

Electron Crystallography  
Workshop

University of California, Davis  
September 7 - 13, 2008



Sponsored By:




<http://2dx.org/workshop>

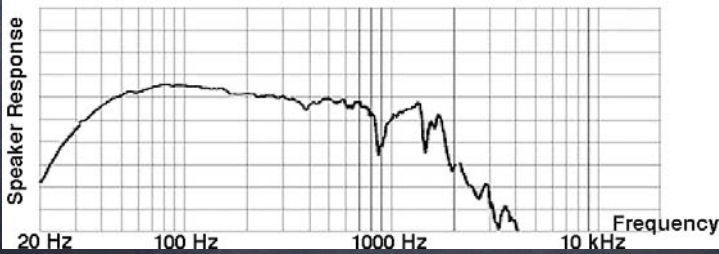
1

# CTF

Henning Stahlberg, UC Davis

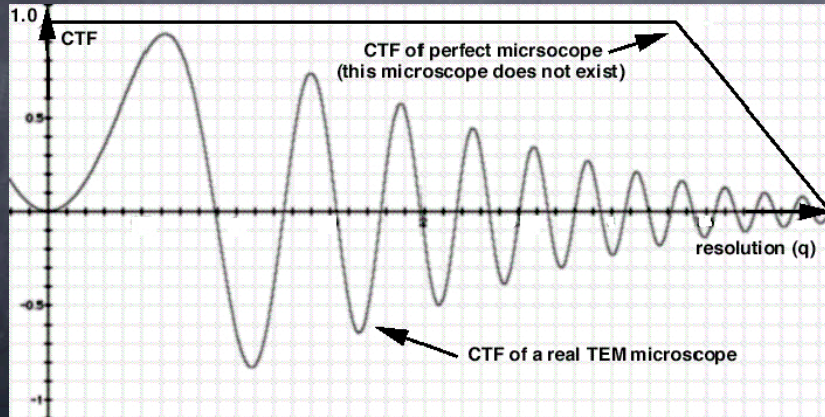


## Transfer Function



Speaker Response vs Frequency (20 Hz to 10 kHz)

## Transfer Function of a TEM



CTF vs resolution (q)

CTF of perfect microscope (this microscope does not exist)

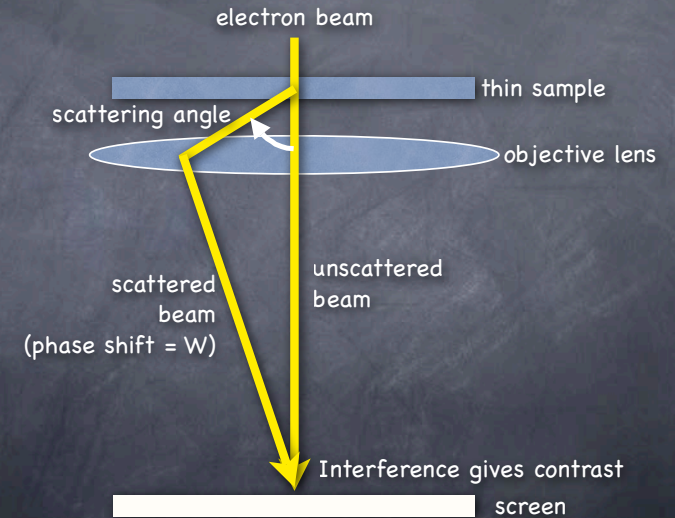
CTF of a real TEM microscope



## Otto Scherzer

(Mar. 9, 1909 - Nov. 15, 1982)

## Contrast

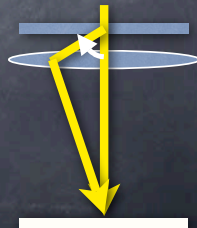


## Scherzer Formula

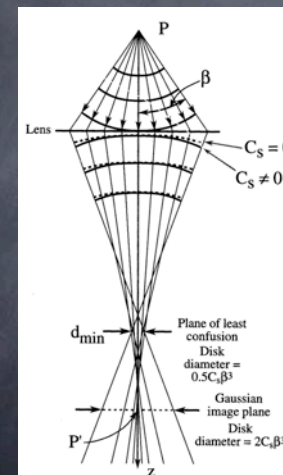
$$\gamma(\mathbf{u}) = \frac{2\pi W}{\lambda} = \frac{\pi}{2} [C_s \lambda^3 \mathbf{u}^4 - 2\Delta z \lambda \mathbf{u}^2]$$

$\sin(\gamma(\mathbf{u}))$ : phase contrast transfer function  
 $\cos(\gamma(\mathbf{u}))$ : amplitude contrast transfer function

$\mathbf{u}$ : scattering vector ( $\approx$ scattering angle)  
 $W$ : wave aberration  
 $\lambda$ : electron wavelength  
 $\Delta z$ : defocus  
 $C_s$ : spherical aberration constant



## Spherical Aberration



$C_s$

Limits resolution beyond  $\sim 2 \text{ \AA}$

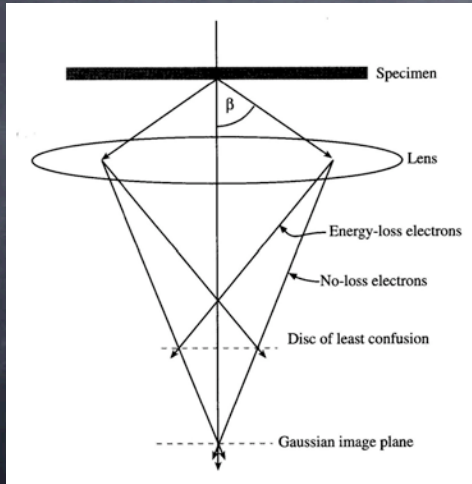
Electrons further from the axis are more strongly bent back towards the axis.

Or:

Electron waves further from the axis receive more phase shift.



# Chromatic Aberration



$C_c$

Limits resolution beyond  
 $\sim 0.5 \text{ \AA}$

Electrons of lower energy are bent more strongly than those of zero-loss energy.

# CTF

$$\text{CTF}(u) = \{ A * \cos(\gamma(u)) - \text{sqrt}(1-A^2) * \sin(\gamma(u)) \} * E(u)$$

$$\gamma(u) = \frac{2\pi W}{\lambda} = \frac{\pi}{2} [ C_s \lambda^3 u^4 - 2\Delta z \lambda u^2 ]$$

$\sin(\gamma(u))$ : phase contrast transfer function

$\cos(\gamma(u))$ : amplitude contrast transfer function

$u$ : scattering vector ( $\approx$ scattering angle)

$A$ : Amplitude contrast fraction. (neg. stain: use 0.07)

# Envelope functions

$$E(u) = E_s(u) \cdot E_c(u) \cdot E_d(u) \cdot E_v(u) \cdot E_D(u)$$

with

$E_s(u)$ : angular spread of the source  $\alpha$  = opening angle

$$E_s(u) = \exp\left[-\left(\frac{\pi\alpha}{\lambda}\right)^2 \left(\frac{\delta X(u)}{\delta u}\right)^2\right] = \exp\left[-\left(\frac{\pi\alpha}{\lambda}\right)^2 (C_s \lambda^3 u^3 + \Delta f \lambda u)^2\right]$$

$E_c(u)$ : chromatic aberration

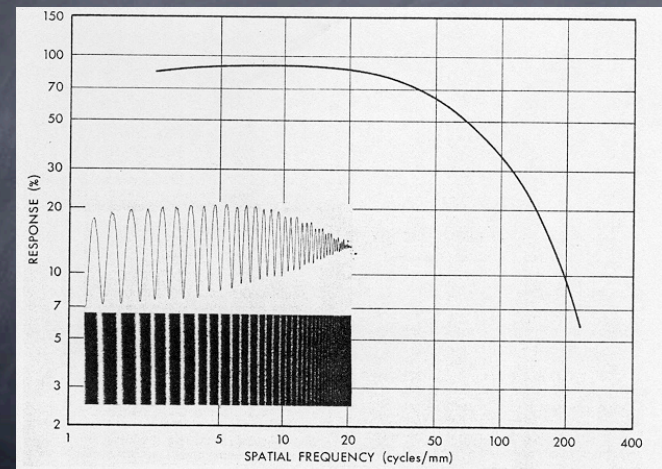
$$E_c(u) = \exp\left[\frac{1}{2}(\pi\lambda\delta)^2 u^4\right] \quad \delta = C_c \sqrt{4\left(\frac{\Delta I_{obj}}{I_{obj}}\right)^2 + \left(\frac{\Delta E}{V_{acc}}\right)^2 + \left(\frac{\Delta V_{acc}}{V_{acc}}\right)^2}$$

$E_d(u)$ : specimen drift

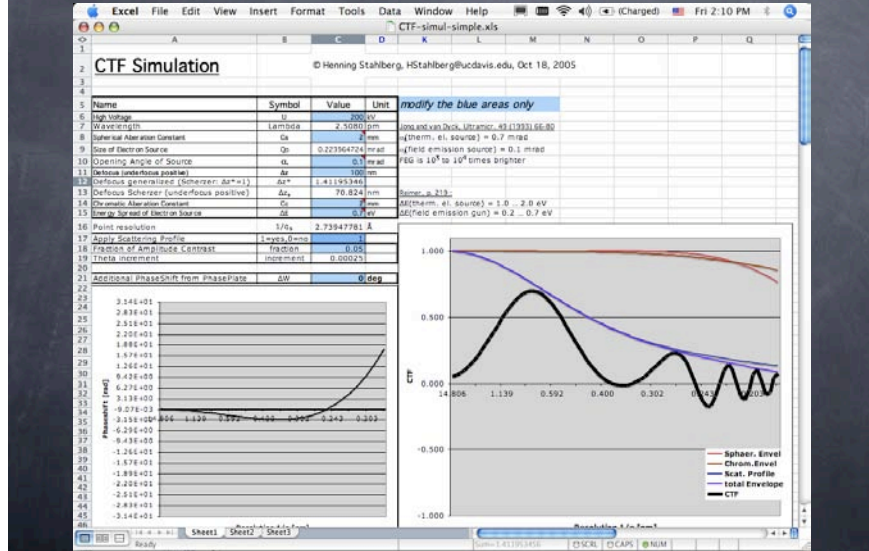
$E_v(u)$ : specimen vibration

$E_D(u) = MTF_D(u)$ : detector

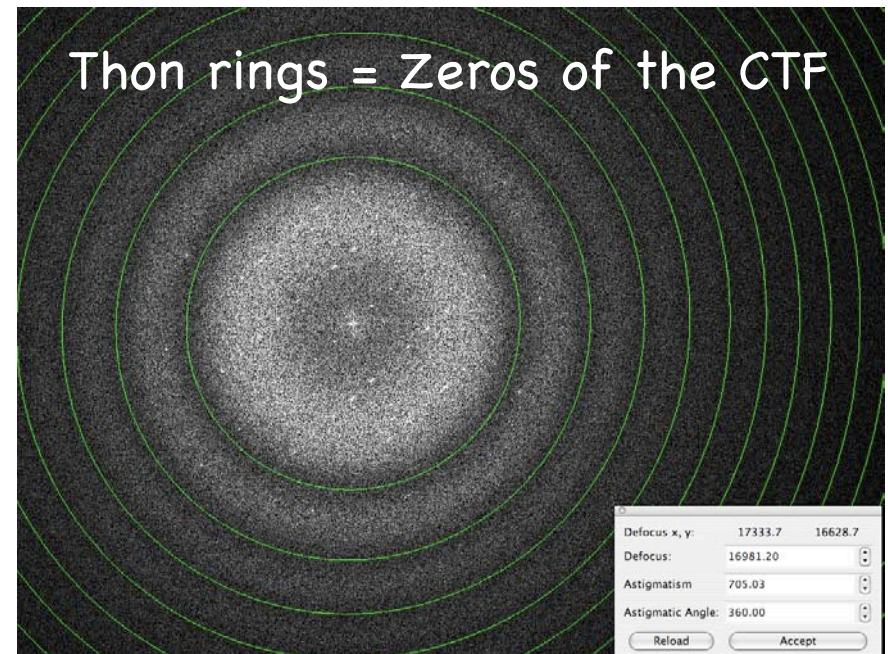
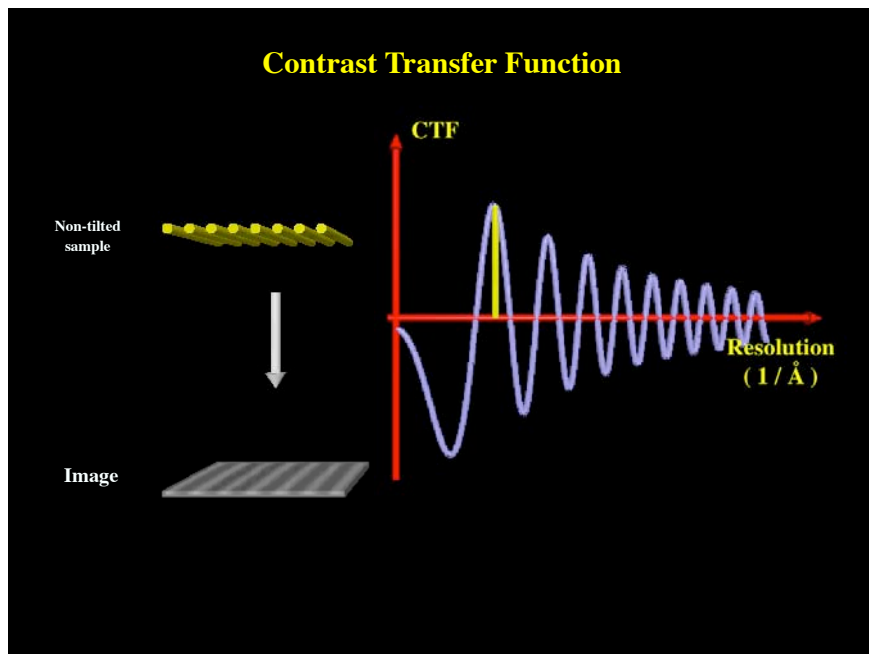
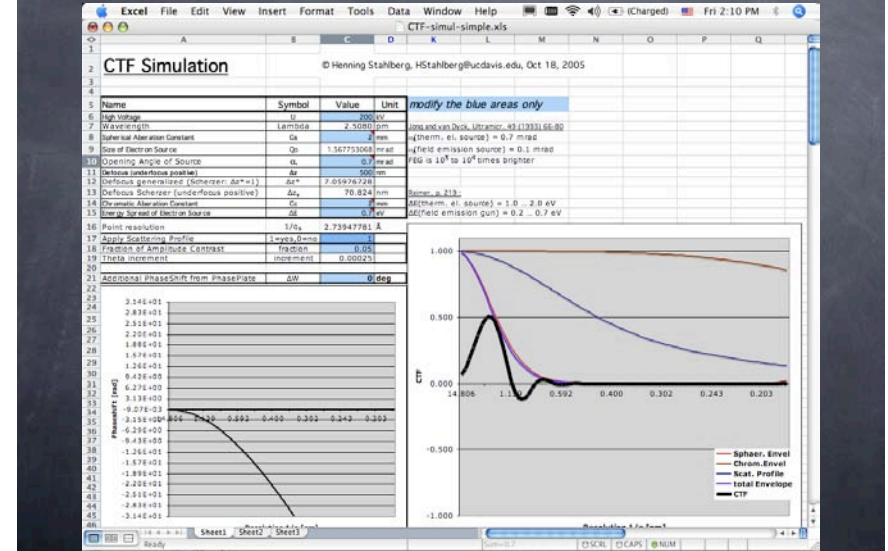
# Modulation Transfer Function (MTF) of photographic film



# CTF simulation



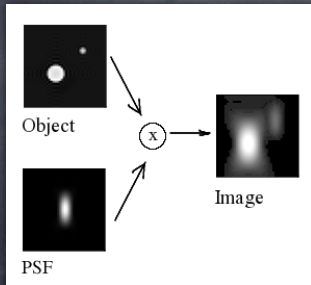
Excel sheet available on <http://stahlberglab.ucdavis.edu>





Real Space:  
Point Spread Function

$$\text{Object} \otimes \text{PSF} = \text{Image}$$

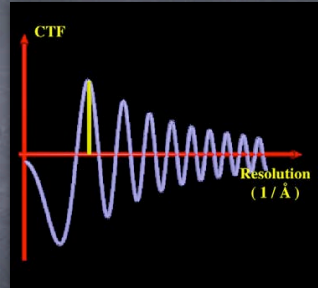


$$\text{FT}(\text{PSF}) \approx \text{CTF}$$

$$\text{PSF} \approx \text{FT}(\text{CTF})$$

Fourier Space:  
Contrast Transfer Function

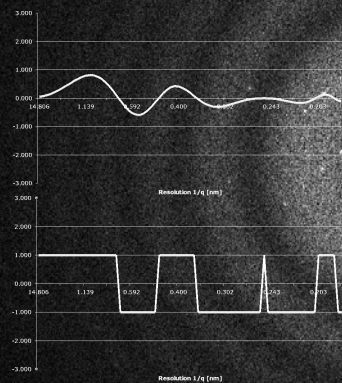
$$\text{FT}(\text{Object}) \cdot \text{CTF} = \text{FT}(\text{Image})$$



Can we correct the CTF ?

CTF correction: Phase flipping

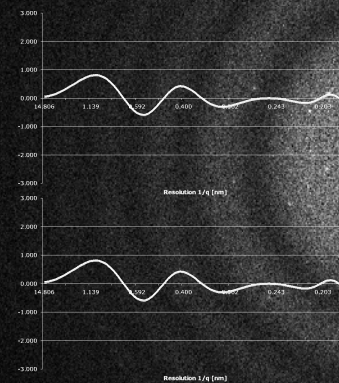
$$\text{NewImage} = \text{FT}^{-1} \{ \text{FT}(\text{Image}) \cdot \text{sign}(\text{CTF}) \}$$



"Invert the phase behind every second Thon ring"

CTF correction: multiply by CTF

$$\text{NewImage} = \text{FT}^{-1} \{ \text{FT}(\text{Image}) \cdot \text{CTF} \}$$

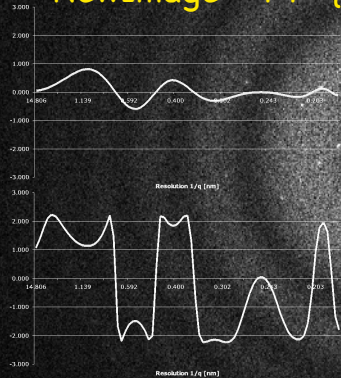


"Multiply by CTF, to increase S/N"



## CTF correction: Wiener Filter

$$\text{NewImage} = \text{FT}^{-1} \left\{ \text{FT}(\text{Image}) \cdot \frac{\text{CTF}}{\text{CTF}^2 + N^2} \right\}$$



"Divide by the CTF (sort of...)"

## Conclusions CTF

- The CTF defines the transfer of contrast from the sample onto the image.
- The PSF defines the impact on the image from a point in the sample.  $\text{PSF} = \text{FFT}(\text{CTF})$ .
- CTF needs to be fitted and corrected.
- CTF for tilted samples is a complicated and important story, which will be told by Ansgar Philippsen after the coffee break.

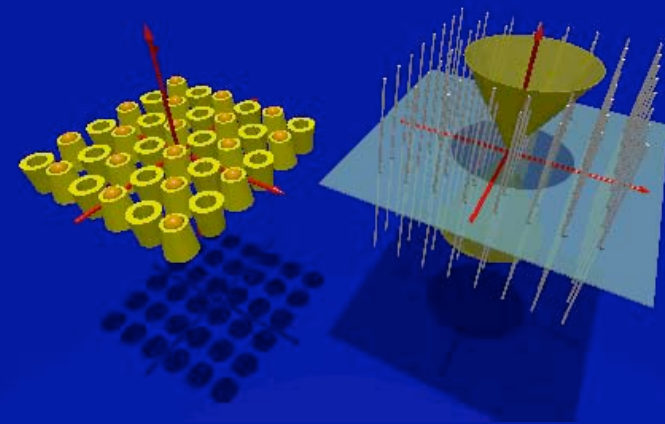
## Tilt Geometry

Henning Stahlberg, UC Davis

## The Missing Cone

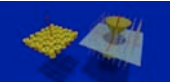

Real Space

Fourier Space



Diffraction Spots in 2D  
Lattice Lines in 3D

## Tilt Geometry

**Coordinate System of the Recorded Image**

- Where is the tilt axis? TLTAXIS: angle from X-axis to tilt-axis
- How much tilt was there? TLTANG: tilt angle of sample
- How is the crystal oriented? TLTAXA: angle from tilt-axis to A\*

**Coordinate System of the Sample**

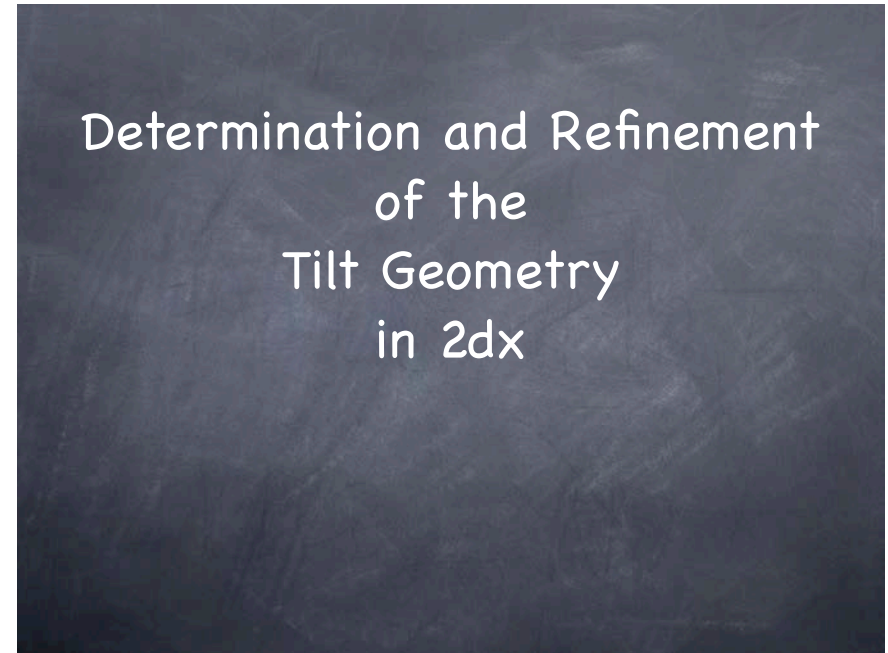
- How much tilt was there? TANGL: tilt angle of sample  
*(Same as TLTANG, but Sign dependent on: Is A\* above tilt axis? Sign of TLTAXA? Handedness of the lattice assignment?)*
- How is the crystal oriented? TAXA: angle from tilt-axis to A\* on sample  
*(different than TLTAXA!)*

Four ways to determine/refine tilt geometry:

- From defocus of negative (ctfsearch3)
- From lattice distortion (lattilt)
- From spot-splitting (ttrefine)
- From comparison with 3D dataset (origiltld)

----- Defocus values in 49 positions on image ----						
12780	11990	10960	9900	9510	8430	7500
12610	10920	10800	9480	9330	8070	7090
11770	11230	10190	8890	8460	7390	6890
11400	10190	9480	8685.35	7770	6790	6590
11090	9630	9360	8590	7570	6730	5930
10590	9540	9260	7850	7240	6060	5670
10580	9460	8330	7310	6640	5930	5200

----- Geometry calculation from -----				
	defocus	lattice	spotsplit	merging
	(ctfsearch3)	(lattilt)	(ttrefine)	(origiltld)
TLTAXIS =	63.1014	64.43140	64.4314	-----
TLTANG =	42.7849	45.38935	45.3894	-----
TLTAXA =	-87.7667	-87.76840	-87.7681	-----
TANGL =	45.3900	45.38935	45.3894	46.998
TAXA =	-86.8216	-86.82402	-86.8236	-93.714



## Determination and Refinement of the Tilt Geometry in 2dx

### Defocus Gradient accross image

Rough Tilt Geometry, but absolute sign of tilt angle (TLTAXIS, TLTANGL)  
No clue about crystal orientation (TAXA, TANGL)

### Lattice Distortion

Precise Tilt Geometry if tilt larger than 25°,  
but no clue about sign of tilt angle (sign taken from above)

### SpotSplitting

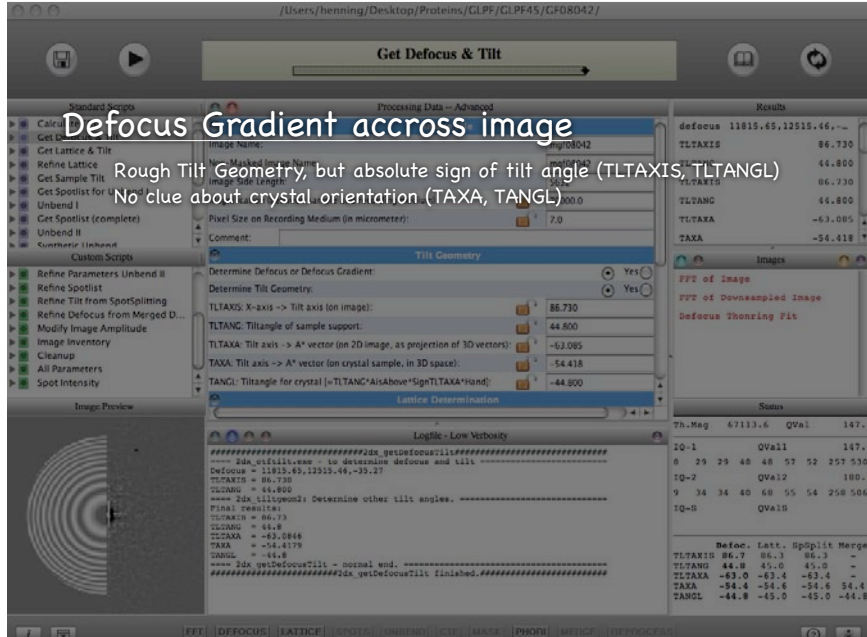
Precise Tilt Geometry if tilt larger than 25°,  
but no clue about sign of tilt angle (sign taken from above)

### 3D Merging

Precise Tilt Geometry for sample (TAXA, TANGL),  
but no clue about carbon film orientation (TLTAXIS, TLTANGL)

## Defocus Gradient accross image

Rough Tilt Geometry, but absolute sign of tilt angle (TLTAXIS, TLTANGL)  
No clue about crystal orientation (TAXA, TANGL)





**Get Lattice & Tilt**

Standard Scripts: Calculate FFT, Get Defocus & Tilt, Get Lattice & Tilt, Refine Lattice, Get Sample Tilt, Get Spotlist for Unbind I, Unbind I, Get Spotlist (complete), Unbind II, Get Spotlist (complete), Unbind II

Processing Data - Standard: Image File, Image Size Length: 5632, Magnification Between sample and recording medium: 70000.0, Pixel Size on Recording Medium (in micrometers): 7.0

Results: lattice "43.075, -36.6", secondLattice "49.061, 43.08", TLTANG 86.39583, TLTAXA -63.41917, TANGL -45.06706

**Lattice Distortion**  
Precise Tilt Geometry if tilt larger than 25°, but no clue about sign of tilt angle (sign taken from above)

**Refine Tilt from SpotSplitting**

Standard Scripts: Initialization, Calculate FFT, Get Defocus & Tilt, Get Lattice & Tilt, Refine Lattice, Get Sample Tilt, Get Spotlist for Unbind I, Unbind I, Get Spotlist (complete), Unbind II, Refine Parameters Unbind II, Refine Spotlist, Refine Tilt from SpotSplitting, Refine Defocus from Merged D., Modify Image Amplitude, Image Inventory, Cleanup, All Parameters, Spot Intensity

Processing Data - Standard: Algorithm Selection: refine tiltaxis, Mode of operation: refine tiltaxis, Number of cycles: 10, MTZ reference for refinement: (none), Overwrite data: Yes

Results: TLTANGIS 86.3960, TLTREFINE\_TLTANG 45.0670, TLTREFINE\_TLTAXA -63.4186, TLTREFINE\_TLTA 54.6847, TLTREFINE\_TANGL -45.0670

**SpotSplitting**  
Precise Tilt Geometry if tilt larger than 25°, but no clue about sign of tilt angle (sign taken from above)

**Initialization**

Standard Scripts: Merge Once, Refine Once, Merge & Refine (Iterative), Generate Image Maps, Final Merge, Generate Merged Map, (Re-)Process all images

Processing Data - Advanced: Directory, PhiRes, TAXA, TANGL, PhiOri, PhiOri Change, rHK

Directory	PhiRes	TAXA	TANGL	PhiOri	PhiOri Change	rHK
GF06653_30	39.92	33.48	29.90	103.9,101.4	-1.20,0.00	
GF06653_30b	43.22	51.21	-30.26	-64.6,-158.5	1.20,-1.80	
GF06666_30	52.87	-80.41	30.79	-72.5,-82.5	0.30,-0.30	
GF06667_30	41.89	17.60	-31.06	146.2,5.2	-1.80,-4.50	
GF06670_30	36.41	38.23	28.56	26.6,52.3	0.90,1.20	
GF06675_30	40.66	60.65	30.16	36.1,-40.6	3.30,1.50	
GF06685_30	33.41	67.07	30.44	37.1,-179.2	1.80,0.30	
GF06685_30b	41.94	-82.07	30.70	-40.8,10.1	0.00,-0.60	
GF06687_30	48.34	53.32	30.93	-103.7,122.5	9.00,6.60	
GF06687_30b	51.04	23.60	-31.32	-55.0,-15.7	-9.60,-14.10	
GF06688_30	39.27	55.84	32.10	-67.6,7.2	0.60,-0.60	
GF06688_30b	42.49	29.05	-32.19	-33.2,123.3	4.50,6.90	
GF06690_30	30.86	57.12	31.95	41.2,175.6	-0.30,-0.60	
GF06690_30b	33.73	31.29	-32.03	-81.5,175.1	-1.20,-3.30	
GF06693_30	38.08	30.94	-30.35	31.8,-104.2	-1.20,1.80	
GF06694_30b	49.34	38.57	31.19	12.1,16.8	-1.20,-0.30	
GLPF45						
GF08005	45.10	6.60	46.97	10.4,-27.6	1.20,-12.30	
GF08013	45.16	45.94	46.09	-175.1,-88.8	-6.00,-11.40	
GF08025	41.07	61.87	-44.26	-171.8,37.0	7.20,3.30	
GF08032	47.00	86.16	45.45	-73.0,-40.9	6.00,0.60	
GF08032b	57.63	86.15	-44.95	-107.2,-81.3	11.70,9.90	
GLPF60	26.83	35.38	59.39	82.0,-8.1	3.00,4.20	
GF08096a	26.83	35.38	59.39	82.0,-8.1	3.00,6.00	
GF08096a-30	27.23	35.38	-59.39	84.4,-8.2	2.40,0.00	

**3D Merging**  
Precise Tilt Geometry for sample (TAXA, TANGL), but no clue about carbon film orientation (TLTAXIS, TLTANG)

**Generate Map**

Standard Scripts: Get Spotlist for Unbind I, Unbind I, Get Spotlist (complete), Unbind II, Synthetic Unbind, Refine Parameters, Correct CTF, Maximum Likelihood, Generate Map

Processing Data - Advanced: Lattice Determination: Real Unit Cell Length (for entire project): 104.0, 104.0; Real Cell Angle (for entire project): 90.0; Crystal Unbinding: Use Synthetic Reference if possible?: Use Fourier filtered reference; Maximum Likelihood Algorithms: Use Maximum Likelihood?: Yes; Common Image Processing: Lower Resolution Limit (RESMIN): 200.00; Upper Resolution Limit (RESMAX): 4.00; ATAT (z-dimension of unit cell to reconstruct): 200.0; Temperature Factor for Map Generation: 20.00; Spider scripts: Generate correctly scaled map with SPIDER? (interpolated and sheared multi-unit-cell): Yes; Scale of the final map (specify the size of one pixel in Angstrom): 0.25; Number of unit cells in final map: 6; Keep Large Temporary Files?: Yes

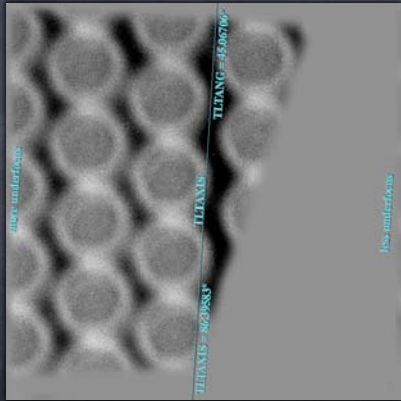
Results: Th.Mag 67113.6 QVal 186.6, IQ-1 OVal1 147.6, IQ-2 OVal2 186.3, IQ-3 OVal3 174.275

**Generate Map**



## Display of Tilt Geometry in 2dx

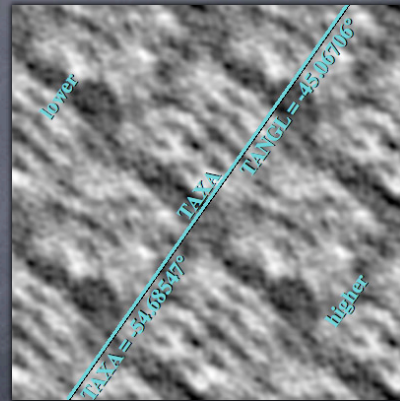
Sample Image



“T”

TLTAXIS, TLTANGL

Final Map



“Shift-T”

TAXA, TANGL

# Maximum Likelihood

Henning Stahlberg, UC Davis

Xiangyan Zeng, FVSU, GA

with Niko Grigorieff, Brandeis Univ. and HHMI

## Cross-Correlational Alignment and Averaging



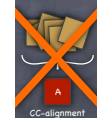
$X_i$  Particles

$\phi_i$  Alignment Parameters

$A$  Average

$$A = \frac{1}{N} \sum_{i=1}^N X_i(\phi_i)$$

Align particles (e.g. via CCF), and then average them.



## Maximum-Likelihood Processing

$$\Theta = (A, \sigma, \xi)$$



$\Theta$  Model Parameters  
 $A$  Structure  
 $\sigma$  Std.dev. of Gaussian noise  
 $\xi$  Std.dev. of distribution of alignment parameters

$X_i$  Observed Particles  
 $X_i = A(-\phi_i) + \sigma R_i$   
 Gaussian Noise R of distribution  $\sigma$

Find model parameters (structure and distributions for alignment and noise), so that the given set of images has the highest likelihood of being photographed.

$$L(\Theta) = \sum_{i=1}^N \ln P(X_i|\Theta)$$

Goal: Maximize  $L$

$$\Theta = (A, \sigma, \xi)$$



## Maximum-Likelihood Processing

$$L(\Theta) = \sum_{j=1}^N \ln P(X_j|\Theta)$$

Goal: Maximize  $L$

$$L(X|\Theta) = \sum_{i=1}^N \ln \int \gamma_i(\phi, \Theta) d\phi,$$

where

$$\gamma_i(\phi, \Theta) = P(X_i|\phi, \Theta) f(\phi|\Theta).$$

$$P(X_i|\phi, \Theta) = \left( \frac{1}{\sigma\sqrt{2\pi}} \right)^M \exp \left[ -\frac{|X_i(\phi) - A|^2}{2\sigma^2} \right]$$

$P$ : Probability density for all images under distribution parameters and alignment parameters

$$f(\phi|\Theta) = \left( \frac{1}{\sqrt{2\pi}} \right)^3 \frac{1}{\sigma_x \sigma_y \sigma_z} \times \exp \left[ -\frac{(q_x - \xi_x)^2}{2\sigma_x^2} - \frac{(q_y - \xi_y)^2}{2\sigma_y^2} - \frac{(q_z - \xi_z)^2}{2\sigma_z^2} \right]$$

$f$ : Probability density of all distribution parameters under alignment parameters

Find model parameters (structure and distributions for alignment and noise),

so that the given set of images has the highest likelihood of being photographed.

$$\Theta = (A, \sigma, \xi)$$



## Maximum-Likelihood Processing

$$L(\Theta) = \sum_{j=1}^N \ln P(X_j|\Theta)$$

Goal: Maximize  $L$

$$\frac{\partial L}{\partial a_j} = 0$$

Goal: Find where derivative of  $L$  is zero

with:

$$\frac{\partial L}{\partial a_j} = \sum_{i=1}^N \frac{\int \left( \frac{X_{ij}(\phi) - a_j}{\sigma^2} \right) \gamma_i(\phi; \Theta) d\phi}{\int \gamma_i(\phi; \Theta) d\phi} = 0$$

Solution:

$$A^{(n+1)} = \frac{1}{N} \sum_{j=1}^N \frac{\int X_{ij}(\phi) \gamma_i(\phi; \Theta^{(n)}) d\phi}{\int \gamma_i(\phi; \Theta^{(n)}) d\phi}$$

$$\Theta = (A, \sigma, \xi)$$



## Maximum-Likelihood Processing

$$L(\Theta) = \sum_{j=1}^N \ln P(X_j|\Theta)$$

Goal: Maximize  $L$

$$\frac{\partial L}{\partial a_j} = 0$$

Solution:

$$A^{(n+1)} = \frac{1}{N} \sum_{j=1}^N \frac{\int X_{ij}(\phi) \gamma_i(\phi, \Theta^{(n)}) d\phi}{\int \gamma_i(\phi, \Theta^{(n)}) d\phi},$$

Structure  $A$

$$\sigma^{(n+1)} = \sqrt{\frac{1}{NM} \sum_{i=1}^N \frac{\int |X_i(\phi) - A^{(n)}|^2 \gamma_i(\phi, \Theta^{(n)}) d\phi}{\int \gamma_i(\phi, \Theta^{(n)}) d\phi}},$$

Noise distribution

$$\xi_x^{(n+1)} = \frac{1}{N} \sum_{j=1}^N \frac{\int q_x \gamma_j(\phi, \Theta^{(n)}) d\phi}{\int \gamma_j(\phi, \Theta^{(n)}) d\phi}$$

$$\xi_y^{(n+1)} = \frac{1}{N} \sum_{j=1}^N \frac{\int q_y \gamma_j(\phi, \Theta^{(n)}) d\phi}{\int \gamma_j(\phi, \Theta^{(n)}) d\phi}$$

$$\xi_z^{(n+1)} = \frac{1}{N} \sum_{j=1}^N \frac{\int q_z \gamma_j(\phi, \Theta^{(n)}) d\phi}{\int \gamma_j(\phi, \Theta^{(n)}) d\phi}$$

Alignment center (x,y,rotation)

$$\sigma_x^{(n+1)} = \sqrt{\frac{1}{2N} \sum_{i=1}^N \frac{\int (q_x - \xi_x^{(n)})^2 \gamma_i(\phi, \Theta^{(n)}) d\phi}{\int \gamma_i(\phi, \Theta^{(n)}) d\phi}}$$

Alignment distribution (x,y,rotation)

$$\sigma_y^{(n+1)} = \sqrt{\frac{1}{2N} \sum_{i=1}^N \frac{\int (q_y - \xi_y^{(n)})^2 \gamma_i(\phi, \Theta^{(n)}) d\phi}{\int \gamma_i(\phi, \Theta^{(n)}) d\phi}}$$

$$\sigma_z^{(n+1)} = \sqrt{\frac{1}{2N} \sum_{i=1}^N \frac{\int (q_z - \xi_z^{(n)})^2 \gamma_i(\phi, \Theta^{(n)}) d\phi}{\int \gamma_i(\phi, \Theta^{(n)}) d\phi}}$$

$$\Theta = (A, \sigma, \xi)$$

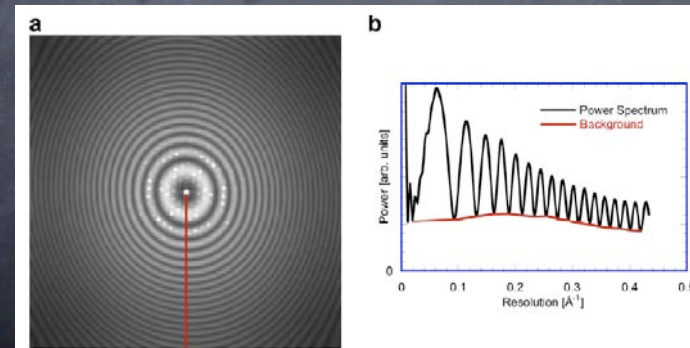


## Maximum-Likelihood Processing

Assumption: Distributions are Gaussian

But: Distributions in TEM images aren't.

e.g.: Thon rings: Background noise is not Gaussian





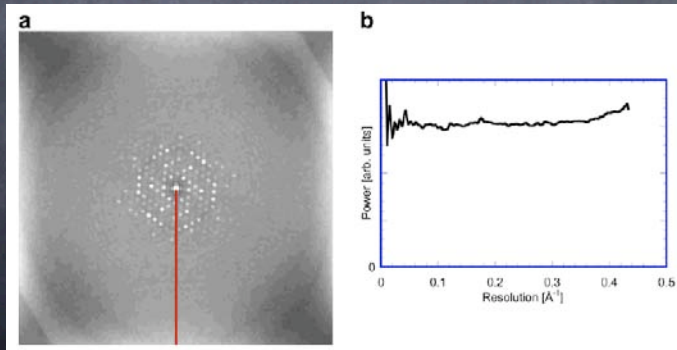
$$\Theta = (A, \sigma, \xi)$$

## Maximum-Likelihood Processing

Assumption: Distributions are Gaussian

But: Distributions in TEM images aren't.

Noise whitening:

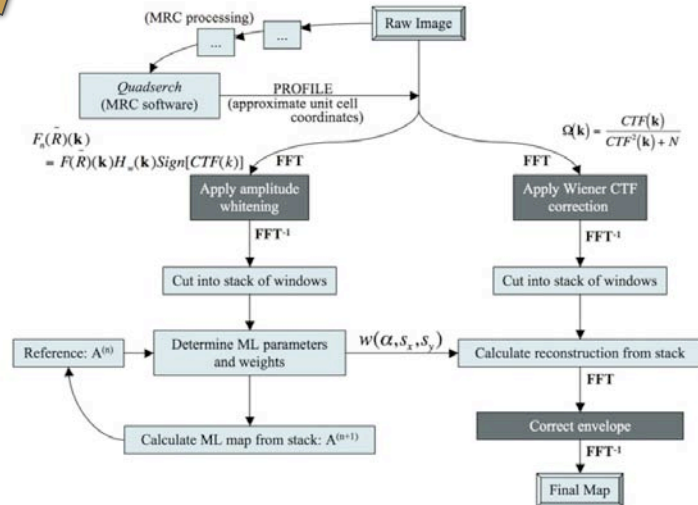


Sigworth, JSB 122 (1998)

Zeng, JSB 160 (2007)

$$\Theta = (A, \sigma, \xi)$$

## Maximum-Likelihood Processing



Sigworth, JSB 122 (1998)

Zeng, JSB 160 (2007)

Maximum Likelihood software interface showing processing data and results.

Processing Data - Standard

Maximum Likelihood Algorithm

- Use Maximum Likelihood?: Yes
- Weighting profile from: maximum likelihood calculation
- Maximum Likelihood Parameters:
  - Total diameter of the windows: 200
  - Diameter of the circular mask: 180
  - Threshold determination method for particle selection: relative percentage given
  - Absolute threshold value for particle selection: 700.0
  - Relative percentage of particles for selection: 90.0
  - First reference from: noise reference
  - Type of low-pass filter: none
  - Low-pass filter radius: 0.0
  - Apply noise whitening: Yes
  - Apply CTF correction: Yes
  - Rotational symmetry: 4-fold

Results

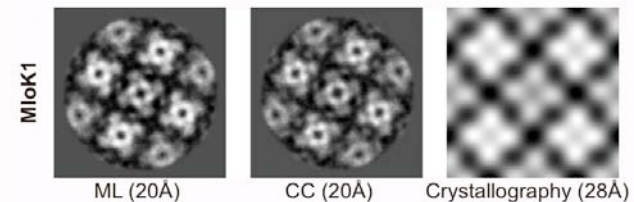
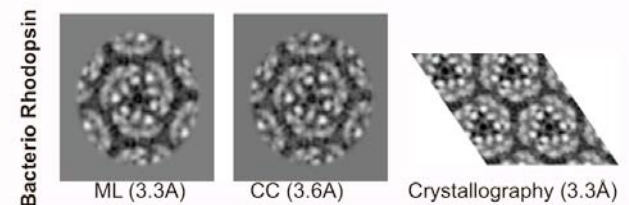
ML\_done: y

Logfile - Moderate Verbosity

```

2dx ML = to run maximum likelihood processing
Starting ML
Calculated x/y center offset = 3.979310, -2.924630
Total number of peaks ..... = 1493
Threshold ..... = 54.049999
Number of peaks above threshold = 1363
Get particle stack
Starting maximum likelihood
    
```

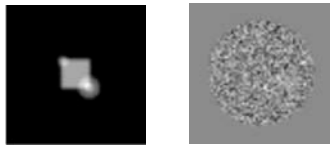
### A maximum likelihood approach to 2D crystals



with Niko Grigorieff, Brandeis

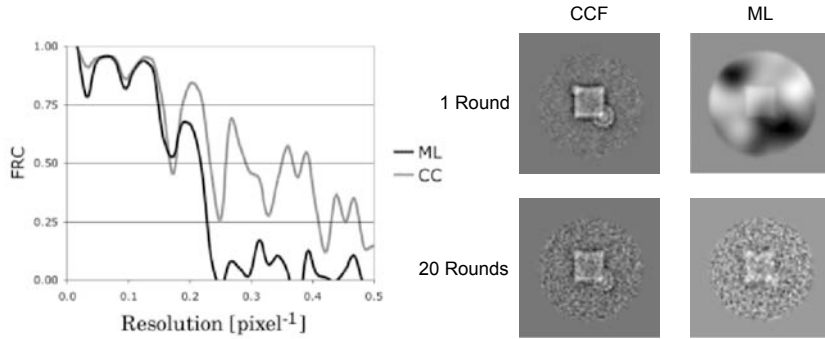
Zeng et al., *J. Struct. Biol.* (2007)

Maximum Likelihood is not entirely free of reference bias

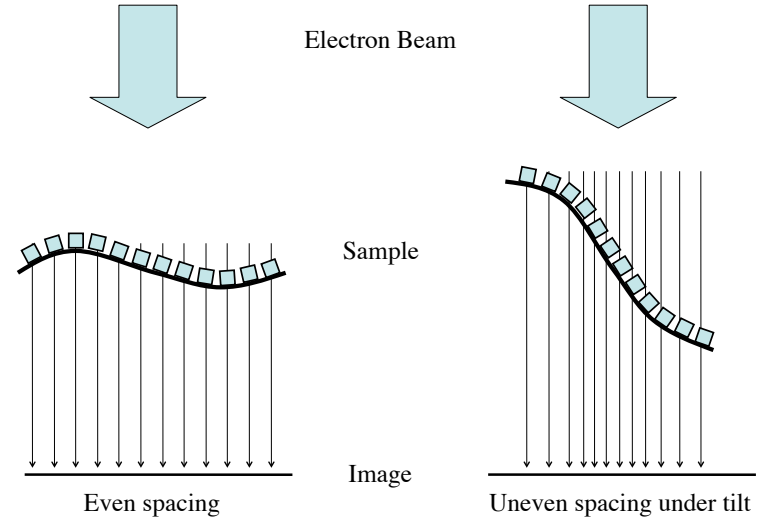


Particle  
Particle with noise  
2000 images,  
SNR=1:200

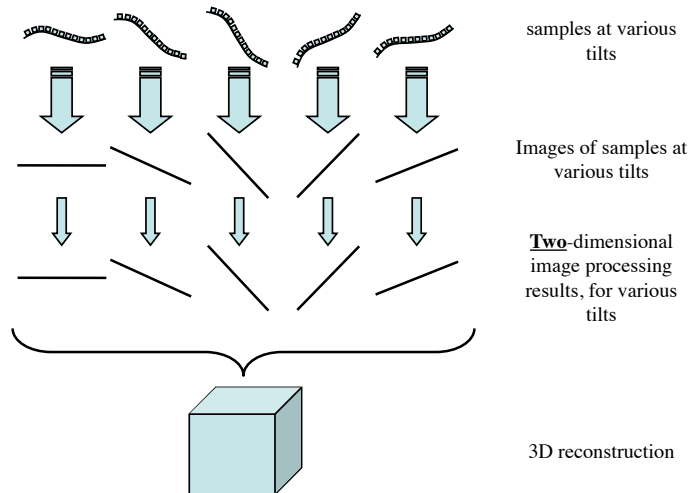
Only the processing result from the noise component is shown:



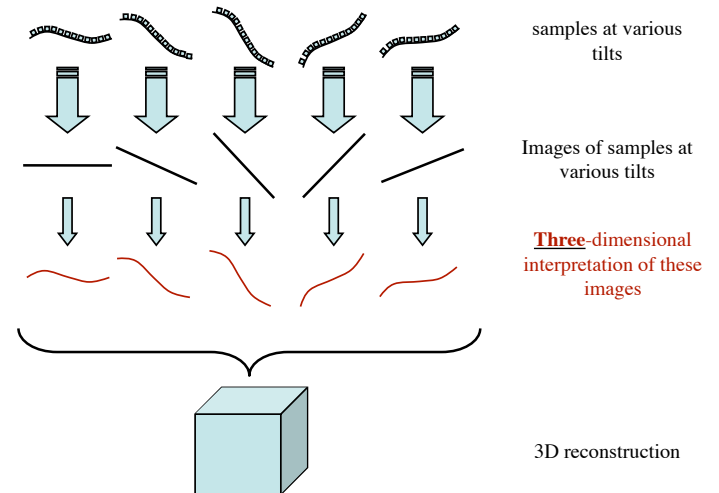
Badly prepared crystals are not flat.



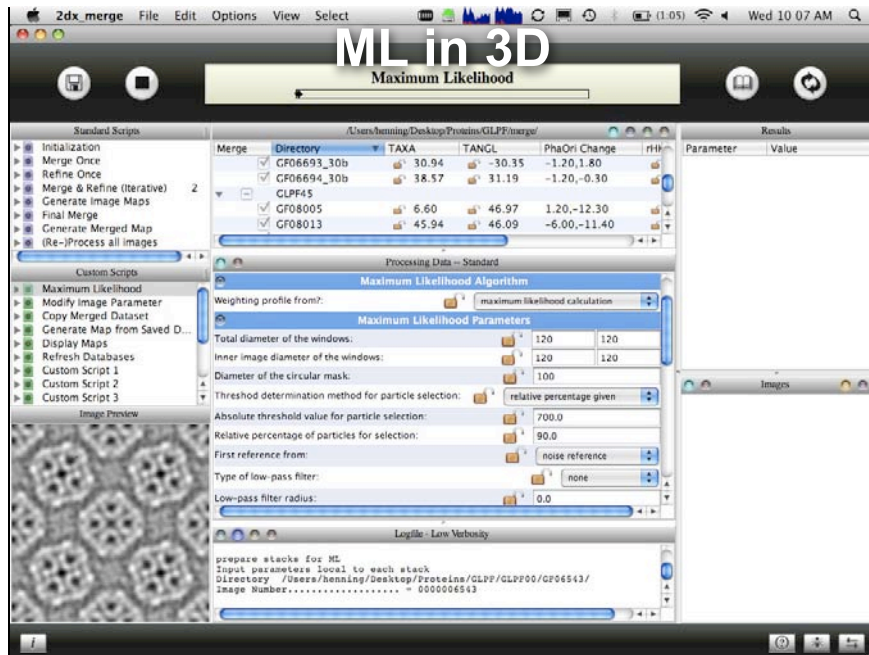
Badly prepared crystals are not flat.  
This leads to resolution loss.



Badly prepared crystals are not flat.  
This should not lead to resolution loss.







**2dx** **ML in 3D** Xiangyan Zeng

**A maximum likelihood approach to 2D crystals**

**here only:**

- 11 micrographs
- \* 10% of unit cells
- \* 120x120 pixels
- \* 4 byte (float)
- = 570 MB

only: 0°... 30°

**Full ML in 3D:**

- 100 micrographs
- \* 100x100 unit cells
- \* 300x300 pixels
- \* 4 byte (float)
- = 360 GB

and: 0°... 70°

**2dx.org** *stahlberglab.ucdavis.edu*

**Expression**  
**Purification**  
**2D Crystallization**  
**Sample Prep**  
**Imaging**  
**Image Processing**  
**Model Building**

**Acknowledgements**

**MloK1:**  
Crina Nimigean

**CIC-ec1:**  
Joe Mindell  
Yoshinori Fujivoshi

**HIO:**  
John Spence

**ML:**  
Niko Grigorieff

**Funding:**  
NSF-Bio CAREER,  
NIH NHLBI R01GM081653,  
NIH NIGMS Protein Structure Initiative.

James Buban  
David Carlson  
Po-Lin Chiu  
Hui-Ting Chou  
Lenin Dominguez  
Sumeet Hayer  
Bryant Gipson  
Mike Nielsen  
Matthew Pagel  
Ludovic Renault