From quantum measurement and estimation towards quantum tomography and back I: Quantum Tomography

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 and others

Full Program

Lecture 1 : General Concepts of MaxLik estimation

- Introduction: inverse problems, quantum measurement and estimation
- MaxLik estimation and implementation in QM
- Fisher information for quantification of noises and diagnostics
- Information principles and MLME estimation
- Full scheme for MaxLik tomography

Lecture 2: Exercise on MaxLik problems

- Radon and Inverse Radon transformation
- Statistical interpretation of measurement
- Fisher information and diffraction on the slit
- MaxLik solution
- Normalization of the likelihood
- Resource analysis for tomography of 5 qbits
- Fisher info in quantum interferometry

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 $k= 1,2,\ldots N$

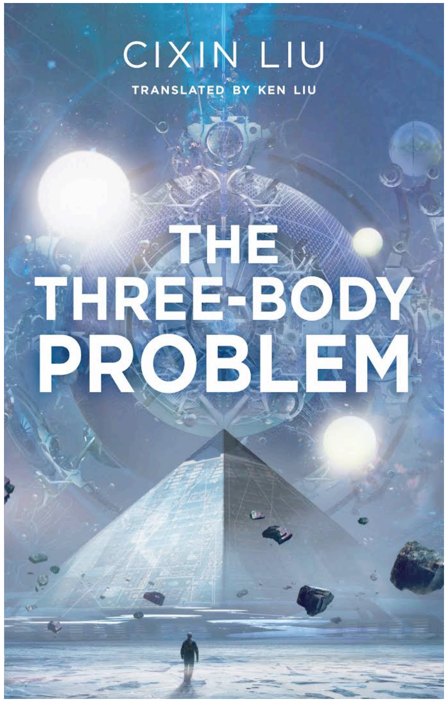
Over-determined problems $M > N$
Well defined problems $M = N$
Under-determined problems $M < N$

If "direct problem" is not solvable?

Probably intractable … follow the sci-fi

知子 *

*according to our secretary Petra Cabišová 韓思夢, she has MA in Chinese ☺



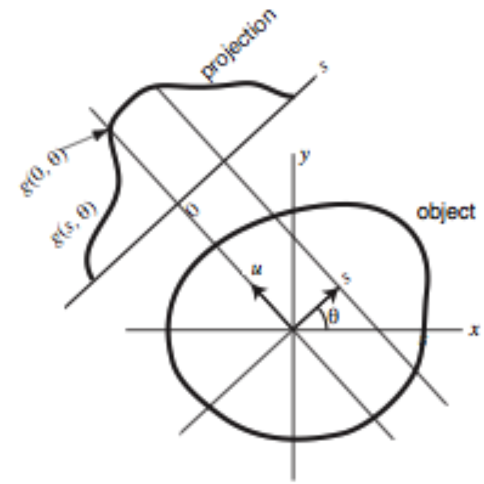
Tomography and Inverse Radon Transformation

Radon transformation

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

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

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

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

Elements of quantum theory

Probability in Quantum Mechanics:

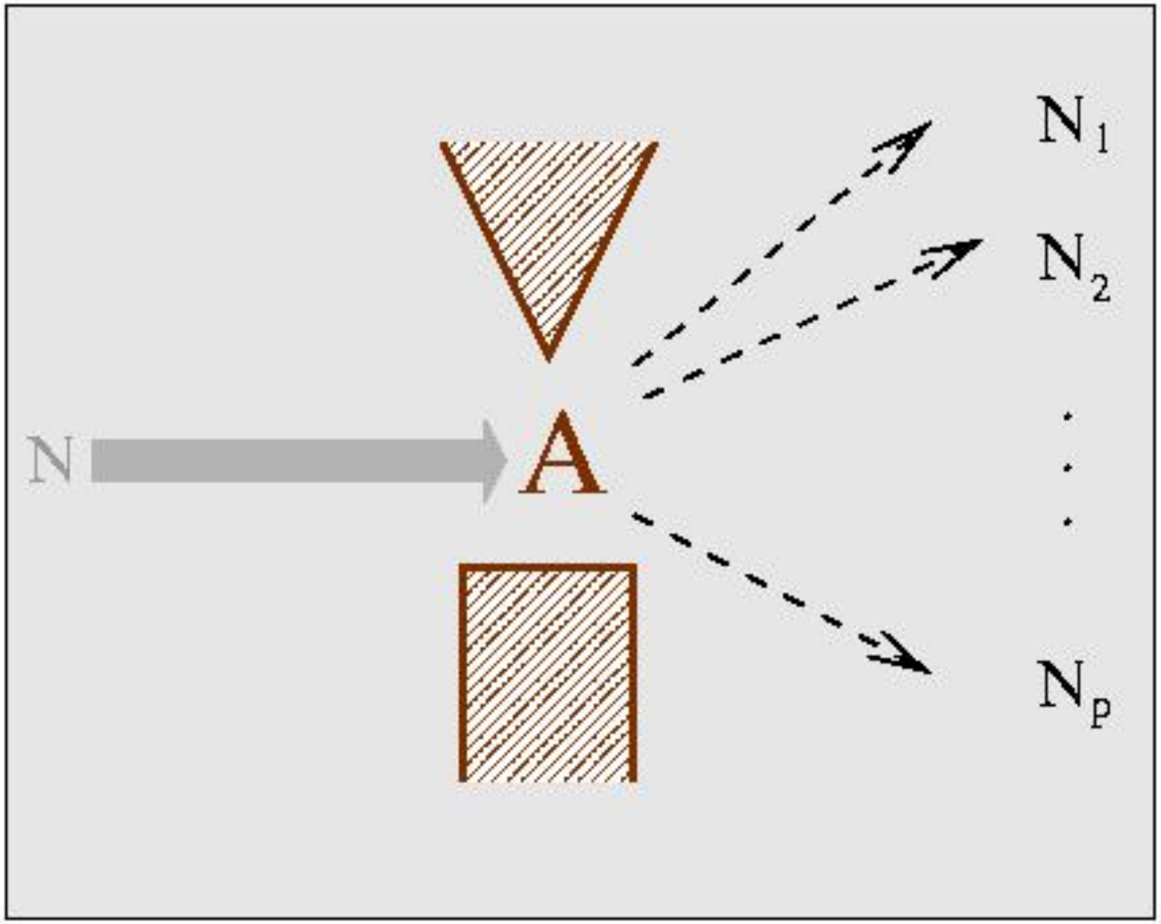
$$p_j = Tr(\rho A_j)$$

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

Relation of completeness $\Sigma_j A_j = 1$

Signal: $\quad$ 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 $\quad p(\theta)$

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

Estimation Theory in Words...

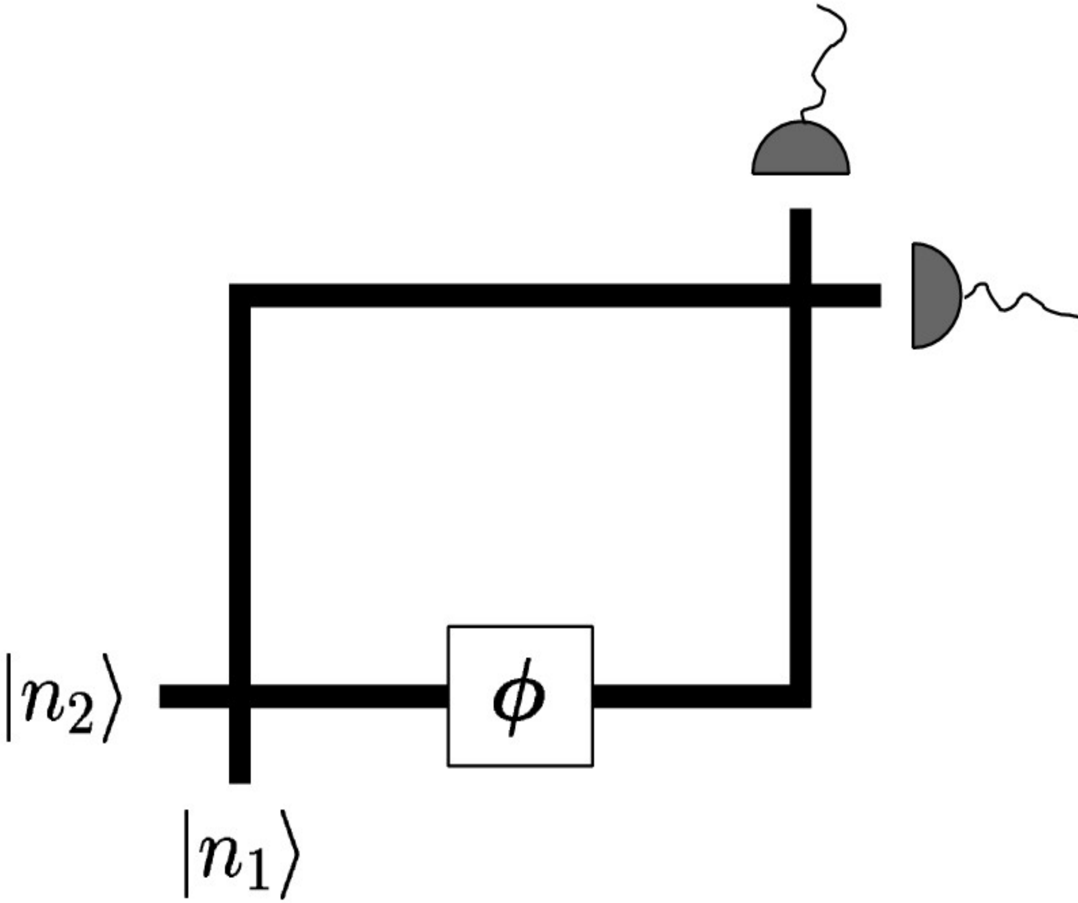
- The quality of estimation should be assessed by **the cost function $C(\theta, \theta_{true})$**
  - -least square fit $C(\theta, \theta_{true}) = (\theta - \theta_{true})^2$
  - -maximum likelihood fit $C(\theta, \theta_{true}) = -\delta(\theta - \theta_{true})$
- **The risk function**

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

- 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

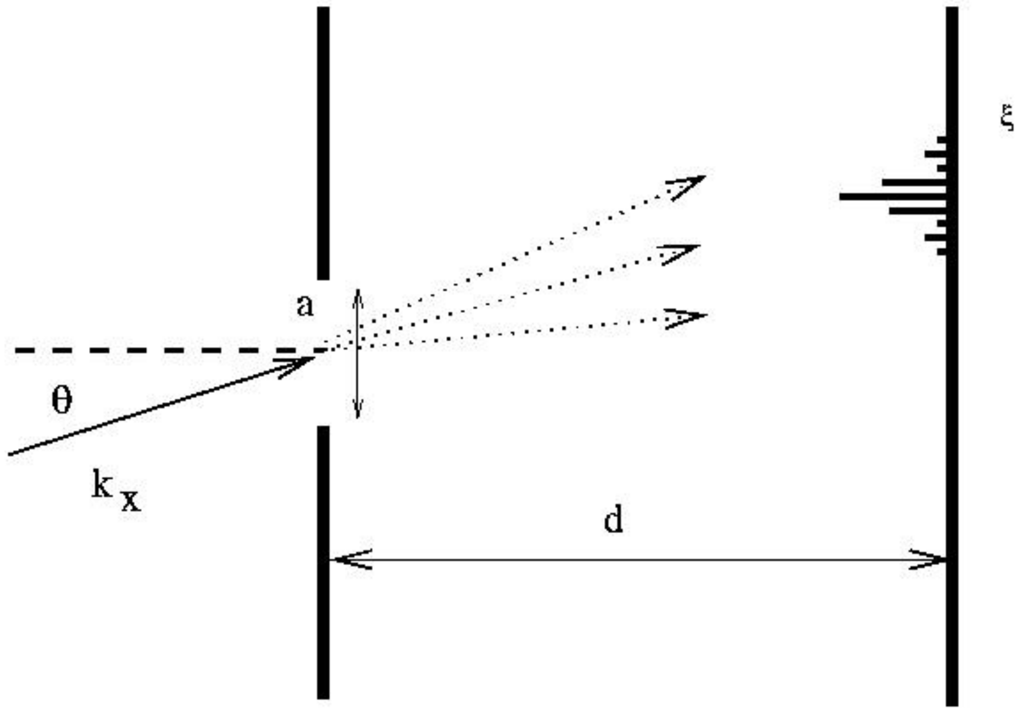
= Quantum Theory + Estimation Theory

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 …

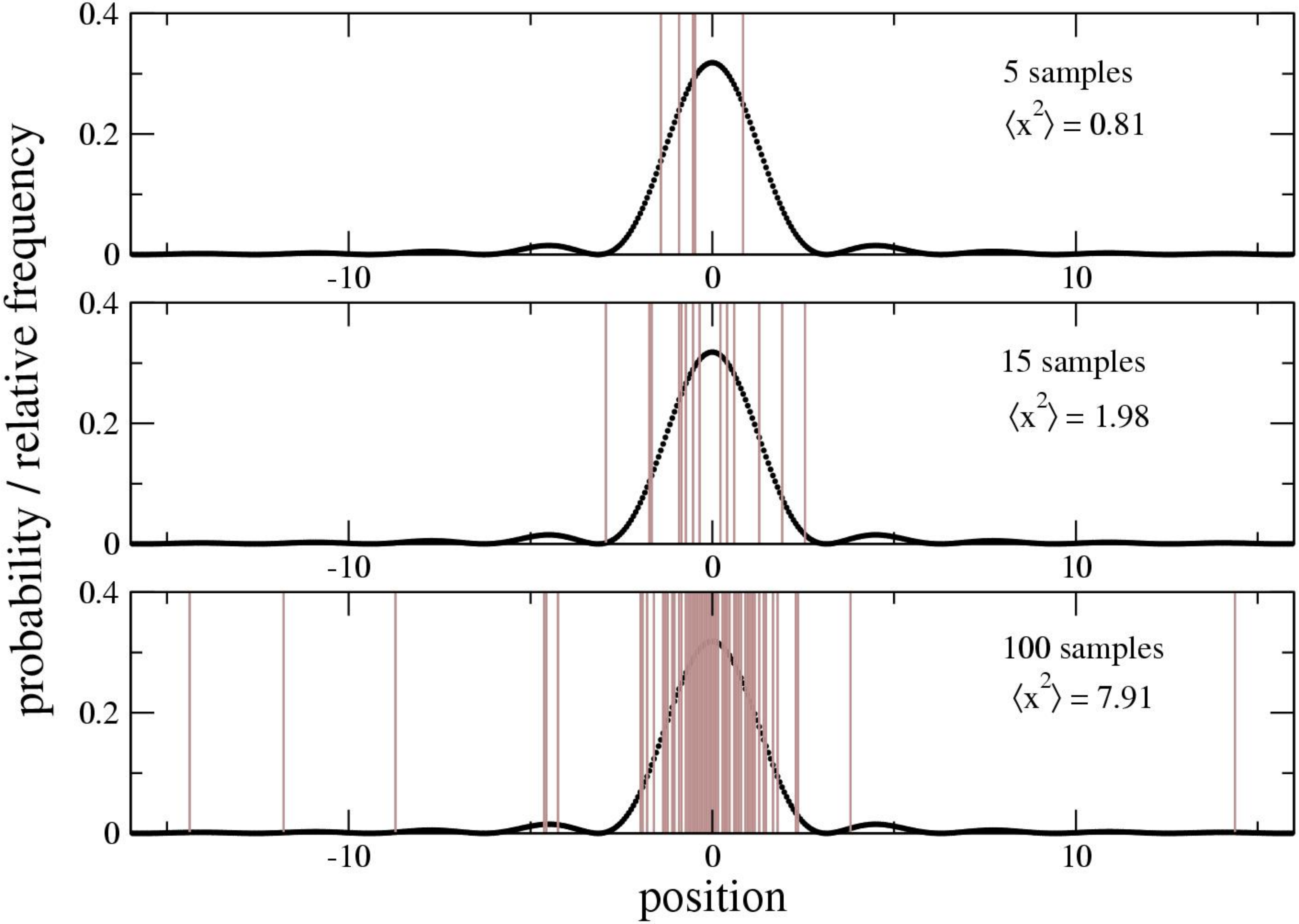
- **The uncertainty is given by wave theory**

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

- **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
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

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) $\quad N_\mu = N\, p(\mu|v_{true})$
- 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) $\quad N\, \Sigma_\mu\, p(\mu|v_{true}) \log p(\mu|\, v_{true})$

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

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

Diffraction continues 3 …

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

2^{nd} term is zero since $\sum_\mu \; p(\mu|v_{true}) \;\; \partial_v \log p(\mu|v)|_{true} \; =$

$\sum_\mu \partial_v \, p(\mu|v)|_{true} \;\; (v-v_{true}) = (v-v_{true}) \; \partial_v \, 1 = 0$

3^{rd} 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 $F = 4\pi^{-1} \int d\mu\, [\partial_\mu \mathrm{sinc}\, \mu]^2 = 4/3$

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

$$\Sigma_\mu\, \partial v_{true}\, p(\mu|v_{true})\, (v - v_{true}) = 1$$ /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 ☺)

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 $\quad p(x) = q(x)^2$

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

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

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

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

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

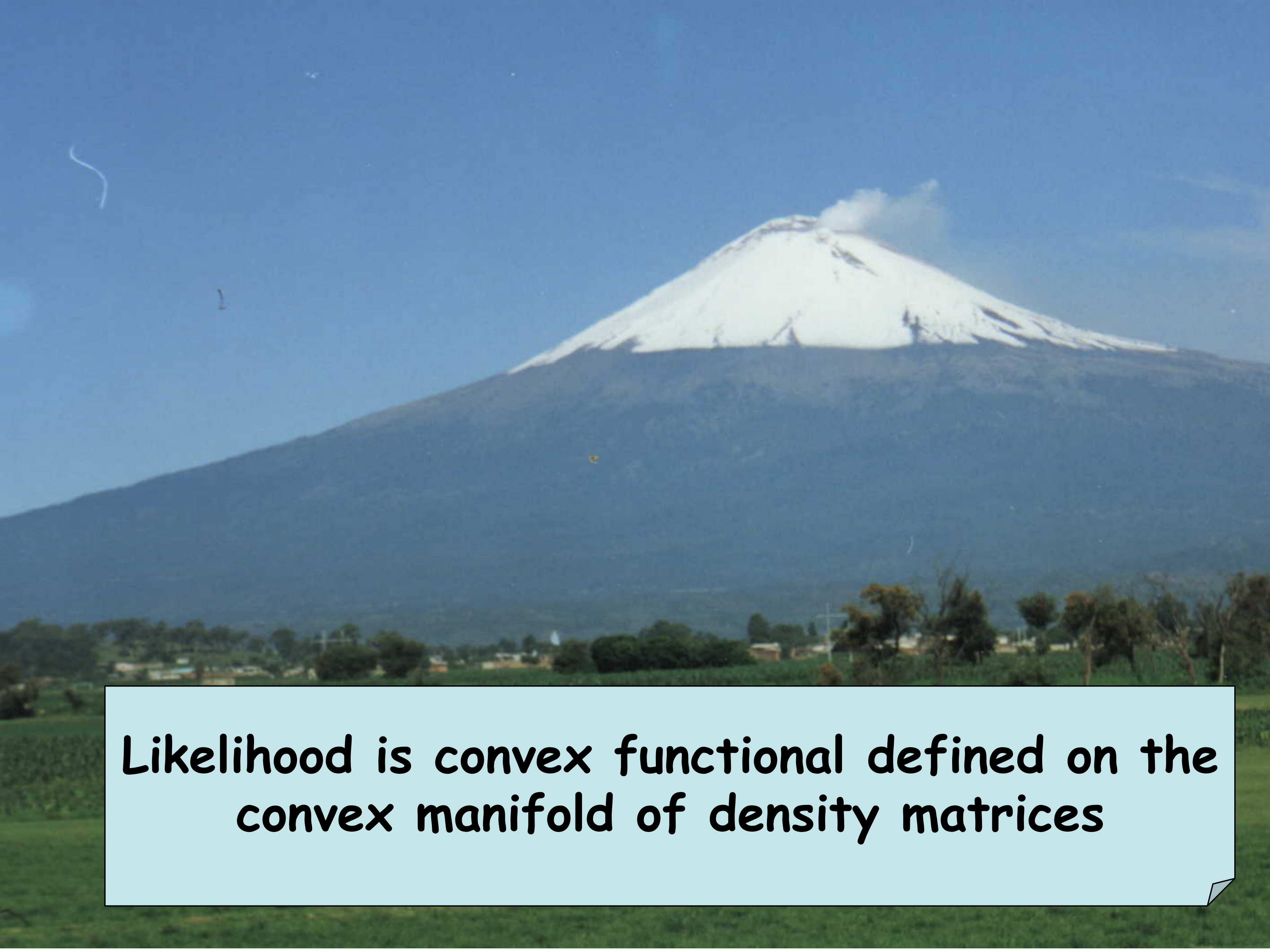
$$R \, \rho = \rho$$

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

MaxLik interpretation

Linear inversion

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

MaxLik inversion

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

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

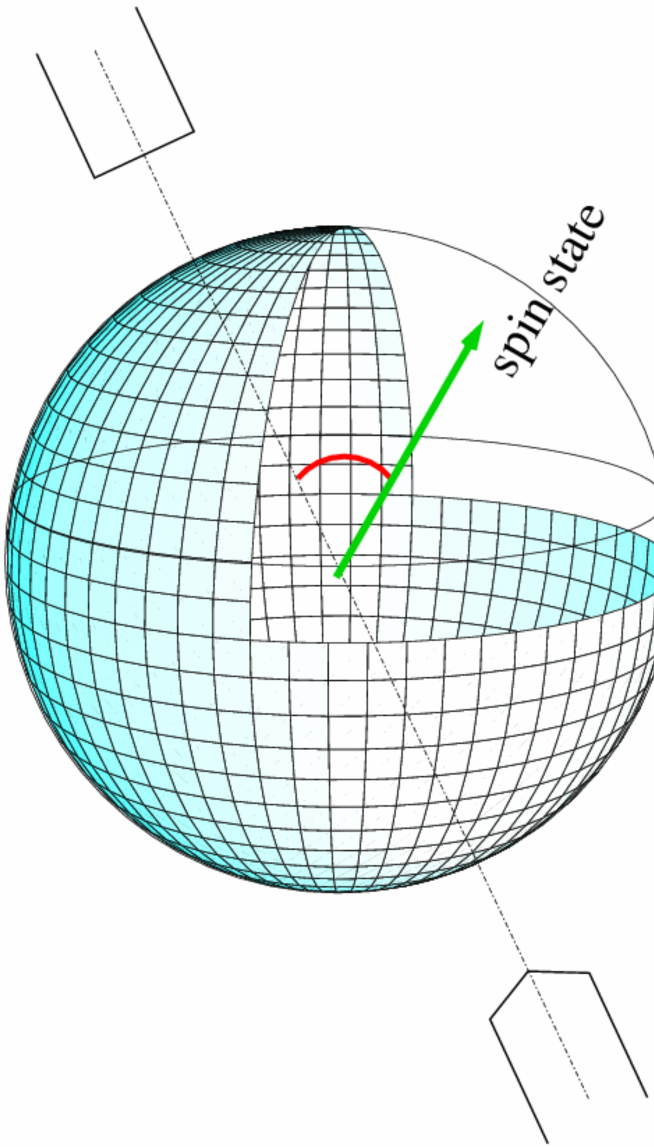
Why the optimal estimation must be nonlinear:

- 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

$$G^{-1/2} \,/ \qquad \ldots\ldots\ldots \qquad /\, G^{-1/2}$$

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

Every measurement is complete…somewhere !!!

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

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

$$\mathbf{R} = (\textstyle\sum_j p_i)\,/(\sum_j N_i)\,\sum_k (N_k/p_k)\,A_k$$

$$G = \textstyle\sum_i A_i$$

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

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

Solution in the iterative form

$$\rho_G = R_G \rho_G R_G$$

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

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

- Experimental uncertainty

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

- Experimental uncertainty relations

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

- Self-consistency check:

measured data f_k should be compared with the mean values $\mathrm{Tr}(\rho A_k)$ 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!

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

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

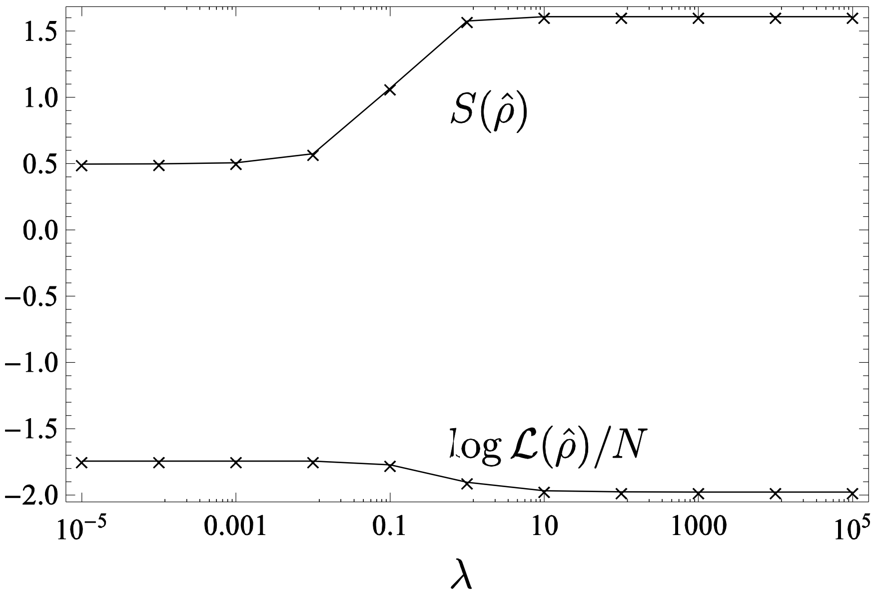
Entropy $S = -Tr(\rho \log \rho)$ quantifies the ignorance

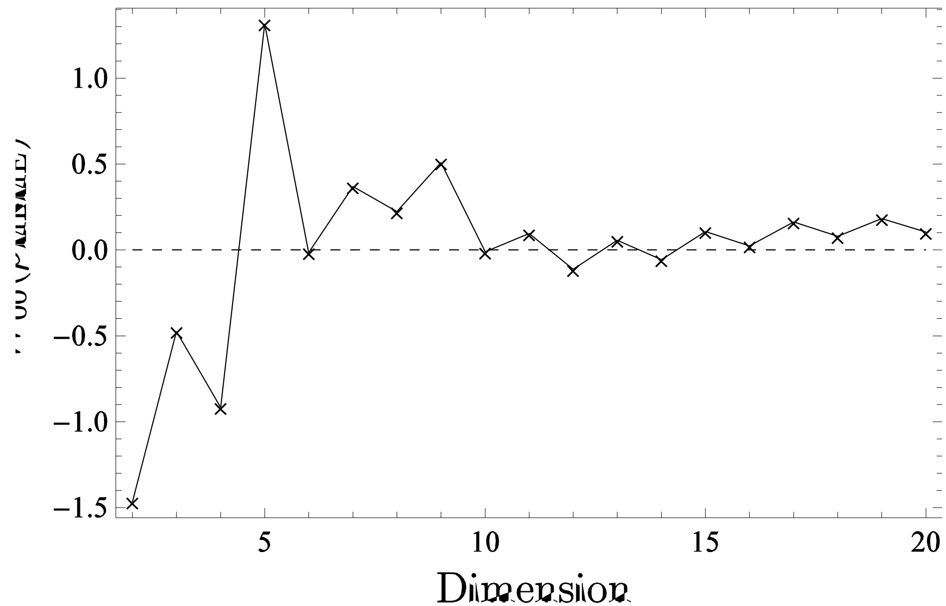
$I(\lambda,\rho) = \lambda S(\rho) + 1/N \log L(\rho)$

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

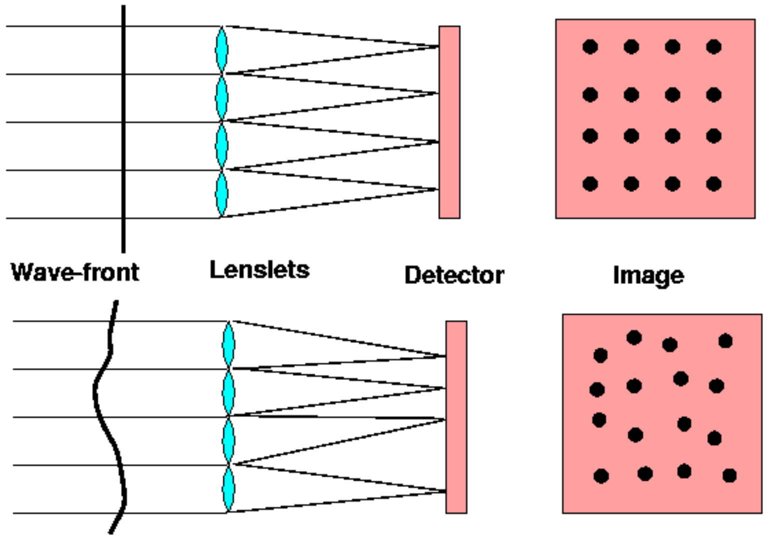
Quantum computation with 5 qbits: $d = 2^5 = 32$
Quantum state: $\sim 10^3$ parameters
Fisher info: $\sim 10^6$ parameters
CP maps: $\sim 10^6$ parameters
Fisher info of CP maps: $\sim 10^{12}$ parameters

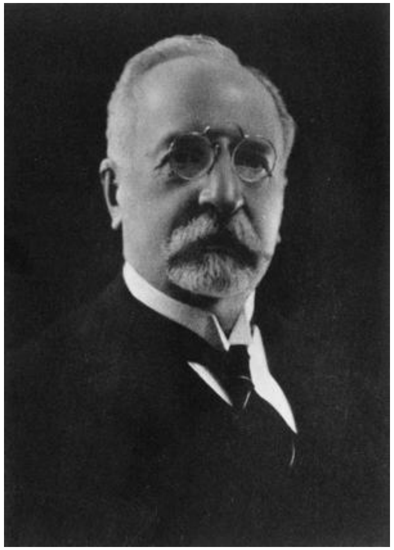
End of General concepts

Several examples

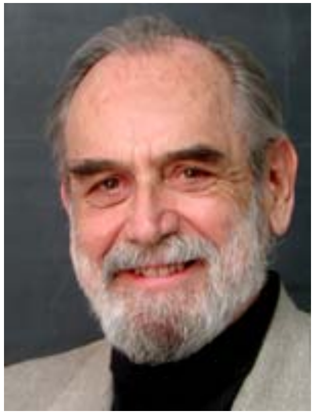
- Phase estimation
- Transmission tomography
- Tomography of CP maps
- Reconstruction of photocount statistics
- Image reconstruction
- Vortex beam analysis
- Quantification of entanglement
- Reconstruction of neutron wave packet
- Reconstruction based on homodyne detection
- Full reconstruction based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

Scanning of the optical field: Hartmann-Shack sensor



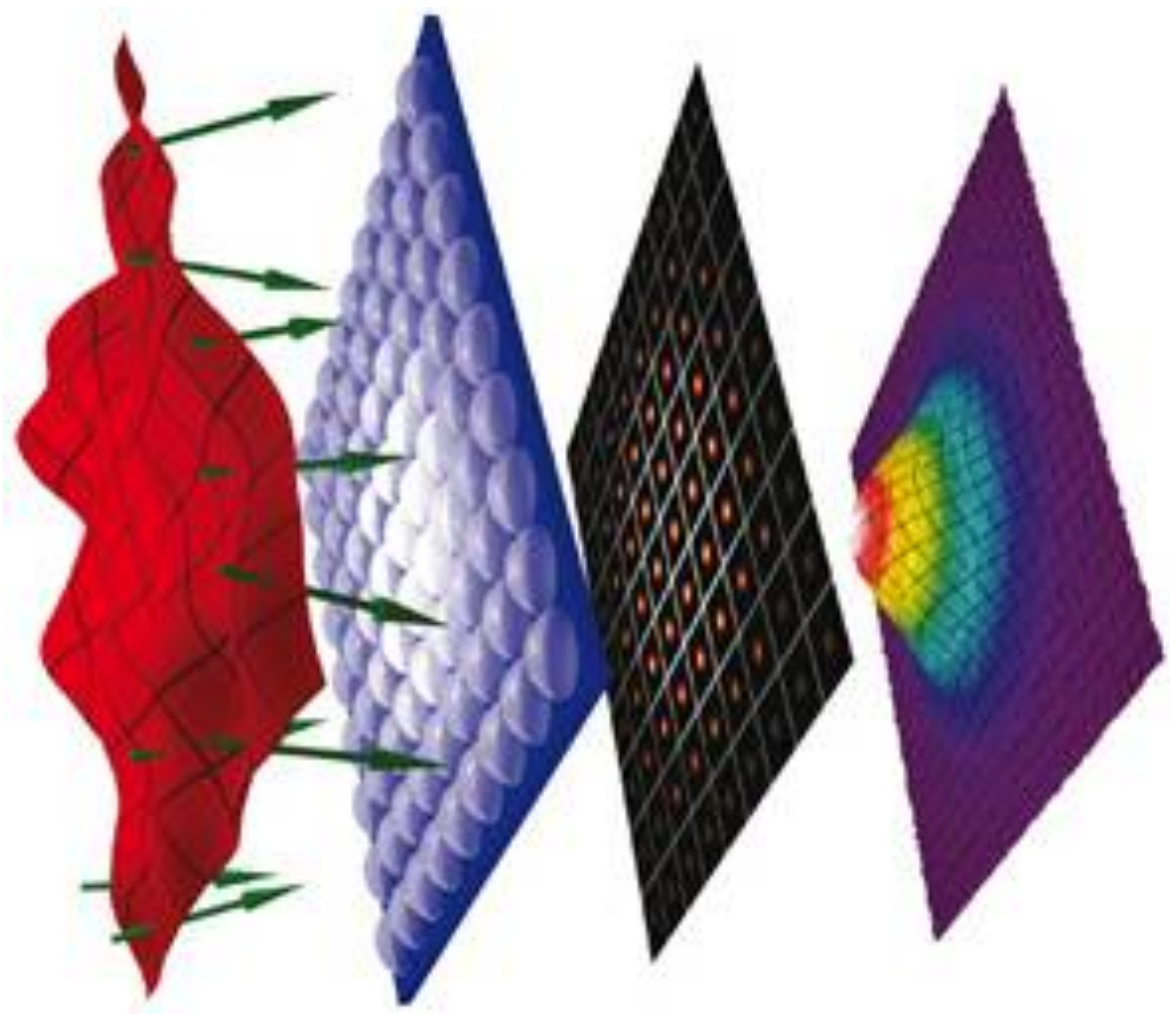


Johannes Hartmann
(1865-1936)



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

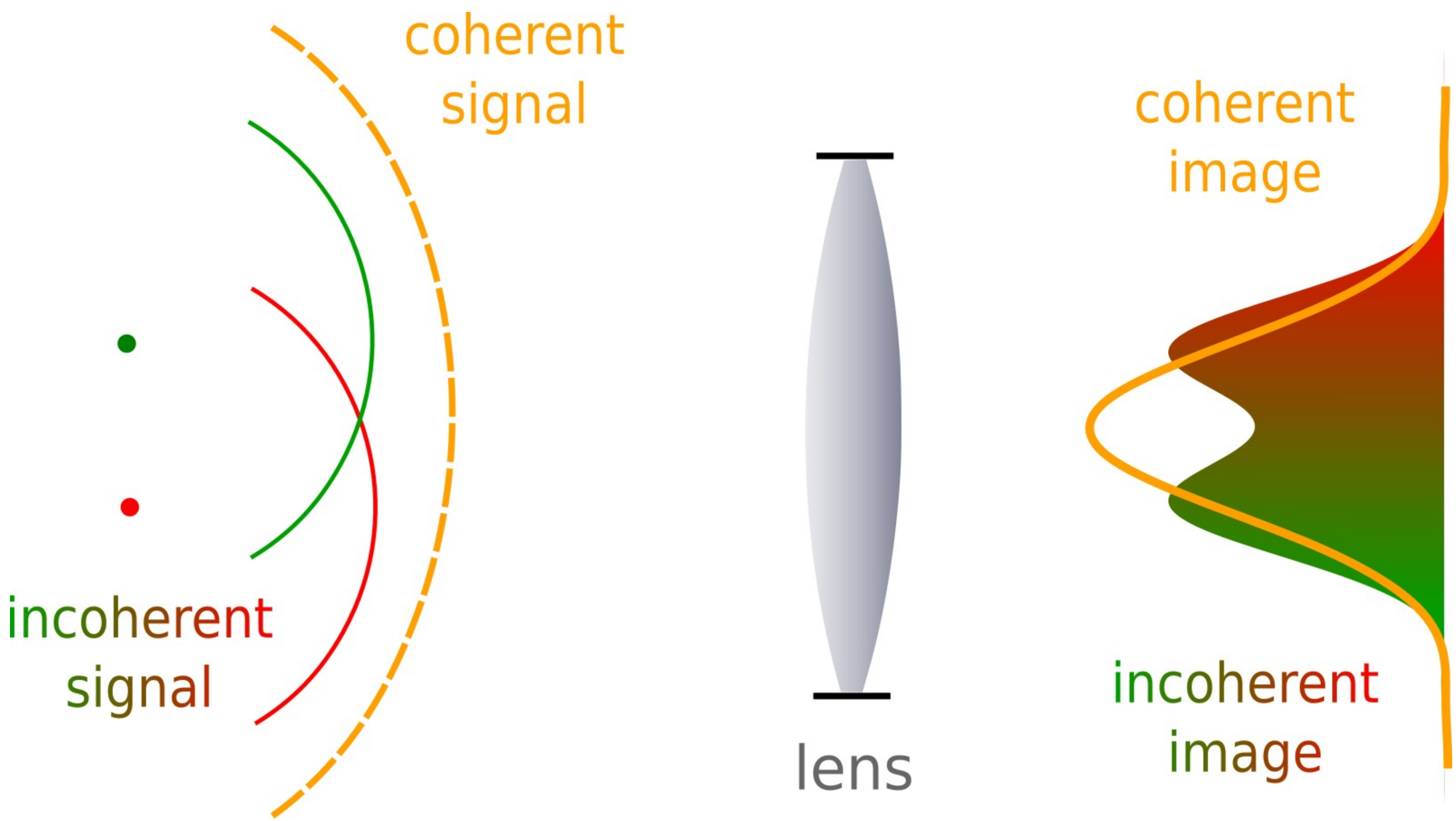
$$\Delta x\;\Delta p \geq \hbar/2$$

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

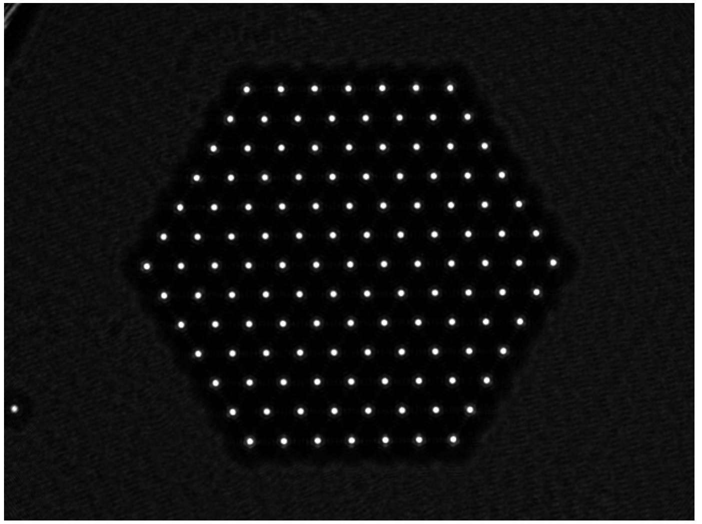
$$\Delta X\;\Delta P \geq \hbar$$

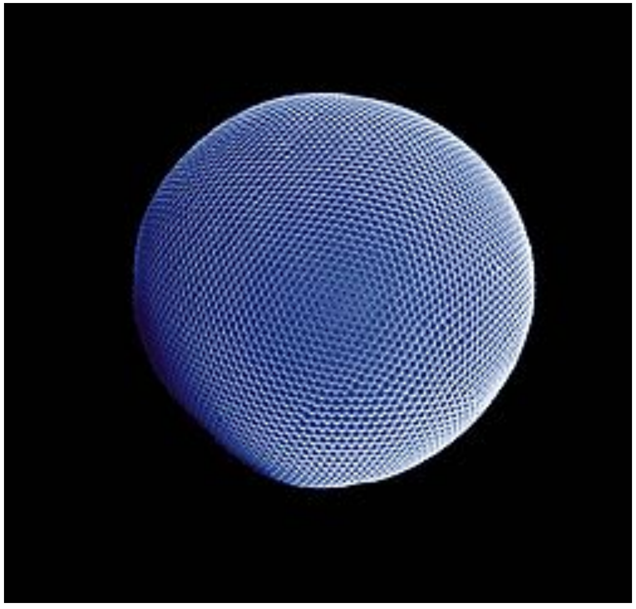
See the excellent paper: S. Stenholm, *Simultaneous measurement of conjugate variables, Annals of Physics 218, 233-254 (1992).*

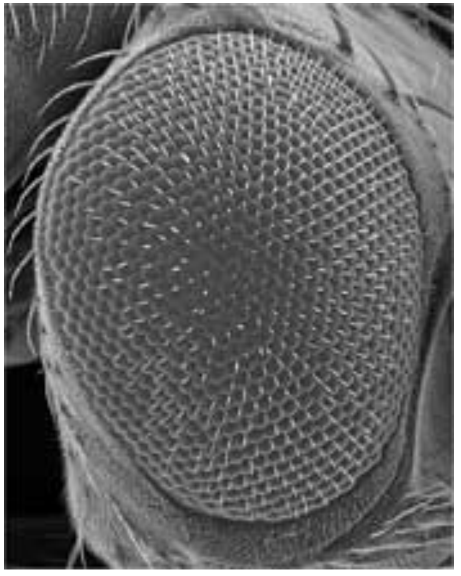
Detection of partially coherent signal



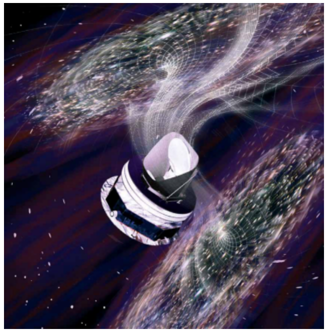
Hartmann-Shack sensor of the wavefront?

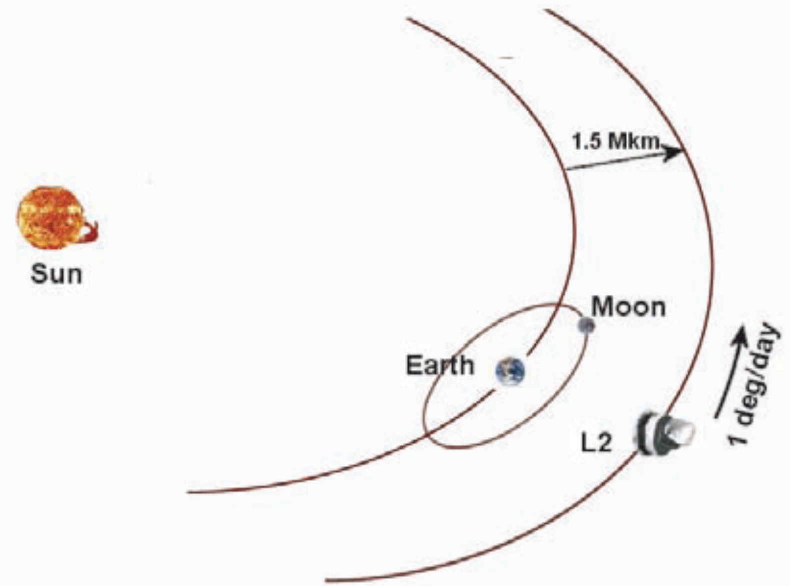


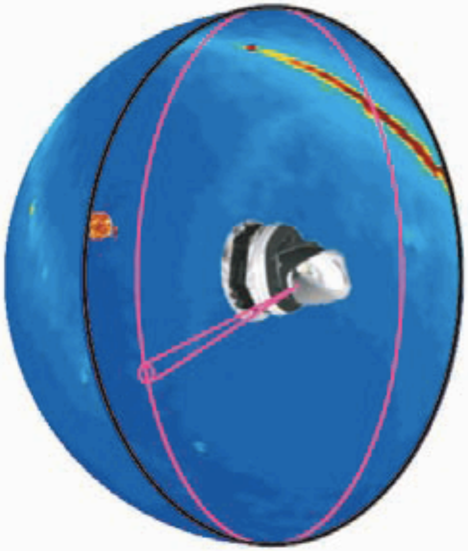




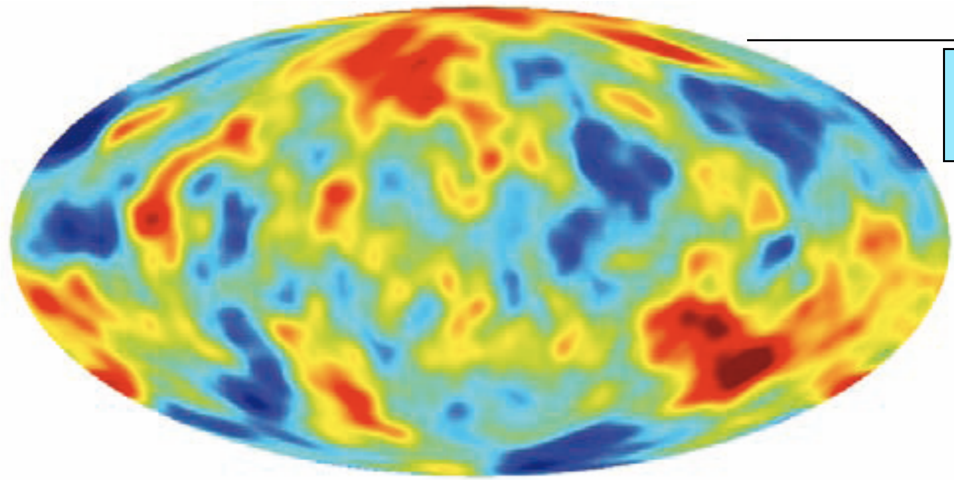
Planck mission of ESA: scanning of cosmic background radiation



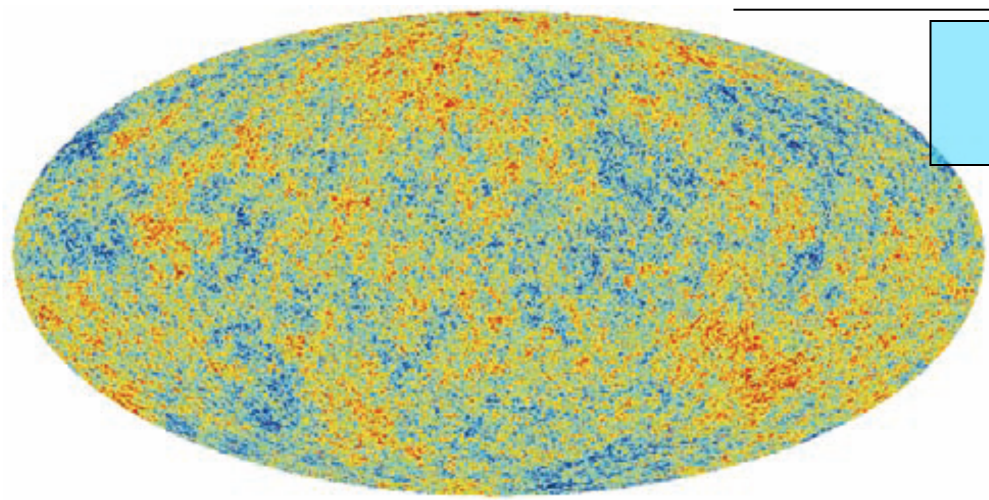




Temperature anisotropies



COBE-DMR resolution



Planck resolution