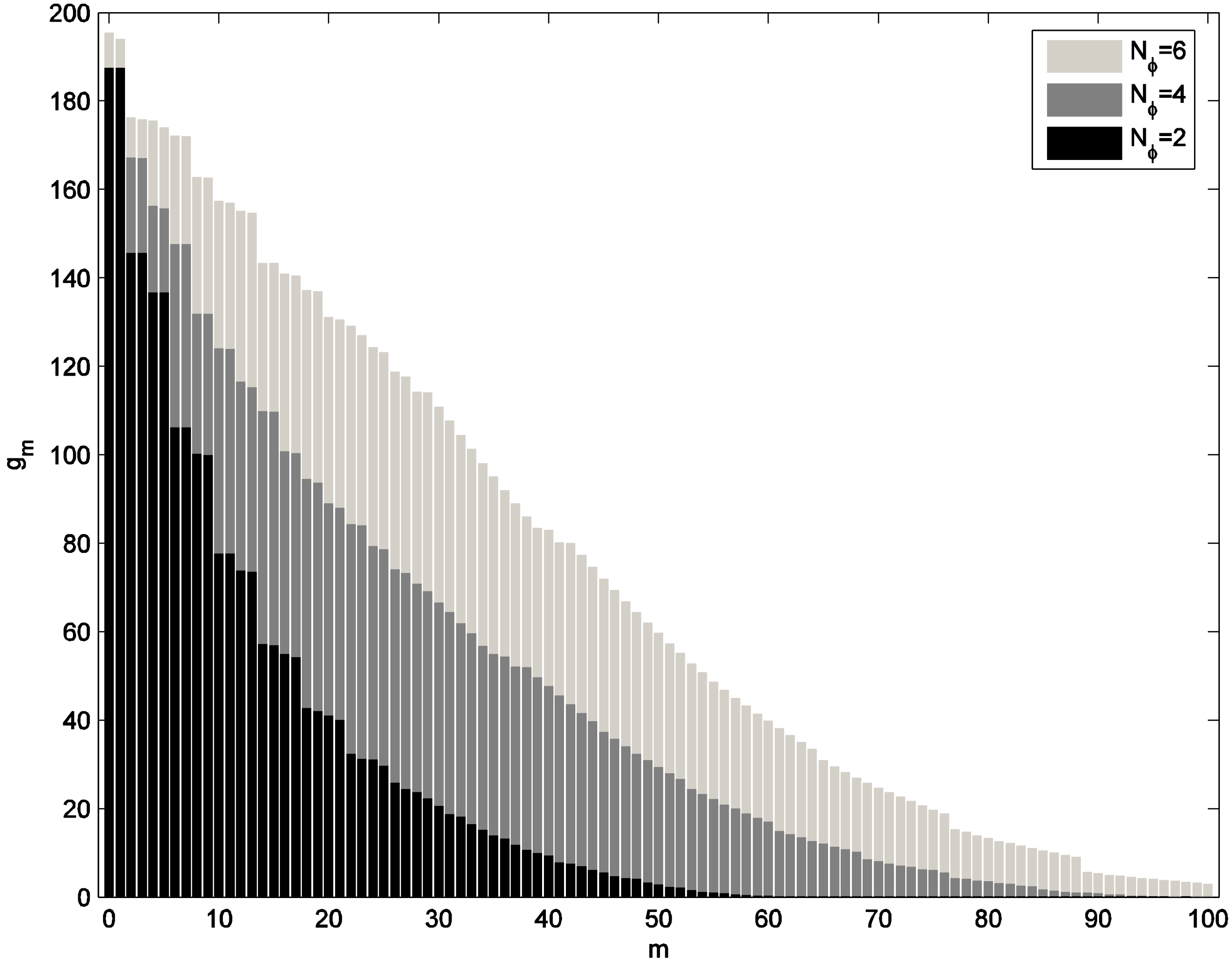
- **Standard setup used for detection of negative Wigner function**

- **6 phase cuts in phase space, efficiency v=0,8**

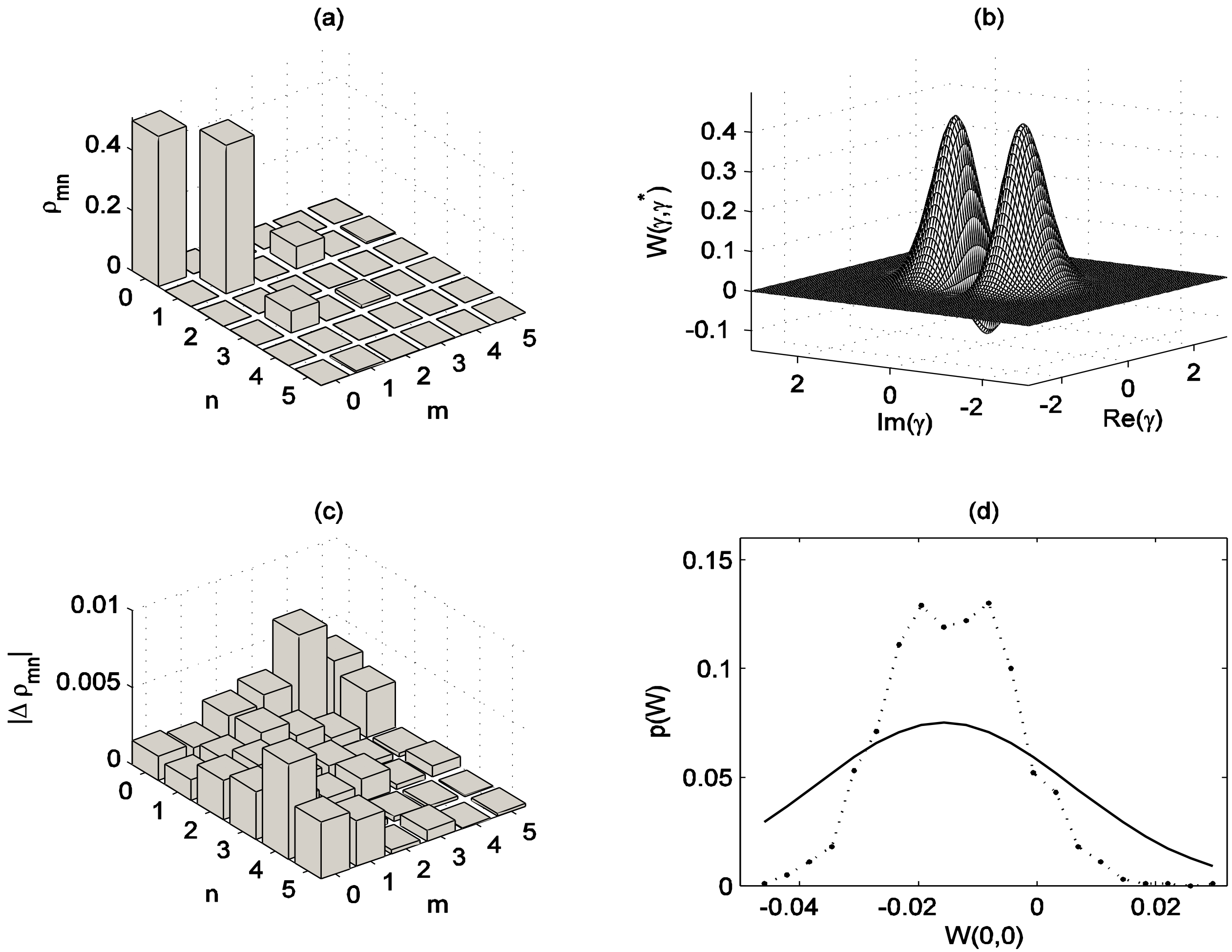
- **1,2 .10^5 detected events accumulated into 64 bins for each phase cut**

- **ML estimation using 1000 iterations**
- **Simulation repeated 1000 times**
- **Simulations from** *Řeháček, Mogilevtsev, Hradil, NJP 10,043022 (2008)*

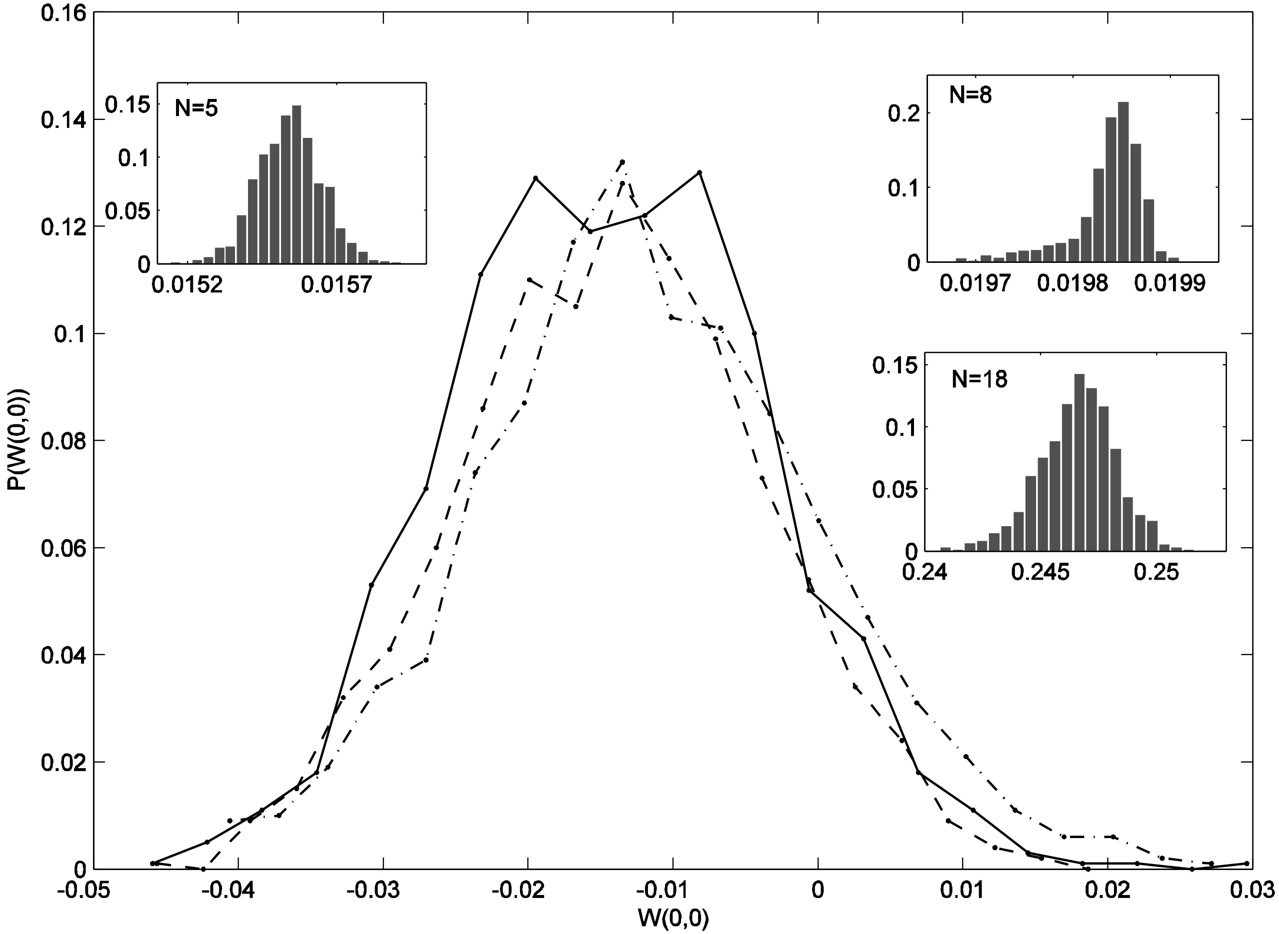
Homodyne tomography for several phase cuts



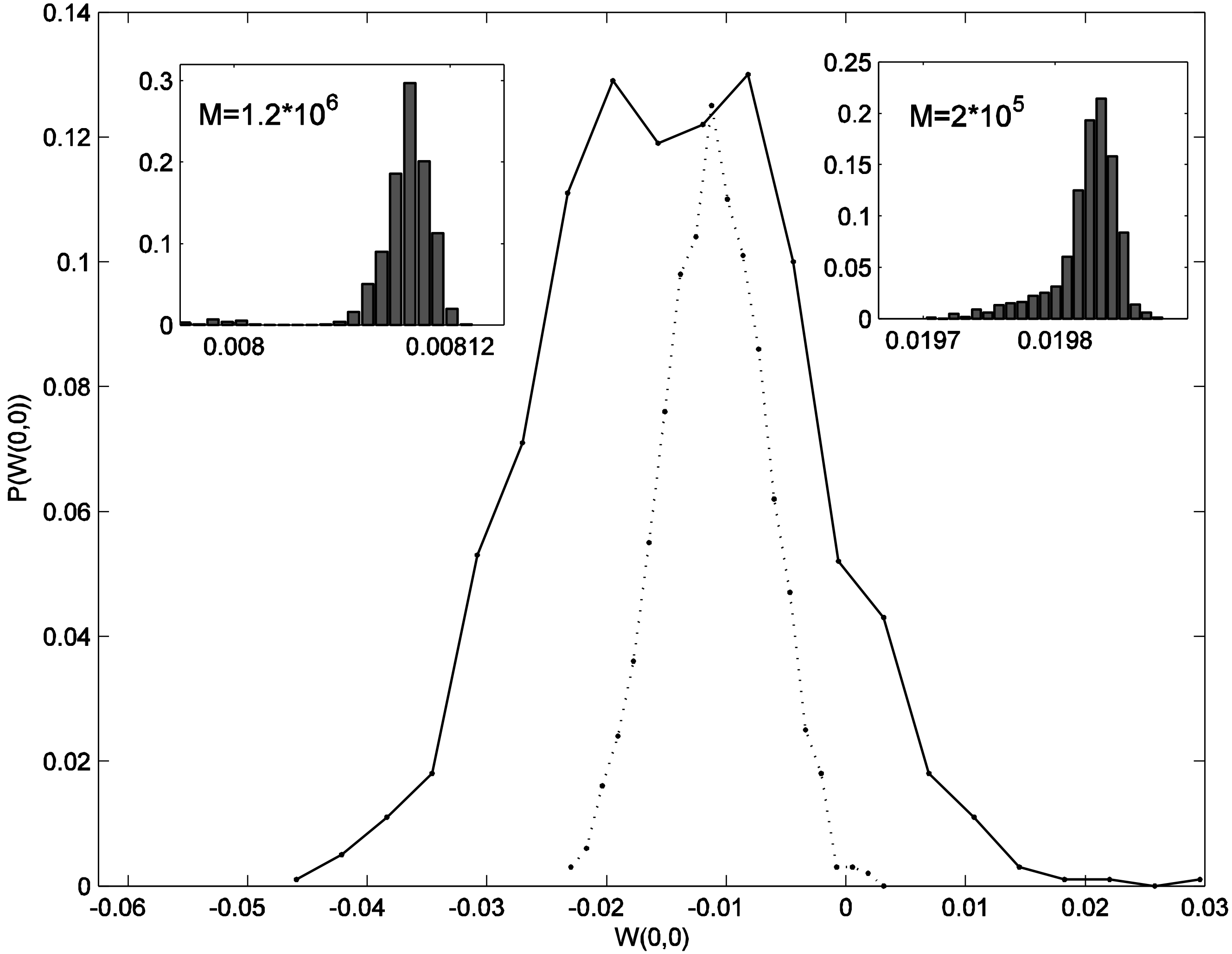
Traditional interpretation-reconstruction of cat-states



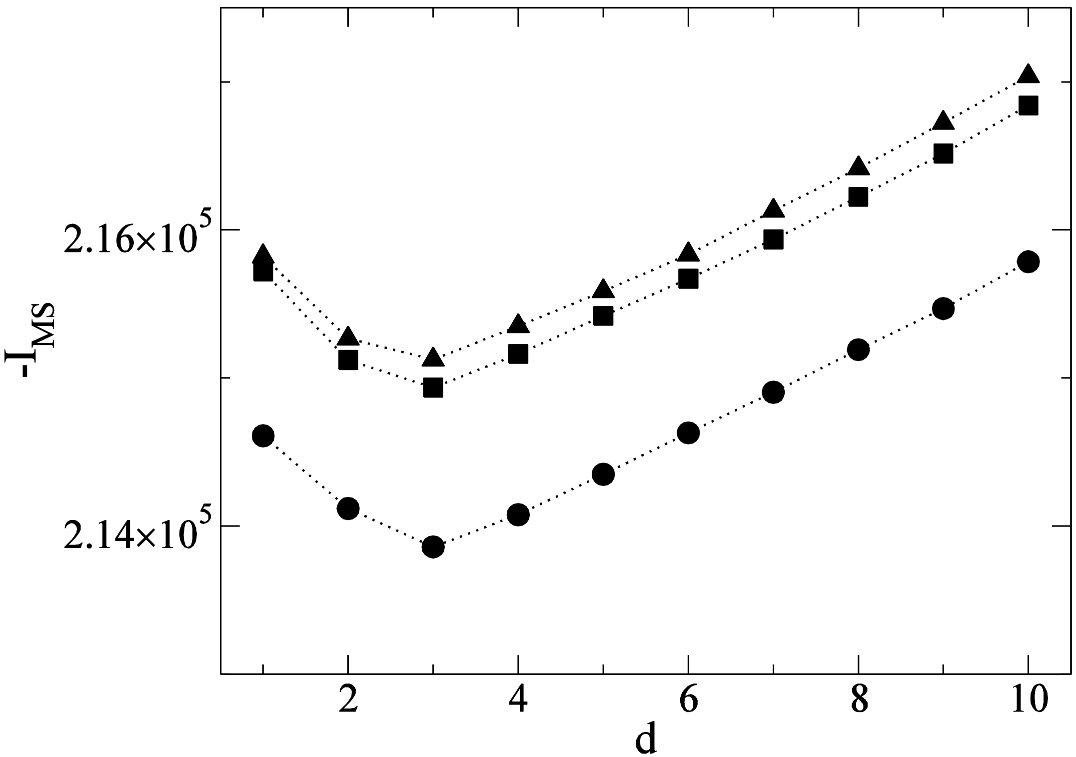
Tomography revisited- effect of dimension



Tomography revisited- detected events



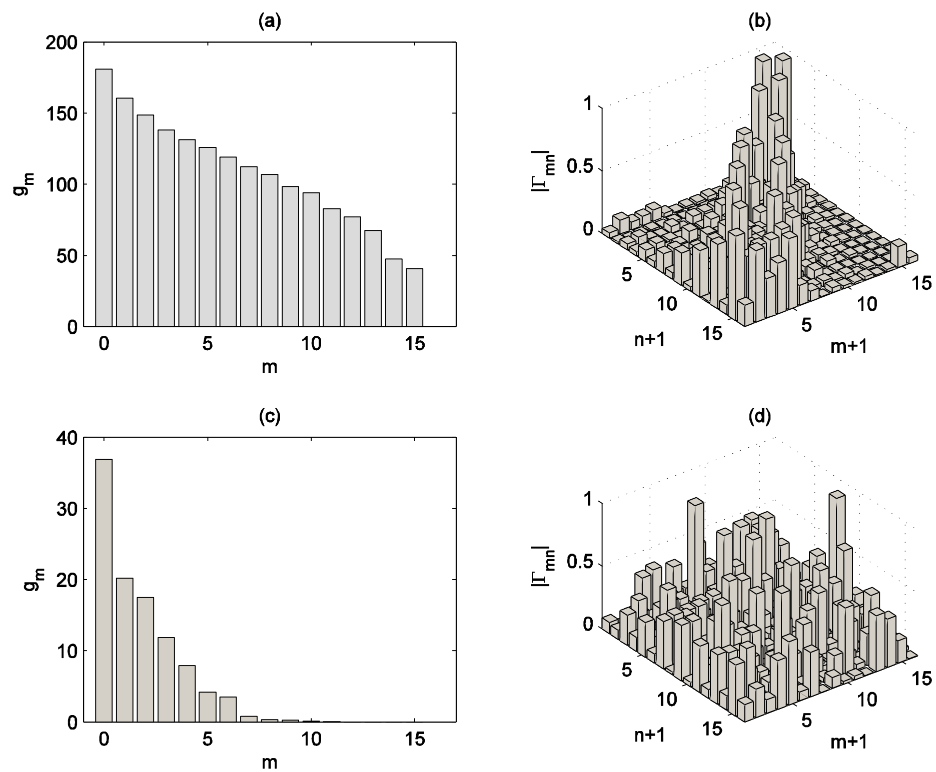
Schwarz info and penalized MaxLik



Simulation of homodyne detection of Schroedinger cat- like state, reconstruction is done in Fock space with cut off dimension d (d^2-1 free parameters). d=3 seems to be enough for fitting.

Řeháček, Mogilevtsev, Hradil, NJP 10,043022 (2008)

data points = # phase cuts ⊗ # sampling points (per each distribution)



256 data points = 16 x 16

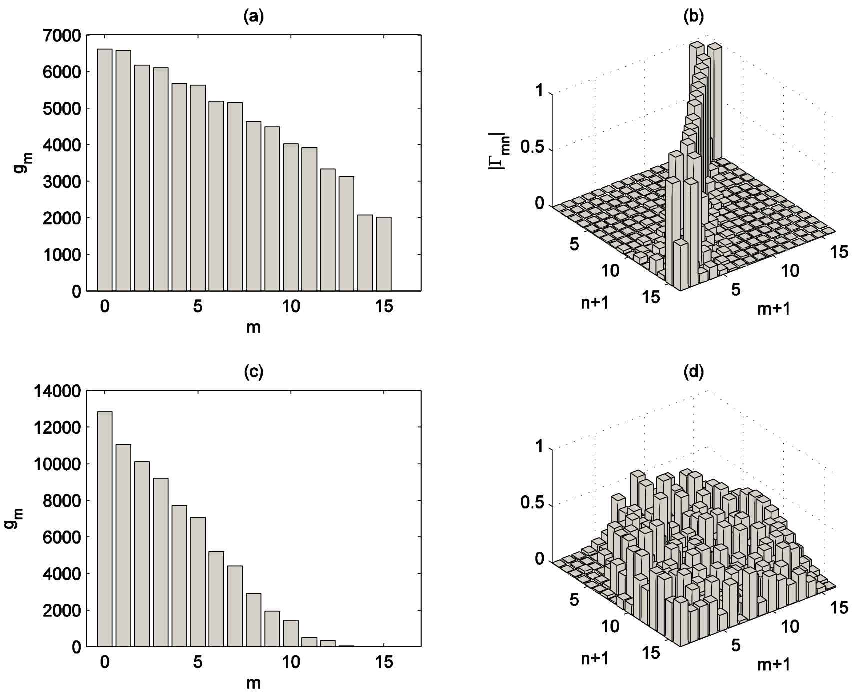
16 data points = 4 x 4

Eigenvalues of G

n-th component of the m-th eigenvector

Data points are distributed in the phase space (0,π) ⊗ (-3,3)

Huge data set



Data set:
12221 = 101 x 121
Phase space
$(0,\pi) \otimes (-3,3)$

Data set:
12221 = 101 x 121
Phase space
$(0,\pi/2) \otimes (0,3)$

Sparse data set

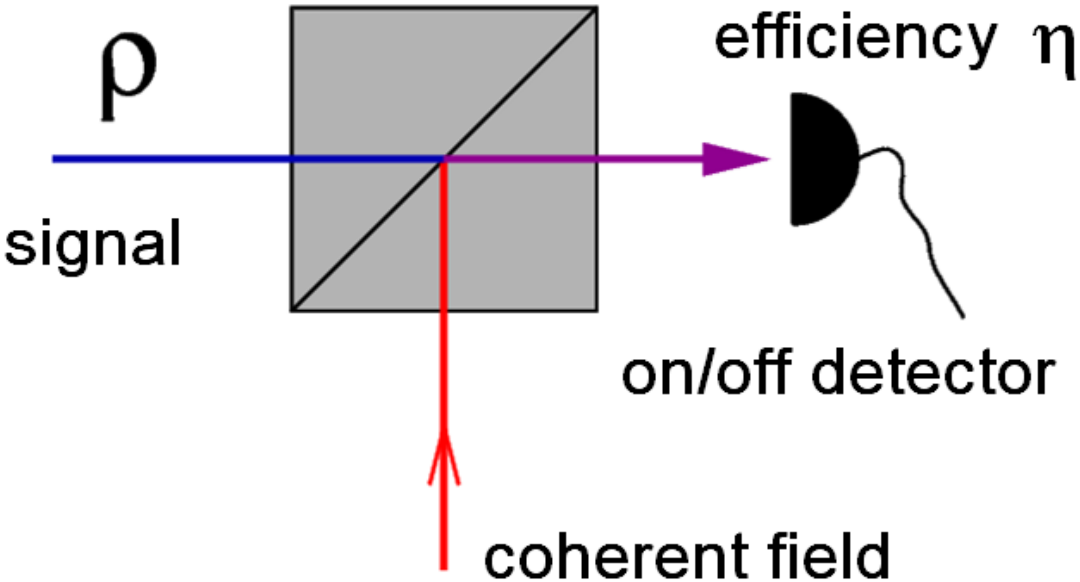
Eigenvalues of G

n-th component of the m-th eigenvector

What we might learn from quantum tomo

- Any piece of measured information may be quantified by means of biased tomography scheme
- Biased scheme = various parts of quantum state are observed (reconstructed) with different accuracy
- Operator ***G*** plays the role of optical transfer function for quantum tomography

Reconstruction based on Inefficient On/Off detection



Recommended approach

Quantum state ρ should be estimated on the subspace spanned by dominant eigenvalues of

$$\mathbf{G} = \Sigma_n \Sigma_\gamma \mathbf{D}(\gamma)|n\rangle \langle n|\mathbf{D}(-\gamma)$$

Log likelihood which should be maximized reads

$$\text{Log L} = \Sigma_{\gamma\eta} N_{\gamma\eta} \log p_\eta(\gamma) - (\Sigma_{\gamma\eta} N_{\gamma\eta}) \log (\Sigma_{\gamma\eta} p_\eta(\gamma))$$

Unabalanced homodyning, Wallentowitz, Vogel, 1996

$P(\alpha, s) = 2/\pi(1-s)\, p_0(\alpha;\eta) \qquad \eta = 2/(1-s)$

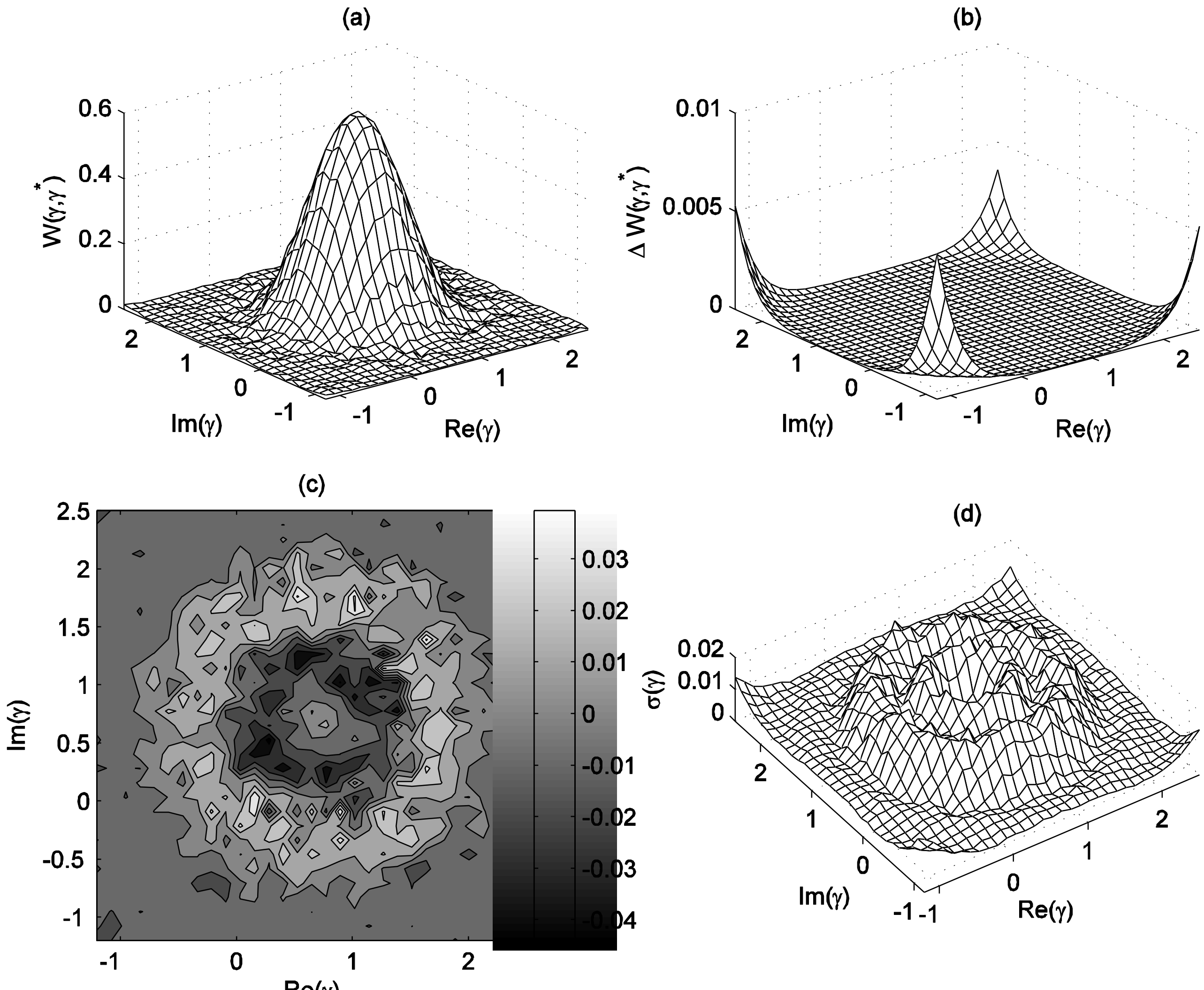
Direct measurement of Wigner function, Banaszek, Wodkiewicz, 1996

$W(\gamma) = (2/\pi) \sum_{n=0} (-1)^n R_n(\gamma)$

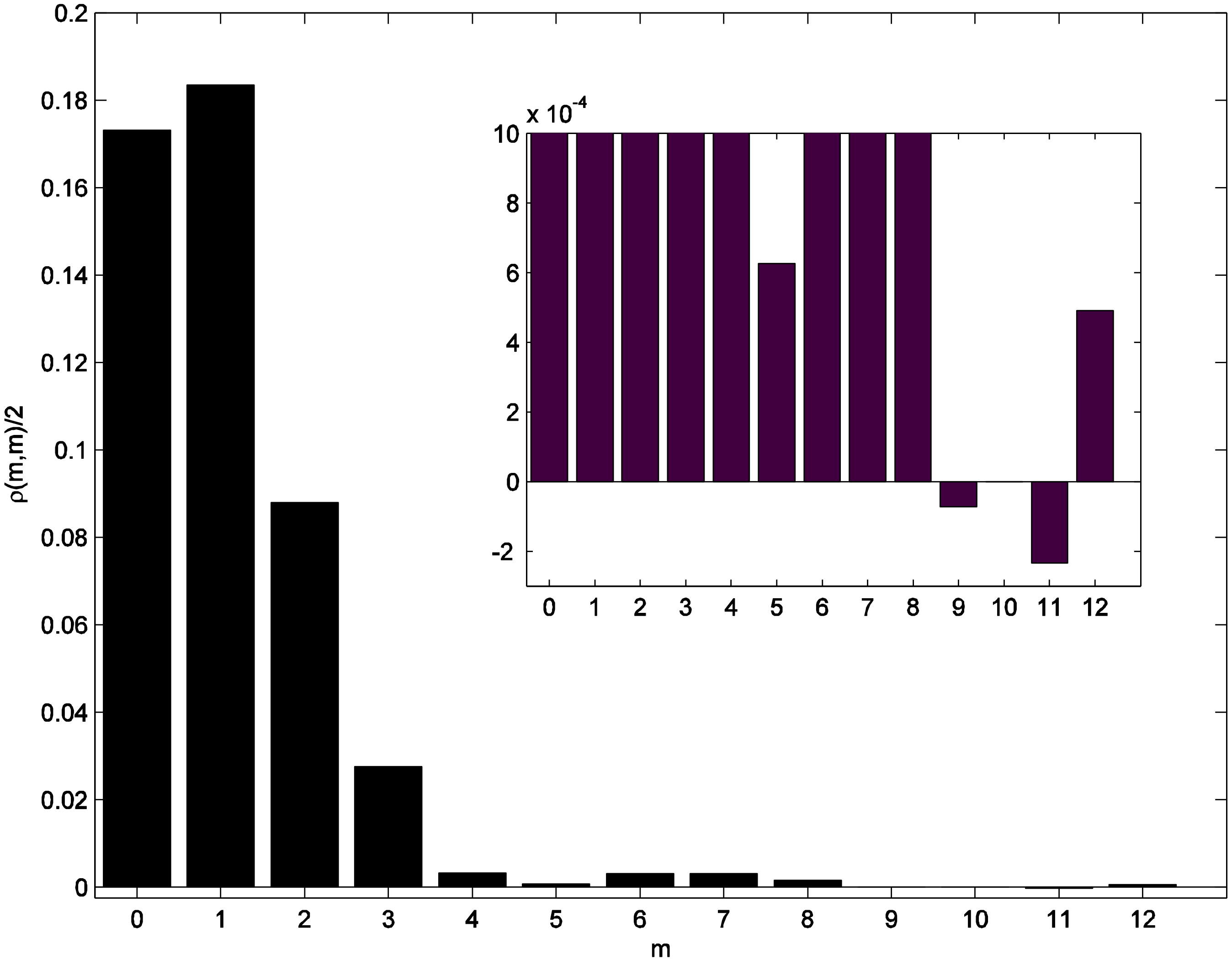
where $R_n(\gamma) = \langle n | \mathbf{D}(-\gamma)\, \rho\, \mathbf{D}(\gamma) | n \rangle$

no count probability: $p_\eta(\gamma) = \sum_{n=0} (1-\eta)^n R_n(\gamma)$

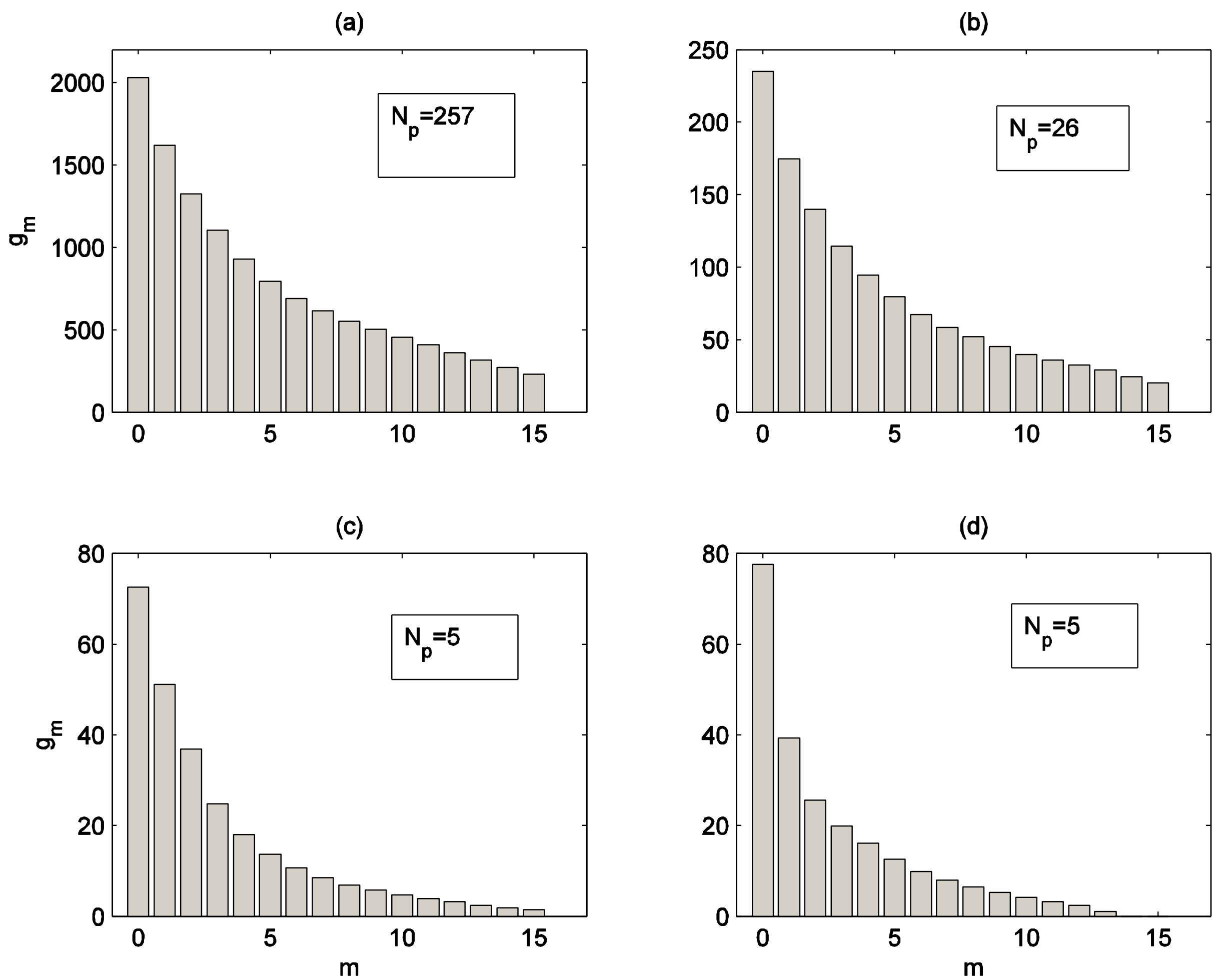
Point-by point reconstruction of coherent state with the amplitude $(1+i)/\sqrt{2}$



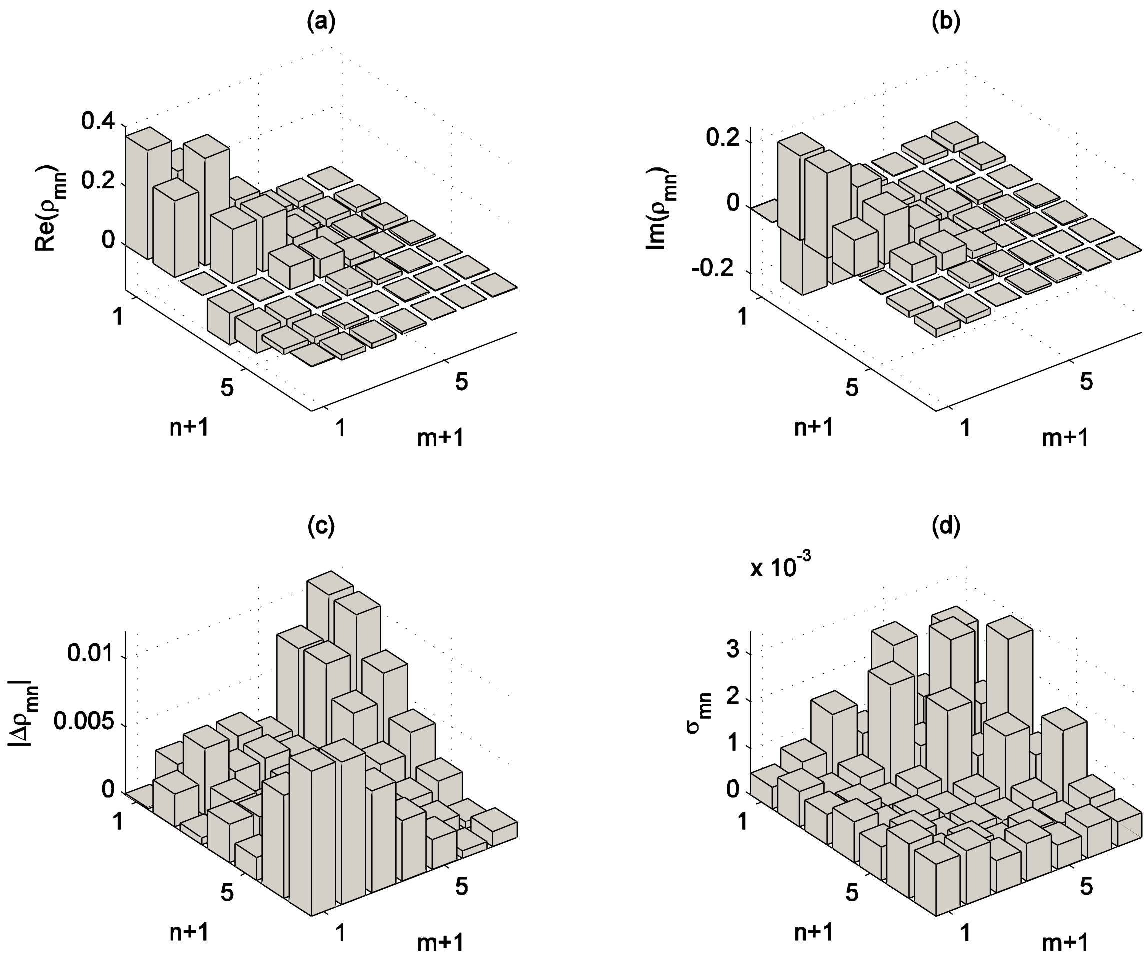
Point-by point reconstruction: diagonal elements of reconstructed density matrix are irregular



All-point reconstruction of the state $(|0\rangle + e^{i\pi/4}\,|2\rangle)/\sqrt{2}$

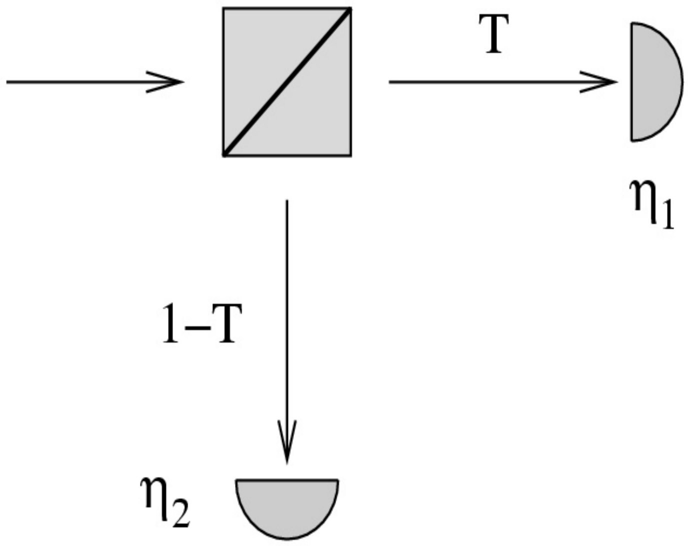


Estimated elements of density matrix



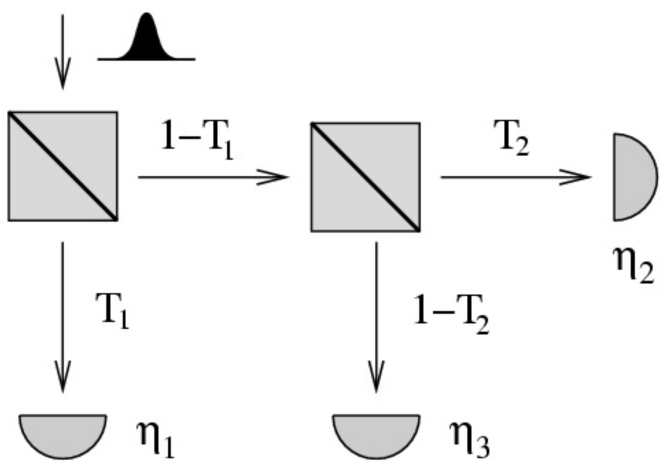
Fiber-loop detector

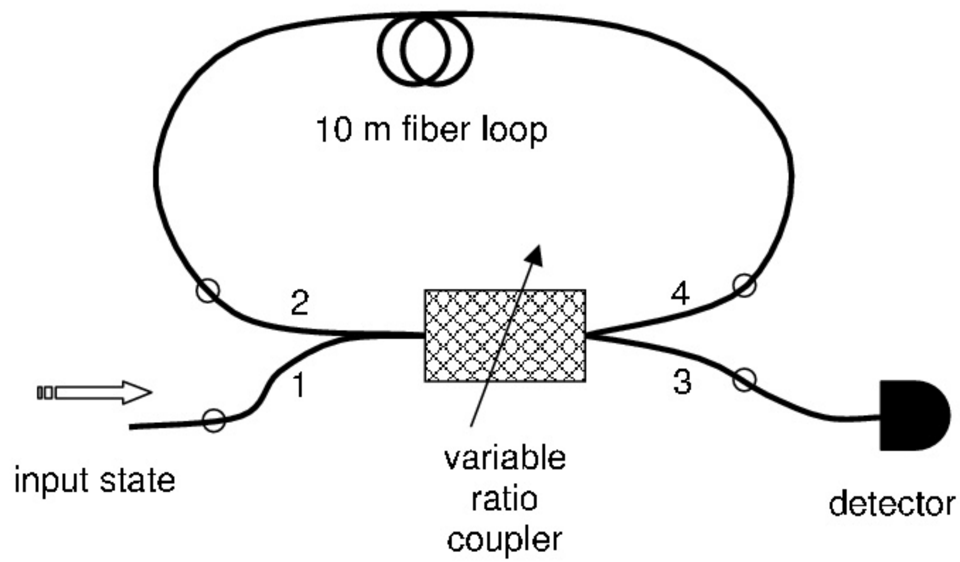
- Commercially available single-photon detectors do not have single-photon resolution
- Cheap (partial) solution: beam splitting
- Coincidences tell us about multi-photon content



J.Řeháček et al.,Multiple-photon resolving fiber-loop detector, Phys. Rev. A (2003) 061801(R)

Fiber loop as a multi-channel photon analyser





Inversion of Bernouli distribution for zero outcome

$$p_0 = \sum_{m=0}^{\infty} (1-\eta)^m \rho_{mm}$$

Example: detection of 2 events = 4 channels

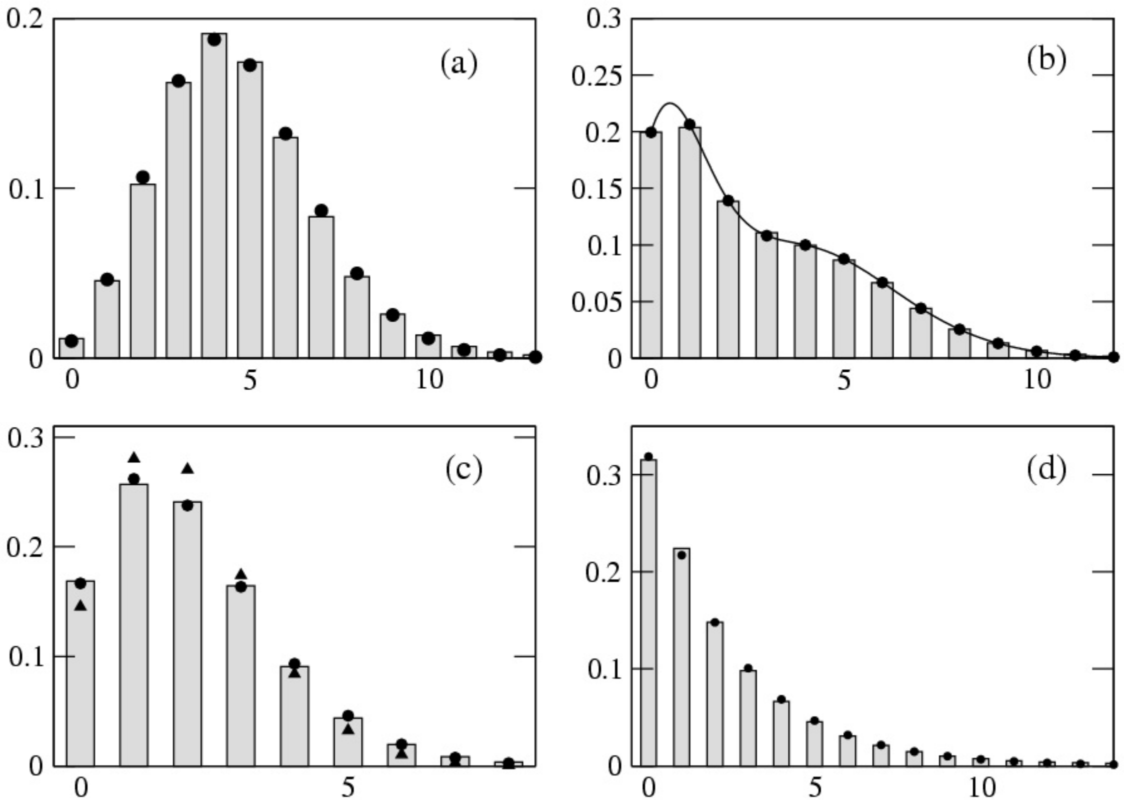
$$p_{00} = \sum_m \rho_m \left[1-\eta_1 T-\eta_2(1-T)\right]^m$$

$$p_{10} = \sum_m \rho_m \left[1-\eta_2(1-T)\right]^m - p_{00}$$

$$p_{01} = \sum_m \rho_m \left[1-\eta_1 T)\right]^m - p_{00}$$

$$p_{11} = 1 - p_{00} - p_{10} - p_{01}$$

Results of MaxLik inversion:



True statistics:

(a) Poissonian

(b) Composite

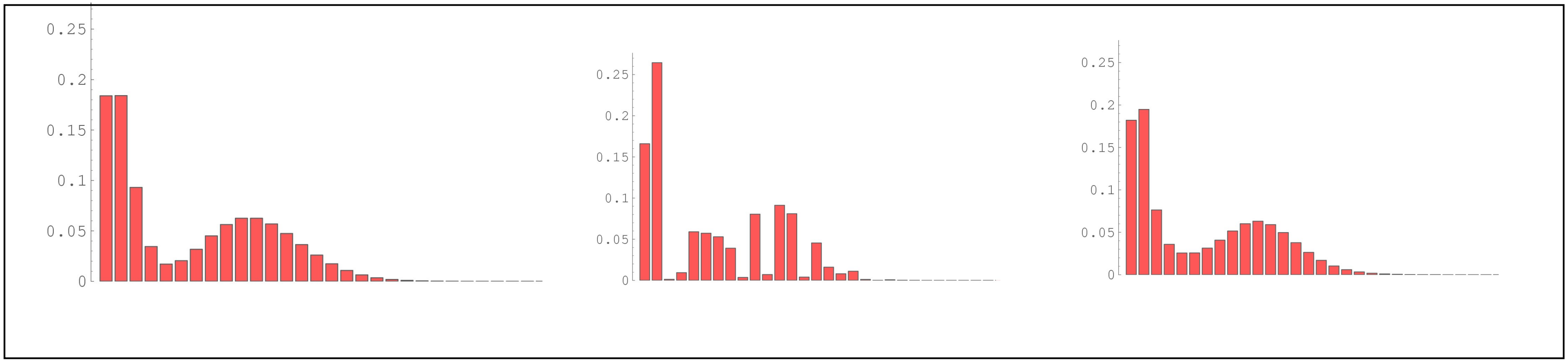
(d) Gamma

(d) Bose-Einstein

True statistics: 50/50 superposition of Poissonian statistics with mean numbers1 and 10

Data: up to 5 counted events (= 32 channels)

Mesh: 100



Original

MaxLik

MaxLik & MaxEnt

Image reconstruction

- **Particles**

Schroedinger equation: $$-i\frac{\partial}{\partial t}\psi = H\psi$$

- **Light**

Paraxial wave equation: $$-i\frac{\partial}{\partial z}\psi = \nabla_T^2\psi$$

pure state ~ complex amplitude

density matrix ~ coherence matrix

- Intensity scan ~ position measurement

$$I(x) \sim \mathrm{Tr}\{\rho |x\rangle\langle x|\}$$

- **Reconstruction of the full density (coherence) matrix:**
  - additional transformation is needed (e.g. free propagation)
  - example: resolution beyond the Rayleigh limit (superresolution)

M. Ježek, Z. Hradil, J. Opt. Soc. Am. 21 (2004) 1407 -1416.

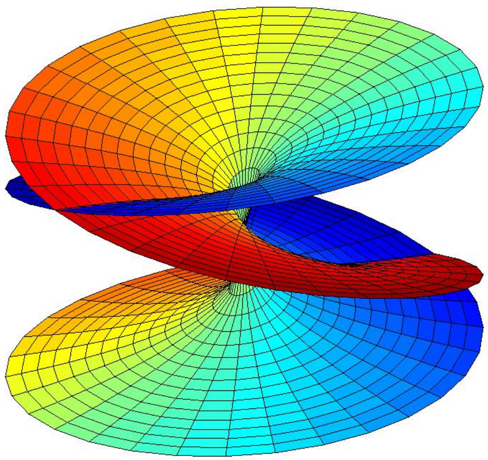
Reconstruction of vortex beams

Laguerre-gaussian modes

$$LG_{lp}(x,y) \propto L_p^{|l|}\left(\frac{2\rho^2}{w_0^2}\right) e^{\frac{-\rho^2}{w_0^2}} e^{il\varphi}$$

Paraxial orbital momentum

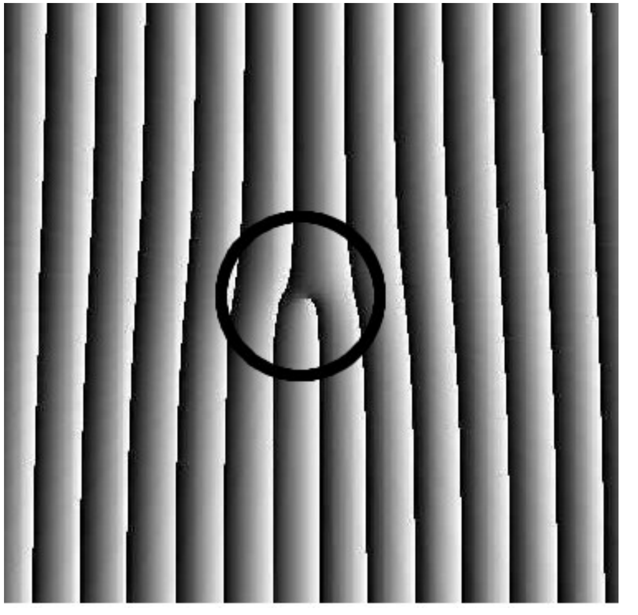
$$l\hbar$$



Mode detection:

- Gaussian mode: single-mode fiber $|0\rangle\langle 0|$
- Higher-order modes: "charged" centered hologram

$$|1\rangle\langle 1| = H_{-1}|0\rangle\langle 0|H_{-1}$$



- Basis in the Hilbert space

$$|m\rangle = H_m(0)|0\rangle$$

$$H_m(\vec{x}) = \sum_i c_i(\vec{x}) H_i(\vec{0}) + \gamma(\vec{x})\Gamma$$

- Off-axis holograms

- Measurement: $\{\Pi_j\},\ \ \Pi_j \sim |j\rangle\langle j|$

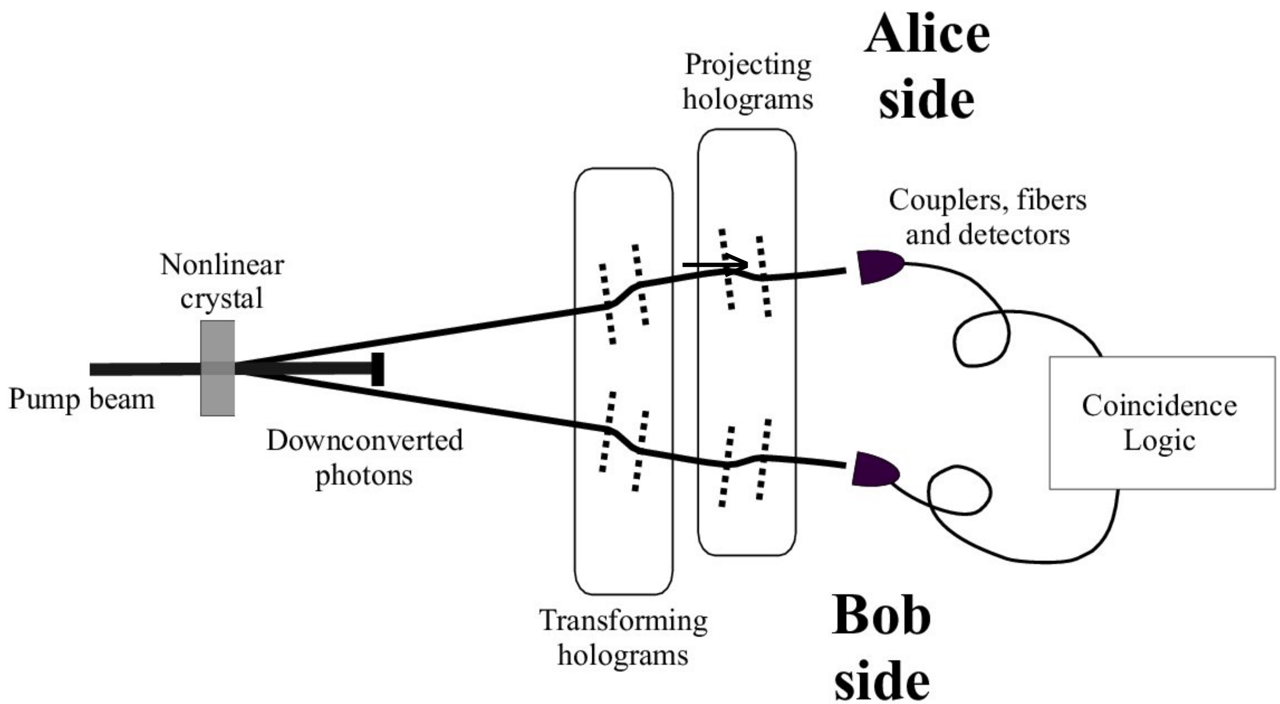
$$|j\rangle = H_1(a^j_{+1}, b^j_{+1})\, H_{-1}(a^j_{-1}, b^j_{-1})|0\rangle$$

- MaxLik reconstruction:

$$L = P(\{f_j\}) = \frac{\prod_j \langle n_j\rangle^{n_j}}{n_j!}\, e^{-\langle n_j\rangle}$$

- **Reconstruction on the 3+8 dim Hilbert space of inner and outer modes from about 2000 projections**

Conditional generation of qutrits



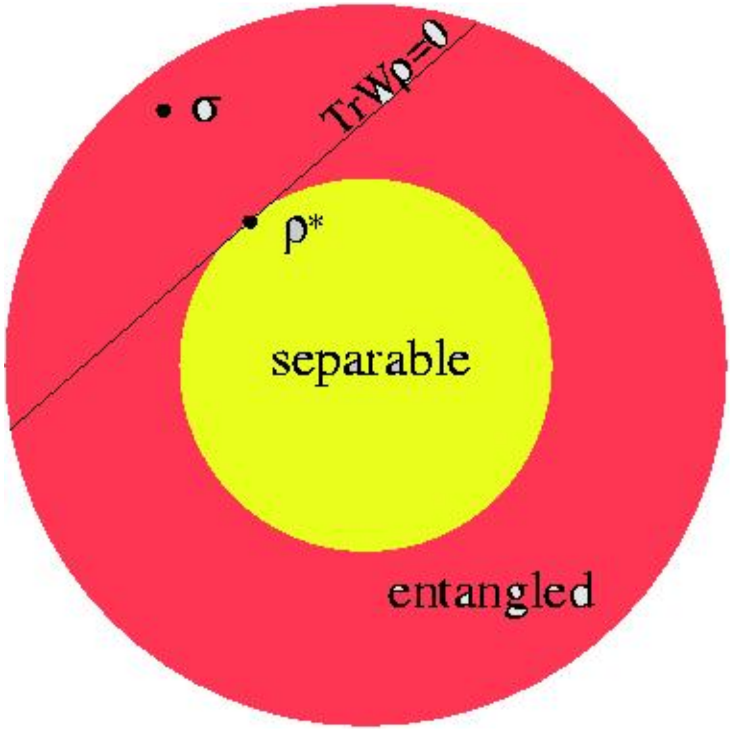
Quantification of entanglement

Quantum entropy

$$S(\sigma \| \rho) = \mathrm{Tr}(\sigma \log \sigma - \sigma \log \rho)$$

Witness of entanglement

$$W(\sigma) = 1 - \int_0^\infty dt\, (\rho + t)^{-1} \sigma (\rho + t)^{-1}$$



J. Řeháček, Z. Hradil: Quantification of entanglement by means of convergent iterations, Phys. Rev. Lett. 90 127904 (2003).

Operational information

Brukner-Zeilinger information: The total lack of information about the system for the set of mutually complementary observables

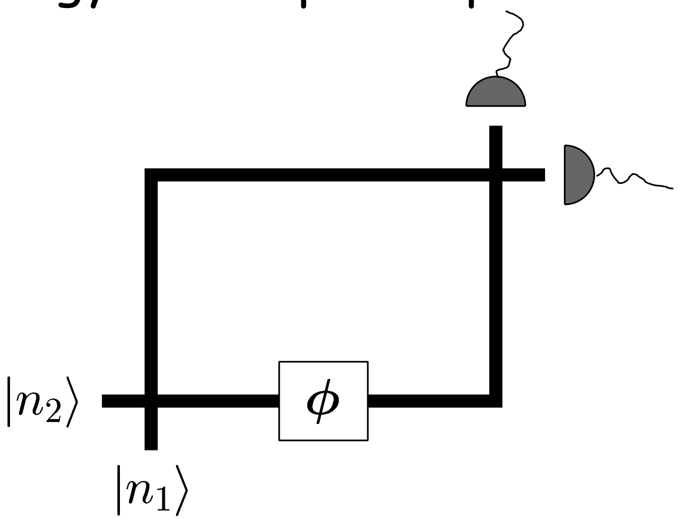
$$E = \sum_{\alpha,j} p_{\alpha,j}(1 - p_{\alpha,j})$$

$$\mathrm{E} = \mathrm{Tr}\left(\rho_{\mathrm{est}} - \rho\right)^2 \geq \mathrm{F}^{-1}$$

J. Řeháček, Z. Hradil: Invariant information and quantum state estimation, Phys. Rev. Lett. 88, 130401 (2002).

Fisher Info in Quantum Interferometry

- Considering the model of Holland, Burnett with the interferometer triggered by two n-Fock states, investigate the resolution limit and discuss the strategy when optimal phase resolution can be achieved.



Detected signal difference: q
($n_1 = n_2 = r$)

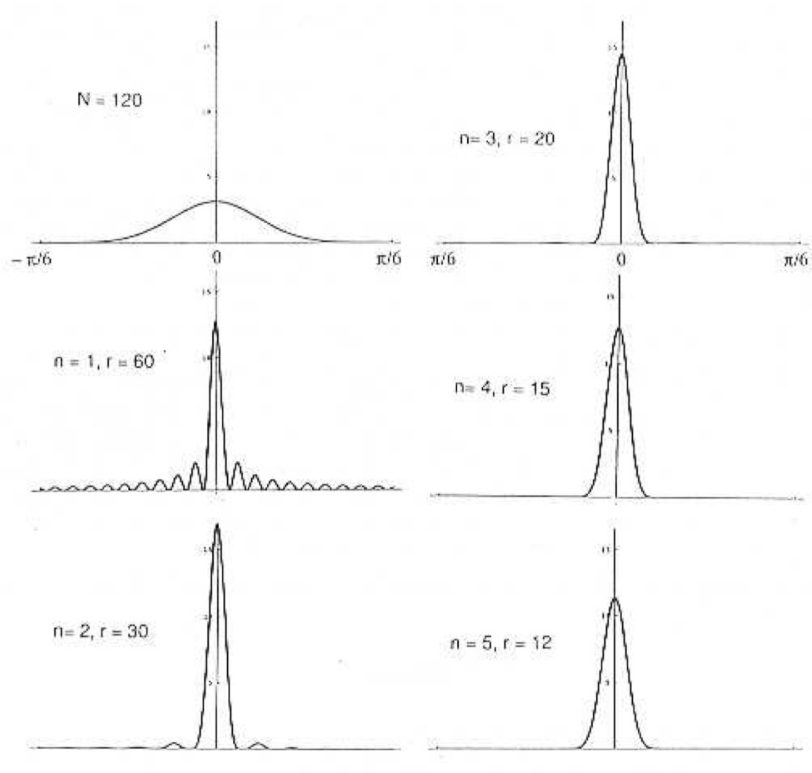
Detected statistics (associated Legendre polynomials):

$$P(2q|\theta) = \frac{(r-q)!}{(r+q)!}[P_r^q(\cos\theta)]^2$$

Approximation for q=0 $P(0|\theta) \approx [J_0(r\theta)]^2$

Phase estimation after "single" detection is not a good idea since

$$(\Delta\phi)^2 \approx (\frac{1}{r})^2 \frac{\int_0^{r\pi/2} dx x^2 J_0^2(x)}{\int_0^{r\pi/2} dx J_0^2(x)} \propto \frac{1}{\log r}$$

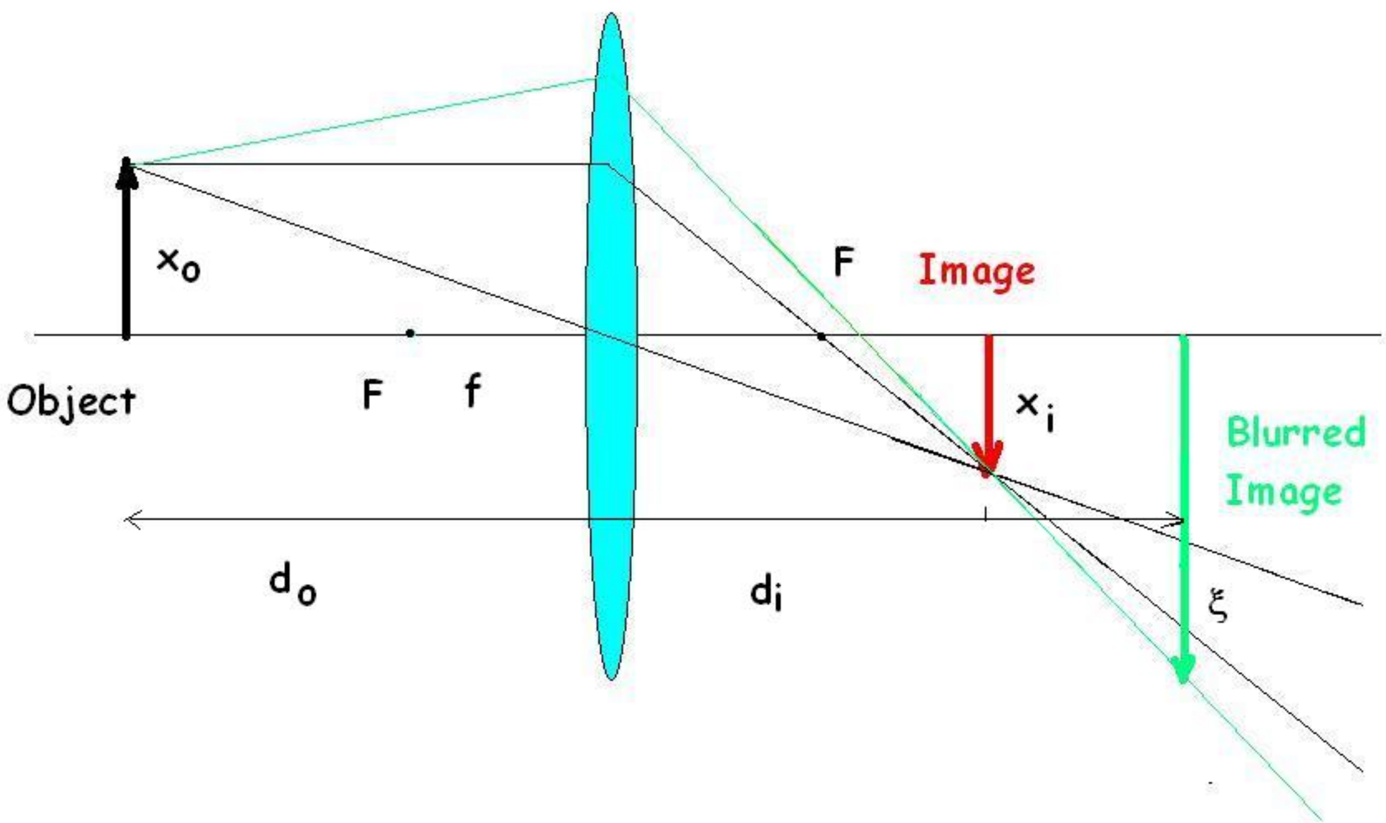


Optimum: repeat n = 4

$$(\Delta\phi)^2 \approx \frac{4n}{N^2}$$

Advanced schemes in image processing: reconstruction of mutual coherence

Optical Imaging: Lens equation in geometrical optics



Lens equation in geometrical optics:

$$1/d_o + 1/d_i = 1/f$$

For sharp image: $x_i = M\, x_o$,
magnification $M = d_i/d_o$

For the blurred image: $\xi = \alpha\, x_o + \beta\, p_o$
x_o … position of the ray
$p_o = 2\pi/\lambda\ \theta$ … direction of the ray

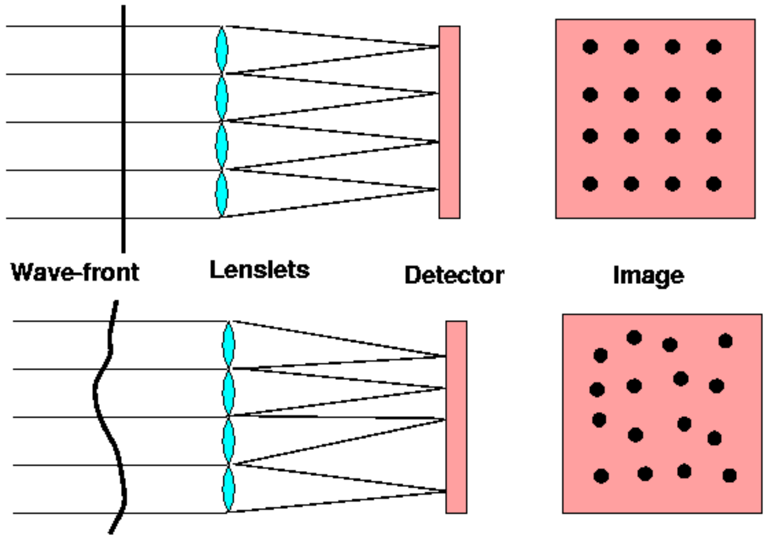
Meaning in quantum mechanics:

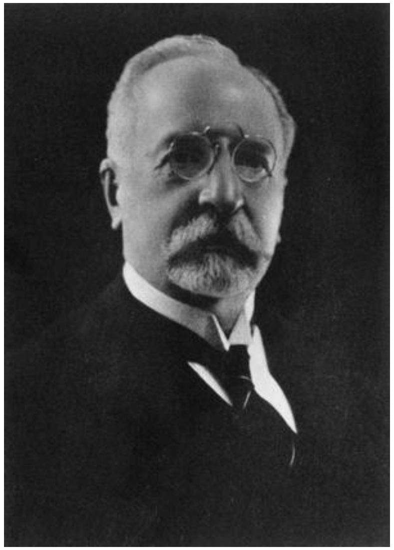
Rotated quadrature operator for $[x_o, p_o] = i\hbar$

See the analogy with the free evolution

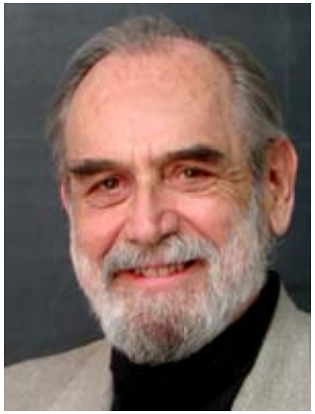
$$x(t) = x(0) + p(0)\ t/m$$

Scanning of the optical field: Hartmann-Shack sensor



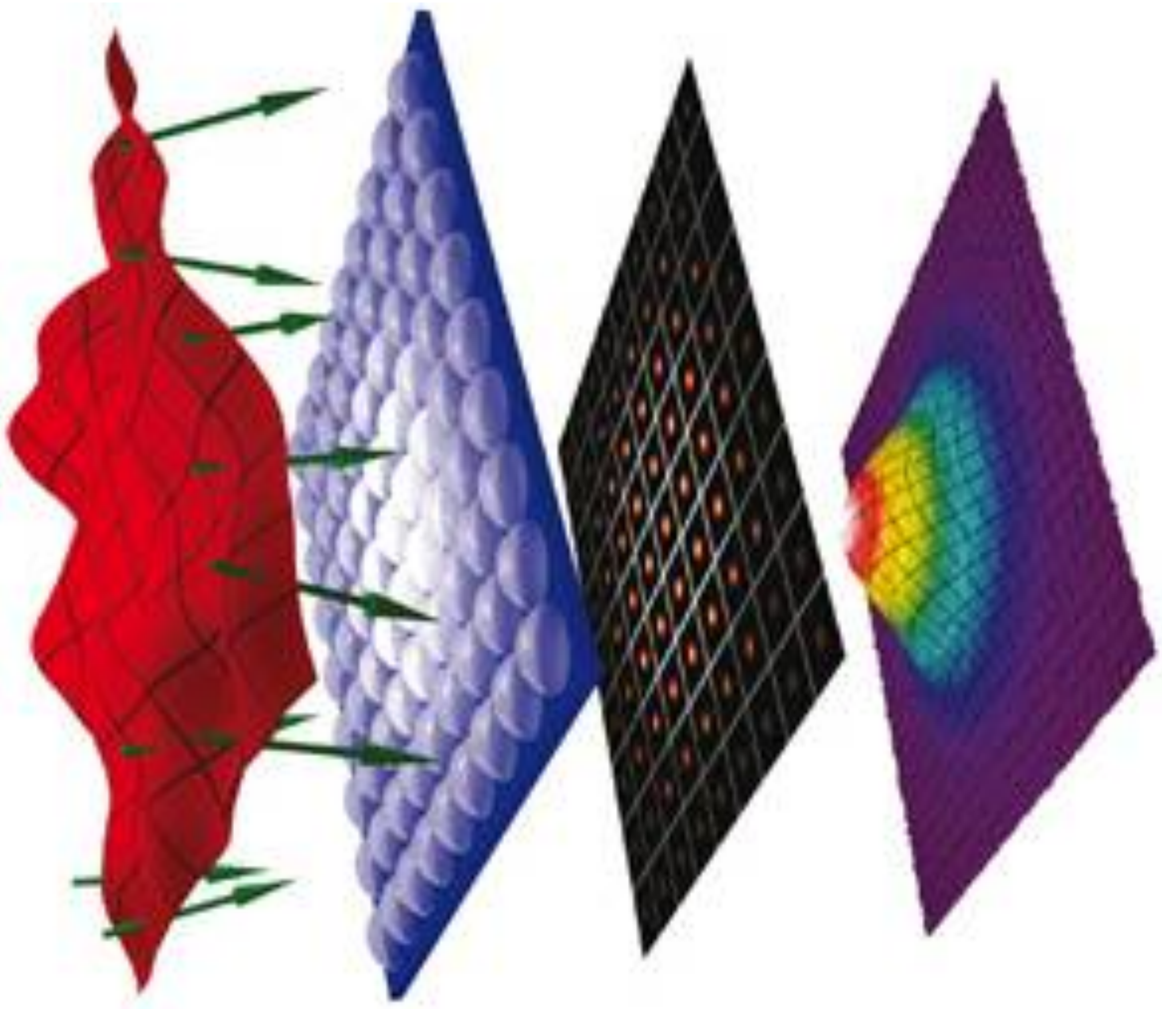


Johannes Hartmann
(1865-1936)



Roland Shack
(1970's)

Scheme of the wave-front reconstruction



Wave theory for HS sensor

- Detected amplitude:

$$\varphi_{det}(\xi)= \int dx'\, dq'\, \varphi(x')h(x'-q')A_i(q')\exp(i\, k\xi q'/f)$$

- Detected signal:

$$S_i(\xi) = \langle |\varphi_{det(}(\xi)|^2 \rangle_{average}$$

$$= \int dx'\, dx'' \int dq'\, dq''\, Q(x',x'')\, h(x'-q')\, a(q',\xi)\, h^*(x''-q'')\, a^*(q'',\xi)$$

where Q… function of mutual coherence

$$a_i(q',\xi) = A_i(q')\exp(i\, k\xi q'/f)$$

- Quantum formulation in x-representation

$$S_i(\xi) = \langle a_{i\xi} | U^\dagger\, Q\, U | a_{i\xi} \rangle$$

$Q(x',x'')= \langle x'|Q|x''\rangle$, $h(x'-q')= \langle q'|U|x'\rangle$, $\langle x'|a_{i\xi}\rangle = a_i(q',\xi)$

HS sensor: Quantum Consequences

- Smooth Gaussian approximation of aperture function:

$$A_i(q') \approx \exp[-(q'-x_i)^2/4(\Delta x)^2]$$

- Detection= Projection into the minimum uncertainty states

$$a_{i,\xi} = \exp[-(q'-x_i)^2/4(\Delta x)^2 + i\,k\xi q'/f\,]$$

- Heisenberg uncertainty relations

$$\Delta x\,\Delta p \geq \hbar/2$$

- Generalized measurement of non-commuting variables x and p, (Arthurs, Kelly 1964)

$$\Delta X\,\Delta P \geq \hbar$$

See the excellent paper: S. Stenholm, *Simultaneous measurement of conjugate variables, Annals of Physics 218, 233-254 (1992).*

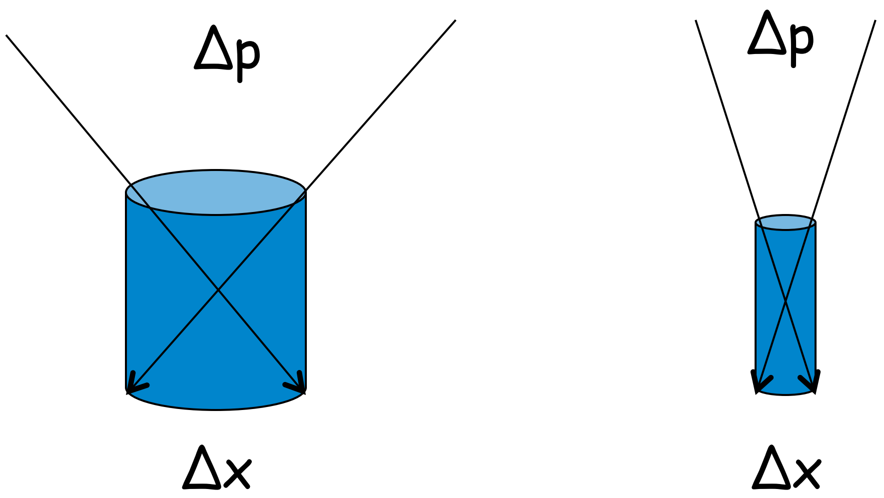
Further Quantum Consequences

- POVM corresponds to detection of annihilation operator

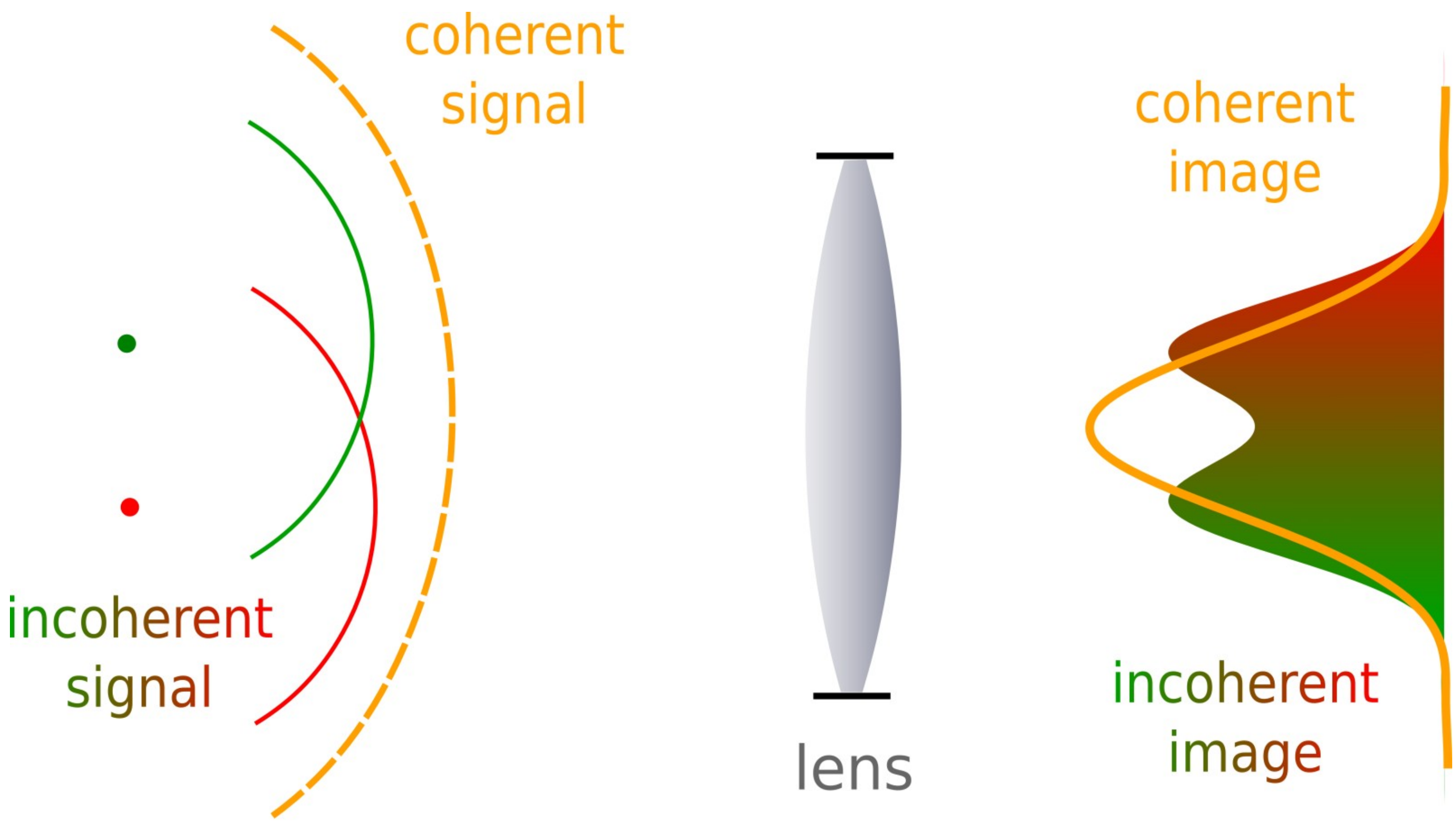
$$a = x + ip$$

$$1/\pi \int d\alpha^2 \; |\alpha\rangle\langle\alpha| = 1$$

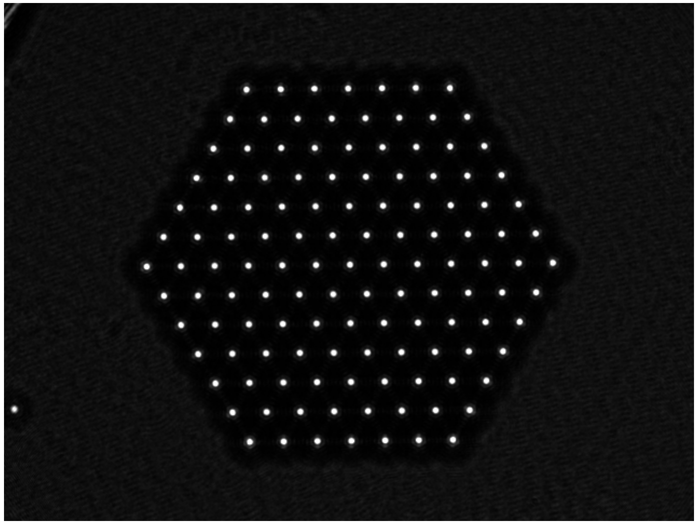
- Q-distribution (Husimi)

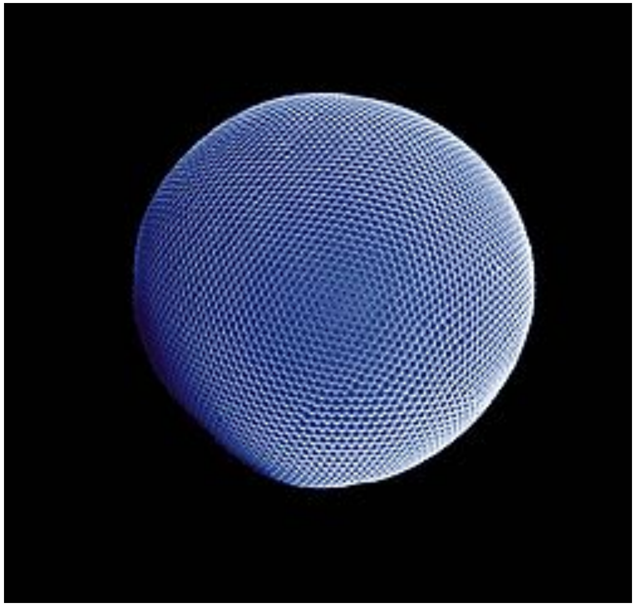


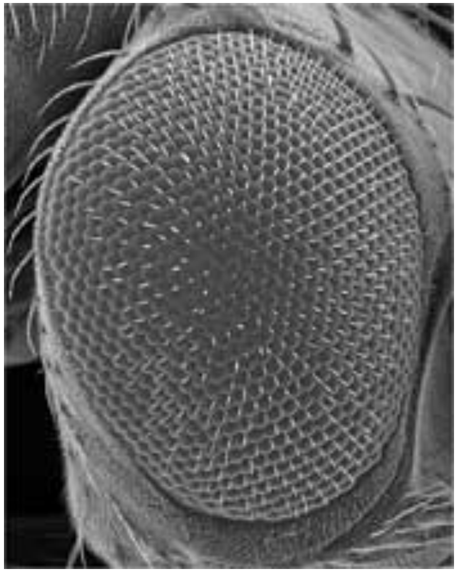
Detection of partially coherent signal



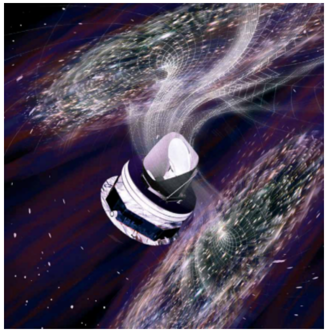
Hartmann-Shack sensor of the wavefront?

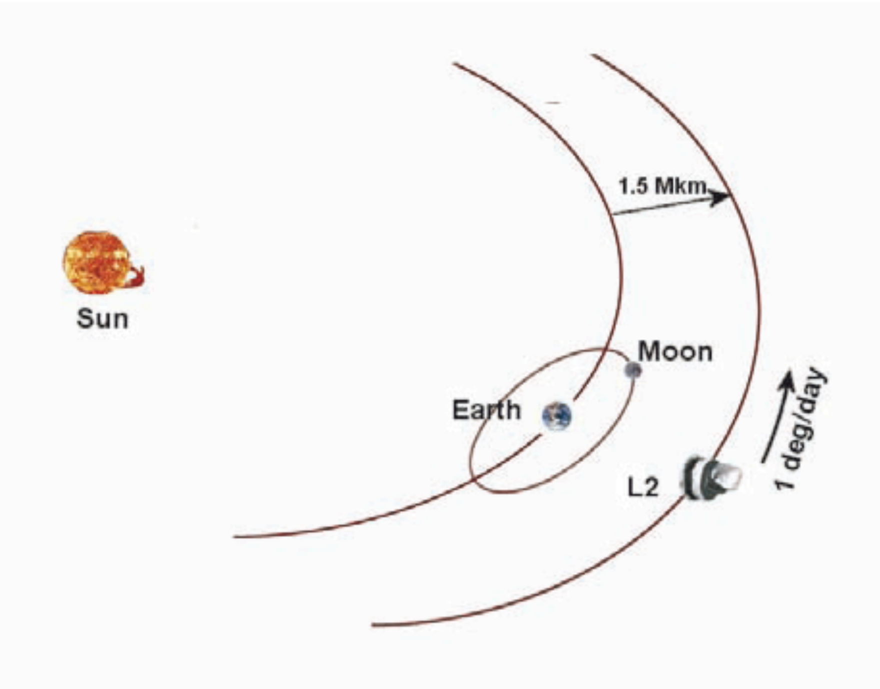


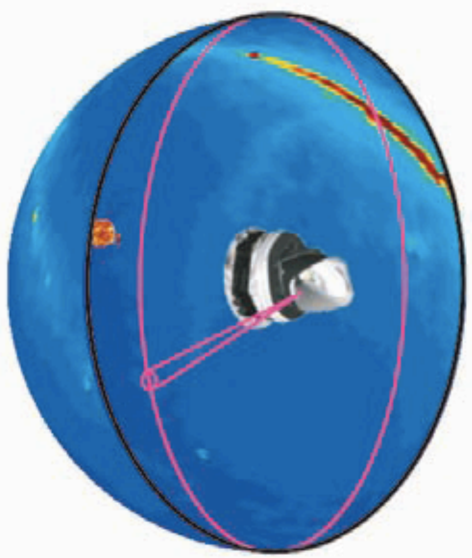




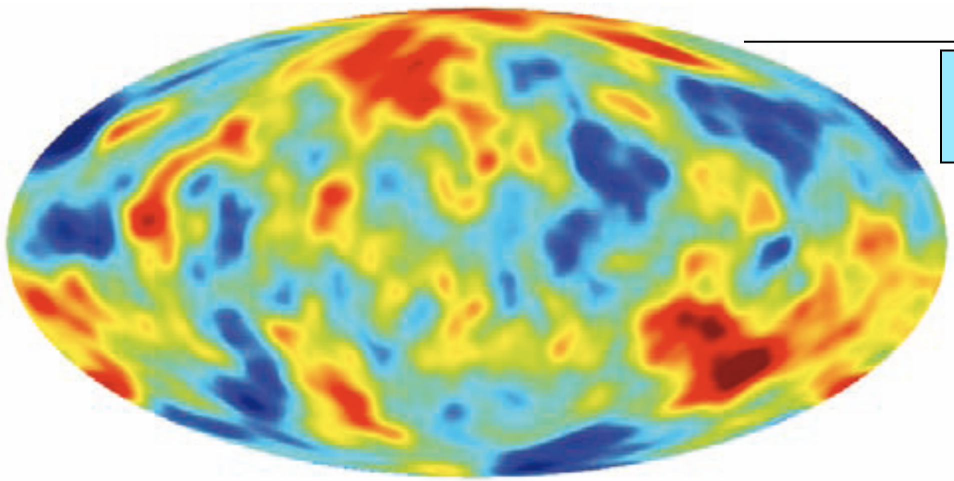
Planck mission of ESA: scanning of cosmic background radiation



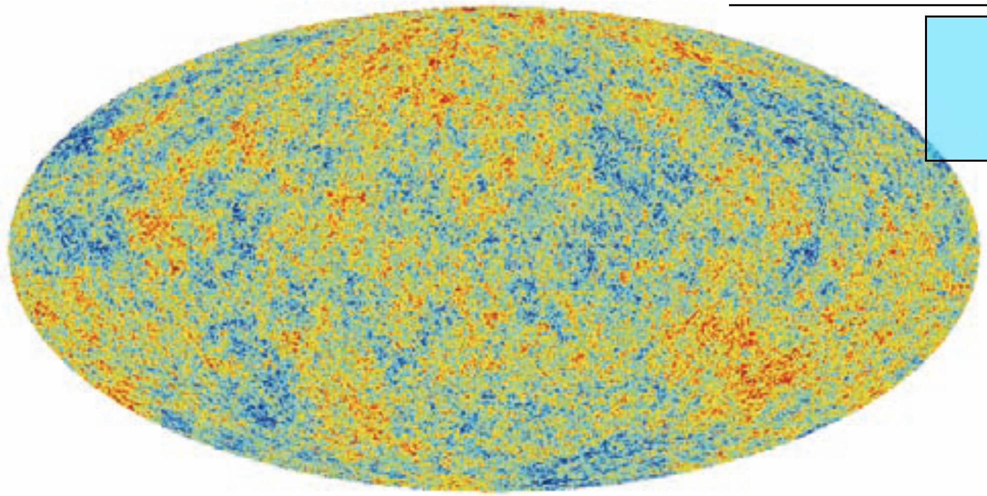




Temperature anisotropies



COBE-DMR resolution



Planck resolution

Statistics matters in the real life!

Think of EUROSTAT and Europe