Basic image processing for EM

Carlos Óscar S. Sorzano

Instruct Image Processing Center

CSIC

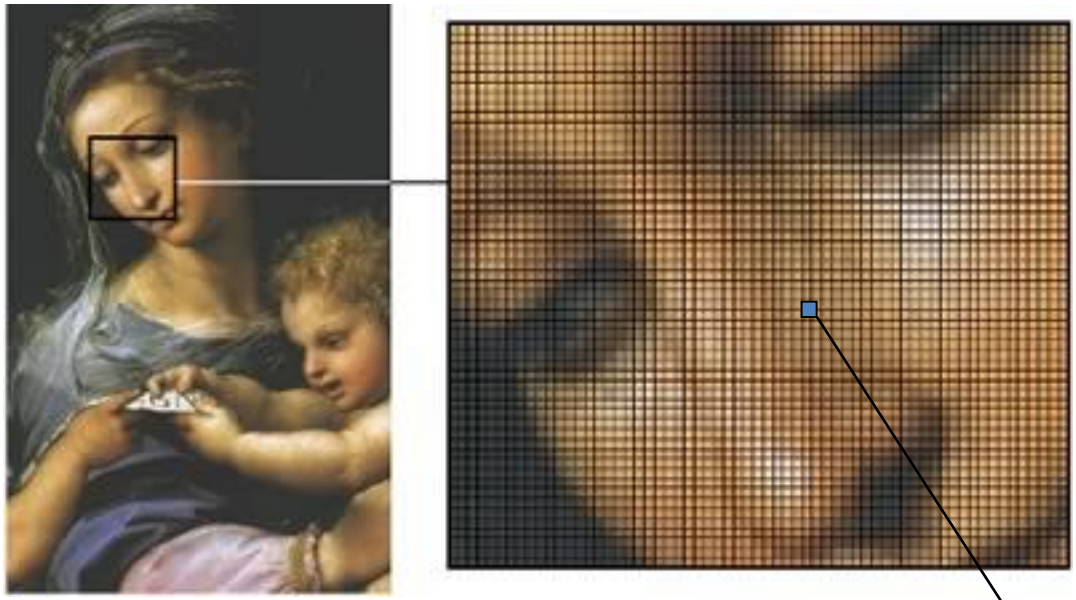
instruct Integrating Biology

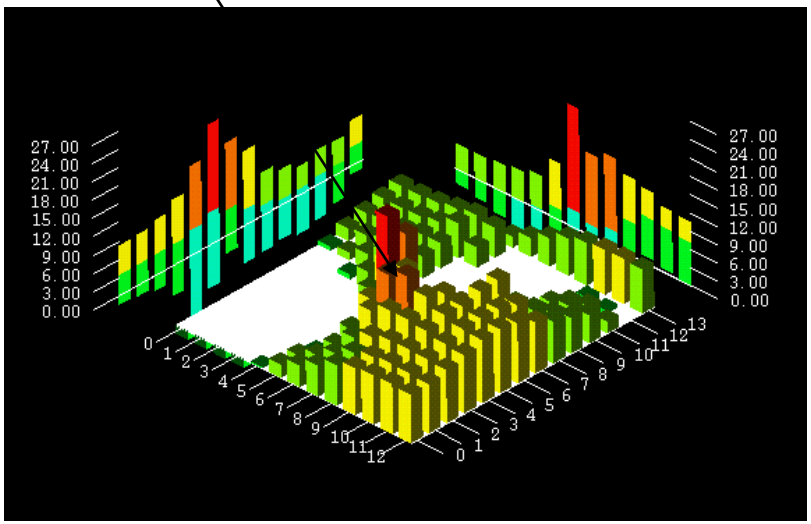
instruct image Processing Center

Basic image processing

- What's an image?
- Image coding
- Sampling
- Basic operations with an image
- Fourier transform

What's an image?





Quantization (bits)

Sampling (pixels)

Image coding

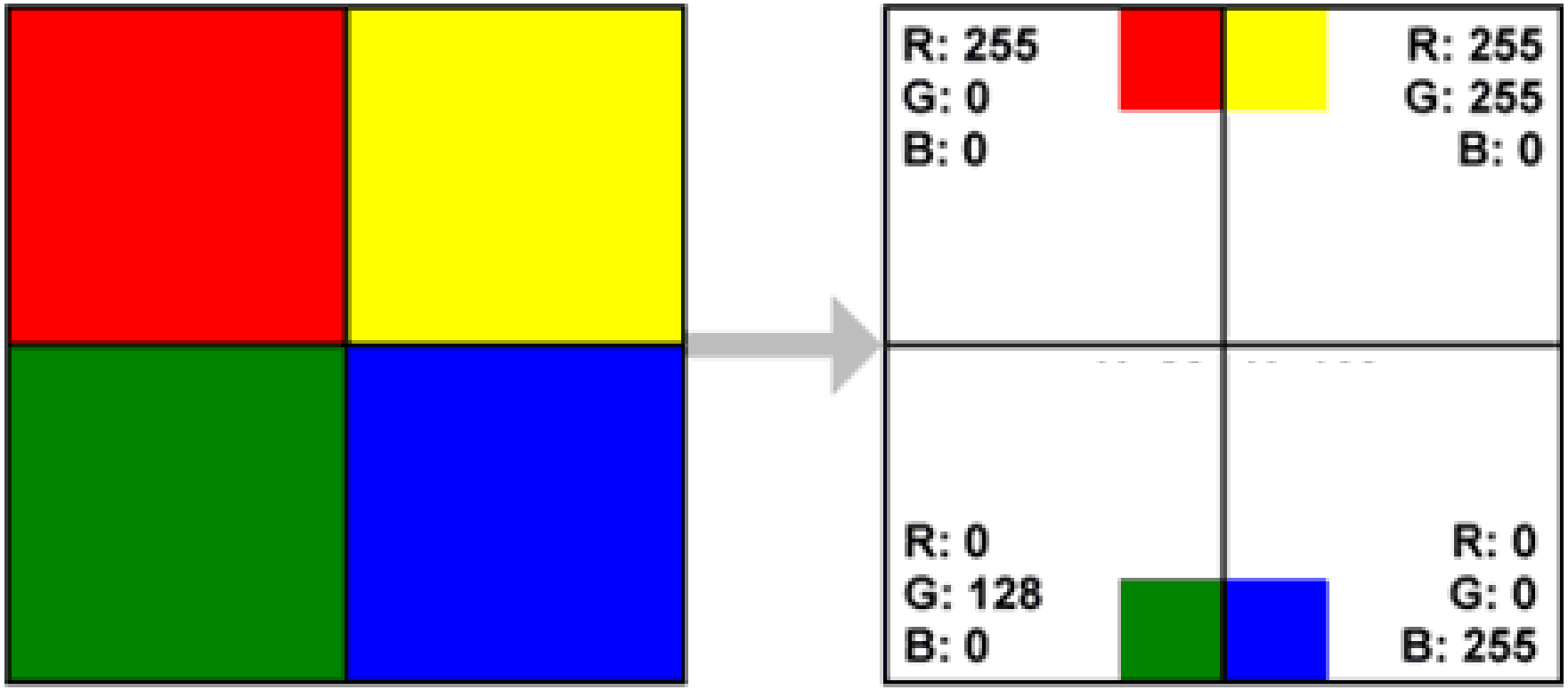
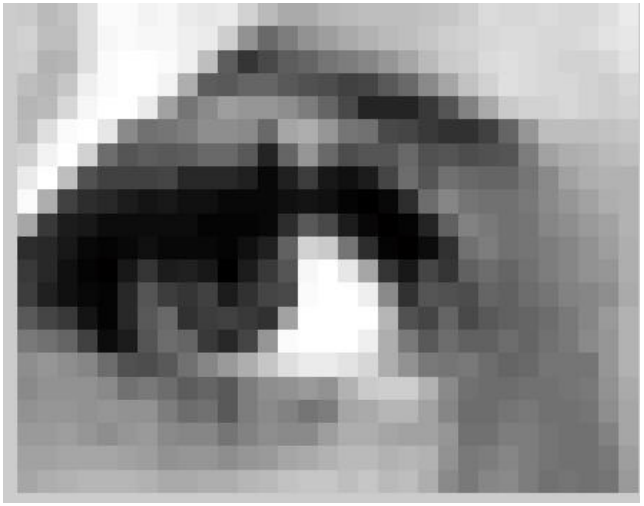


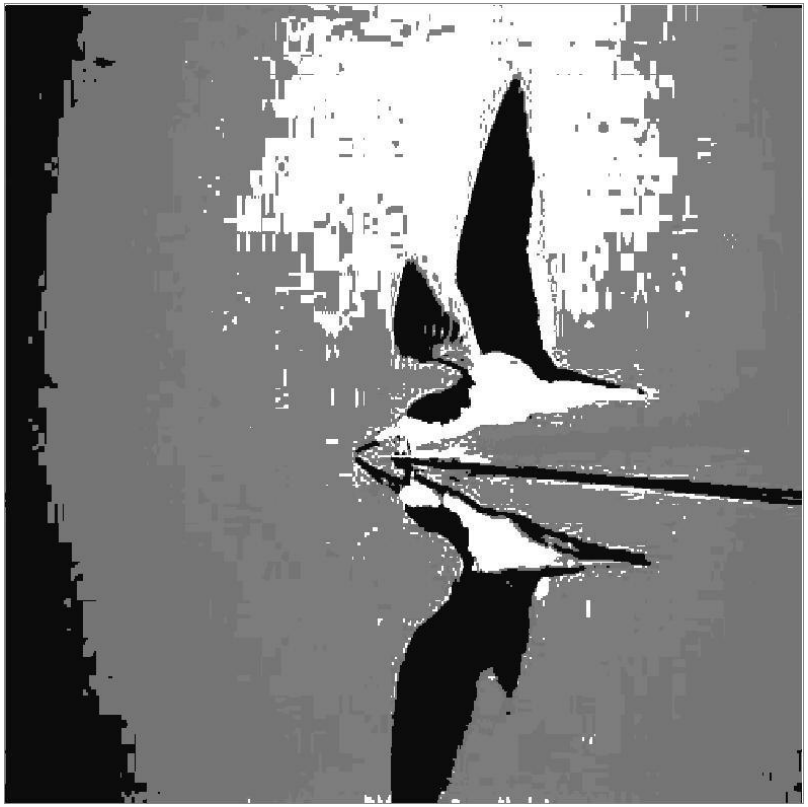
Image coding



202	189	195	221	248	248	240	244	232	228	234	212	189	172	171	182	186
211	191	191	222	247	250	247	245	238	234	205	148	110	128	148	159	171
205	181	184	222	250	255	255	250	238	191	97	58	80	97	106	123	130
192	182	197	237	250	255	255	240	175	97	96	104	104	94	87	84	70
184	188	214	244	253	255	232	163	123	140	159	153	153	143	128	107	77
188	221	250	255	230	139	116	122	128	145	145	139	110	166	175	152	120
209	241	255	211	113	90	97	96	99	110	109	94	43	104	153	110	84
235	255	191	87	70	70	69	73	47	46	47	38	15	41	66	33	25
250	176	70	51	46	35	27	31	24	23	18	17	18	24	30	23	11
166	64	35	31	20	20	20	21	21	14	11	14	18	51	93	77	35
51	46	33	18	12	18	23	18	17	15	11	25	40	103	212	204	122
63	46	24	12	8	33	61	35	40	33	8	37	64	89	244	255	231
66	38	24	17	12	47	90	67	58	71	35	54	74	80	248	255	255
94	76	63	30	14	47	96	109	56	67	69	63	57	163	255	255	255
125	119	107	71	20	37	90	113	106	63	47	74	153	250	255	255	255
143	151	145	135	115	87	84	93	117	135	155	175	199	222	234	234	227
142	148	149	149	139	119	125	115	97	107	94	130	145	148	155	149	181
148	151	151	149	151	148	148	128	112	116	93	129	140	132	122	129	168
156	158	161	151	146	152	158	142	138	135	123	133	143	142	140	159	174
165	163	169	168	166	165	159	149	153	151	146	151	158	168	171	181	186
169	174	181	184	182	182	178	171	172	176	174	182	188	194	197	198	192

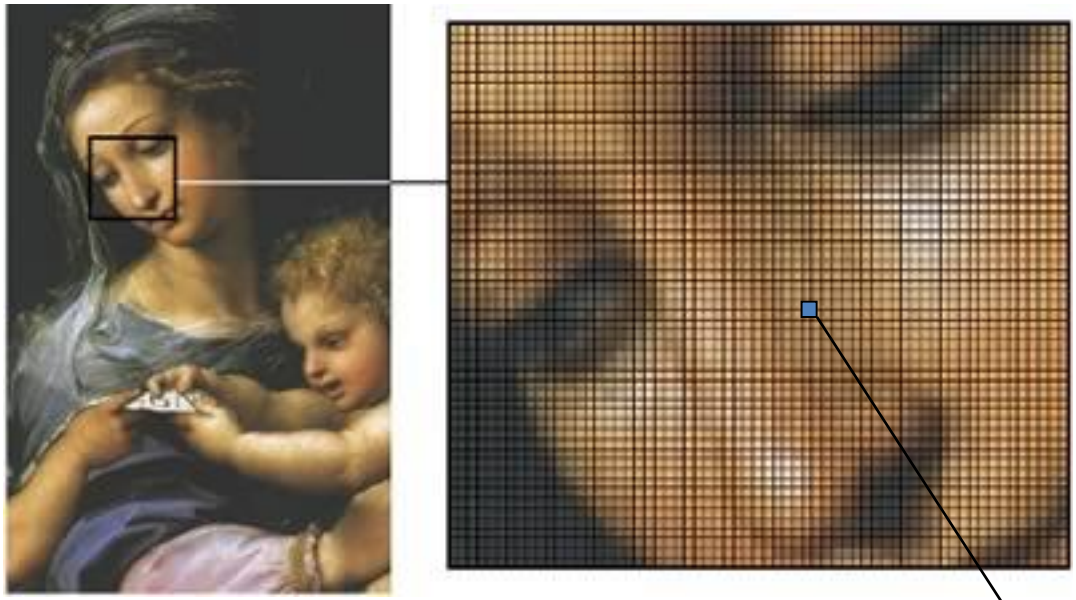
instruct image Processing Center

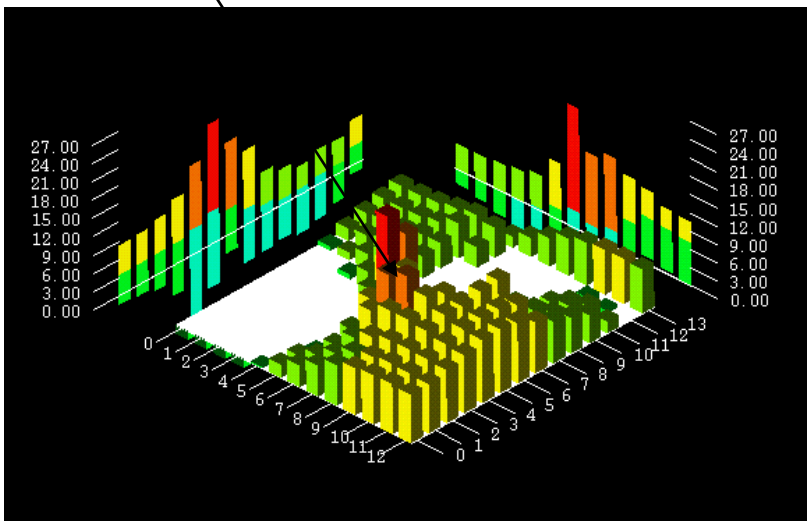
Image coding



Demo

What's an image?

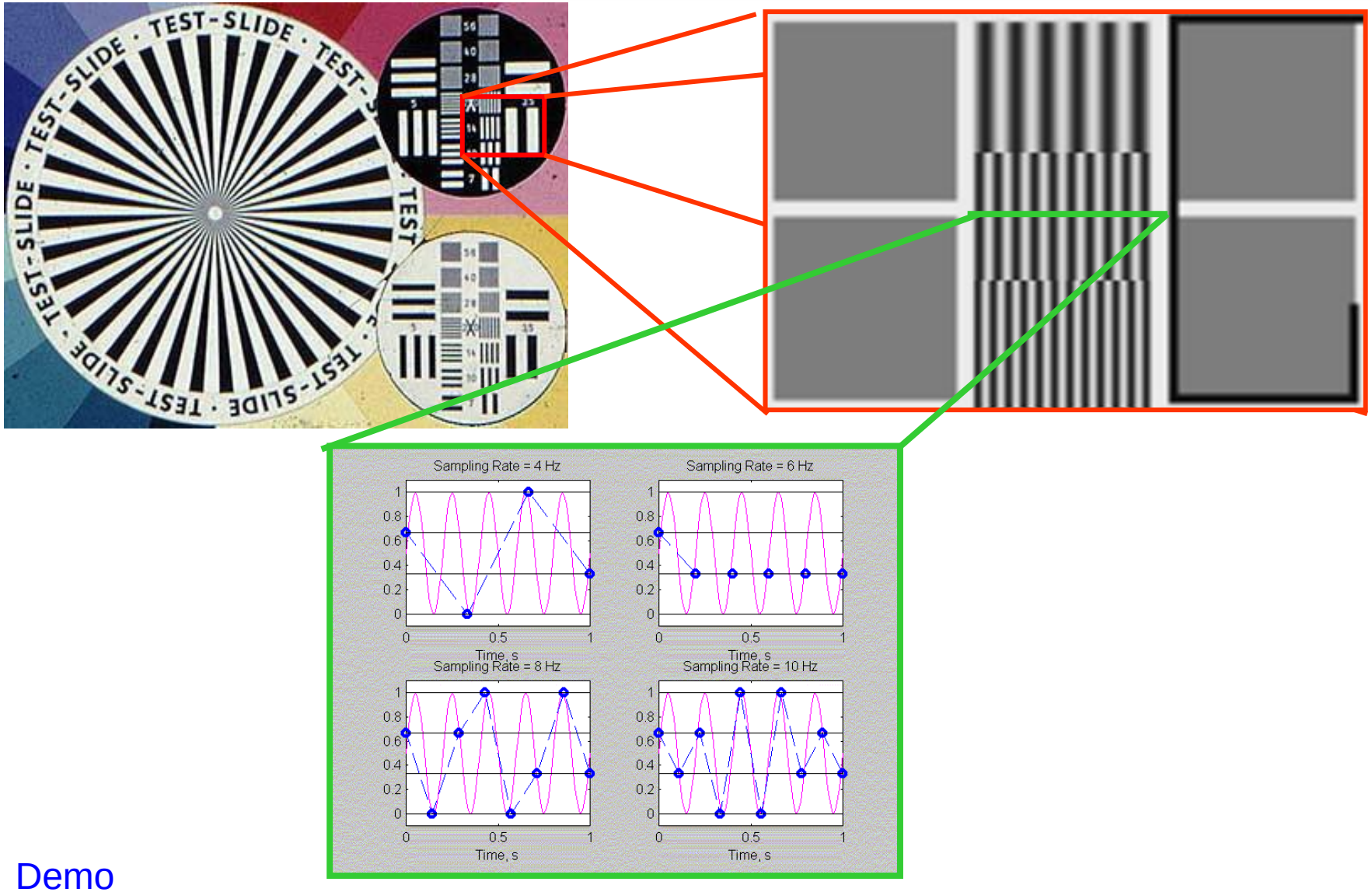




Quantization (bits)

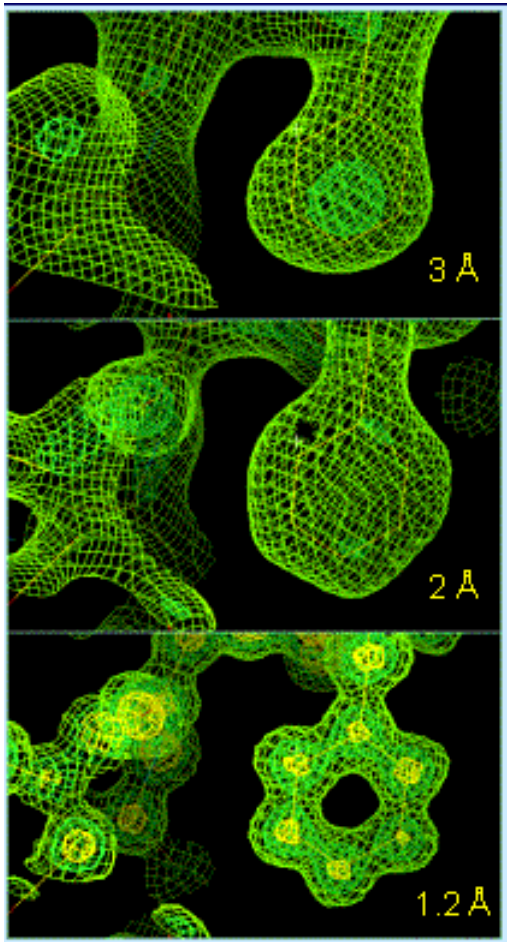
Sampling (pixels)

Spatial sampling



Demo

Pixel size for EM

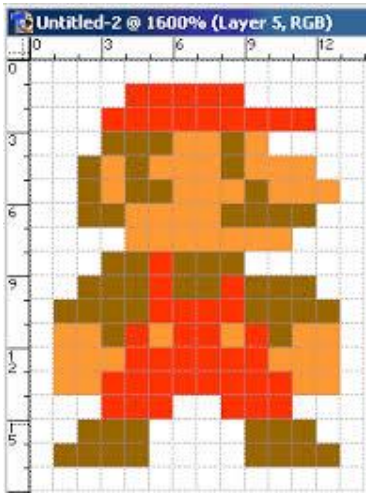


Desired resolution: R (e.g., 4Å)
Nyquist Sampling rate:

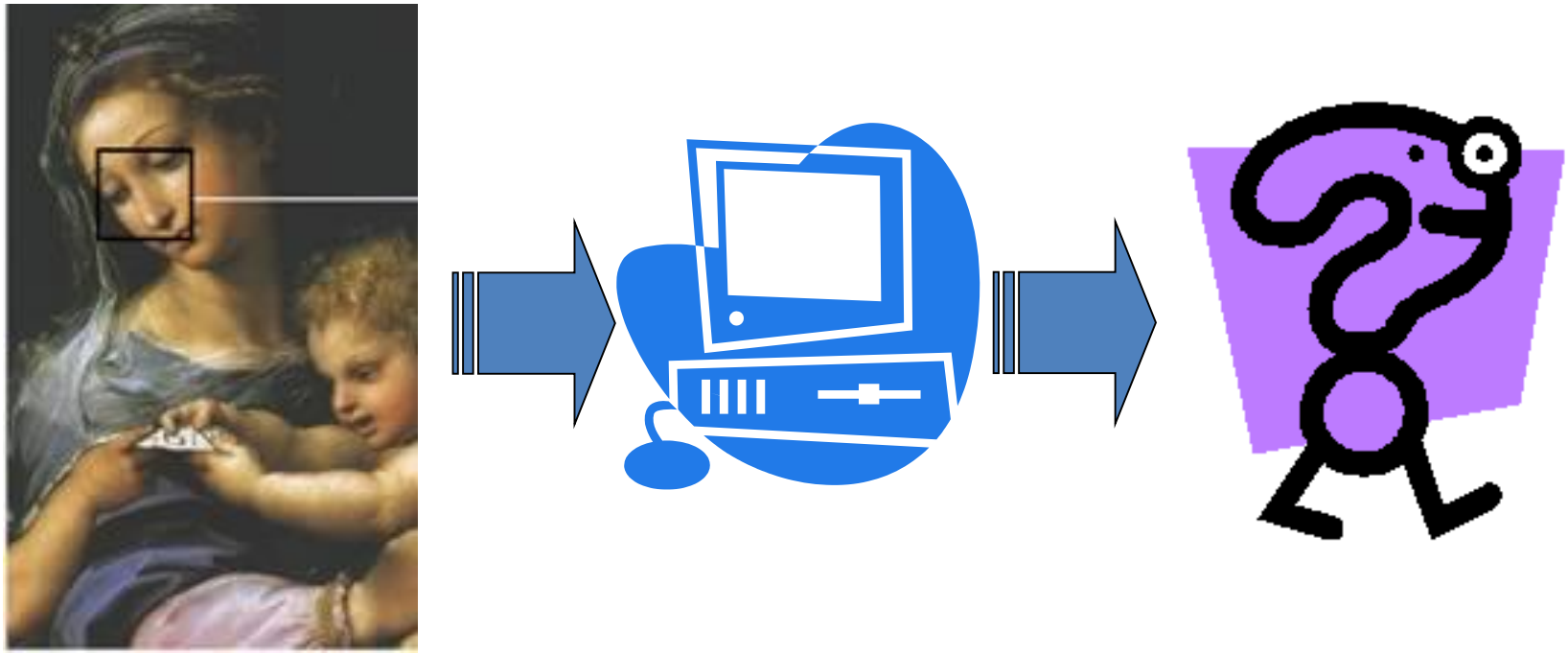
$$T_s < \frac{1}{2} R \quad \text{(e.g., <2Å/pixel)}$$

Typical Sampling rate:

$$T_s < \frac{1}{3} R \quad \text{(e.g., <1.67Å/pixel)}$$



So, what?



CSIC

instruct Integrating Biology

instruct image Processing Center

Image level operations

$$I = k_1 I_1 + k_2 I_2$$

$$I = 0.75 I_1 + 0.25 I_2 \qquad I = 0.5 I_1 + 0.5 I_2 \qquad I = 0.25 I_1 + 0.75 I_2$$

Image level operations

$$\bar{I}(x,y) = \frac{1}{1000}\sum_{i=1}^{1000} I(x,y)$$

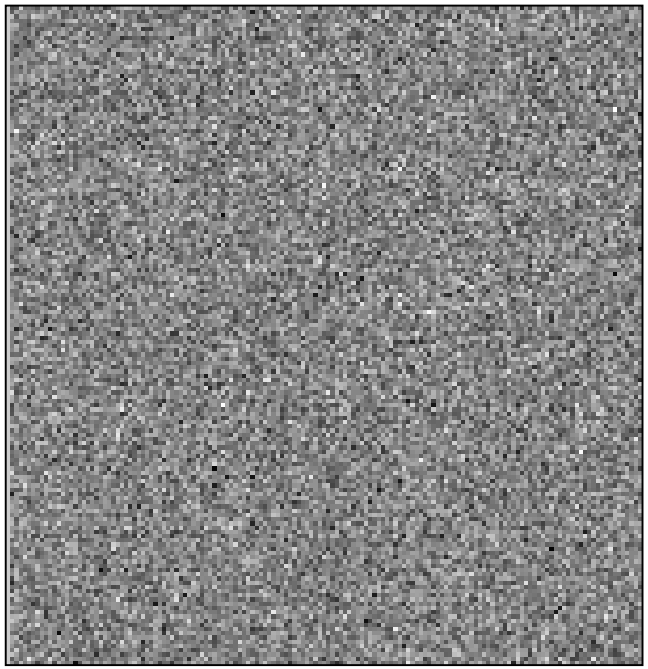
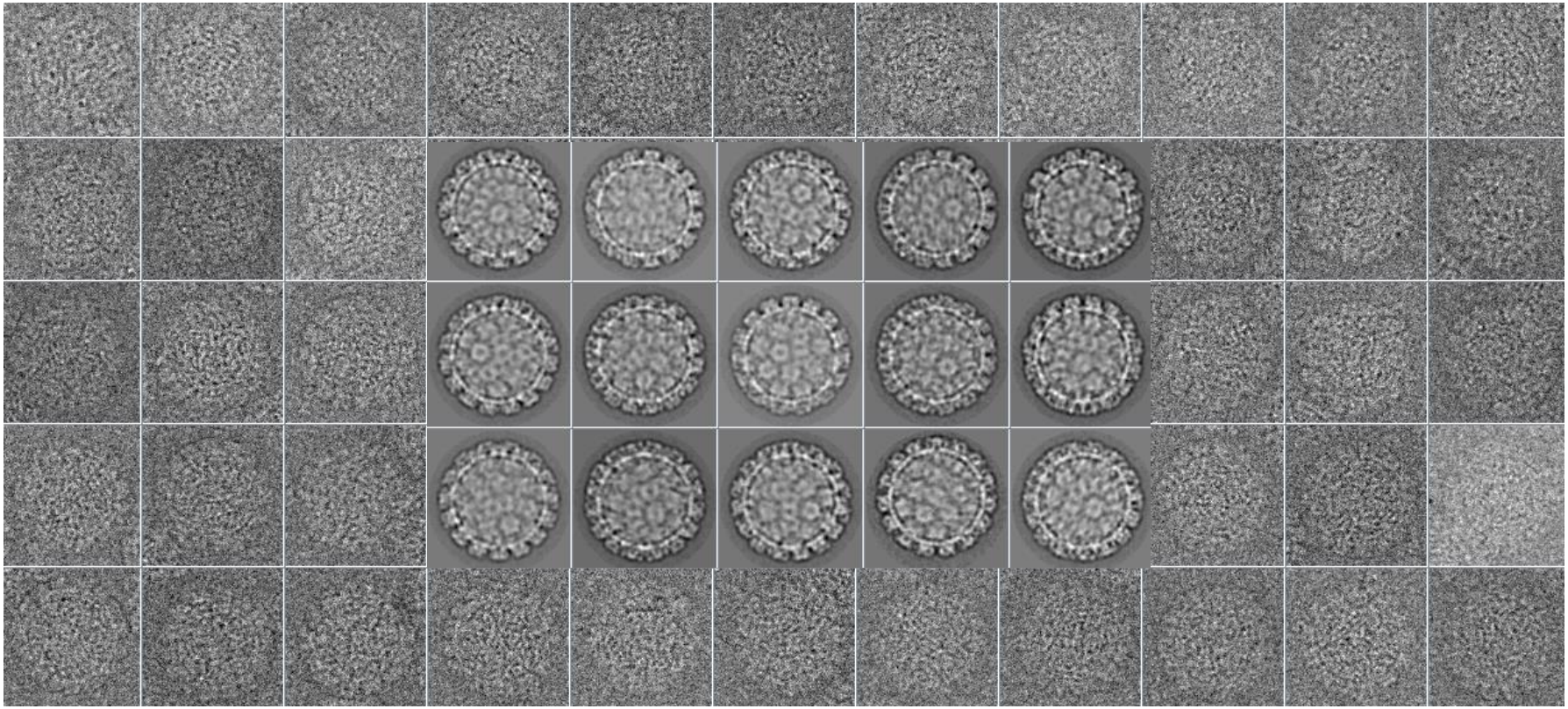
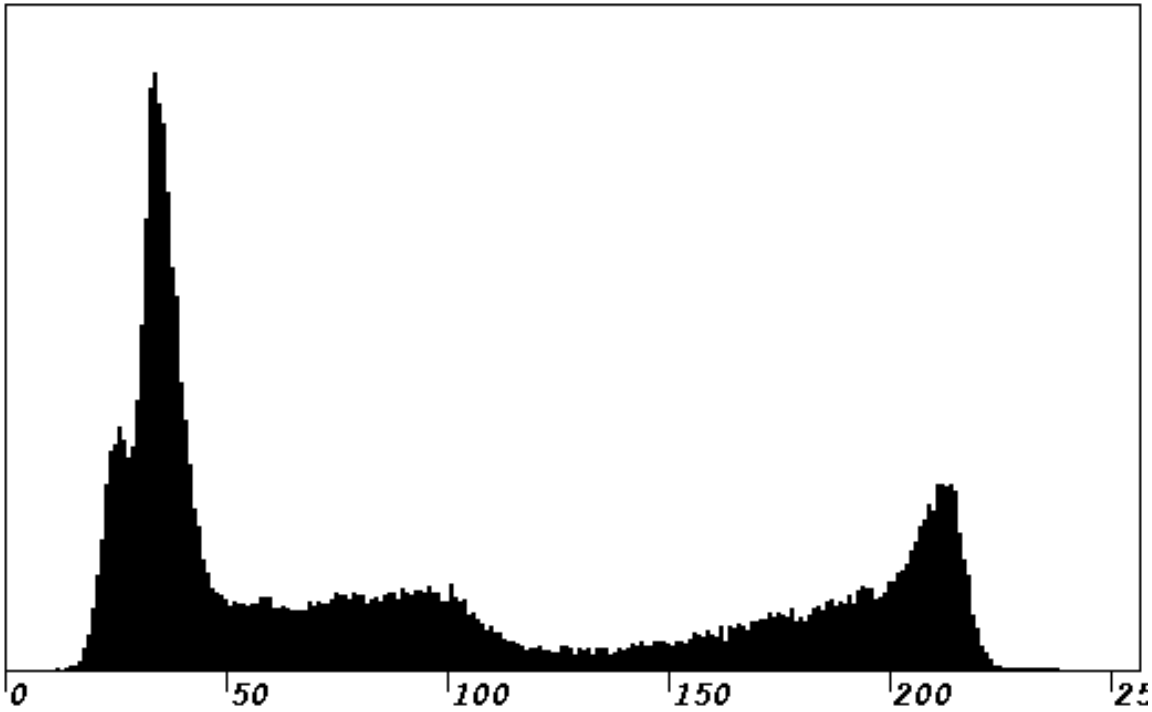


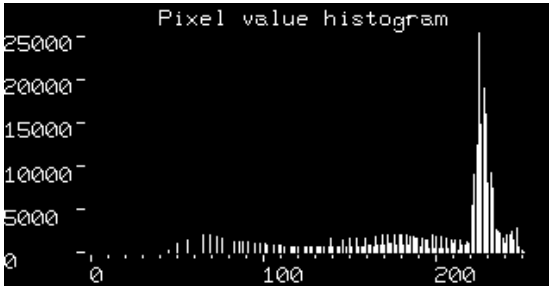
Image level operations

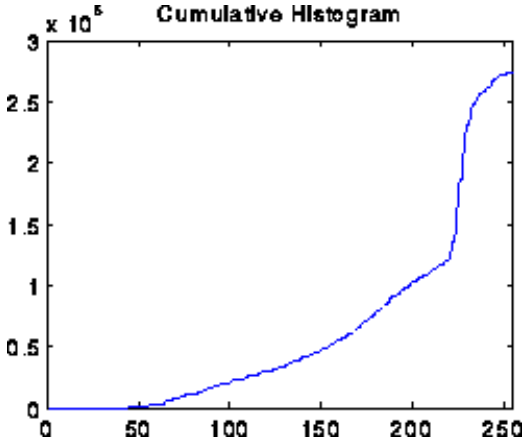


Pixel level operations



Pixel level operations

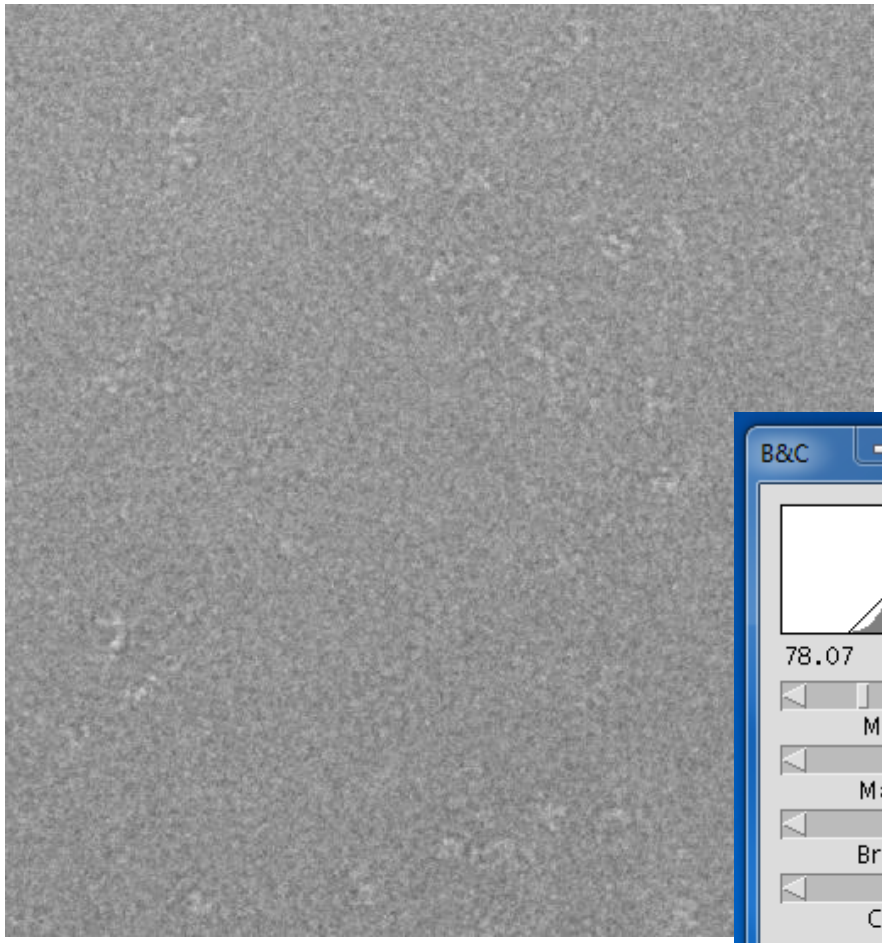


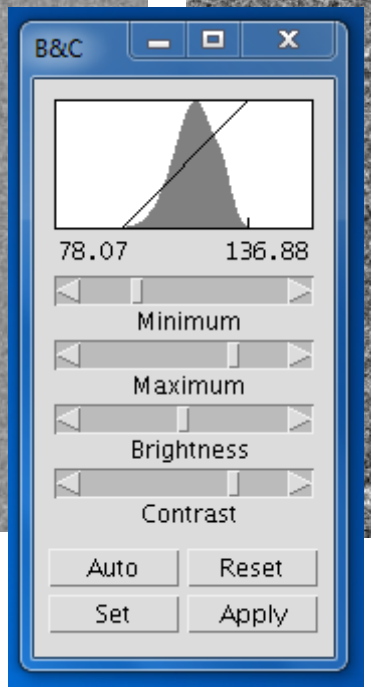


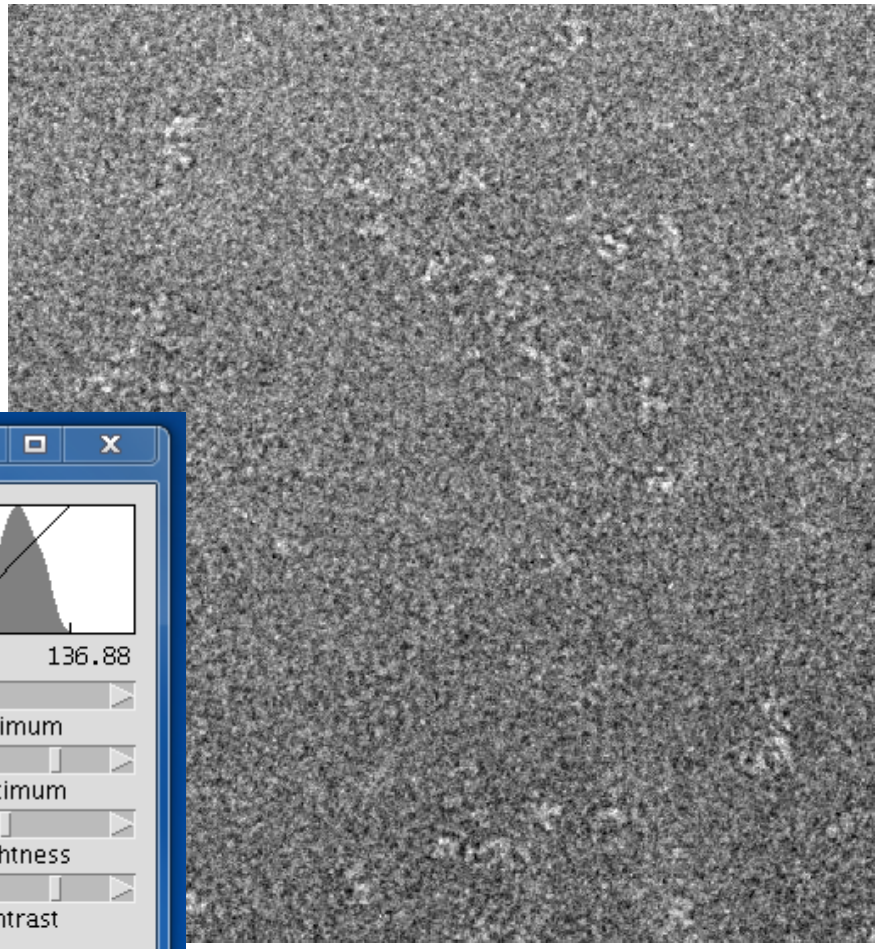
Histogram normalization

Demo

Pixel level operations







Histogram stretching

Pixel level operations

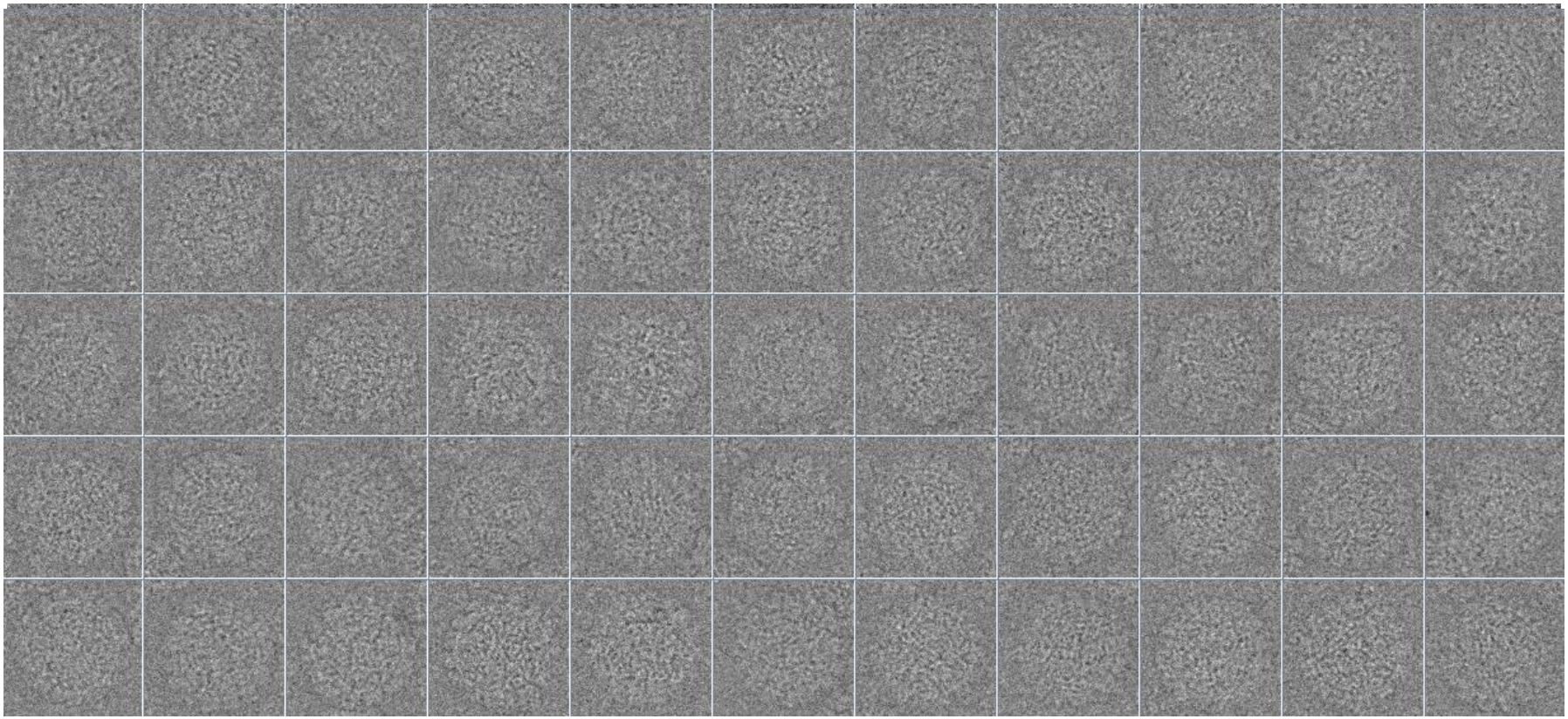
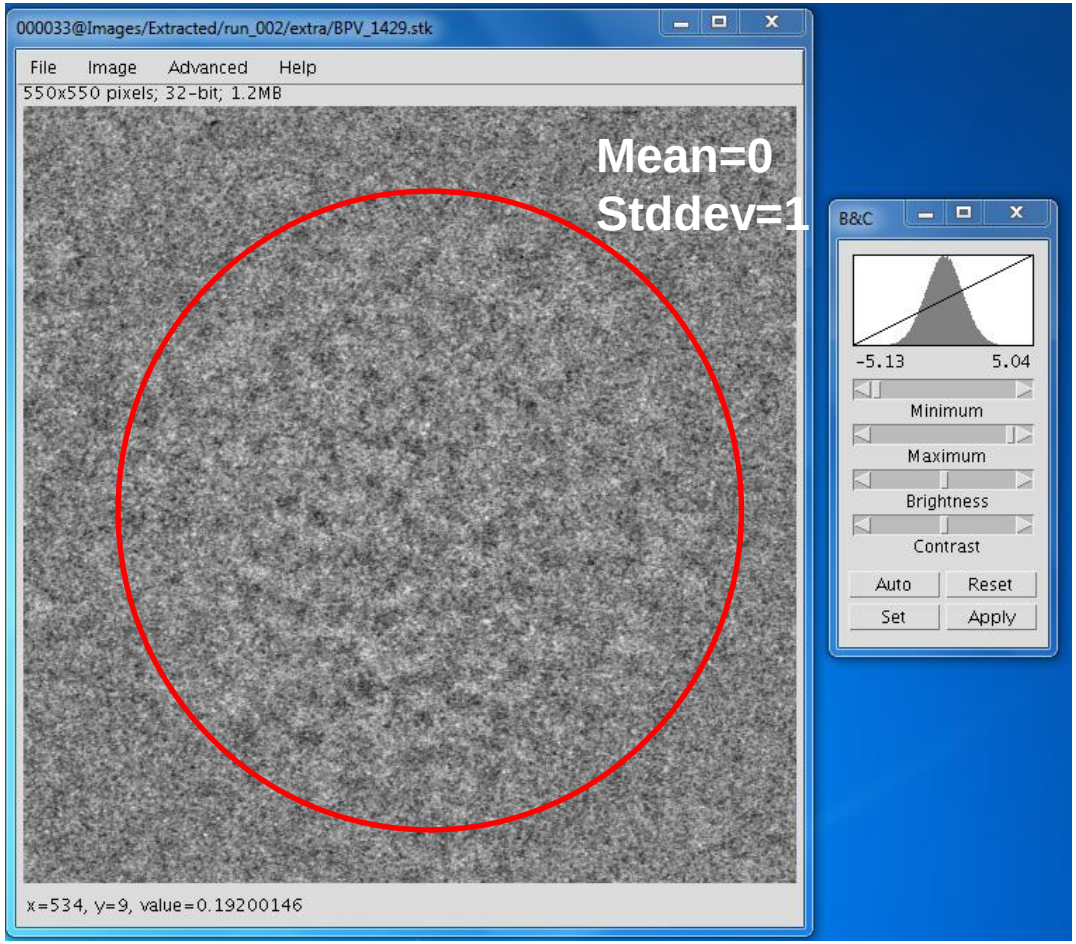


Image normalization

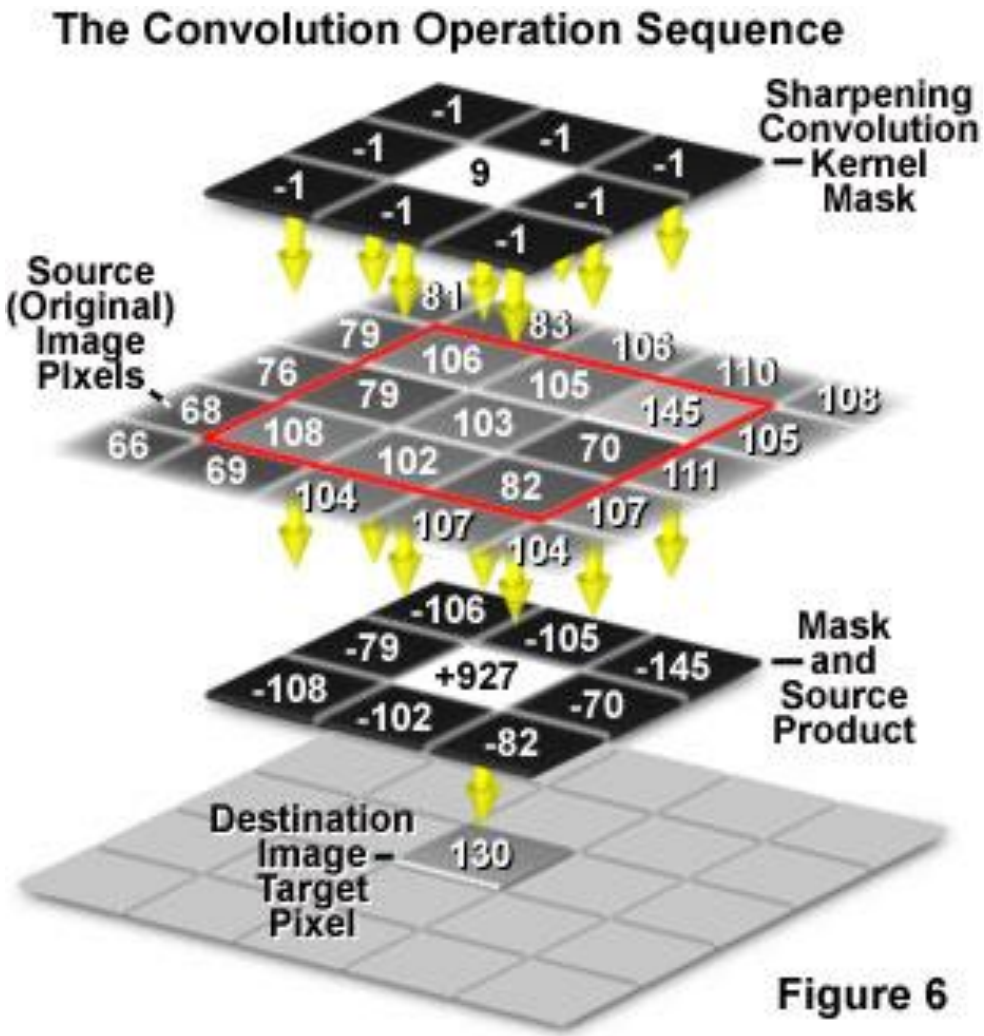
Pixel level operation



Normalization

Sorzano, C. O. S.; de la Fraga, L. G.; Clackdoyle, R. & Carazo, J. M.
Normalizing projection images: A study of image normalizing procedures for single particle three-dimensional electron microscopy
Ultramicroscopy, **2004**, *101*, 129-138

Group level operations

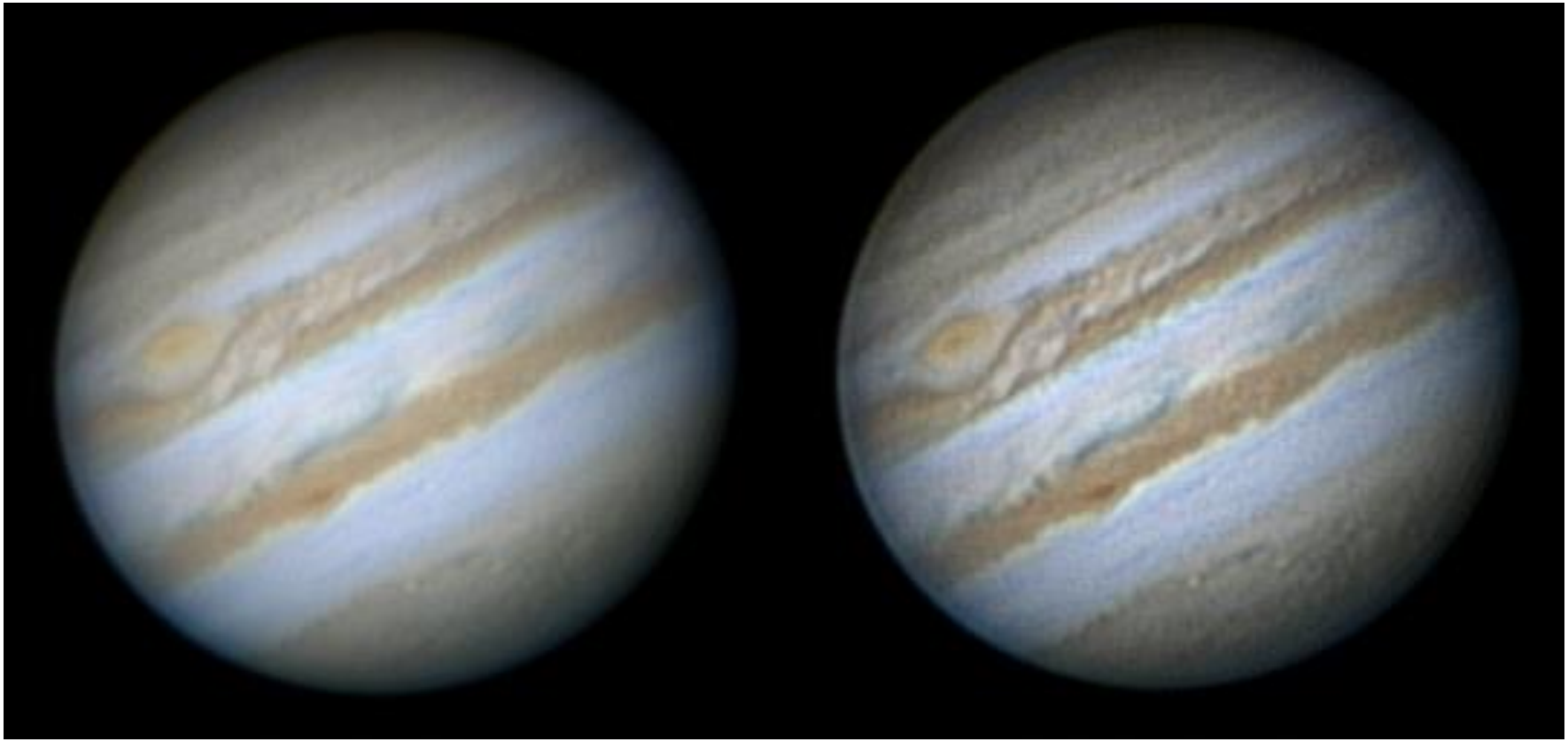


Demo

Group level operations

FIGURE 10.13 Convolution illustrated. The observed image (*left*) is the convolution of the true object (*center*) with the PSF (*right*).

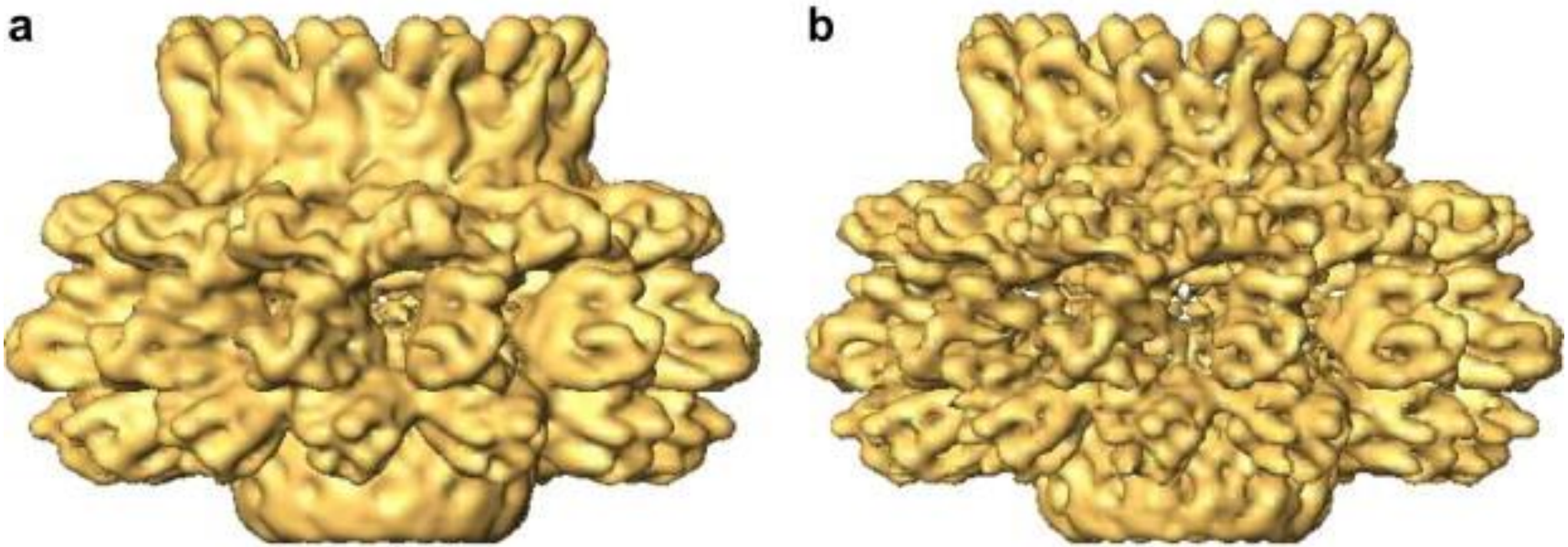
Group level operations



Sharpening

Demo

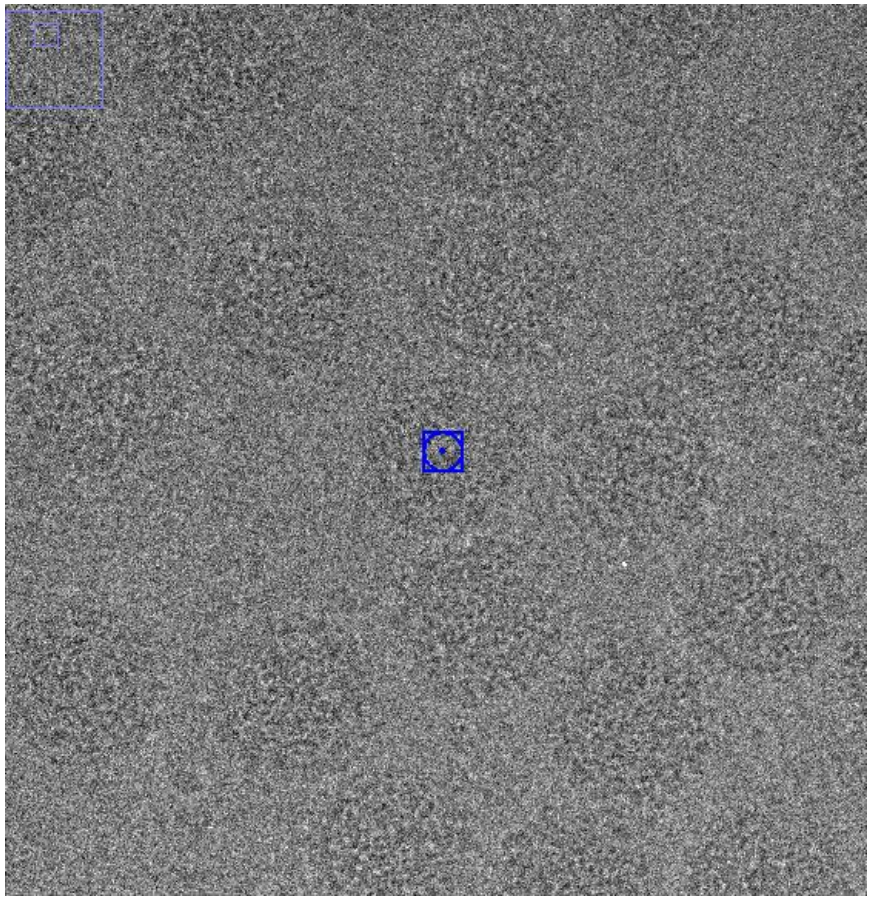
Group level operations

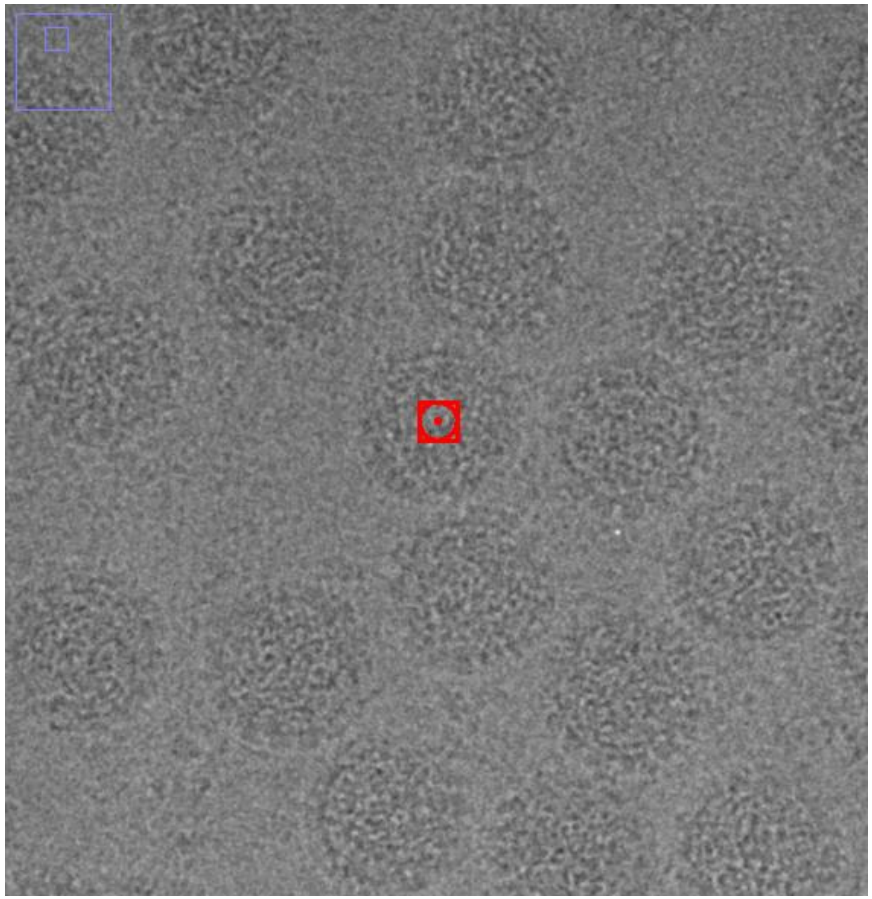


Fernández, J. J.; Luque, D.; Castón, J. R. & Carrascosa, J. L.
Sharpening high resolution information in single particle electron cryomicroscopy.
J Struct Biol, **2008***, 164*, 170-175

Sharpening

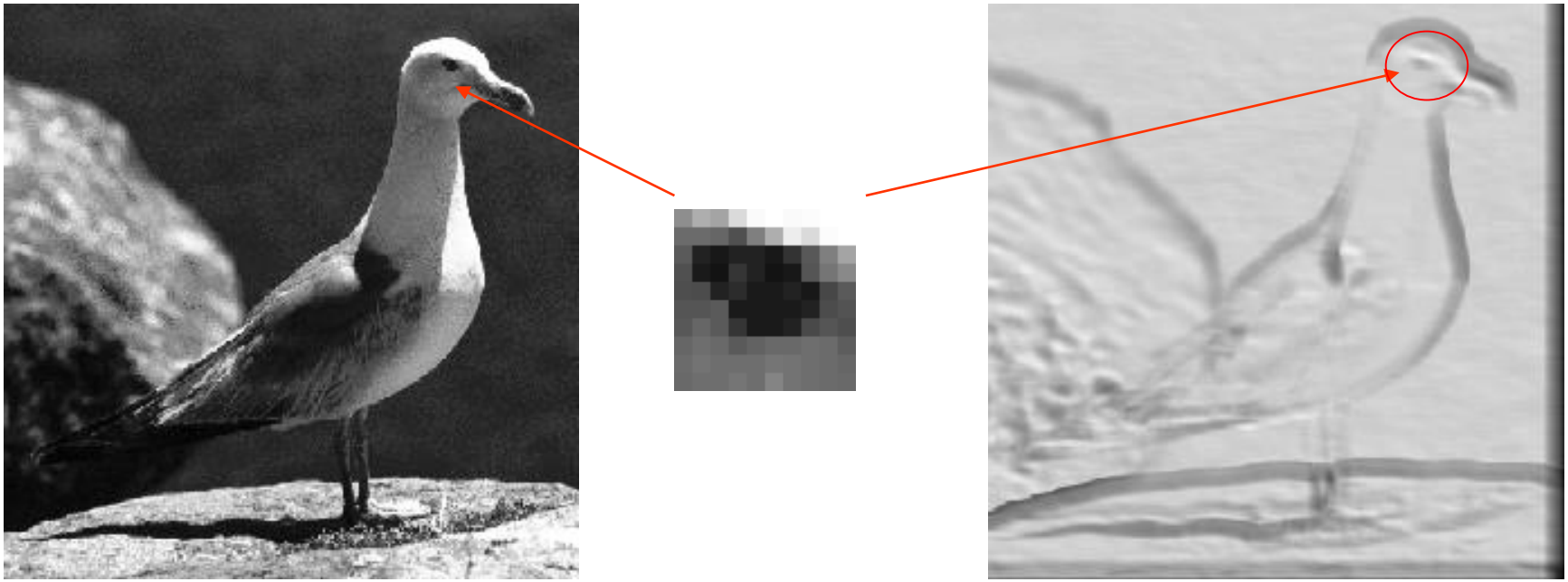
Group level operations





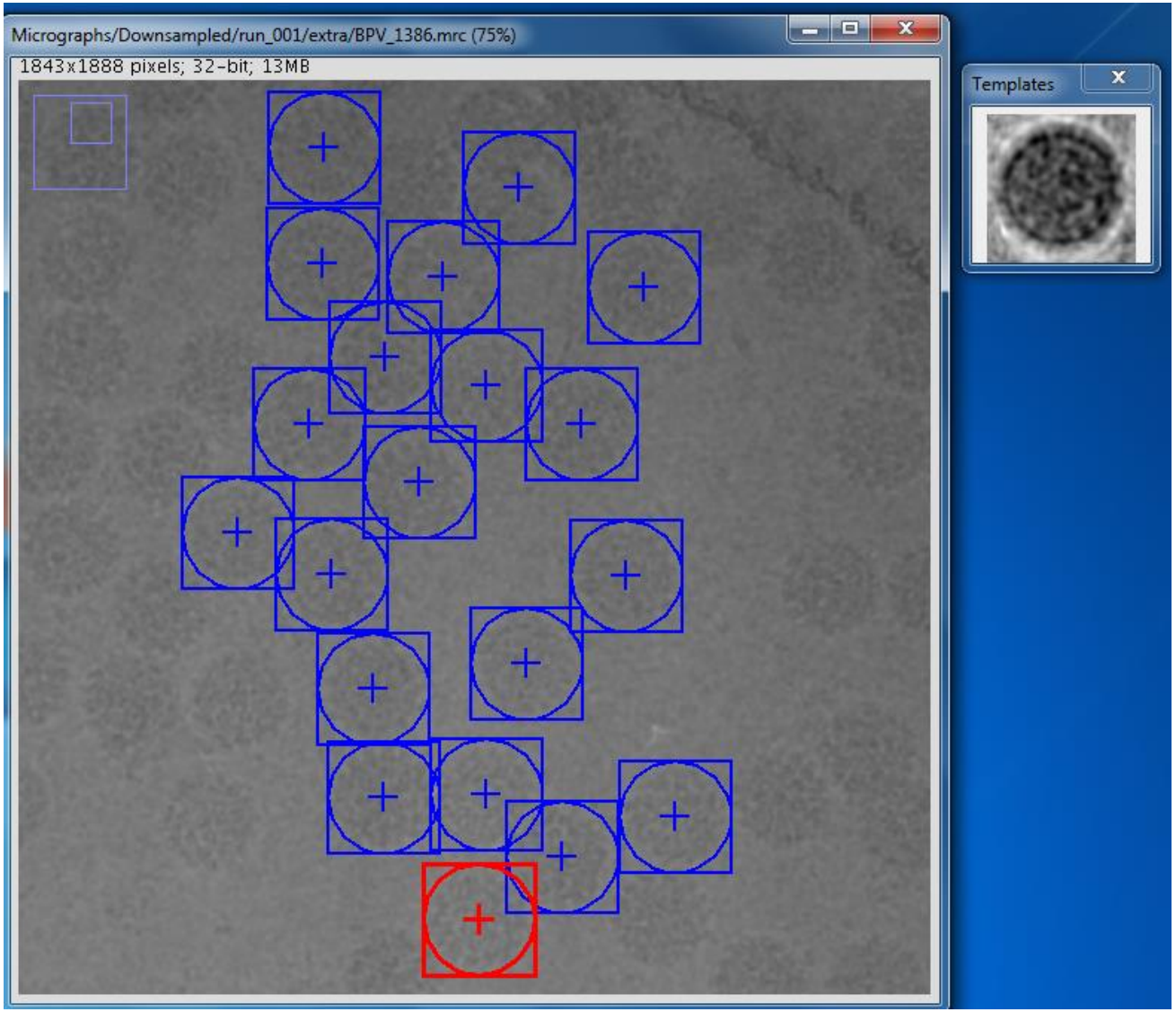
Smoothing

Group level operations



Correlation

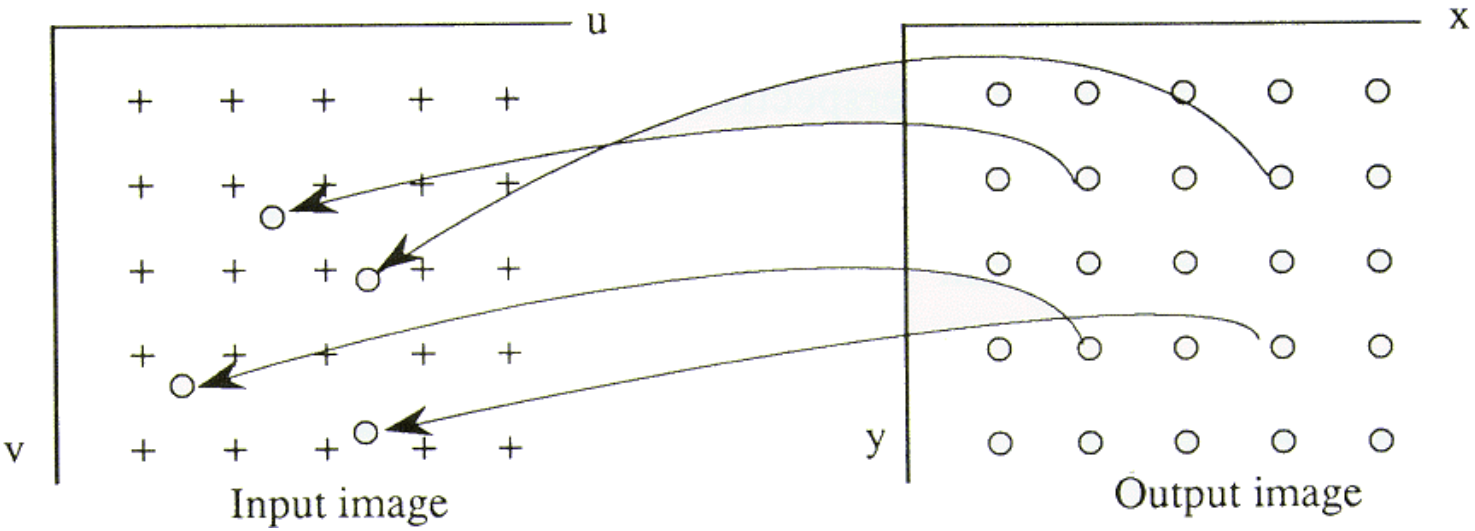
Group level operations



Correlation

Geometric transformations

Geometric transformations



Interpolation

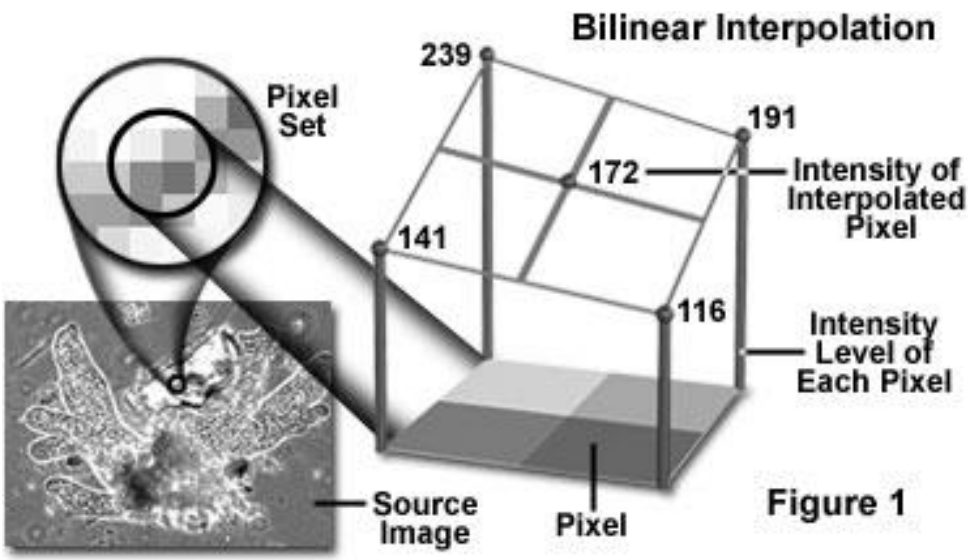


Figure 1

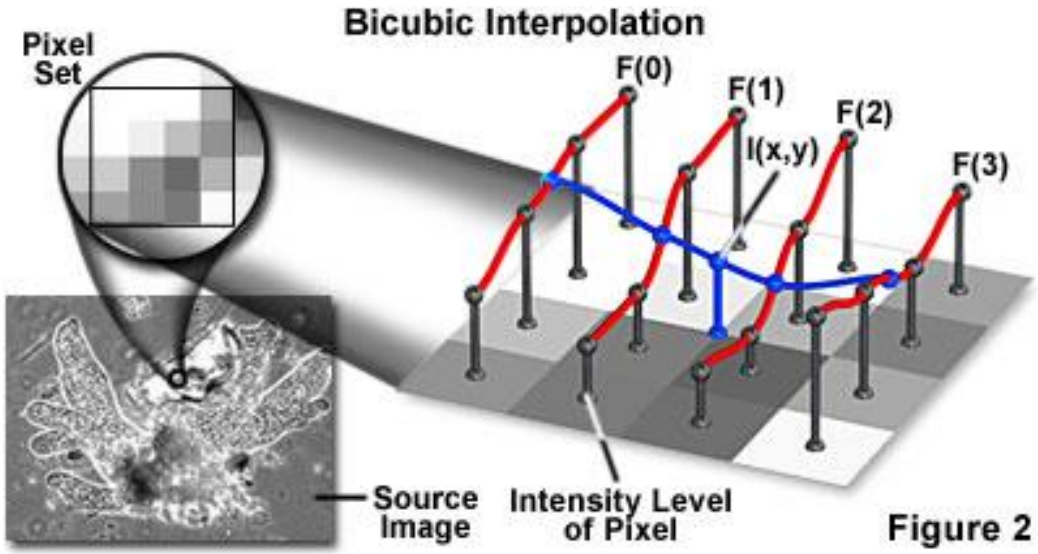
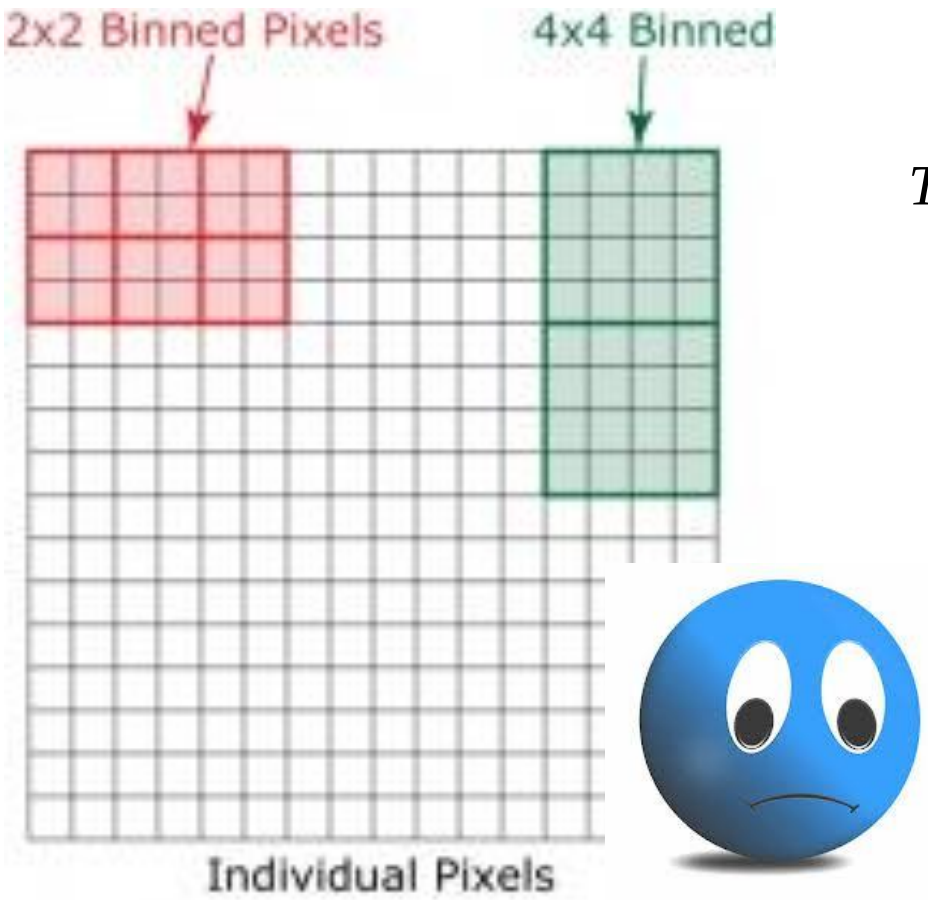


Figure 2

Demo

Downsampling/Binning



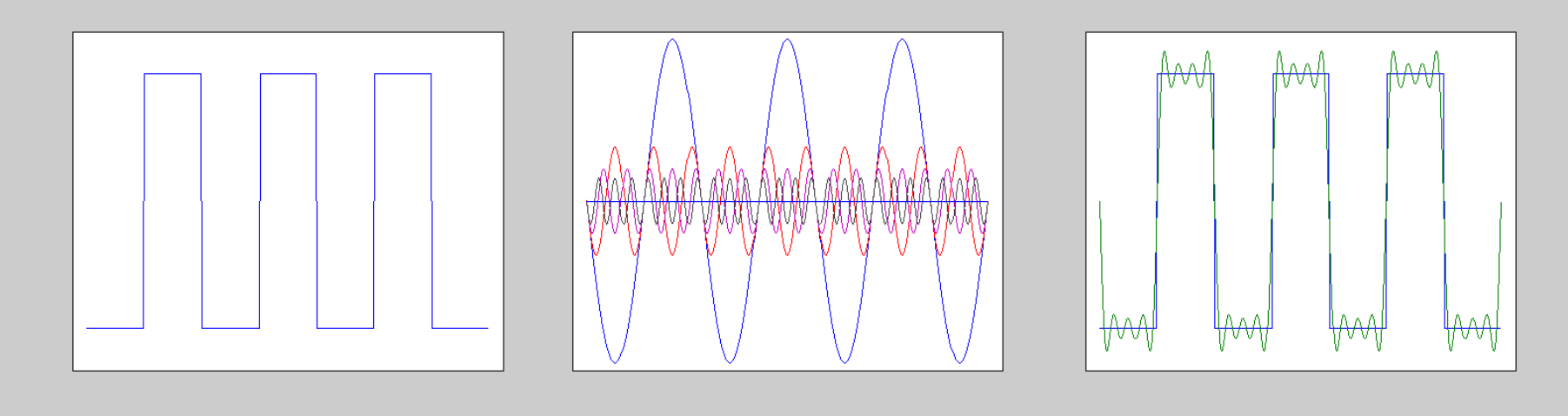
$$T_s^{'} = kT_s$$

Remind Nyquist $R > 2T_s^{'}$

Sorzano, C. O. S.; Iriarte-Ruiz, A.; Marabini, R. & Carazo, J. M.
Effects of the downsampling scheme on three-dimensional electron microscopy of single particles
Proc. of IEEE Workshop on Intelligent Signal Processing, **2009**

Fourier Transform

Original wave | Decomposition | Approximation



Demo Sine waves

Demo Fourier Transform 1D

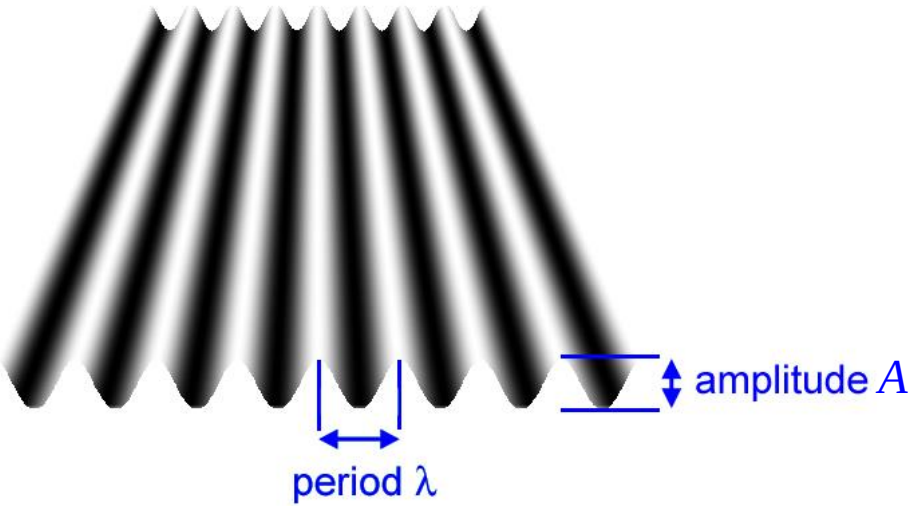
Demo Fourier Transform 2D

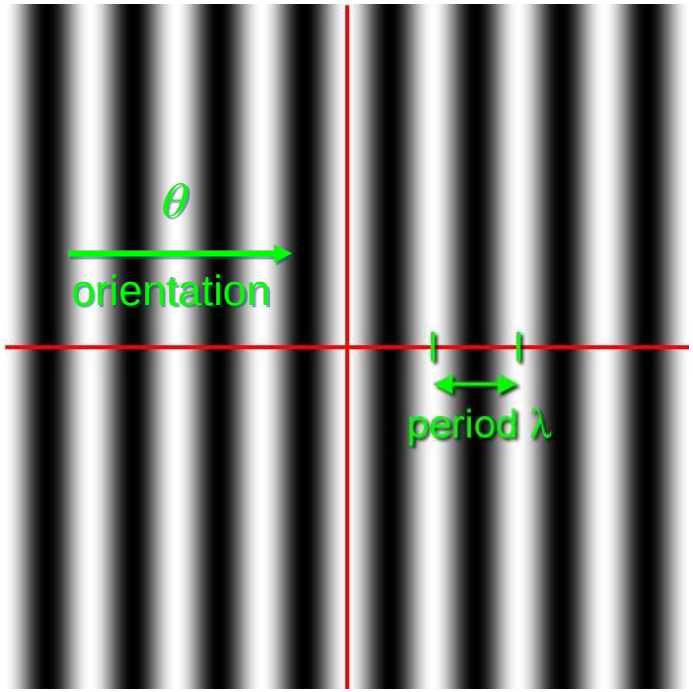
2D Sinusoids

2D Sinusoids:

$$\mathbf{I}(r,c) = \frac{A}{2}\left\{\cos\left[\frac{2p}{l}(r \times \sin\theta + c \times \cos\theta) + j\right] + 1\right\}$$

... are plane waves with grayscale amplitudes, periods in terms of lengths, ...



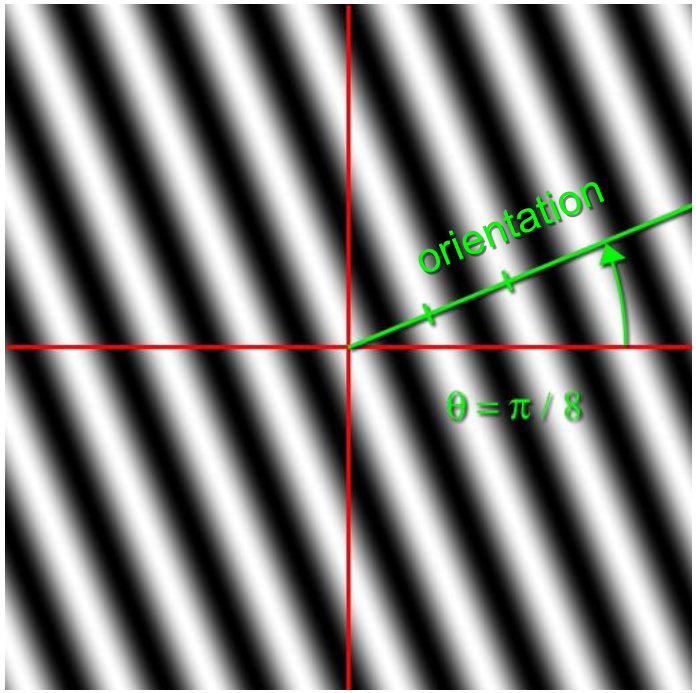


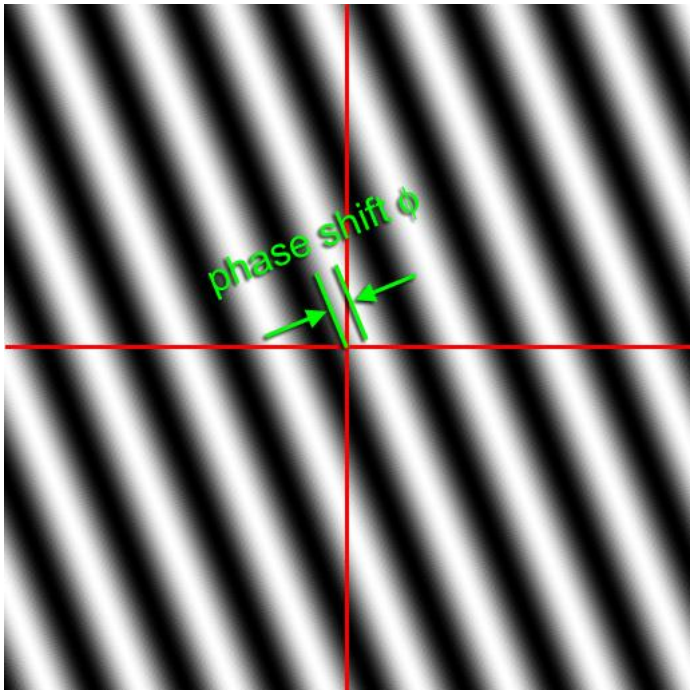
ϕ = phase shift

2D Sinusoids

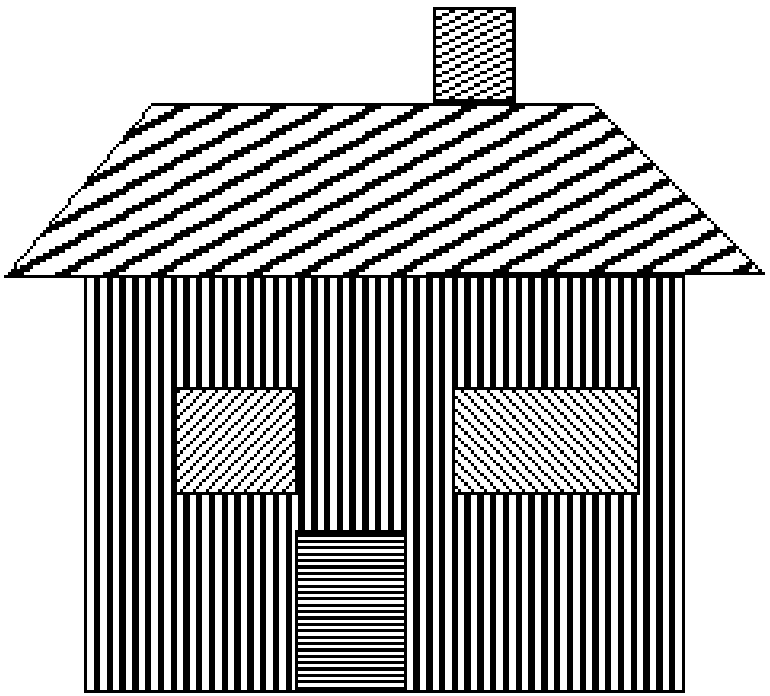
2D Sinusoids:

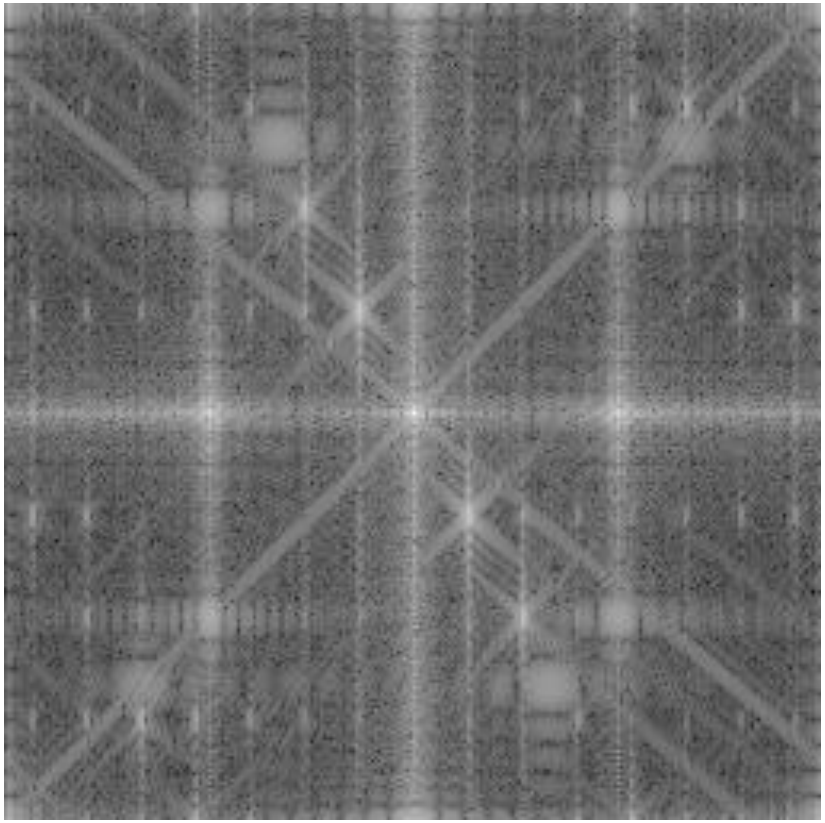
... specific orientations, and phase shifts.



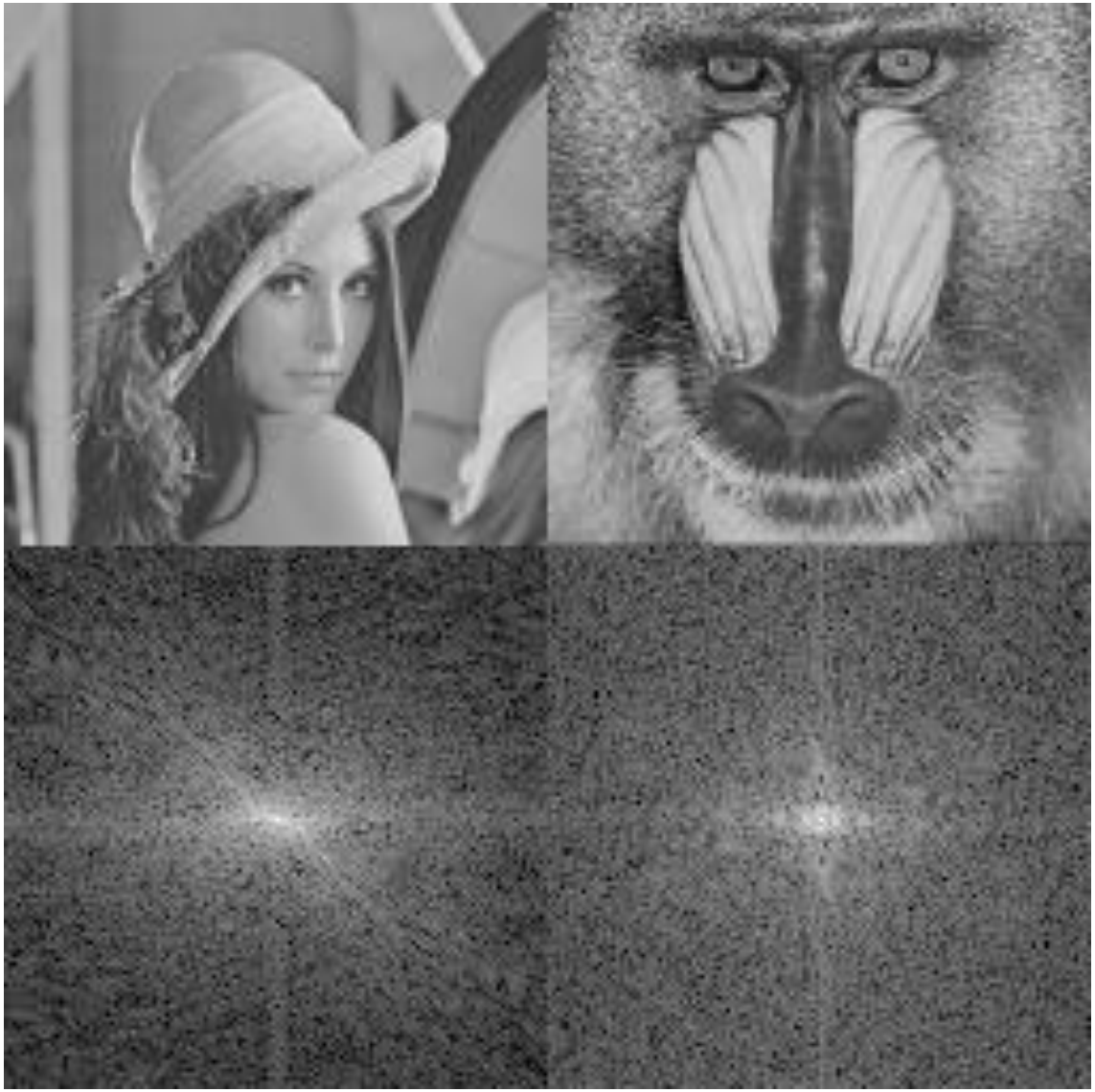


Fourier Transform

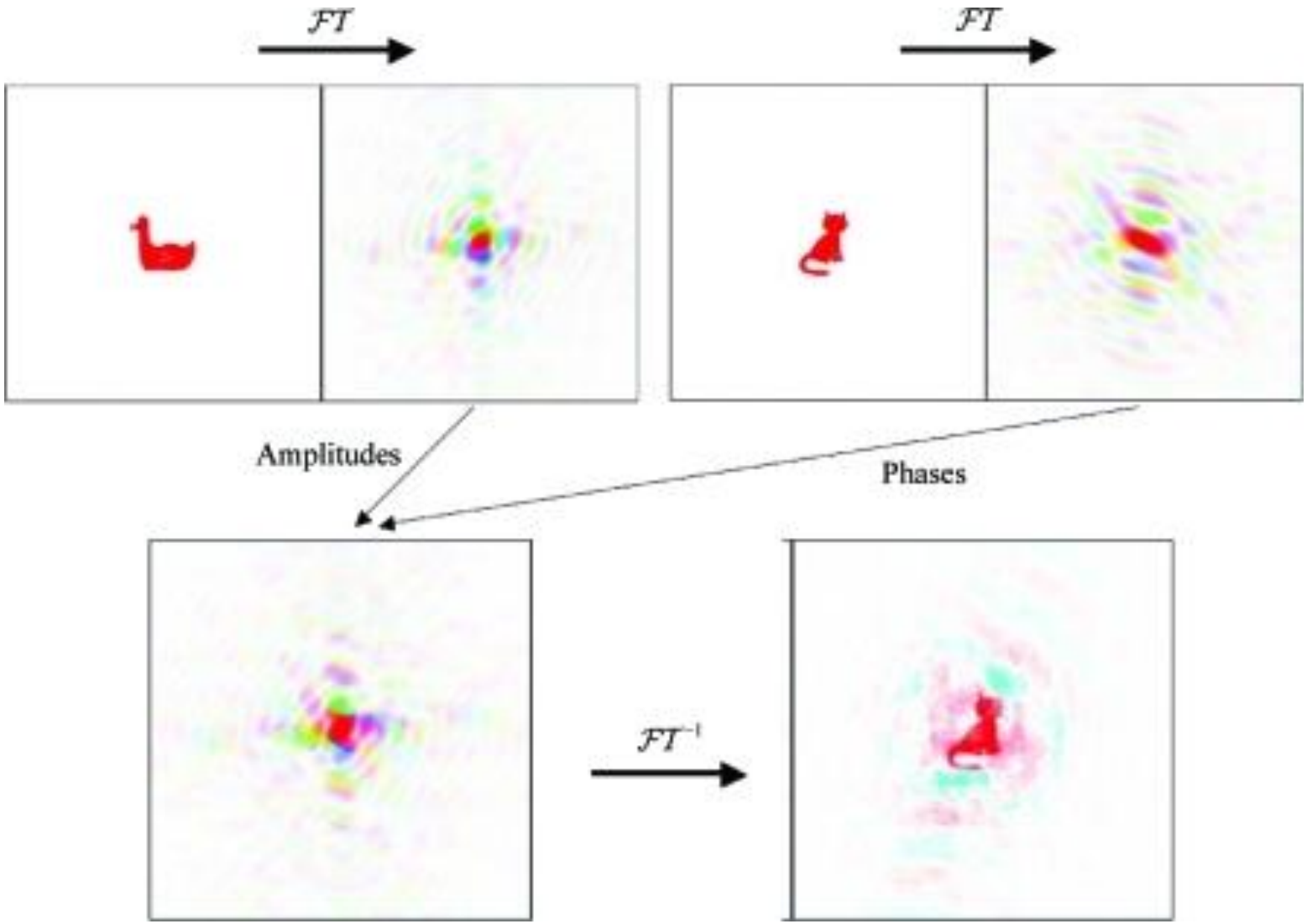




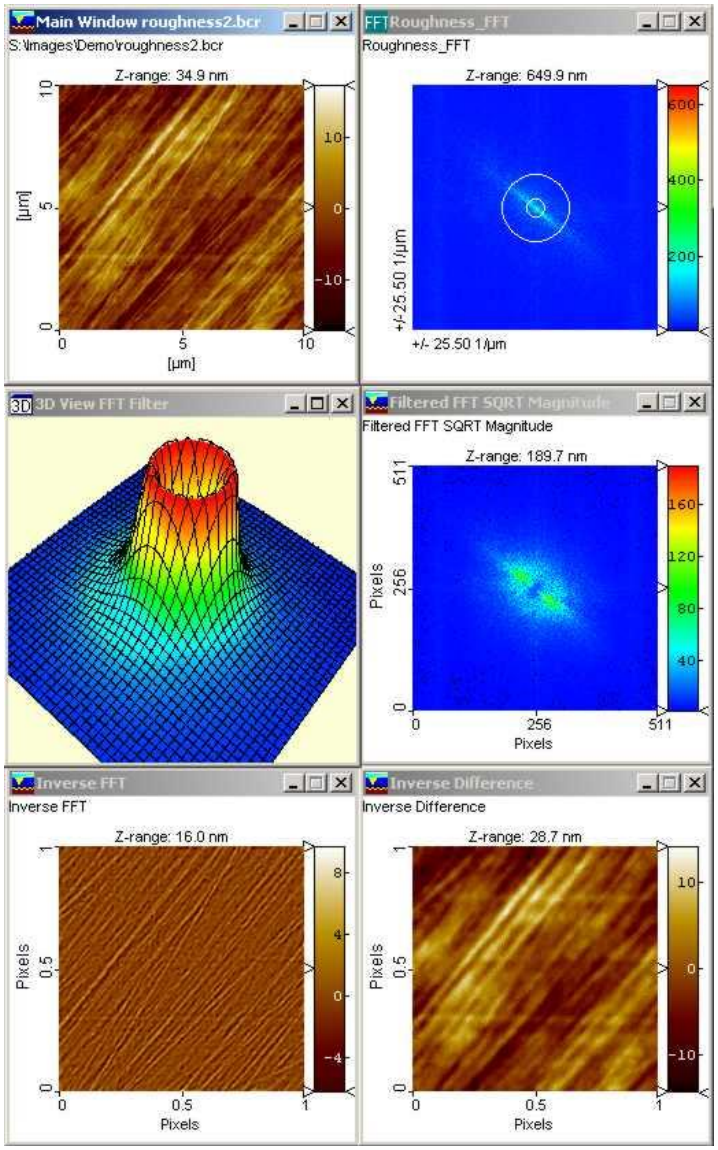
Fourier Transform

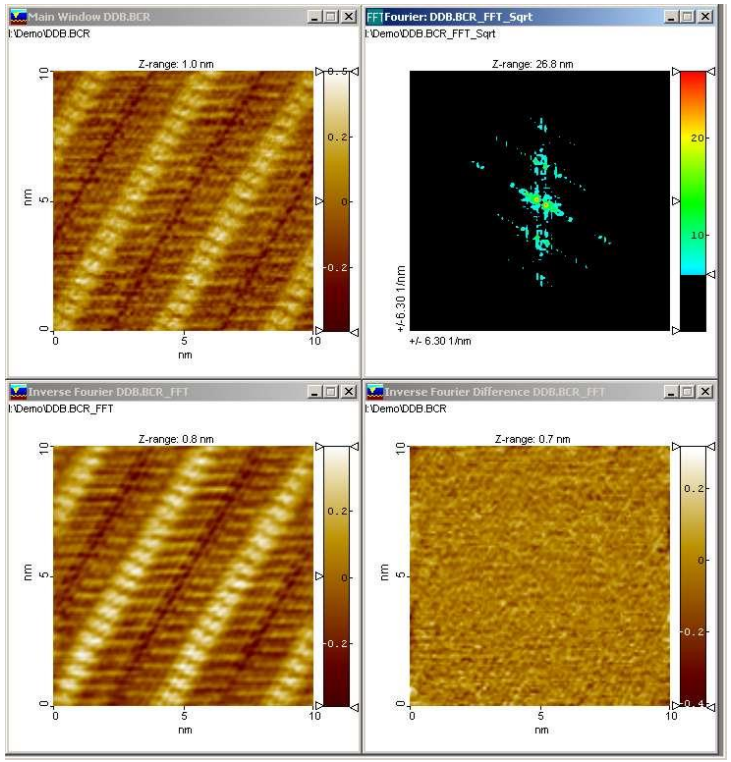


Fourier Transform



Filters in Fourier space



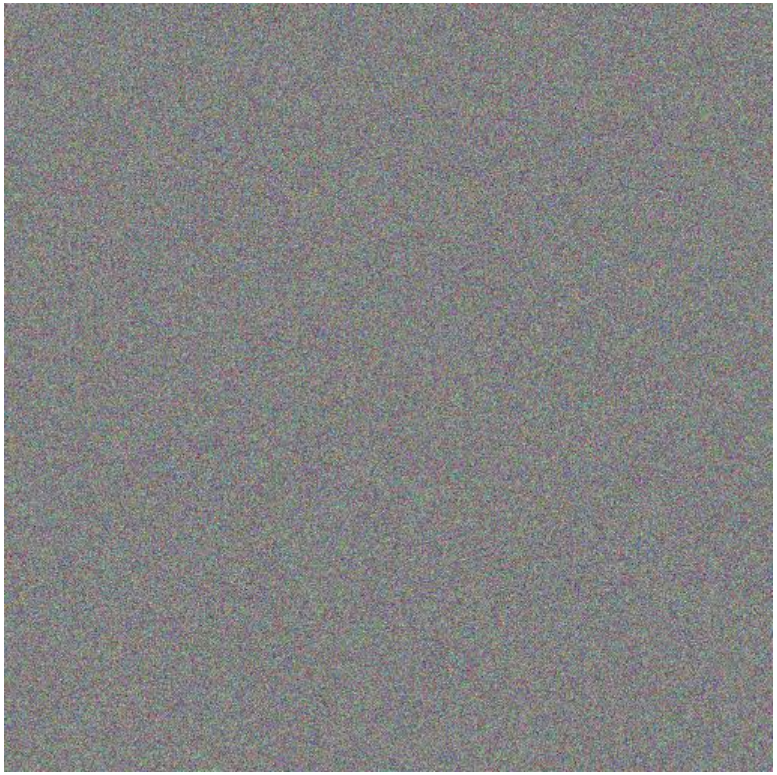


Demo

Spectral SNR

Effects of Noise on Images

image



noise field

Spectral SNR

Effects of Noise on Images

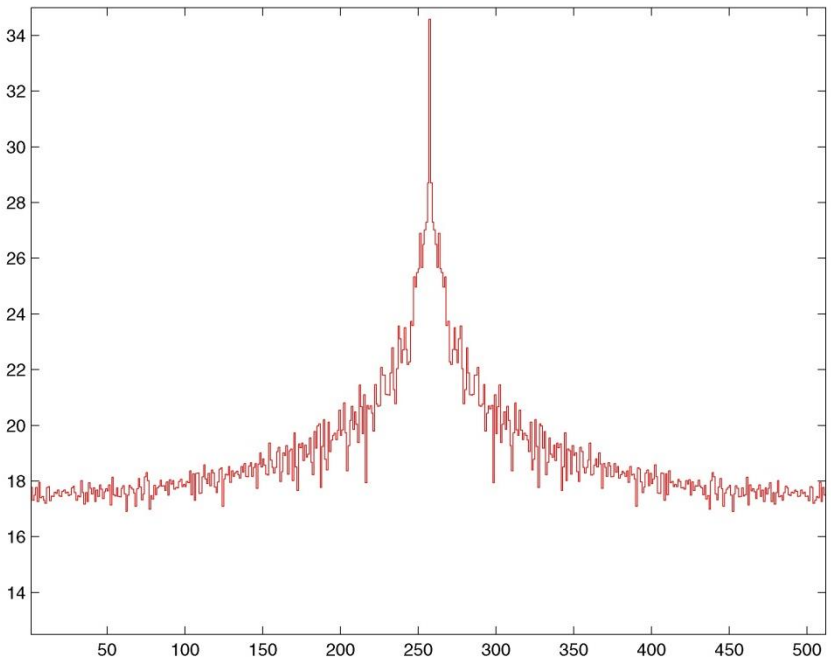
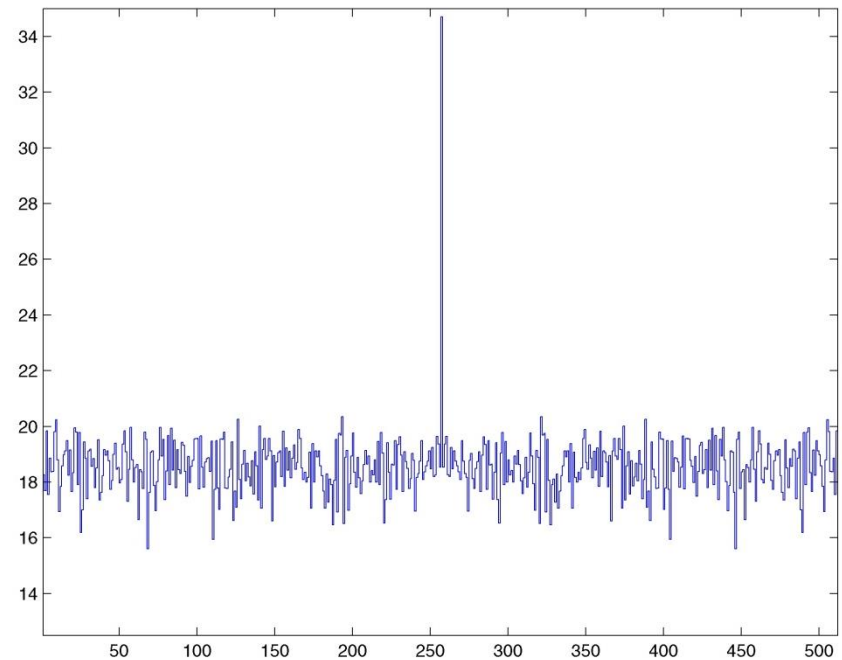


image center row log power spectrum



noise field center row log power spectrum

Spectral SNR

Effects of Noise on Images

image + noise field

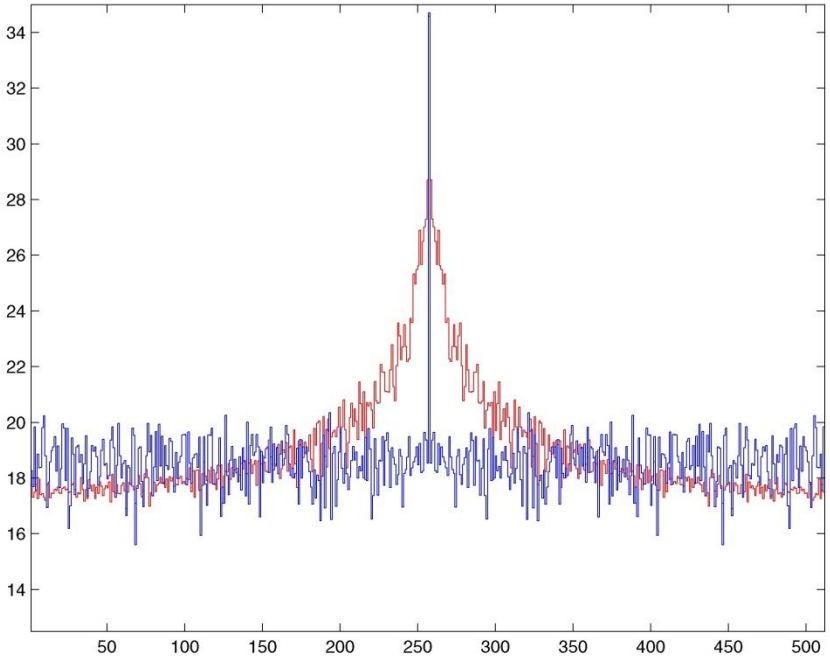
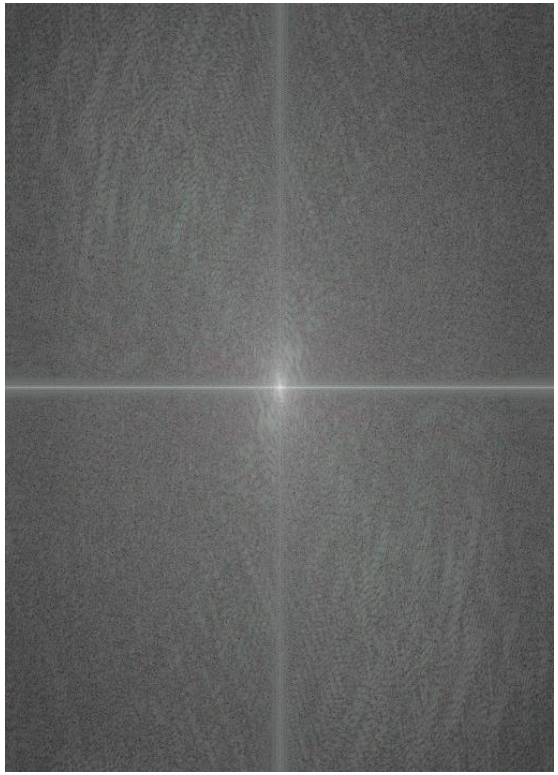


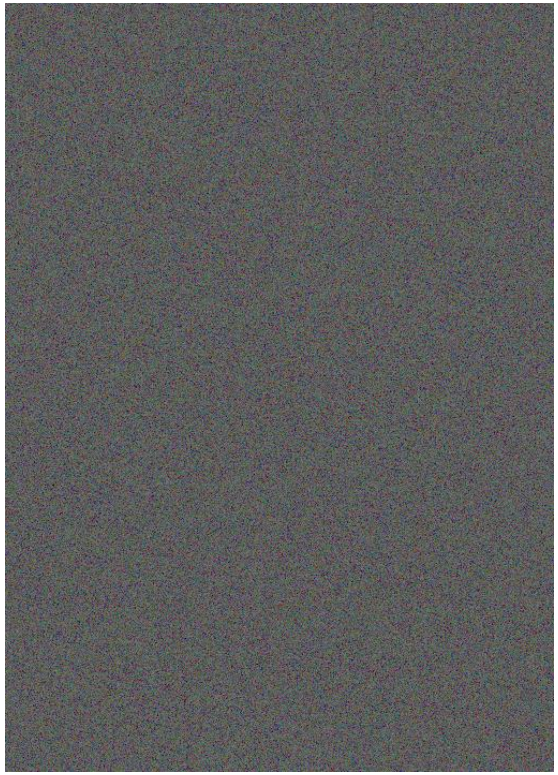
image + noise field center row log PS

Spectral SNR

Effects of Noise on Images (Power Spectra)



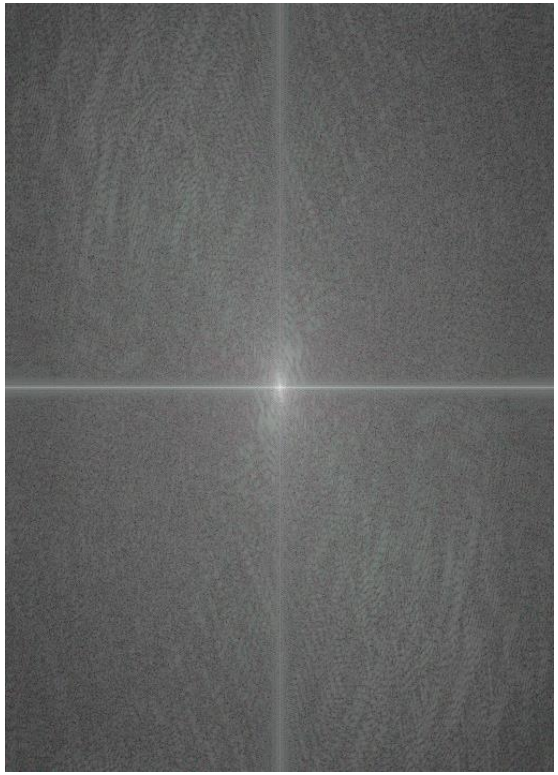
original image



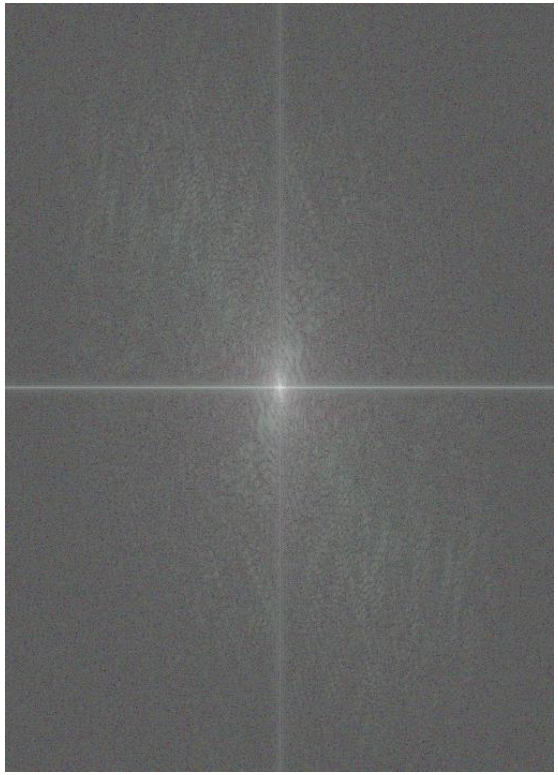
noise image

Spectral SNR

Effects of Noise on Images (Power Spectra)



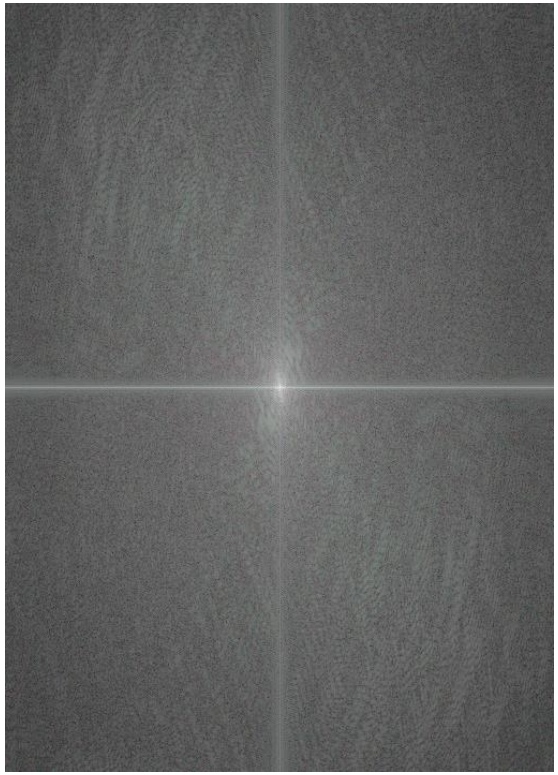
original image



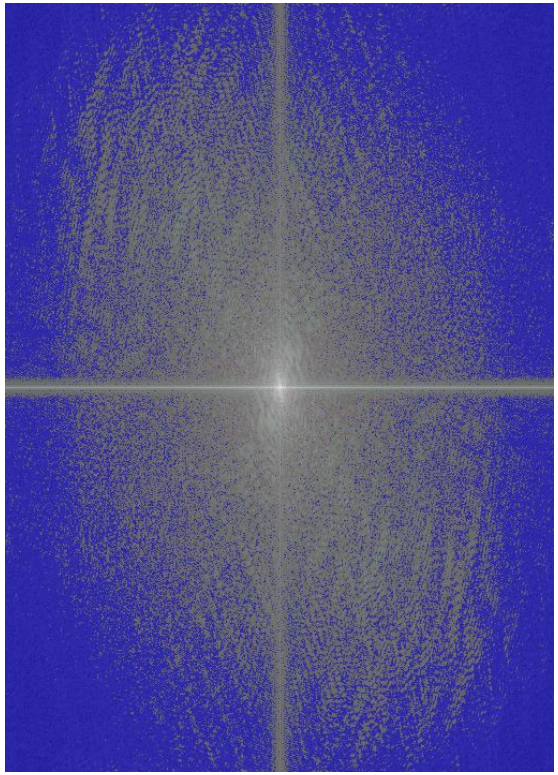
noisy image

Spectral SNR

Effects of Noise on Images (Power Spectra)



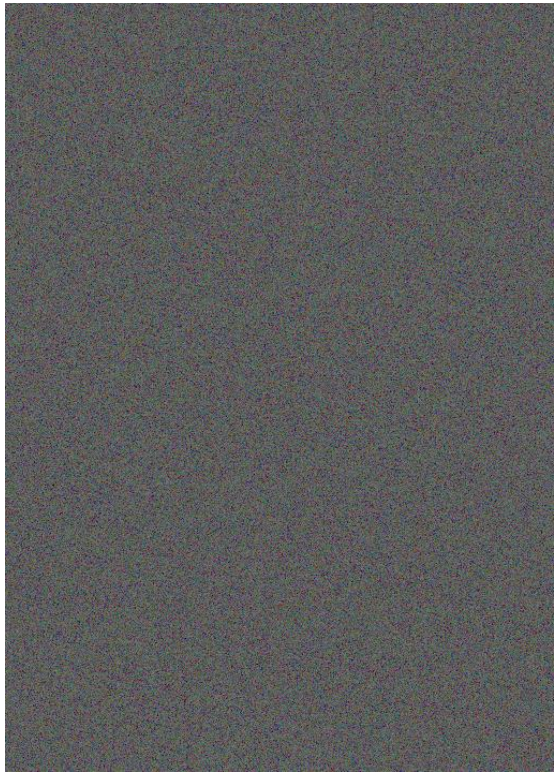
original image



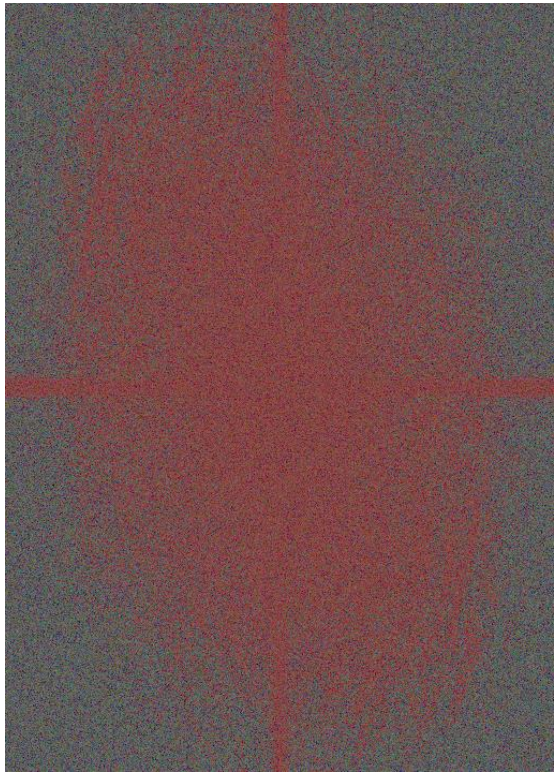
blue indicates noise > image

Spectral SNR

Effects of Noise on Images (Power Spectra)

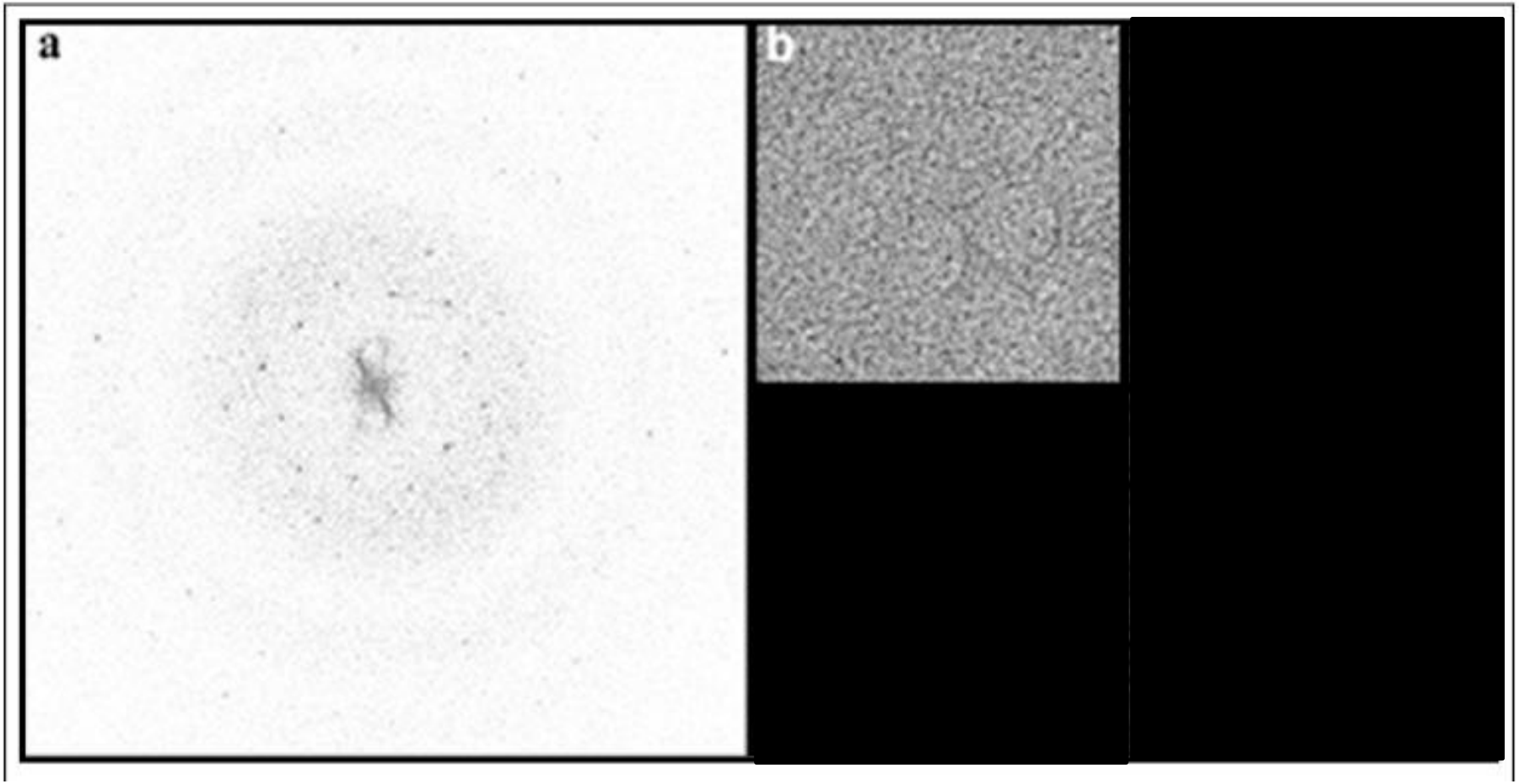


noise image

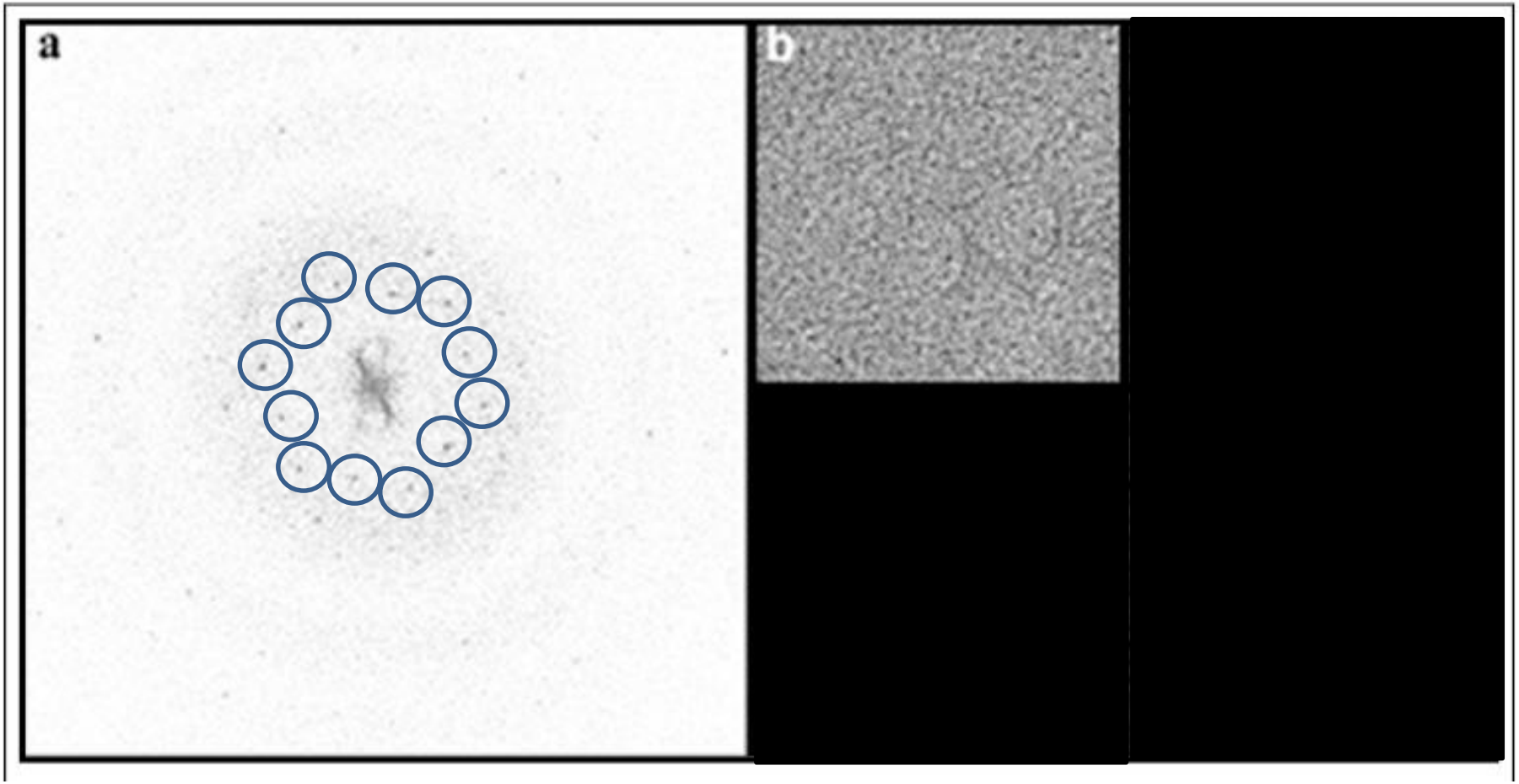


red indicates image > noise

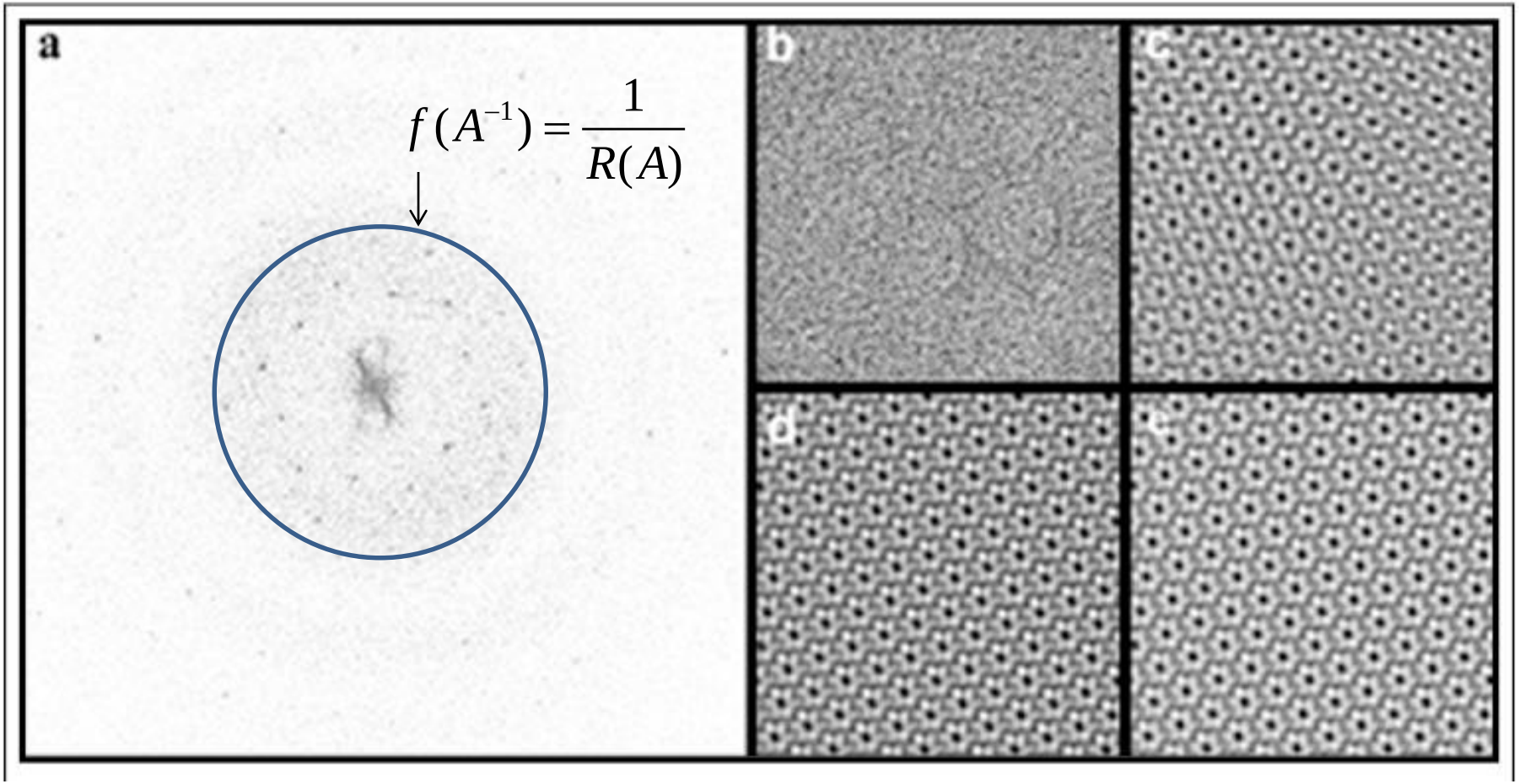
Quasi optical filtering



Quasi optical filtering



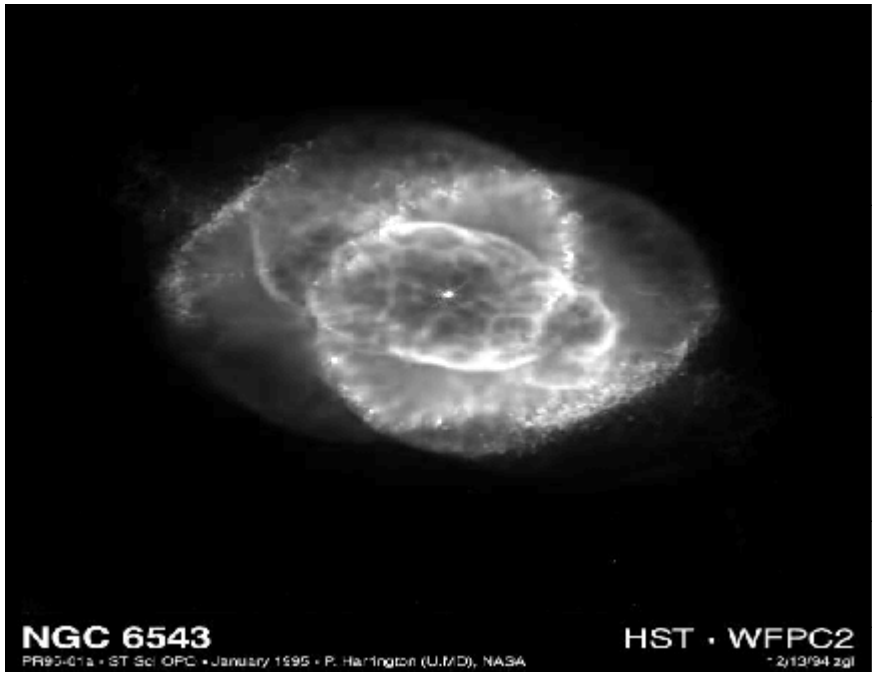
Quasi optical filtering



$$f(A^{-1}) = \frac{1}{R(A)}$$

Deconvolution in Fourier space

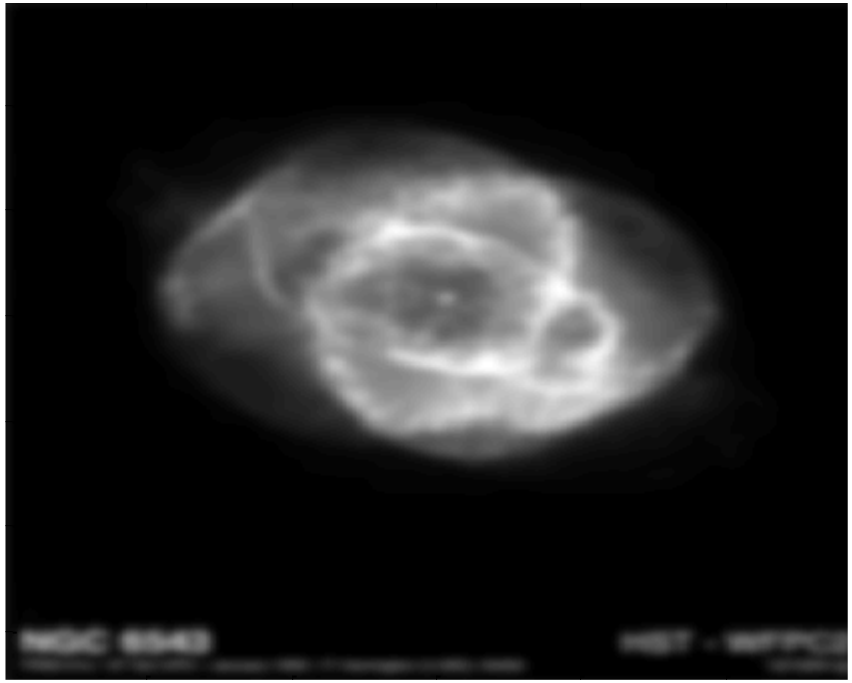
I



Deconvolution in Fourier space

$$I$$

$$I_{obs} = I \otimes PSF \Rightarrow FT\left\{I_{obs}\right\} = FT\left\{I\right\} FT\left\{PSF\right\}$$



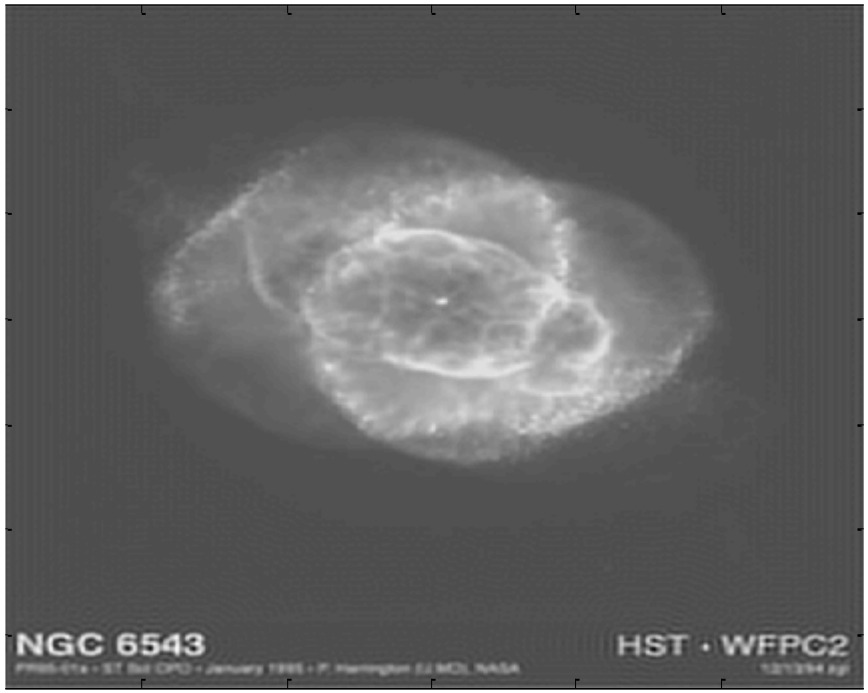
Deconvolution in Fourier space

$$I$$

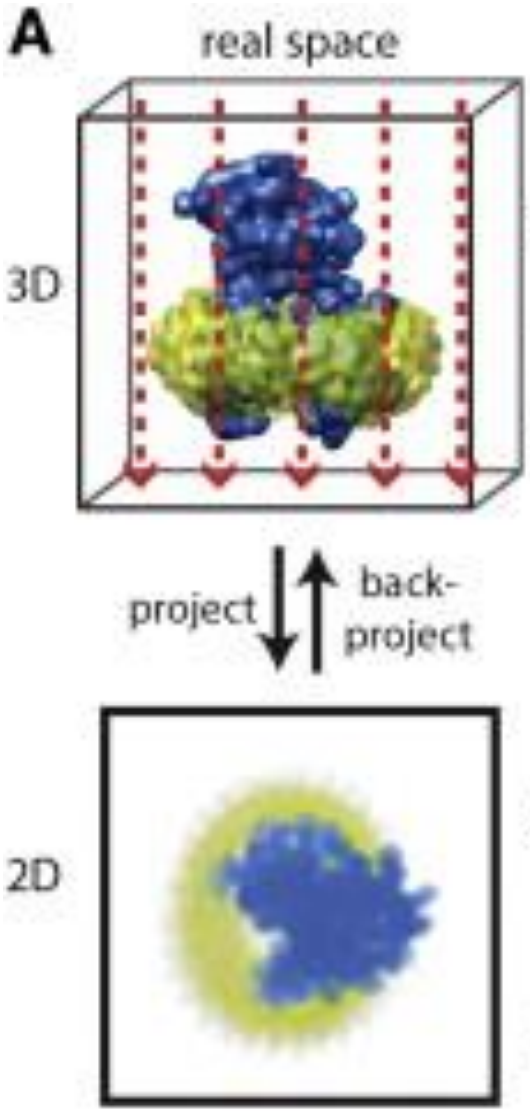
$$I_{obs} = I \otimes PSF \Rightarrow FT\left\{I_{obs}\right\} = FT\left\{I\right\} FT\left\{PSF\right\}$$

$$FT\left\{I\right\} = \frac{FT\left\{I_{obs}\right\}}{FT\left\{PSF\right\}}$$

$$I = FT^{-1}\left\{\frac{FT\left\{I_{obs}\right\}}{FT\left\{PSF\right\}}\right\}$$

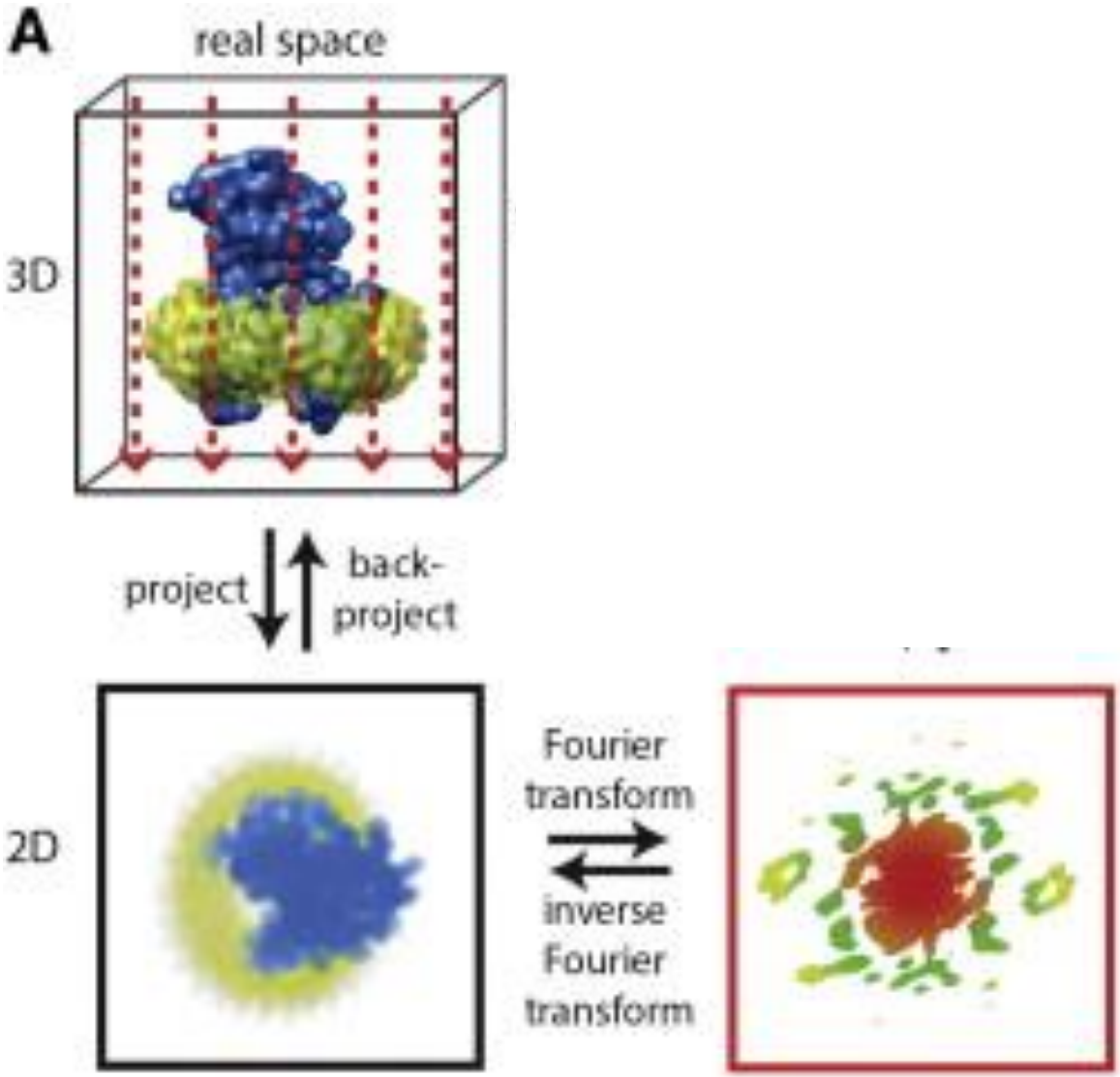


Central slice theorem



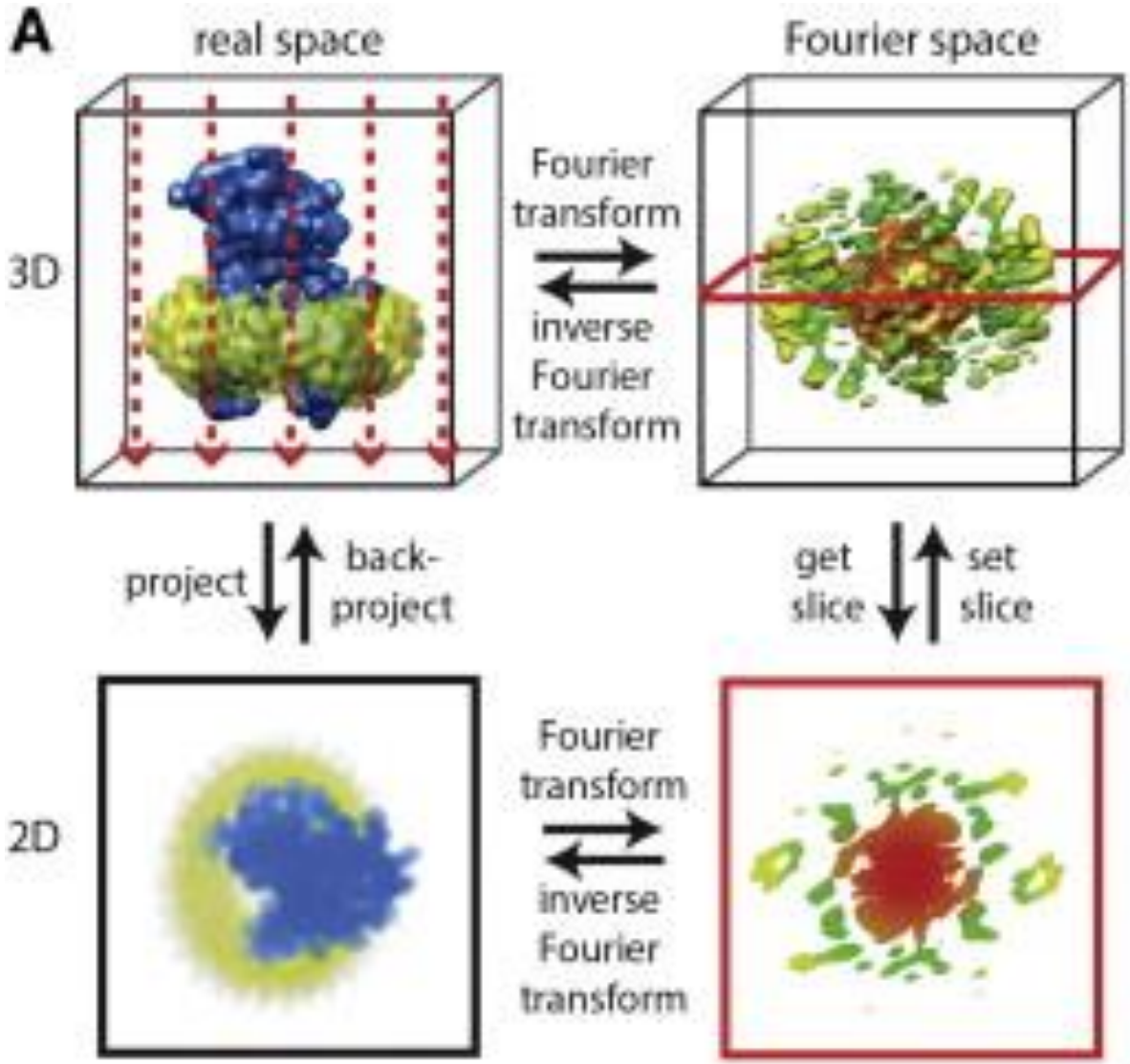
Nogales, Scheres. Molecular Cell, 58: 677-689 (2015)

Central slice theorem



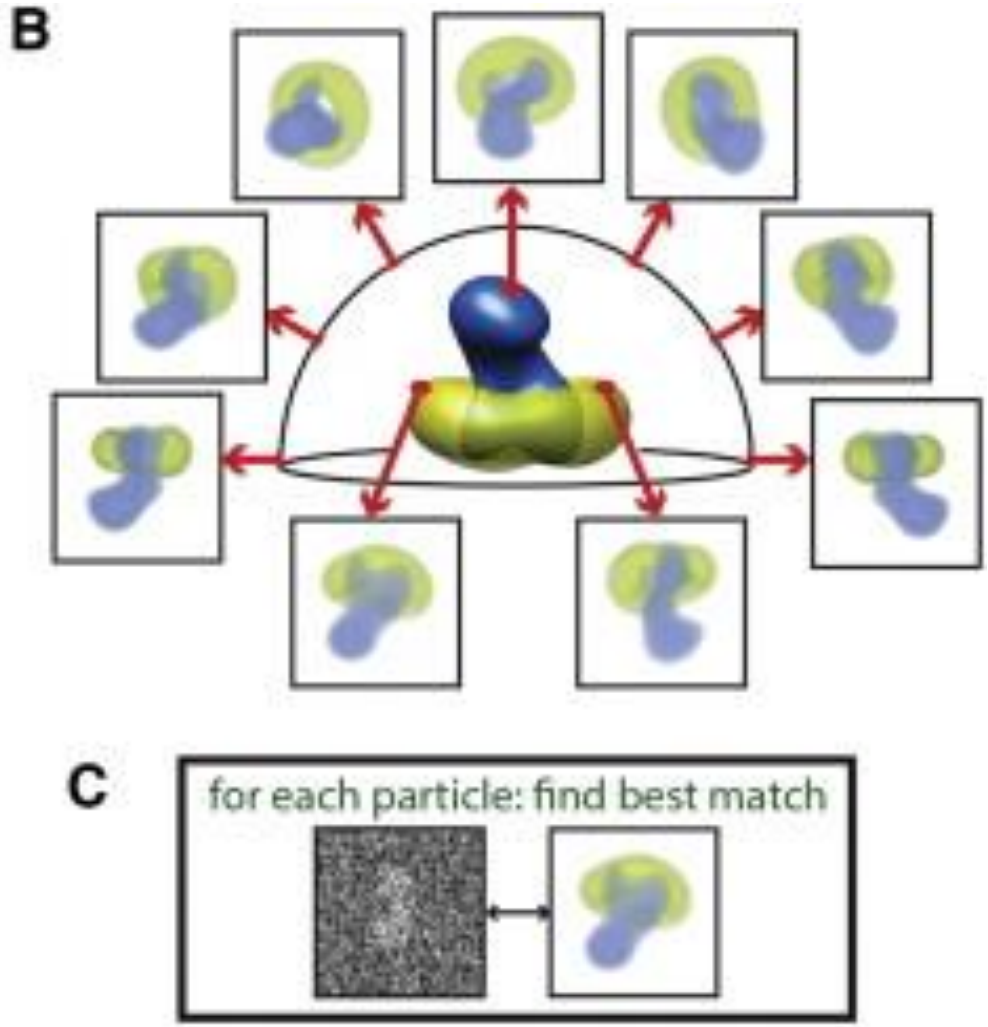
Nogales, Scheres. Molecular Cell, 58: 677-689 (2015)

Central slice theorem



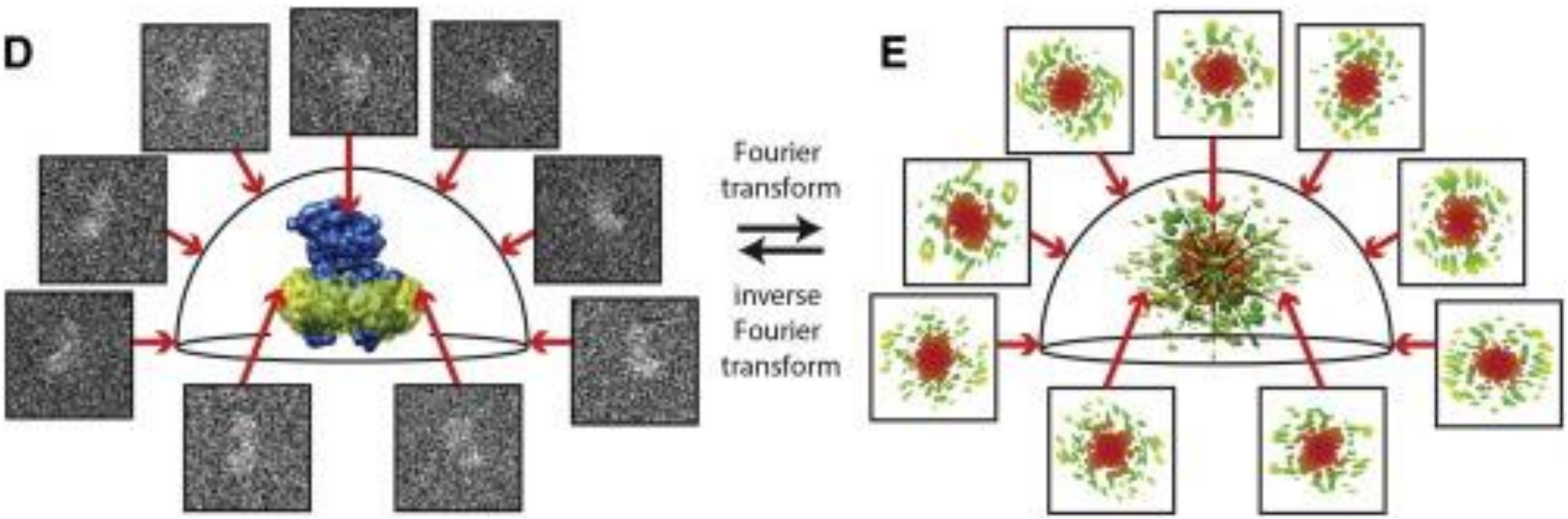
Nogales, Scheres. Molecular Cell, 58: 677-689 (2015)

Projection Matching



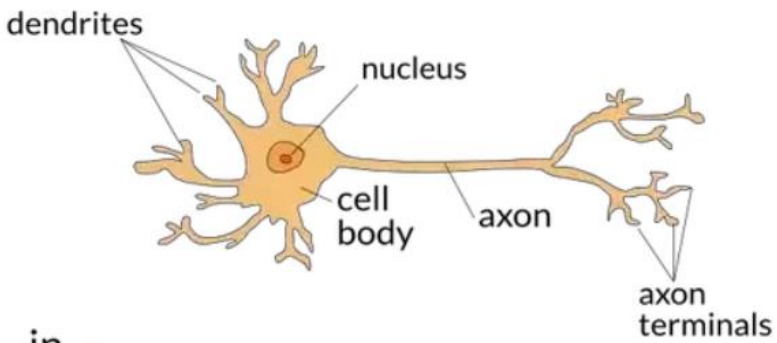
Nogales, Scheres. Molecular Cell, 58: 677-689 (2015)

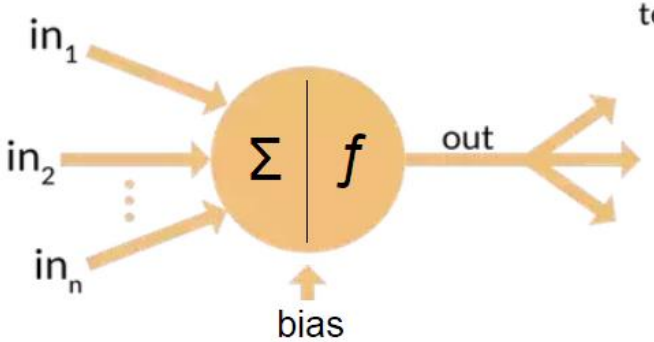
Projection matching and Central Slice Theorem

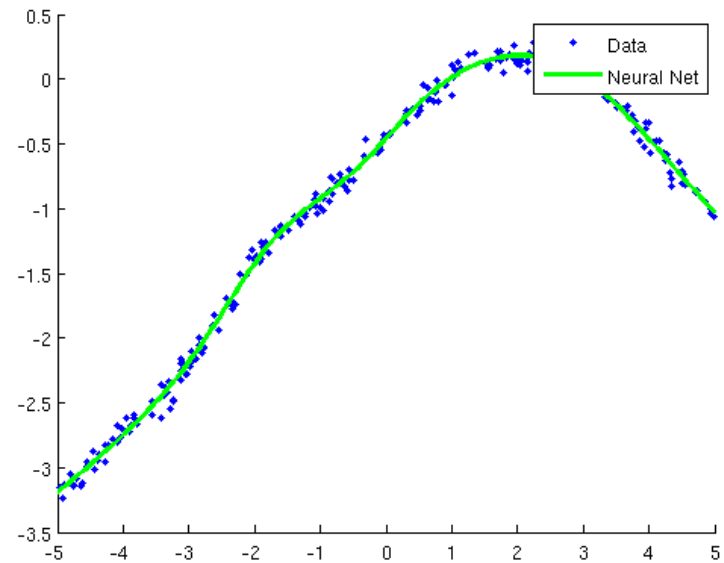


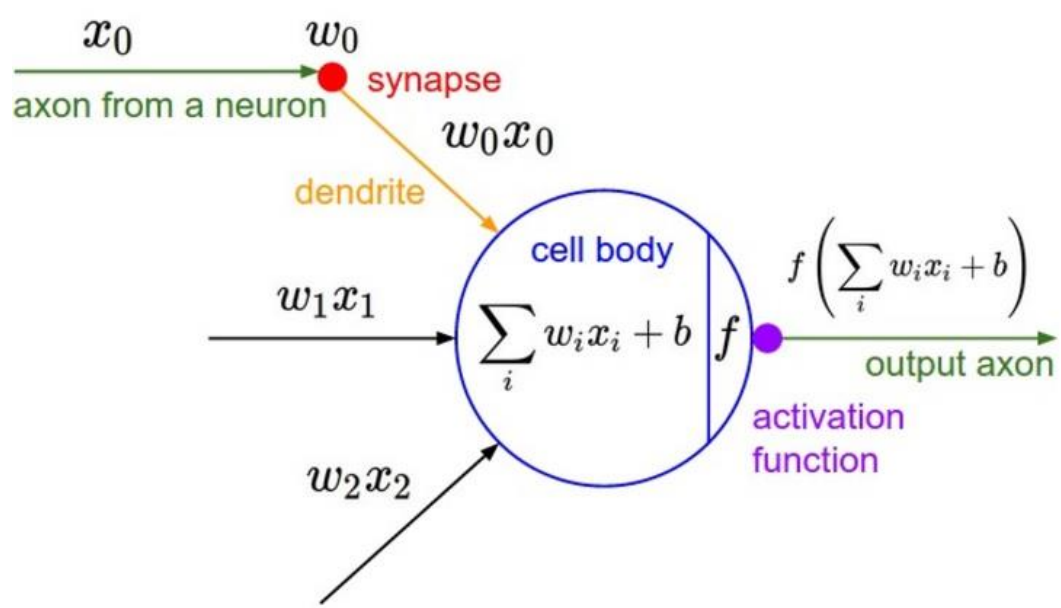
Nogales, Scheres. Molecular Cell, 58: 677-689 (2015)

Deep Learning

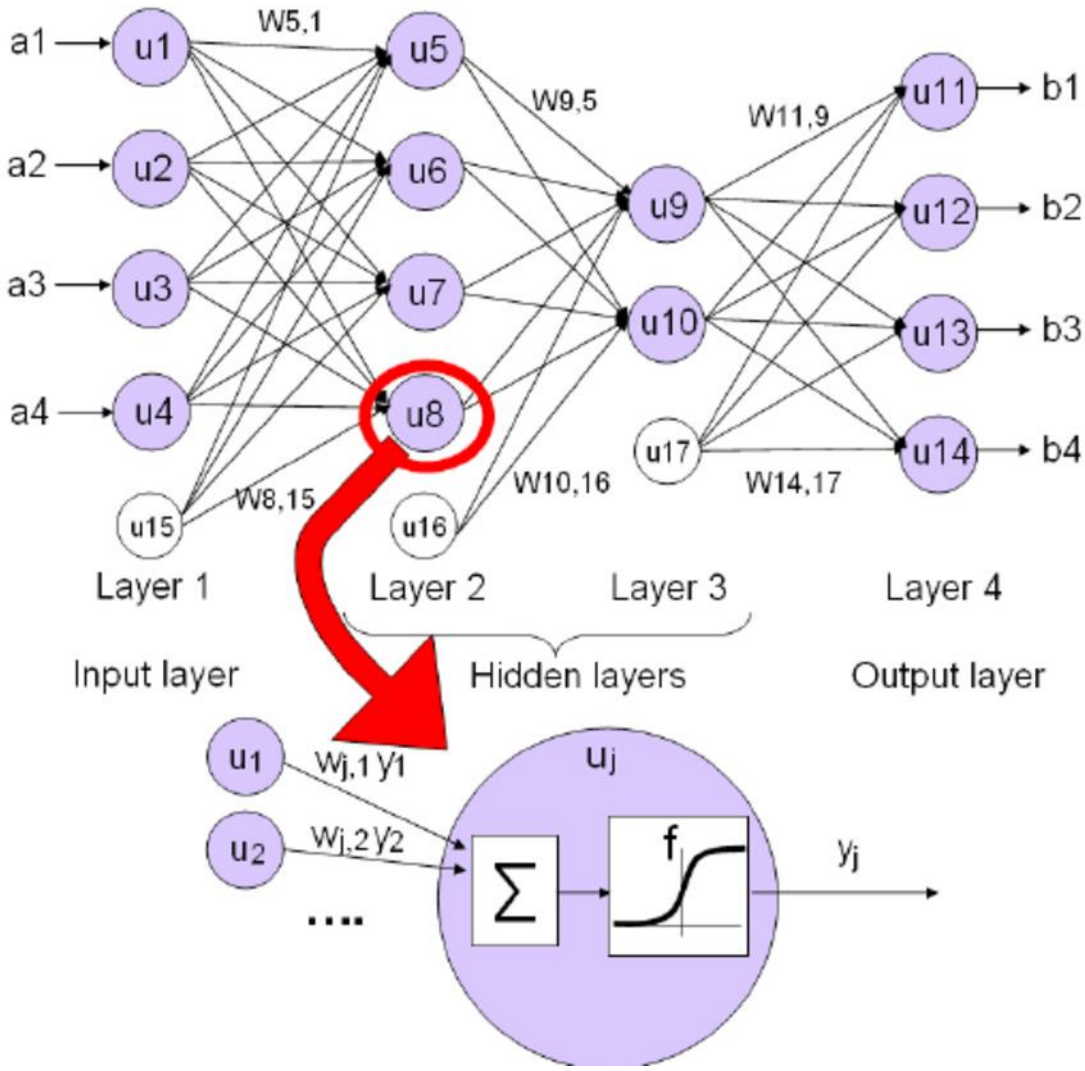




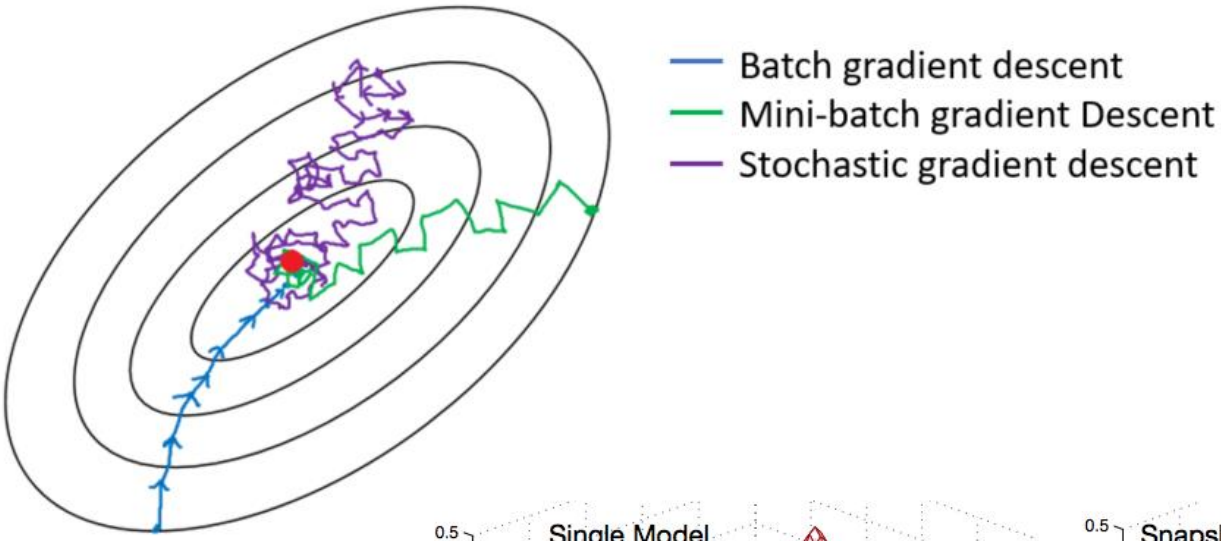


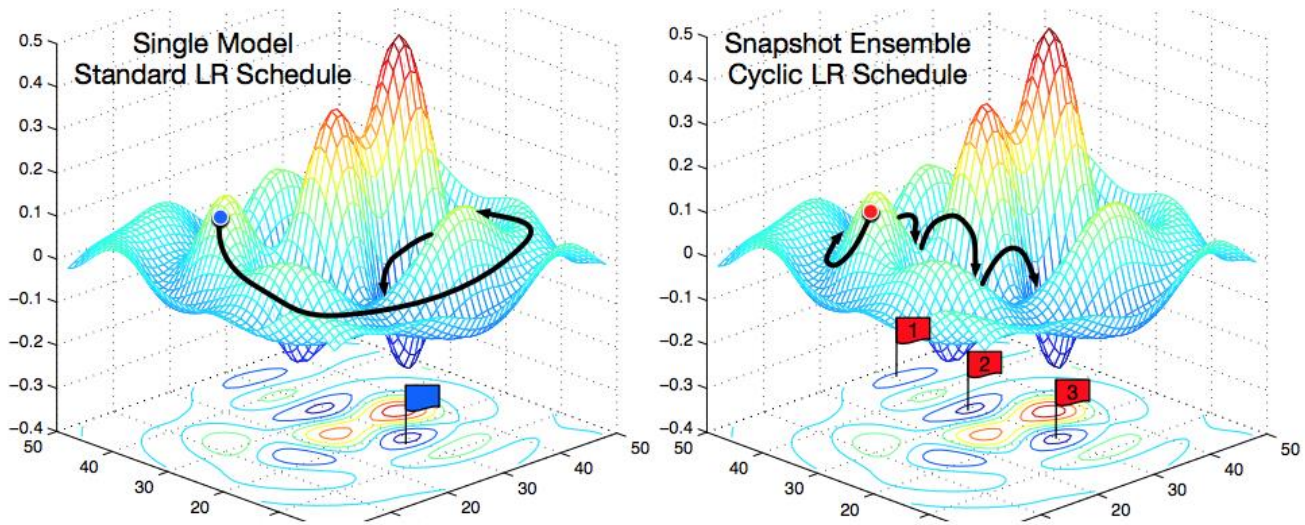


Deep Learning

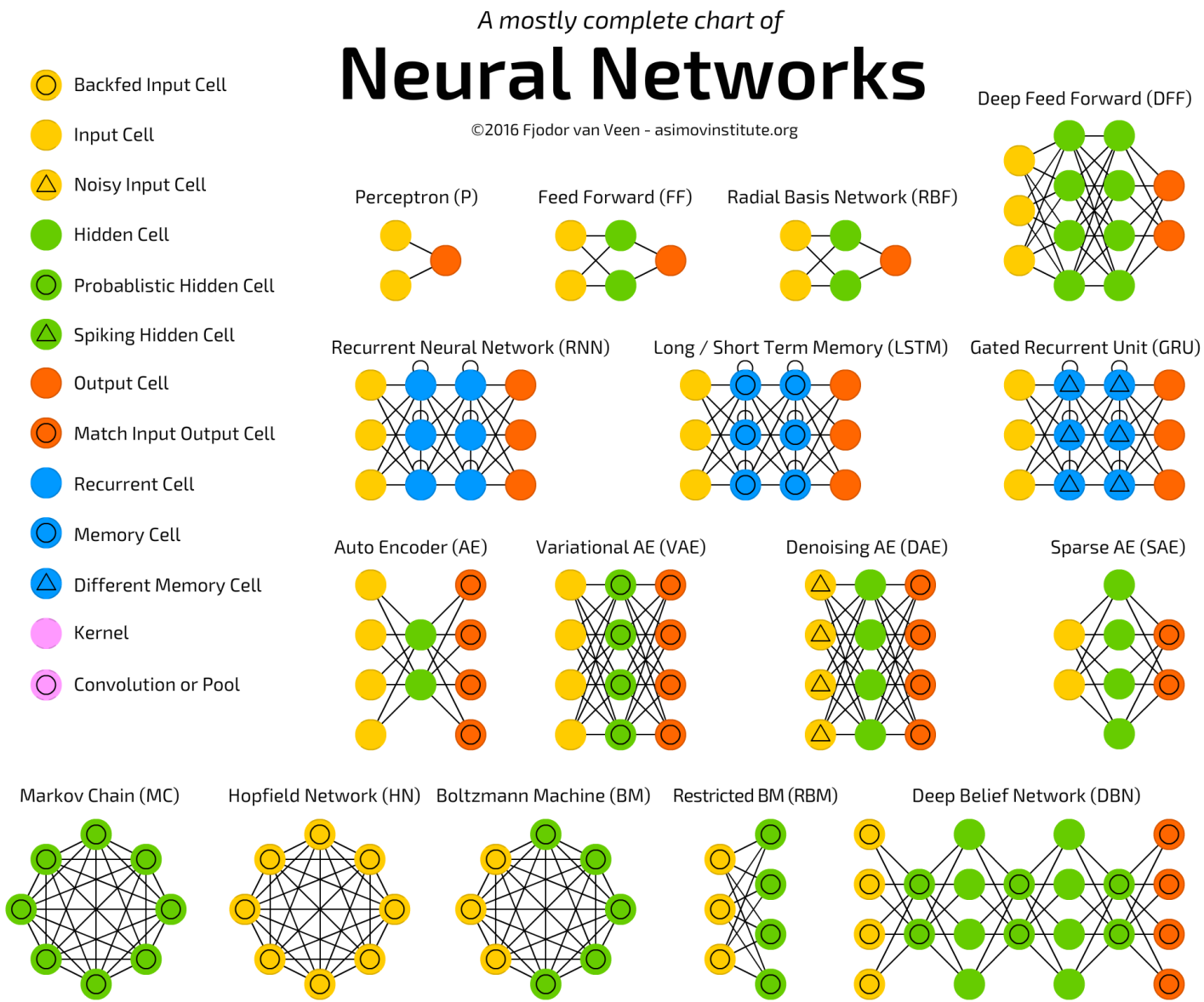


Stochastic optimization

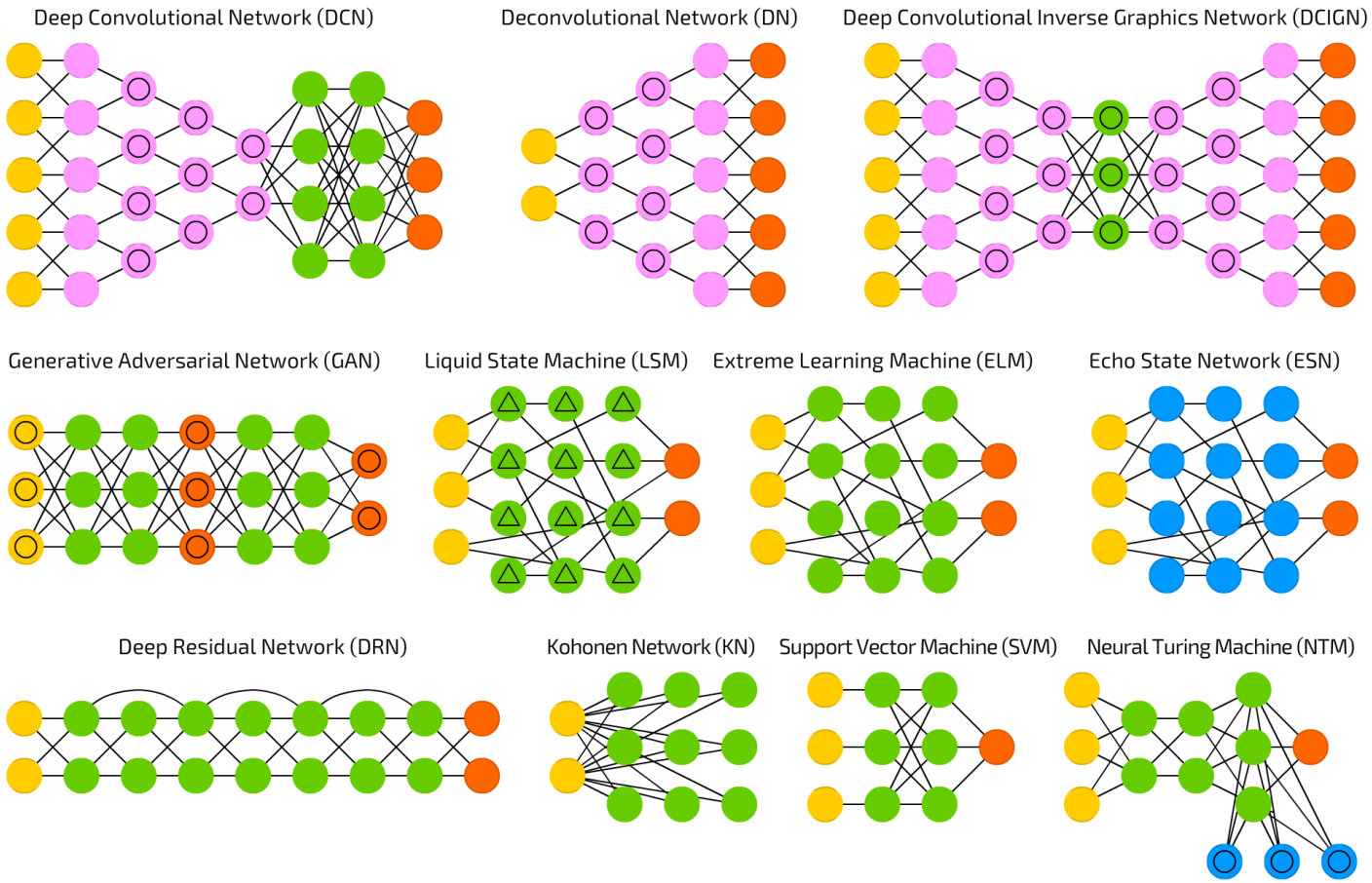




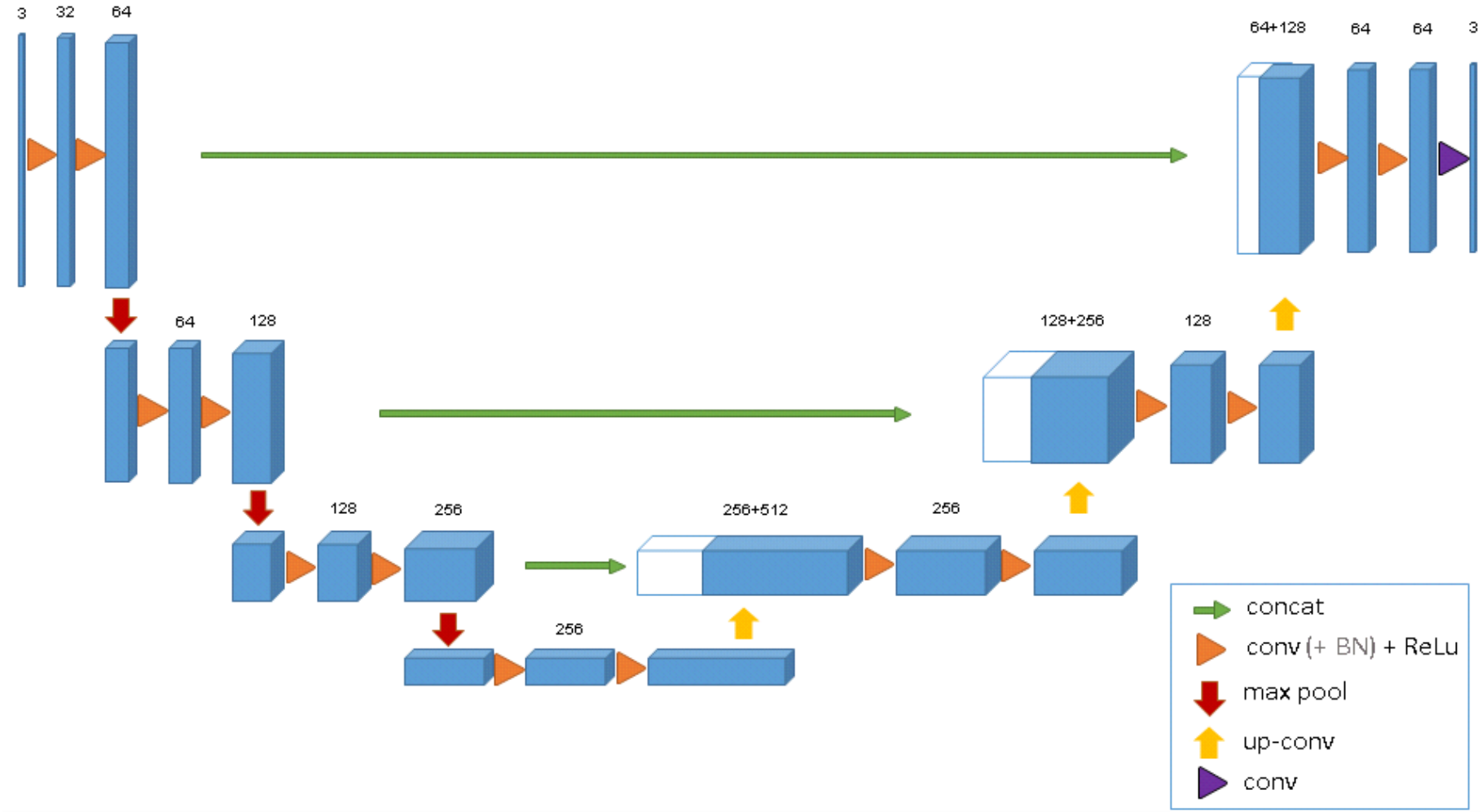
Deep neural network



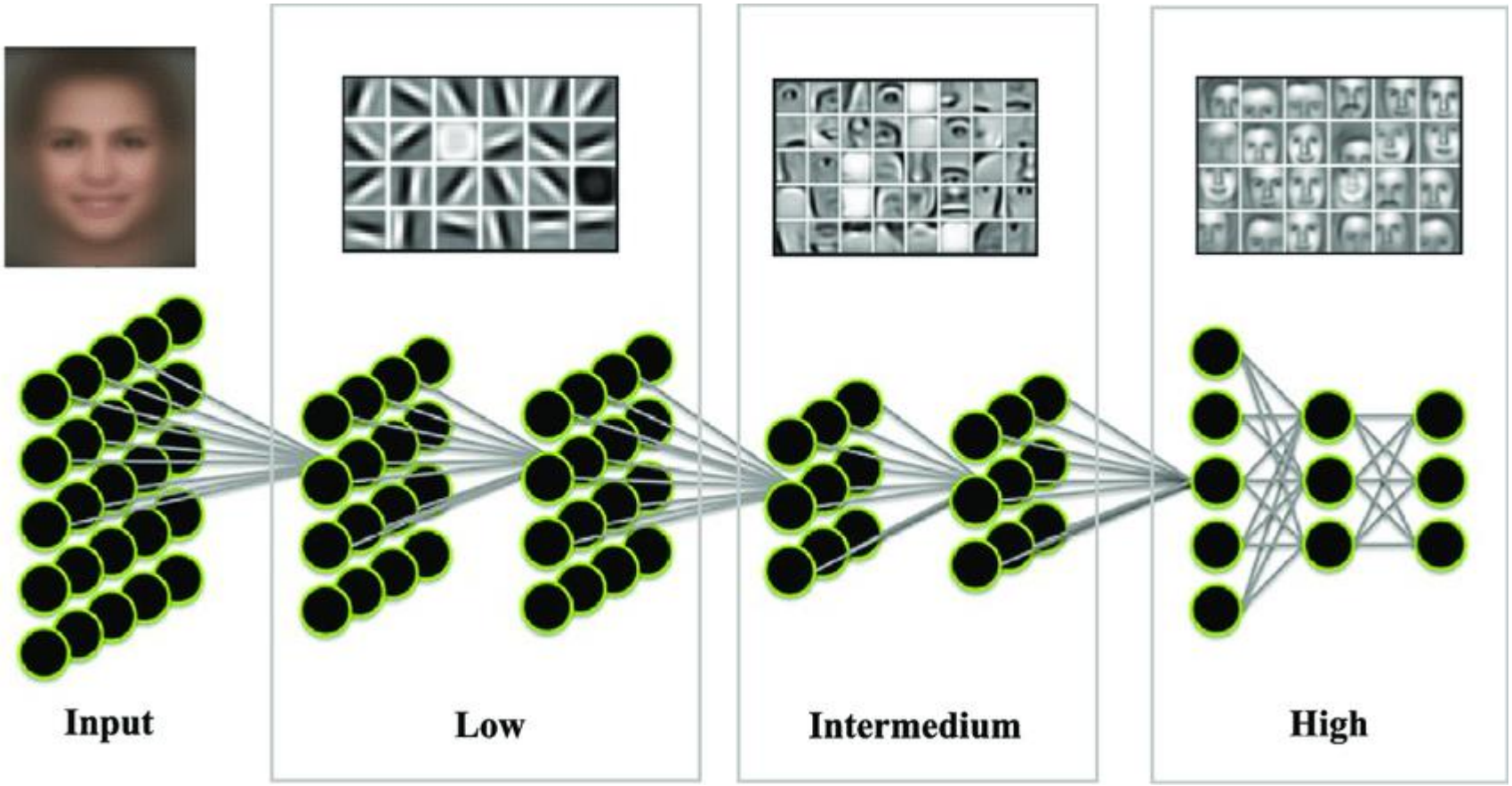
Deep neural network



Deep neural network



Deep neural network



Conclusions

- The quality of an image depends on its bit depth and its sampling rate
- Being a matrix of numbers, we can perform many operations with images at the level of:
  - Full images
  - Pixels
  - Groups of pixels
  - Geometrical transformations
  - Transformations
- Fourier transforms are one of the most important transformations for EM
- Deep learning has revolutionized image processing

Acquisition automation and movie alignment

Carlos Oscar S. Sorzano

Instruct Image Processing Center

CSIC

instruct Integrating Biology

instruct image Processing Center

Google: 3DEM Methods (Wiki+Mail list)

main page | discussion | edit | history | mov

3D EM Methods

navigation
- Main Page
- Recent changes
- All articles

search
Search
Go | Search

tools
- What links here
- Related changes
- Special pages
- Printable version
- Permanent link
- Page information

Main Page

Contents [hide]

page | discussion | edit | history | move | unwatch

3D EM Methods

navigation
- Main Page
- Recent changes
- All articles

search
Search
Go | Search

tools
- What links here
- Related changes
- Special pages
- Printable version
- Permanent link
- Page information

2016Liu TiltControl

Contents [hide]

Citation [edit]

Liu, J.; Li, H.; Zhang, L.; Rames, M.; Zhang, M.; Yu, Y.; Peng, B.; Celis, C. D.; Xu, A.;

Abstract [edit]

Knowledge of three-dimensional (3D) structures of each individual particles of asym
provides a tool for imaging a single and unique biological object from a series of tilt
specimen goniometer under both a medium to high magnification (approximately 50
using beam tilt/shift processes. This method could reduce the accumulation of beam
error of the target object center during the tilting process through a closed-loop prop
comparable capability to other ET methods in tracking target proteins while maintair

Keywords [edit]

Links [edit]

http://www.nature.com/articles/srep29231

CSIC

instruct Integrating Biology

instruct image Processing Center

Automation

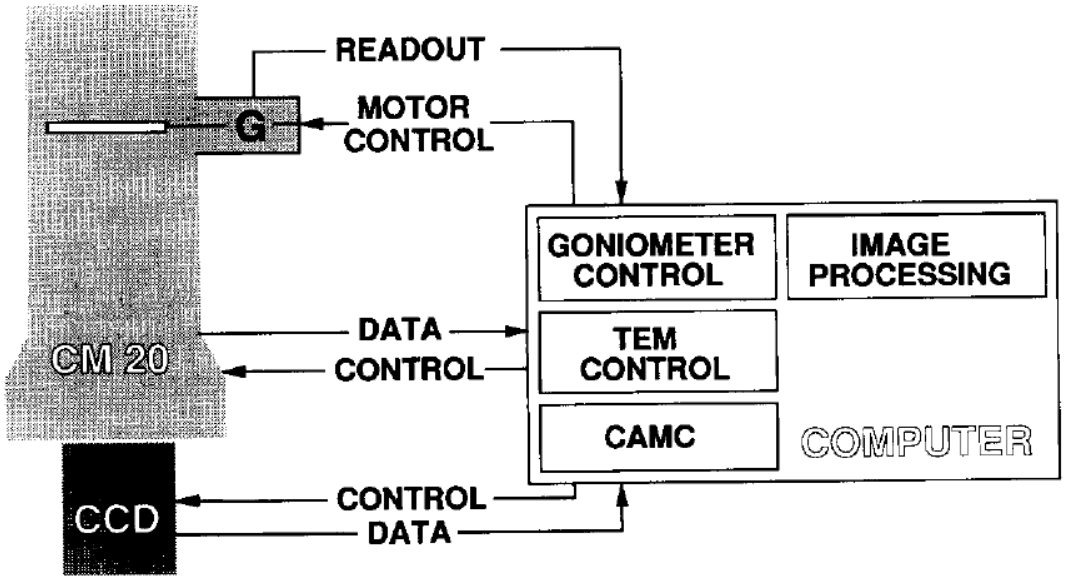
1992

Ultramicroscopy 40 (1992) 71–87
North-Holland

ultramicroscopy

Towards automatic electron tomography

K. Dierksen, D. Typke *, R. Hegerl, A.J. Koster [1] and W. Baumeister
Max-Planck-Institut für Biochemie, W-8033 Martinsried, Germany

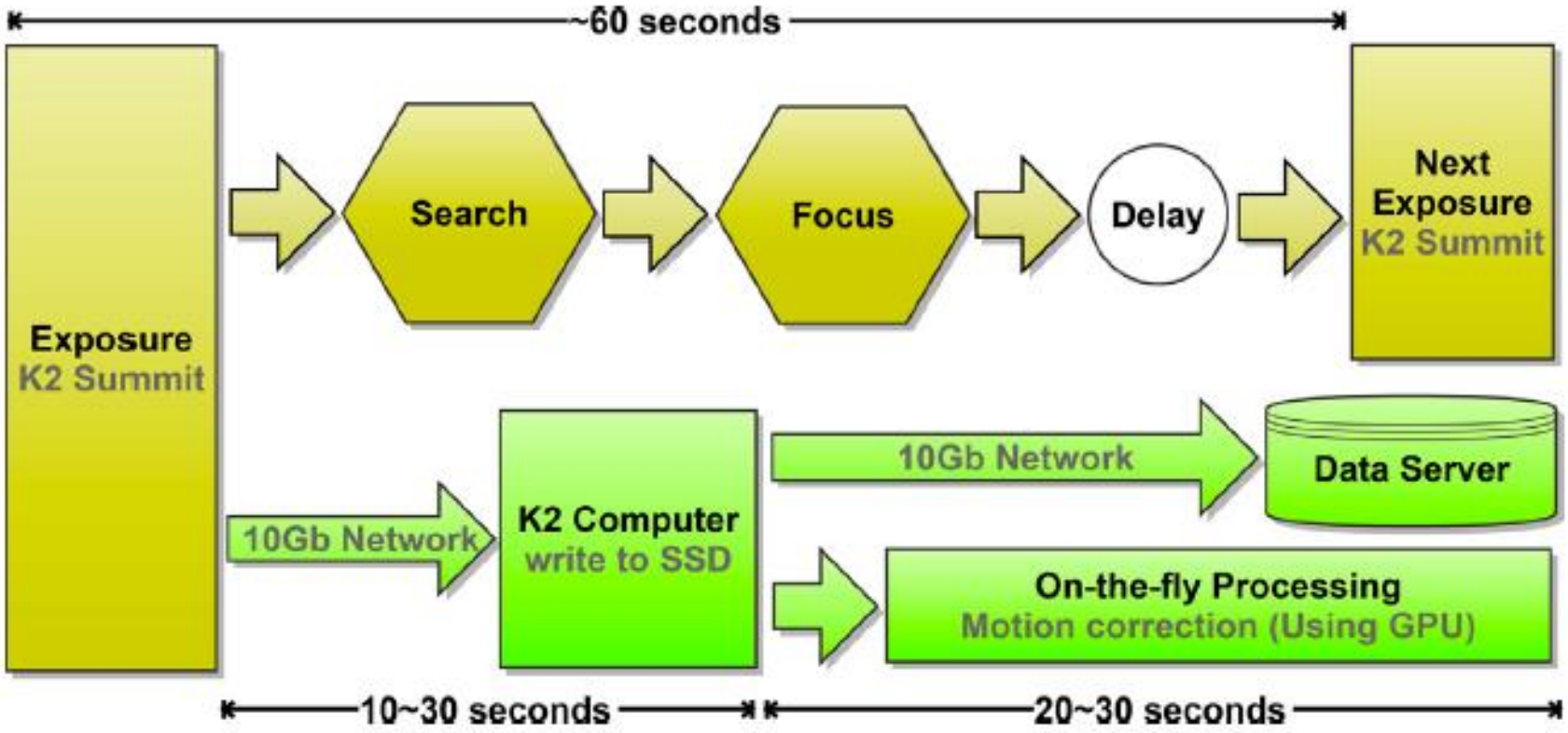


Automation

2017=1992+25

- Autoloader
- Thermal stabilization
- Automatic focusing
- Acquisition software
  - EPU (FEI)
  - SerialEM (Mastronarde, Colorado)
  - Leginon (Carragher, NY)
  - (Camera support)

Streaming Processing



Li, X.; Zheng, S.; Agard, D. A. & Cheng, Y. Asynchronous data acquisition and on-the-fly analysis of dose fractionated cryoEM images by UCSFImage. J Struct Biol, 2015, 192, 174-178

Streaming Processing

Acquisition
~60s (4k x 4k x 34) -2Gb HD

Microscope

Transfer
~16s - 1Gb/s
~4s – 4Gb/s
~1.6 – 10Gb/s

Scipion Box

Import Movies

Movie Alignment

CTFFIND4

~17s
1 Core
1Gb RAM
gCTF

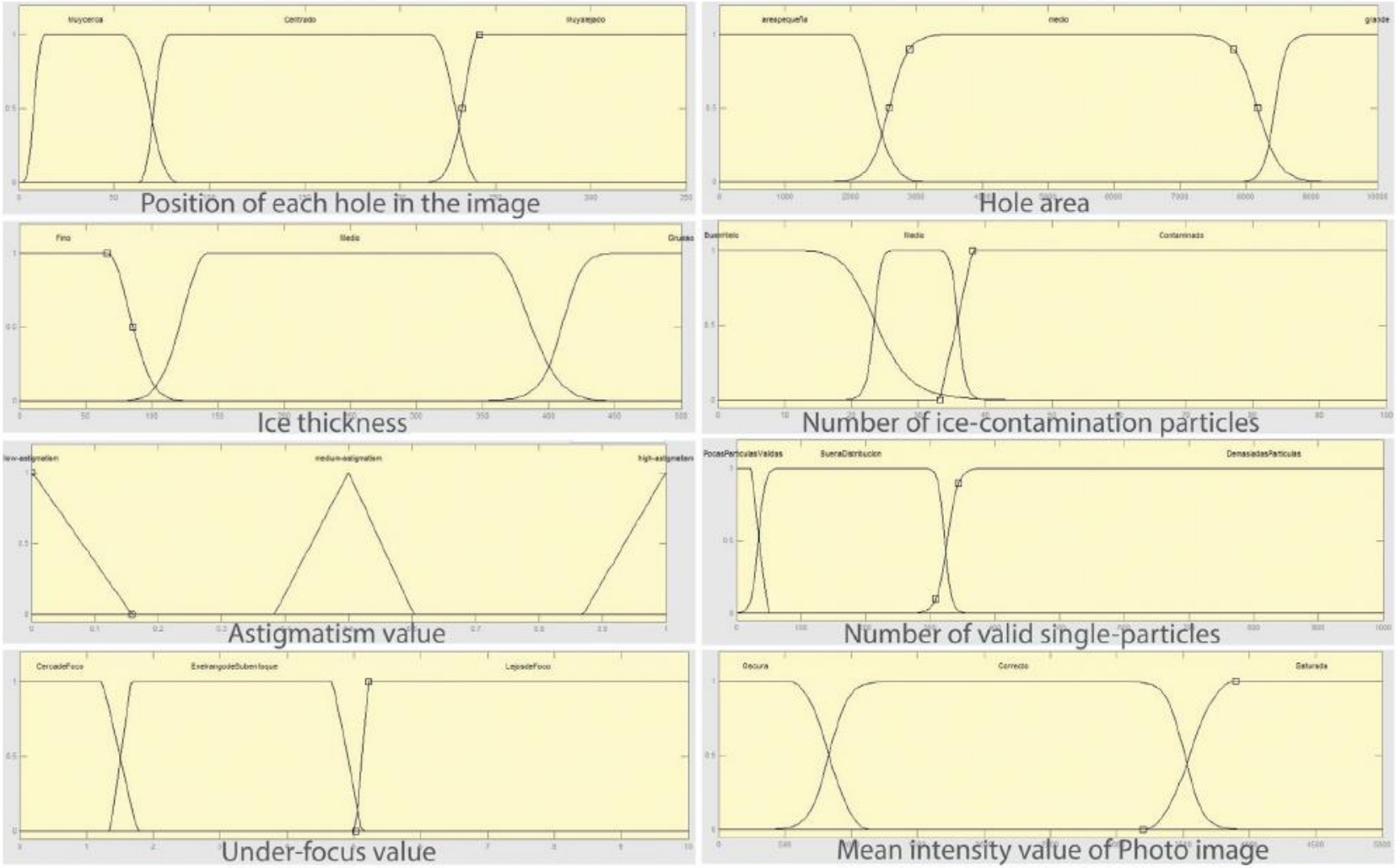
MotionCorr

~14s – (22 frames)
2 Cores
2Gb RAM
0.4 Gb VRAM
GPU

Optical Flow

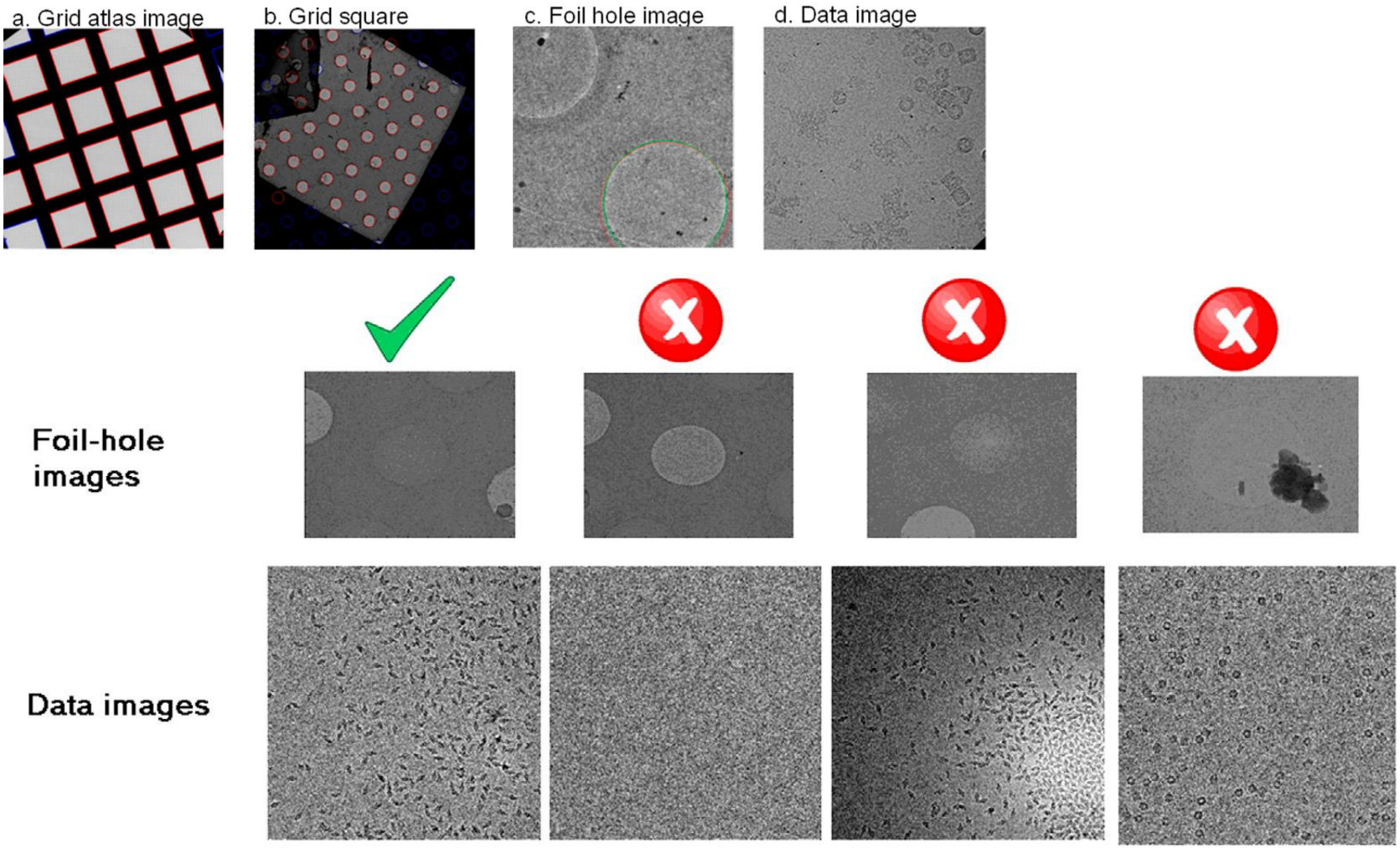
~660s – (22 frames)
1 Core
2Gb RAM
12 Cores < 60s

Intelligent acquisition



Gil-Carton, D.; Zamora, M.; Sutherland, J. D.; Barrio, R.; Garrido, I.; Valle, M. & Garrido, A. J. Real-time and decision taking selection of single-particles during automated cryo-EM sessions based on neuro-fuzzy method. Expert Systems with Applications, 2016, 55, 403-416

Intelligent acquisition



Vargas, J.; Franken, E.; Sorzano, C. O. S.; Gomez-Blanco, J.; Schoenmakers, R.; Koster, A. J. & Carazo, J. M. Foil-hole and data image quality assessment in 3DEM: Towards high-throughput image acquisition in the electron microscope. Journal of structural biology, 2016, 196, 515-524

Intelligent acquisition

Dataset	Num	% Bad	ErrorDI	ErrorFHI	FHI → DI
β-Gal	43	(DI)81% (FHI)81%	0%	2%	0%
	41	(DI)32% (FHI)34%	0%	3%	3%
	156	(DI)5% (FHI)6%	0%	1%	1%
GroEL	128	(DI)21% (FHI)27%	0%	1%	0%
	105	(DI)32% (FHI)27%	1%	5%	0%
KLH	140	(DI)1% (FHI)3%	0%	2%	0%
Hemoglobin	204	(DI)12% (FHI)11%	1%	3%	2%

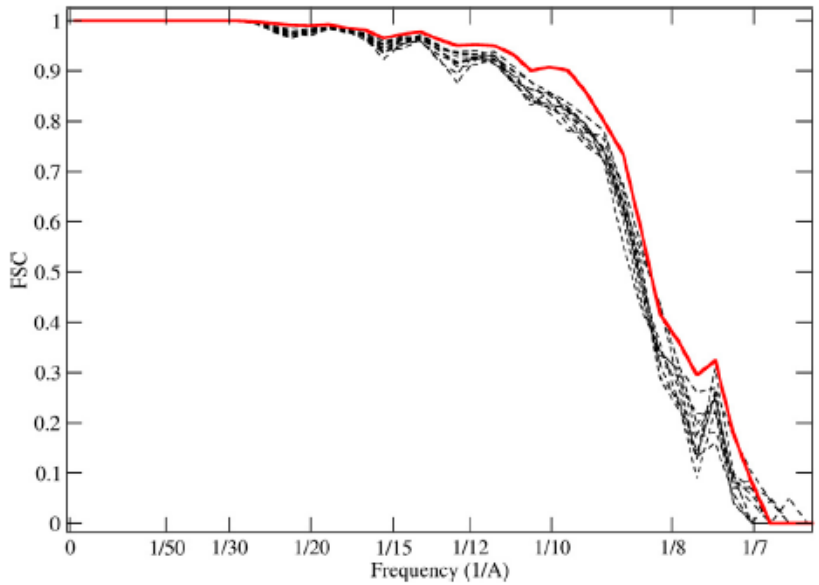
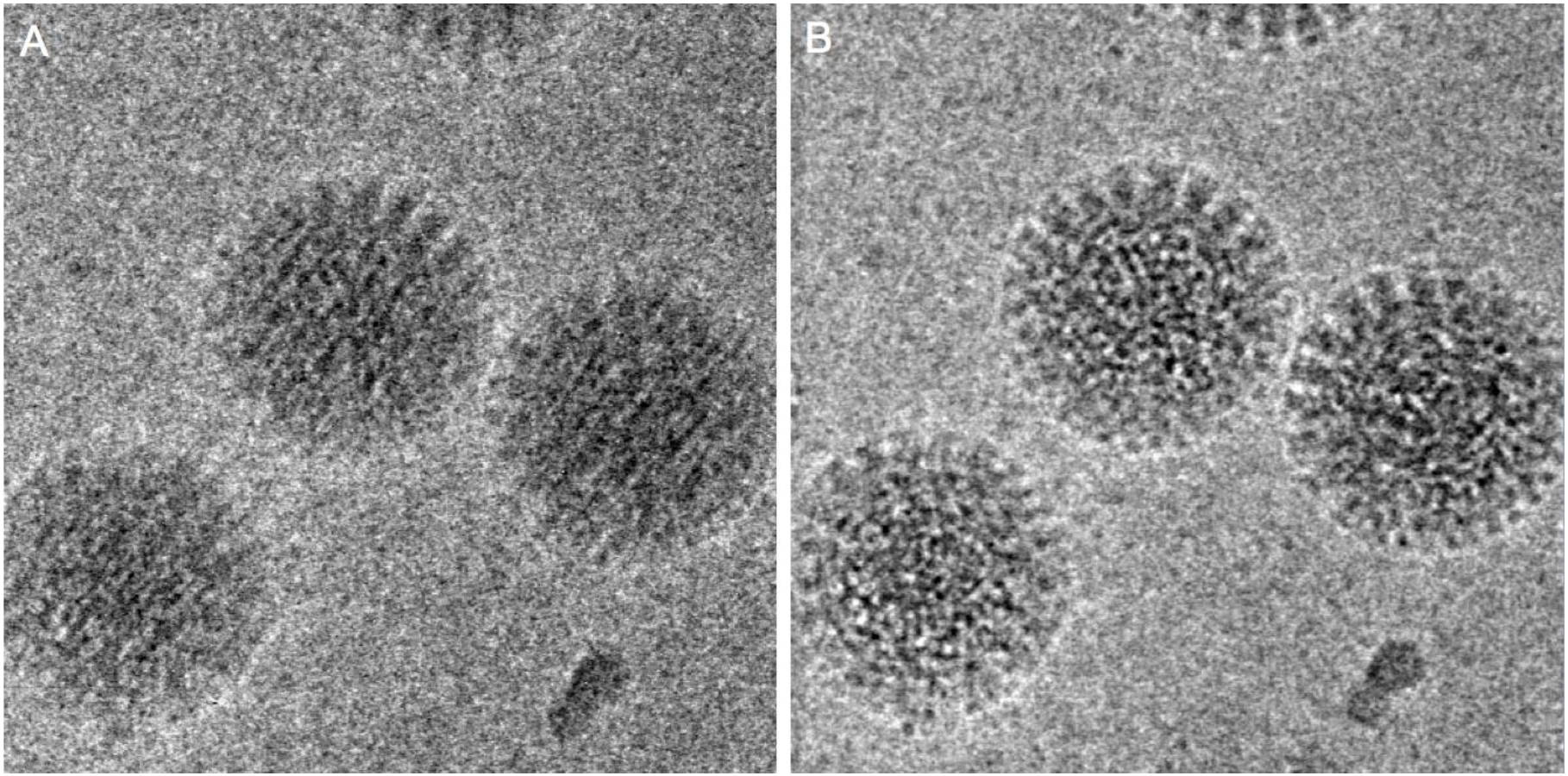


Fig. 9. The dashed black lines represent FSC curves obtained by gold-standard procedure computed using 13,242 particles randomly picked from the set of 27,347. The red curve is the respective FSC curve obtained by the 13,242 particles picked from the selected high quality set of micrographs.

1) Histogram based indicators: for ice thickness characterization and contaminant detection.
2) Local homogeneity ice descriptors: for ice homogeneity characterization.
3) Autocorrelation descriptors: for accurate ice contaminants detection.

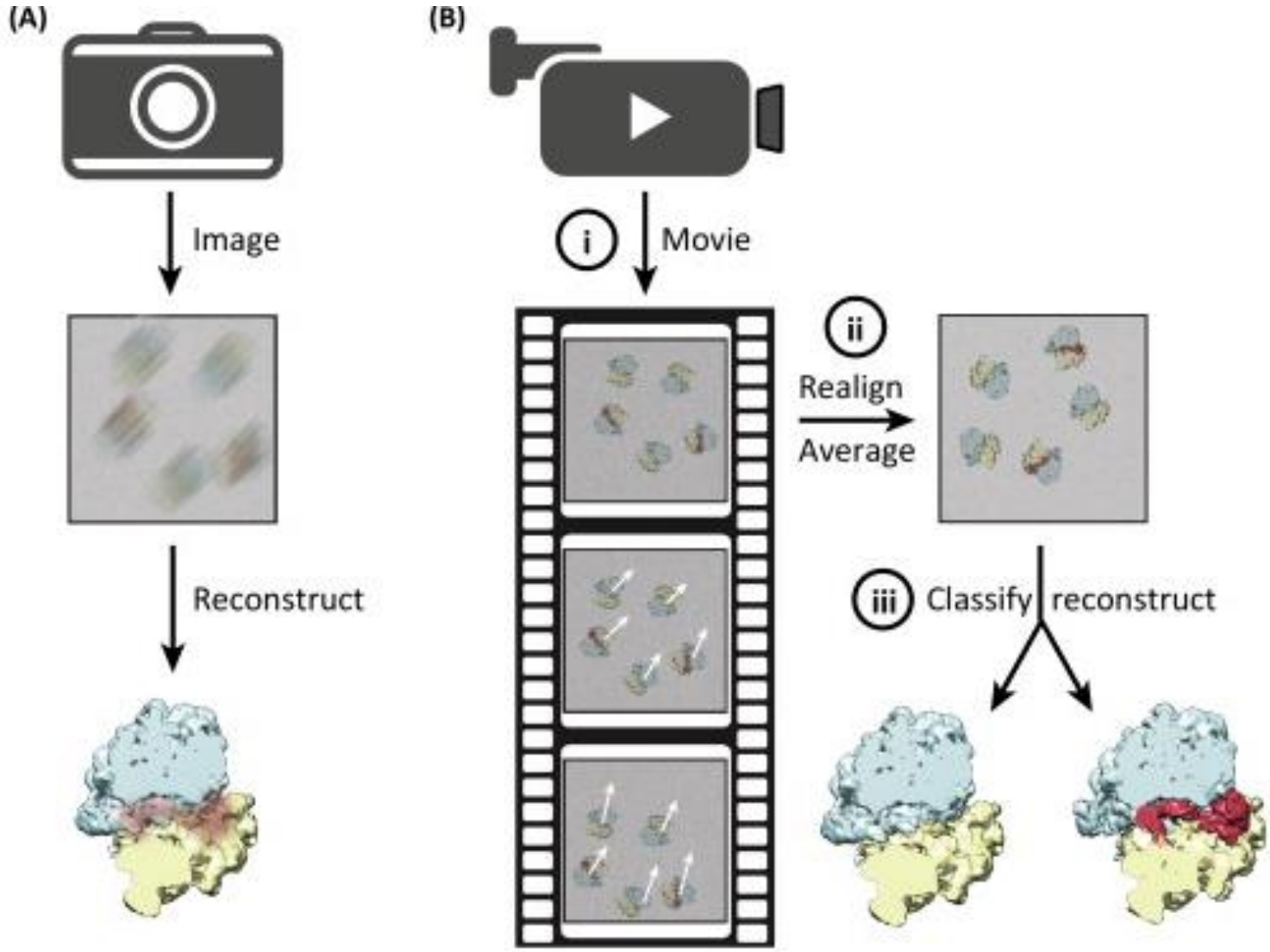
Vargas, J.; Franken, E.; Sorzano, C. O. S.; Gomez-Blanco, J.; Schoenmakers, R.; Koster, A. J. & Carazo, J. M. Foil-hole and data image quality assessment in 3DEM: Towards high-throughput image acquisition in the electron microscope. Journal of structural biology, 2016, 196, 515-524

Beam Induced Movement



Brilot, A. F.; Chen, J. Z.; Cheng, A.; Pan, J.; Harrison, S. C.; Potter, C. S.; Carragher, B.; Henderson, R. & Grigorieff, N. Beam-induced motion of vitrified specimen on holey carbon film. J Struct Biol, 2012, 177, 630-637

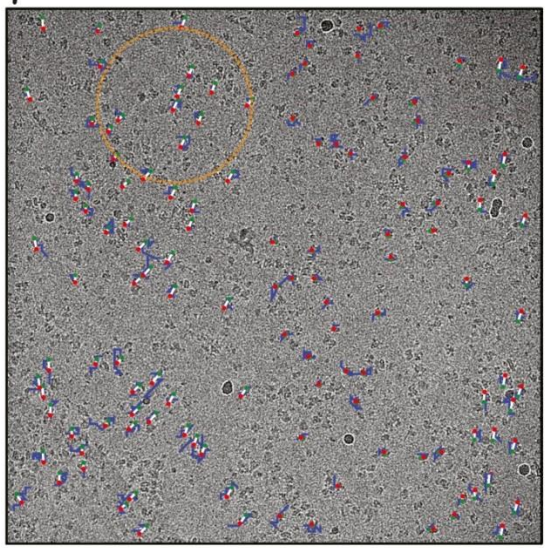
Beam Induced Movement



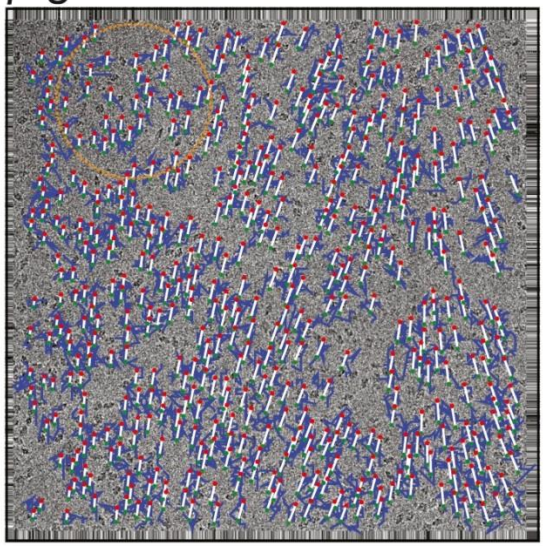
Bai et al. Trends in Biochemical Sciences, Elsevier, 2014, 40, 49-57

Beam Induced Movement

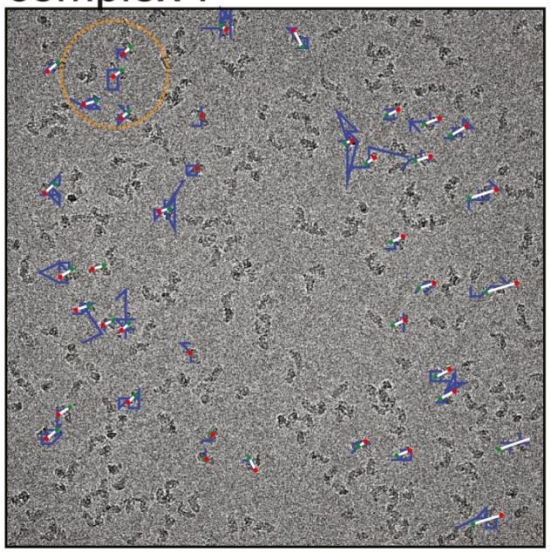
γ-secretase



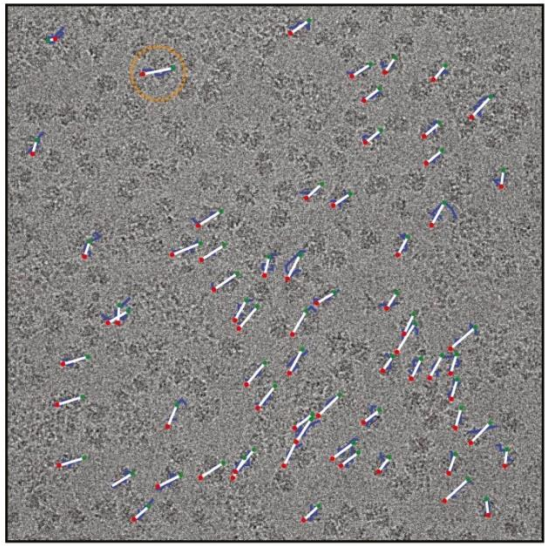
β-galactosidase



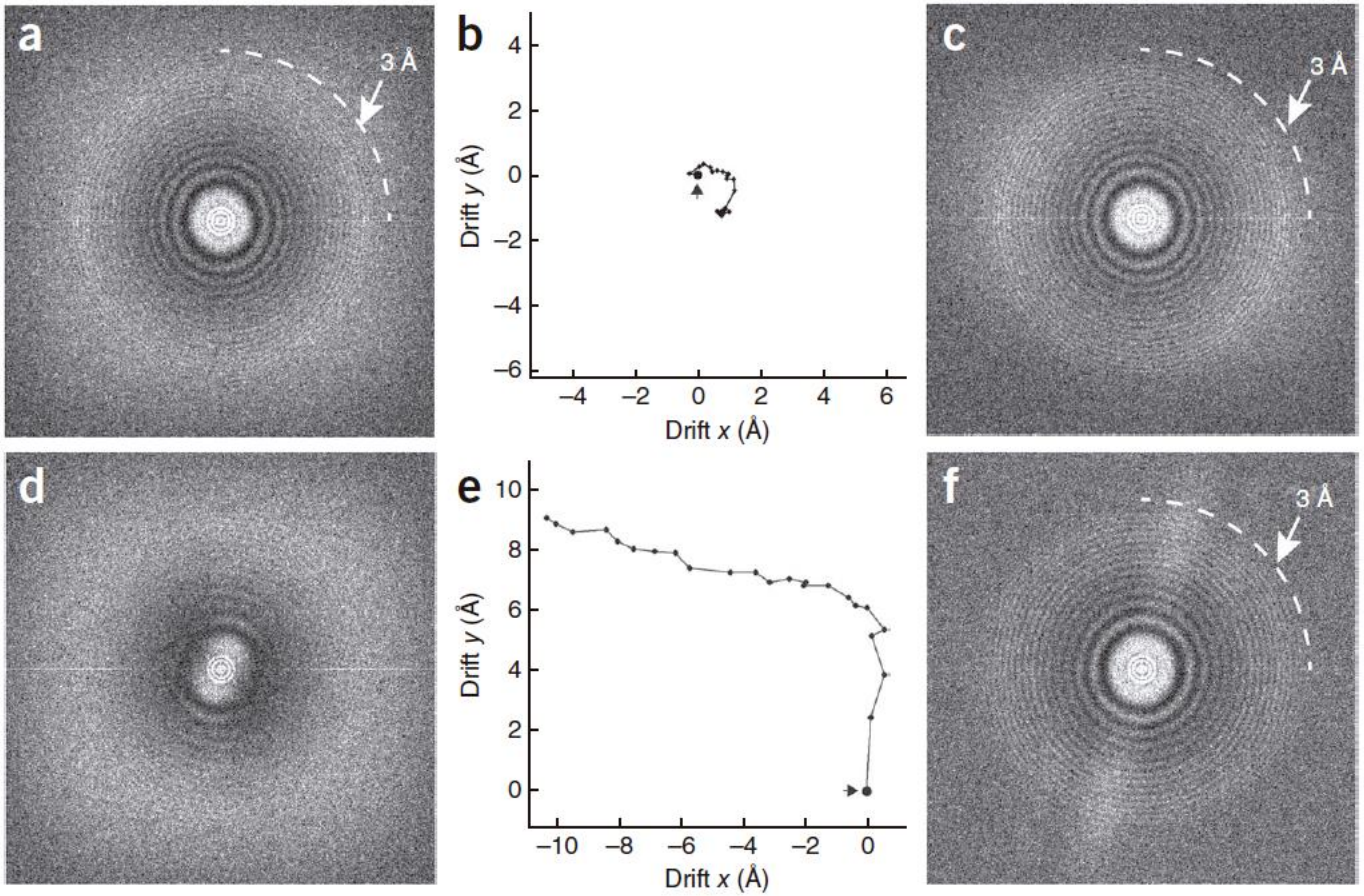
complex-I



mitoribosome

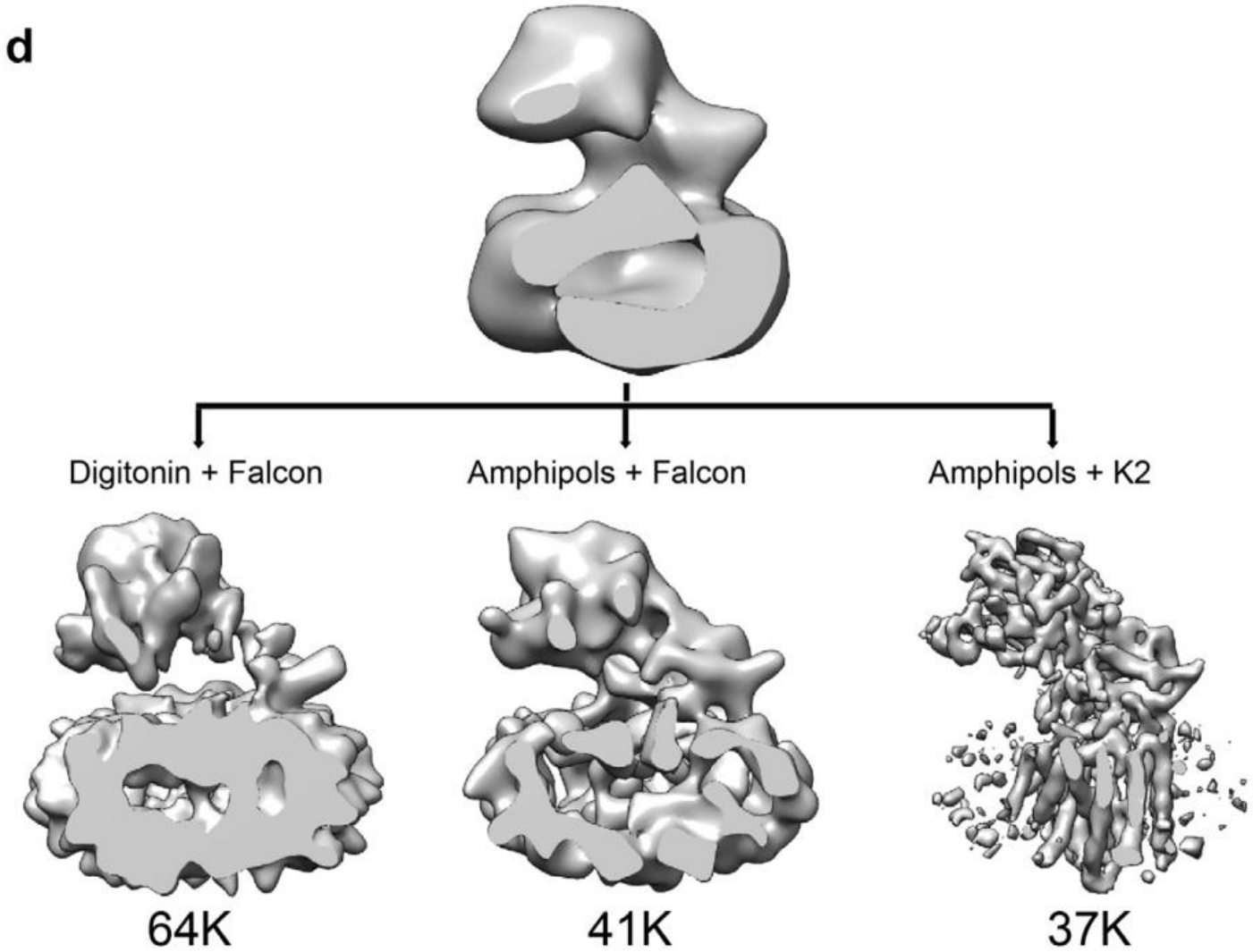


Beam Induced Movement



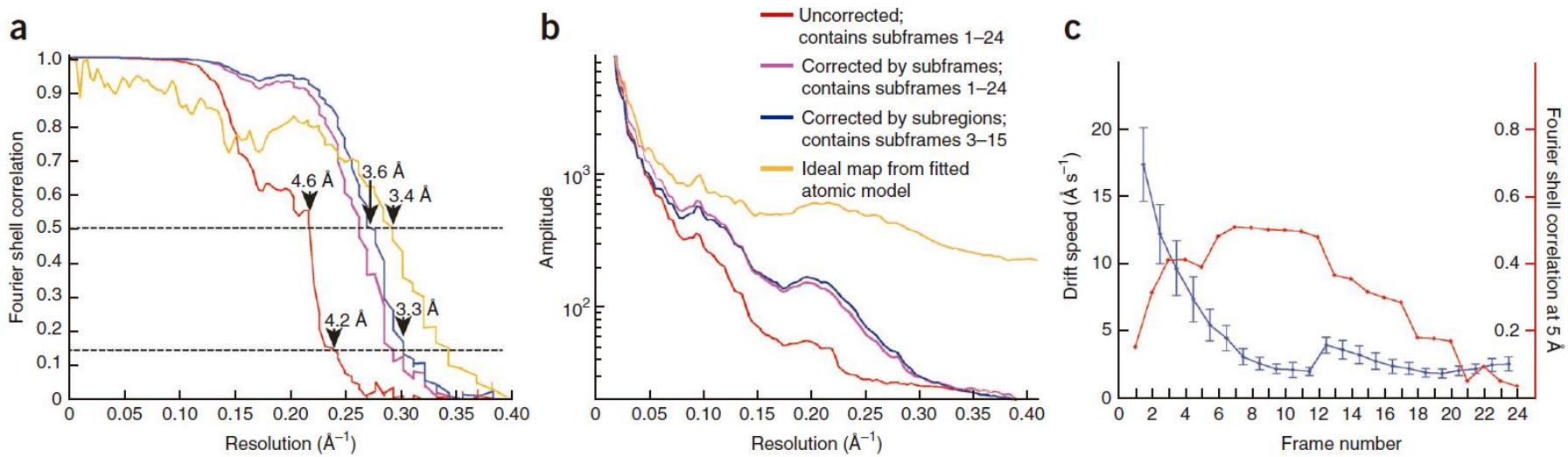
Li, X.; Mooney, P.; Zheng, S.; Booth, C. R.; Braunfeld, M. B.; Gubbens, S.; Agard, D. A. & Cheng, Y. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nature Methods, 2013

Beam Induced Movement



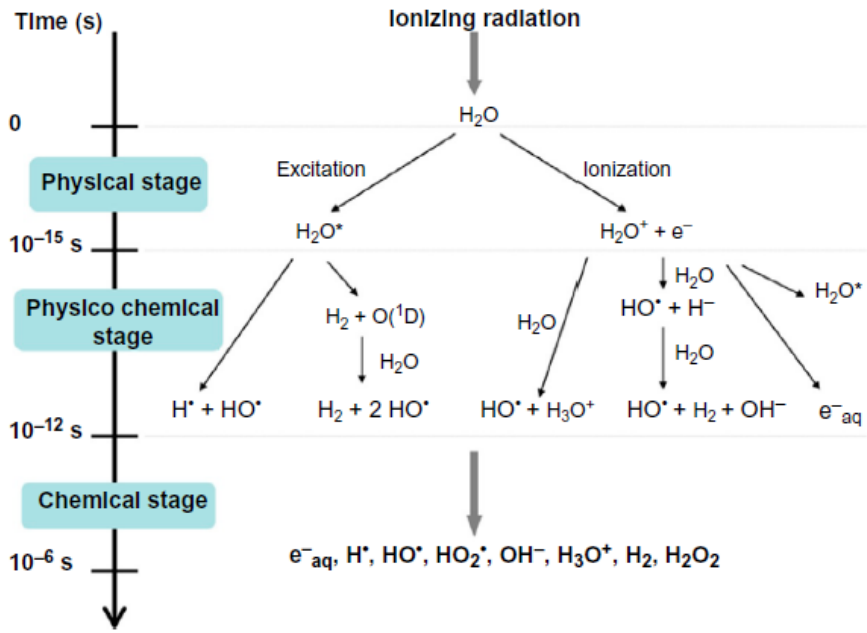
Lu et al. Nature 2014, 512, 166-170

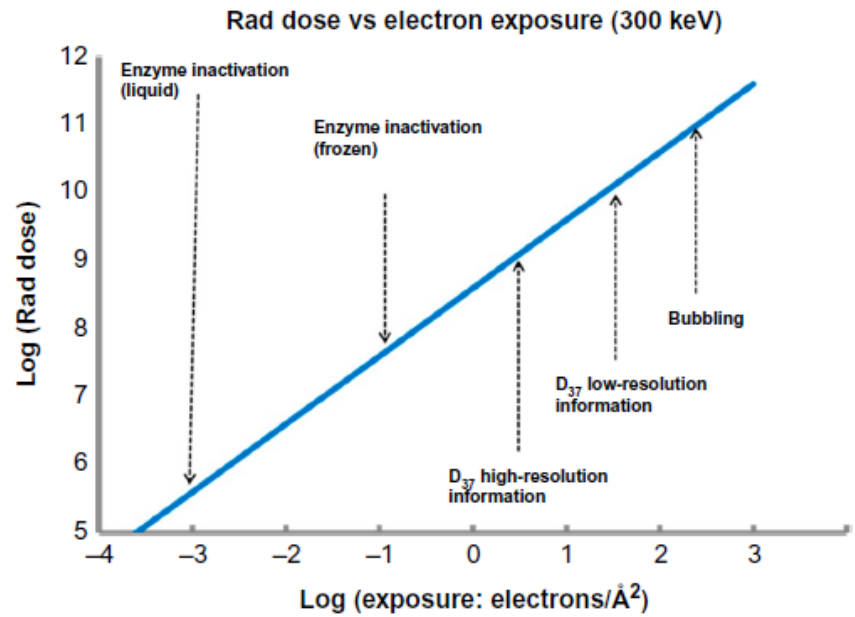
Beam Induced Movement

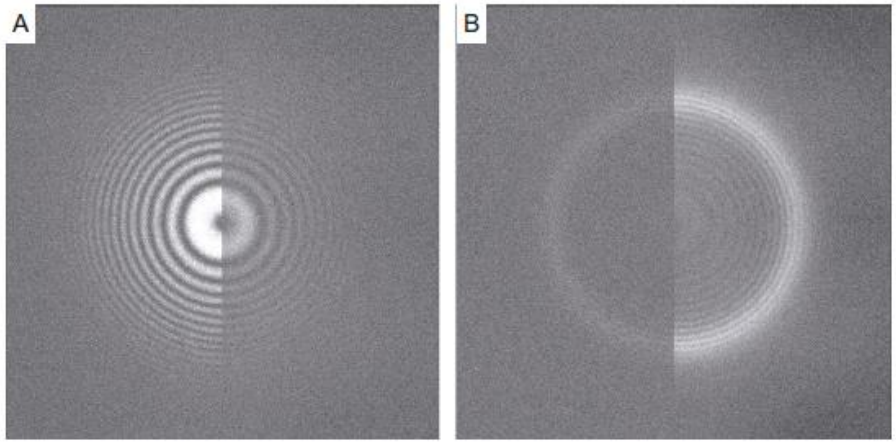


Physical reasons

- Inelastic scattering deposits energy
- Protein is damaged
- Vitreous ice is damaged
- Ice gets thinner

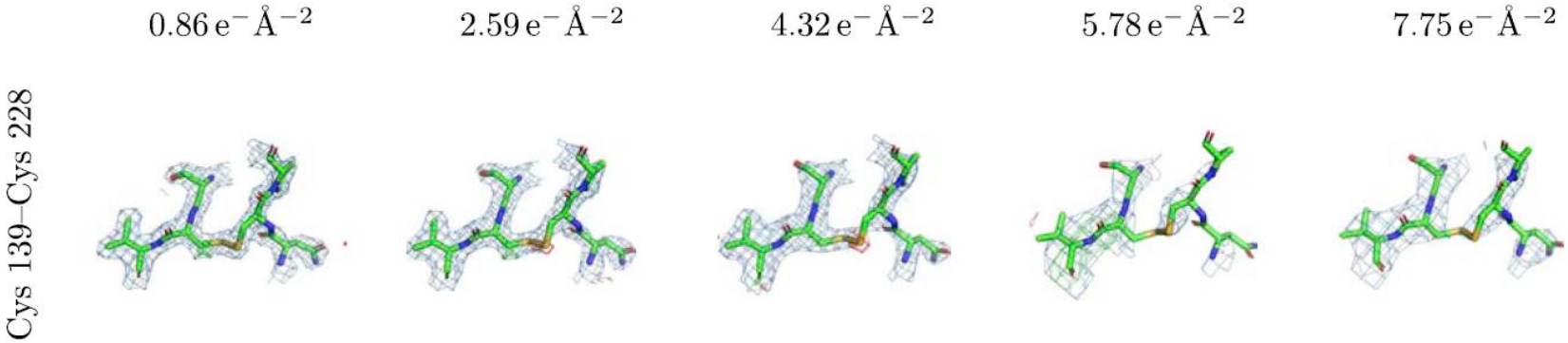


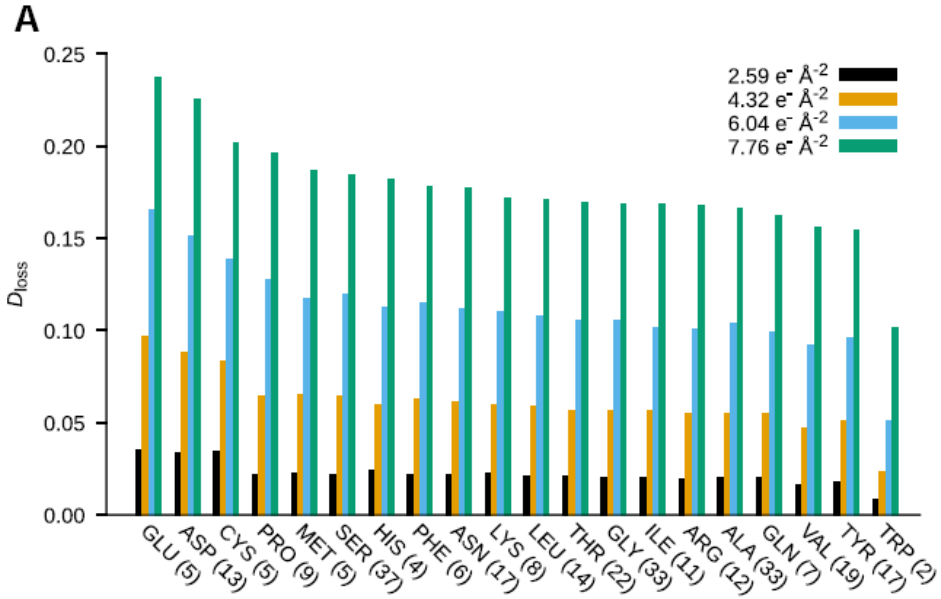




Glaeser, R. M. Specimen behavior in the electron beam. Methods in Enzymology. The Resolution Revolution: Recent Advances In cryoEM, 2016, 19-50

Radiation damage is aa dependent

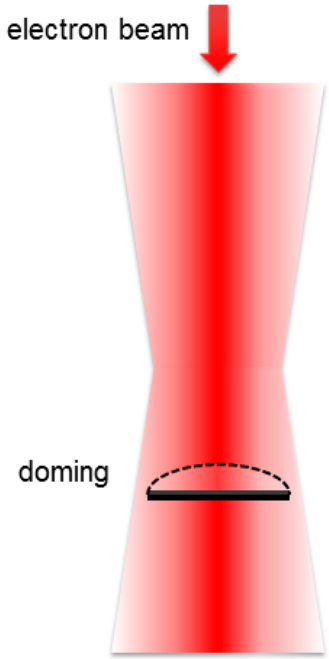


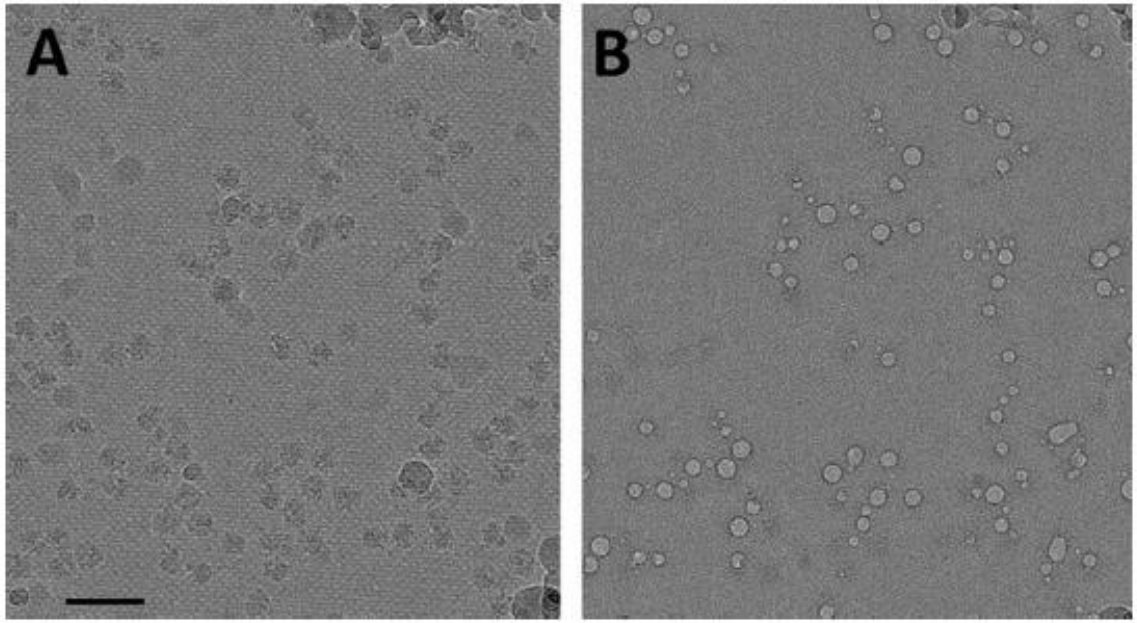


Hattne, J.; Shi, D.; Glynn, C.; Zee, C.-T.; Gallagher-Jones, M.; Martynowycz, M. W.; Rodriguez, J. A. & Gonen, T. Analysis of Global and Site-Specific Radiation Damage in Cryo-EM. Structure, 2018 , 26 , 759-766.e4

Physical reasons

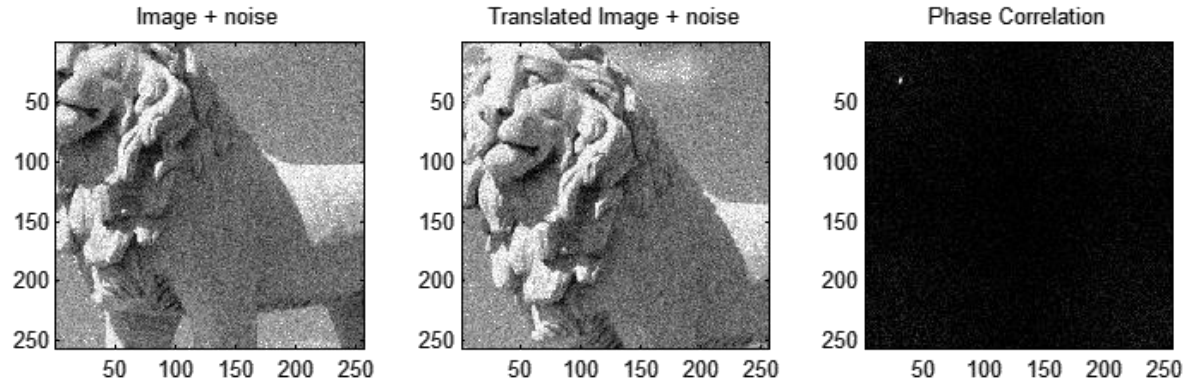
- Ice bubbles of hydrogen are formed
- Release/Creation of mechanical stress
- Charging deflects incoming electrons
- Doming





- Glaeser, R. M. Specimen behavior in the electron beam. Methods in Enzymology. The Resolution Revolution: Recent Advances In cryoEM, 2016, 19-50
- S.Zheng, E. Palovcak, J.P. Armache, Y. Cheng, D. Agard. Anisotropic Correction of Beam-induced Motion for Improved Single-particle Electron Cryo-microscopy. bioArxiv
- J. Vargas. On the physical causes of the beam-induced specimen movement in cryo-electron microscopy

MotionCorr



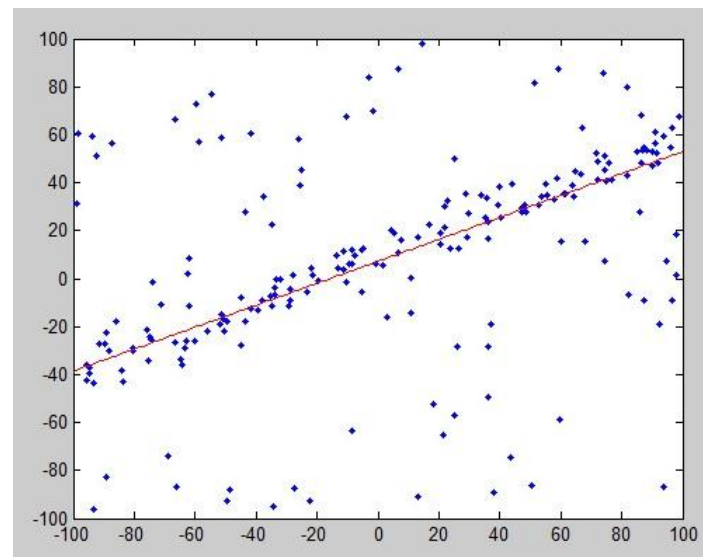
$$\mathbf{r}_i + \mathbf{r}_{i+1} + \cdots + \mathbf{r}_{j-1} = \mathbf{b}_{ij}$$

To be determined

Experimentally measured through correlation

$$A\mathbf{r} = \boldsymbol{b}$$

$$\mathbf{r}_s = (A^T A)^{-1} A^T \mathbf{b}$$

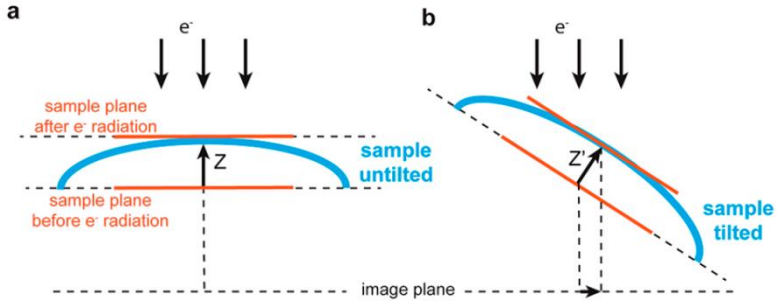


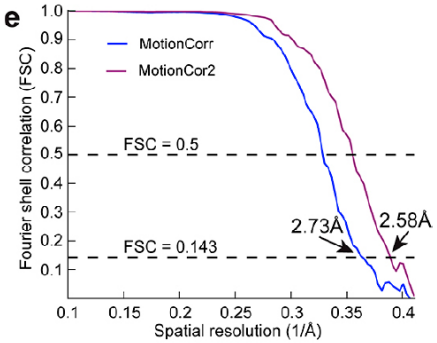
MotionCorr: Least Squares solution (GPU)
Xmipp: Outlier detection and robust solution (CPU)

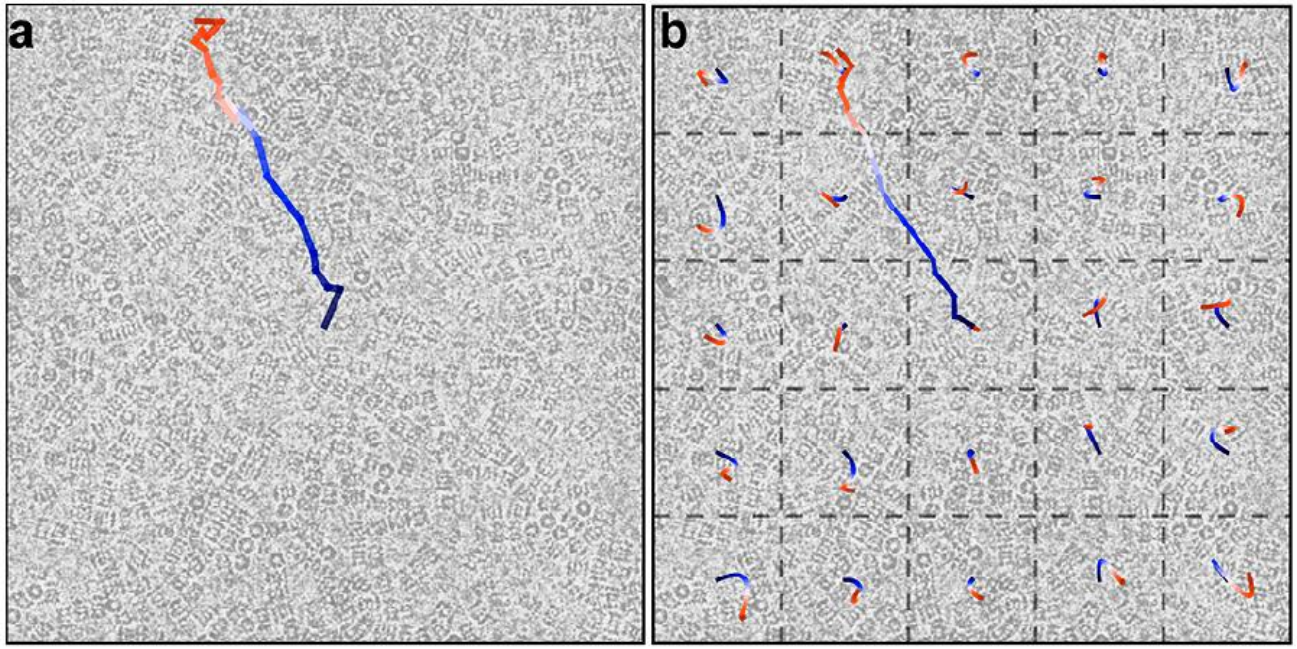
Li, X.; Mooney, P.; Zheng, S.; Booth, C. R.; Braunfeld, M. B.; Gubbens, S.; Agard, D. A. & Cheng, Y. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nature Methods, 2013

MotionCorr 2

$$S(x,y,t) = (a_0 + a_1 x + a_2 x^2 + a_3 y + a_4 y^2 + a_5 xy)\cdot(b_1 t + b_2 t^2 + b_3 t^3)$$





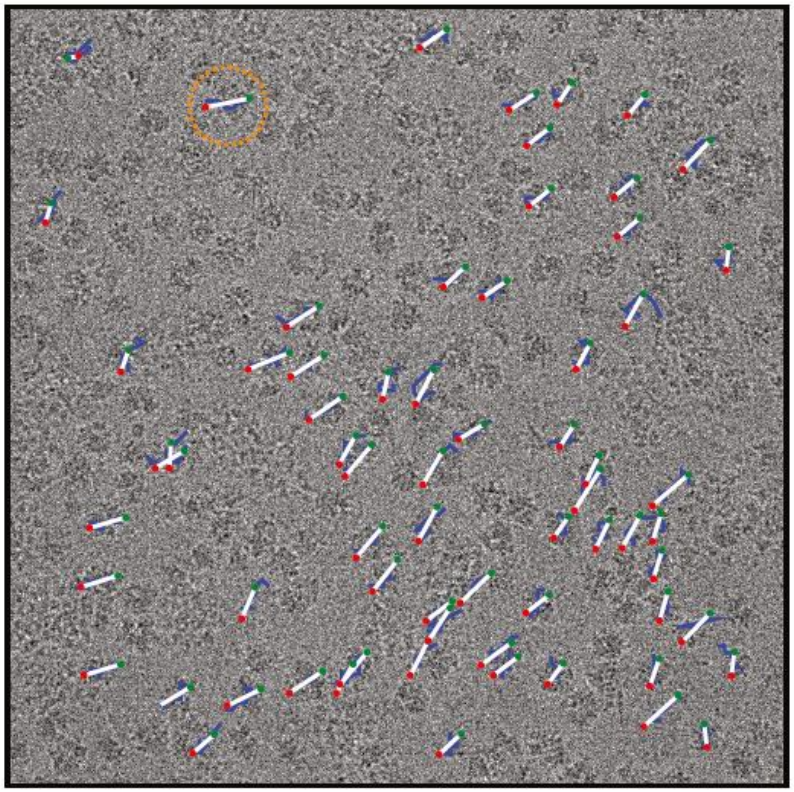


S.Zheng, E. Palovcak, J.P. Armache, Y. Cheng, D. Agard. Anisotropic Correction of Beam-induced Motion for Improved Single-particle Electron Cryo-microscopy. bioArxiv

Relion

$$\min_{\alpha_p,\beta_p}\left(\sum_{p'} w_{p'} \sum_{f} \left(\Delta x_{p'} - \alpha_p - \beta_p f\right)^2\right)$$

mitoribosome



Scheres, S. H. W. Beam-induced motion correction for sub-megadalton cryo-EM particles. Elife, 2014, 3, e0366

Optical Flow

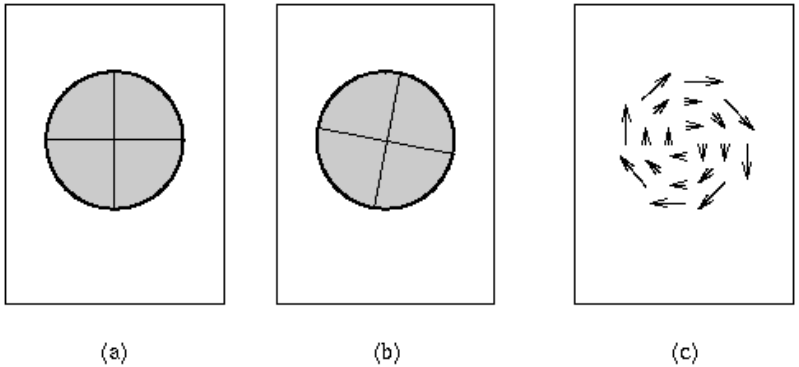
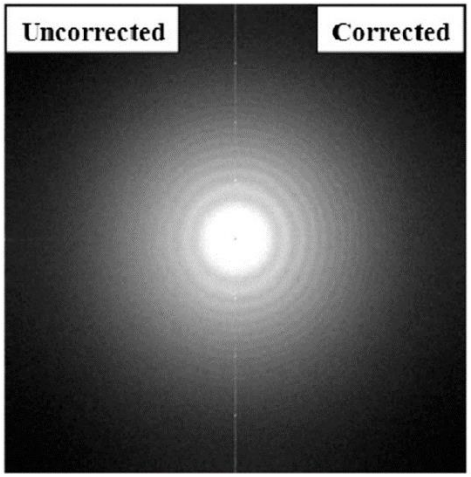
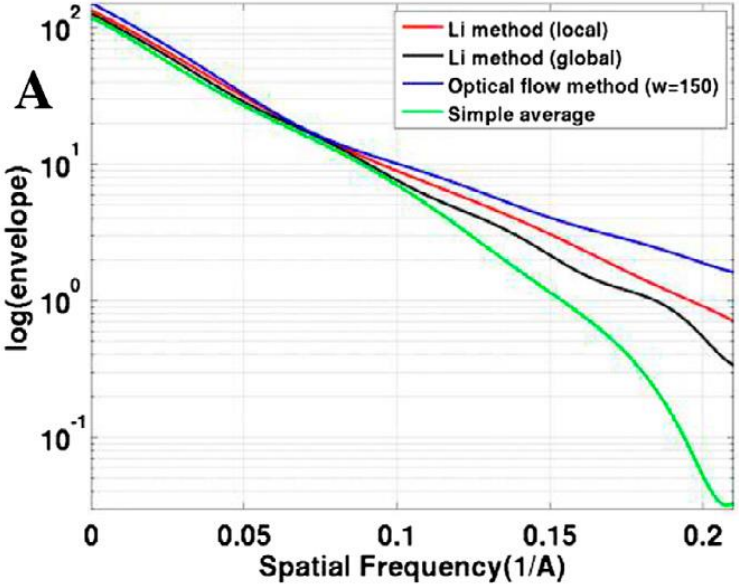
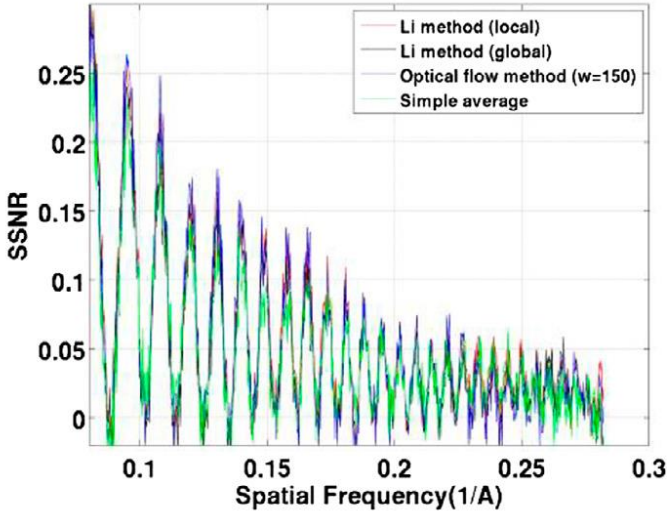


Figure 14.6 *Optical flow: (a) Time t_1, (b) time t_2, (c) optical flow.*



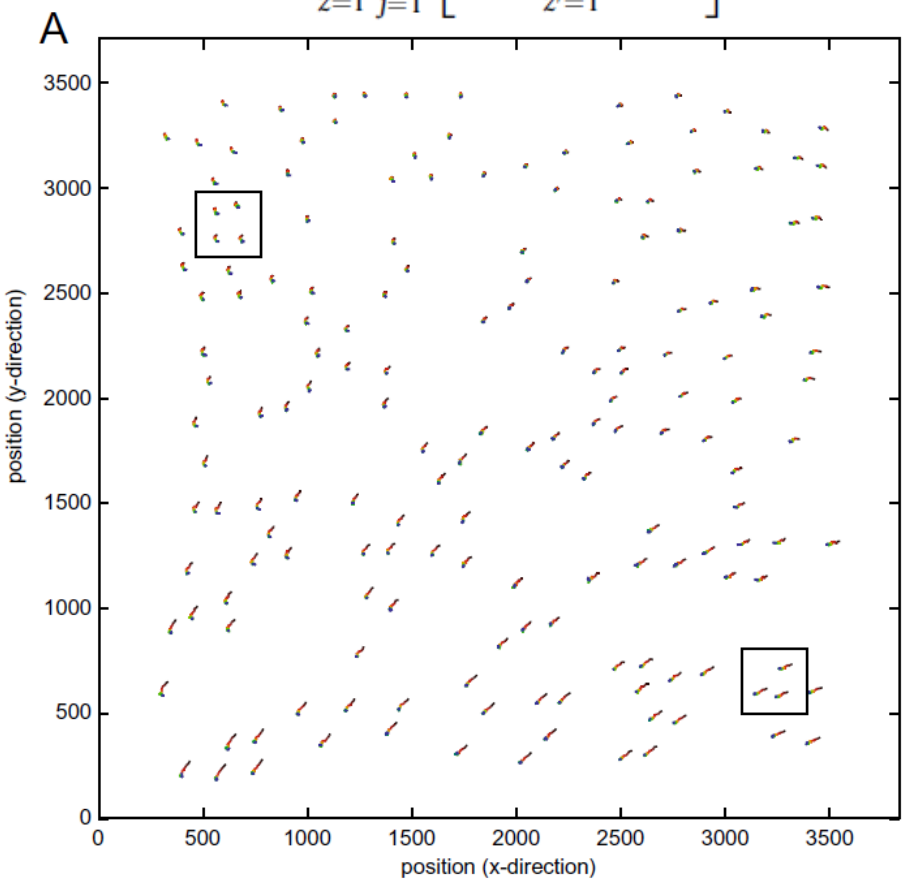


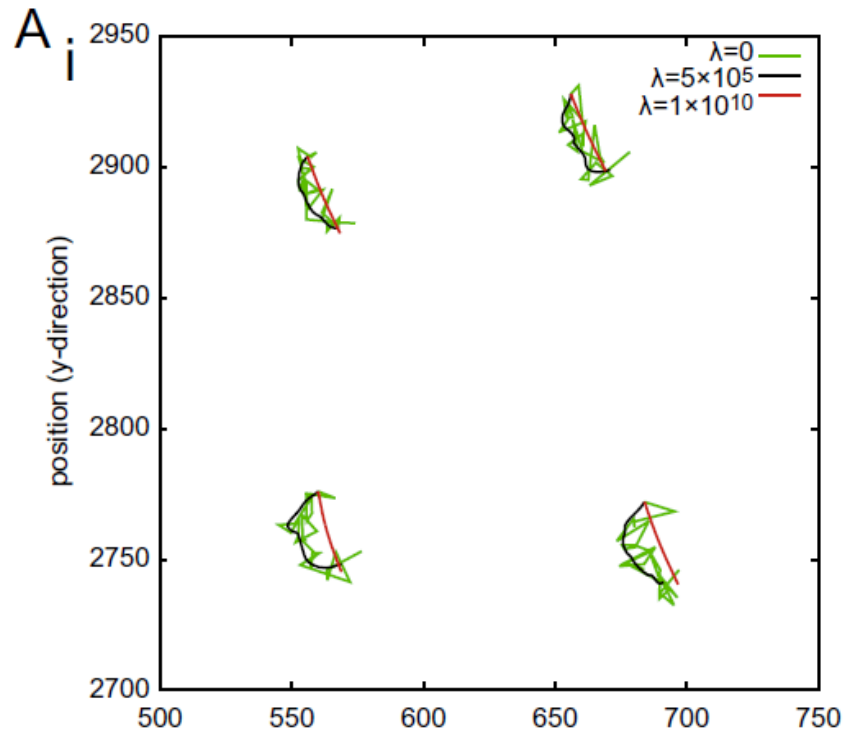


Abrishami, V.; Vargas, J.; Li, X.; Cheng, Y.; Marabini, R.; Sorzano, C. O. S. & Carazo, J. M. Alignment of direct detection device micrographs using a robust Optical Flow approach. J. Structural Biology, 2015, 189, 163-176

"Global" shift at the particle level

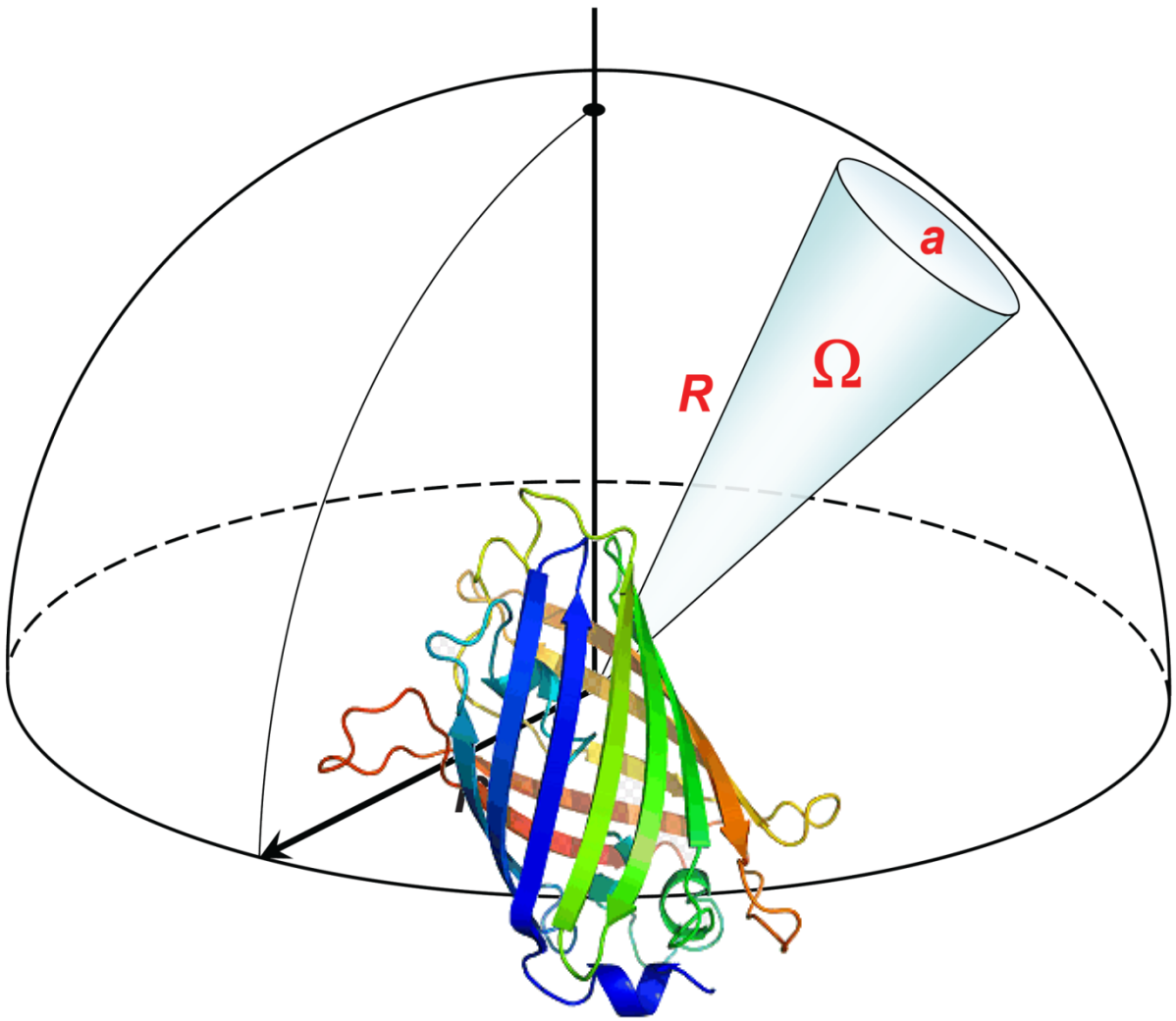
$$O(\Theta) = -Re\sum_{z=1}^{Z}\sum_{j=1}^{J}\left[F_{jz}^{*}S_{jz}^{*}\sum_{z'=1}^{Z}F_{jz'}S_{jz'}\right]$$



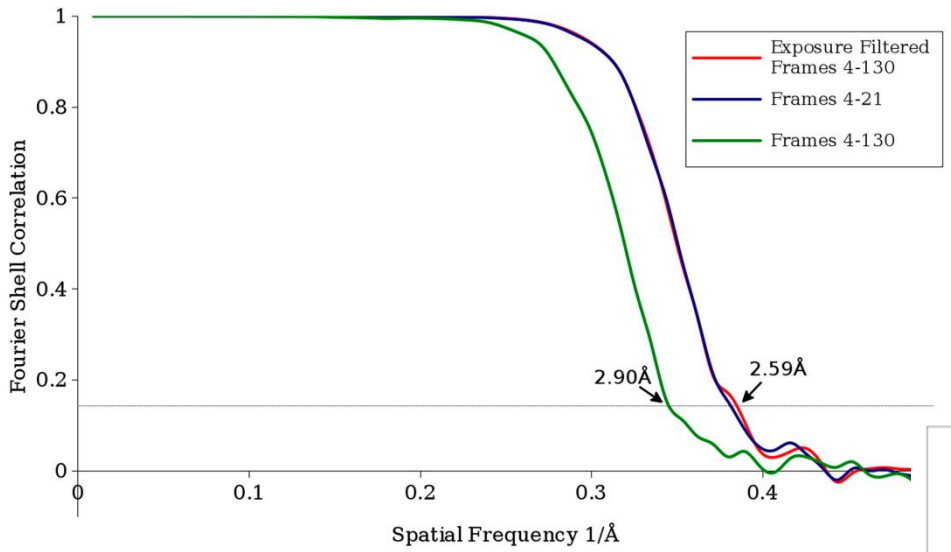


Rubinstein, J. L. & Brubaker, M. A. Alignment of cryo-EM movies of individual particles by optimization of image translations. J Struct Biol,, 2015, 192, 188-195

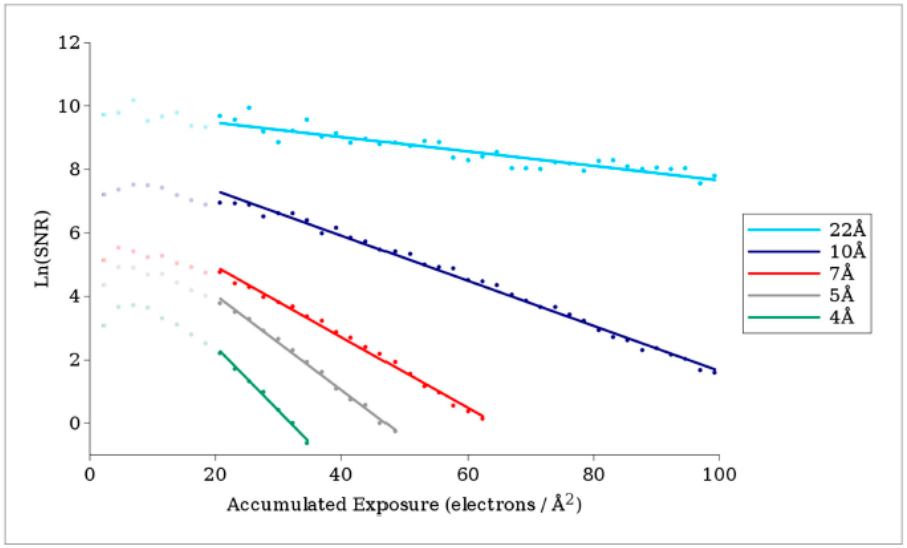
3D BIM Correction

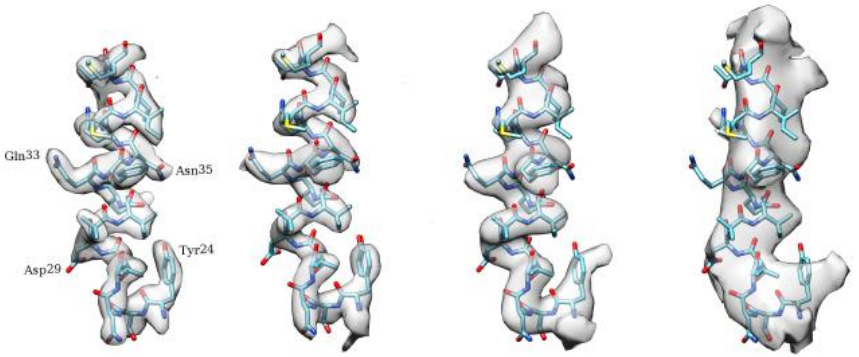


Dose compensation



$$A(k,N) = A(k,0)e^{-\frac{N}{2N_e(k)}}.$$





Frames 4-6
3.1 - 4.6 e⁻/Å²
2.9Å

Frames 25-27
19.2 - 20.8 e⁻/Å²
3.4Å

Frames 46-48
35.4 - 36.9 e⁻/Å²
4.1Å

Frames 67-69
51.5 - 53.1 e⁻/Å²
5.7Å

Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. Elife, 2015, 4

Dose compensation

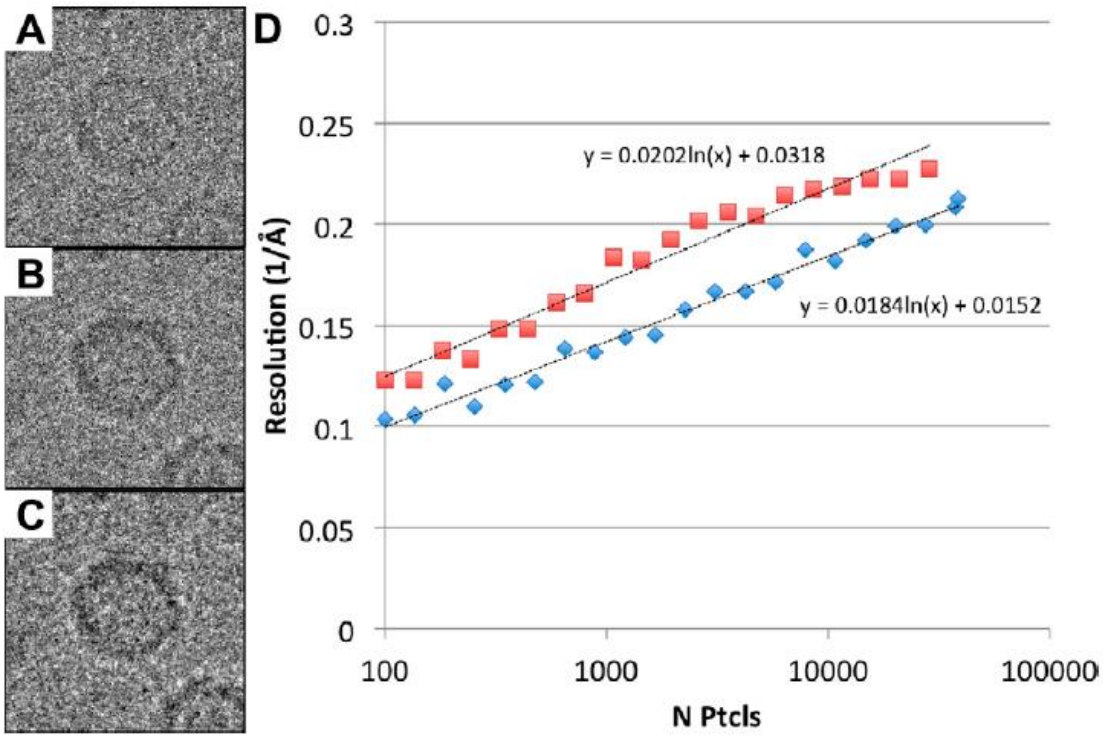
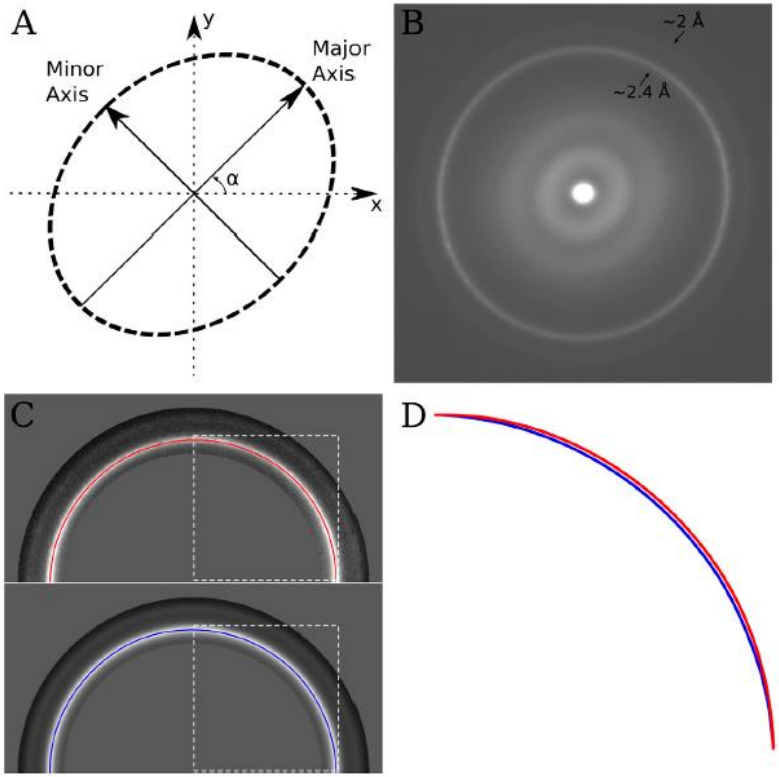
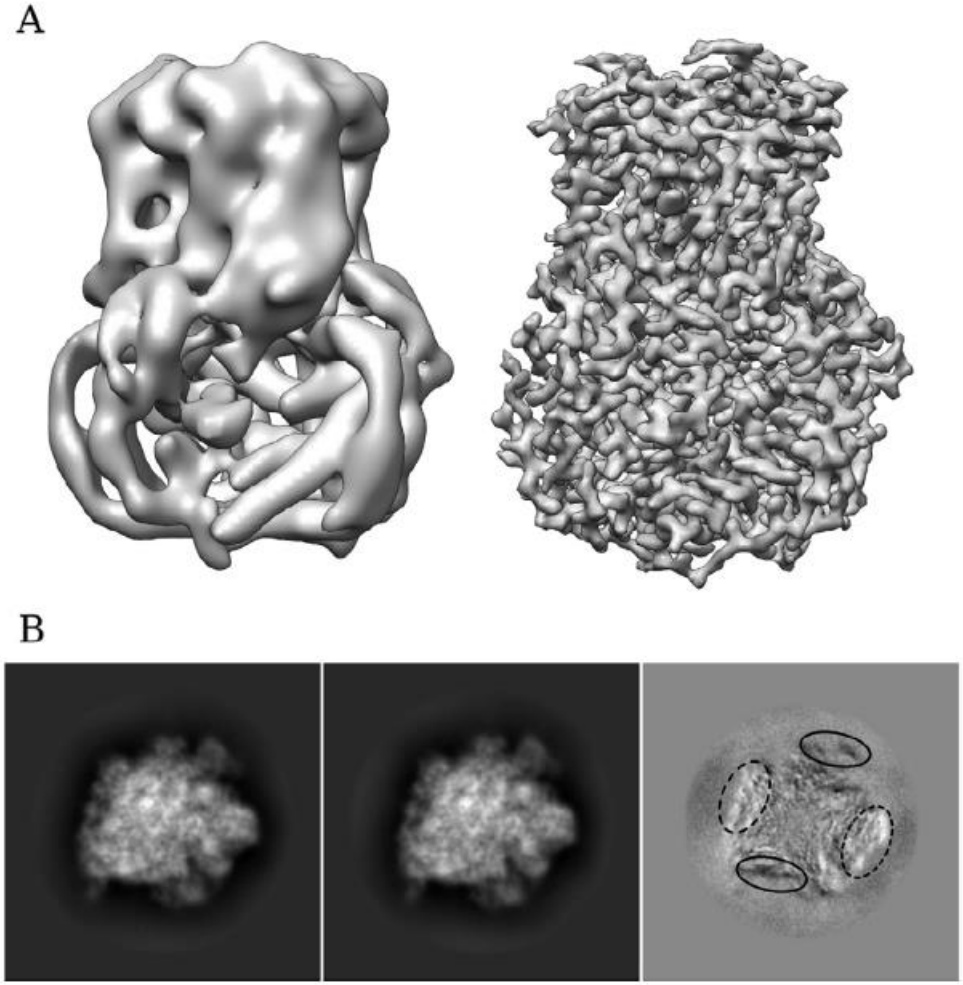


Fig. 2. Effect of dose compensation on particle contrast. (A) Aligned/uncompensated particle with 15 $e^-/Å^2$ dose. (B) Aligned/uncompensated particle with 60 $e^-/Å^2$ dose. (C) Aligned/compensated particle with 60 $e^-/Å^2$ dose. (D) ResLog plots comparing refinements with aligned/uncompensated particles (blue) with aligned/compensated particles (red).

Spear, J. M.; Noble, A. J.; Xie, Q.; Sousa, D. R.; Chapman, M. S. & Stagg, S. M. The influence of frame alignment with dose compensation on the quality of single particle reconstructions. J. Structural Biology, 2015, 192, 196-203

Anisotropic magnification

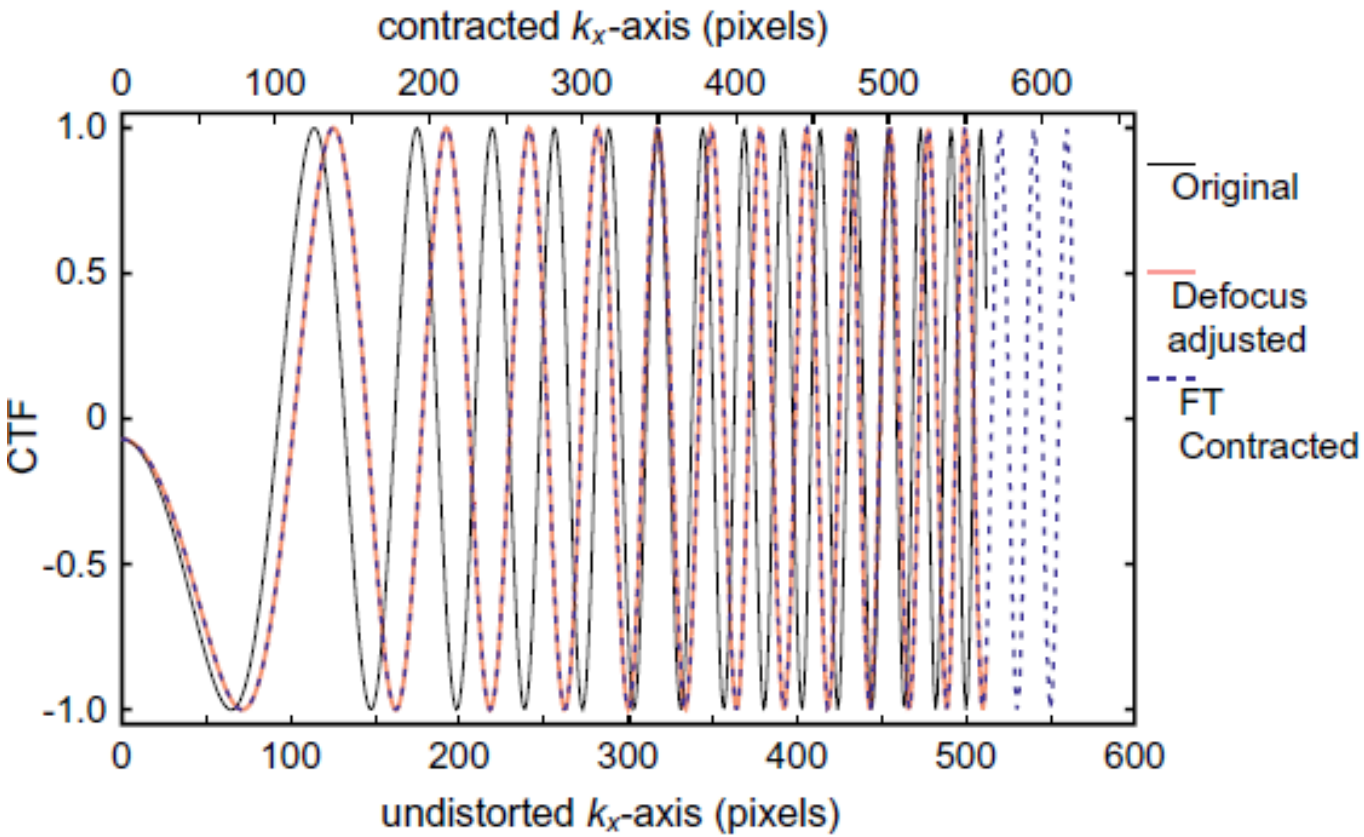




1-3% anisotropy
Experimental calibration with polycrystalline gold or Xmipp: highres

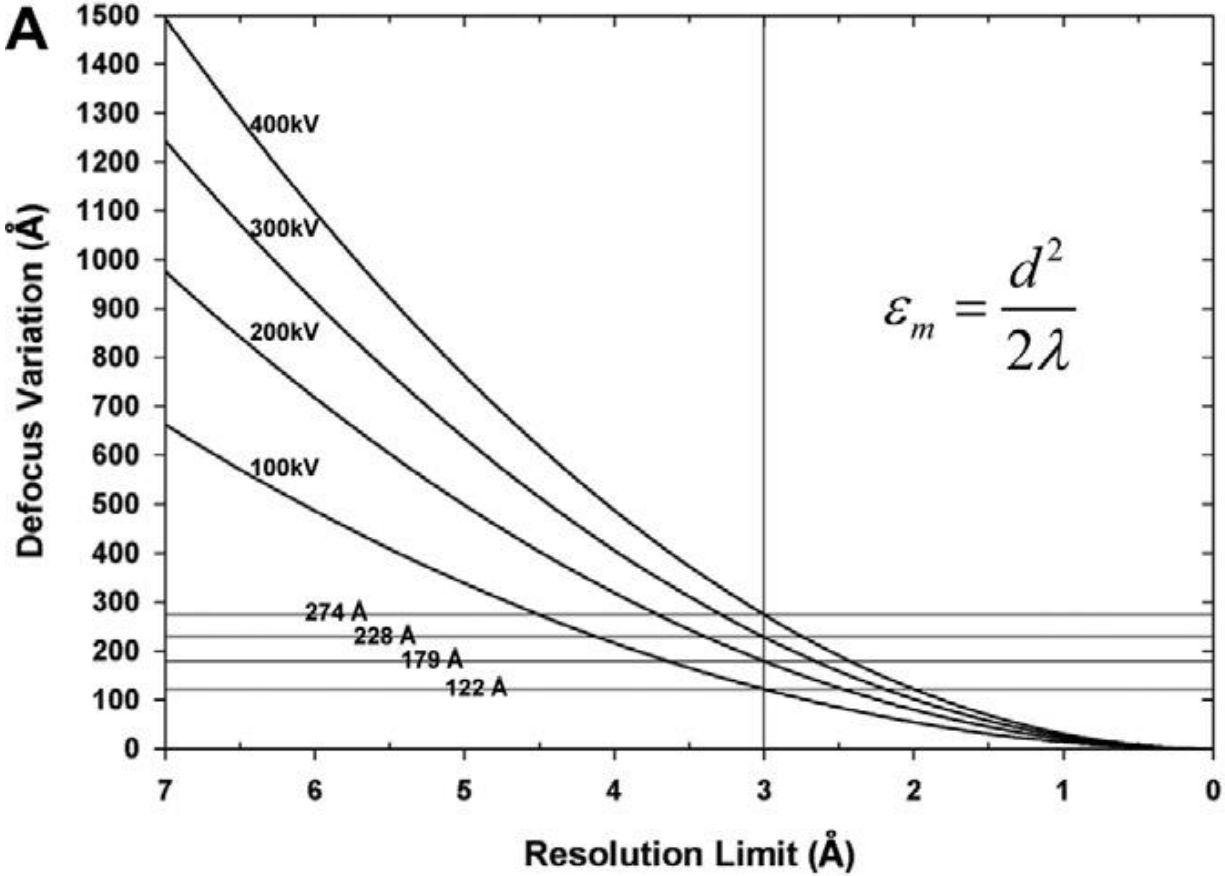
Grant, T. & Grigorieff, N. Automatic estimation and correction of anisotropic magnification distortion in electron microscopes. J. Structural Biology, 2015, 192, 204-208

Anisotropic magnification



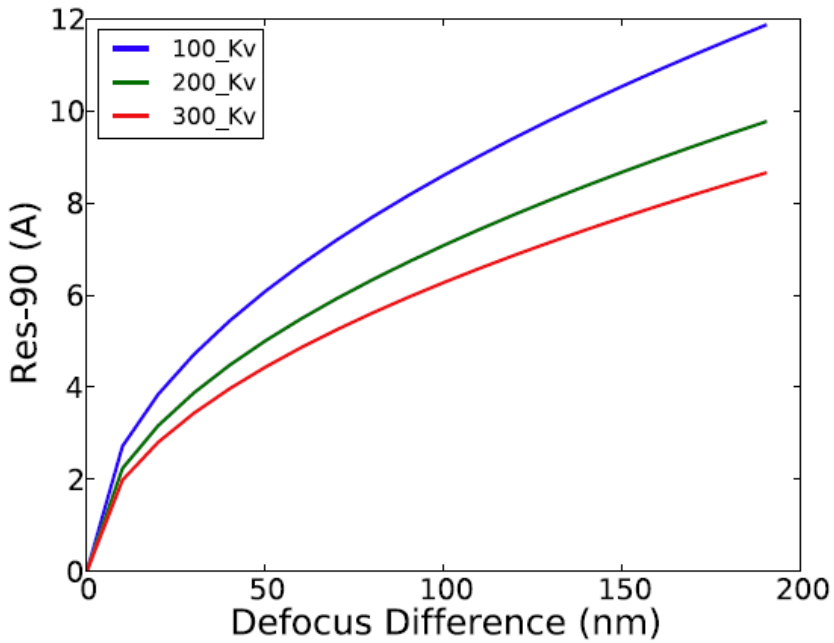
Zhao, J.; Brubaker, M. A.; Benlekbir, S. & Rubinstein, J. L. Description and comparison of algorithms for correcting anisotropic magnification in cryo-EM images. J. Structural Biology, 2015, 192, 209-215

Defocus accuracy



Zhang et al. J. Structural Biology, 2011, 175, 253-263

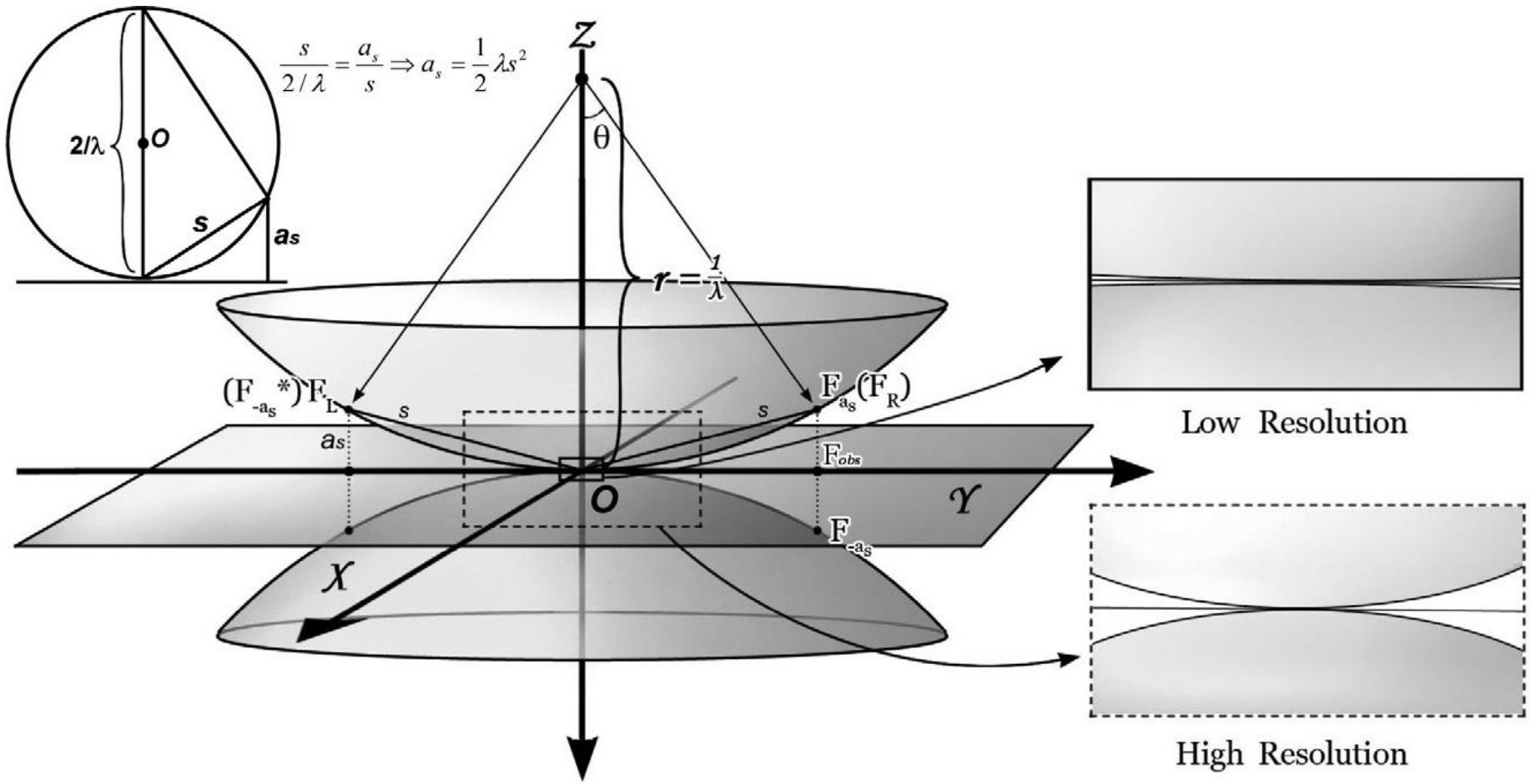
Defocus estimation



R. Marabini, B. Carragher, S. Chen, J. Chen, A. Cheng, K.H. Downing, J. Frank, R.A. Grasucci, J.B. Heymann, W. Jiang, S. Jonic, H. Liao, S.J. Ludtke, S. Patwari, A.L. Piotrowski, A. Quintana, C.O.S. Sorzano, H. Stahlberg, J. Vargas, N.R. Voss, W. Chiu, J.M. Carazo. CTF Challenge: Result summary. J. Structural Biology, 190: 348-359 (2015)

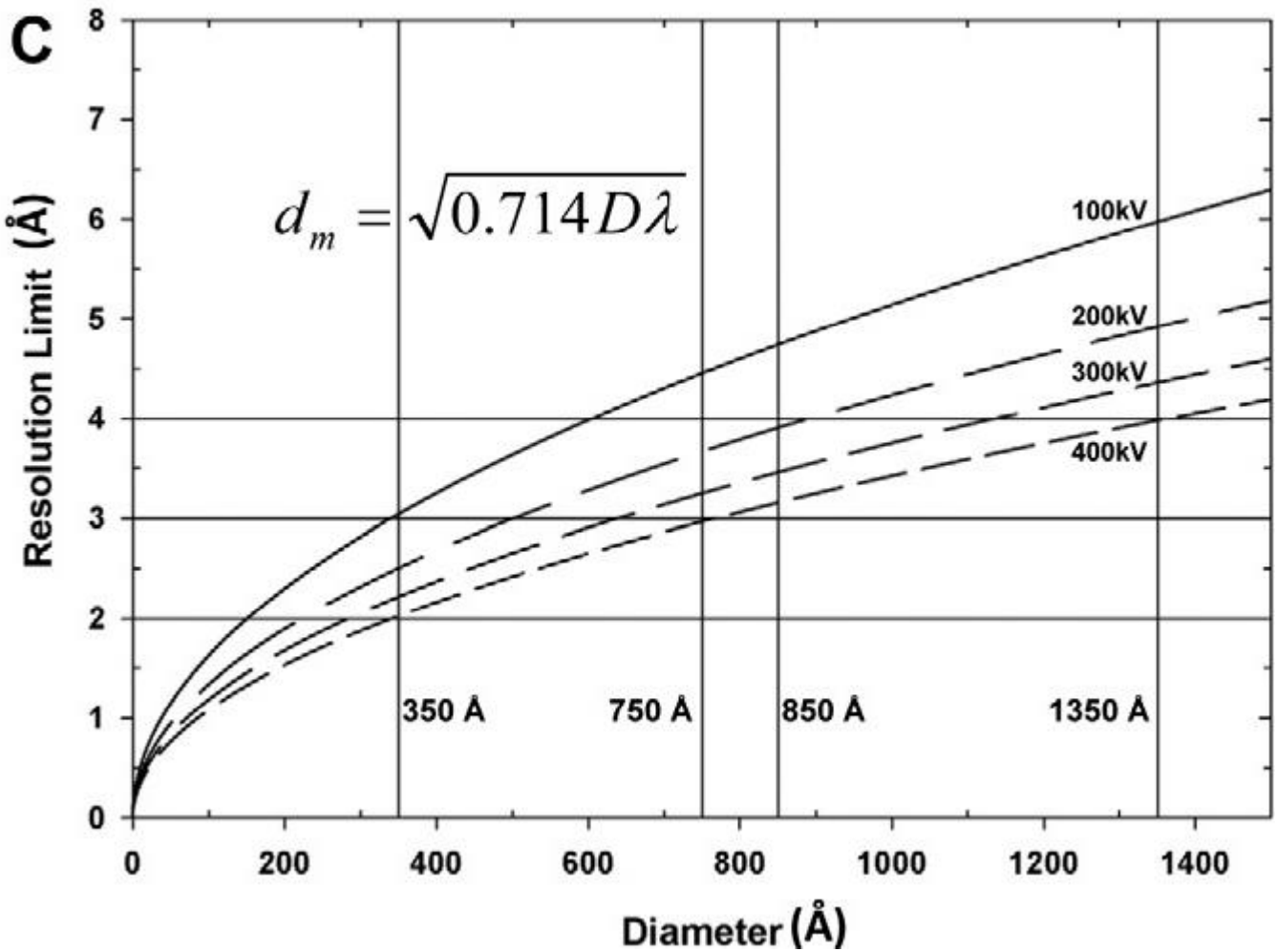
#	data sets 9	data sets 1,2&6 Pool 2	data sets 3,4,5&7 Pool 1	data sets 1, 2, 3, 4, 5, 6 & 7
1	dudelft (337)	ctffind(287)	ctffind(287)	ctffind(287)
2	xmipp(291)	appion(310)	appion(310)	appion(310)
3	particle(299)	xmipp(282)	spider(318)	xmipp(282)
4	sparx(292)	xmipp(291)	bsoft(312)	sparx(292)
5	ace(300)	sparx(292)	sparx(292)	xmipp(291)

Defocus gradient



Zhang et al. J. Structural Biology, 2011, 175, 253-263

Defocus gradient

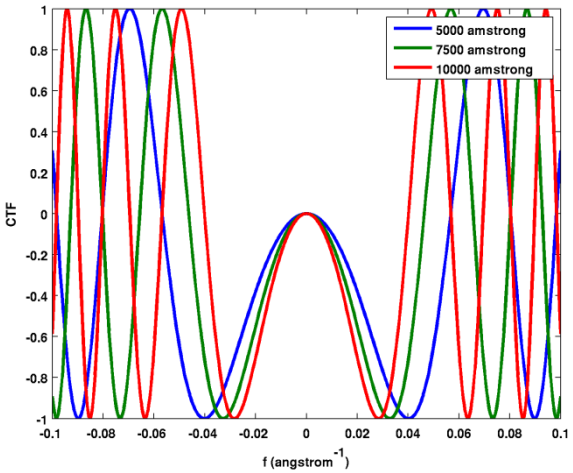


Zhang et al. J. Structural Biology, 2011, 175, 253-263

Defocus gradient

CTF Z-dependent

$$\hat{c}(\hat{\mathbf{x}}, z_l) = -k\cos\left(-2\pi\left(\frac{C_s\lambda^3 s^4}{4} - \frac{z_l\lambda s^2}{2}\right) - \cos^{-1} Q\right)$$

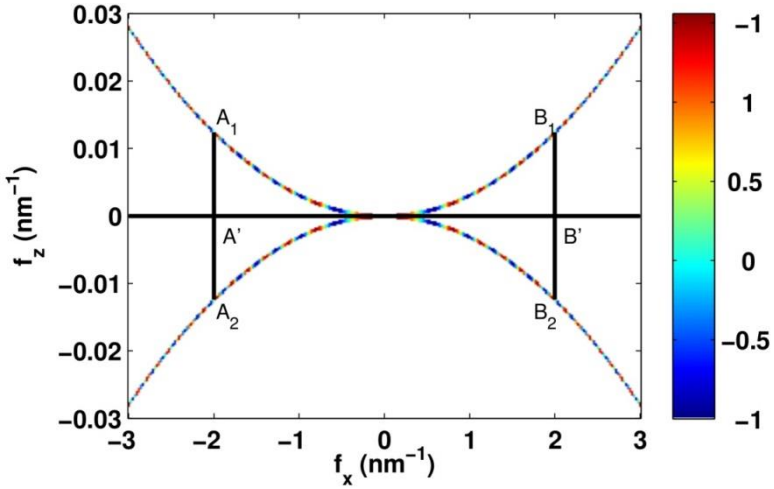


3DCTF

$$\hat{c}_z(\hat{\mathbf{x}}, \hat{z}) = \frac{-k}{2}\left(e^{-2\pi\left(\frac{C_s\lambda^3 s^4}{4}\right) - \cos^{-1} Q}\delta\left(\hat{z} - \frac{\lambda s^2}{2}\right) + e^{2\pi\left(\frac{C_s\lambda^3 s^4}{4}\right) + \cos^{-1} Q}\delta\left(\hat{z} + \frac{\lambda s^2}{2}\right)\right)$$

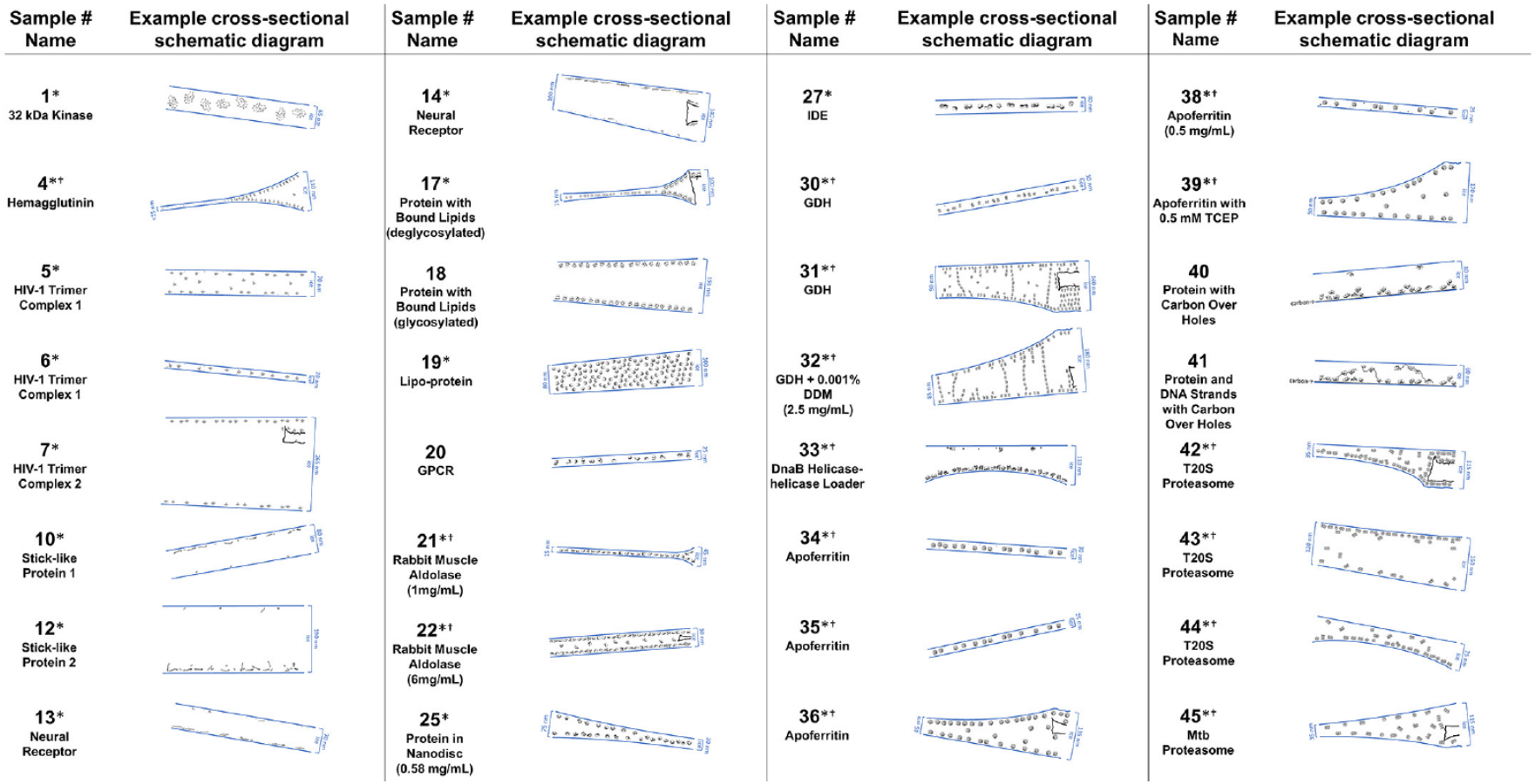
Projection

$$\hat{p}(\hat{\mathbf{x}}, z_l) = \frac{-k}{2}\left(e^{-j\left(2\pi\left(\frac{C_s\lambda^3 s^4}{4} - \frac{z_l\lambda s^2}{2}\right) - \cos^{-1} Q\right)}\hat{v}\left(\hat{\mathbf{x}}, \hat{z} - \frac{\lambda s^2}{2}\right) + e^{j\left(2\pi\left(\frac{C_s\lambda^3 s^4}{4} + \frac{z_l\lambda s^2}{2}\right) + \cos^{-1} Q\right)}\hat{v}\left(\hat{\mathbf{x}}, \hat{z} + \frac{\lambda s^2}{2}\right)\right)$$



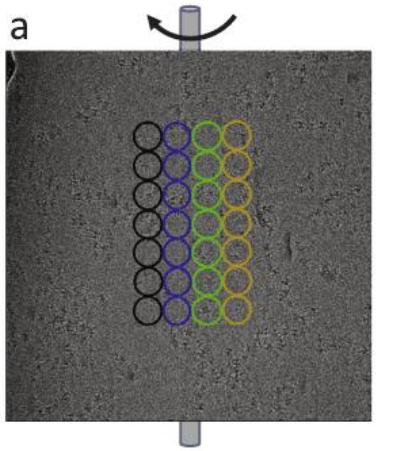
Otón, …, Sorzano, … J Struct Biol, 2012, 178, 29-37

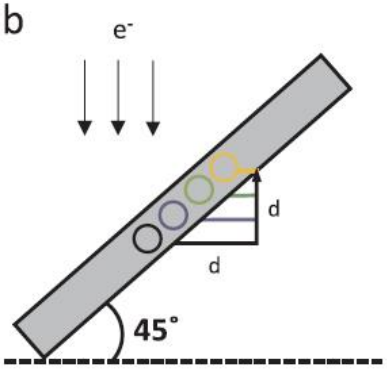
Defocus gradient

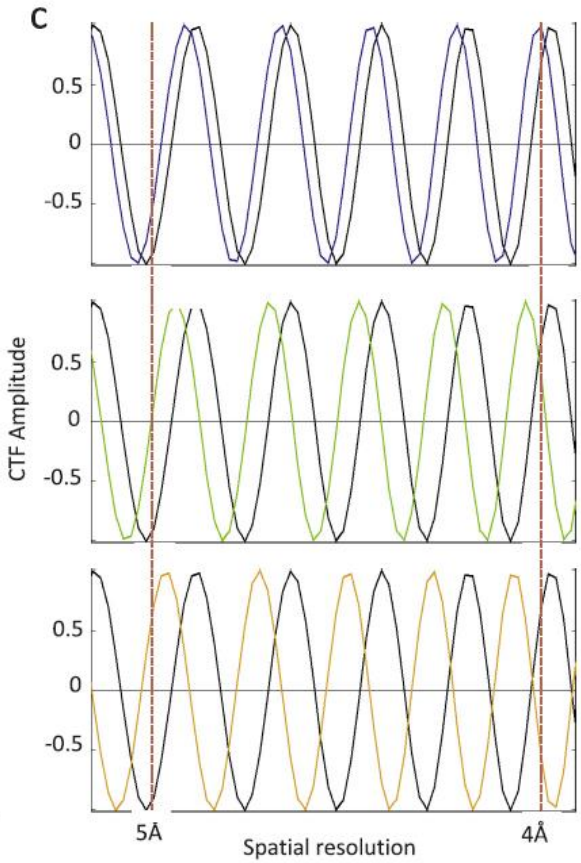


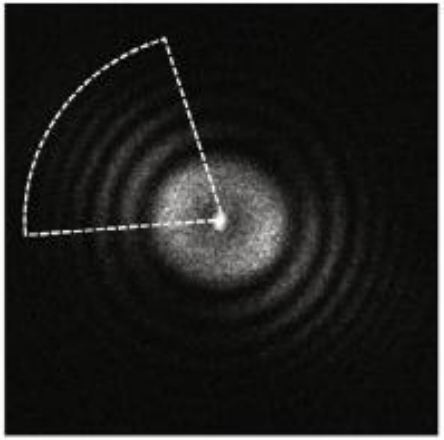
Noble, A. J.; Dandey, V. P.; Wei, H.; Brasch, J.; Chase, J.; Acharya, P.; Tan, Y. Z.; Zhang, Z.; Kim, L. Y.; Scapin, G.; Rapp, M.; Eng, E. T.; Rice, W. J.; Cheng, A.; Negro, C. J.; Shapiro, L.; Kwong, P. D.; Jeruzalmi, D.; des Georges, A.; Potter, C. S. and Carragher, B. Routine single particle CryoEM sample and grid characterization by tomography. eLife, 2018, 7

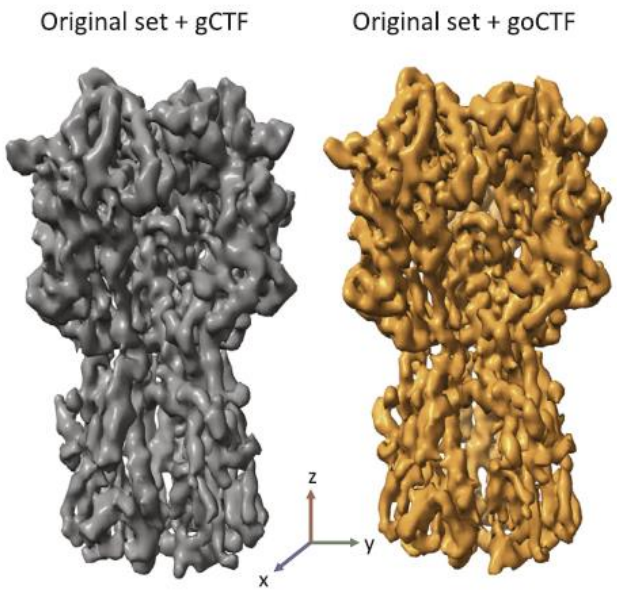
Defocus gradient





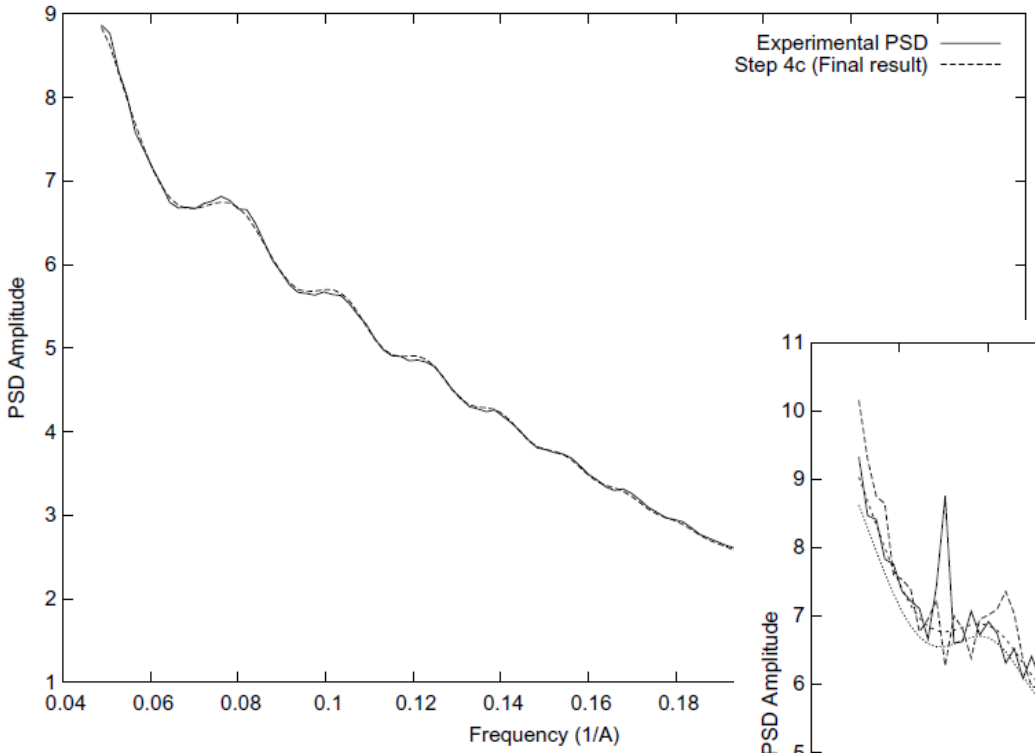


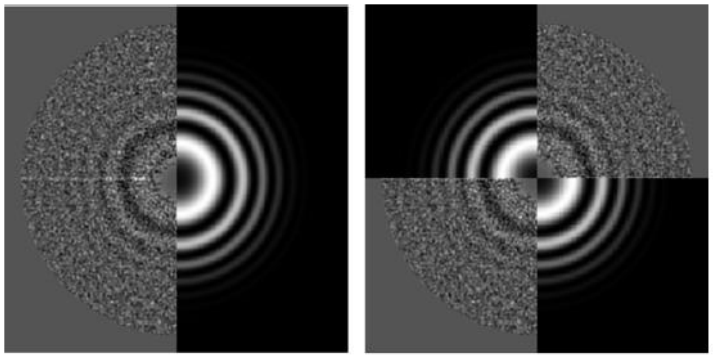


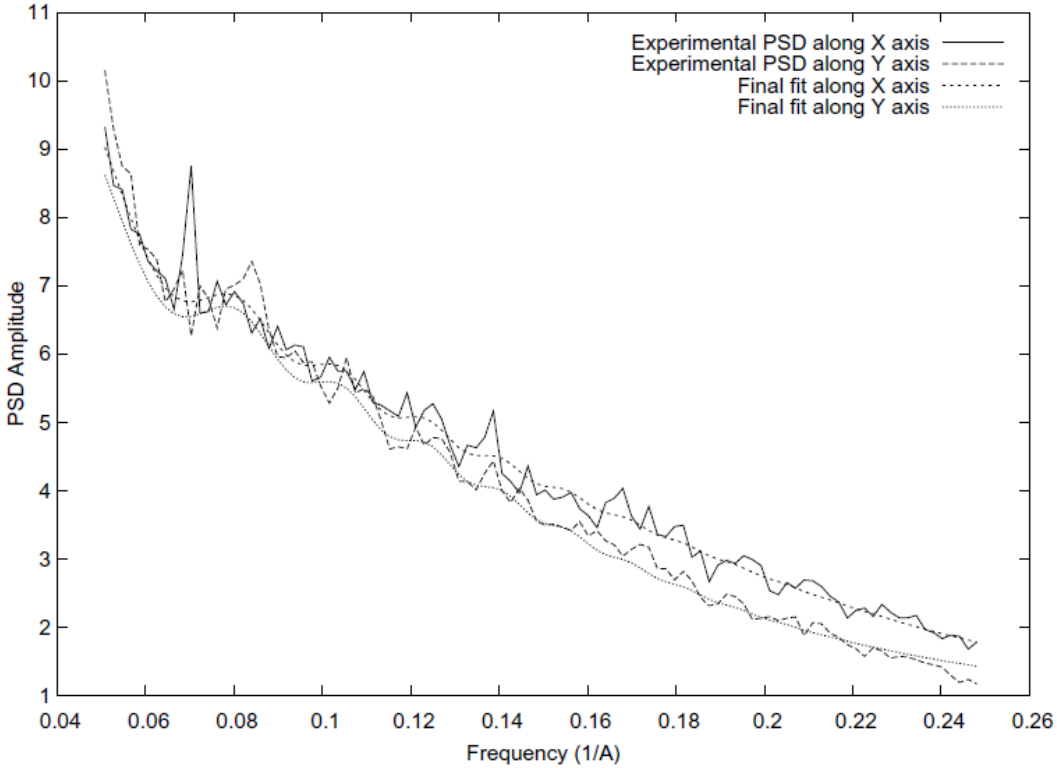


Su, M. goCTF: Geometrically optimized CTF determination for single-particle cryo-EM. Journal of structural biology, 2019, 205: 22-29

Envelope and background estimation

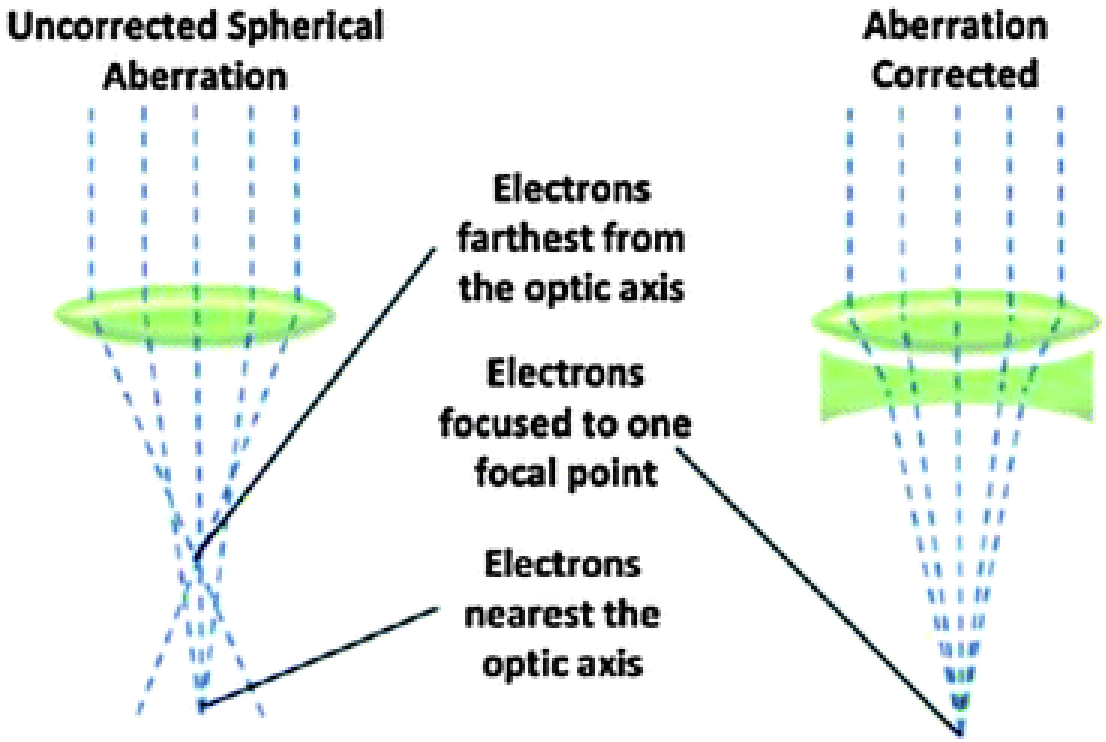


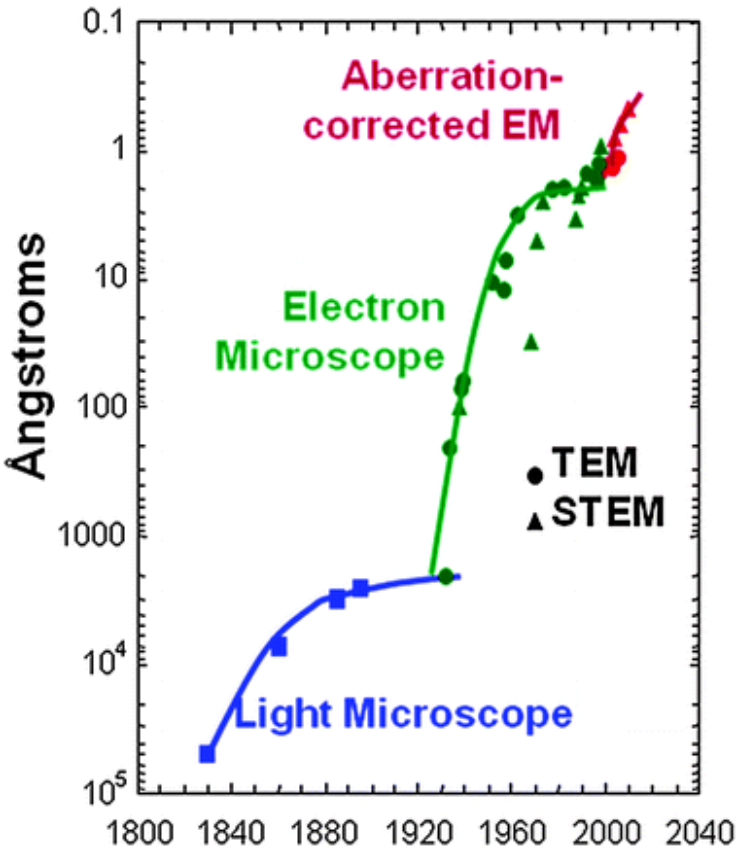




C.O.S. Sorzano, S. Jonic, R. Núñez, N. Boisset, J.M. Carazo. Fast, robust and accurate determination of transmission electron microscopy contrast transfer function. Journal Structural Biology 160: 249-262 (2007)

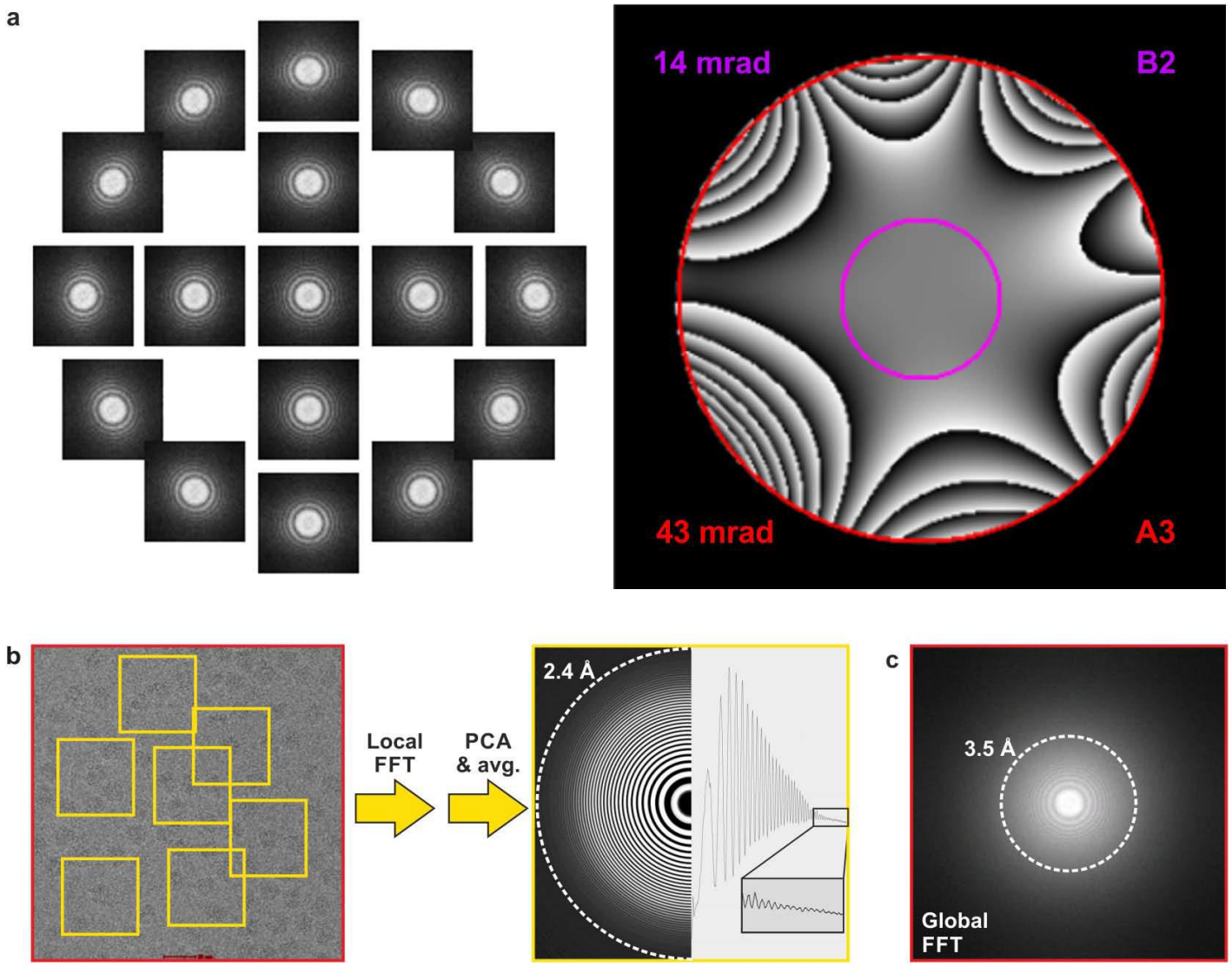
Cs Correction





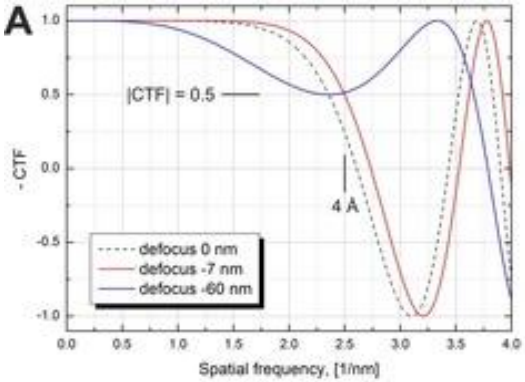
Yang, et al. Chem. Soc. Rev., 2012, 41, 8179–8194

Cs Correction and Local CTFs

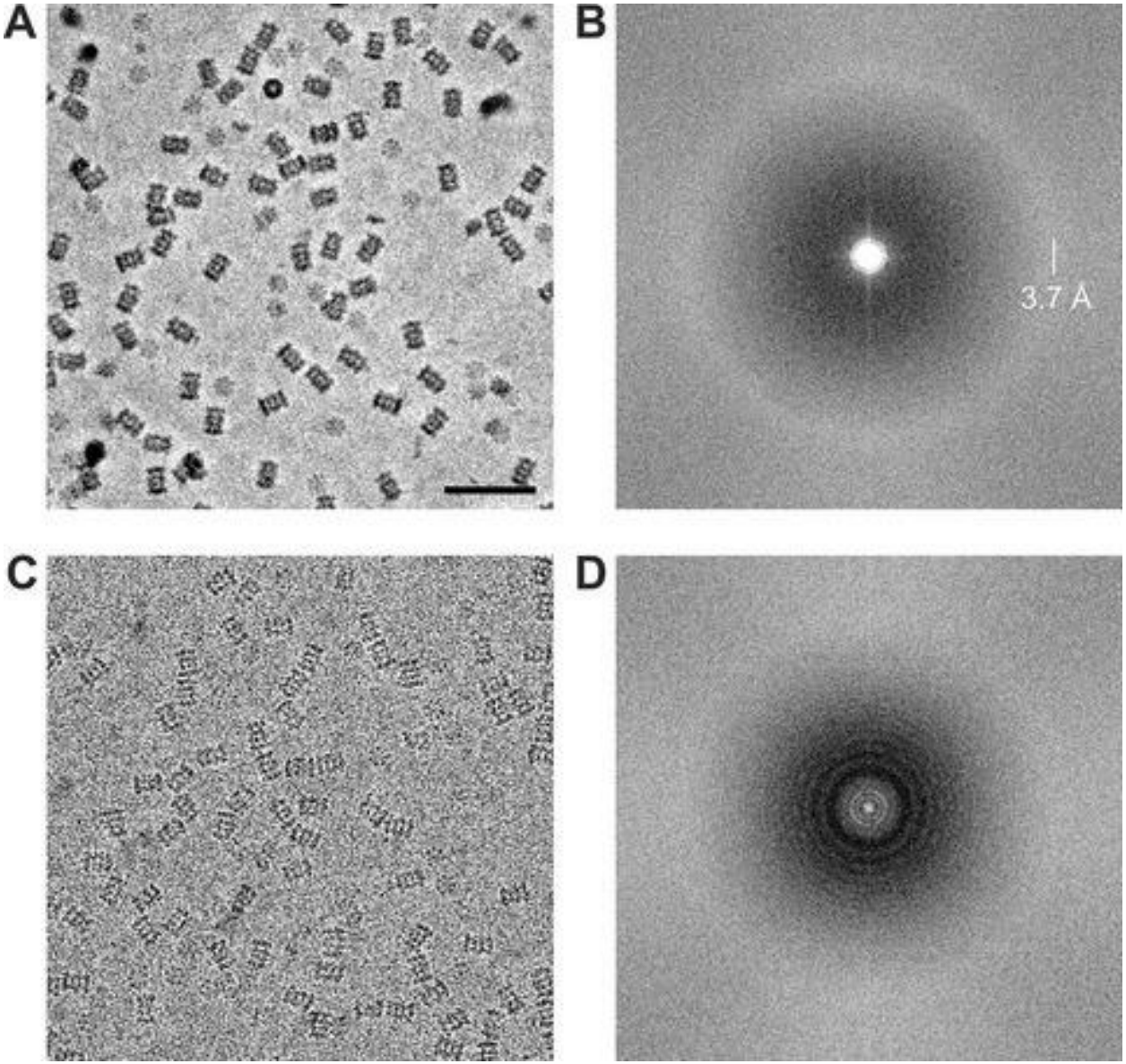


Fischer et al. Nature, (2015) 23 Feb

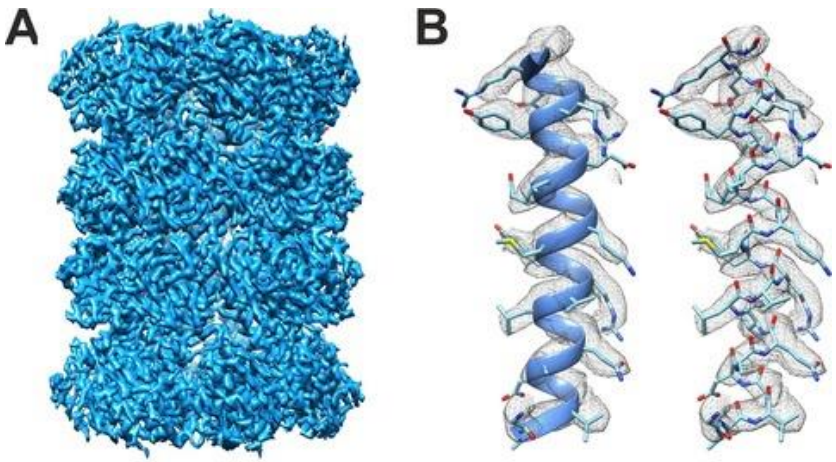
Phase plates

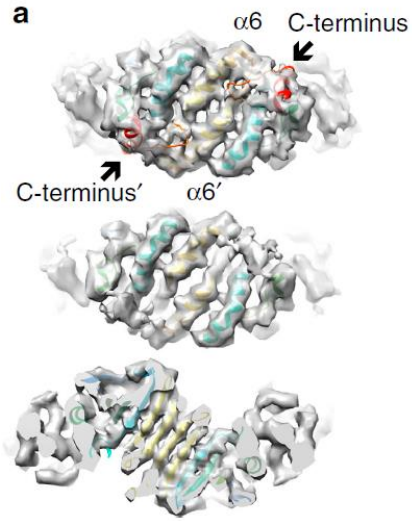


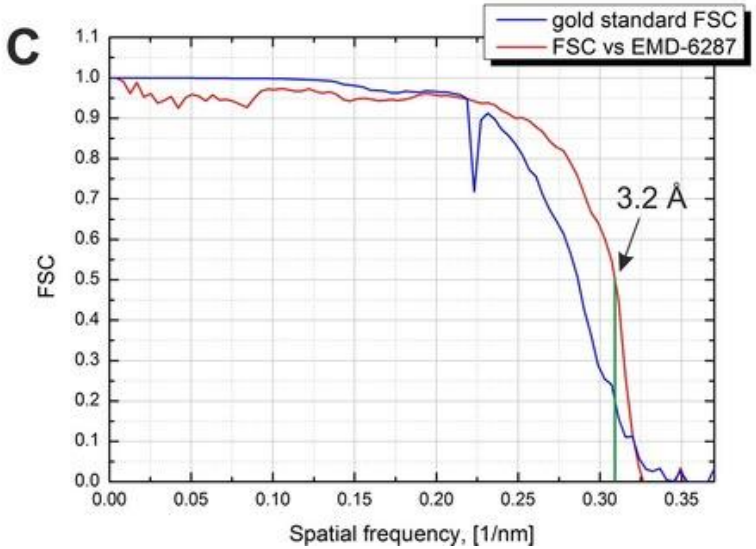
Danev, R. & Baumeister, W. Elife, 2016, 5



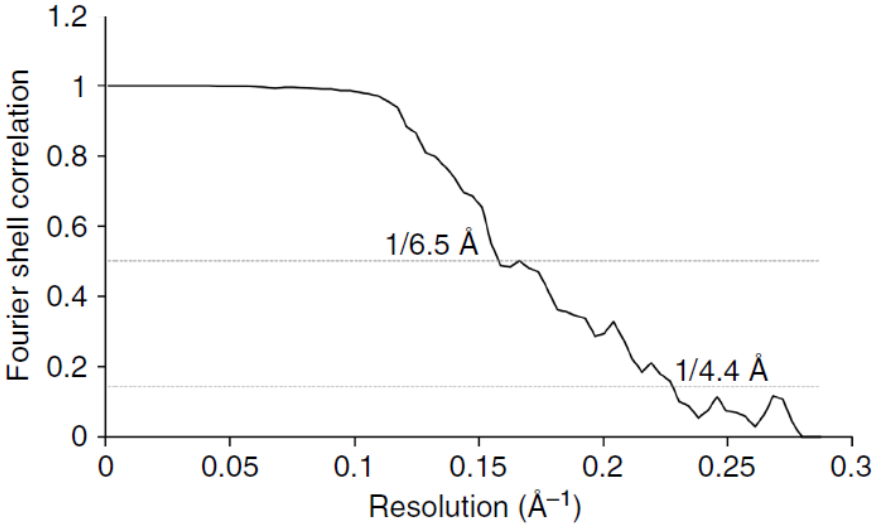
Phase plates





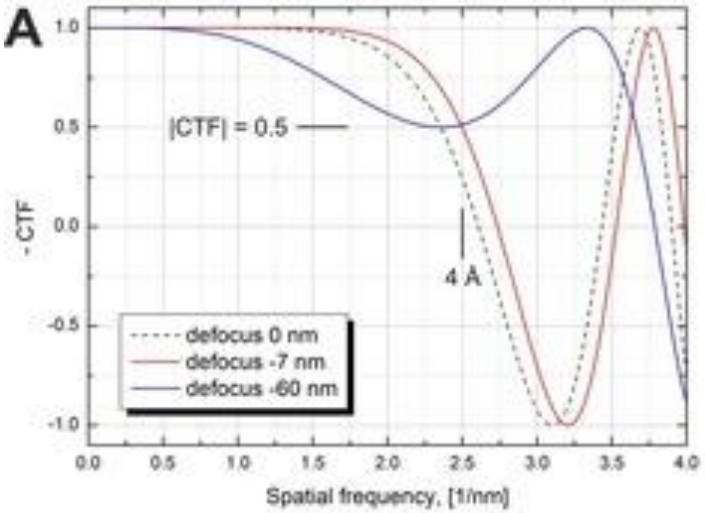


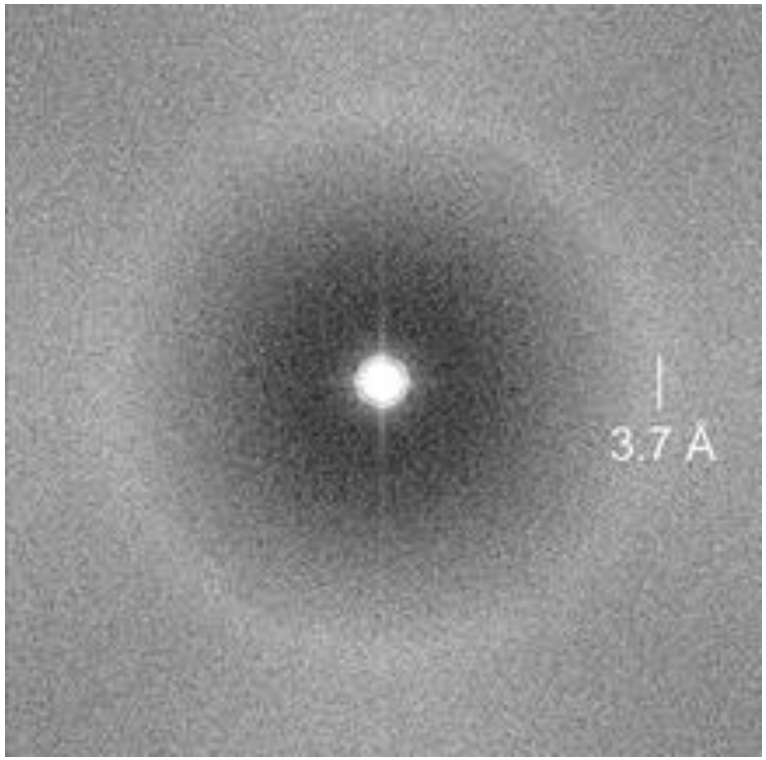
Danev, R. & Baumeister, W. Elife, 2016, 5



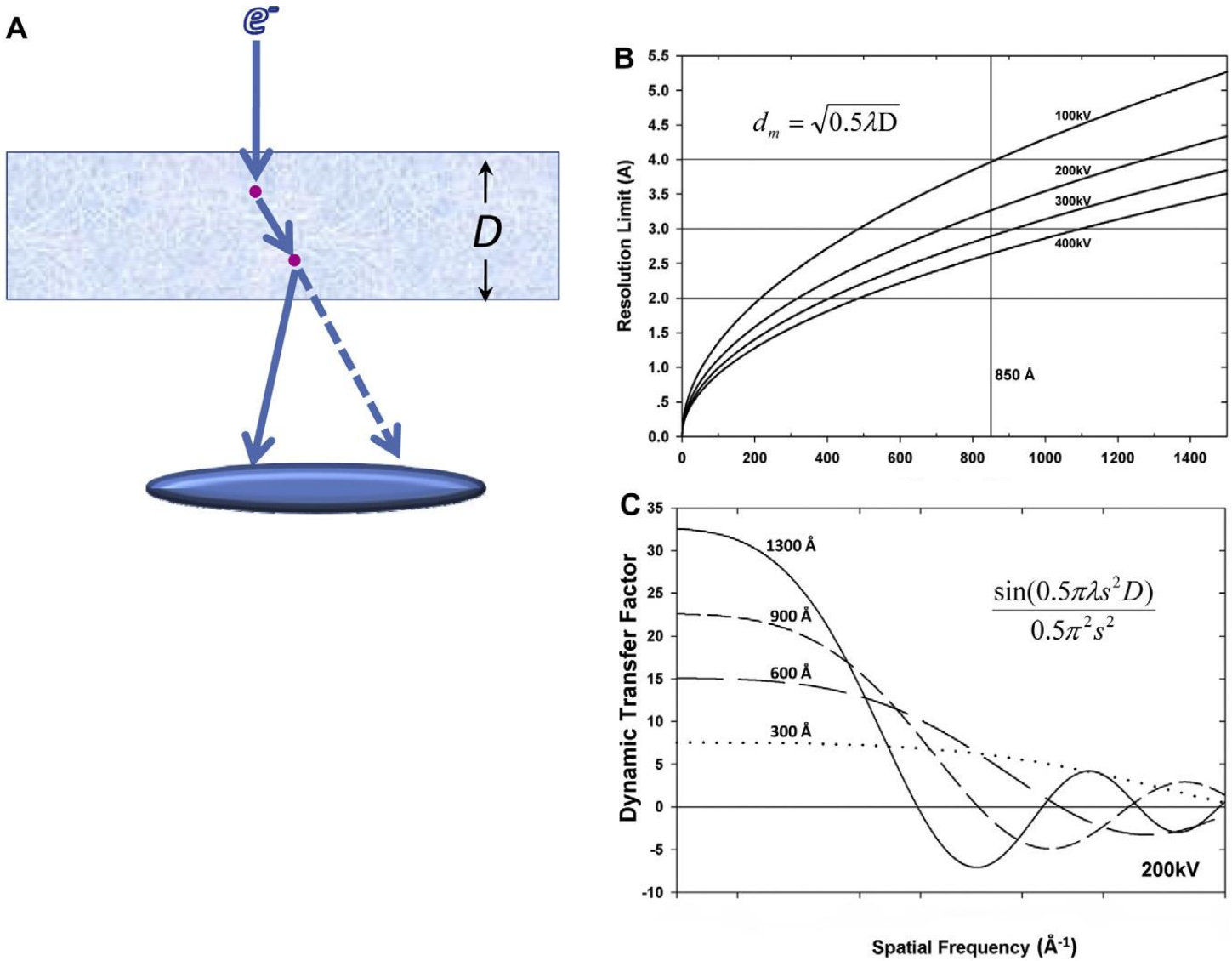
Khoshouei, et al. Nat Commun, 2016, 7, 10534

Phase plates



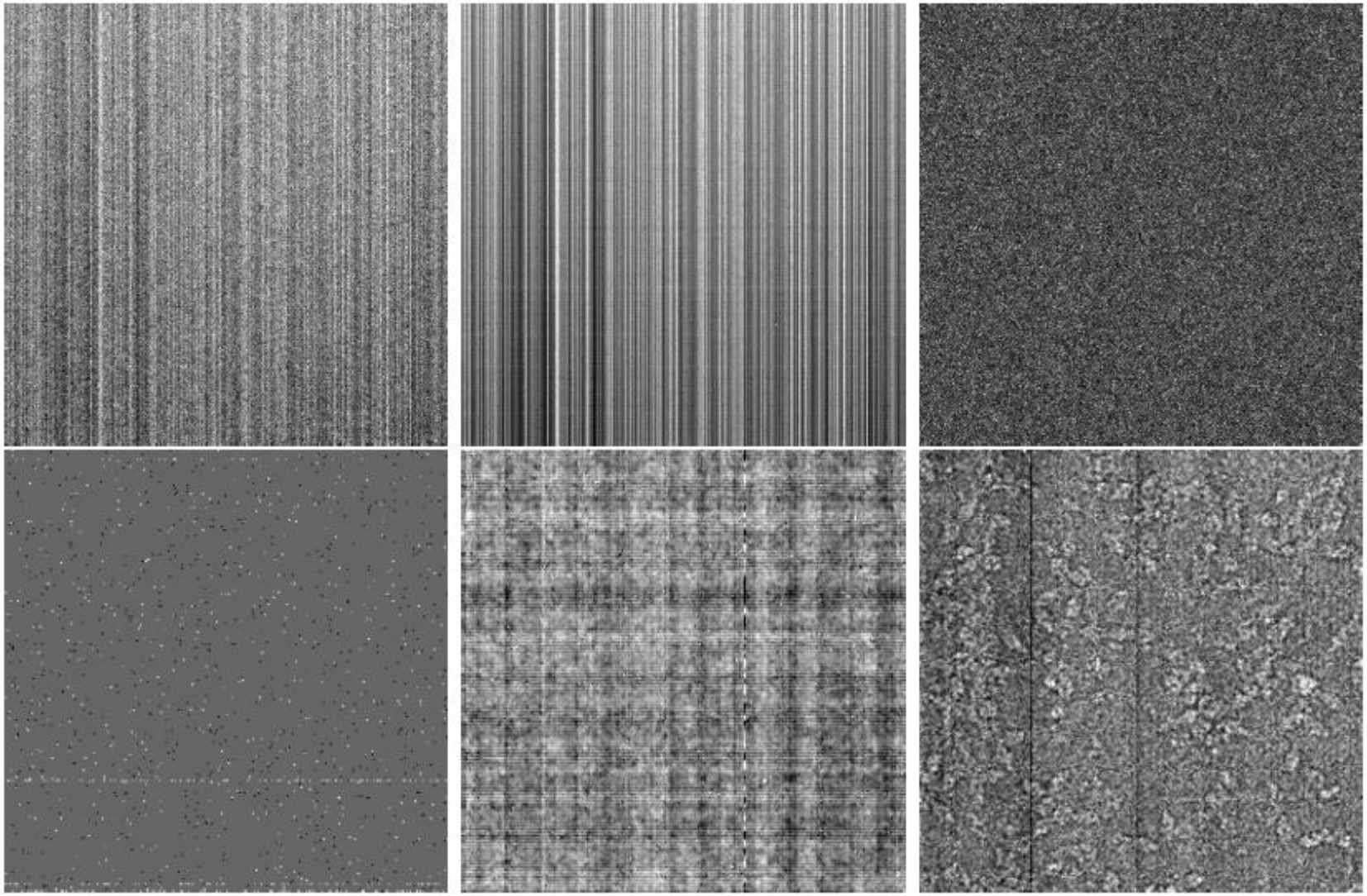


Dynamic electron scattering

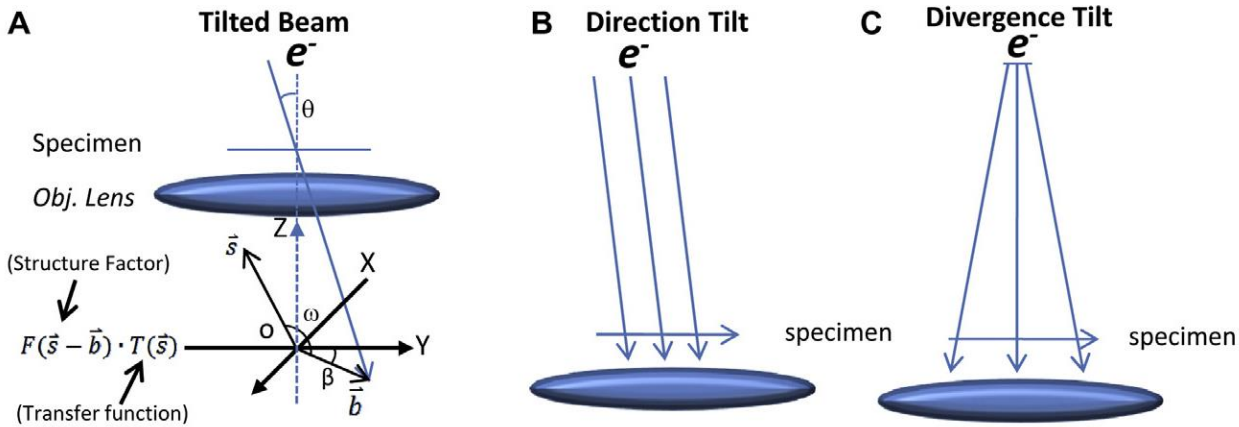


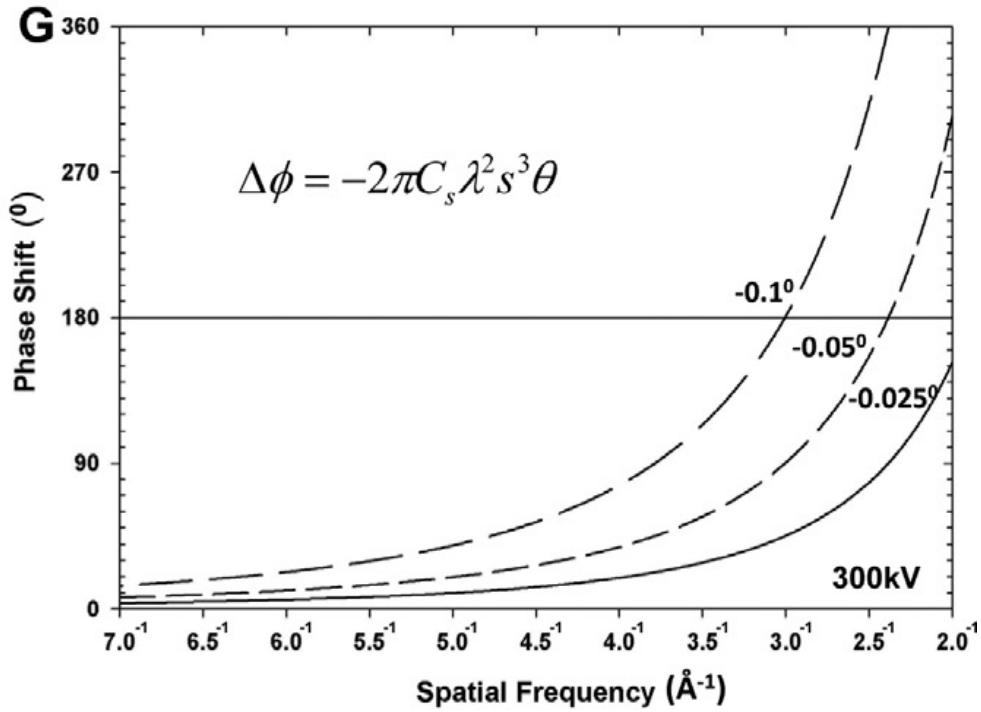
Zhang et al. J. Structural Biology, 2011, 175, 253-263

Efficient automated acquisition



Beam tilt





Zhang et al. J. Structural Biology, 2011, 175, 253-263

Beam tilt estimation

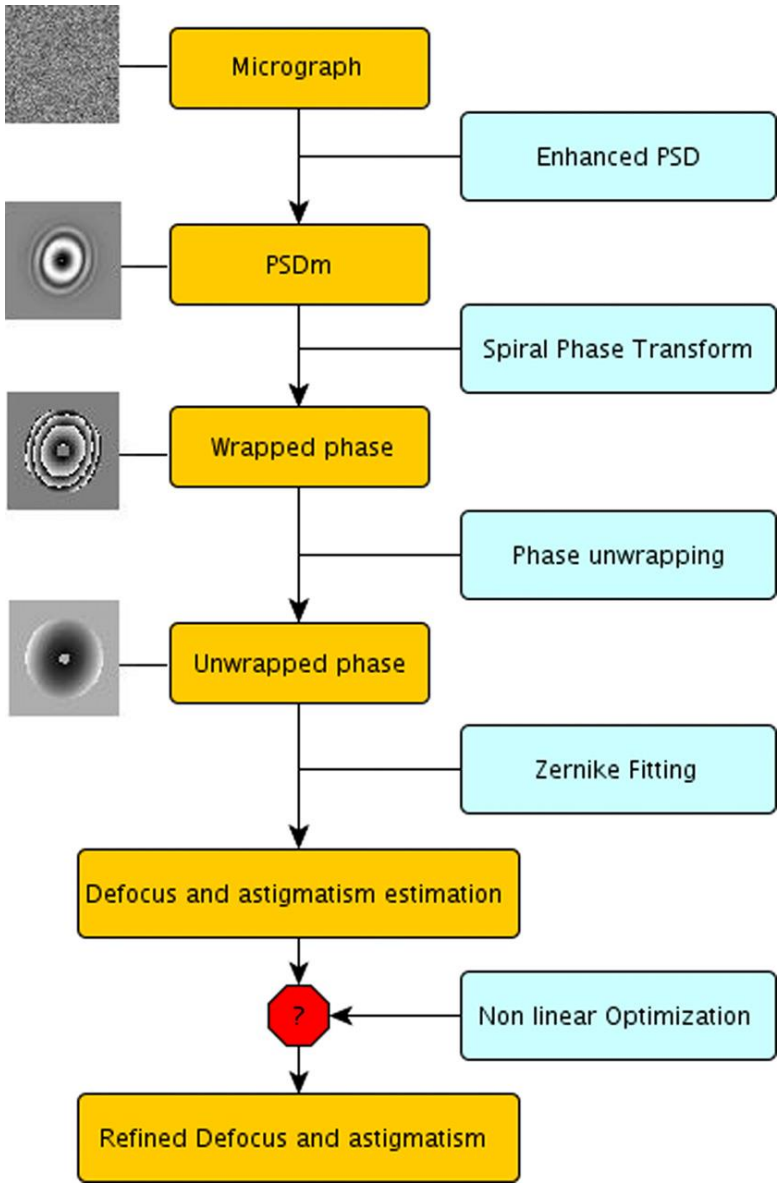
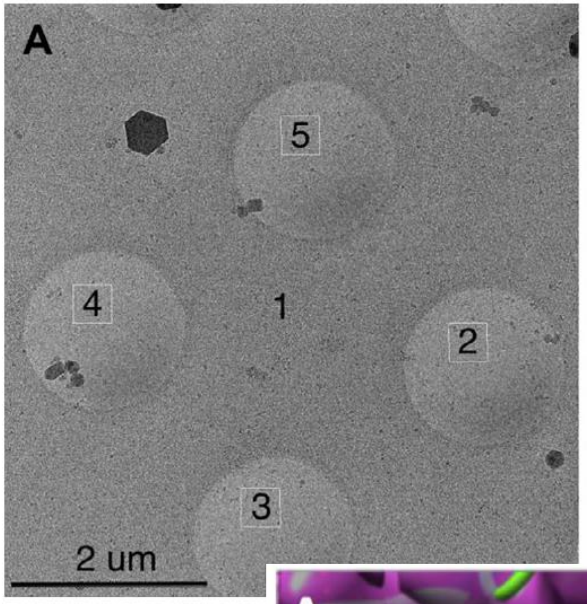


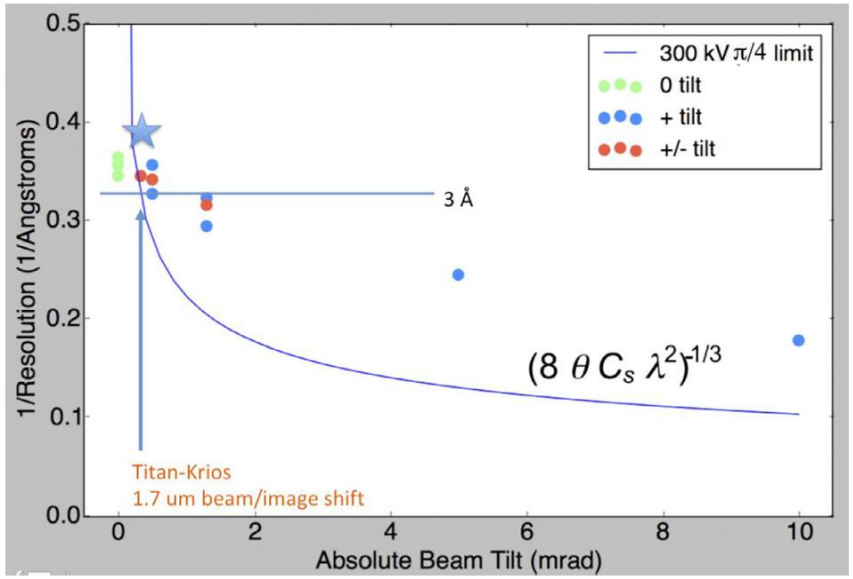
Table 2
Third-order Seidel aberrations coefficients obtained from Zernike polynomials coefficients.

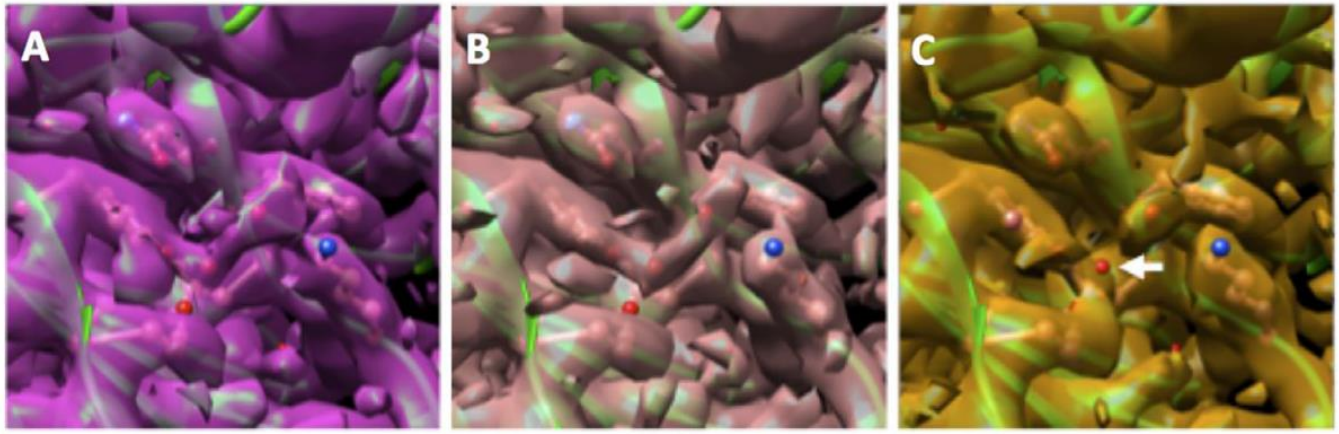
Piston	$Z_0 - Z_3 + Z_8$
Tilt	$r\sqrt{(Z_1 - 2Z_6)^2 + (Z_2 - 2Z_7)^2}\cos\left[\theta - \arctan\left(\frac{((Z_2 - 2Z_7))}{(Z_1 - 2Z_6)}\right)\right]$
Defocus	$r^2 2(Z_3 - 6Z_8) \pm \sqrt{Z_4^2 + Z_5^2}$
Astigmatism	$\pm 2r^2\sqrt{Z_4^2 + Z_5^2}\cos^2\left[\alpha - \frac{1}{2}\arctan\left(\frac{Z_5}{Z_4}\right)\right]$
Coma	$3r^3\sqrt{Z_6^2 + Z_7^2}\cos\left[\alpha - \arctan\left(\frac{Z_7}{Z_6}\right)\right]$
Spherical	$6r^4 Z_8$

Vargas, Sorzano, … J. Structural Biology, 2013, 181, 136-148

Beam shift as fast acquisition



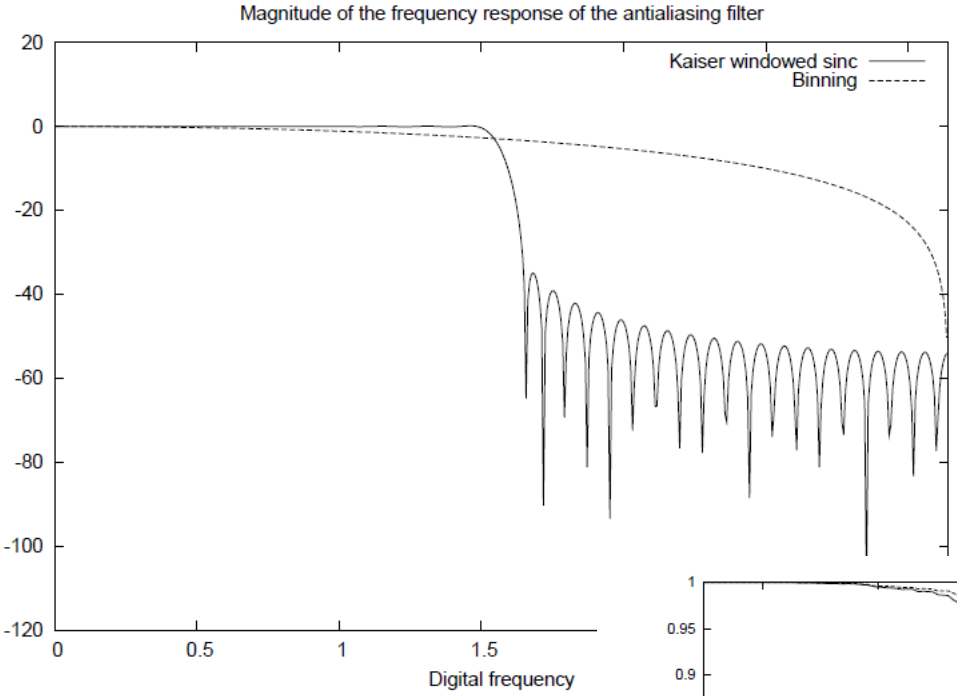


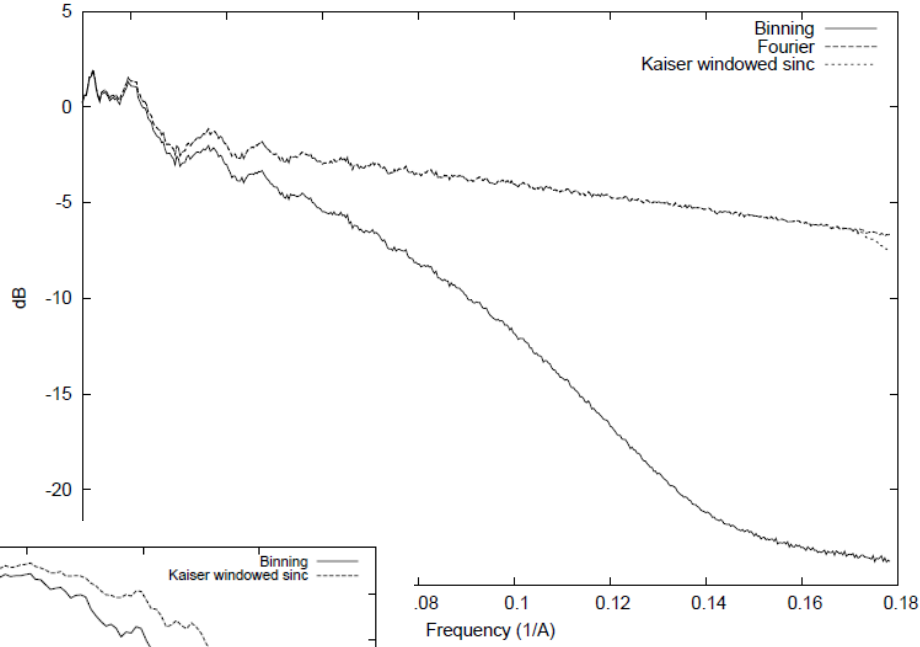


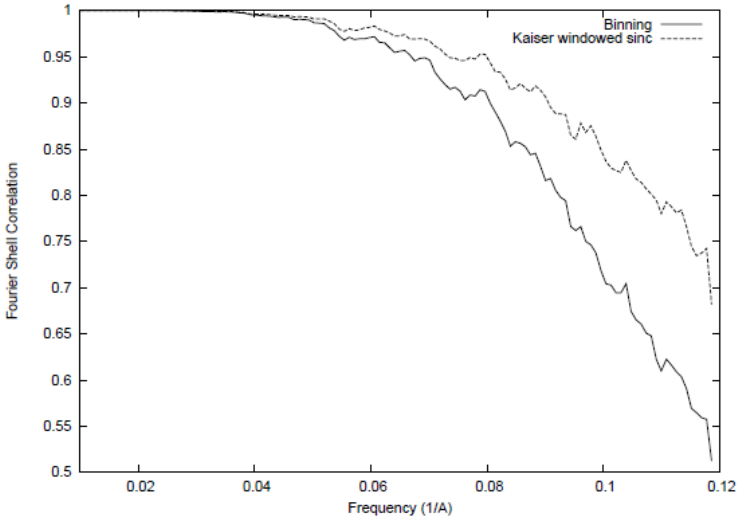
0 mrad 0.5 mrad 1.33 mrad

Cheng, A.; Eng, E. T.; Alink, L.; Rice, W. J.; Jordan, K. D.; Kim, L. Y.; Potter, C. S. & Carragher, B. High resolution single particle cryo-electron microscopy using beam-image shift. Journal of structural biology, 2018, 204, 270-275

Downsampling vs binning





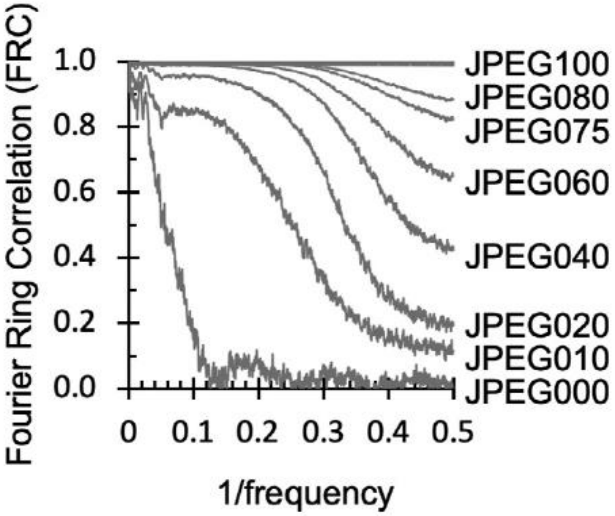


C.O.S. Sorzano, A. Iriarte, R. Marabini, J.M. Carazo. Effects of the downsampling scheme on three-dimensional electron microscopy of single particles. IEEE Workshop on Intelligent Signal Processing (2009)

Image compression

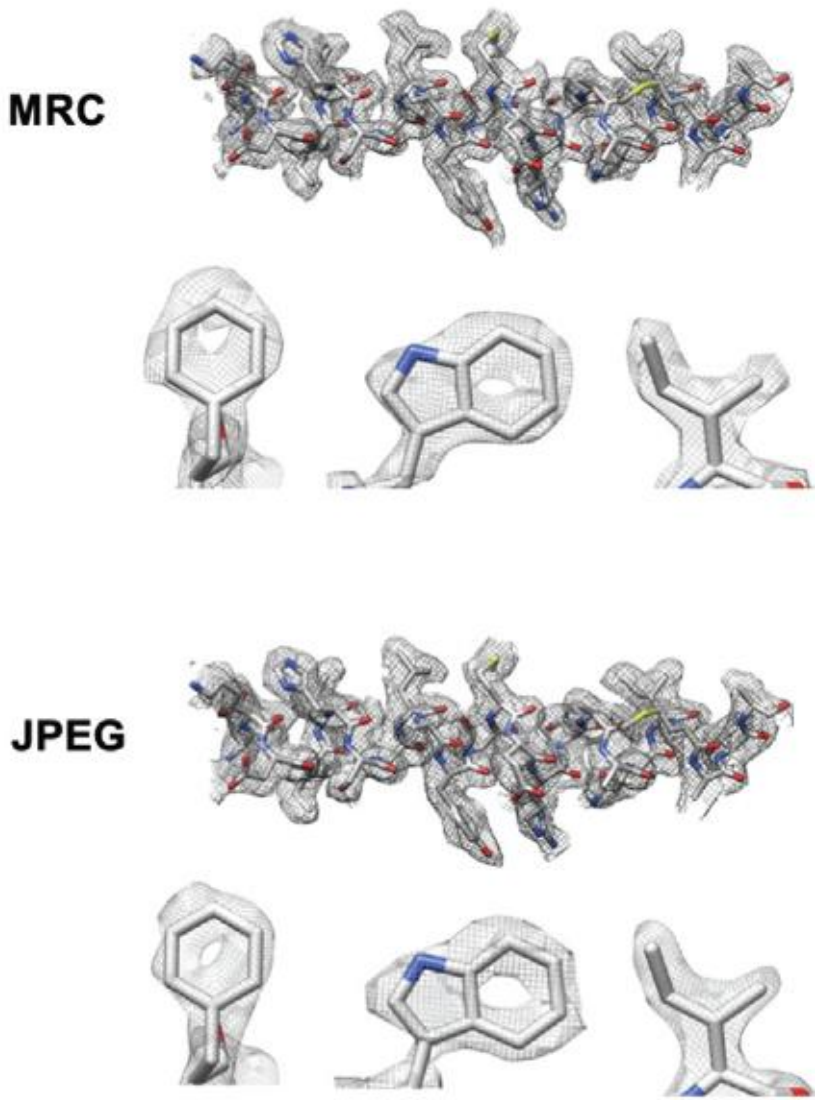
Comparison of image file size in MB for various file formats. In parenthesis is the compression ratio as compared against the original MRC file size.

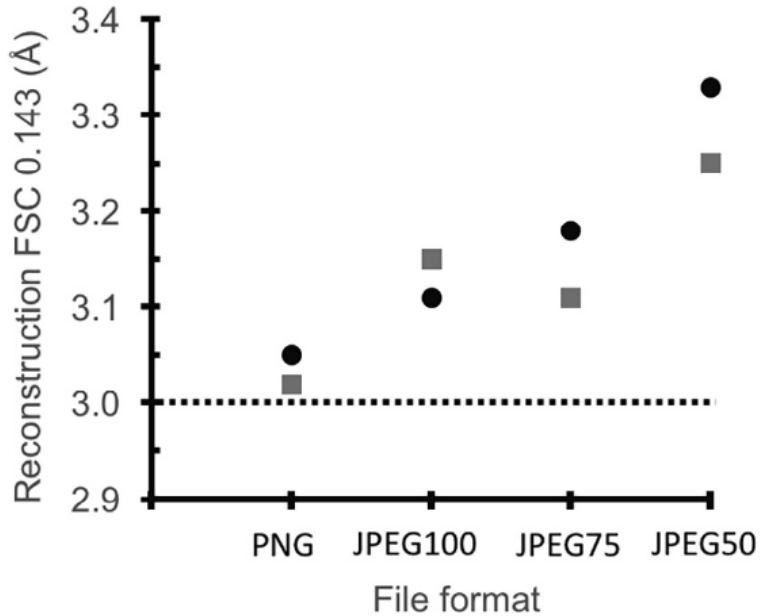
File format	K2 EC. 16-bit integer 3838 × 3710 pixels × 50 frames	K2 Summed image 32-bit floating point 3838 × 3710 pixels
MRC	1400	55
MRC with bz2[a]	221 (15.8%)	42 (76.4%)
Tiff	1400 (100%)	55 (100%)
Tiff ZIP	282 (20.1%)	42 (76.4%)
Tiff LZW	286 (20.4%)	51 (92.7%)
PNG (8-bit)	223 (15.9%)	12 (21.8%)
JPEG000 (most loss)	8.5 (0.6%)	0.2 (0.4%)
JPEG050	188 (13.4%)	2.7 (4.8%)
JPEG075	273 (19.5%)	4.2 (7.6%)
JPEG100 (least loss)	842 (60.1%)	15 (27.3%)



Eng, E. T.; Kopylov, M.; Negro, C. J.; Dallaykan, S.; Rice, W. J.; Jordan, K. D.; Kelley, K.; Carragher, B. & Potter, C. S. Reducing cryoEM file storage using lossy image formats. Journal of structural biology, 2019

Image compression





instruct image Processing Center

Conclusions

- Intelligent acquisition
- Quality filters
- Correction of Beam Induced Movement
- Dose compensation
- Anisotropic magnification
- Accurate CTF estimation
- Accurate downsampling

- Accuracy 1st, then speed

Particle selection and 2D classification

Carlos Oscar S. Sorzano

Instruct Image Processing Center

CSIC

instruct Integrating Biology

instruct image Processing Center

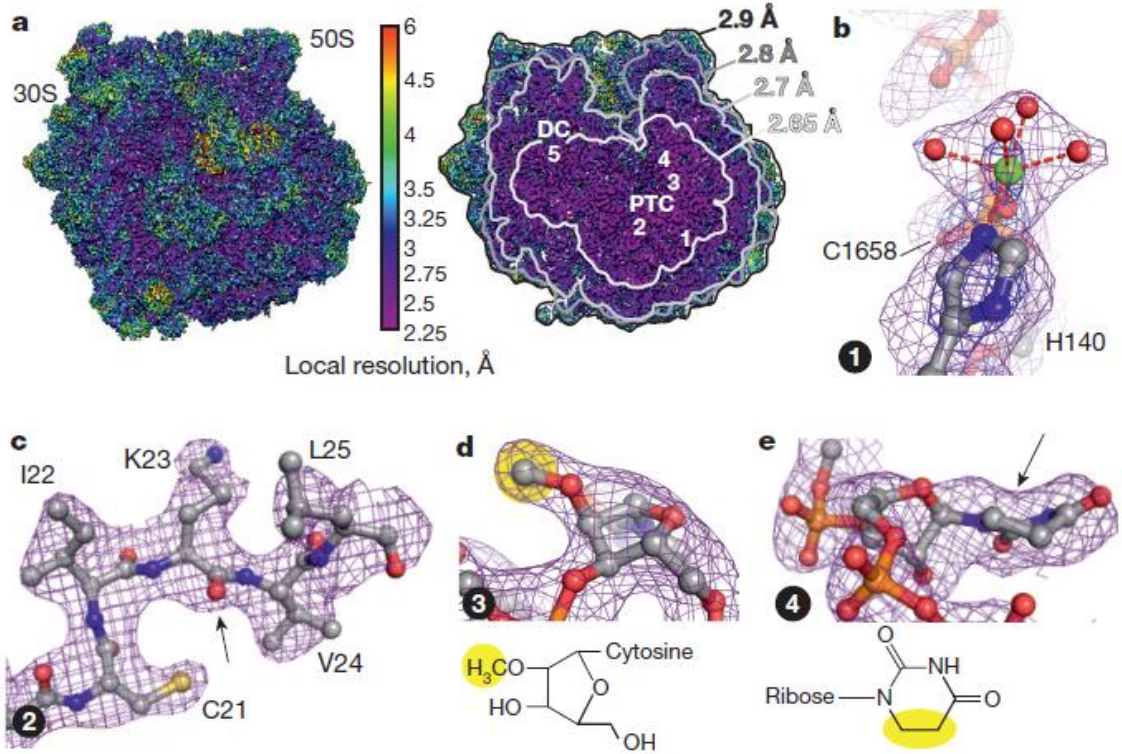
High-resolution, High-throughput EM

LETTER

doi:10.1038/nature14275

Structure of the *E. coli* ribosome–EF–Tu complex at <3 Å resolution by C_s-corrected cryo-EM

Niels Fischer[1]*, Piotr Neumann[2]*, Andrey L. Konevega[3,4,5], Lars V. Bock[6], Ralf Ficner[2], Marina V. Rodnina[5] & Holger Stark[1,7]

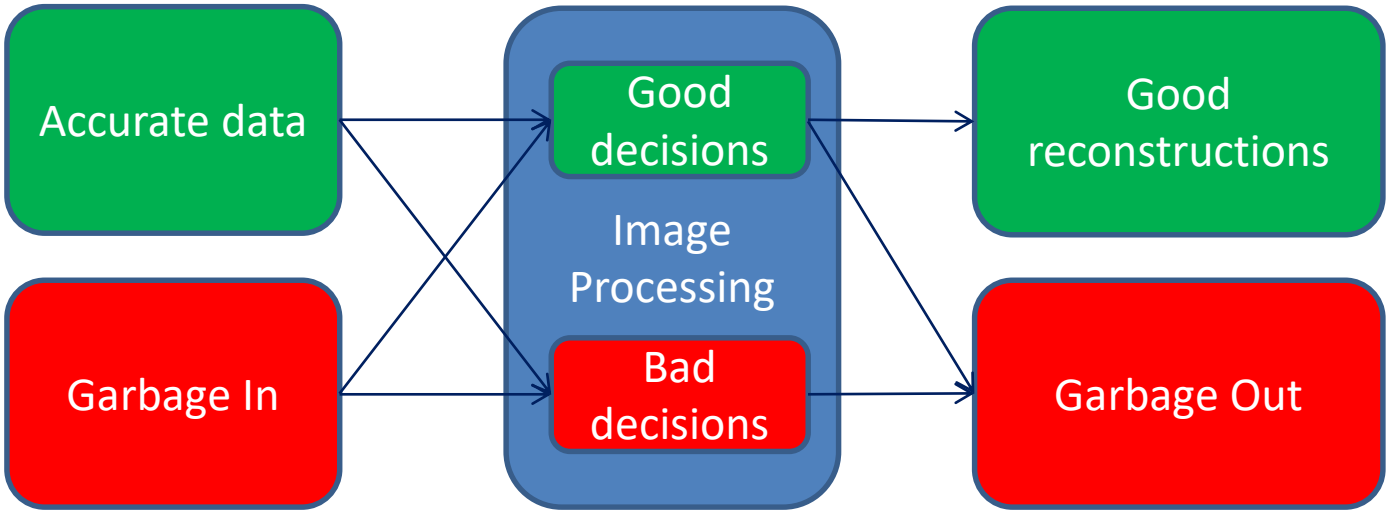


1.6M particles

Fischer et al. Nature, (2015) 23 Feb

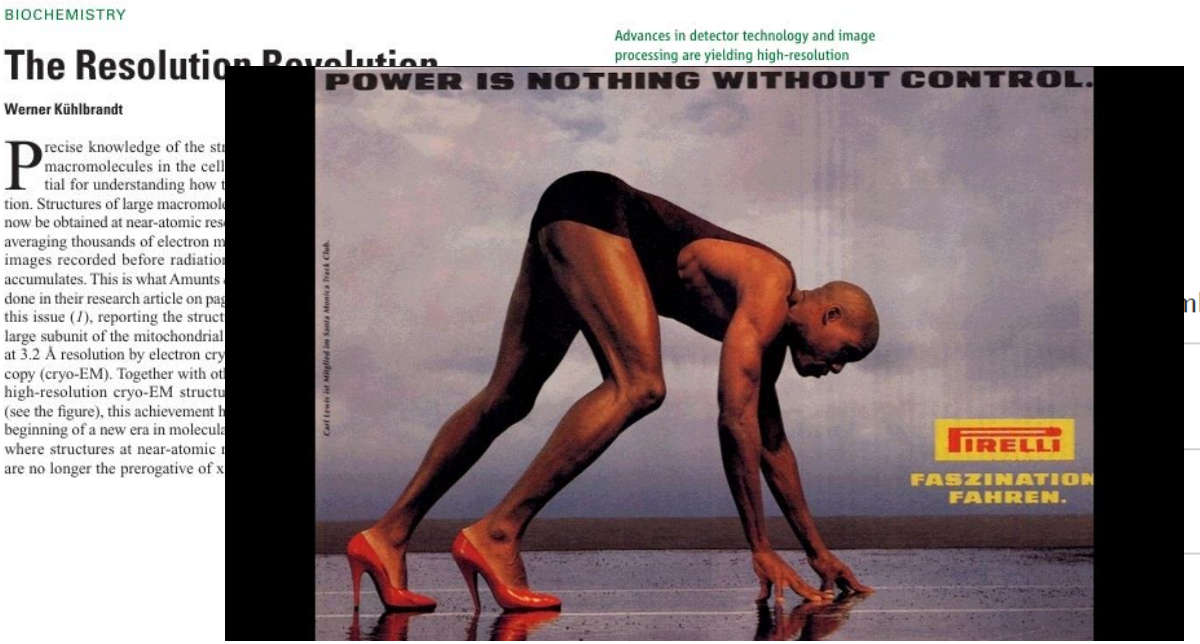
EM is a GIGO system

- The good thing about EM is that <u>it always gives a volume</u>
- The bad thing about EM is that <u>it always gives a volume</u>



EM is a GIGO system

- Do not go to the next step till you are sure of the current step

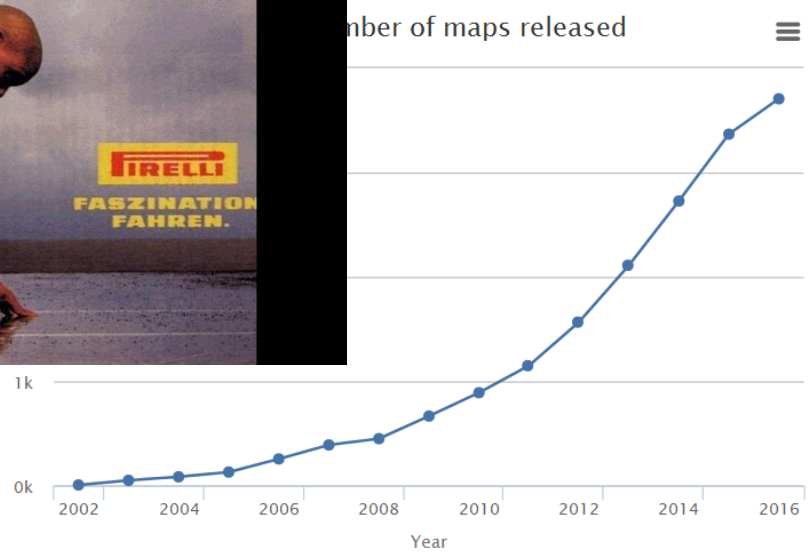

BIOCHEMISTRY

The Resolution Revolution

Werner Kühlbrandt

Precise knowledge of the str
macromolecules in the cell
tial for understanding how t
tion. Structures of large macromol
now be obtained at near-atomic res
averaging thousands of electron m
images recorded before radiatio
accumulates. This is what Amunts
done in their research article on pag
this issue (*1*), reporting the struct
large subunit of the mitochondrial
at 3.2 Å resolution by electron cry
copy (cryo-EM). Together with ot
high-resolution cryo-EM structu
(see the figure), this achievement h
beginning of a new era in molecula
where structures at near-atomic r
are no longer the prerogative of x

Advances in detector technology and image processing are yielding high-resolution



Controversial reconstructions

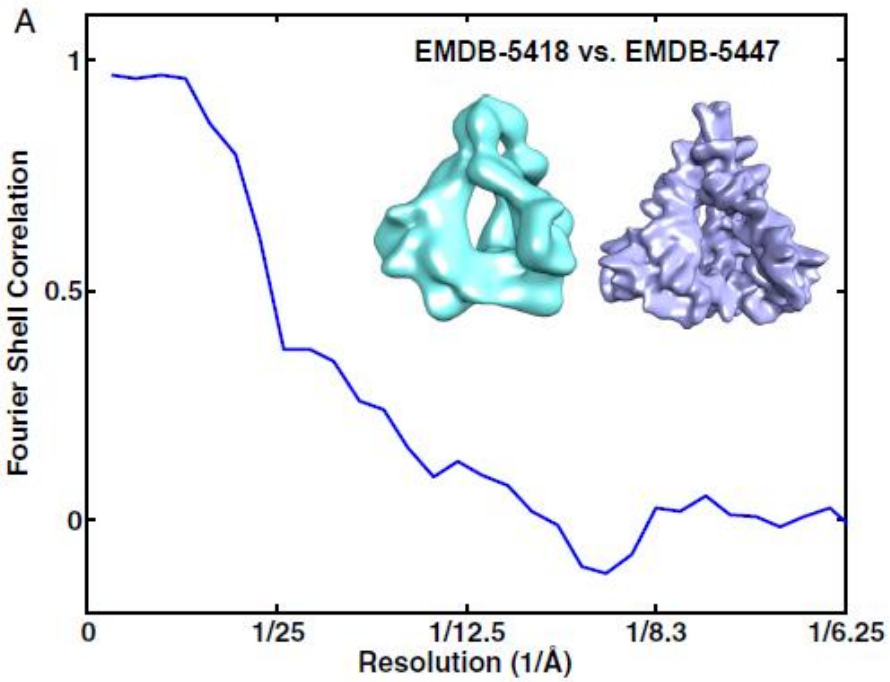
PNAS

LETTER

Structure of trimeric HIV-1 envelope glycoproteins

Mao et al. (1) published a paper describing a 3D structure of uncleaved, trimeric HIV-1 envelope glycoprotein (Env) at ~6-Å resolution, following a similar paper last year on the same structure at ~11-Å resolution (2).

uncleaved trimeric HIV-1 Env a
in the work of Mao et al. (1, 2)
ment with the structure of nati
trimeric HIV-1 Env presented i
(4) is incorrect. EMD-5019 repr

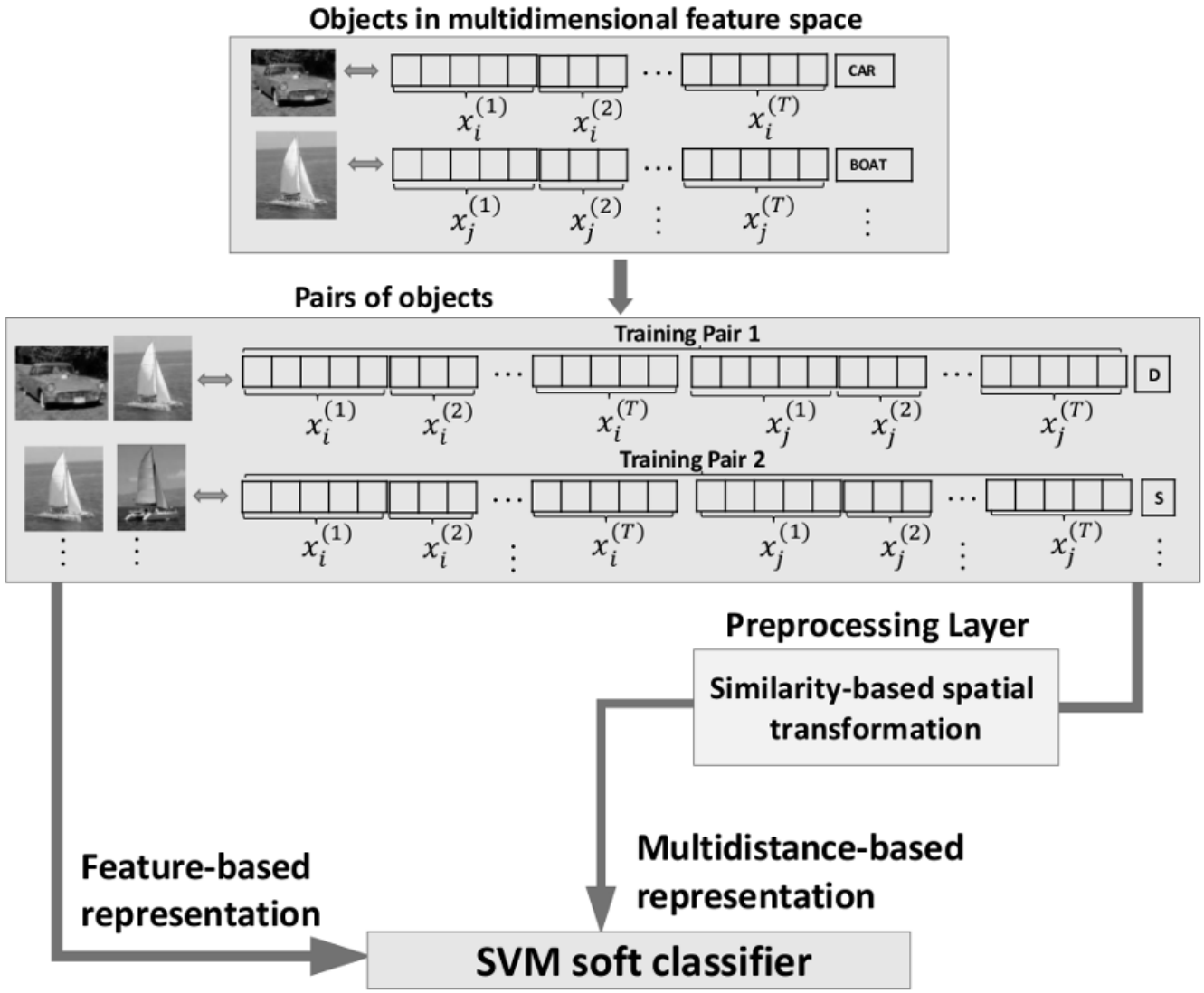


Particle picking algorithms

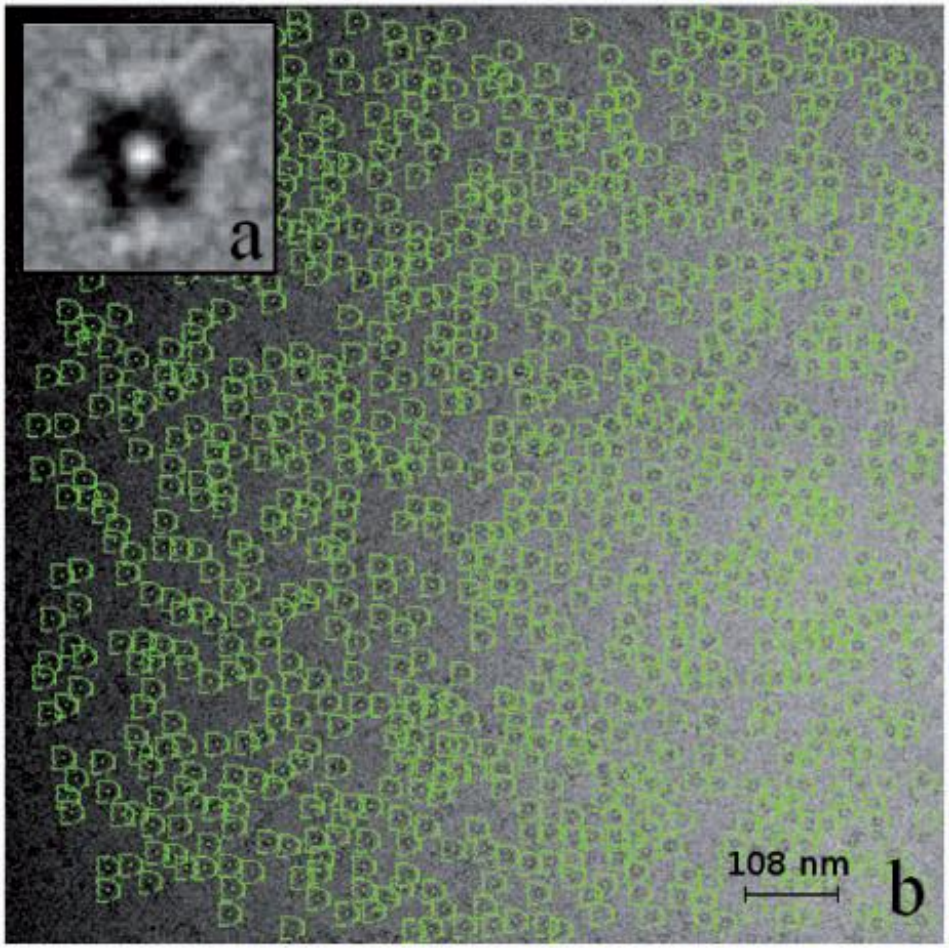
Automatic particle picking [edit]

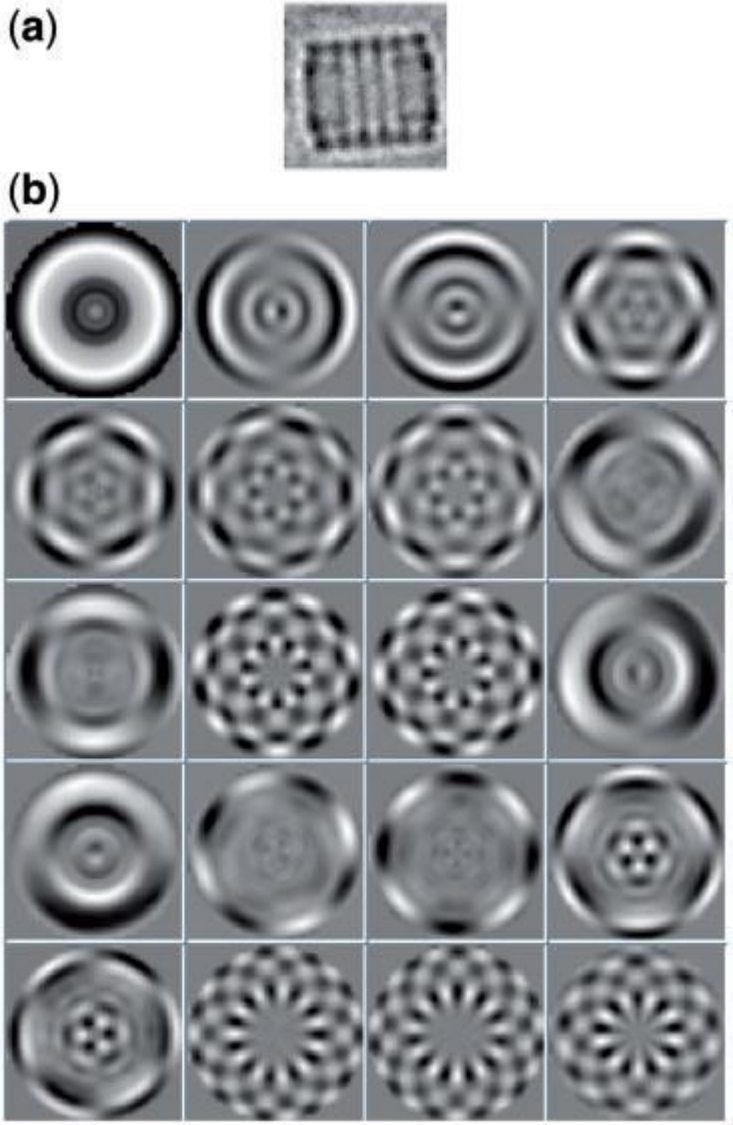
Paper 1982VanHeel_Detection	Detection of particles in micrographs
Paper 2001Nicholson_Review	Review on automatic particle picking
Paper 2001Zhu_Filaments	Automatic identification of filaments in micrographs
Paper 2004Sigworth_Detection	Classical detection theory and the cryo-EM particle selection problem
Paper 2004Volkmann_ParticlePicking	An approach to automated particle picking from electron micrographs based on reduced representation templates
Paper 2004Wong_ParticlePicking	Model-based particle picking for cryo-electron microscopy
Paper 2004Zhu_Review	Review on automatic particle picking
Paper 2007Chen_Signature	Automatic particle picking program: Signature
Paper 2007Woolford_SwarmPS	Automatic particle picking with several criteria, implemented in EMAN Boxer
Paper 2009Sorzano_MachineLearning	Automatic particle picking based on machine learning of rotational invariants
Paper 2011Arbelaez_Comparison	Evaluation of the performance of software for automated particle-boxing
Paper 2013Abrishami_MachineLearning	A pattern matching approach to the automatic selection of particles from low-contrast electron micrographs
Paper 2013Hauer_2013	Automatic tilt pair detection in Random Conical Tilt
Paper 2013Shatsky_ParticlePicking	Automated particle correspondence and accurate tilt-axis detection in tilted-image pairs
Paper 2013Vargas_ParticleQuality	Automatic determination of particle quality
Paper 2014Langlois_ParticlePicking	Automatic particle picking
Paper 2016Vilas_AutomaticTilt	Automatic identification of image pairs in untilted-tilted micrograph pairs

Feature based classification



Automatic Particle Picking





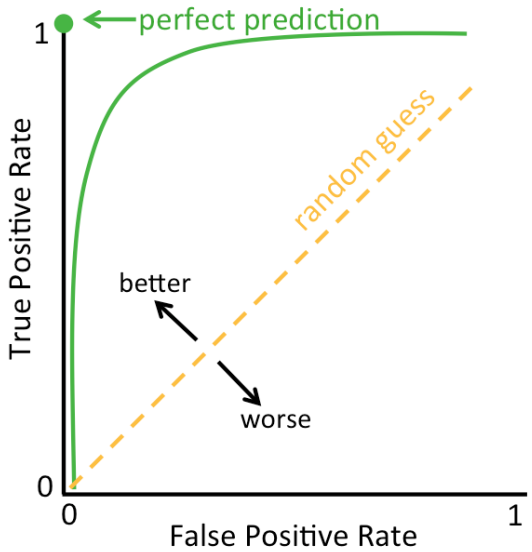
Abrishami, … Sorzano. Bioinformatics 2013, 29, 2460-2468

Automatic Particle Picking

Median/Mean	FNR	11.4/13.3	28.0/27.9	16.2/16.2	21.5/22.0	43.4/44.5	33.7/34.4	20.5/20.8	48.3/47.9	7.8/10.9	27.1/28.0	30.1/30.7	17.1/17.2
	FPR	33.9/34.7	21.5/25.3	23.2/21.7	18.4/18.7	27.1/28.9	30.1/33.6	23.1/23.8	33.9/35.4	30.3/29.8	10.3/15.1	15.7/20.6	28.0/26.3

False Negative Rate (Particles missed): 10-50%
False Positive Rate (False particles): 15-35%

Relative Operating Characteristic (ROC)



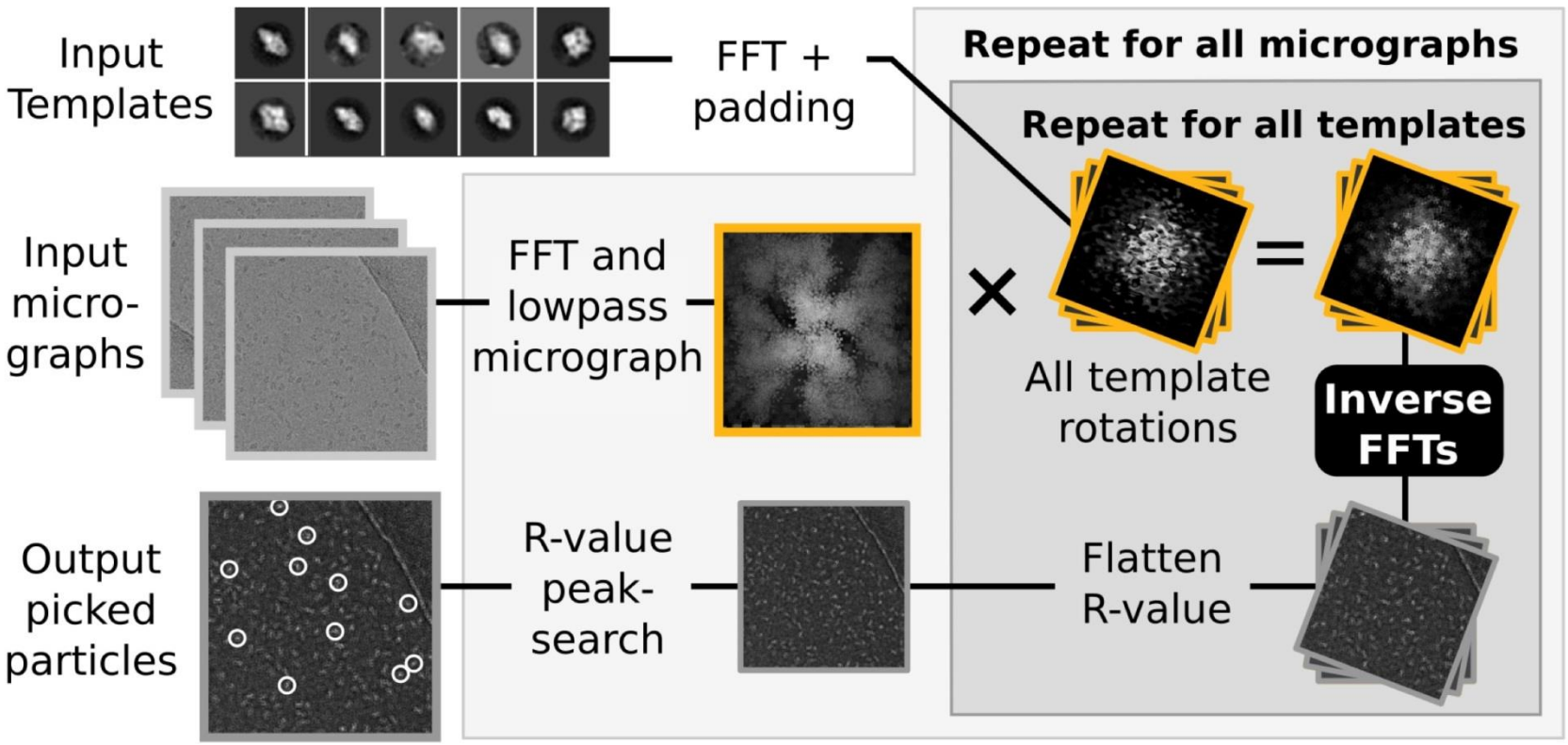
AUC = area under the curve

0.5-0.7 = poor model performance
0.7-0.9 = moderate
> 0.9 = excellent

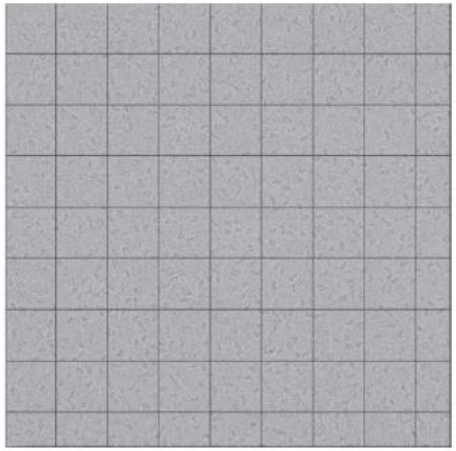
- $\text{Precision} = \frac{TP}{FP+TP}$
- $\text{Recall} = \frac{TP}{FN+TP}$
- $F\text{-measure} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$

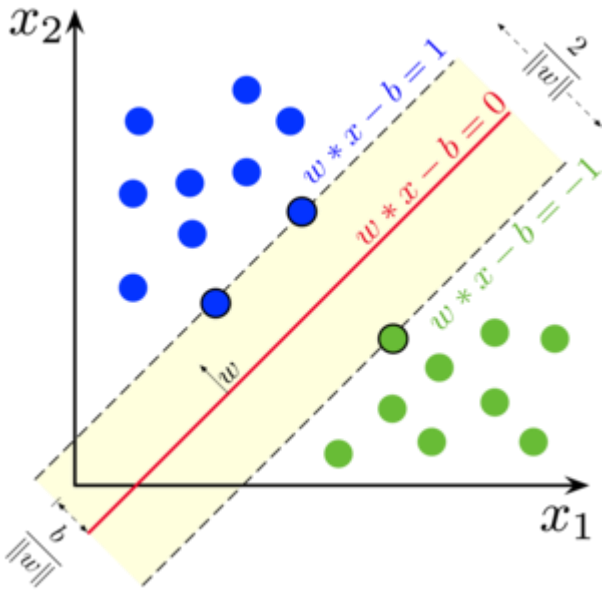
Zhu, Y.; Carragher, B.; Glaeser, R. M.; Fellmann, D.; Bajaj, C.; Bern, M.; Mouche, F.; de Haas, F.; Hall, R. J.; Kriegman, D. J.; Ludtke, S. J.; Mallick, S. P.; Penczek, P. A.; Roseman, A. M.; Sigworth, F. J.; Volkmann, N. & Potter, C. S. Automatic particle selection: results of a comparative study J. Structural Biology, 2004, 145, 3-14

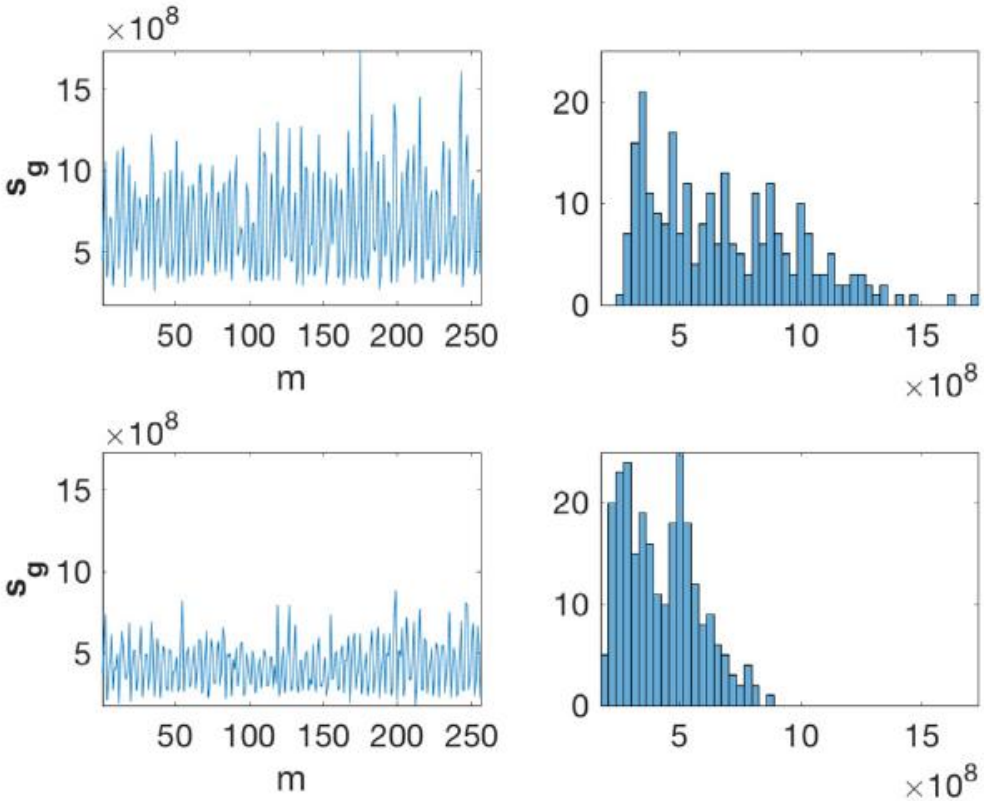
Template based matching



Automatic template detection

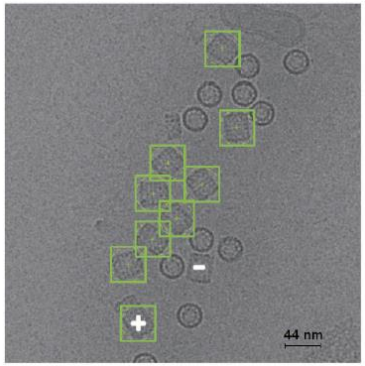






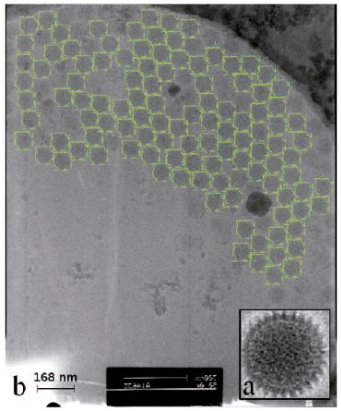
Heimowitz, A., Andén, J. & Singer, A. APPLE Picker: Automatic Particle Picking, a Low-Effort Cryo-EM Framework. Journal of Structural Biology, 2018, 204(2), 215-227.

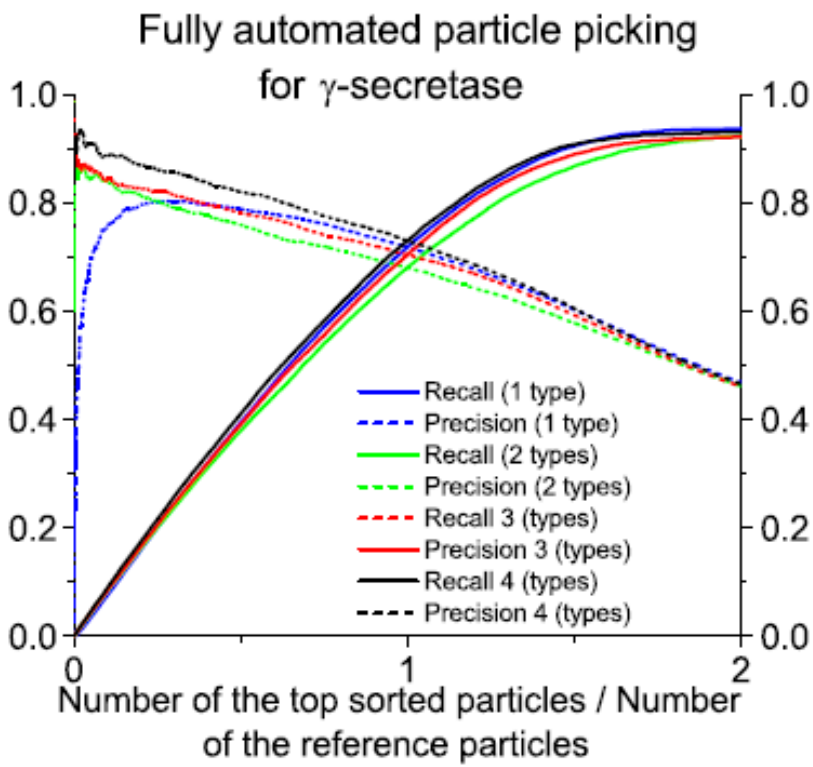
Automatic Particle Picking



Precision: 85% (15% are not particles)
Recall: 75% (25% particles lost)

Precision: 92% (8% are not particles)
Recall: 90% (10% particles lost)

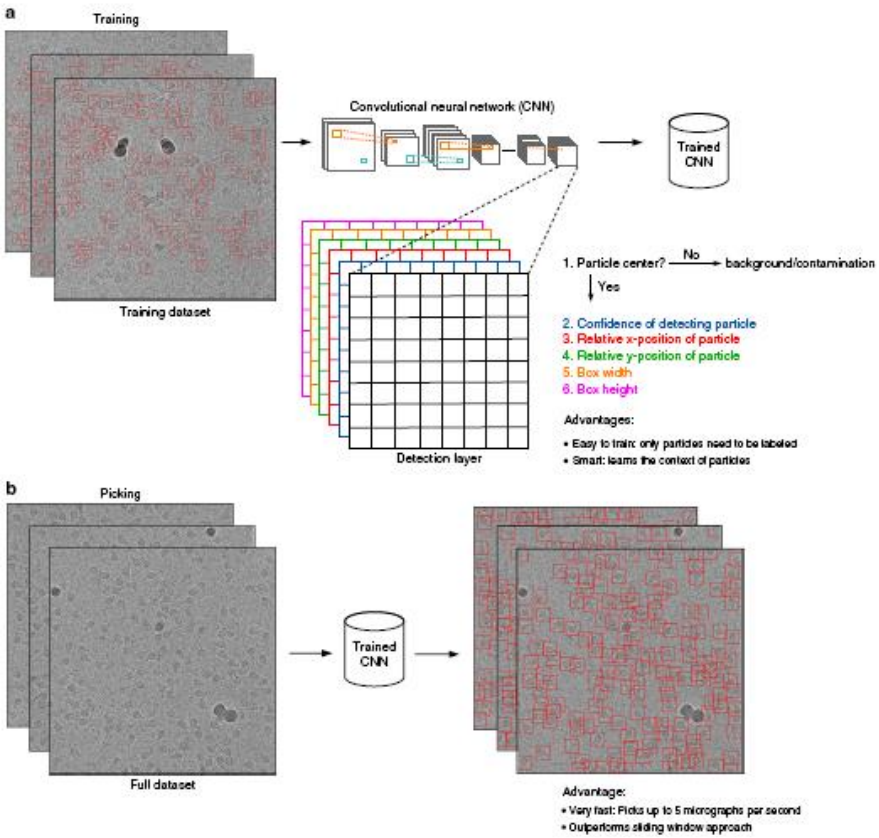




Abrishami, V.; Zaldívar-Peraza, A.; de la Rosa-Trevín, J. M.; Vargas, J.; Otón, J.; Marabini, R.; Shkolnisky, Y.; Carazo, J. M. & Sorzano, C. O. S. A pattern matching approach to the automatic selection of particles from low-contrast electron micrographs. Bioinformatics, 2013, 29, 2460-2468

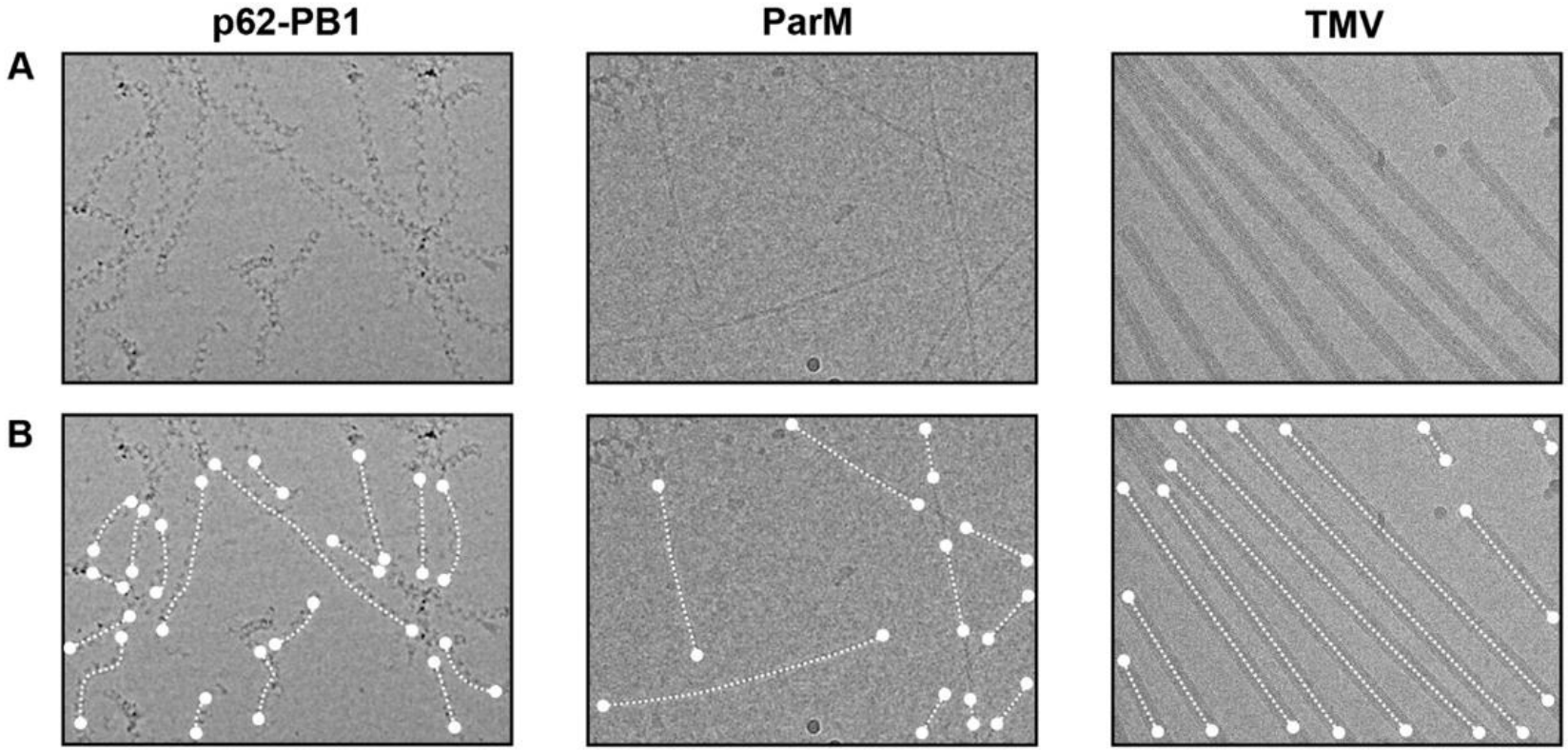
Wang, F.; Gong, H.; Liu, G.; Li, M.; Yan, C.; Xia, T.; Li, X. & Zeng, J. DeepPicker: A deep learning approach for fully automated particle picking in cryo-EM. J. Structural Biology, 2016, 195, 325-336

Deep Learning



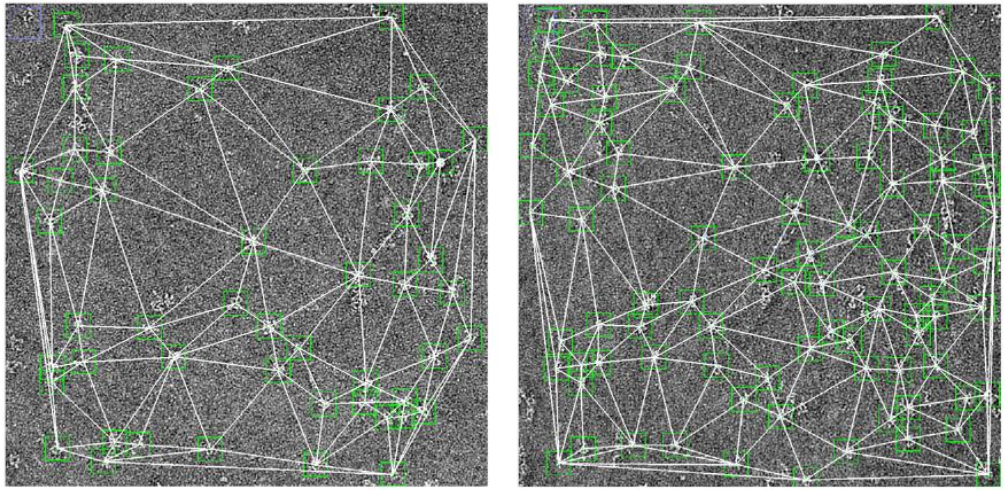
1. Wang, F., Gong, H., Liu, G., Li, M., Yan, C., Xia, T., ... & Zeng, J. (2016). DeepPicker: A deep learning approach for fully automated particle picking in cryo-EM. Journal of structural biology, 195(3), 325-336.
2. Zhu, Y., Ouyang, Q., & Mao, Y. (2017). A deep convolutional neural network approach to single-particle recognition in cryo-electron microscopy. BMC bioinformatics, 18(1), 348.
3. Wagner, T.; Merino, F.; Stabrin, M.; Moriya, T.; Antoni, C.; Apelbaum, A.; Hagel, P.; Sitsel, O.; Raisch, T.; Prumbaum, D.; Quentin, D.; Roderer, D.; Tacke, S.; Siebolds, B.; Schubert, E.; Shaikh, T. R.; Lill, P.; Gatsogiannis, C. & Raunser, S. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. Communications biology, 2019 , 2 , 218
4. Zhang, J.; Wang, Z.; Chen, Y.; Han, R.; Liu, Z.; Sun, F. Zhang, F. PIXER: an automated particle-selection method based on segmentation using a deep neural network. BMC bioinformatics, 2019, 20, 41

Helix detection

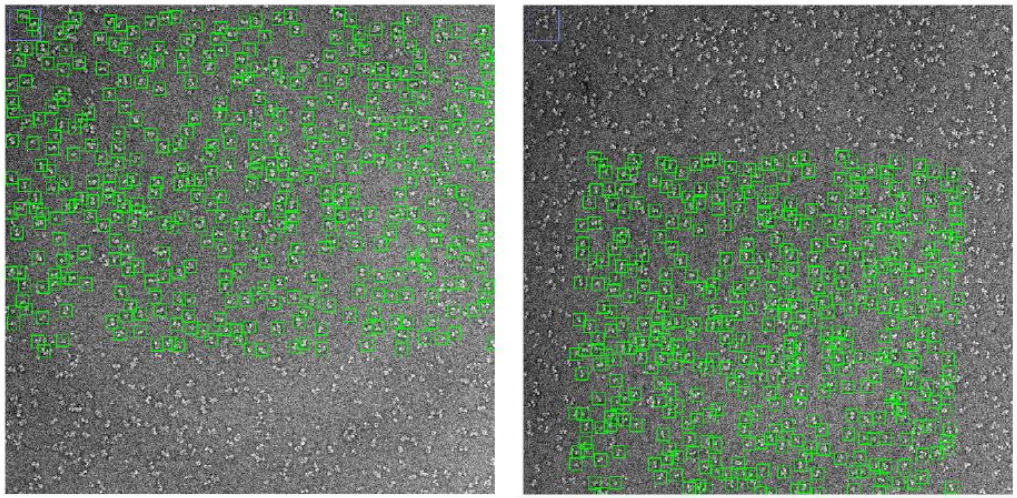


Huber, S. T.; Kuhm, T. & Sachse, C. Automated tracing of helical assemblies from electron cryo-micrographs. Journal of structural biology, 2018, 202, 1-12

Tilt pair correspondences

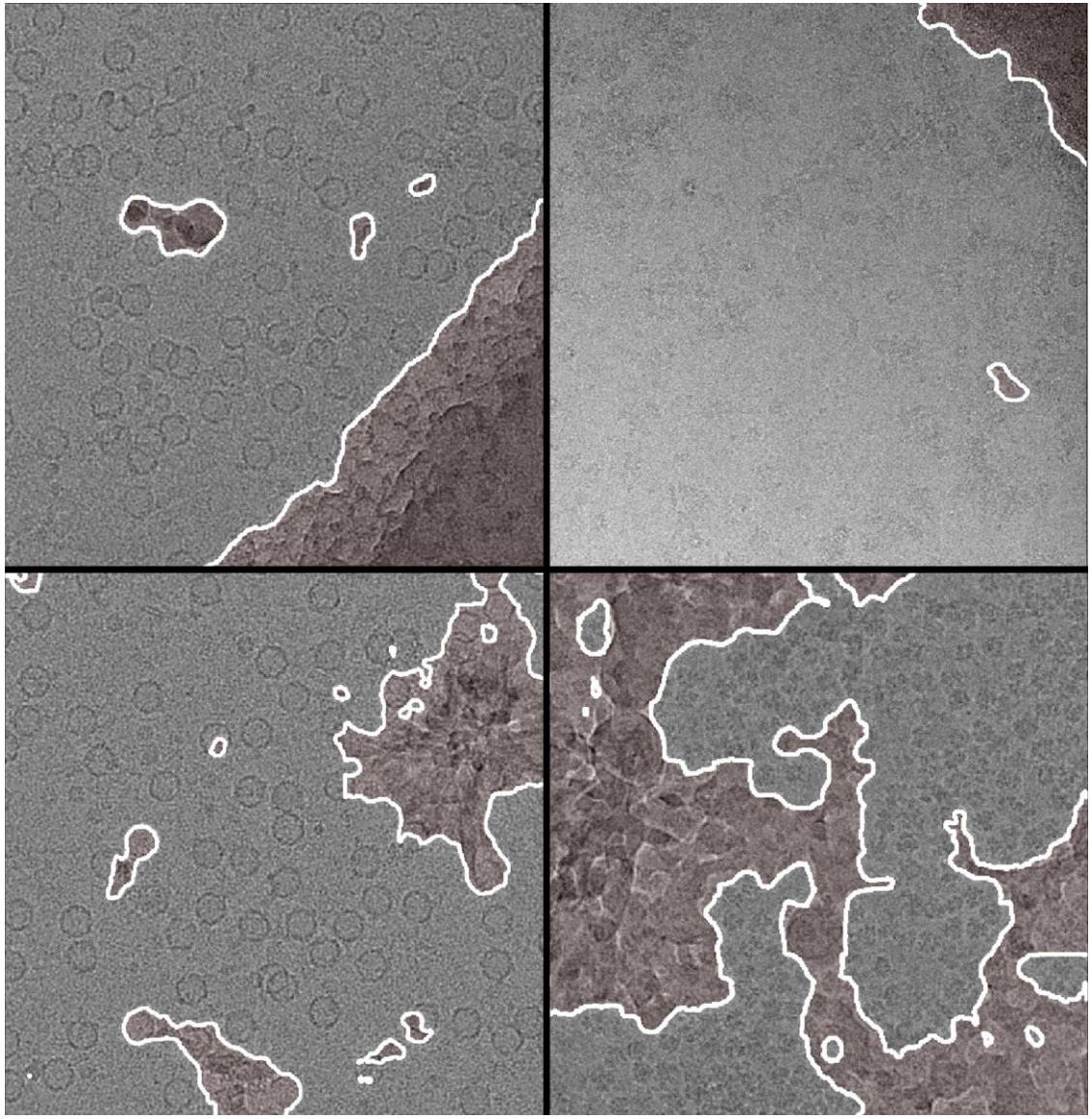


(a) Untilted (b) Tilted



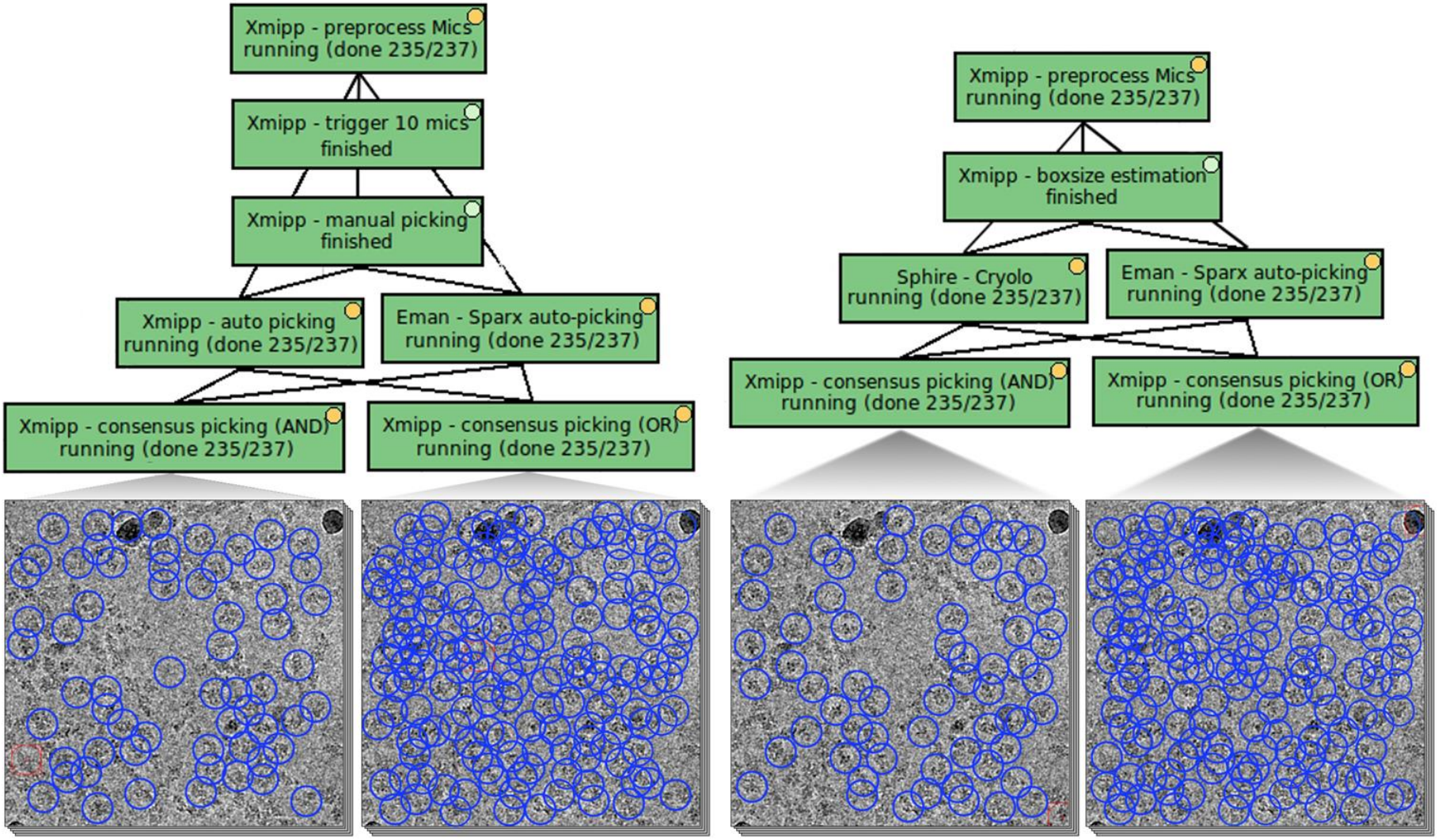
Vilas, J. L.; Navas, J.; Gómez-Blanco, J.; de la Rosa-Trevín, J. M.; Melero, R.; Peschiera, I.; Ferlenghi, I.; Cuenca, J.; Marabini, R.; Carazo, J. M.; Vargas, J. & Sorzano, C. O. S. Fast and automatic identification of particle tilt pairs based on Delaunay triangulation. Journal of structural biology, 2016, 196, 525-533

Bad areas recognition

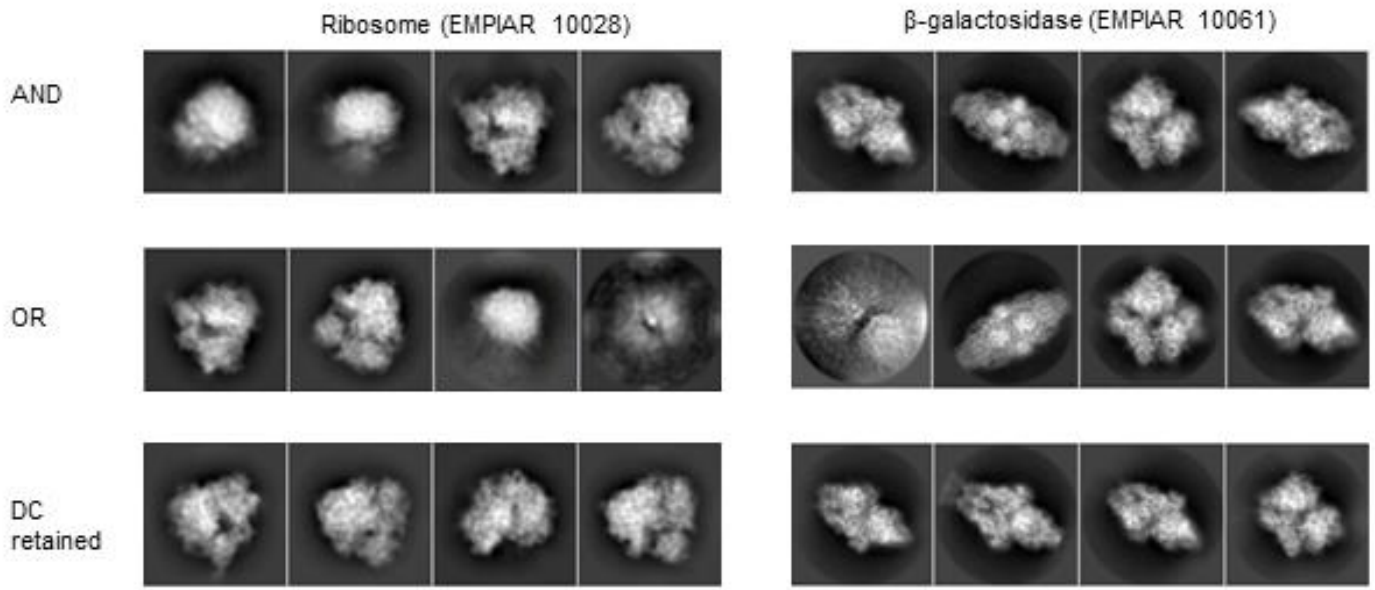


Consensus picking

Combining several particle pickings



Deep consensus

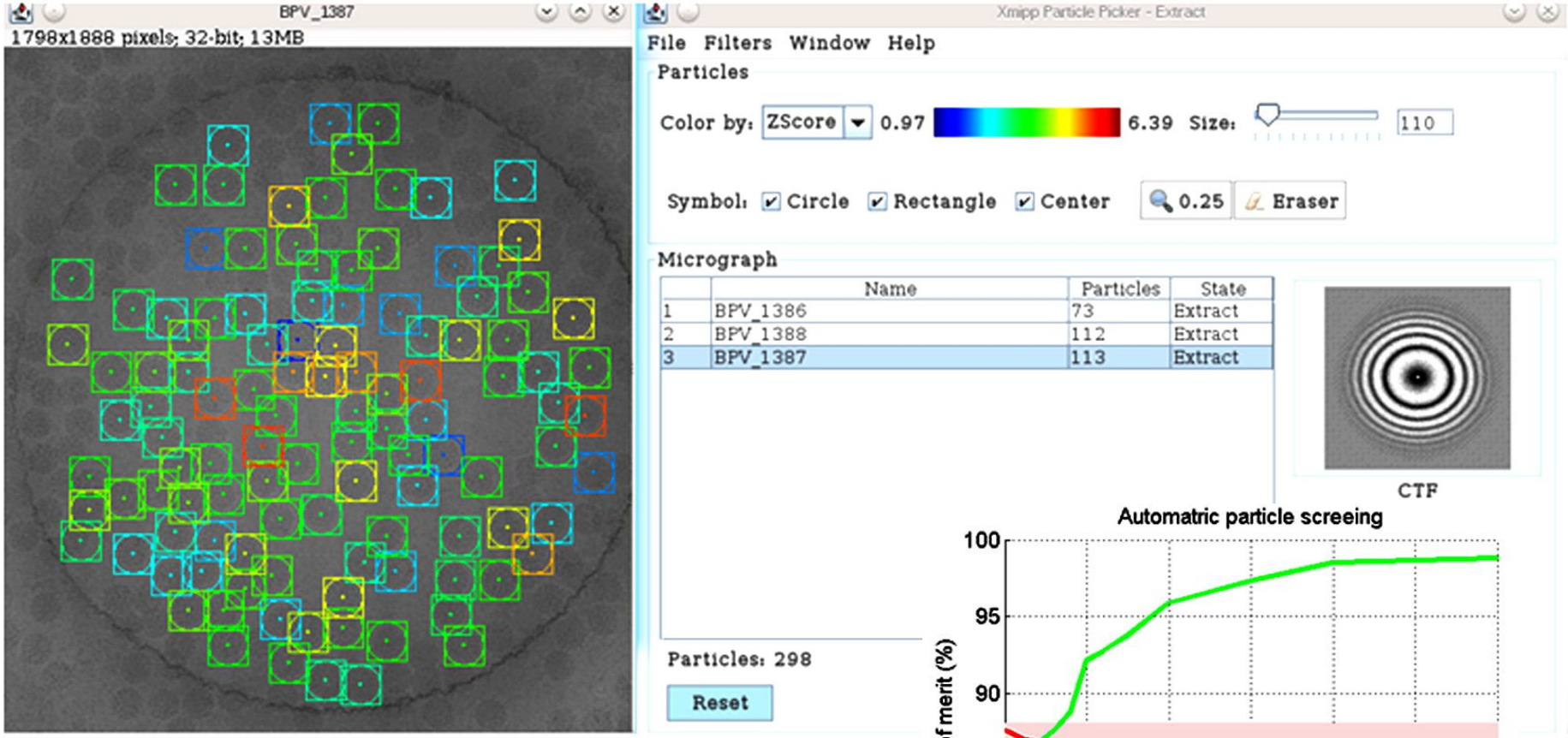


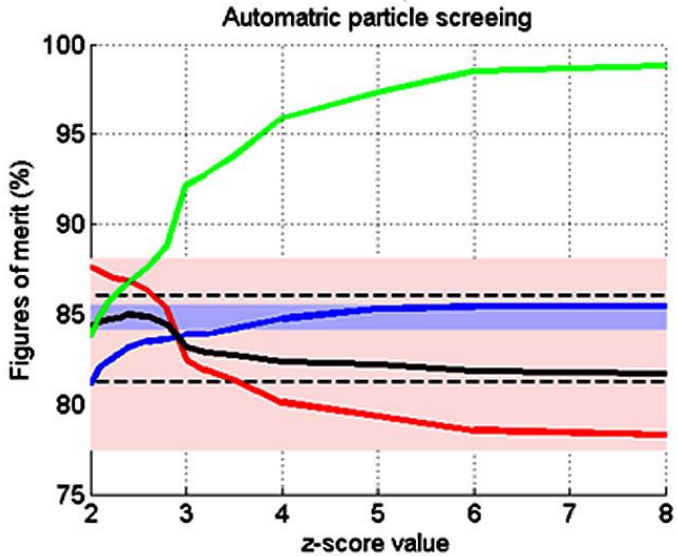
<2h

EMPIAR dataset	OR		AND		Best Picker		DC retained		DC pruned	
	R(Å)	N	R(Å)	N	R(Å)	N	R(Å)	N	R(Å)	N
10061	3.76	**231251**	3.32	25600	2.92	117047	**2.83**	**125586**	12.74	105665
10028	3.83	119171	3.87	67043	3.70	97561	**3.65**	88622	33.50	30549

Sanchez-Garcia, R.; Segura, J.; Maluenda, D.; Carazo, J. M. & Sorzano, C. O. S. Deep consensus, a deep learning-based approach for particle pruning in cryo-electron microscopy. IUCrJ, 2018, 5, 854-865

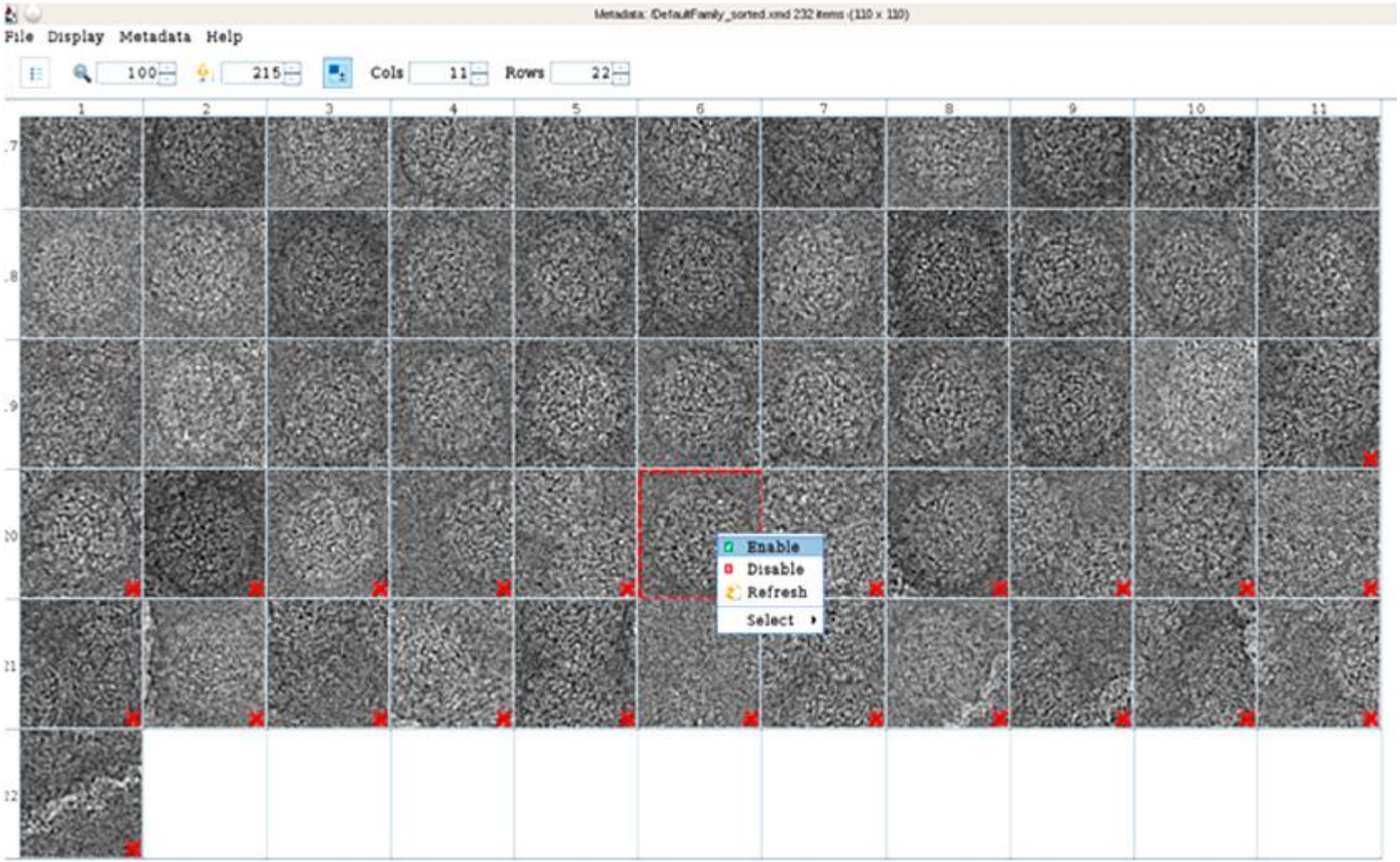
Detection of contaminants





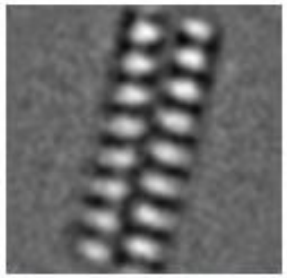
Vargas, … Sorzano. J. Structural Biology, 183: 342-353 (2013)

Detection of contaminants

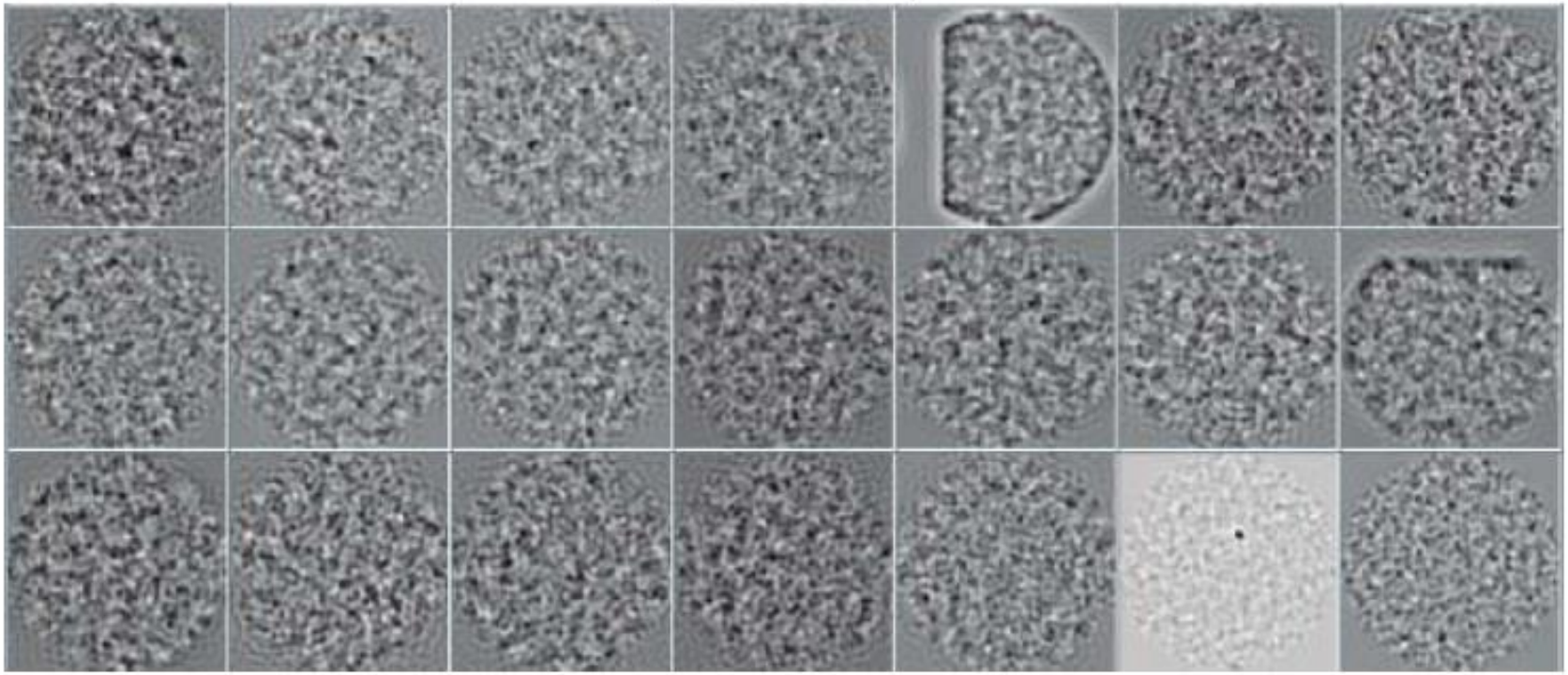


Vargas, … Sorzano. J. Structural Biology, 183: 342-353 (2013)

Detection of contaminants

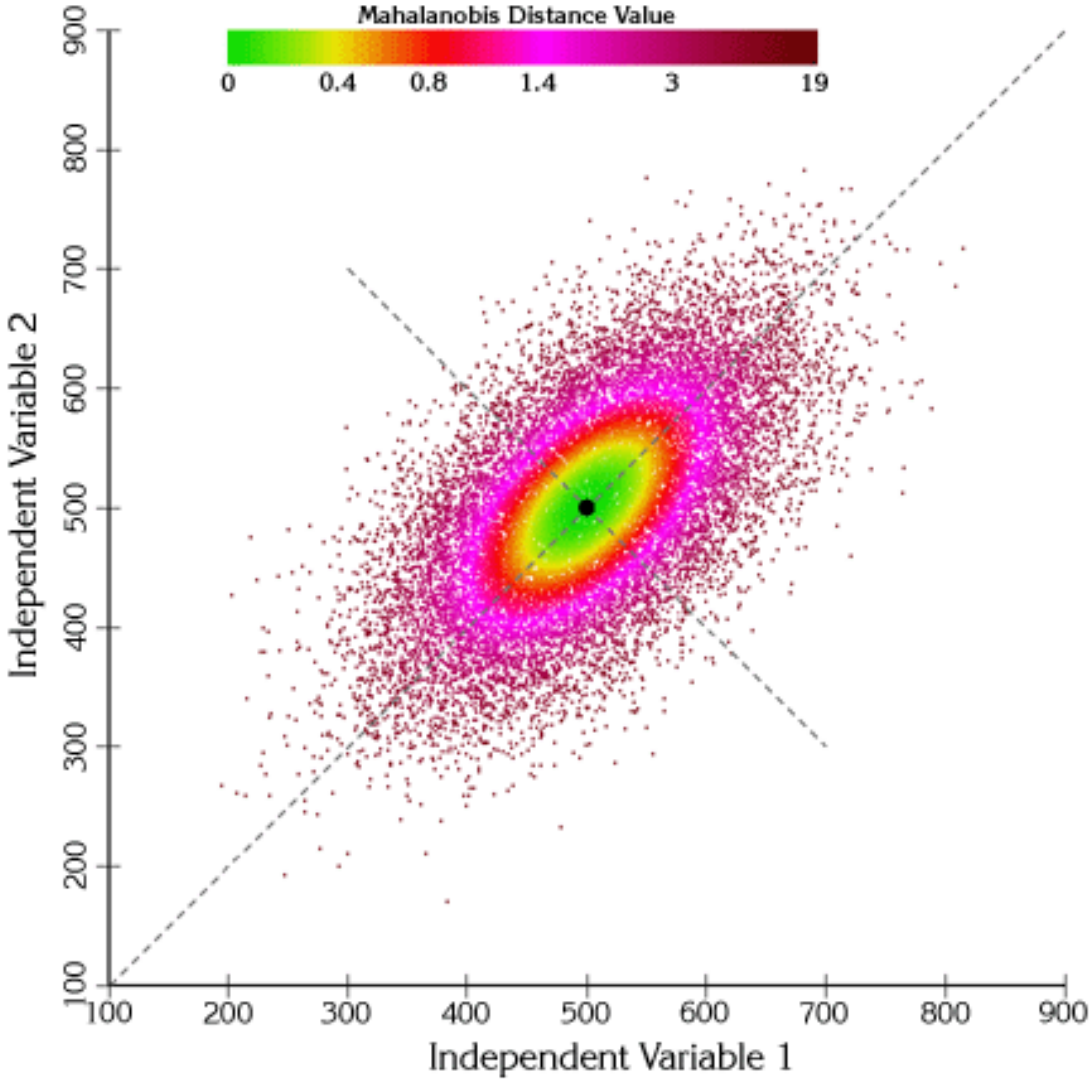


Class core
Stable core

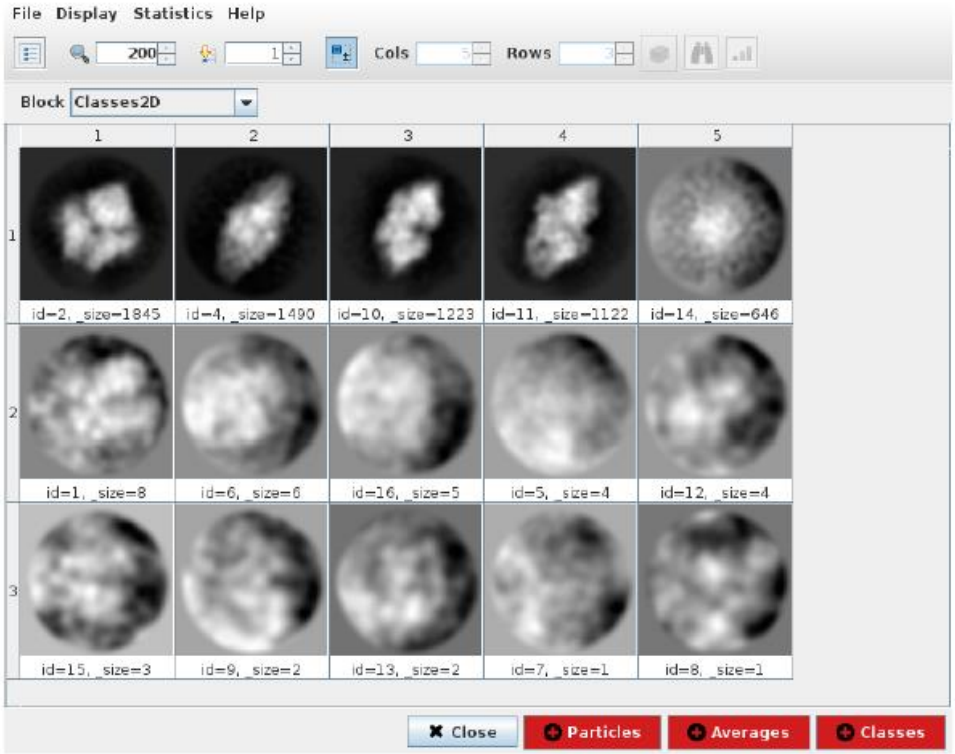


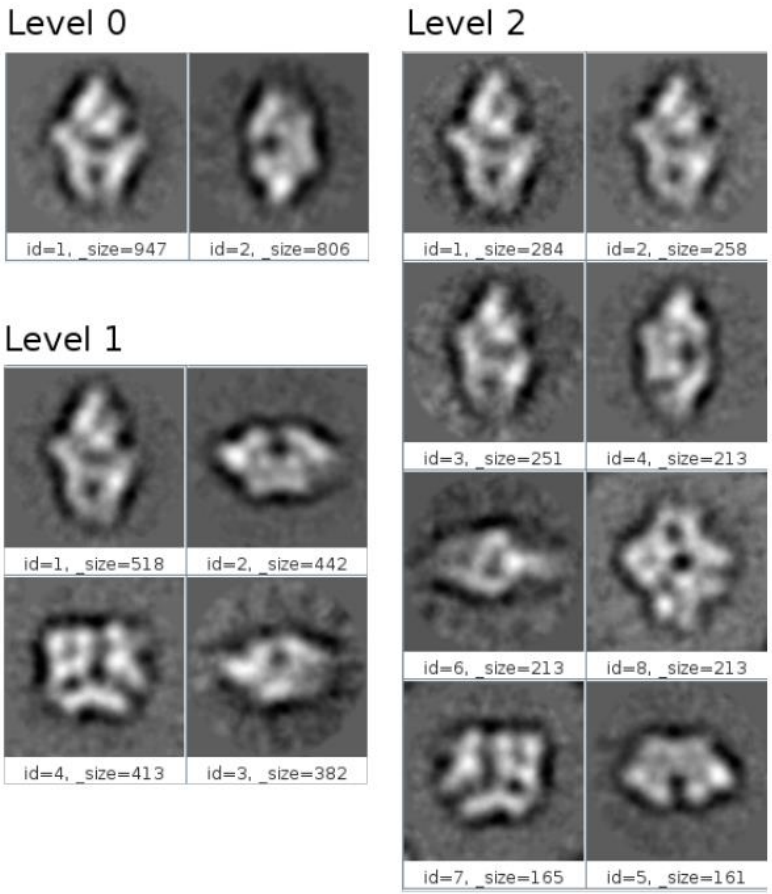
Sorzano et al., Proc. IWBBIO 2014, 950

Detection of contaminants



2D Classification



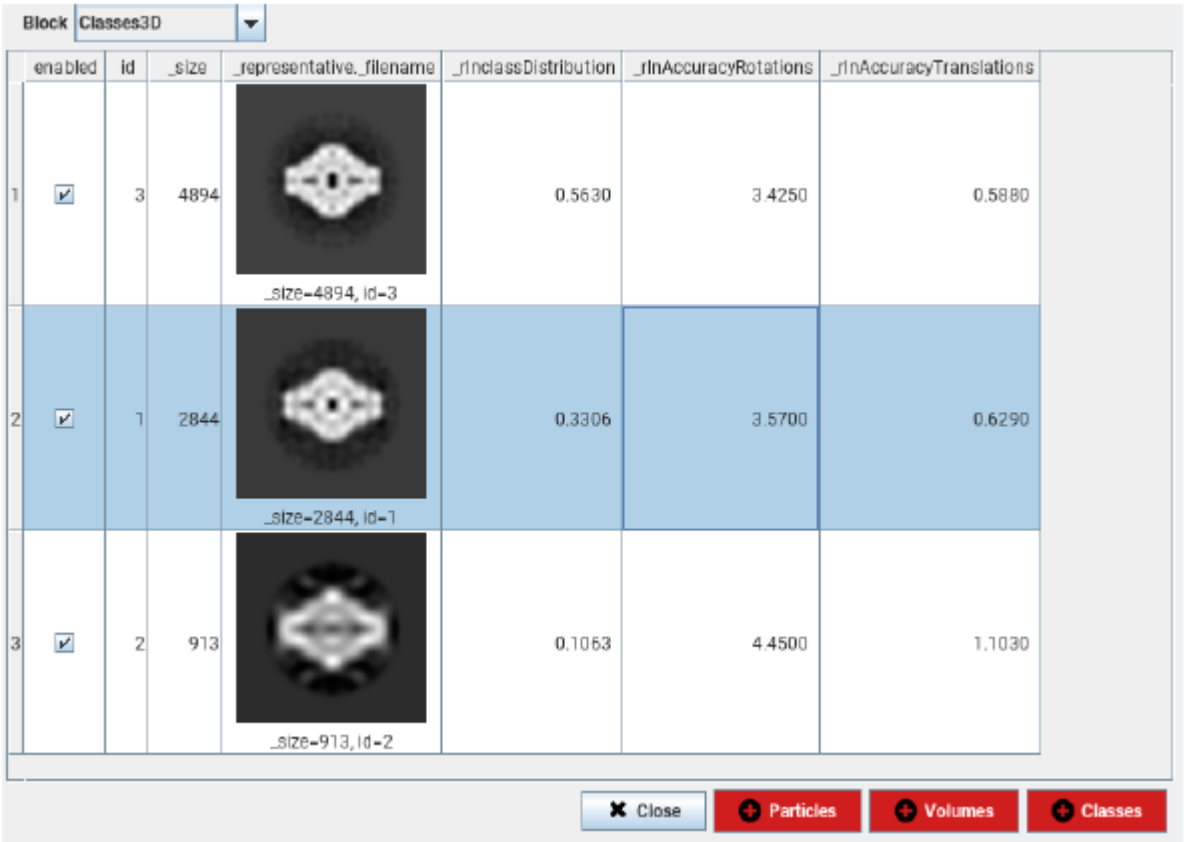


- Relion to detect bad classes

- CL2D to split classes
- Core and Stable core to remove bad particles

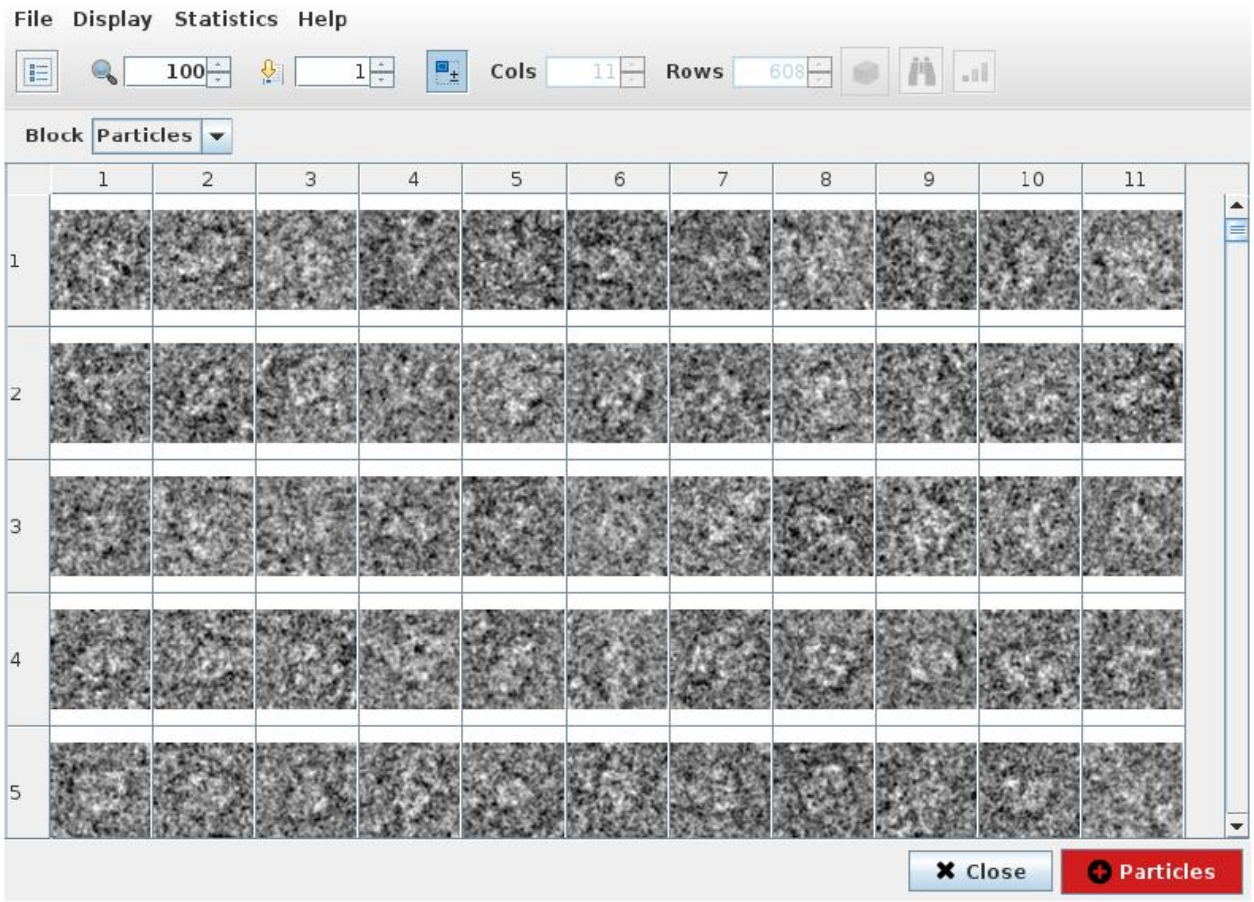
C.O.S. Sorzano, J.R. Bilbao-Castro, Y. Shkolnisky, M. Alcorlo, R. Melero, G. Caffarena-Fernández, M. Li, G. Xu, R. Marabini, J.M. Carazo. A clustering approach to multireference alignment of single-particle projections in electron microscopy. J. Structural Biology 171: 197-206 (2010)

Pruning through 3D Classification



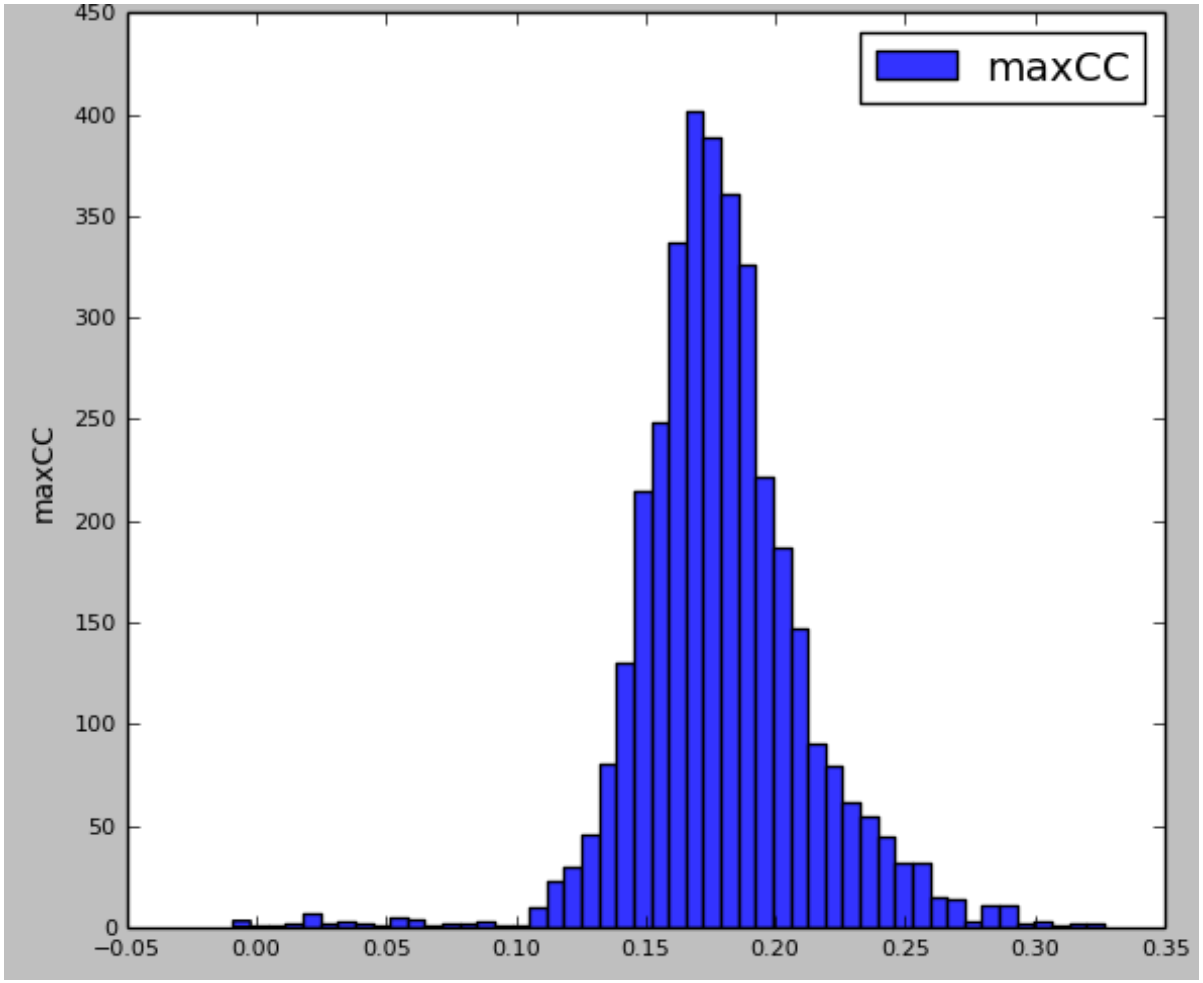
Scheres, S. H. W. Methods in Enzymology. The Resolution Revolution: Recent Advances In cryoEM Processing of structurally heterogeneous cryo-EM data in RELION Academic Press, 2016, 125-157

Pruning through Highres



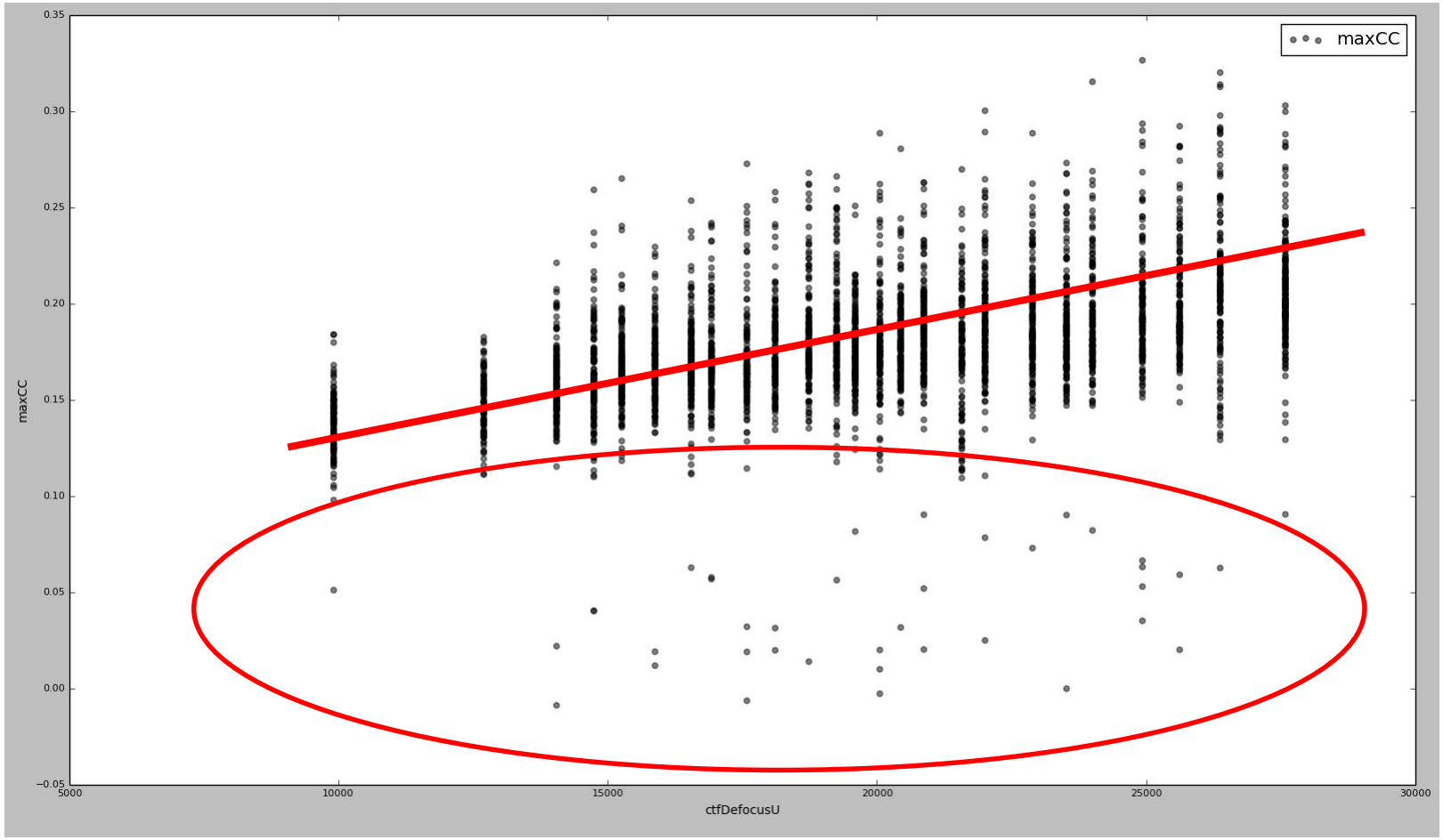
- Correlation
- Correlation percentile by defocus
- Cost
- Shift
- Gray scale correction

Correlation histograms

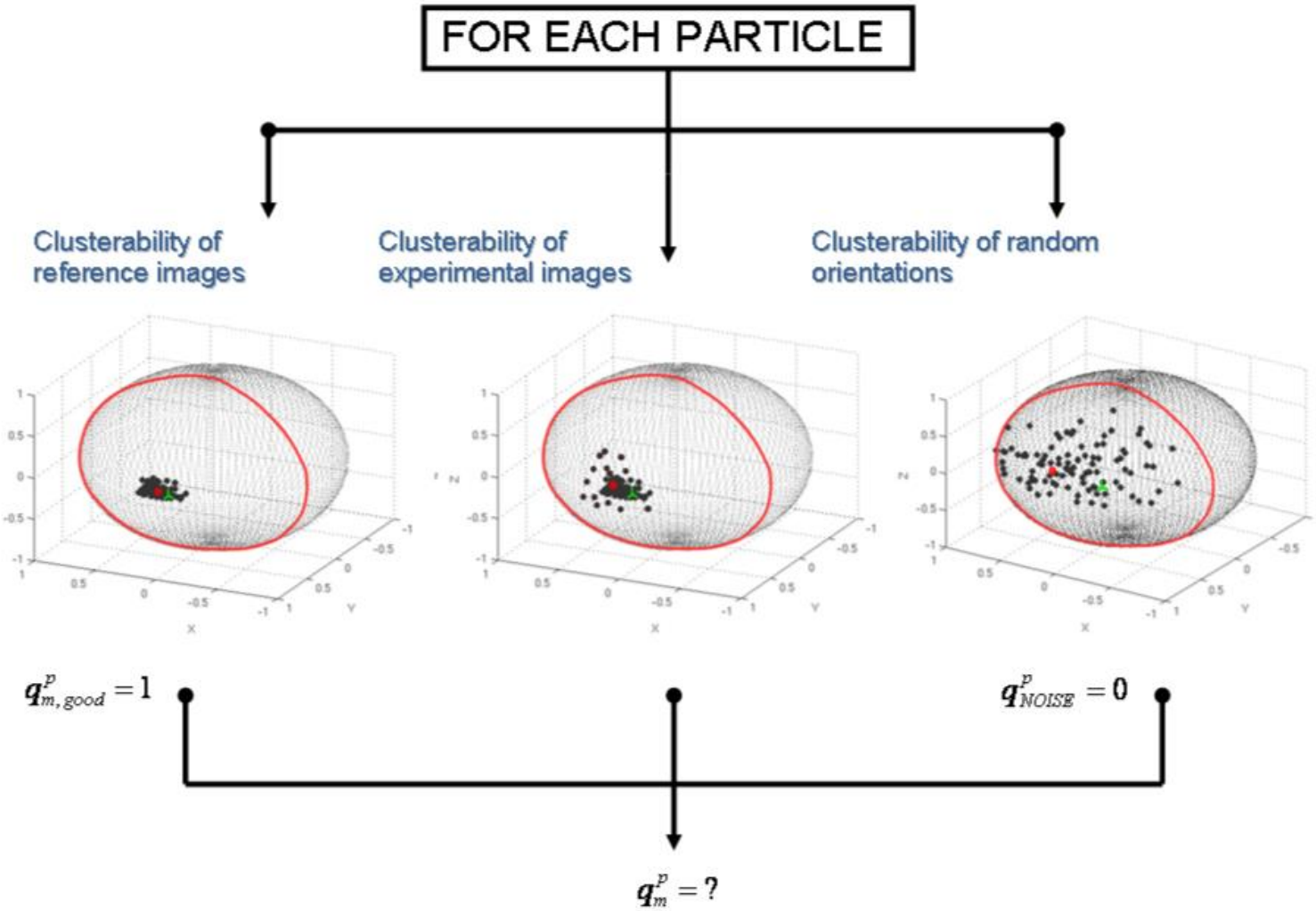


Scipion Prot: Highres, compare reprojections

Correlation and defocus

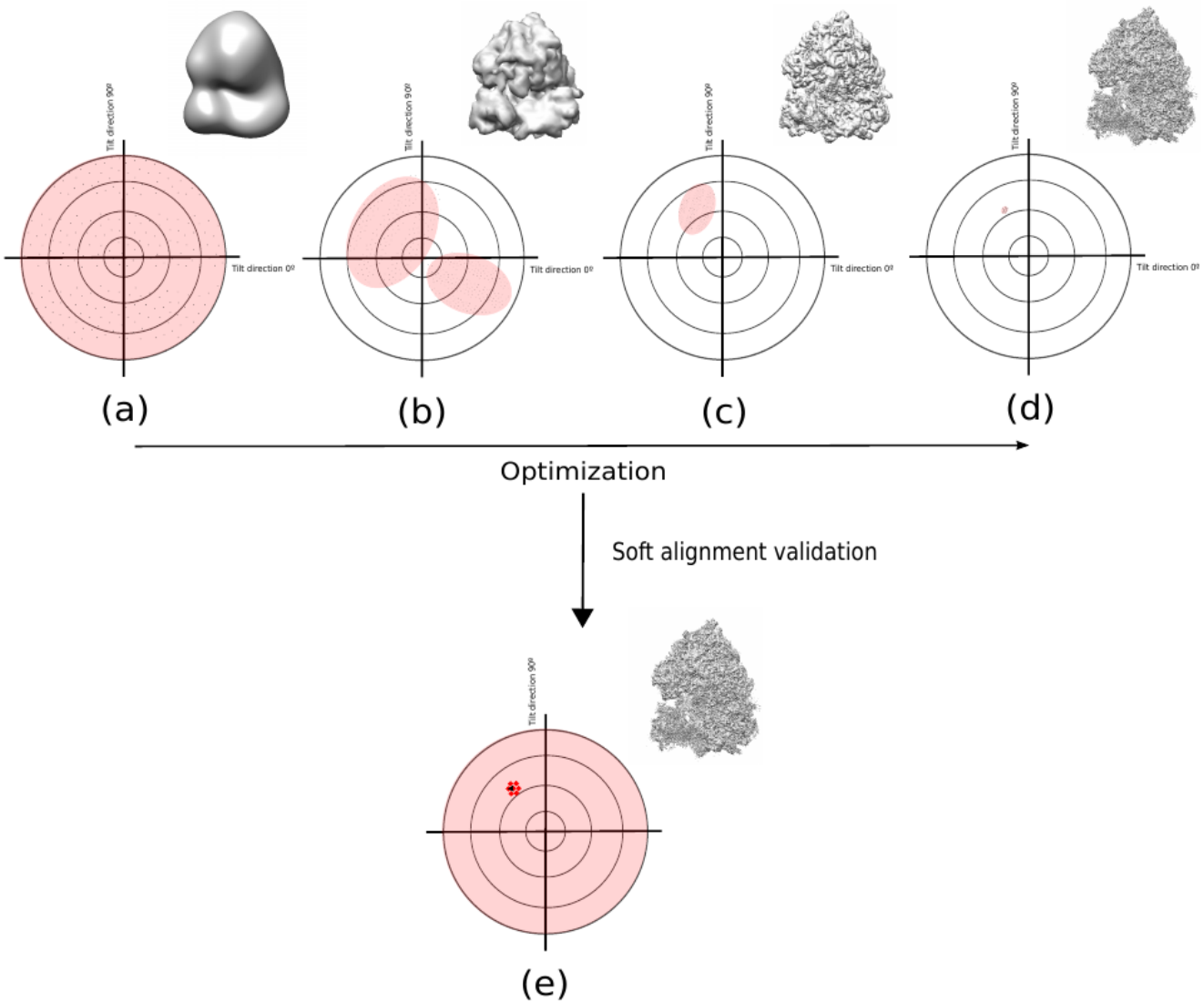


Pruning through soft-validation



J. Vargas, J. Otón, R. Marabini, J.M. Carazo, C.O.S. Sorzano. Particle alignment reliability in single particle electron cryomicroscopy: a general approach. Scientific Reports 6: 21626 (2016)

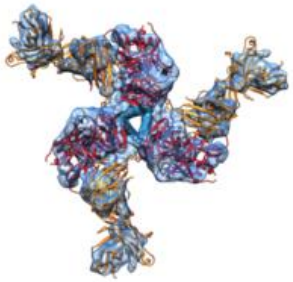
Pruning through soft-validation



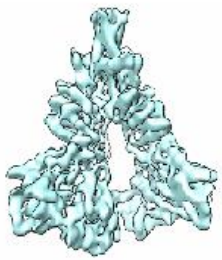
J. Vargas, R. Melero, J. Gómez-Blanco, J.M. Carazo, C.O.S. Sorzano. Quantitative analysis of 3D alignment quality: its impact on soft-validation, particle pruning and homogeneity analysis. Scientific Reports, 7: 6307 (2017)

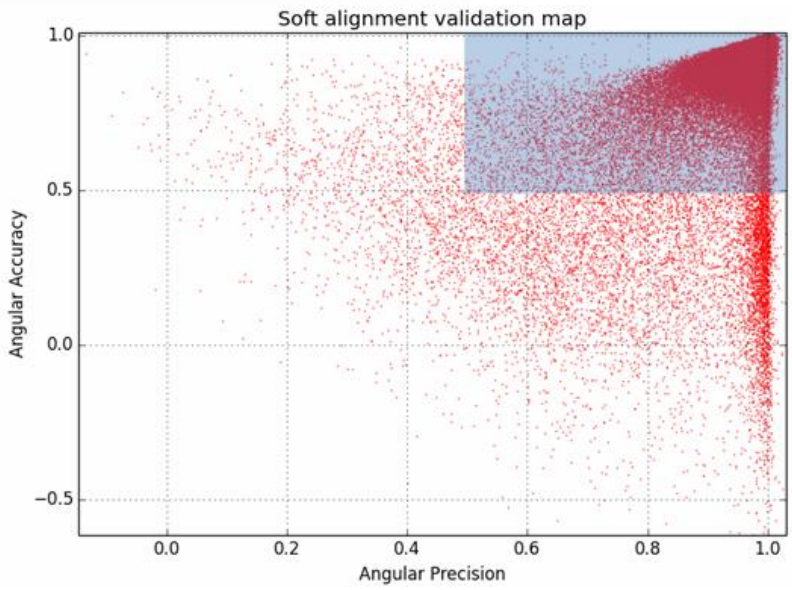
Pruning through soft-validation

EMDB 2484
EMPIAR 10008

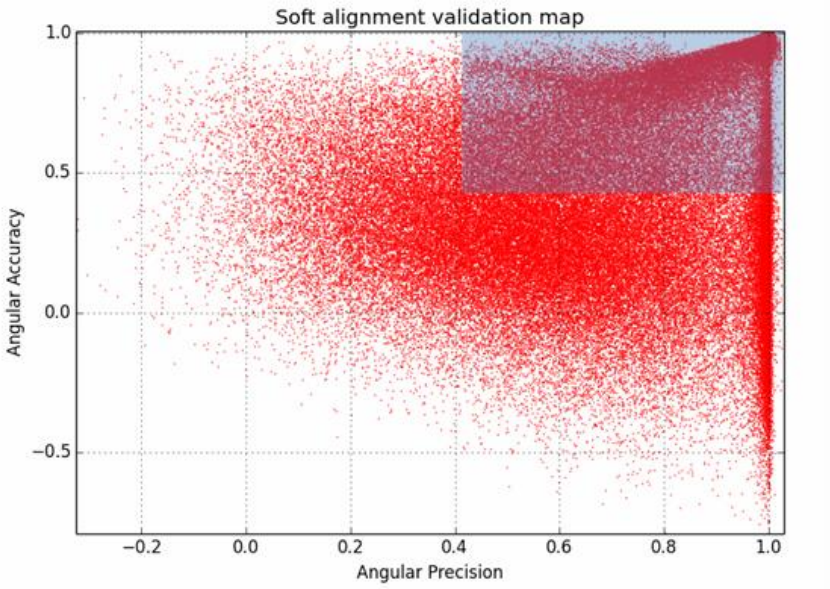


EMDB 5447
EMPIAR 10004



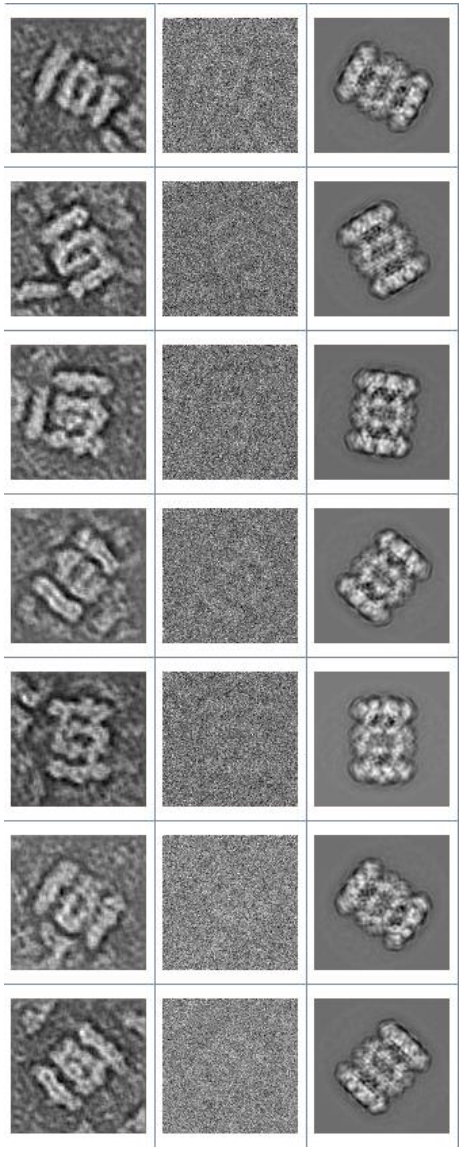


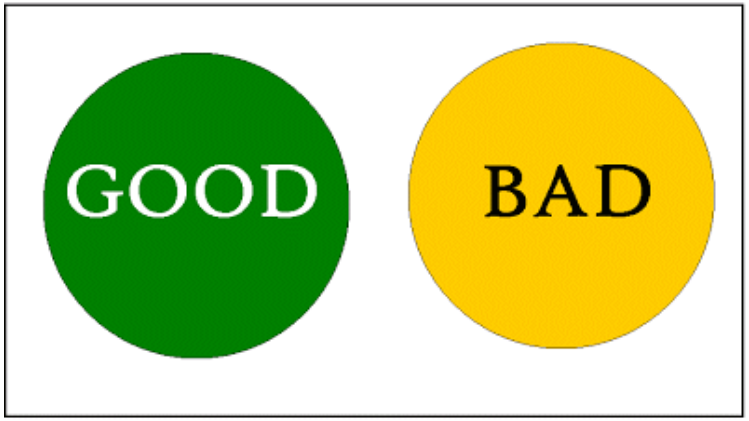
(a)

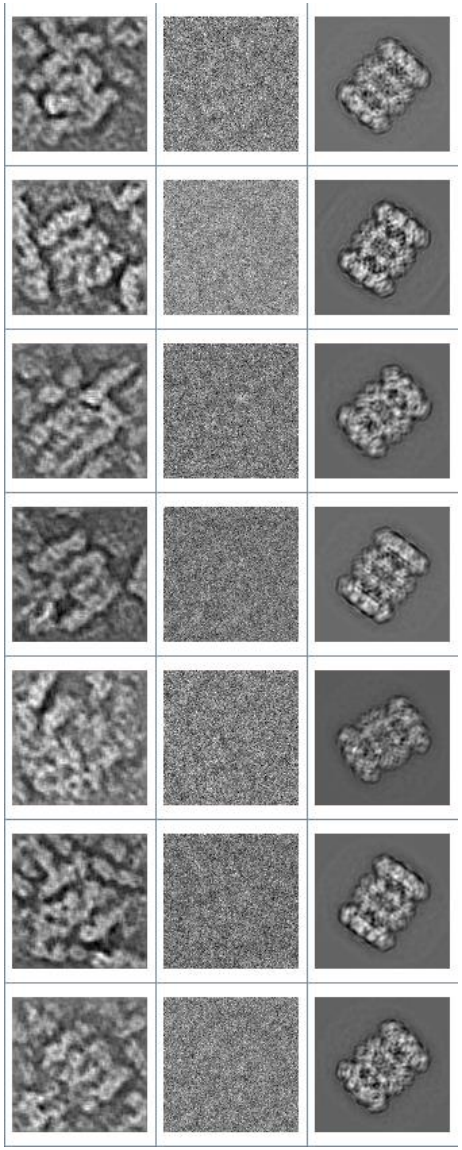


(b)

Deep denoising



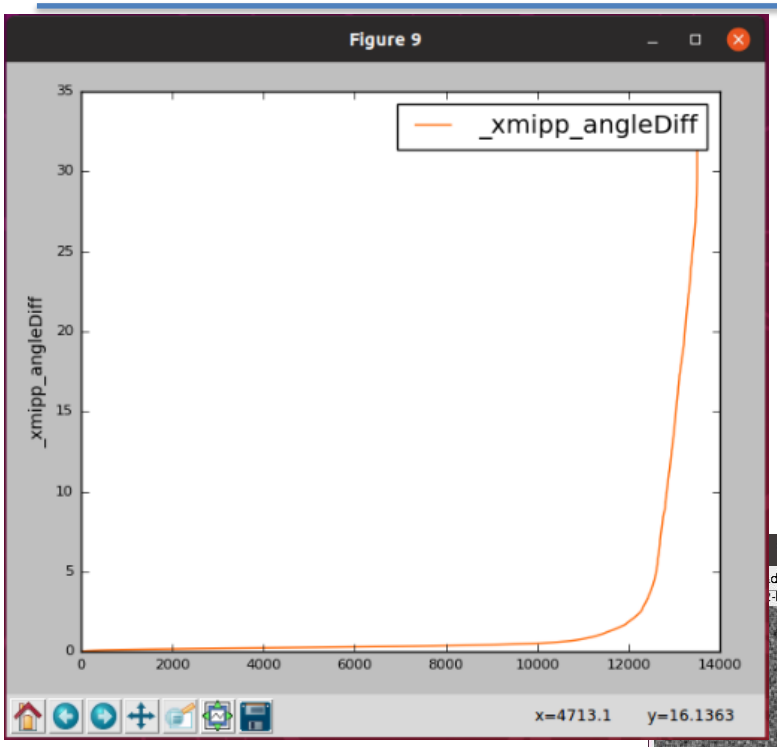




It can be used to:

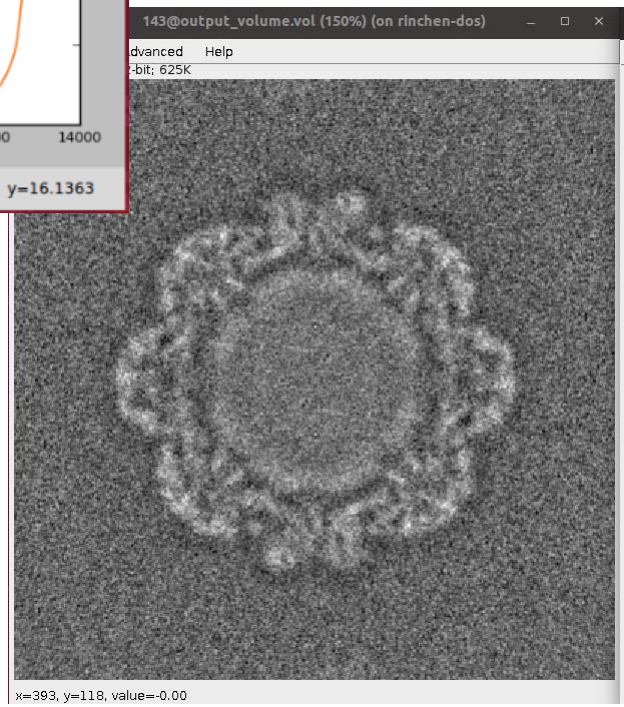
- Pruning
- Angular assignment
- Check angular assignment

Angular assignment consensus

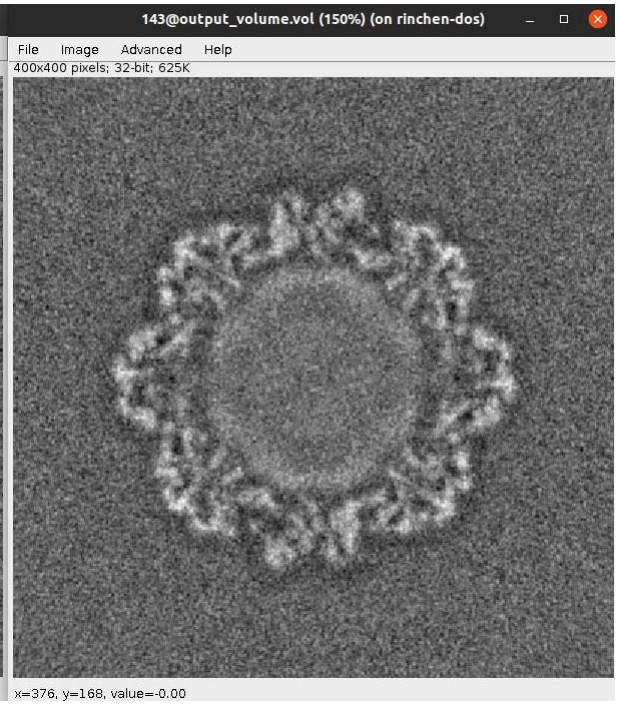


11% of differently assigned images!!!

1,500 most dissimilar Relion



1,500 most dissimilar highres



Particle pruning is still an open problem

EMBL-EBI

Protein Data Bank in Europe

Bringing Structure to Biology

EMDB statistics

EM Resources

- Home
- Statistics
- Validation
- EMDataBank
- EMPIAR
- Test data

EMDB

- Latest maps
- Latest headers
- Latest updates
- Search
- Browse
- FTP archive
- Deposit EM map/model
- EMDB data model

Select another chart: Single-particle released maps - resolution versus number of projections

Single-particle released maps - resolution versus number of projections

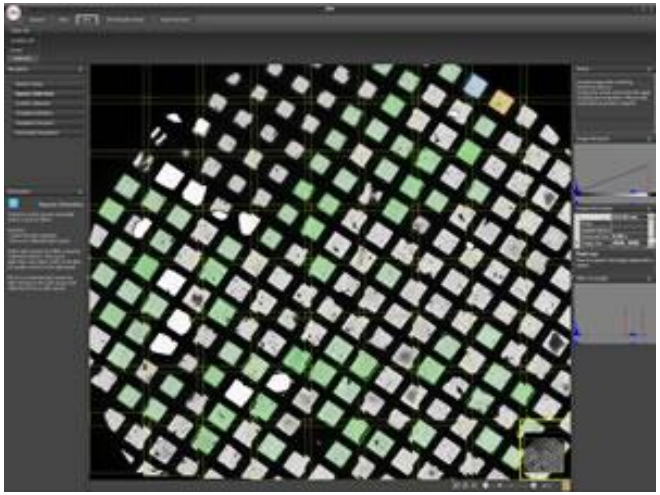
Initial volume problem in EM

C.O.S. Sorzano

Biocomputing Unit, CNB-CSIC

Instruct Image Processing Center

Towards high-throughput and high-resolution



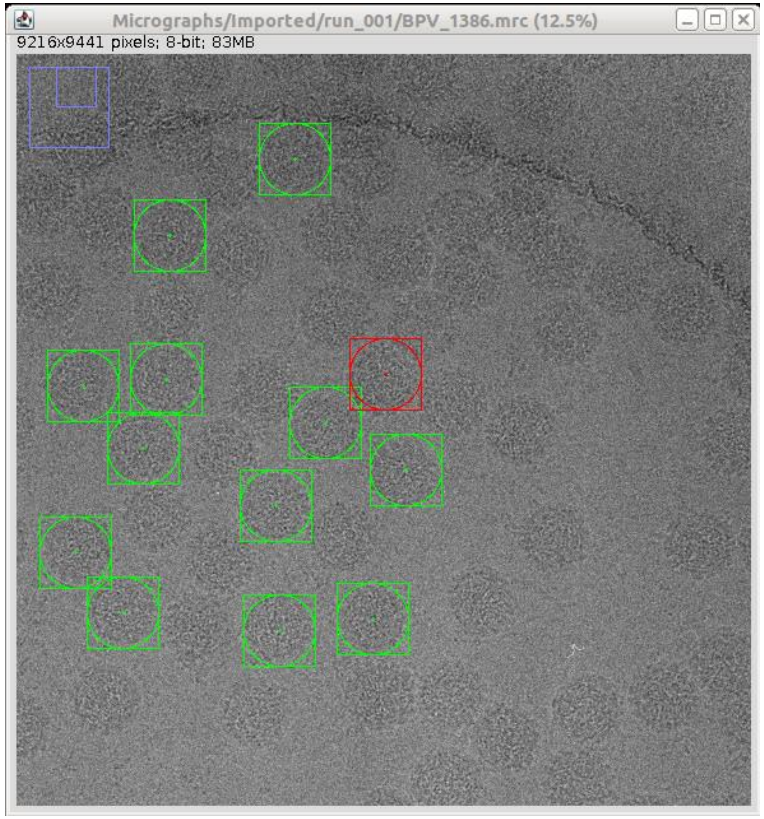
CSIC

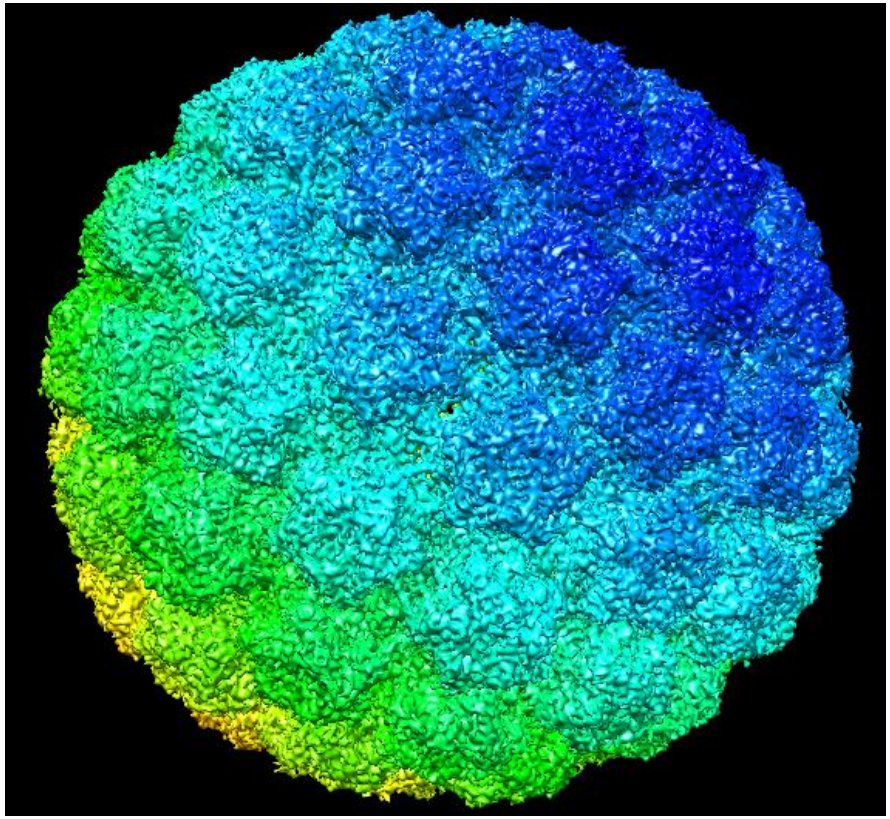
instruct Integrating Biology

BIO COMPUTING Unit

instruct image Processing Center

Towards high-throughput and high-resolution





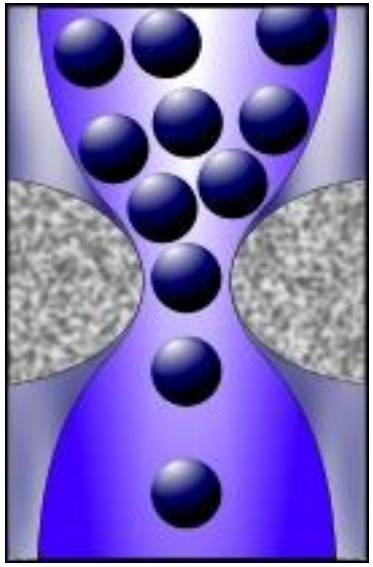
C.O.S. Sorzano, J. M. de la Rosa Trevín, J. Otón, J. J. Vega, J. Cuenca, A. Zaldívar-Peraza, J. Gómez-Blanco, J. Vargas, A. Quintana, R. Marabini, J. M. Carazo. *Semiautomatic, high-throughput, high-resolution protocol for three-dimensional reconstruction of Single Particles in Electron Microscopy*. Nanoimaging: Methods and Protocols. Methods in Molecular Biology, 950: 171-193. Eds. Alioscka Sousa, Michael Kruhlak. Humana Press. (2012)

CSIC

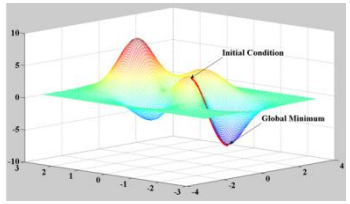
instruct Integrating Biology

instruct image Processing Center

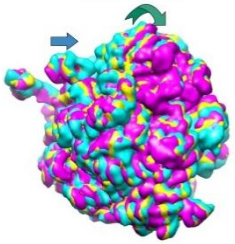
But … there are a few bottlenecks



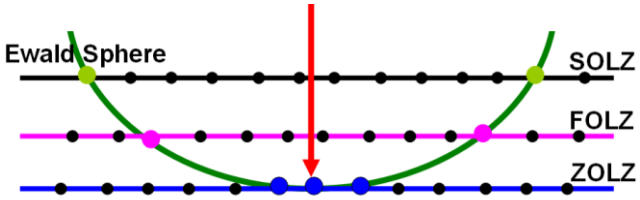
- Initial volume



- 3D heterogenity



- High-resolution modeling

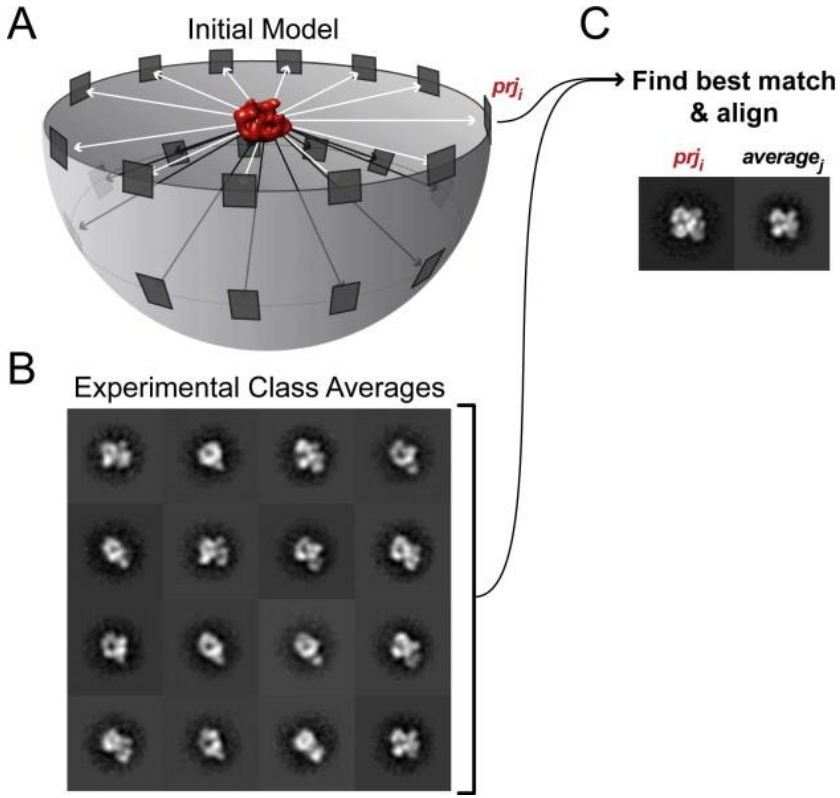


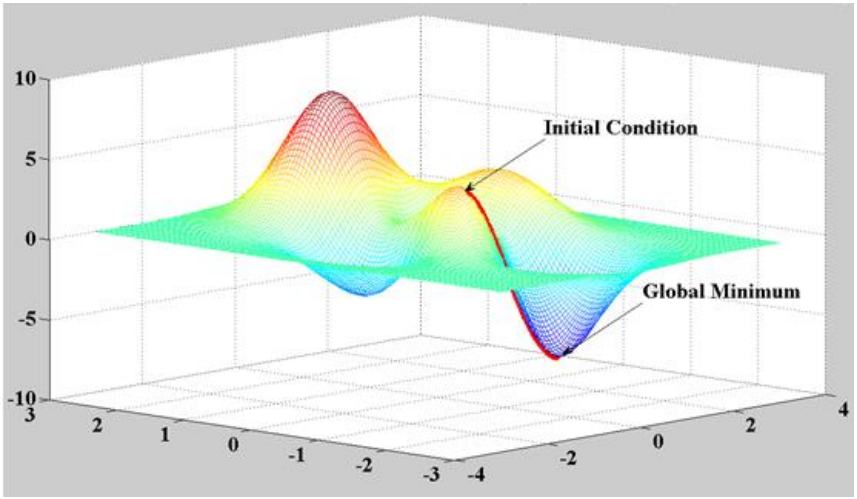
CSIC

instruct Integrating Biology

instruct image Processing Center

3D Refinement is a greedy algorithm





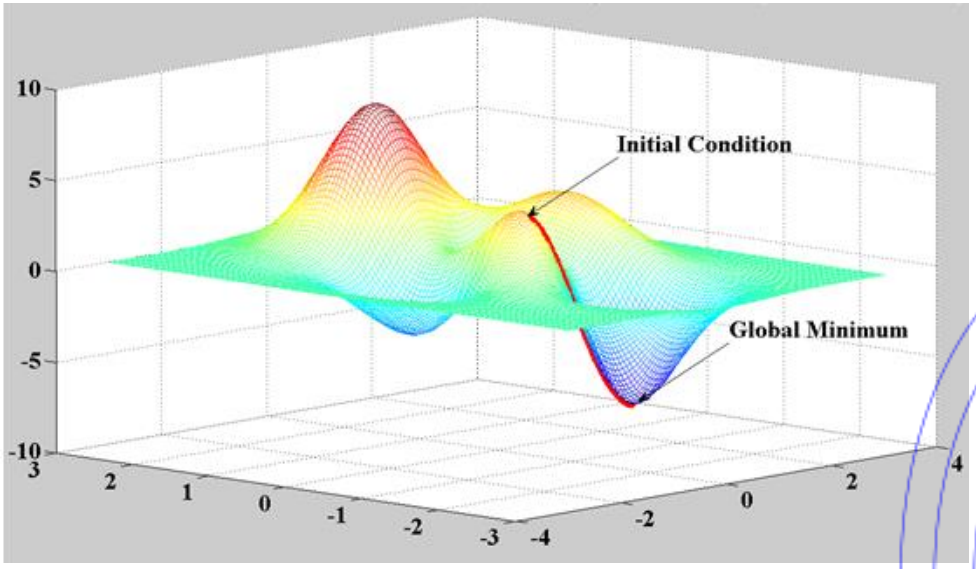
Chandramouli, P.; Hernandez-Lopez, R.; Wang, H.-W. & Leschziner, A. E. Validation of the orthogonal tilt reconstruction method with a biological test sample. J Struct Biol, 2011, 175, 85-96

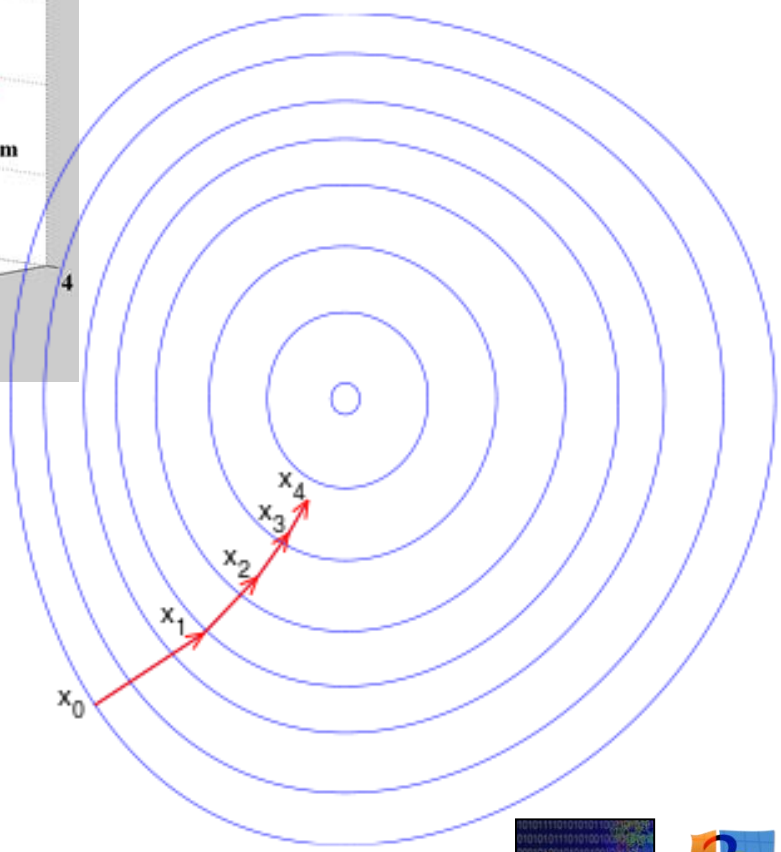
CSIC

instruct Integrating Biology

instruct image Processing Center

3D Refinement is a greedy algorithm





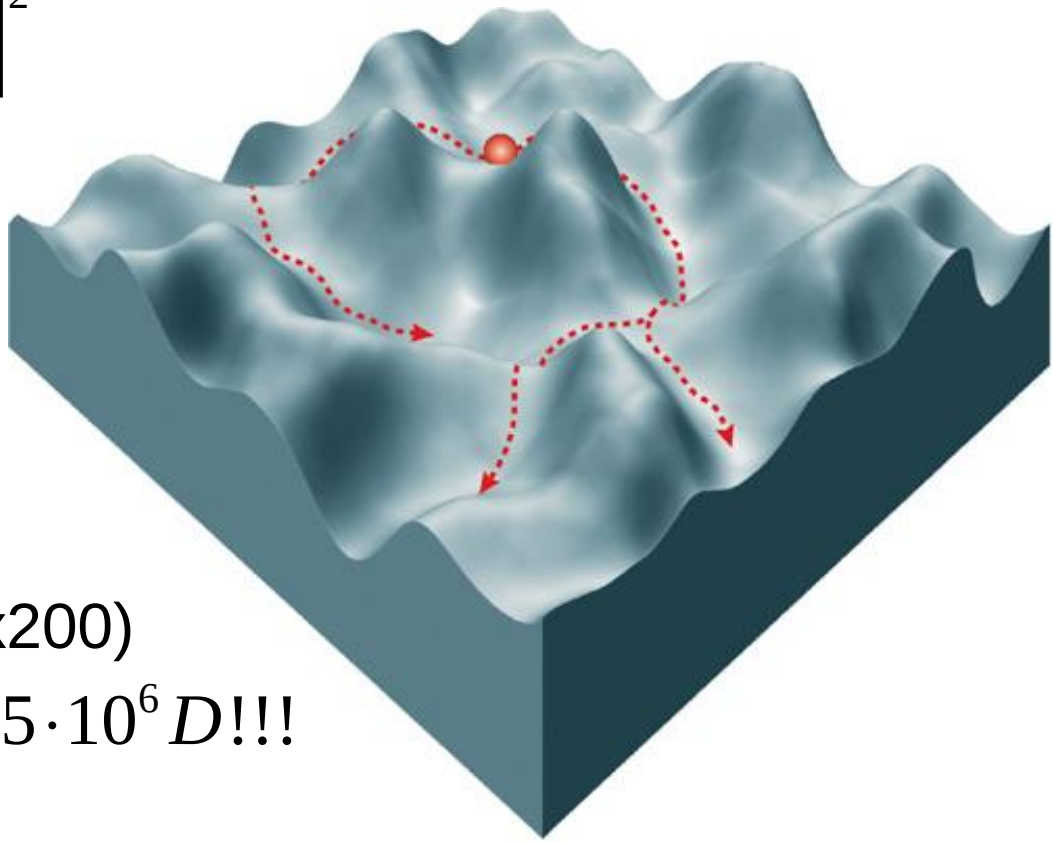
CSIC

instruct Integrating Biology

instruct image Processing Center

Goal function landscape

$$\min_{V,\theta} \sum_i \left\| I_i - P_{\theta_i} V \right\|^2$$



100k projections (200x200)

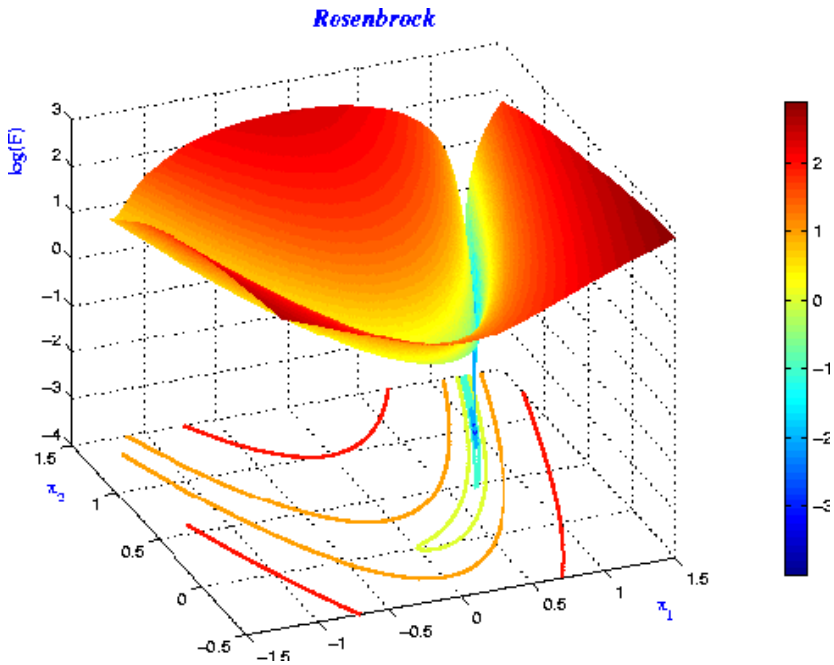
$$\Rightarrow 200^3 + 5 \cdot 10^5 = 8.5 \cdot 10^6 D!!!$$

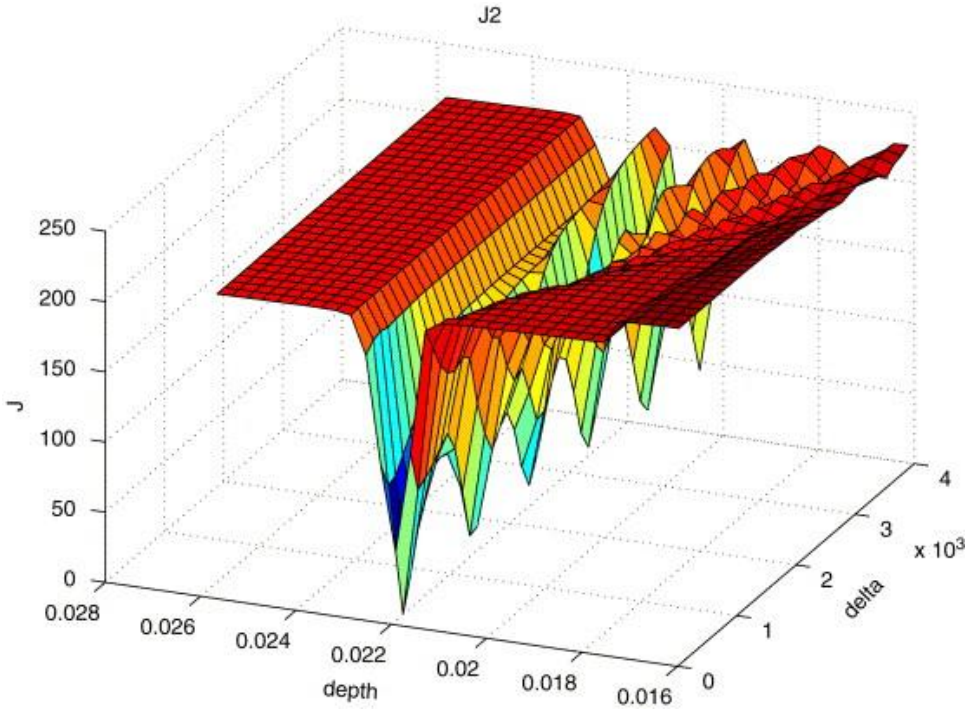
CSIC

instruct Integrating Biology

instruct image Processing Center

Goal function landscape



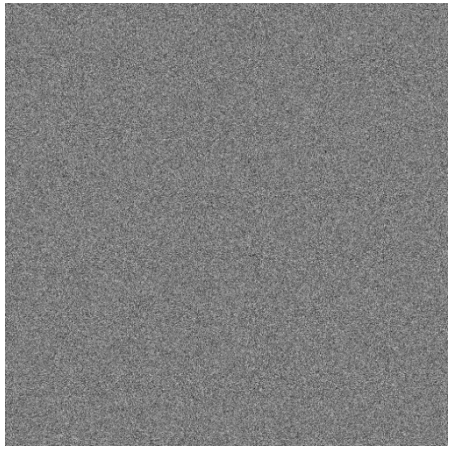


CSIC

instruct Integrating Biology

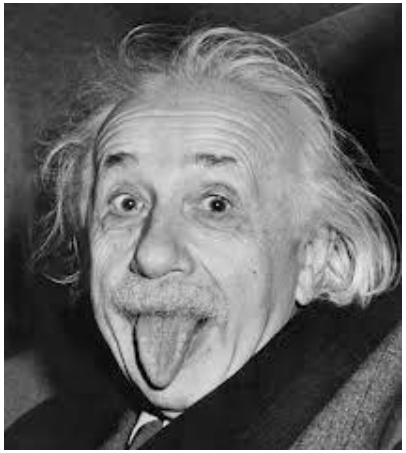
instruct image Processing Center

Model bias



1000 images

$\otimes$



$=$

Model bias=local minimum!!

Shatsky, M.; Hall, R. J.; Brenner, S. E. & Glaeser, R. M. A method for the alignment of heterogeneous macromolecules from electron microscopy. J Struct Biol, 2009, 166, 67-78

CSIC

instruct Integrating Biology

instruct image Processing Center

Initial volume problem

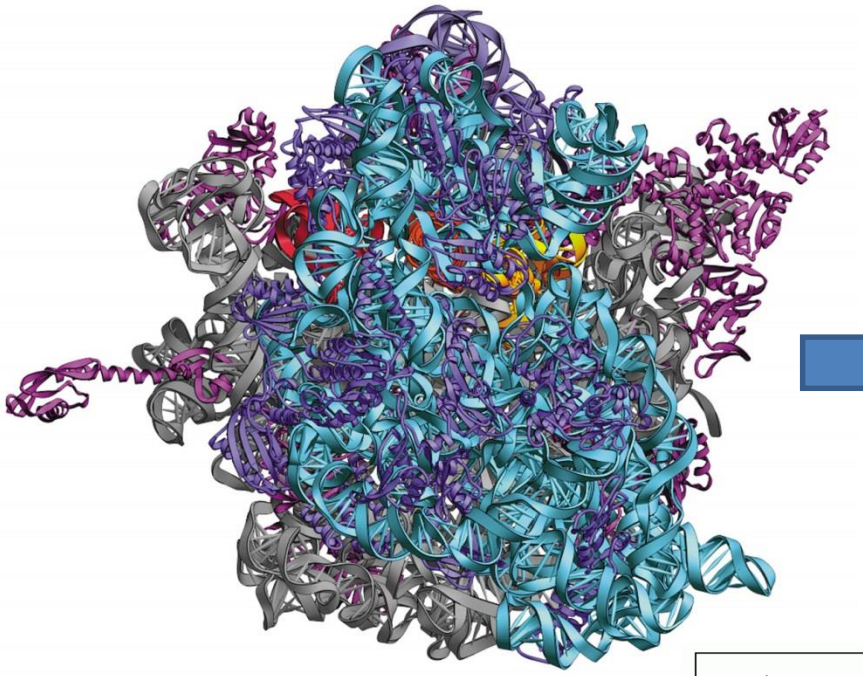
How to choose the initial volume?

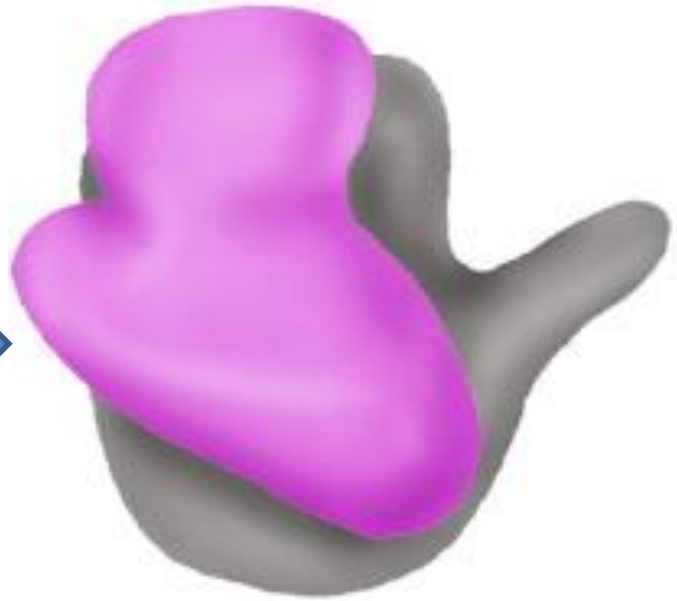
CSIC

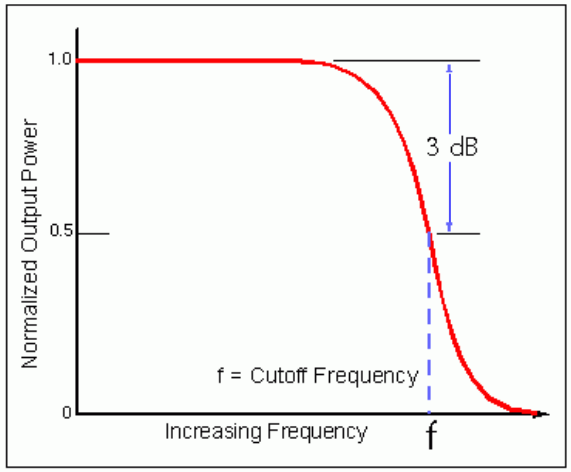
instruct Integrating Biology

instruct image Processing Center

Option 1: Filtered volume







low pass filter

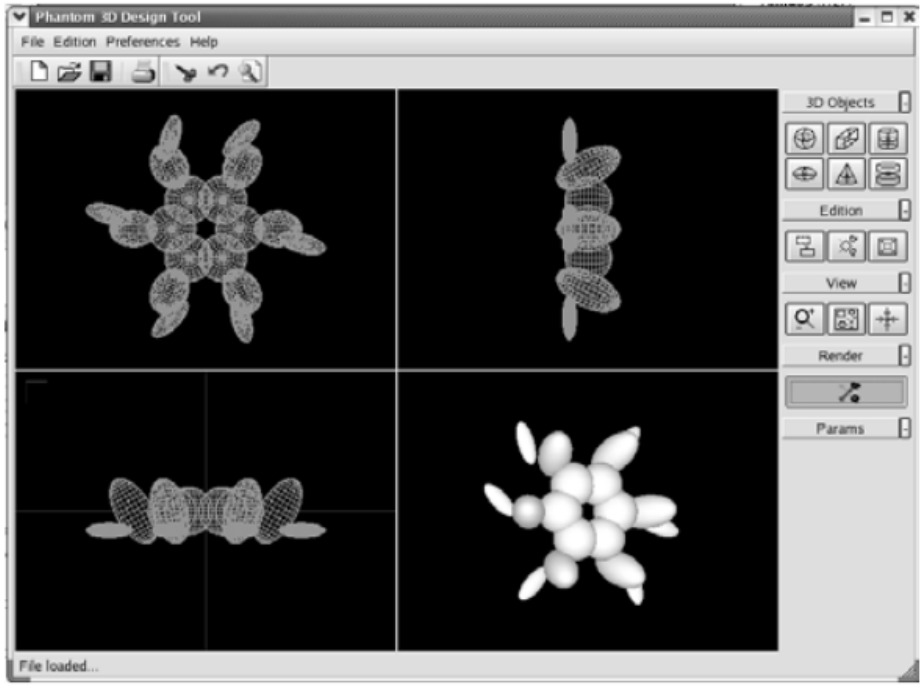
CSIC

instruct Integrating Biology

instruct image Processing Center

Option 2: Geometrical description

```
#Xdim Ydim Zdim Background_density
128 128 128 0
# Feature Parameters:
# 6 main ellipses
ell = 1 30.00 0.00 0 15 15 30 0 30 0
ell = 1 15.00 25.9808 0 15 15 30 60 30 0
ell = 1 -15.00 25.9808 0 15 15 30 120 30 0
ell = 1 -30.00 0.00 0 15 15 30 0 -30 0
ell = 1 -15.00 -25.9808 0 15 15 30 -120 30 0
ell = 1 15.00 -25.9808 0 15 15 30 -60 30 0
# 6 inner ring features
sph = 1 14.00 0.00 0 10
sph = 1 7.00 12.1235 0 10
sph = 1 7.00 -12.1235 0 10
sph = 1 -7.00 12.1235 0 10
sph = 1 -7.00 -12.1235 0 10
sph = 1 -14.00 0.00 0 10
# 6 outer features providing chirality
ell = 1 40.00 -6.00 -8 20 8 6 -15 0 0
ell = 1 14.80 37.64 -8 20 8 6 45 0 0
ell = 1 -14.80 37.64 -8 20 8 6 75 0 0
ell = 1 -40.00 6.00 -8 20 8 6 -15 0 0
ell = 1 -14.80 -37.64 -8 20 8 6 45 0 0
ell = 1 14.80 -37.64 -8 20 8 6 -75 0 0
```



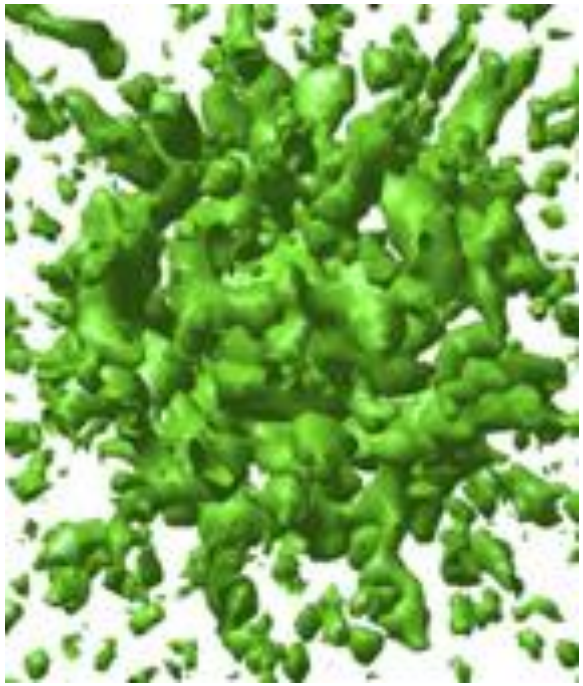
Bilbao-Castro, J. R.; Sorzano, C. O. S.; García, I. & Fernández, J. J.
Phan3D: design of biological phantoms in 3D electron microscopy
Bioinformatics, 2004, 20, 3286-3288

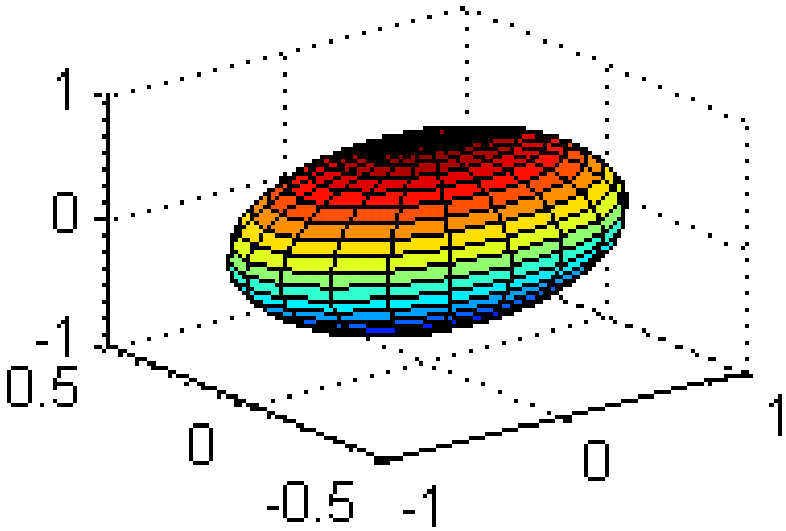
CSIC

instruct Integrating Biology

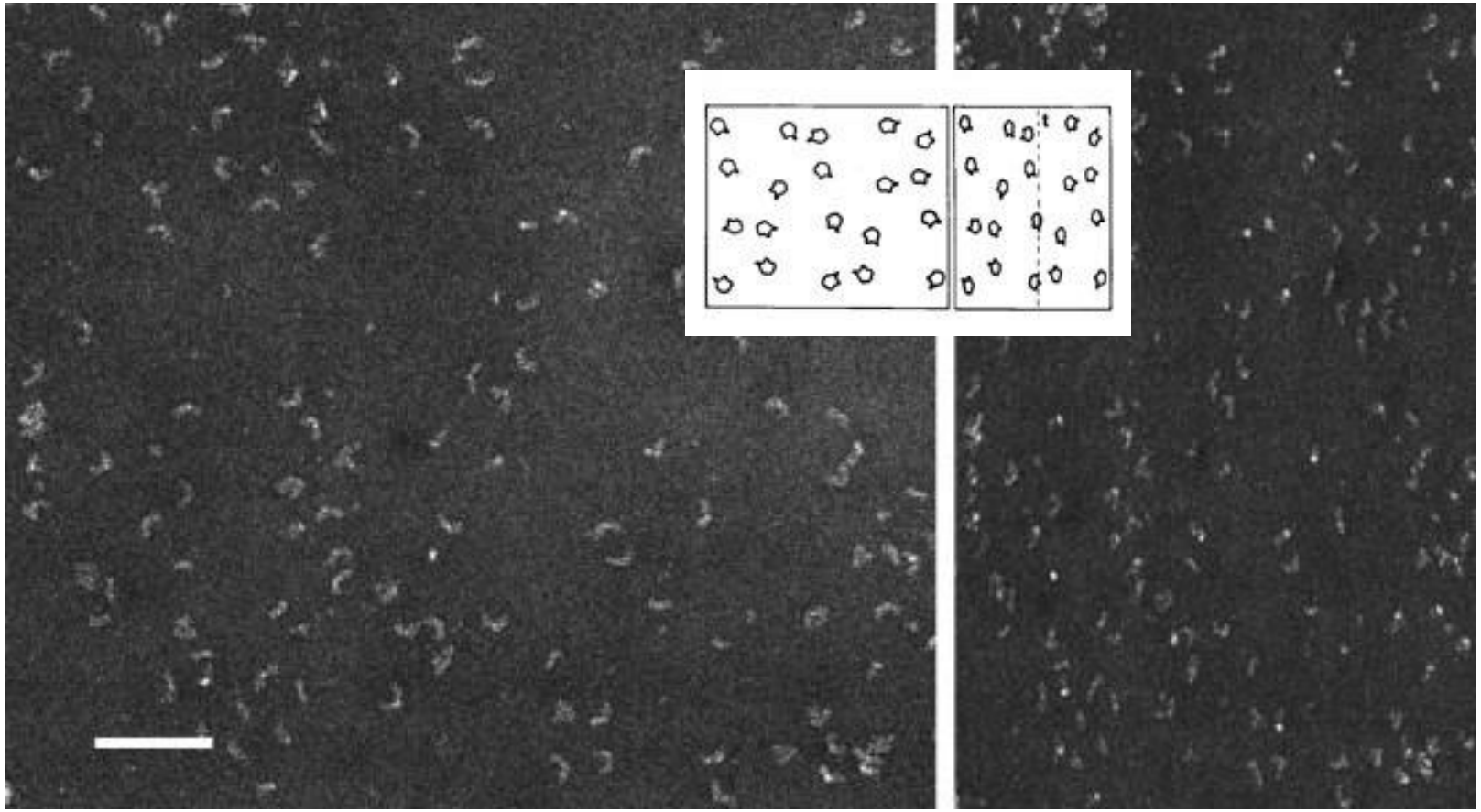
instruct image Processing Center

Option 3: Noisy blob or blob





Option 4: Random Conical Tilt

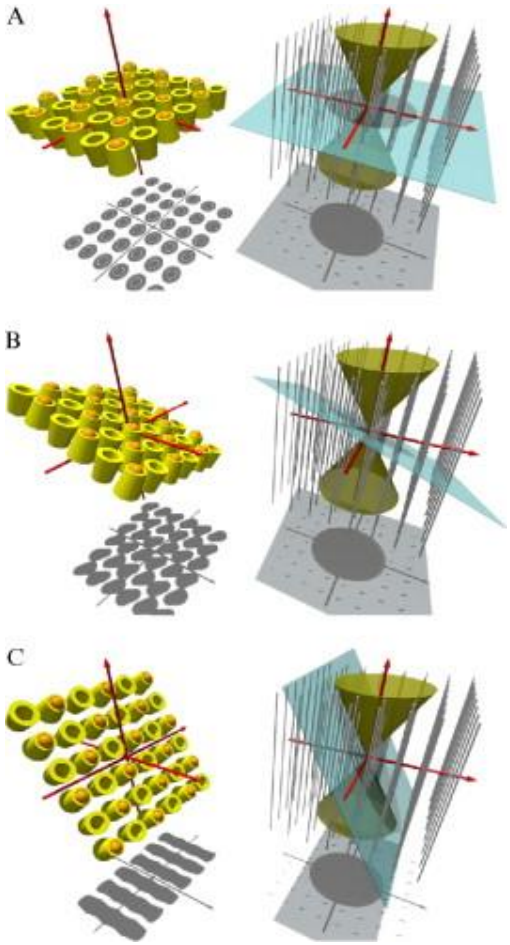


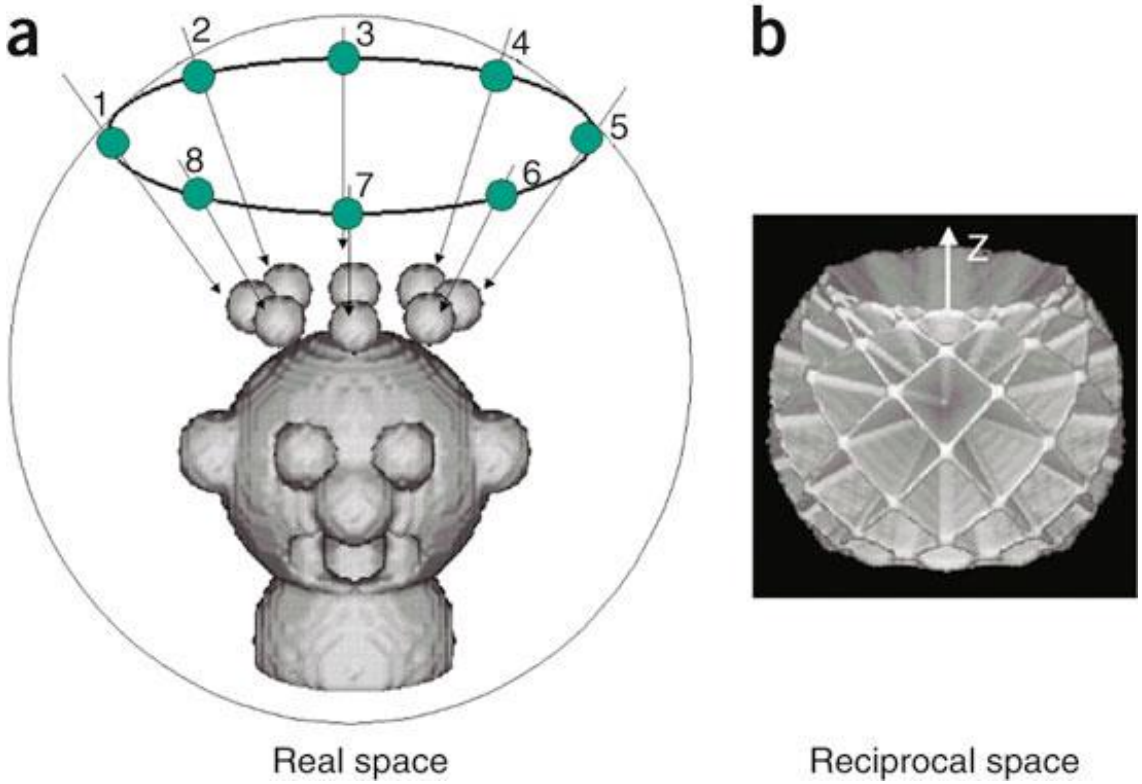
CSIC

instruct Integrating Biology

instruct image Processing Center

Option 4: Random Conical Tilt

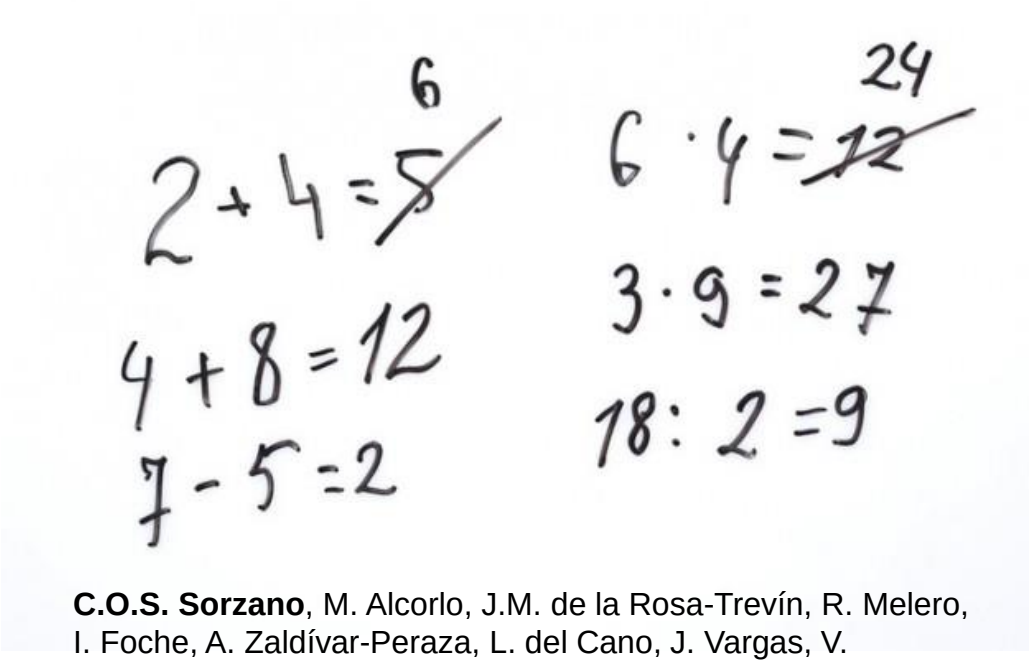




Schenk, A. D.; Castaño-Díez, D.; Gipson, B.; Arheit, M.; Zeng, X. & Stahlberg, H. 3D reconstruction from 2D crystal image and diffraction data. Methods Enzymol, 2010, 482, 101-129

Option 4: Random Conical Tilt

In practice: Not always work
Important simplifications in the derivation



C.O.S. Sorzano, M. Alcorlo, J.M. de la Rosa-Trevín, R. Melero, I. Foche, A. Zaldívar-Peraza, L. del Cano, J. Vargas, V. Abrishami, J. Otón, R. Marabini, J.M. Carazo. *Cryo-EM and the elucidation of new macromolecular structures: Random Conical Tilt revisited*. Nature Scientific Reports, 5: 14290 (2015)

CSIC

instruct Integrating Biology

instruct image Processing Center

Option 5: Computational means

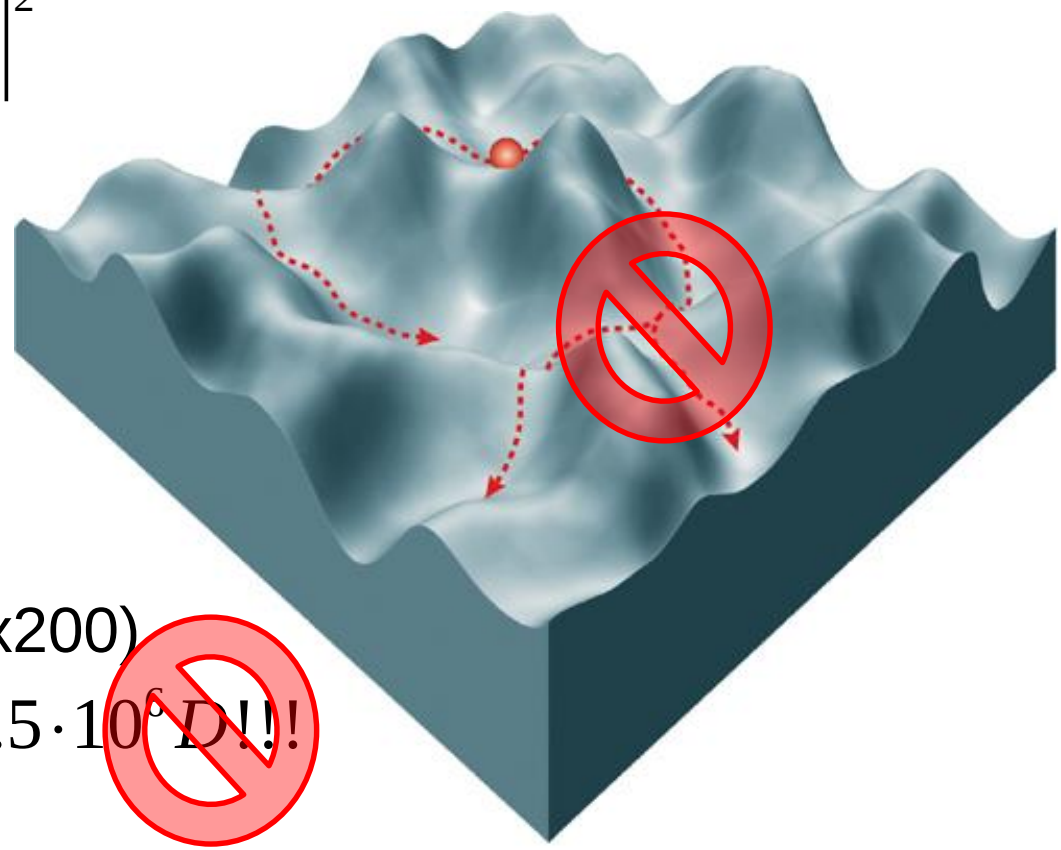
CSIC

instruct Integrating Biology

instruct image Processing Center

Option 5: Computational means

$$\min_{V,\theta} \sum_i \left\| I_i - P_{\theta_i} V \right\|^2$$



100k projections (200x200)

$$\Rightarrow 200^3 + 5 \cdot 10^5 = 8.5 \cdot 10^6 D!!!$$

CSIC

instruct Integrating Biology

instruct image Processing Center

Option 5: Computational means

Work with (a few) classes

100k projections (200x200)

$$\Rightarrow 200^3 + 5 \cdot 10^5 = 8.5 \cdot 10^6 \, D!!!$$

CSIC

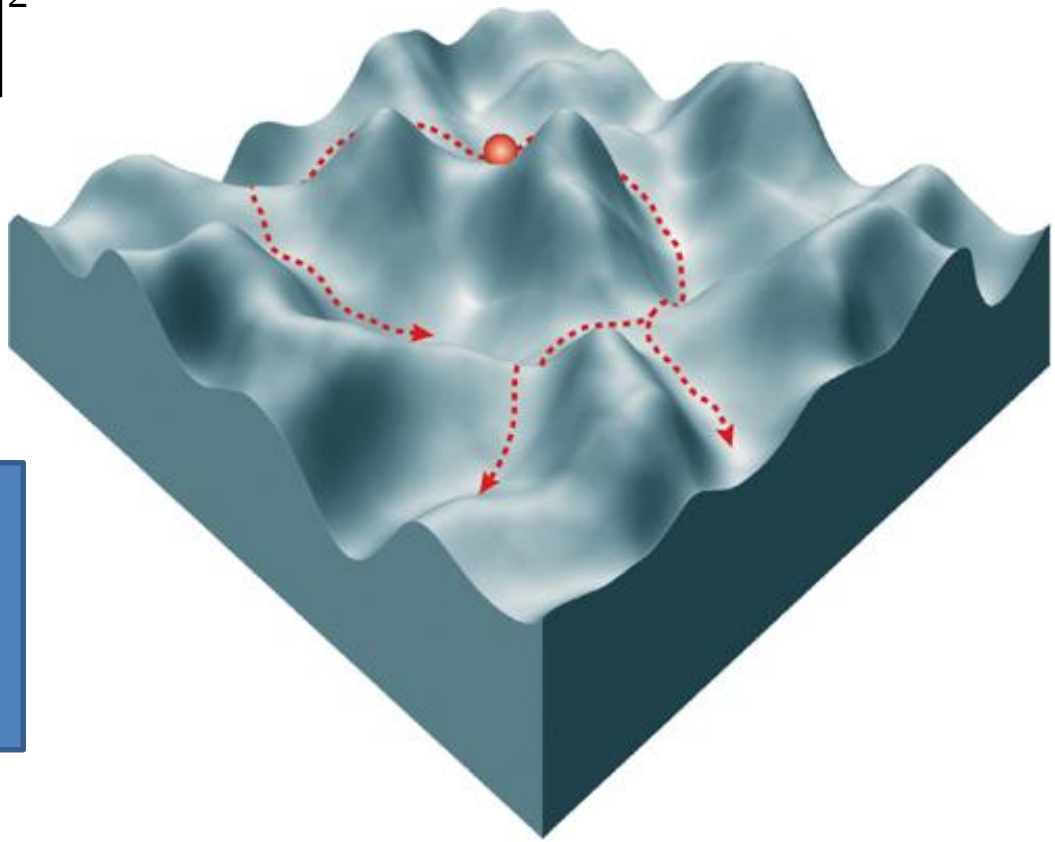
instruct Integrating Biology

instruct image Processing Center

Option 5: Computational means

$$\min_{V,\theta} \sum_i \left\| I_i - P_{\theta_i} V \right\|^2$$

1) Increase SNR
2) Common lines



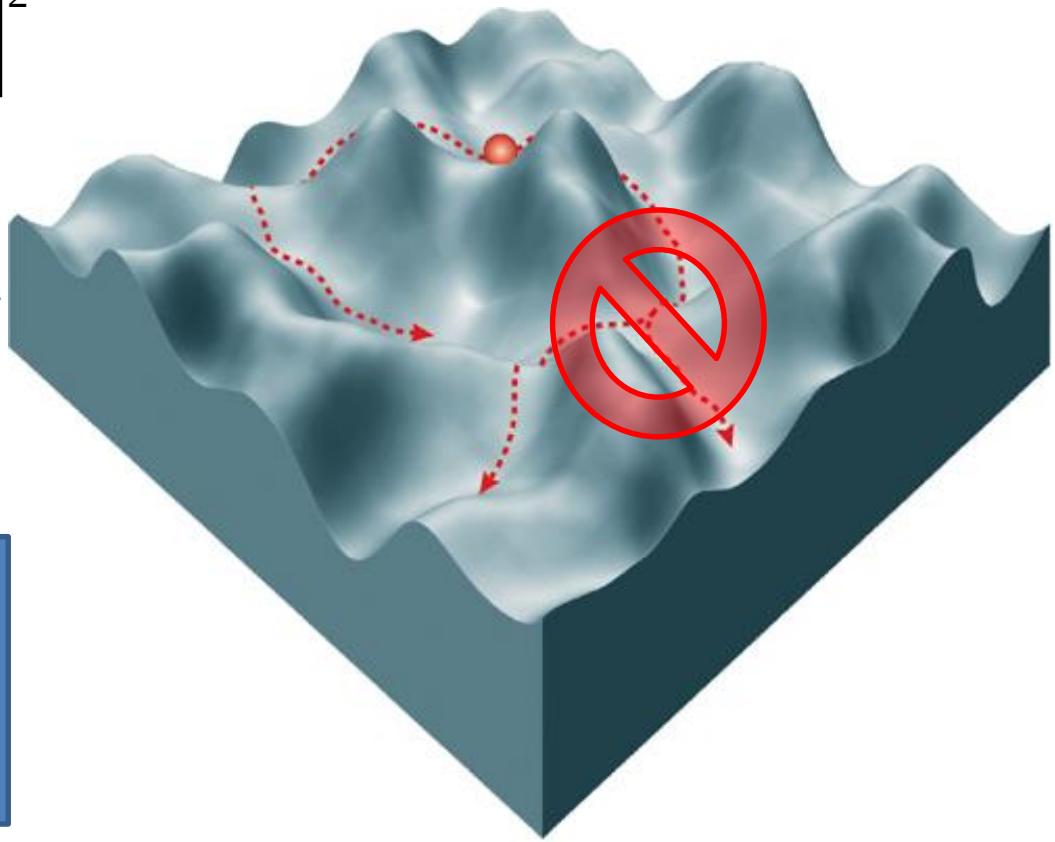
CSIC

instruct Integrating Biology

instruct image Processing Center

Option 5: Computational means

$$\min_{V,\theta} \sum_i \left\| I_i - P_{\theta_i} V \right\|^2$$



Don’t be greedy

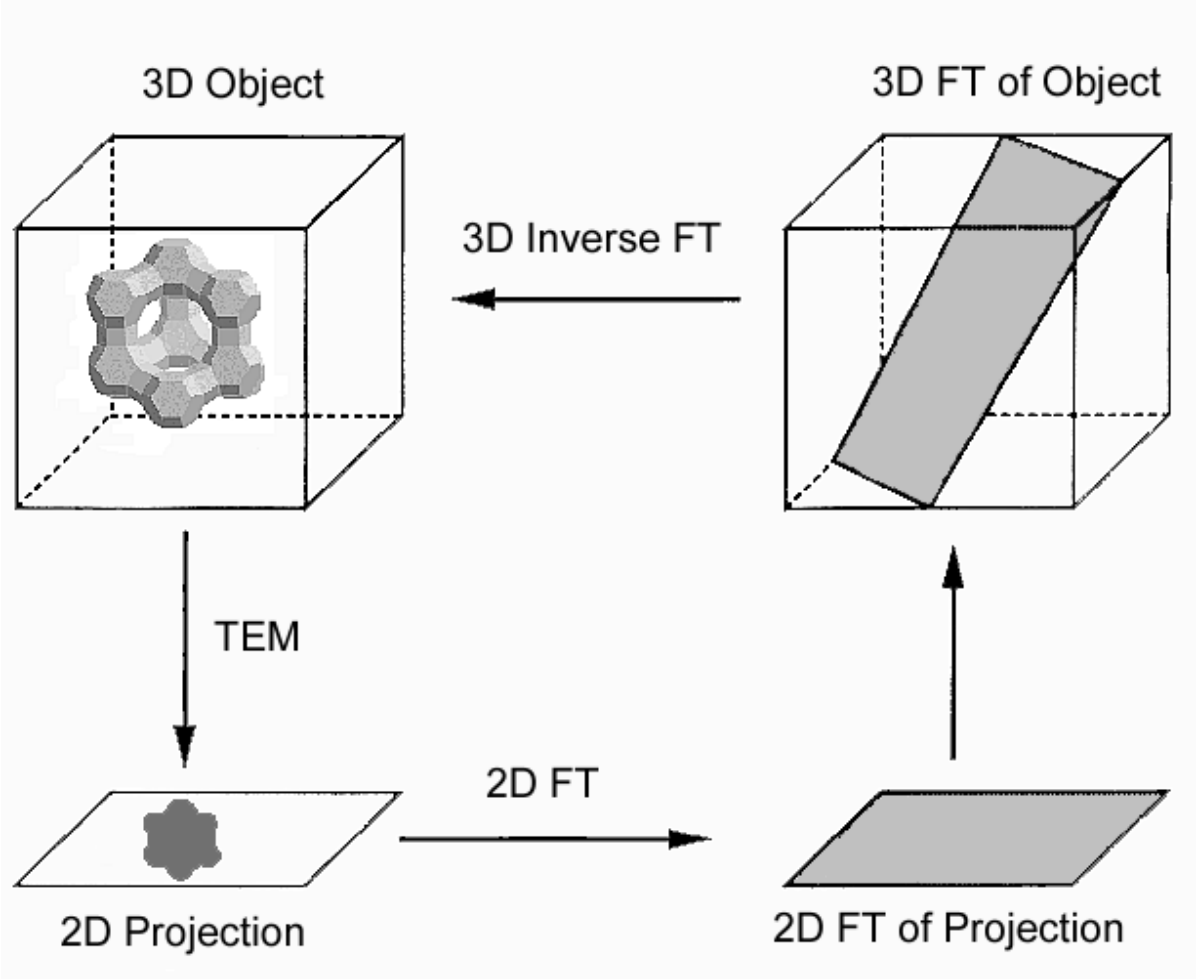
CSIC

instruct Integrating Biology

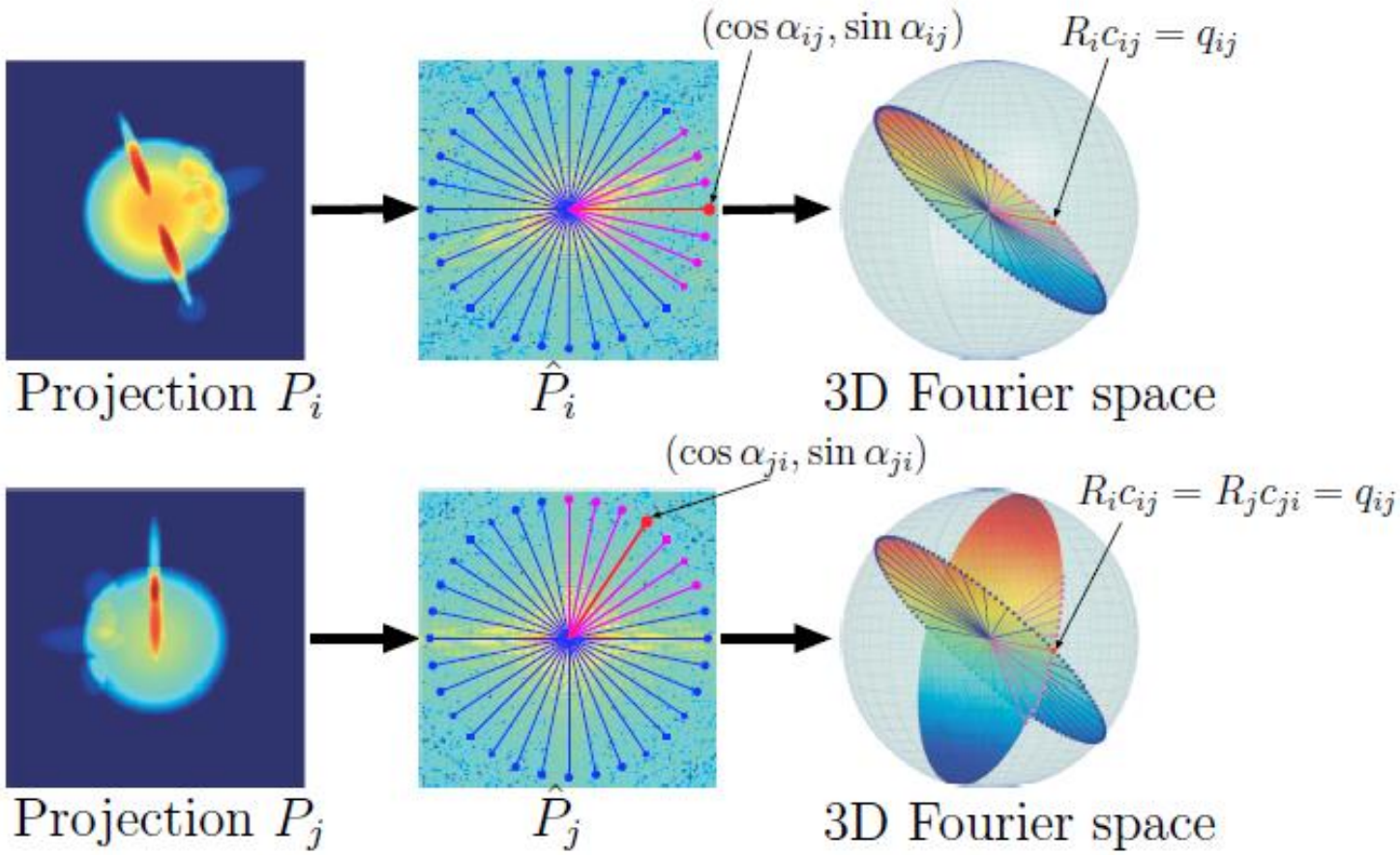
BIO COMPUTING Unit

instruct image Processing Center

Option 5: Central Slice Theorem



Option 5: Common lines

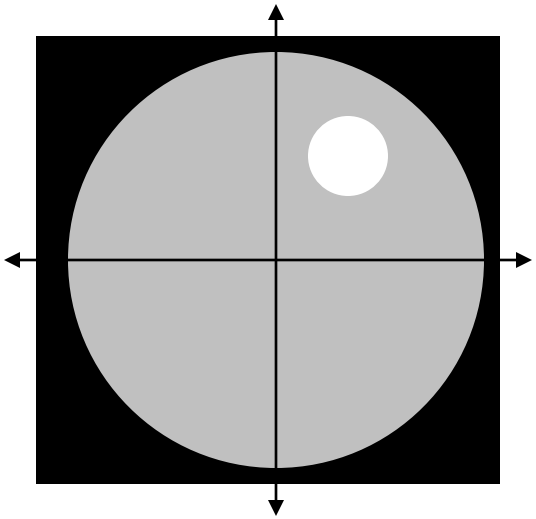


CSIC

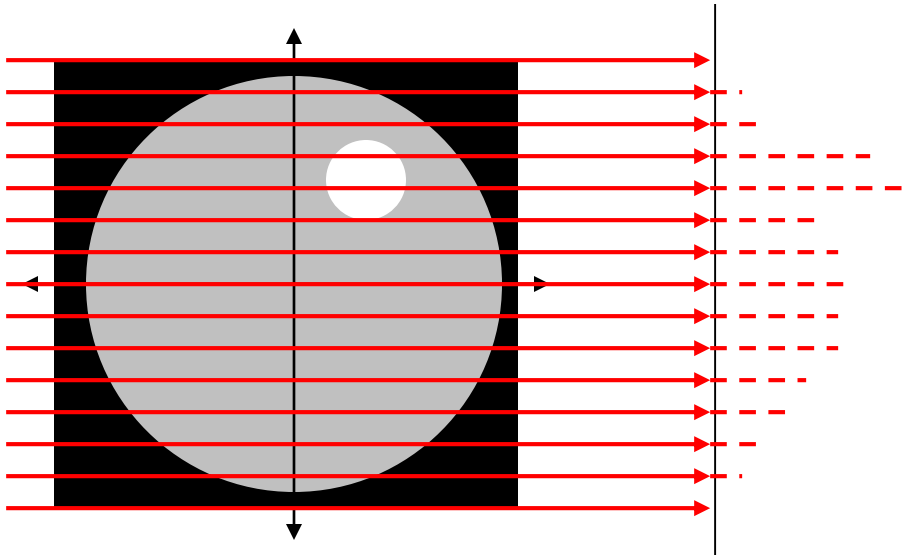
instruct Integrating Biology

instruct image Processing Center

Option 5: Sinogram



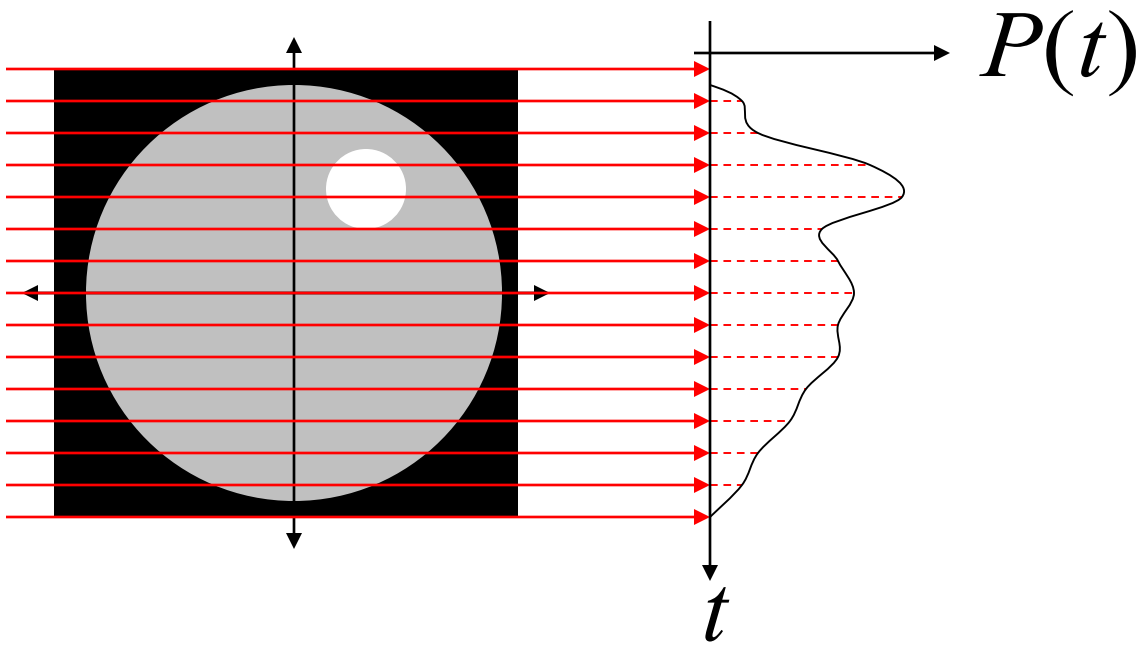
Option 5: Sinogram



$$P(t) = \int_{-\infty}^{\infty} \rho(x,t)dx$$

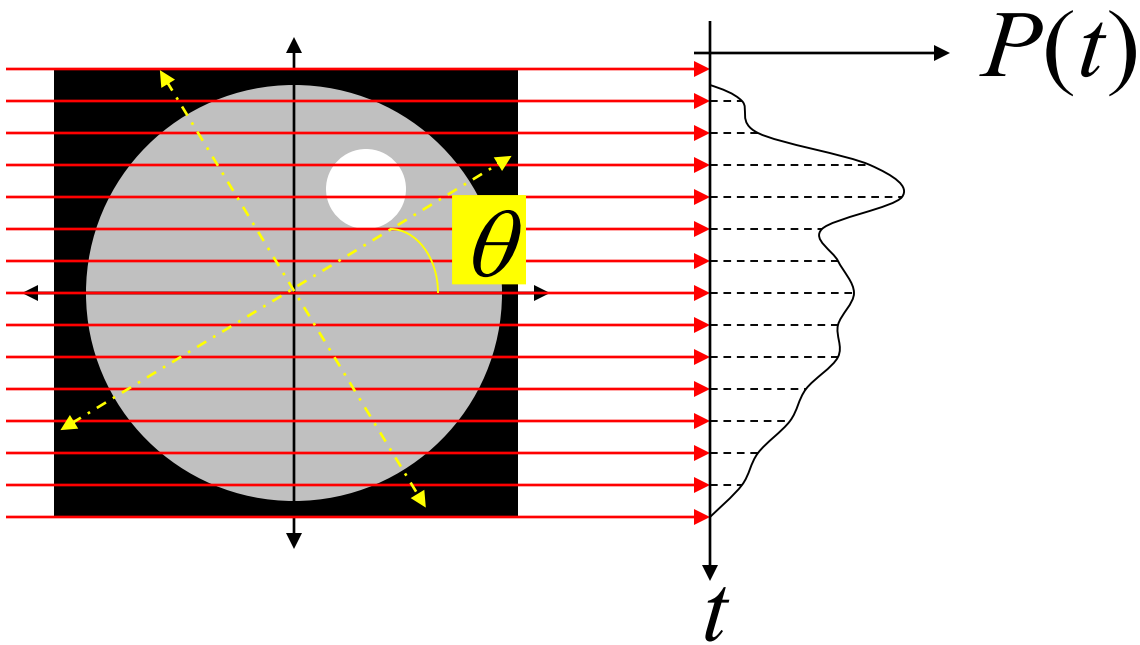
Our First Projection

Option 5: Sinogram



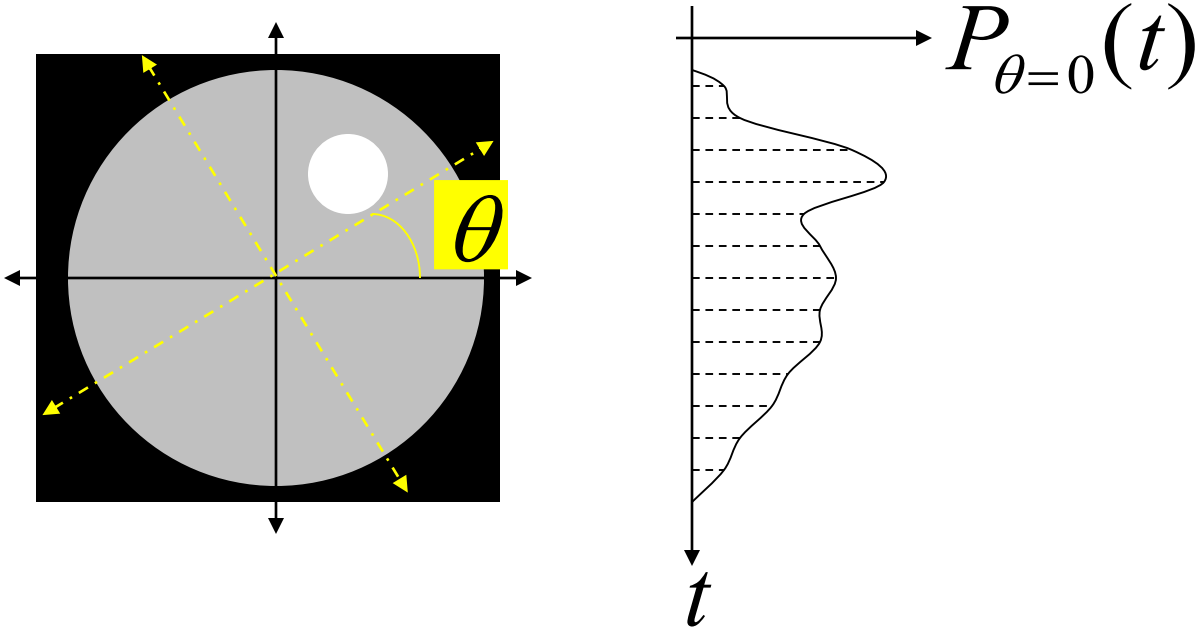
Our First Projection

Option 5: Sinogram



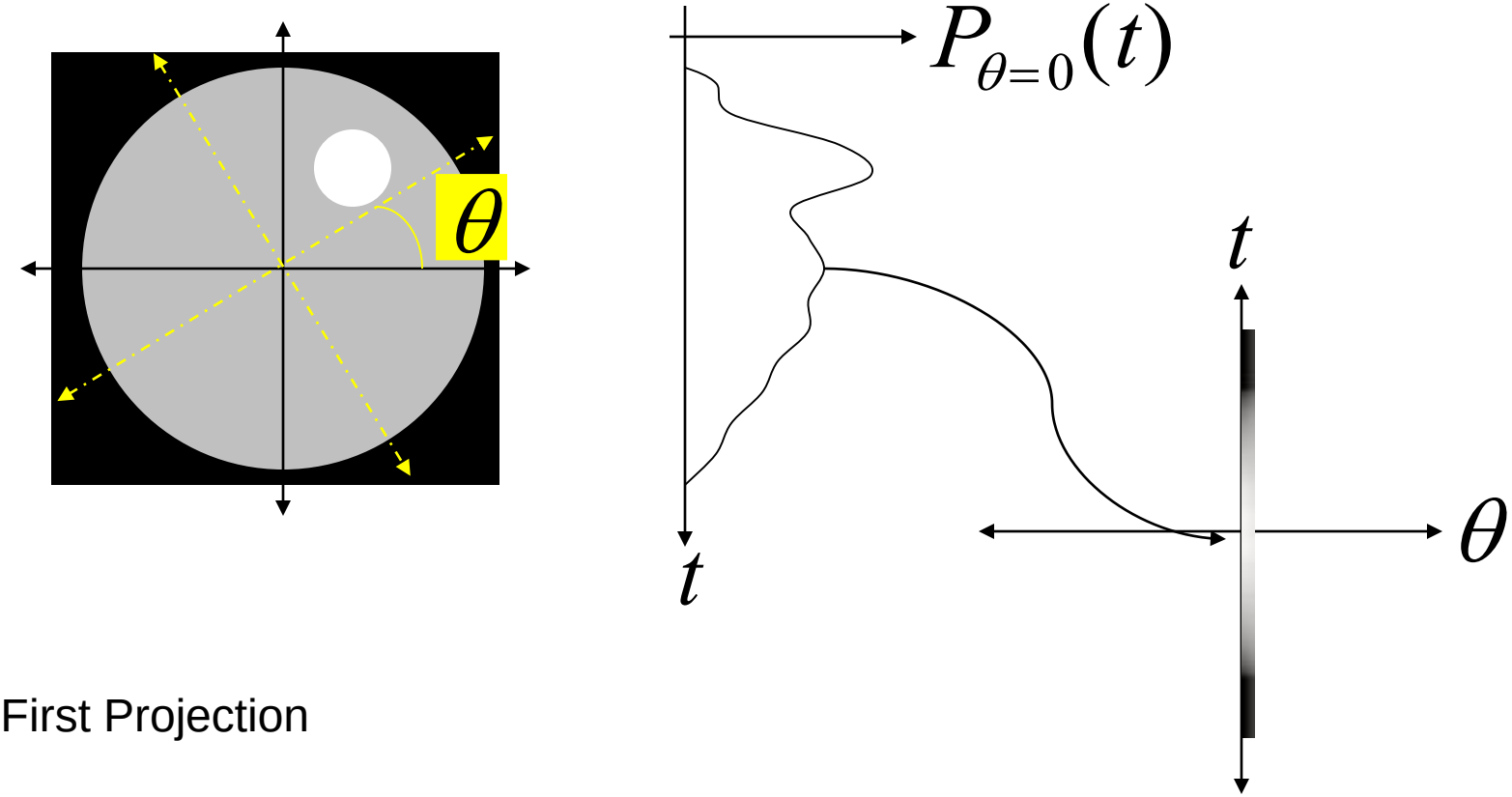
Our First Projection

Option 5: Sinogram

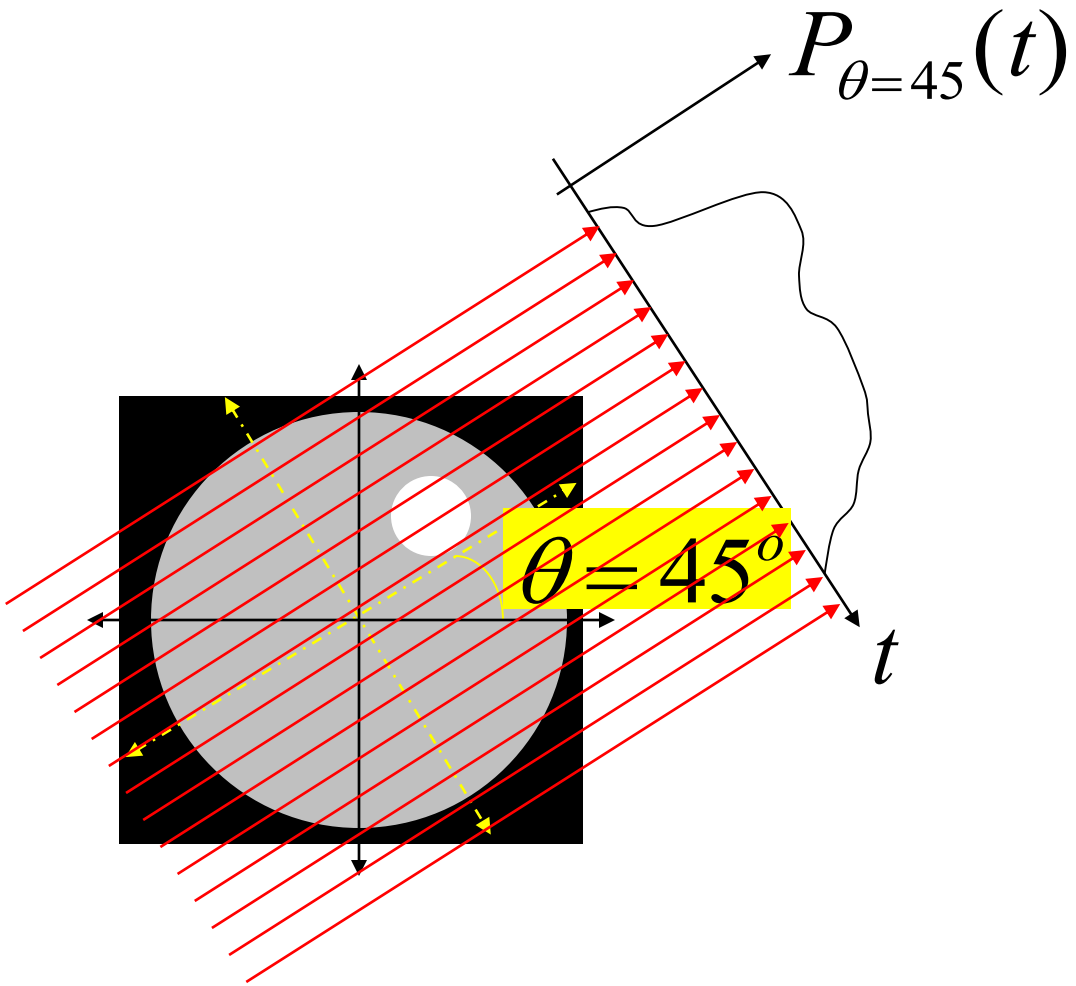


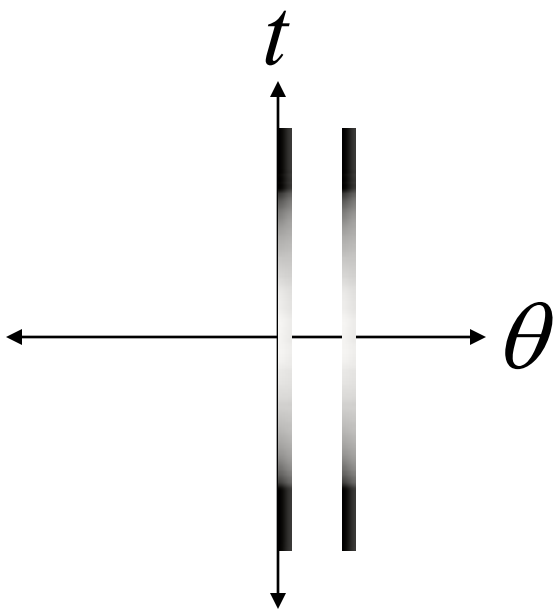
Our First Projection

Option 5: Sinogram

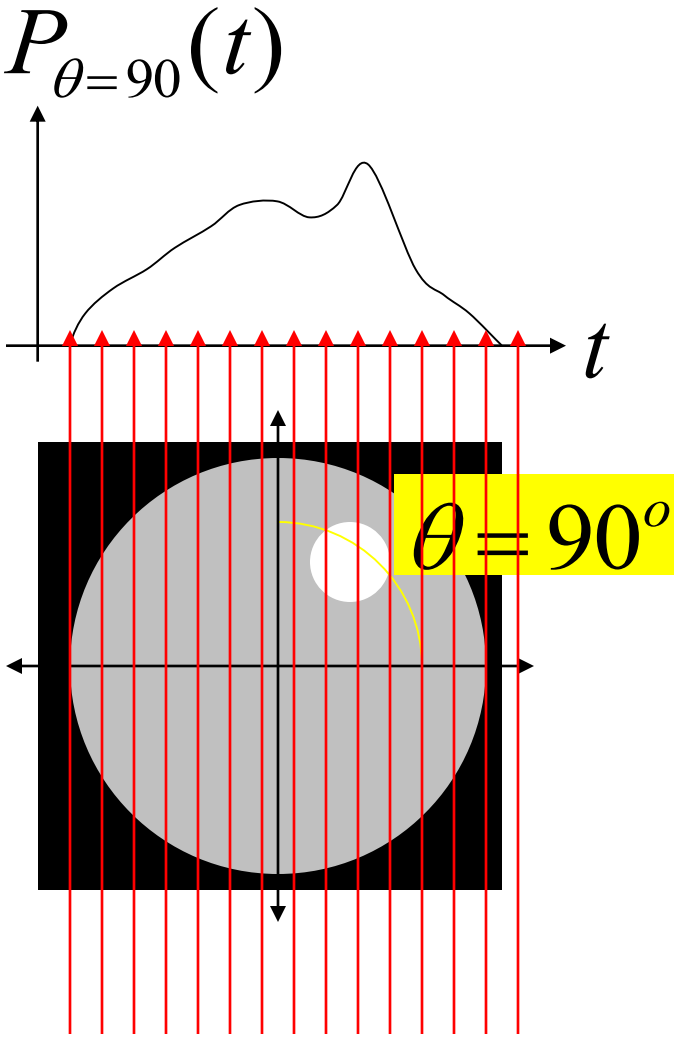


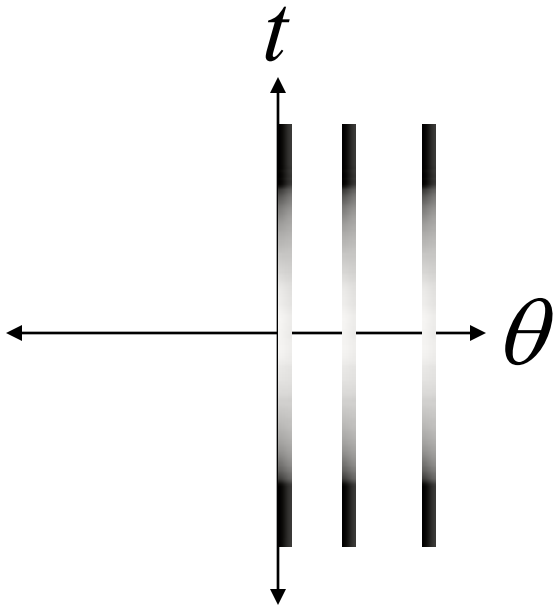
Our First Projection





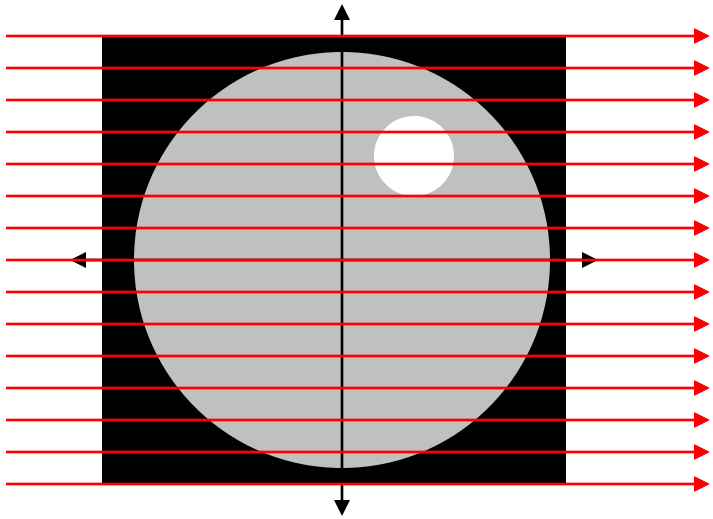
Our Second Projection

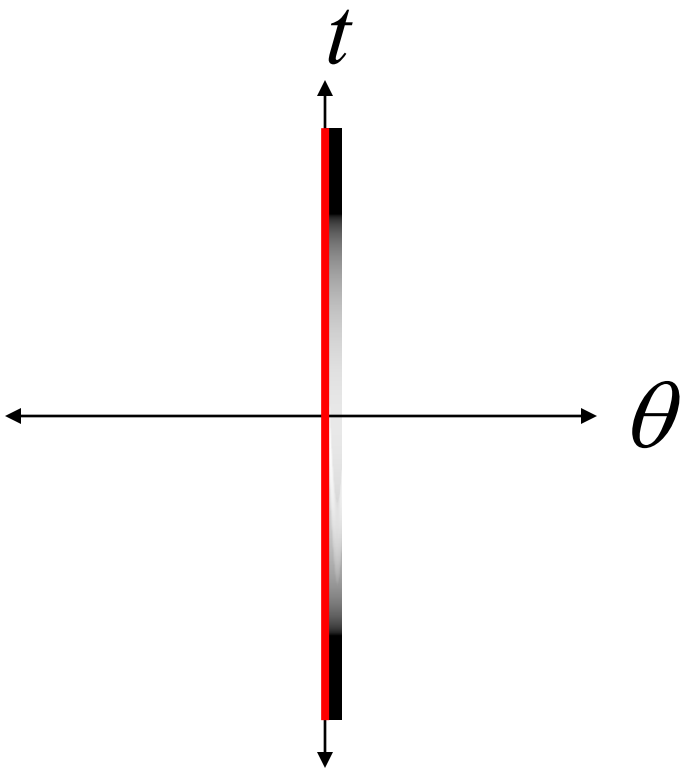




Our Third Projection

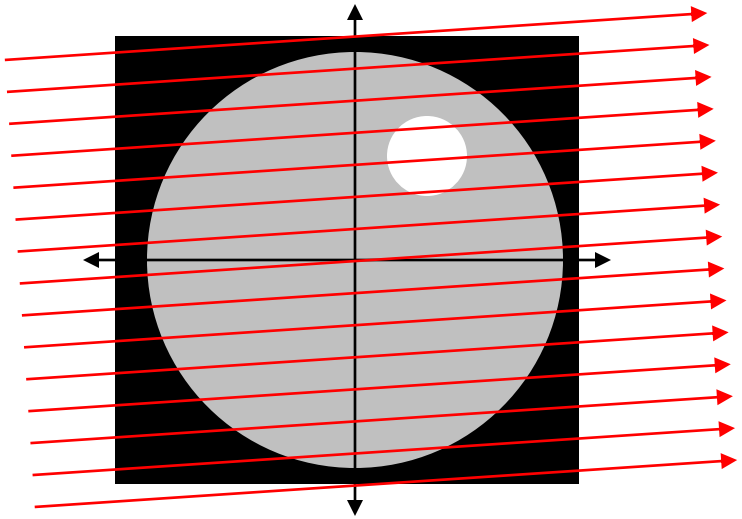
Option 5: Sinogram

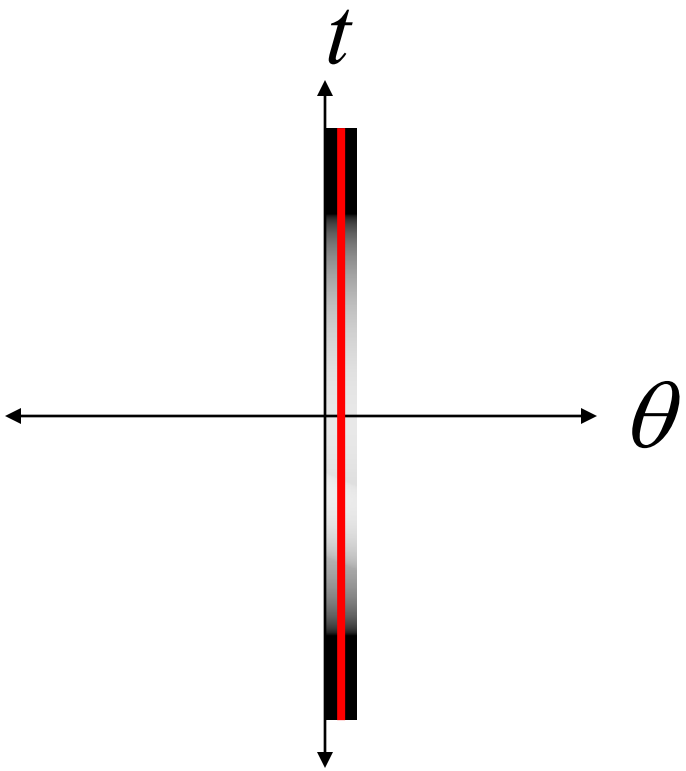




Sinogram

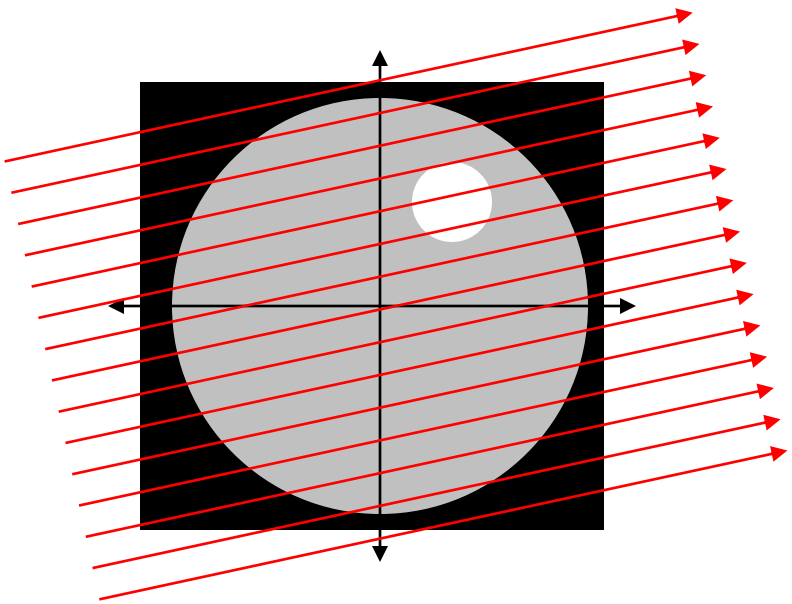
Option 5: Sinogram

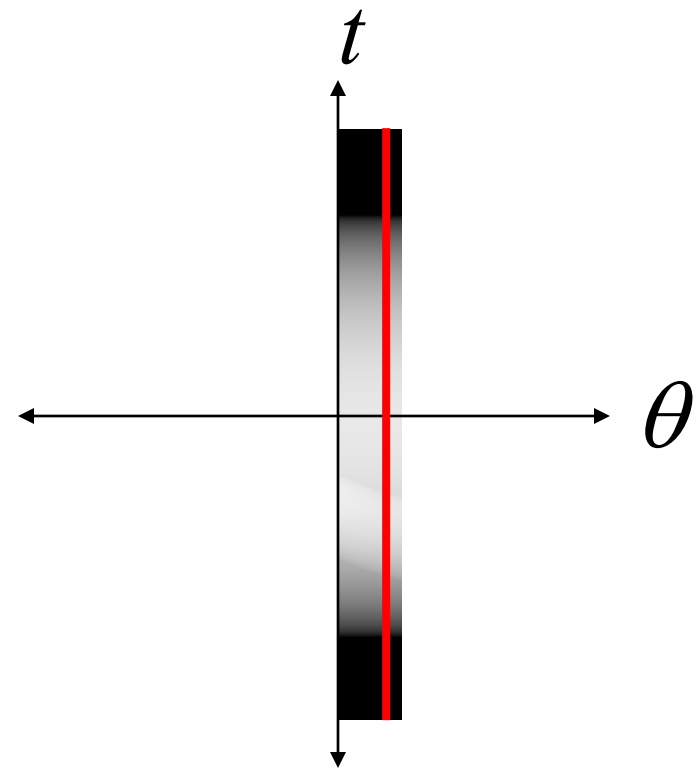




Sinogram

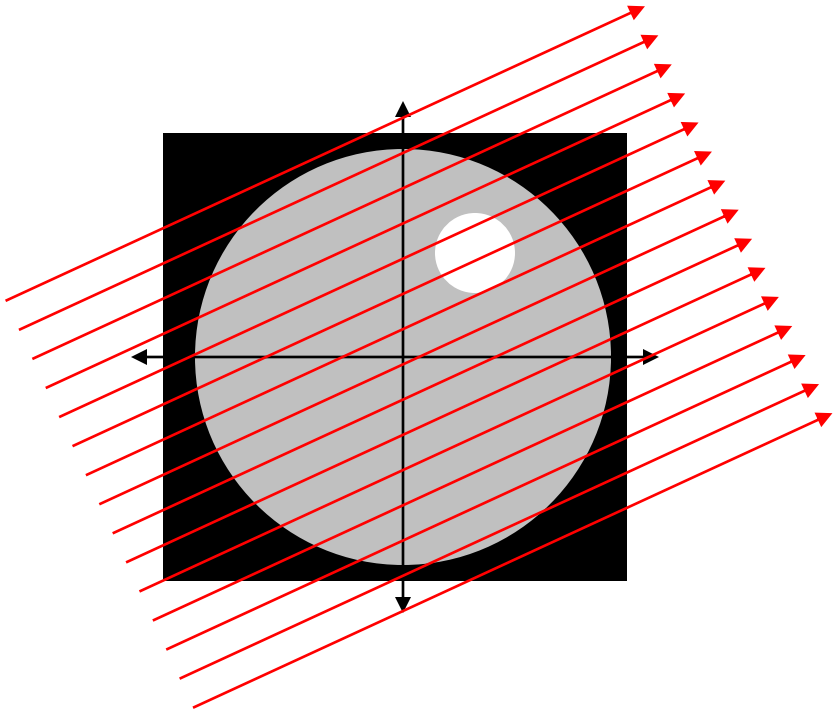
Option 5: Sinogram

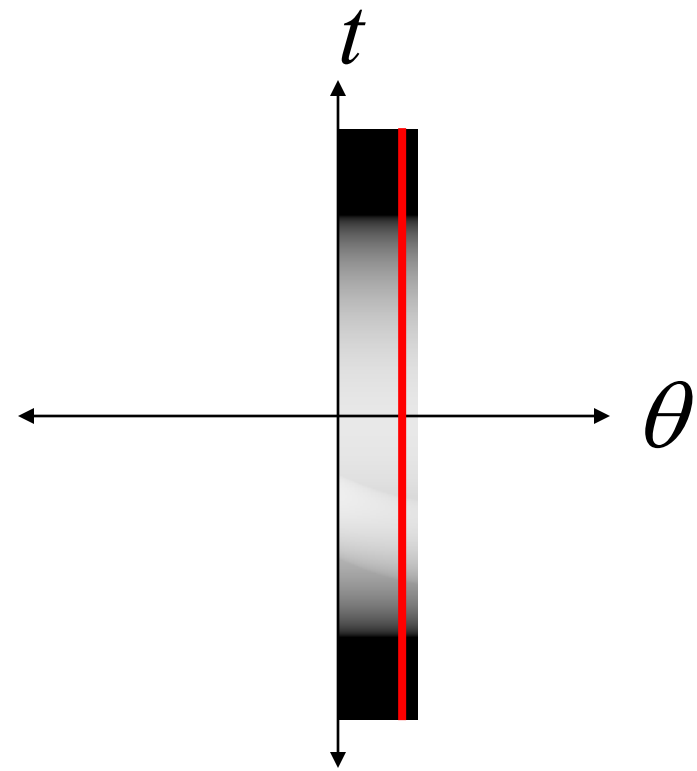




Sinogram

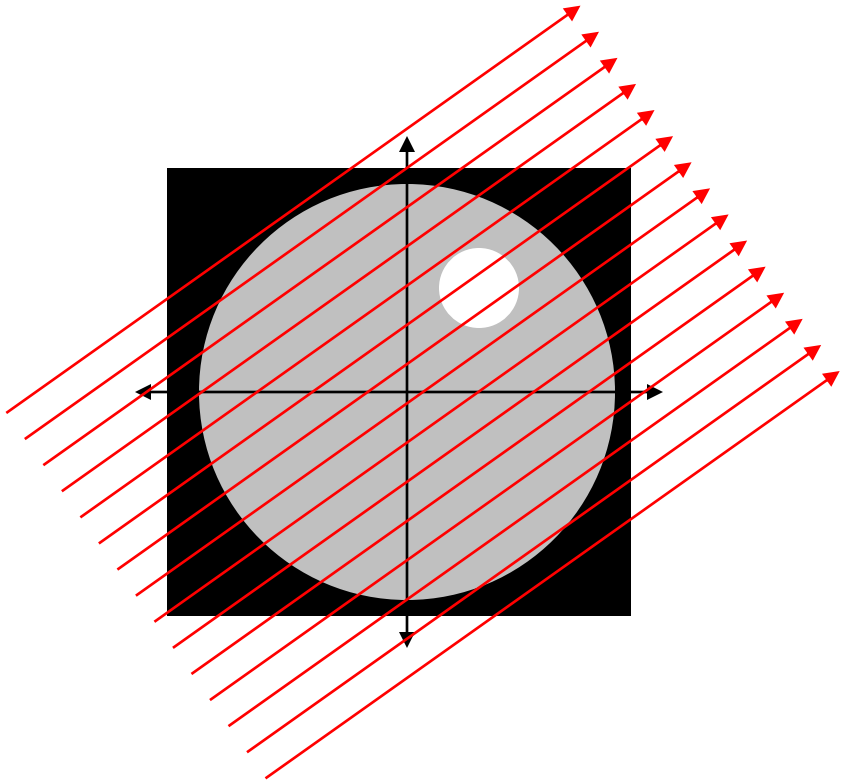
Option 5: Sinogram

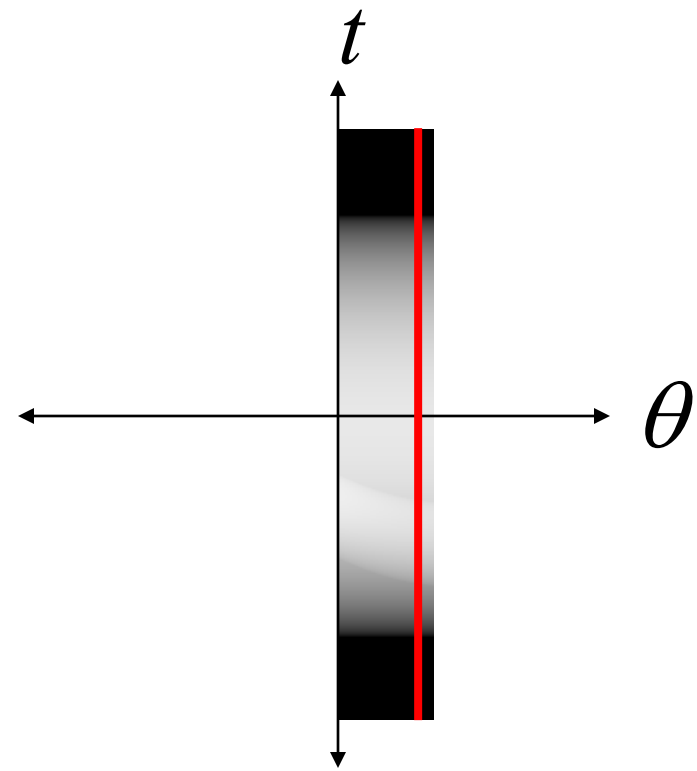




Sinogram

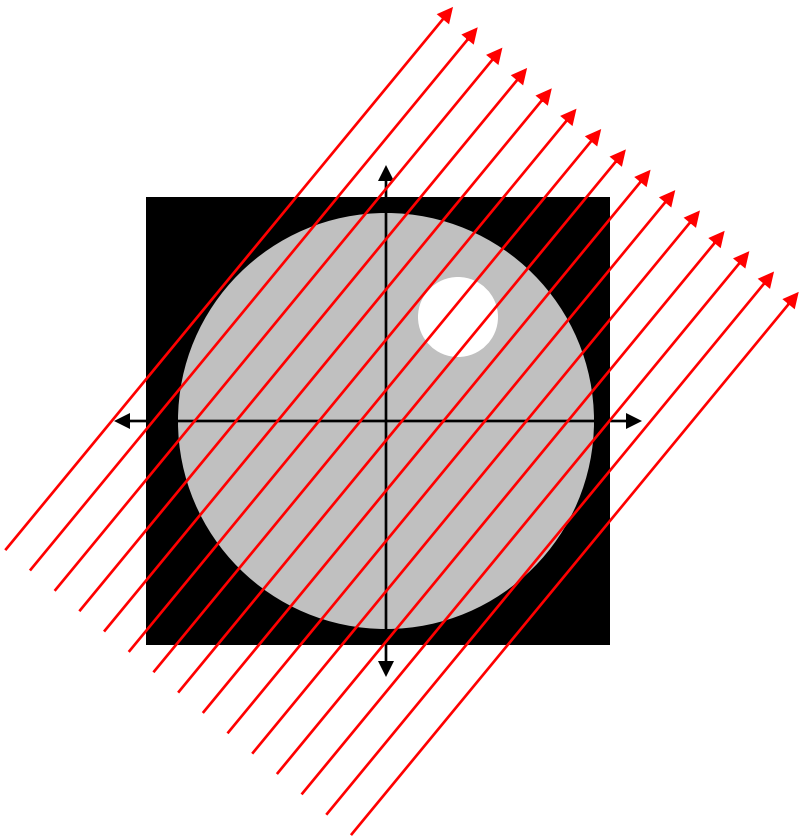
Option 5: Sinogram

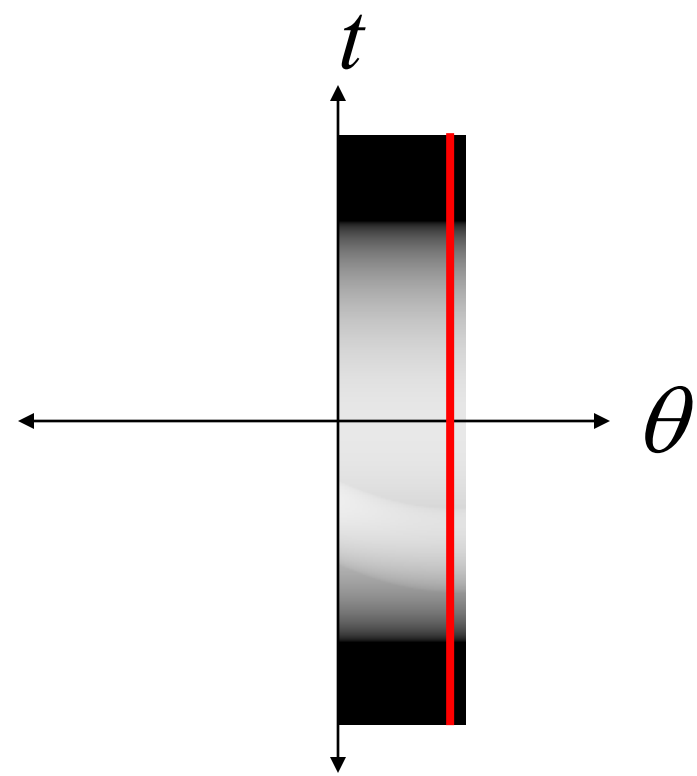




Sinogram

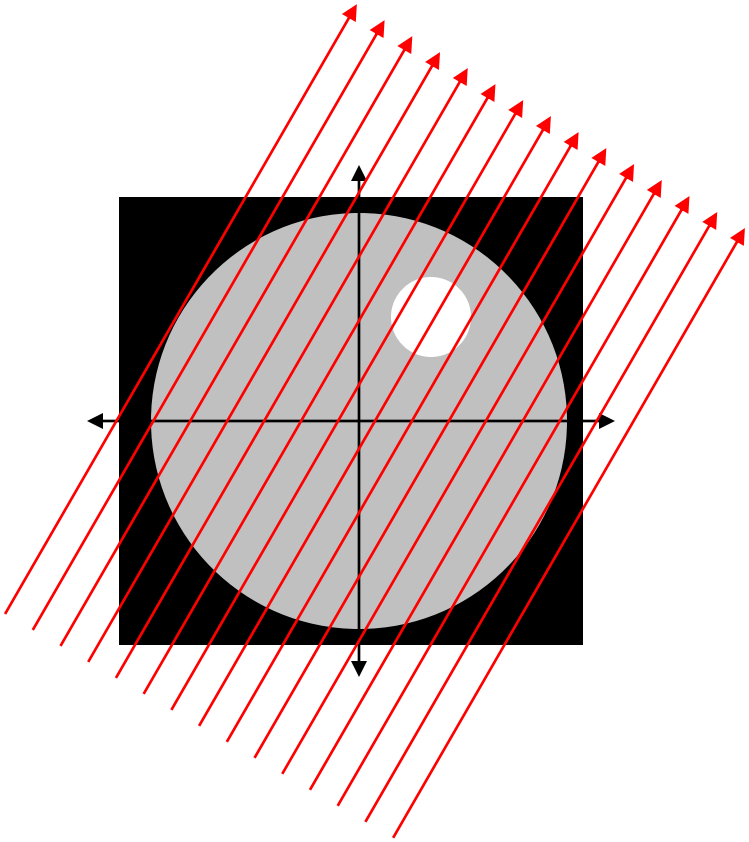
Option 5: Sinogram

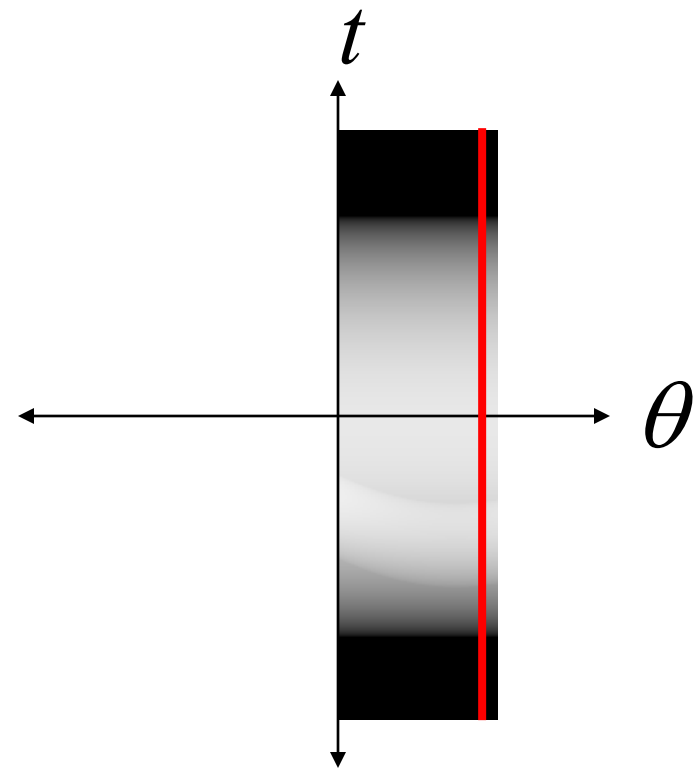




Sinogram

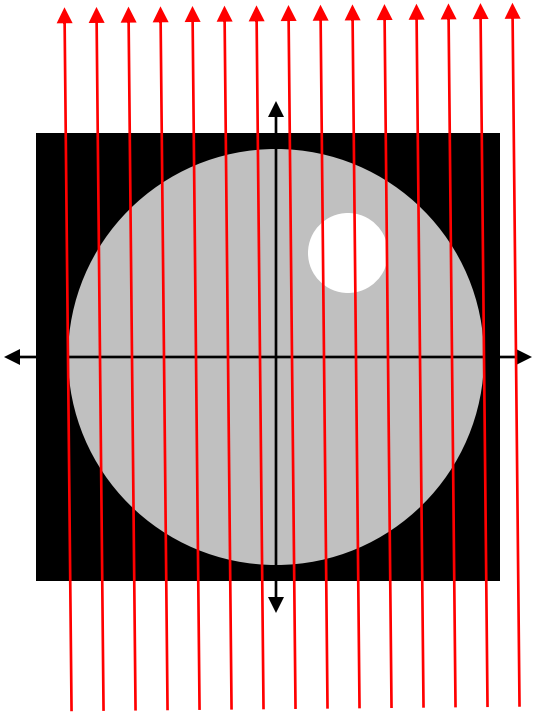
Option 5: Sinogram

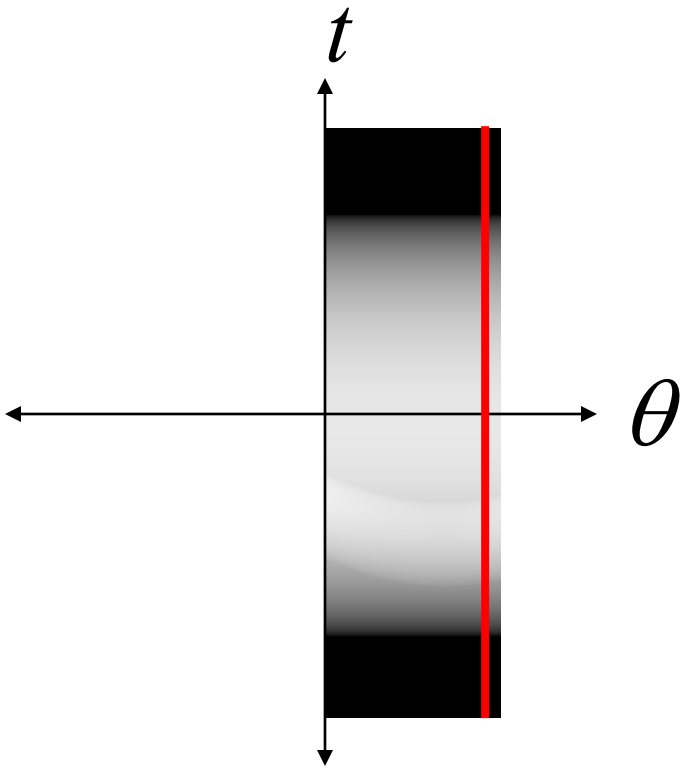




Sinogram

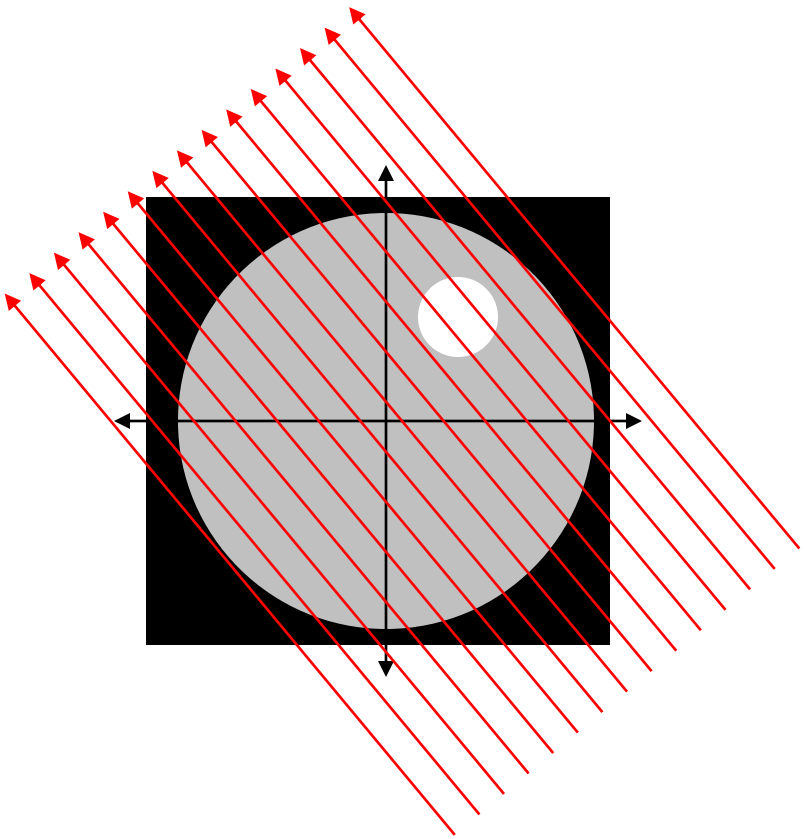
Option 5: Sinogram

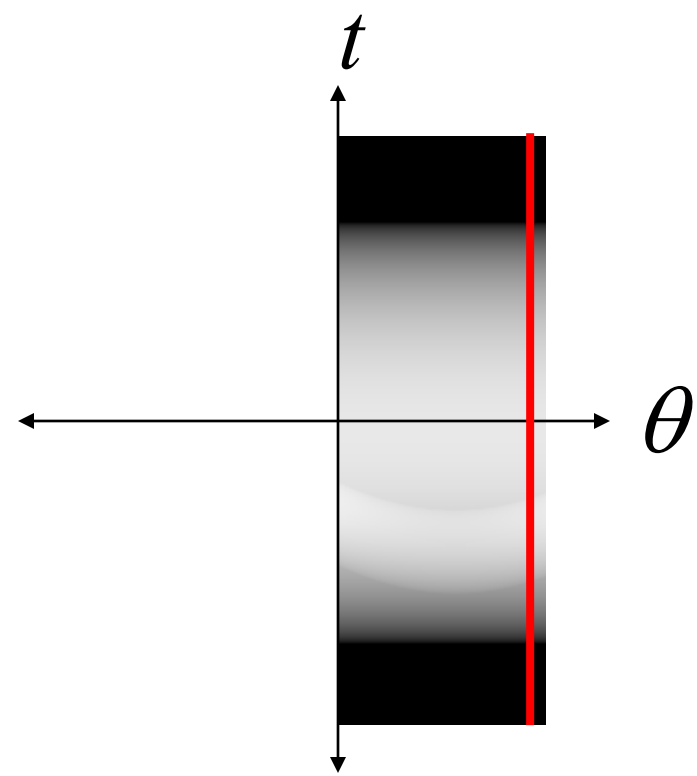




Sinogram

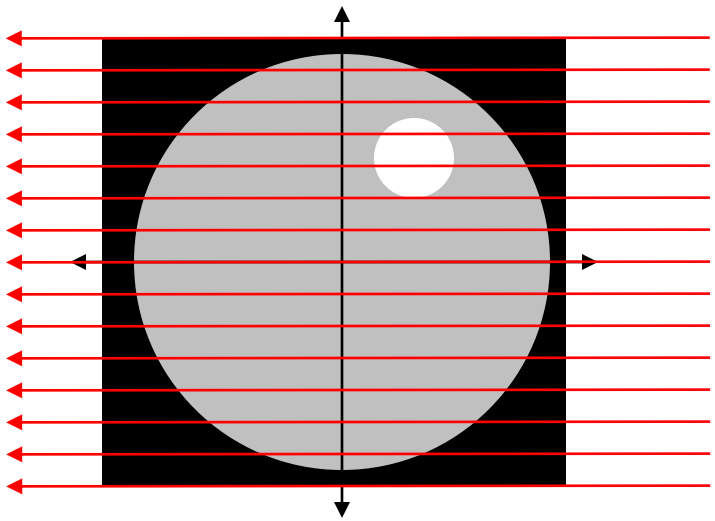
Option 5: Sinogram

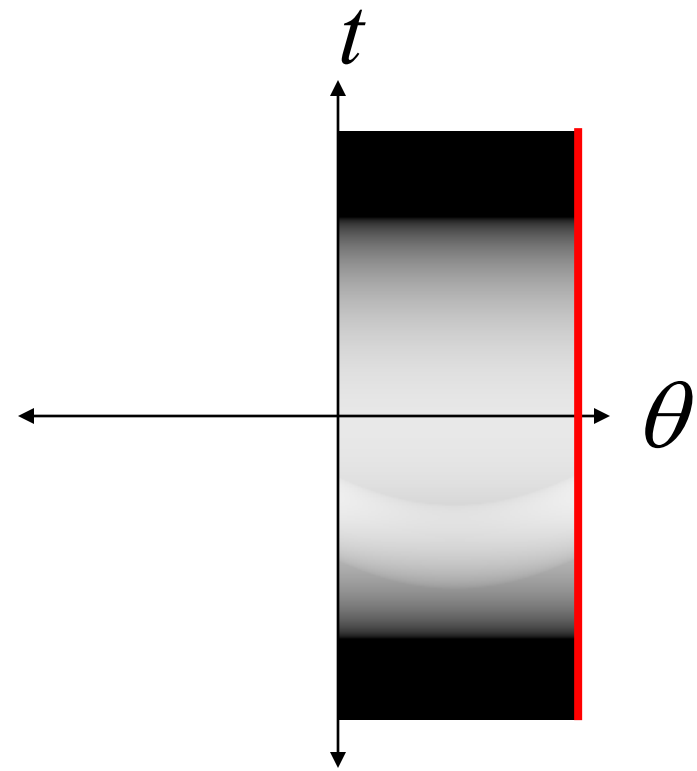




Sinogram

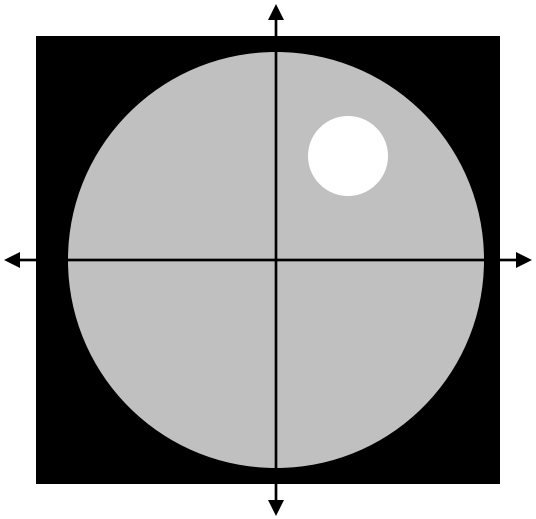
Option 5: Sinogram

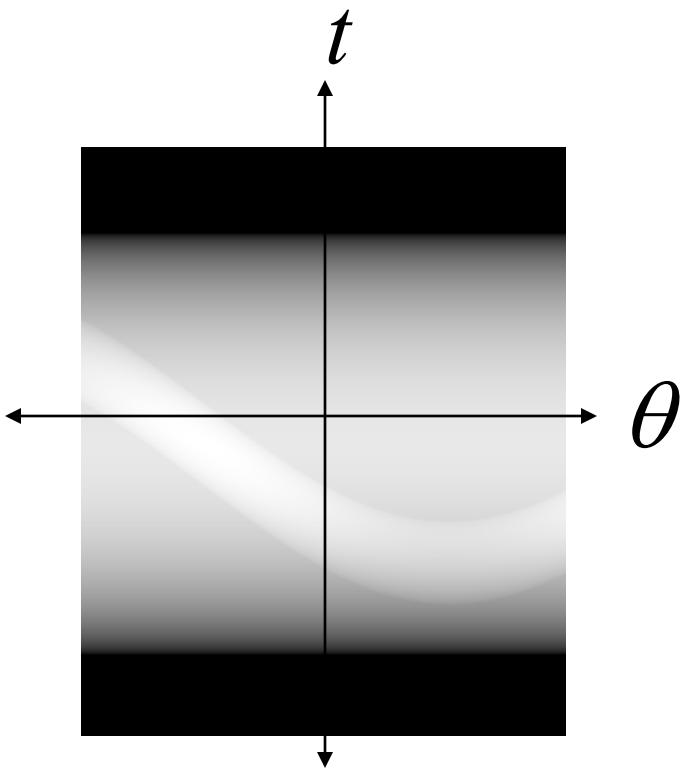




Sinogram

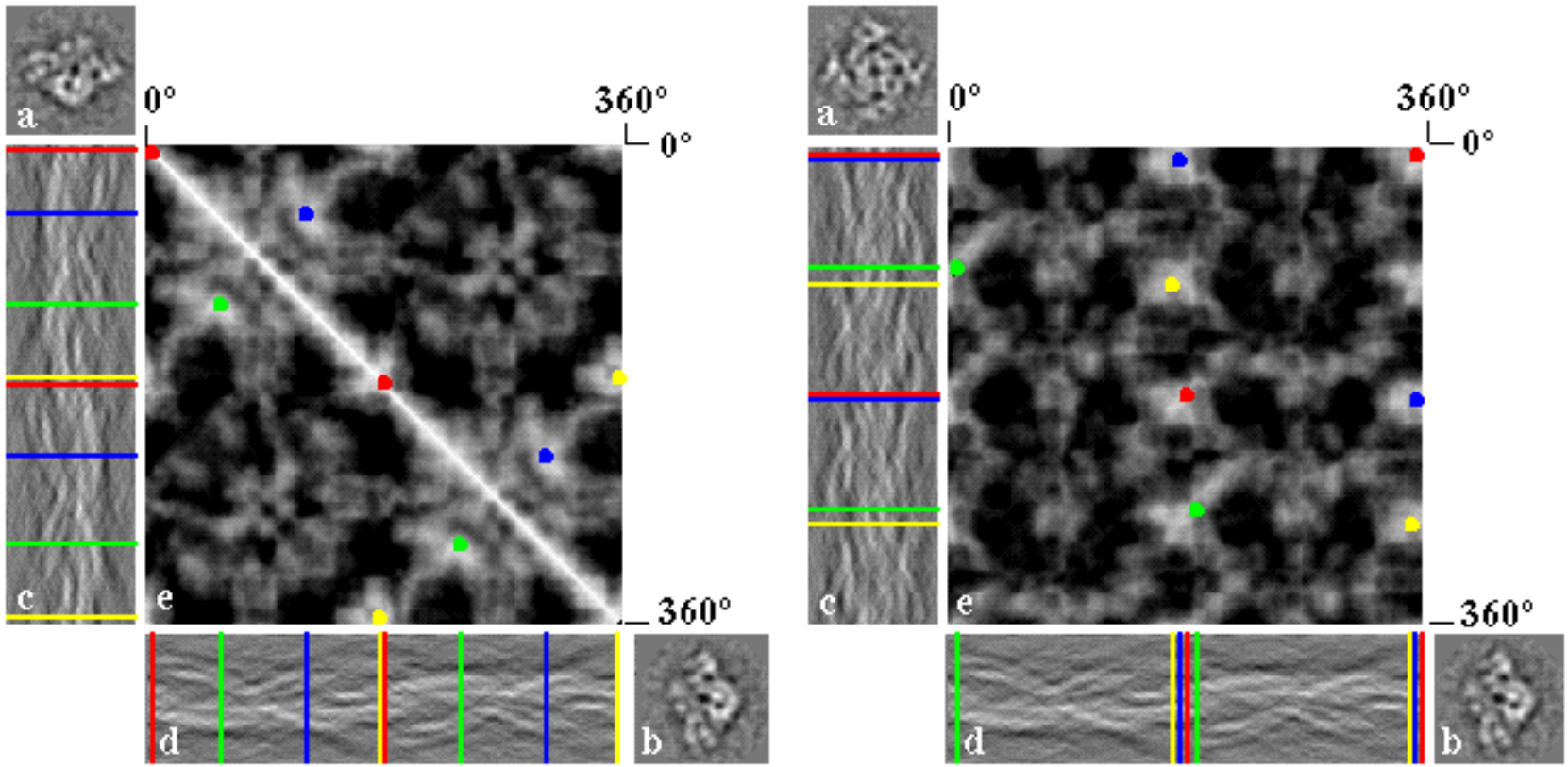
Option 5: Sinogram





Sinogram

Option 5: Common lines and sinograms



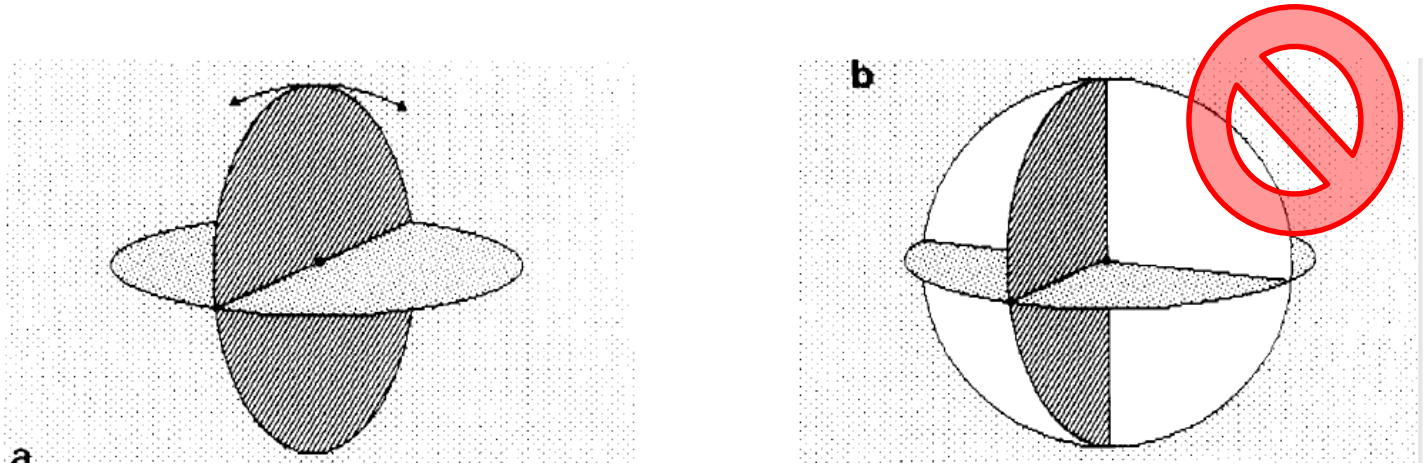
Serysheva, I. I.; Orlova, E. V.; Chiu, W.; Sherman, M. B.; Hamilton, S. L. & van Heel, M. Electron cryomicroscopy and angular reconstitution used to visualize the skeletal muscle calcium release channel. Nat Struct Biol, 1995, 2, 18-24

CSIC

instruct Integrating Biology

instruct image Processing Center

Option 5: Common lines



Shifts=0
Common line 1-2 along Z
Common line 2-3 in XY plane

$$\Rightarrow 3D \Rightarrow \begin{pmatrix} 3 \\ 2 \end{pmatrix} = 3 \text{ constraints}$$

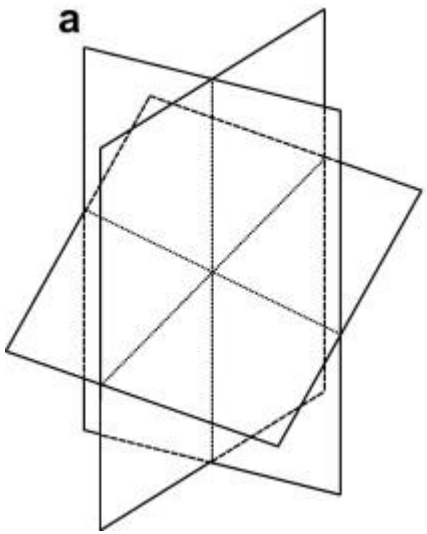
van Heel, M. Angular reconstitution: A posteriori rm assignment of projection directions for 3D reconstruction. Ultramicroscopy, 1987, 21, 111-124

CSIC
instruct Integrating Biology
instruct image Processing Center

Option 5: Common lines

$$\min_{\theta} \sum_{ij=1}^{N} \left\| CL_{\theta_i}(I_i) - CL_{\theta_j}(I_j) \right\|^2$$

Shifts=0 $\Rightarrow 3N\,D \Rightarrow \binom{3N}{2} = 9N^2 - 3N$ comparisons



Greedy optimization:
Iterate till convergence
- Leave N-1 images fixed, optimize 1 image

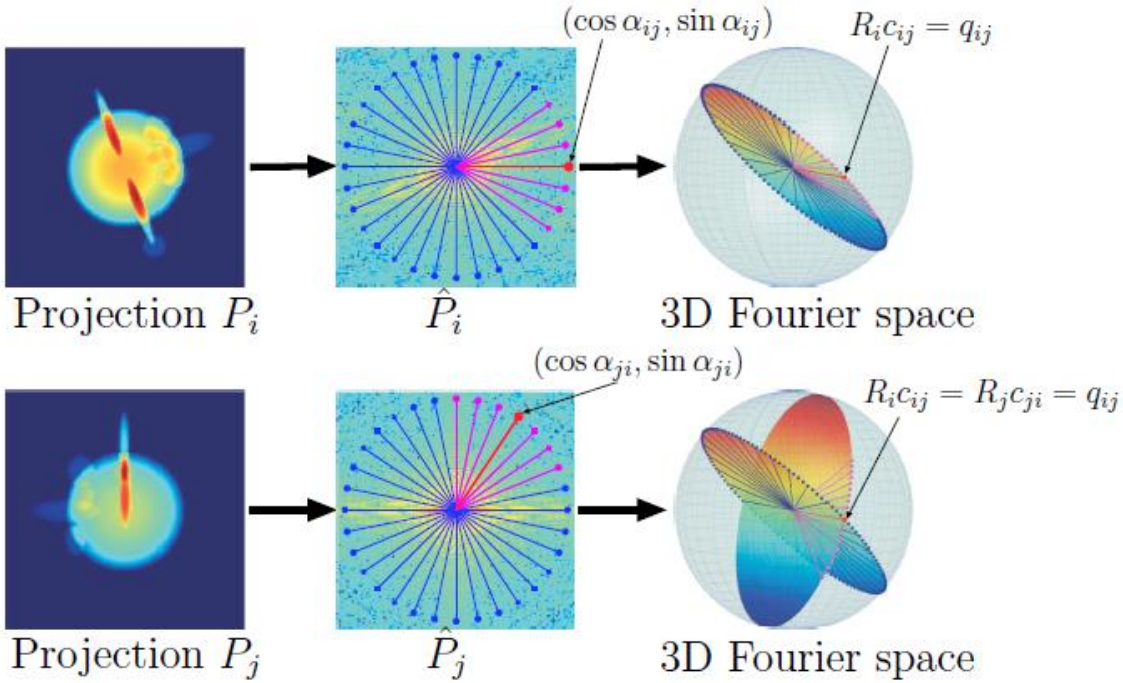
Penczek, P. A.; Zhu, J. & Frank, J. A common-lines based method for determining orientations for N>3 particle projections simultaneously Ultramicroscopy, 1996, 63, 205-218

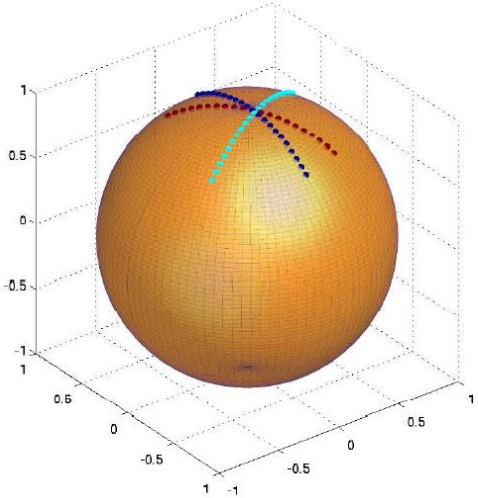
CSIC

instruct Integrating Biology

instruct image Processing Center

Option 5: Common lines





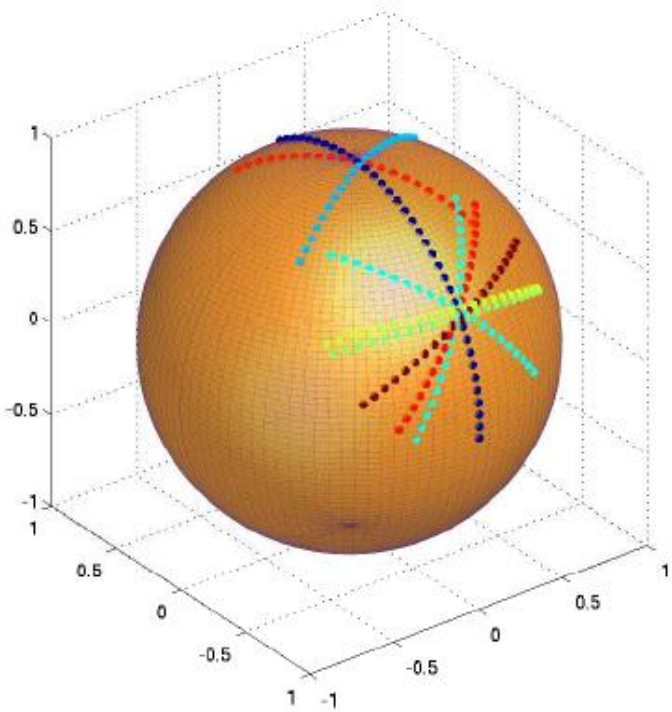
1. Coifman, R. R.; Shkolnisky, Y.; Sigworth, F. J. & Singer, A. Reference Free Structure Determination through Eigenvectors of Center of Mass Operators. Appl Comput Harmon Anal, 2010, 28, 296-312
2. Singer, A. & Shkolnisky, Y. Three-Dimensional Structure Determination from Common Lines in Cryo-EM by Eigenvectors and Semidefinite Programming.*SIAM J Imaging Sci,* **2011***, 4*, 543-572
3. Singer, A.; Zhao, Z.; Shkolnisky, Y. & Hadani, R. Viewing Angle Classification of Cryo-Electron Microscopy Images Using Eigenvectors. *SIAM J Imaging Sci,* **2011***, 4*, 723-759
4. Greenberg, I. & Shkolnisky, Y. Common lines modeling for reference free Ab-initio reconstruction in cryo-EM. Journal of structural biology, 2017, 200, 106-117

CSIC

instruct Integrating Biology

instruct image Processing Center

Option 5: Common lines



Eigenvectors of an adjacency matrix

$$\theta \in \square^{3N}$$

$$Shifts \in \square^{2N}$$

In practice?

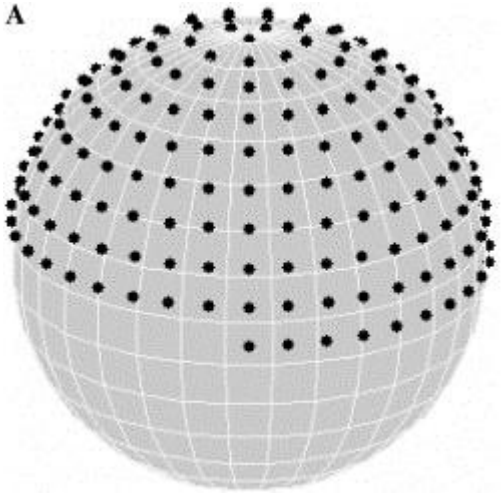
1. Coifman, R. R.; Shkolnisky, Y.; Sigworth, F. J. & Singer, A. Reference Free Structure Determination through Eigenvectors of Center of Mass Operators. Appl Comput Harmon Anal, 2010, 28, 296-312
2. Singer, A. & Shkolnisky, Y. Three-Dimensional Structure Determination from Common Lines in Cryo-EM by Eigenvectors and Semidefinite Programming.*SIAM J Imaging Sci,* **2011***, 4*, 543-572
3. Singer, A.; Zhao, Z.; Shkolnisky, Y. & Hadani, R. Viewing Angle Classification of Cryo-Electron Microscopy Images Using Eigenvectors. *SIAM J Imaging Sci,* **2011***, 4*, 723-759

CSIC

instruct Integrating Biology

instruct image Processing Center

Option 5: Manifold analysis



Local transition probability matrix

Eigenvalues

Euler angles

Number of images

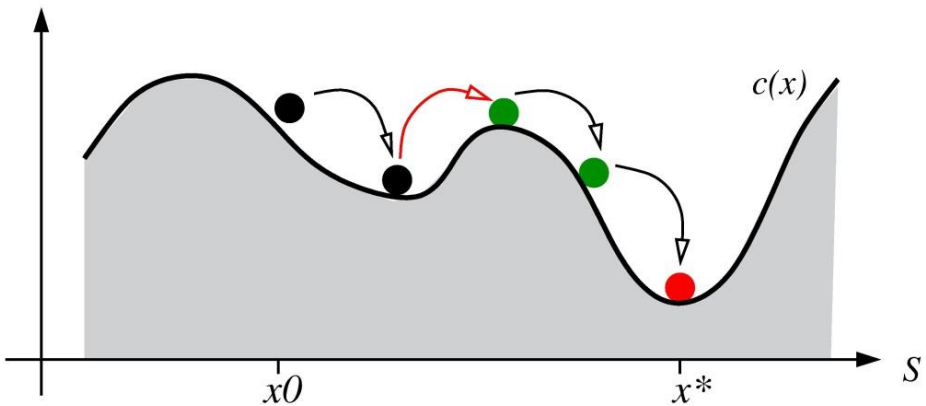
Giannakis, D.; Schwander, P. & Ourmazd, A. The symmetries of image formation by scattering. I. Theoretical framework. Opt Express, 2012, 20, 12799-12826

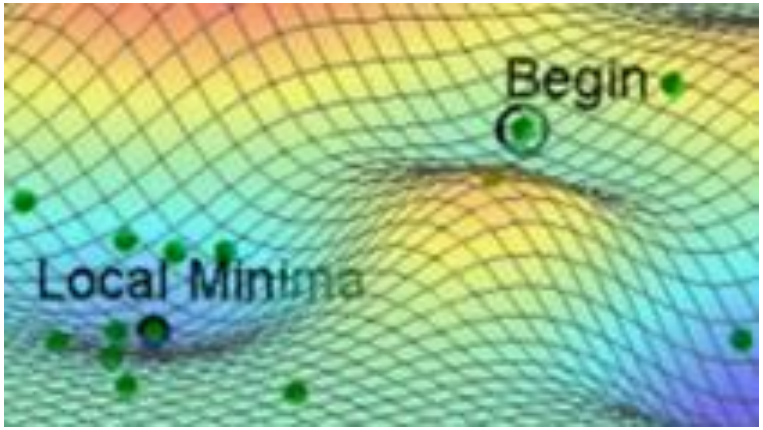
CSIC

instruct Integrating Biology

instruct image Processing Center

Option 5: Stochastic optimization





- Simulated annealing
- Stochastic hill climbing
- Simultaneous Perturbation Stochastic Approximation (SPSA)
- Swarm algorithms
- Genetic algorithms
- Differential Evolution

Not to be greedy

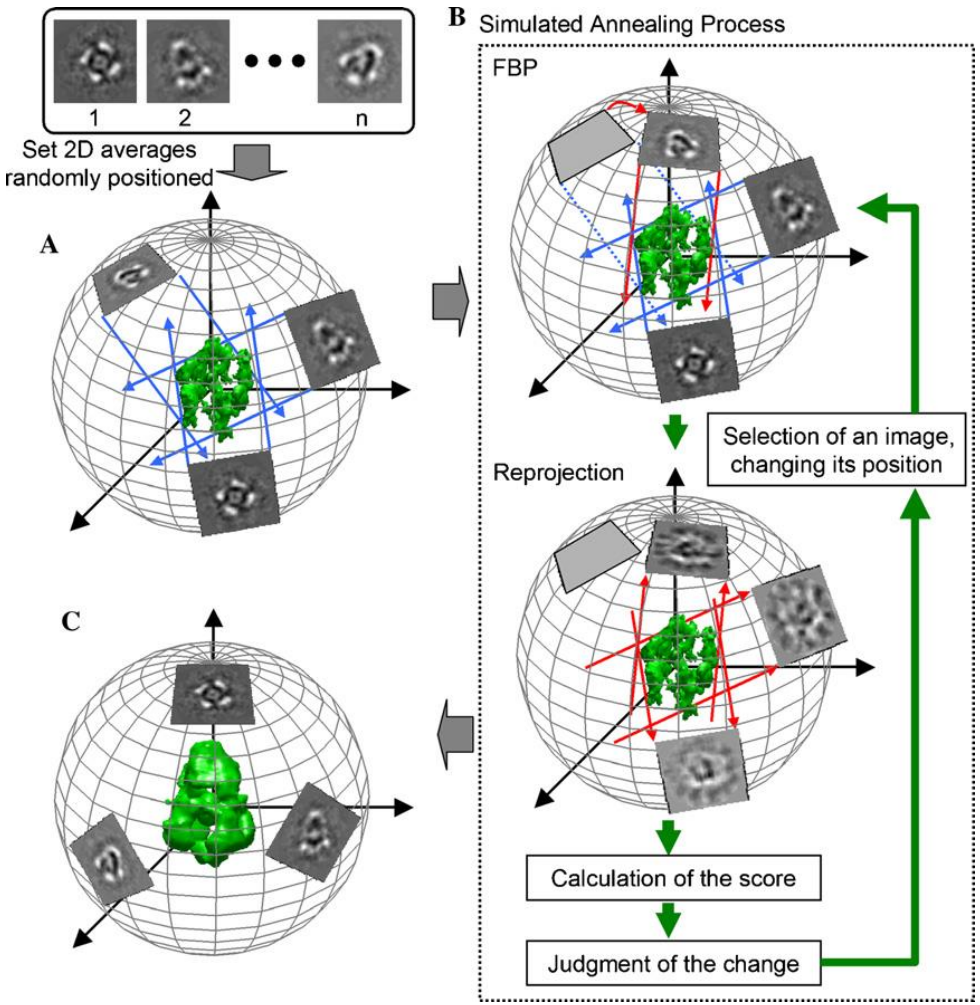
CSIC

instruct Integrating Biology

instruct image Processing Center

Option 5: Stochastic optimization

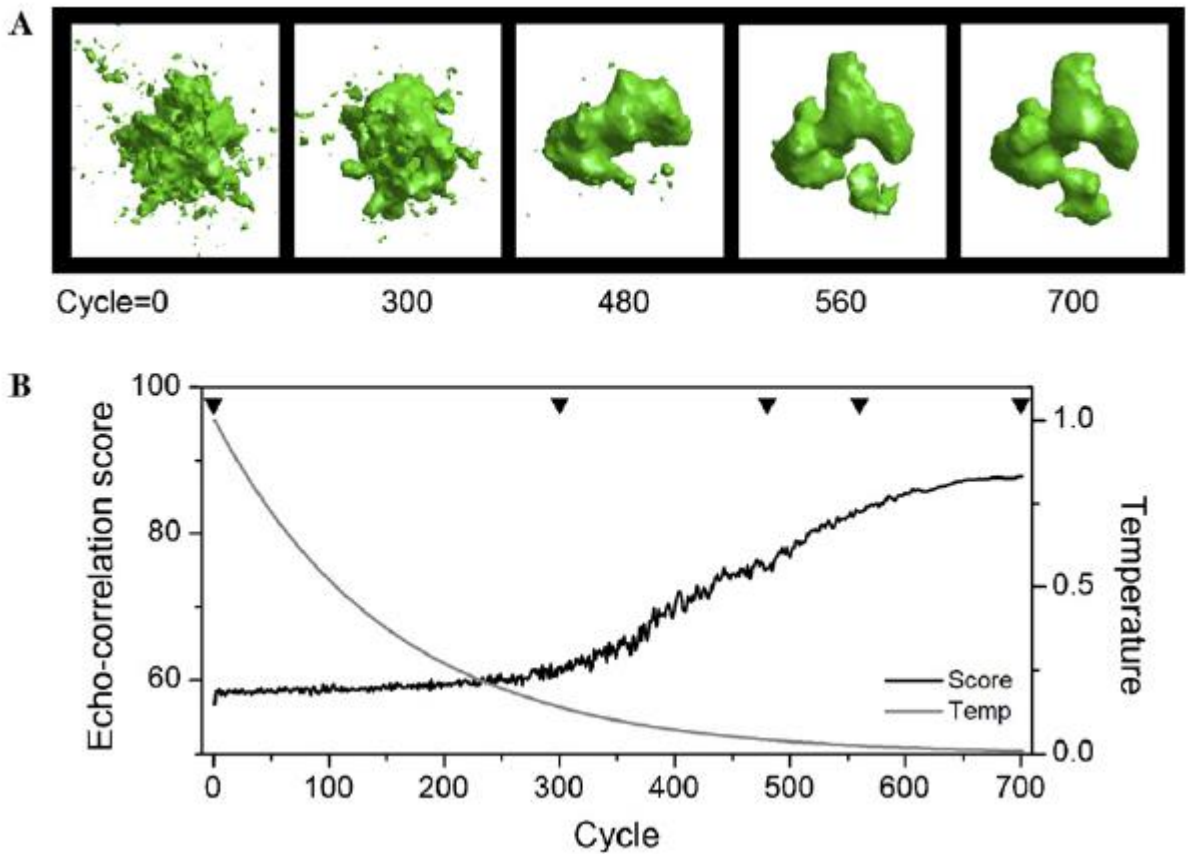
$$\min_{V,\theta} \sum_i \left\| I_i - P_{\theta_i} V \right\|^2$$



$$\Rightarrow 200 \cdot 3 = 600D$$

Ogura, T. & Sato, C. Posterior Euler angle assignment using simulated annealing J. Structural Biology, 2006, 156, 371-386

Option 5: Stochastic optimization



CSIC

instruct Integrating Biology

instruct image Processing Center

Option 5: Stochastic optimization

$$\min_{\theta} \sum_{ij=1}^{N} \left\| CL_{\theta_i}(I_i) - CL_{\theta_j}(I_j) \right\|^2$$

Shifts=0 $\Longrightarrow$ Simulated annealing: 3N D
+ Differential evolution: 5N D

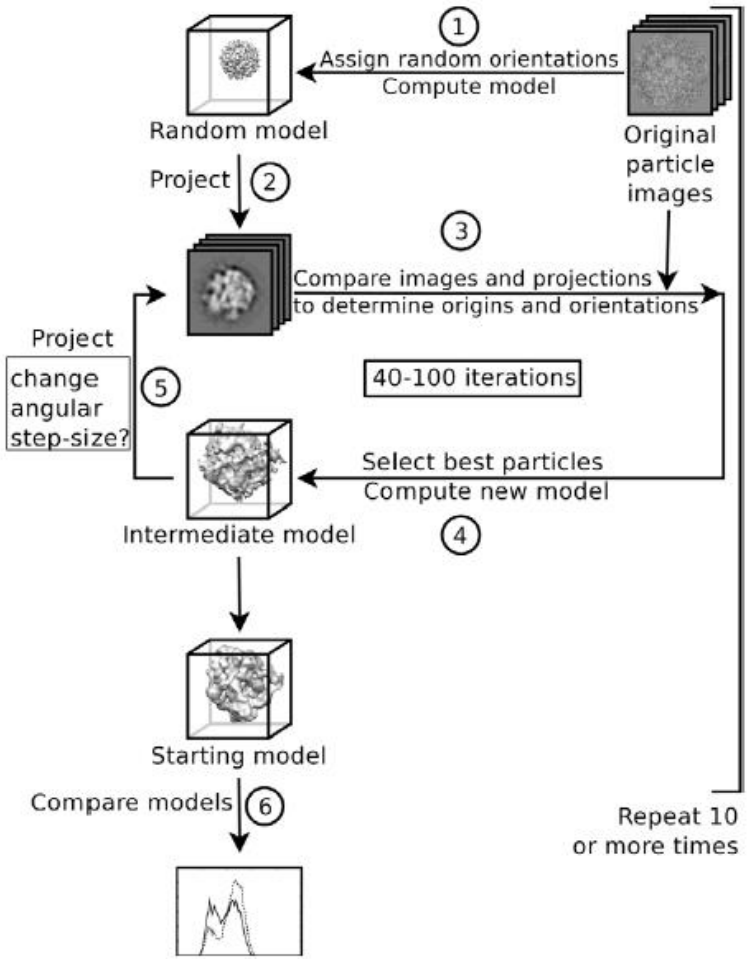
1. Elmlund, H.; Lundqvist, J.; Al-Karadaghi, S.; Hansson, M.; Hebert, H. & Lindahl, M. A new cryo-EM single-particle ab initio reconstruction method visualizes secondary structure elements in an ATP-fueled AAA+ motor. J Mol Biol, 2008, 375, 934-947
2. Elmlund, D.; Davis, R. & Elmlund, H. Ab initio structure determination from electron microscopic images of single molecules coexisting in different functional states. *Structure,* **2010***, 18*, 777-786
3. Elmlund, D. & Elmlund, H. SIMPLE: Software for ab initio reconstruction of heterogeneous single-particles. J Struct Biol, 2012, 180, 420-42
4. Reboul, C. F.; Eager, M.; Elmlund, D. & Elmlund, H. Single-particle cryo-EM-Improved ab initio 3D reconstruction with SIMPLE/PRIME. Protein science, 2018, 27, 51-61

CSIC

instruct Integrating Biology

instruct image Processing Center

Option 5: Stochastic optimization



Relationship of success rate to the number of particle images.[a]

Number of particle images	Success rate (%)	Iteration at convergence[b]
500	20	–
1000	20	–
3000	40	80 (±23)
10,000	40	58 (±15)
76,000	90	57 (±23)

Execution time

$$\Rightarrow 100k \cdot 3 = 300kD$$

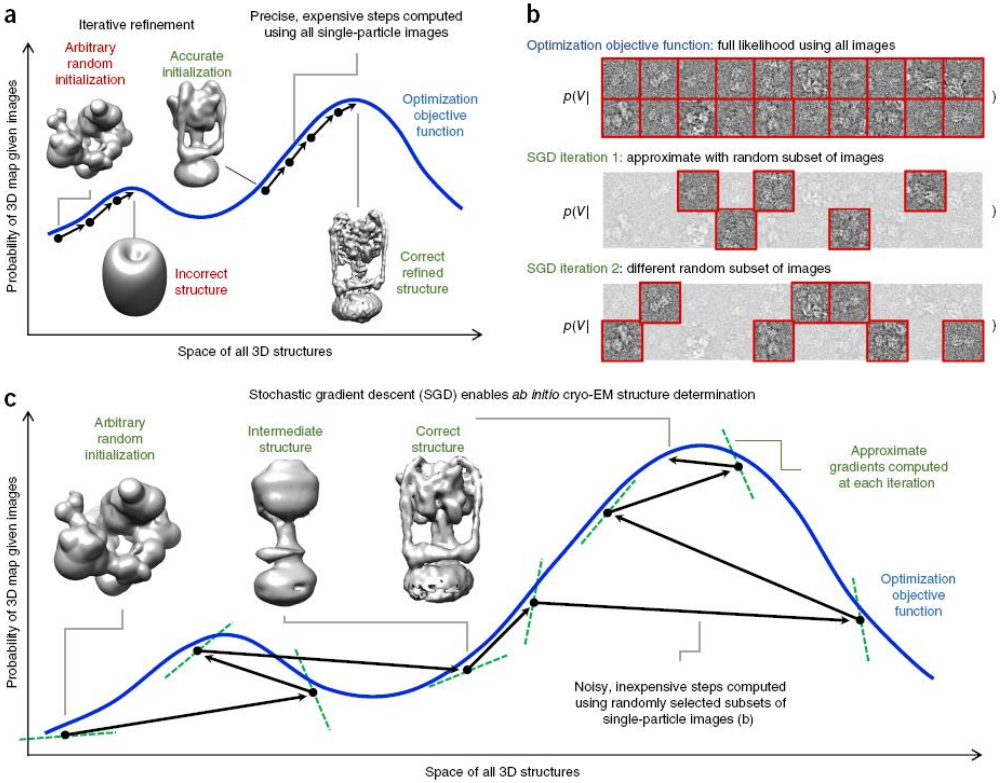
Sanz-García, E.; Stewart, A. B. & Belnap, D. M. The random-model method enables ab initio 3D reconstruction of asymmetric particles and determination of particle symmetry. J Struct Biol, 2010, 171, 216-222

CSIC

instruct Integrating Biology

instruct image Processing Center

Option 5: Stochastic Gradient Descent



Punjani, A.; Brubaker, M. A. & Fleet, D. J. Building Proteins in a Day: Efficient 3D Molecular Structure Estimation with Electron Cryomicroscopy. IEEE transactions on pattern analysis and machine intelligence, 2017, 39, 706-718

Punjani, A.; Rubinstein, J. L.; Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nature methods, 2017

CSIC

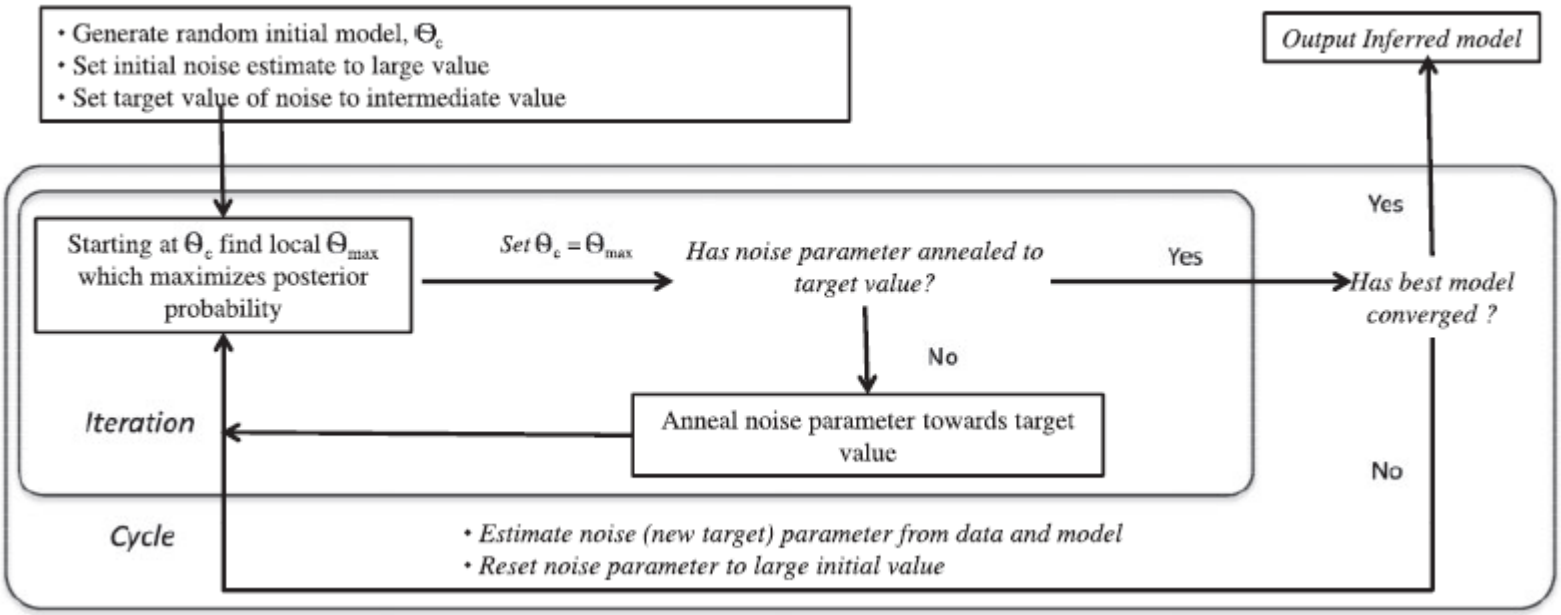
instruct Integrating Biology

instruct image Processing Center

Option 5: Deterministic Annealing+ Bayesian prior

$$p(\mathbf{I}|\mathbf{R},\theta)=\mathbf{N}\left(\mathbf{I}|\mathbf{P}(\theta,\mathbf{R}),\sigma^2\right)$$

$$p(\theta)\propto\exp\left(-\lambda\sum_{(i,j)\in\text{Neighbours}(\theta)}(\alpha_i-\alpha_j)^2\right)$$ Low local variance



Jaitly, N.; Brubaker, M. A. & Rubinstein J. L. snf Lilien, R. H. A Bayesian method for 3-D macromolecular structure inference using class average images from single particle electron microscopy Bioinformatics, 2010

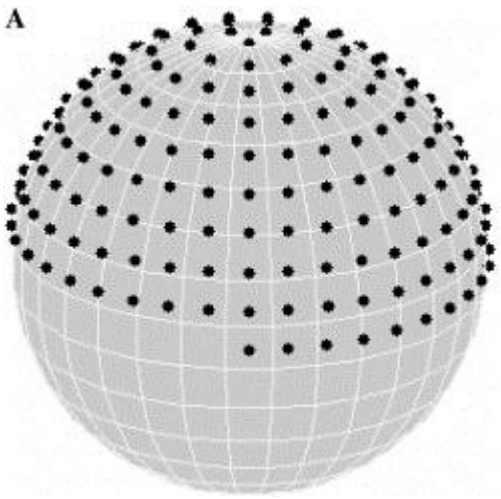
CSIC

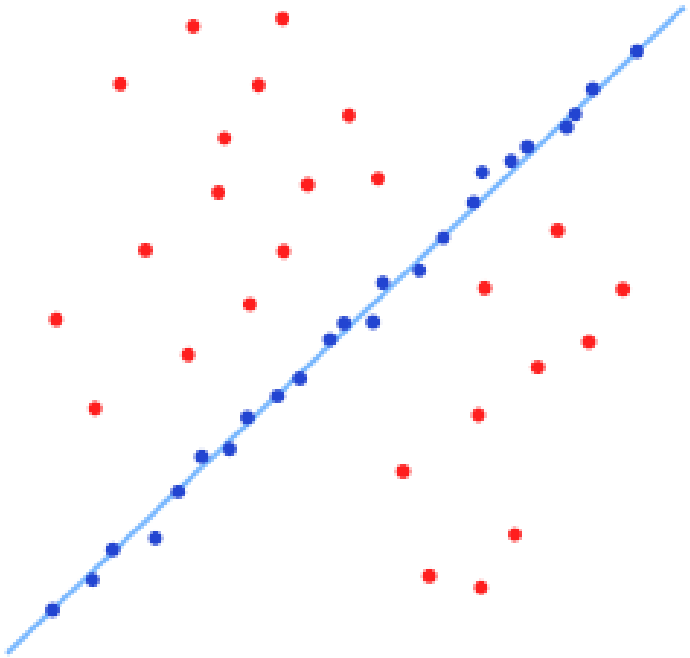
instruct Integrating Biology

instruct image Processing Center

Option 5: RANSAC

$$\min_{V,\theta} \sum_i \left\| I_i - P_{\theta_i} V \right\|^2$$





J. Vargas, A.L. Álvarez-Cabrera, R. Marabini, J.M. Carazo, **C.O.S. Sorzano**. *Efficient initial volume determination from electron microscopy images of single particles*. Bioinformatics, 30: 2891-2898 (2014)

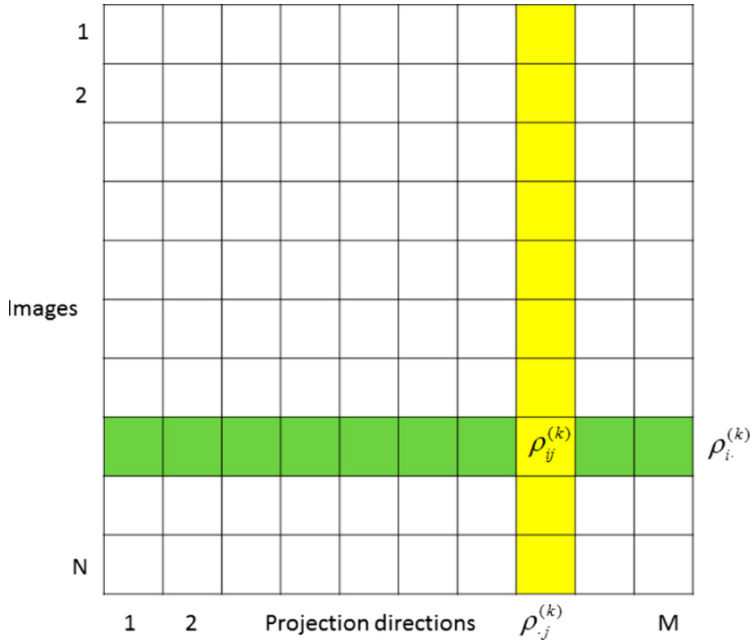
Random subset: 3x9 D
+ Local refinement: 5N D (greedy)

CSIC

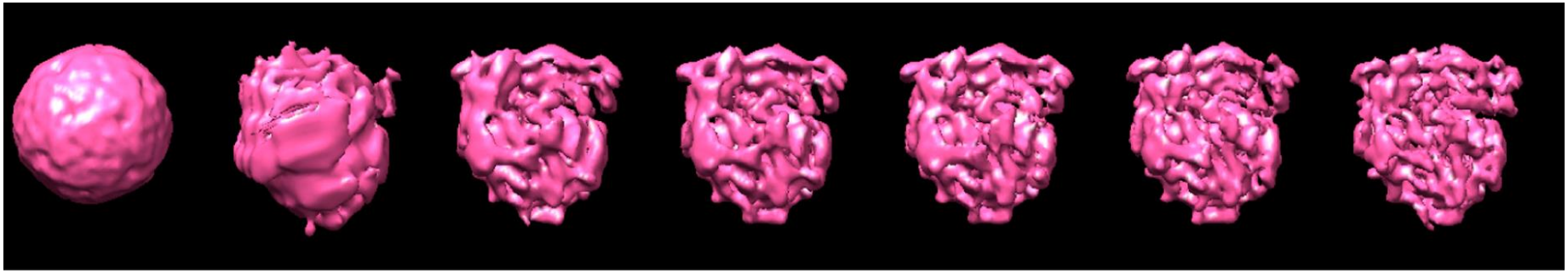
instruct Integrating Biology

instruct image Processing Center

Option 6: Reconstruct Significant



C.O.S. Sorzano, J. Vargas, J.M. de la Rosa-Trevín, J. Otón, A.L. Álvarez-Cabrera, V. Abrishami, E. Sesmero, R. Marabini, J.M. Carazo. A Statistical approach to the initial volume problem in Single Particle Analysis by Electron Microscopy. J. Structural Biology, 189: 213-219 (2015)



CSIC

instruct Integrating Biology

instruct image Processing Center

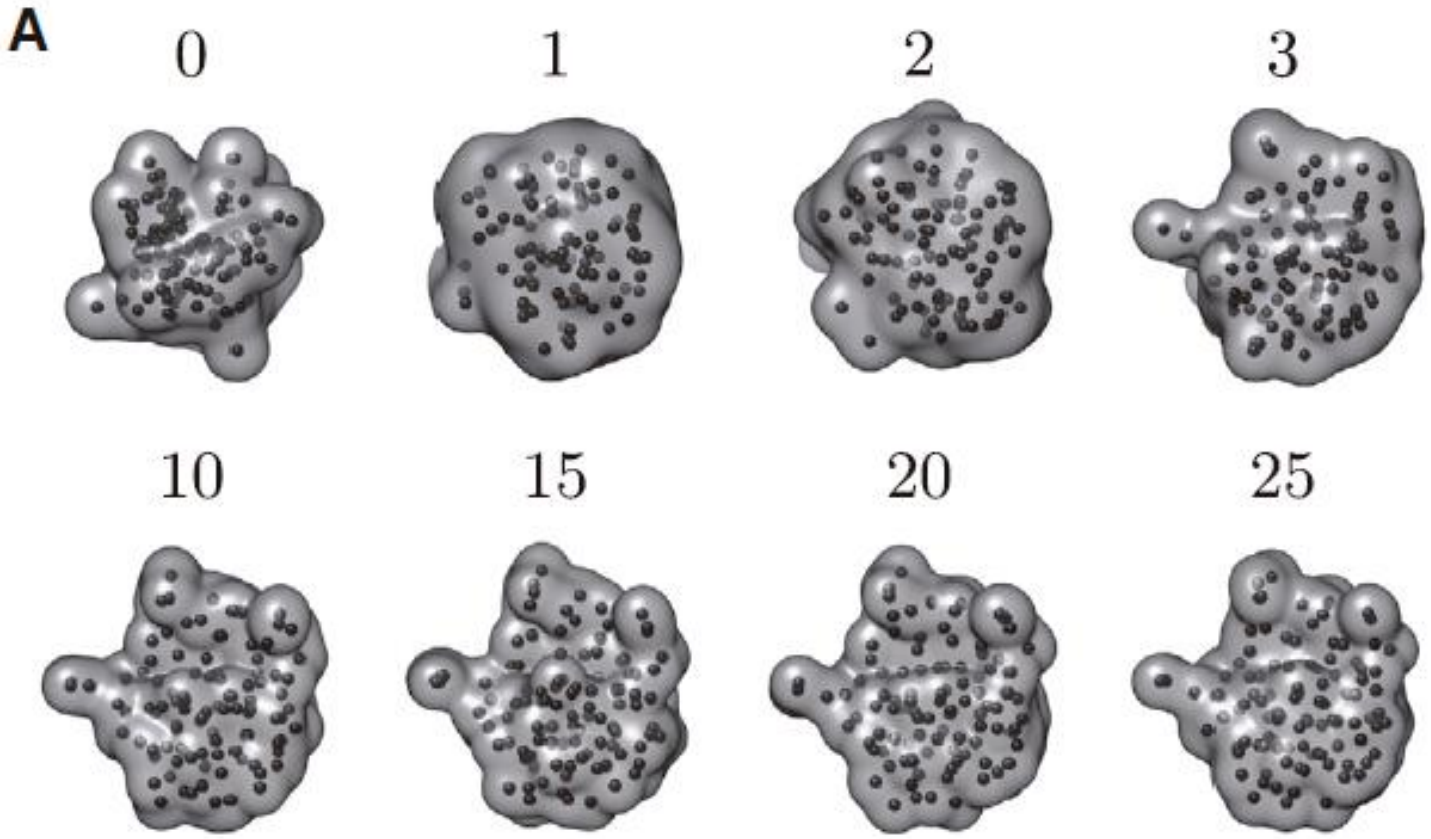
Option 7: EMAN

CSIC

instruct Integrating Biology

instruct image Processing Center

Option 8: Pseudoatoms



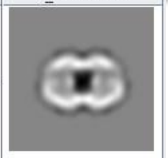
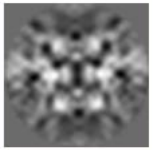
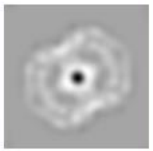
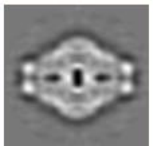
Joubert, P. & Habeck, M. Bayesian Inference of Initial Models in Cryo-Electron Microscopy Using Pseudo-atoms. *Biophys J,* **2015**, *108*, 1165-1175

CSIC

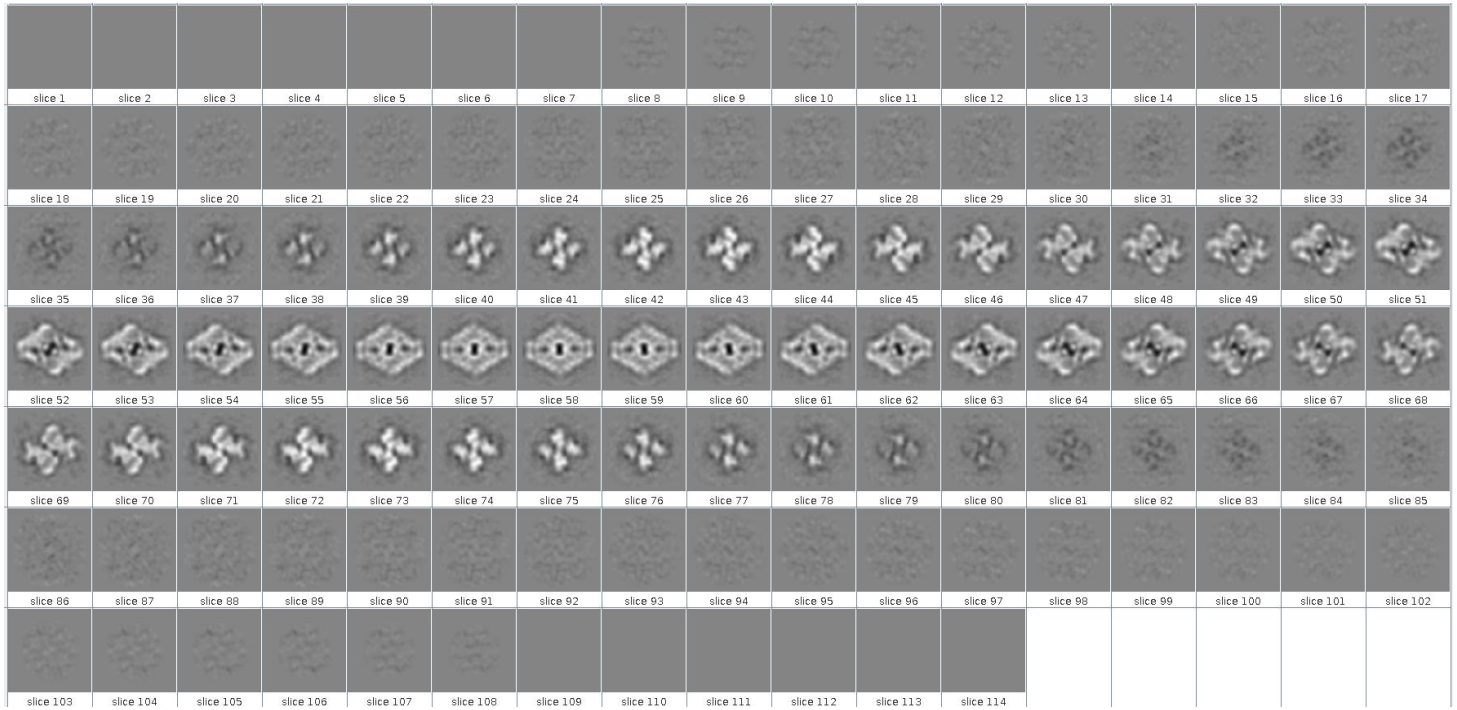
instruct Integrating Biology

instruct image Processing Center

Swarm consensus

comment	_filename
eman initial model ...	
eman initial model ...	
eman initial model ...	
eman initial model ...	
eman initial model ...	
eman initial model ...	
significant volume 1	

[Swarm Consensus]



instruct
image
Processing
Center

Furthermore … reality is heterogeneous

CSIC

instruct Integrating Biology

BIO COMPUTING Unit

instruct image Processing Center

Current situation (2013)

… but there is …

CSIC

instruct Integrating Biology

instruct image Processing Center

Current situation (2016)

… it is no longer a major bottleneck …except for very heterogeneous samples

CSIC

instruct Integrating Biology

instruct image Processing Center

Random Conical Tilt

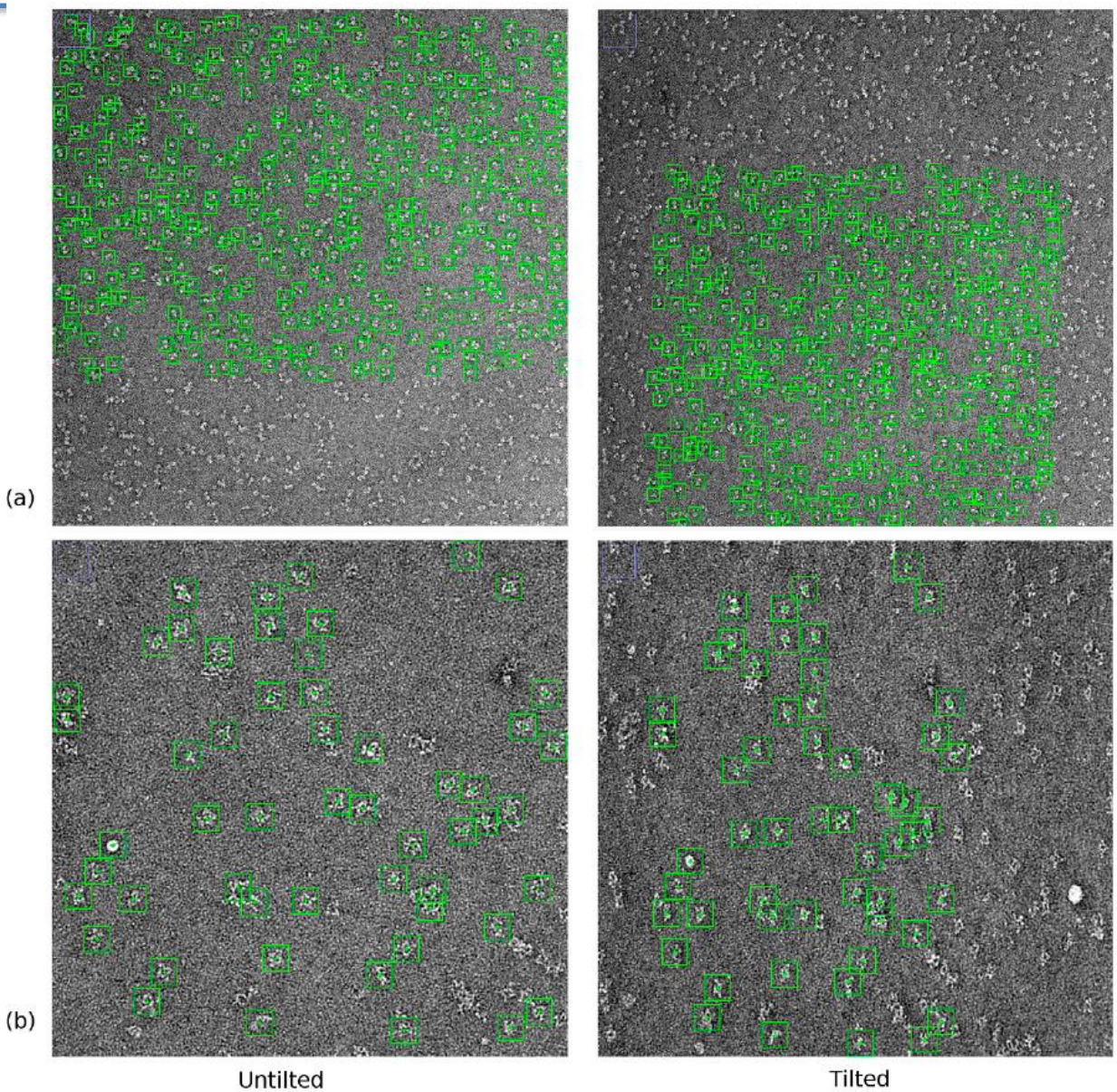
Carlos Oscar S. Sorzano
Instruct Image Processing Center

CSIC

instruct Integrating Biology

instruct image Processing Center

Random Conical Tilt

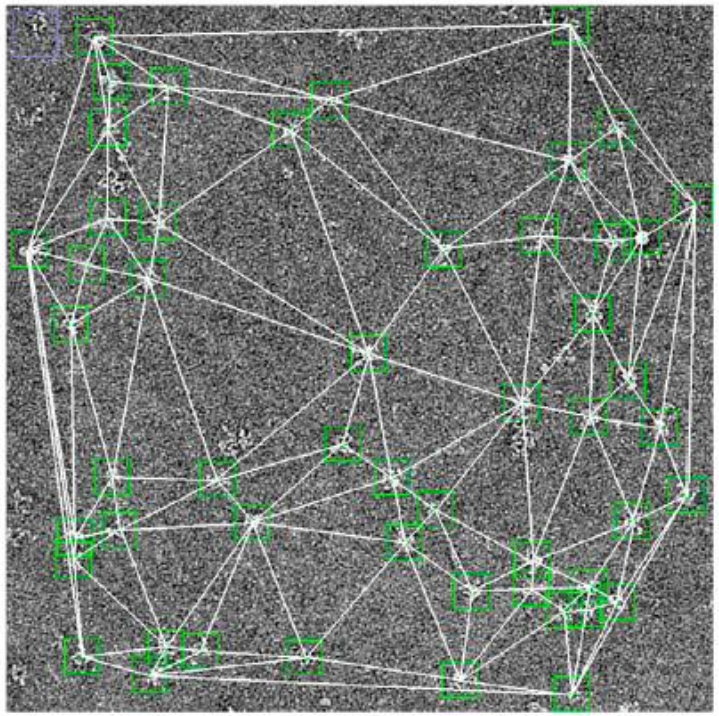


J.L. Vilas, J. Navas, J. Gomez-Blanco, J.M. de la Rosa-Trevin, R. Melero, I. Peschiera, I. Ferlenghi, J. Cuenca, J.M. Carazo, J.Vargas, C.O.S. Sorzano. Fast and automatic identification of particle tilt pairs based on Delaunay triangulation. J. Structural Biology 196: 525-533 (2016)

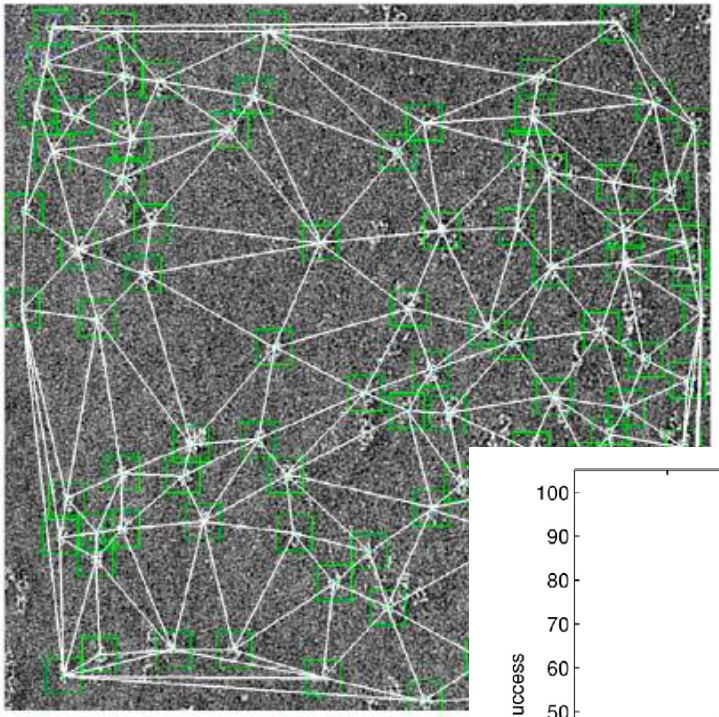
CSIC

Instruct Integrating Biology

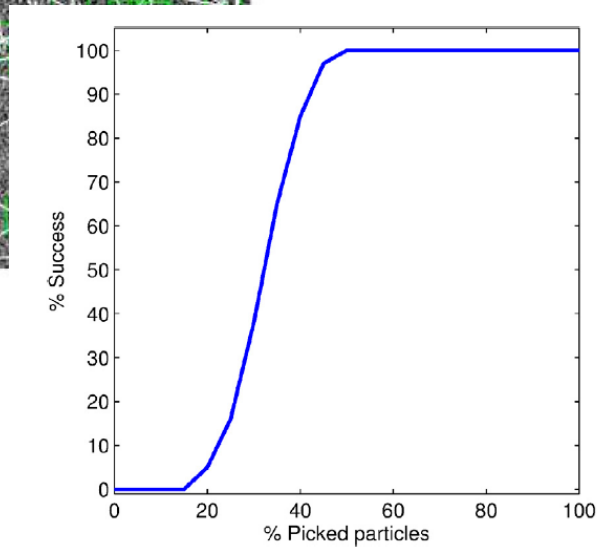
Random Conical Tilt



(a) Untilted

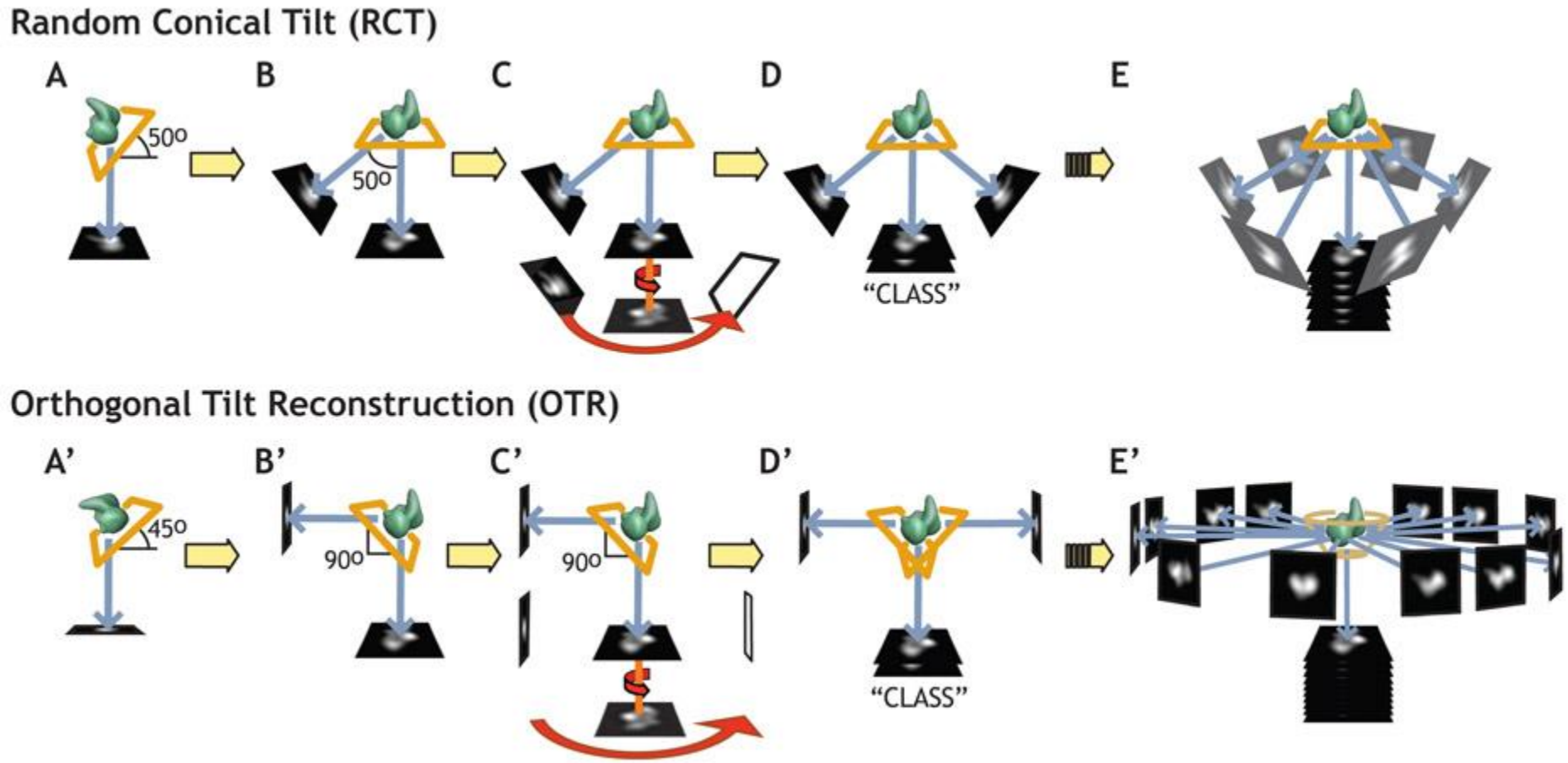


(b) Tilted



J.L. Vilas, J. Navas, J. Gomez-Blanco, J.M. de la Rosa-Trevin, R. Melero, I. Peschiera, I. Ferlenghi, J. Cuenca, J.M. Carazo, J.Vargas, C.O.S. Sorzano. Fast and automatic identification of particle tilt pairs based on Delaunay triangulation. J. Structural Biology 196: 525-533 (2016)

Random Conical Tilt



Chandramouli P, Hernandez-Lopez R, Wang HW, Leschziner AE (2011). Validation of the ortogonal tilt reconstruction method with a biological test sample. J Struct Biol. 175:85-96
Leschziner A E (2010). The ortogonal tilt reconstruction method. Methods in Enzymology. 2010 482:237-262
Leschziner AE and Nogales E (2006). The Orthogonal Tilt Reconstruction method: an approach to generating single-class volumes with no missing cone for ab initio reconstruction of asymmetric particles. J Struct Biol. 153(3):284-99.

Random Conical Tilt

C.O.S. Sorzano, M. Alcorlo, J.M. de la Rosa-Trevín, R. Melero, I. Foche, A. Zaldívar-Peraza, L. del Cano, J. Vargas, V. Abrishami, J. Otón, R. Marabini, J.M. Carazo. Cryo-EM and the elucidation of new macromolecular structures: Random Conical Tilt revisited. Nature Scientific Reports, 5: 14290 (2015)

$$I_{\text{micrograph},u}(\tilde{\mathbf{s}}) = \int^{\infty} V_{\text{micrograph}}(\tilde{R}_Z^{-1}(\alpha_u)\tilde{H}^T\tilde{\mathbf{s}})dz$$

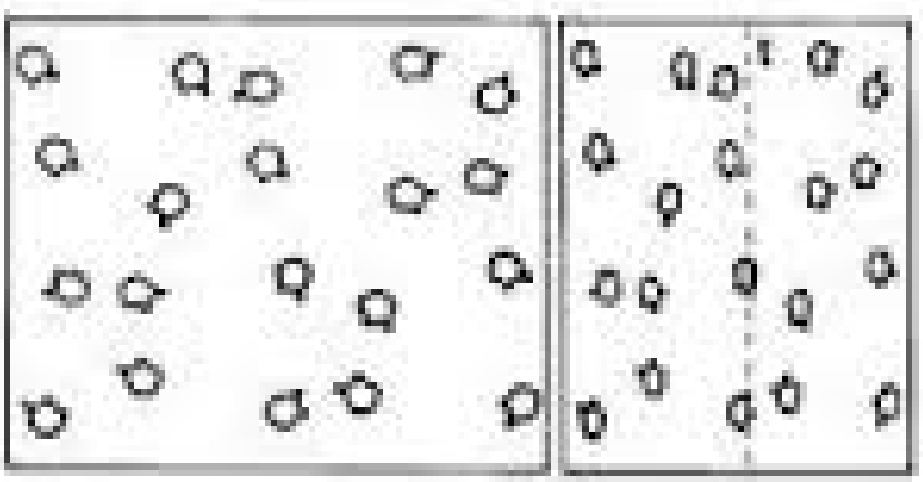
$$I_{\text{micrograph},t}(\tilde{\mathbf{s}}) = \int_{-\infty}^{\infty} V_{\text{micrograph}}((\tilde{T}\tilde{R}_Z(\alpha_t)\tilde{R}_Y(\theta))^{-1}\tilde{H}^T\tilde{\mathbf{s}})dz$$

$$\tilde{\mathbf{s}}_t = \tilde{H}_0\tilde{T}\tilde{R}_Z(\alpha_t)\tilde{R}_Y(\theta)\tilde{R}_Z^{-1}(\alpha_u)\tilde{H}_0^T\tilde{\mathbf{s}}_u$$

Random Conical Tilt

Radermacher, M.; Wagenknecht, T.; Verschoor, A. & Frank, J. Three-Dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of Escherichia coli J. Microscopy, 1987, 146, 113-136

relative to the projecting direction. They can be created from the orthogonal projections by stretching the latter in the direction perpendicular to the tilt axis by $1/\cos(\vartheta)$, with ϑ being the tilt angle. The relative shift between two projections is then determined by the location of the center of mass of the cross-correlation function.



Random Conical Tilt

C.O.S. Sorzano, M. Alcorlo, J.M. de la Rosa-Trevín, R. Melero, I. Foche, A. Zaldívar-Peraza, L. del Cano, J. Vargas, V. Abrishami, J. Otón, R. Marabini, J.M. Carazo. Cryo-EM and the elucidation of new macromolecular structures: Random Conical Tilt revisited. Nature Scientific Reports, 5: 14290 (2015)

$$\begin{aligned} I_u(\tilde{\mathbf{s}}) &= V'(\tilde{H}_0^T\tilde{\mathbf{s}}) \\ I_t(\tilde{\mathbf{s}}) &= \int_{-\infty}^{\infty} V'(\tilde{E}^{-1}\tilde{H}^T\tilde{T}_t^{-1}\tilde{\mathbf{s}})dz = V'(\tilde{H}_0^T(\tilde{H}_0\tilde{E}\tilde{H}_0^T)^{-1}\tilde{T}_t^{-1}\tilde{\mathbf{s}}) \\ &= I_u((\tilde{H}_0\tilde{E}\tilde{H}_0^T)^{-1}\tilde{T}_t^{-1}\tilde{\mathbf{s}}) \end{aligned} \tag{12}$$

In the particular case that $\alpha_u = \alpha_t = 0$ (i.e., the tilt is performed around the Y axis in both images), then the previous equation becomes

$$\begin{aligned} I_u(\tilde{\mathbf{s}}) &= V'(\tilde{H}_0^T\tilde{\mathbf{s}}) \\ I_t(\tilde{\mathbf{s}}) &= V'(\tilde{H}_0^T S\tilde{T}_t^{-1}\tilde{\mathbf{s}}) = I_u(S\tilde{T}_t^{-1}\tilde{\mathbf{s}}) \end{aligned} \tag{13}$$

where

$$S = \begin{pmatrix} \frac{1}{\cos(\theta)} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{14}$$

Modifications to RCT

- Extend the theory for arbitrary tilt axis orientation.
- Align tilted images with respect to classes (not untilted images).
- Explicitly consider mirrors.
- Explictly consider picking errors.

C.O.S. Sorzano, M. Alcorlo, J.M. de la Rosa-Trevín, R. Melero, I. Foche, A. Zaldívar-Peraza, L. del Cano, J. Vargas, V. Abrishami, J. Otón, R. Marabini, J.M. Carazo. Cryo-EM and the elucidation of new macromolecular structures: Random Conical Tilt revisited. Nature Scientific Reports, 5: 14290 (2015)

Modifications to RCT

$$
\begin{aligned}
I_u(\tilde{\mathrm{s}}) &= \int_{-\infty}^{\infty} V_{\mathrm{ref}}(\tilde{M}_u \tilde{H}^T \tilde{\mathrm{s}}) dz \\
I_t(\tilde{\mathrm{s}}) &= \int_{-\infty}^{\infty} V_{\mathrm{ref}}(\tilde{M}_u \tilde{R}_Z(\alpha_u)(\tilde{R}_Z(\alpha_t)\tilde{R}_Y(\theta))^{-1} \tilde{H}^T \tilde{T}_t^{-1} \tilde{\mathrm{s}}) dz \\
&= \int_{-\infty}^{\infty} V_{\mathrm{ref}}(\tilde{T}_u \tilde{R}_Z(\alpha'_u)\tilde{R}_Z(\alpha_u)(\tilde{R}_Z(\alpha_t)\tilde{R}_Y(\theta))^{-1} \tilde{H}^T \tilde{T}_t^{-1} \tilde{\mathrm{s}}) dz \\
&= \int_{-\infty}^{\infty} V_{\mathrm{ref}}(\tilde{T}_u (\tilde{R}_Z(\alpha_t)\tilde{R}_Y(\theta)\tilde{R}_Z(-(\alpha'_u + \alpha_u)))^{-1} \tilde{H}^T \tilde{T}_t^{-1} \tilde{\mathrm{s}}) dz \\
&= \int_{-\infty}^{\infty} V_{\mathrm{ref}}(\tilde{T}_u \tilde{E}^{-1} \tilde{H}^T \tilde{T}_t^{-1} \tilde{\mathrm{s}}) dz,
\end{aligned}
$$

$$
I_t(\tilde{\mathrm{s}}) = \int_{-\infty}^{\infty} V_{\mathrm{ref}}(\tilde{E}^{-1} \tilde{T}'_u \tilde{H}^T \tilde{T}_t^{-1} \tilde{\mathrm{s}}) dz = \int_{-\infty}^{\infty} V_{\mathrm{ref}}(\tilde{E}^{-1} \tilde{H}^T \tilde{T}'_{u,2D} \tilde{T}_t^{-1} \tilde{\mathrm{s}}) dz
$$

$$
\tilde{T}'_t = \tilde{T}_t \left(\tilde{T}'_{u,2D}\right)^{-1}
$$

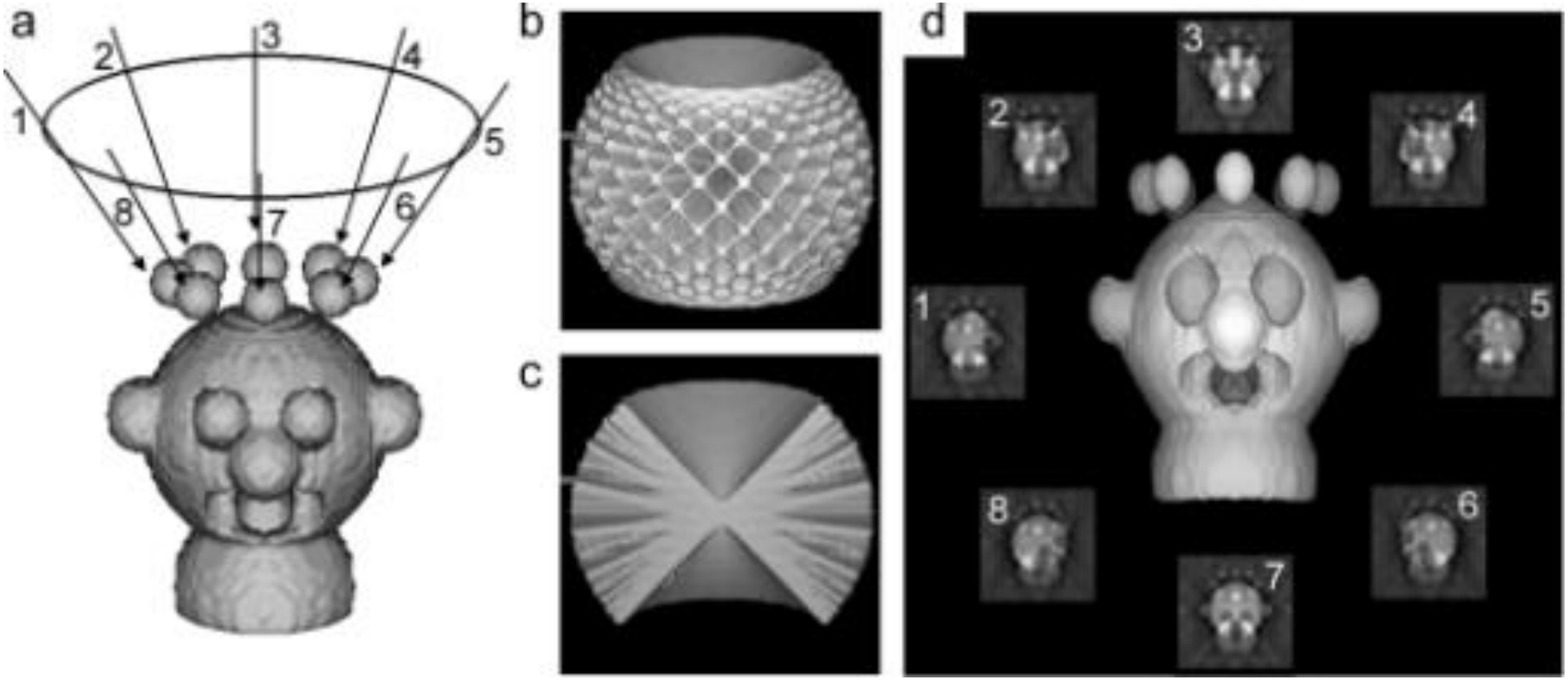
$$
\tilde{T}_t = (\tilde{A}^{\mathrm{2Dclass}})^{-1} (\tilde{T}_t^{\mathrm{2Dclass}})^{-1} \tilde{A}^{\mathrm{2Dclass}}
$$

CSIC

instruct Integrating Biology

instruct image Processing Center

Random Conical Tilt



Orlova, E. V. & Saibil, H. R. Structural analysis of macromolecular assemblies by electron microscopy. Chem Rev, 2011, 111, 7710-7748

RCT with only 13 tilt pairs ("ground truth")

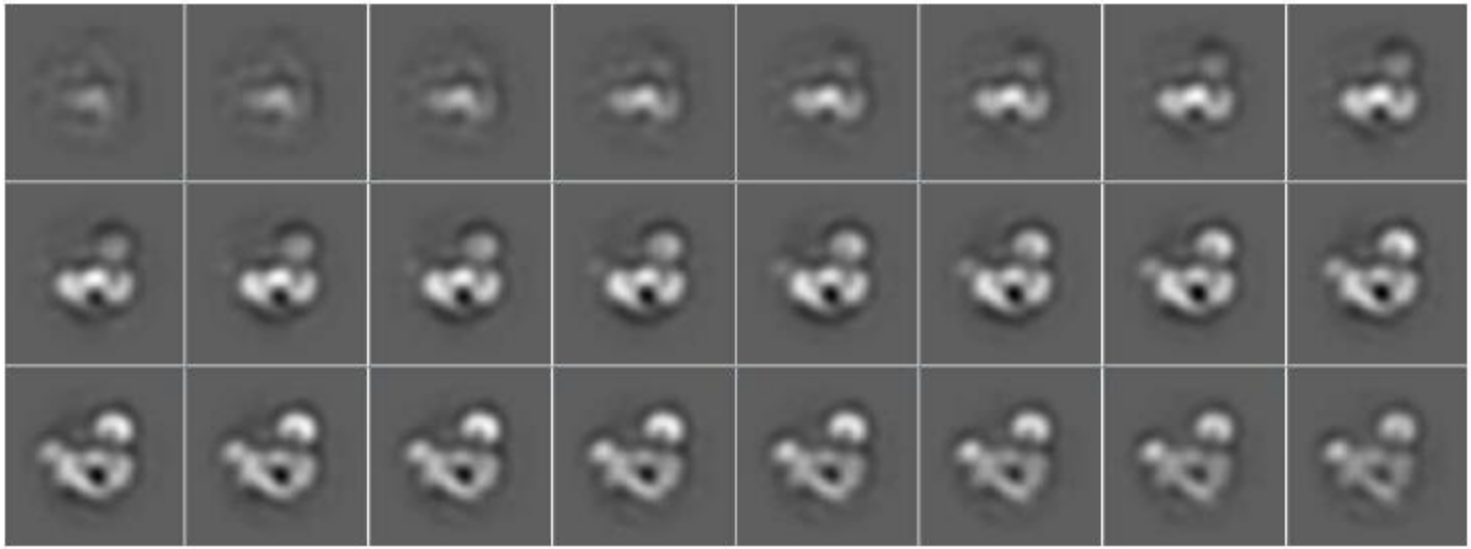


Figure 1: Control map: Central slices of the control map of C3b computed from 32,595 images and a 3D classification procedure (see main text for details).

CSIC

instruct Integrating Biology

instruct image Processing Center

RCT with only 13 tilt pairs ("old RCT")

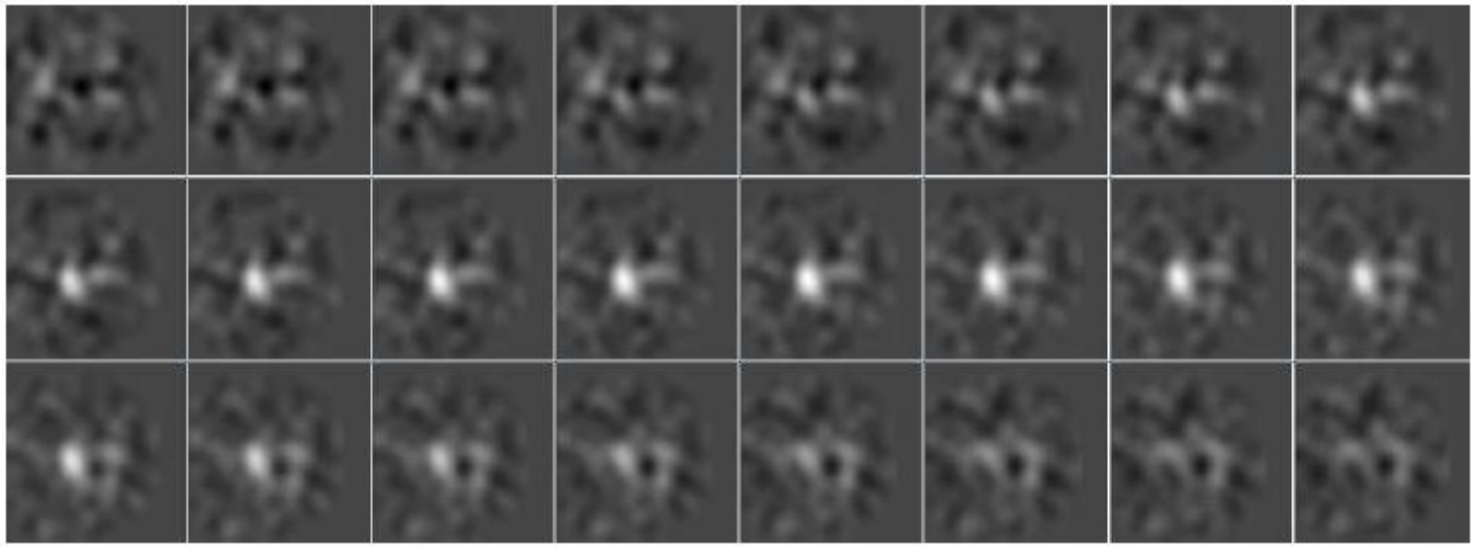


Figure 2: Standard RCT workflow for small-sized 2D classes (20 image pairs): Central slices of the structure of C3b using standard RCT, as implemented in Spider. Compare them to Suppl. Fig. 1.

RCT with only 13 tilt pairs ("revised RCT")

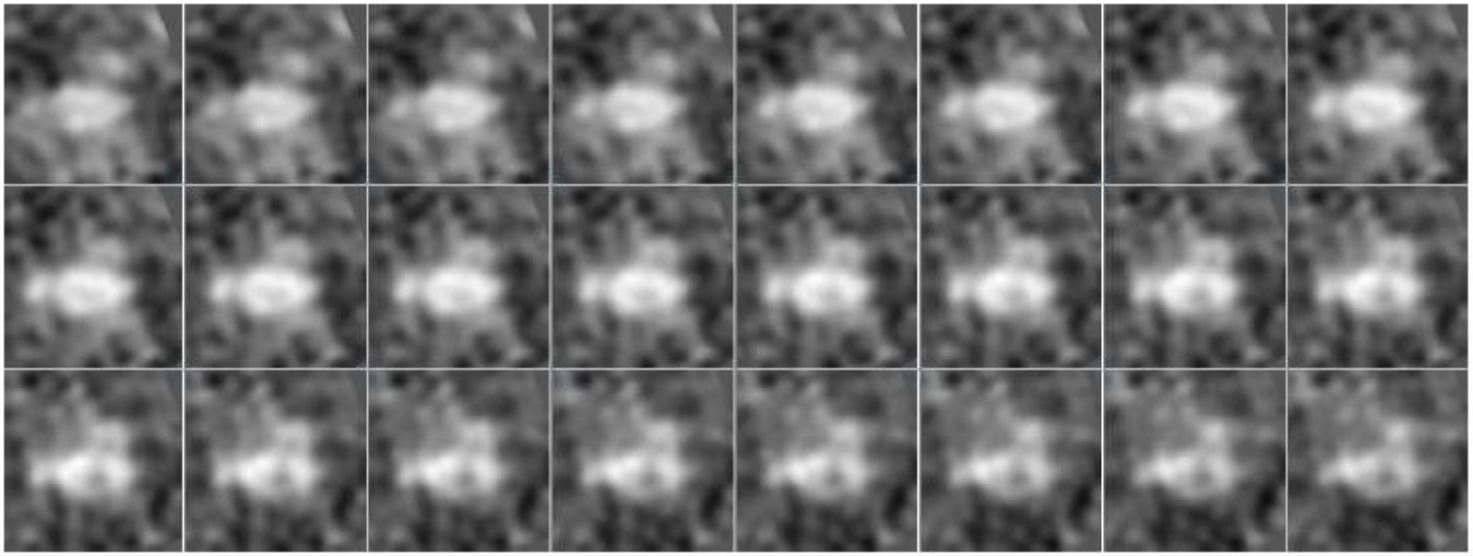


Figure 3: Modified RCT workflow for small-sized 2D classes (13 image pairs): Central slices of the structure of C3b using modified RCT, as implemented in Xmipp. Compare them to Suppl. Fig. 1.

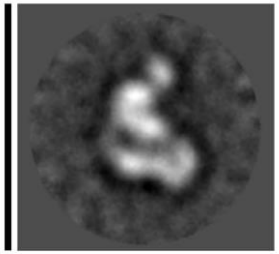
CSIC

instruct Integrating Biology

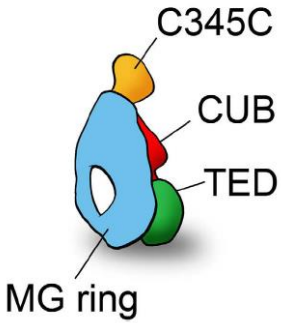
instruct image Processing Center

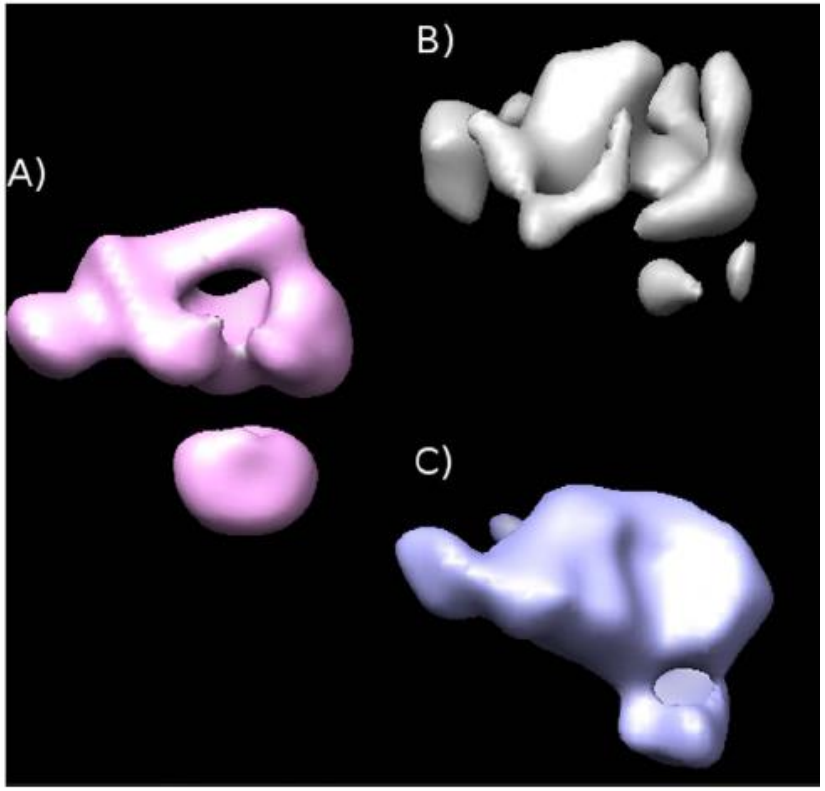
RCT with 13 tilt pairs

2D reference-free average



Schematic cartoon



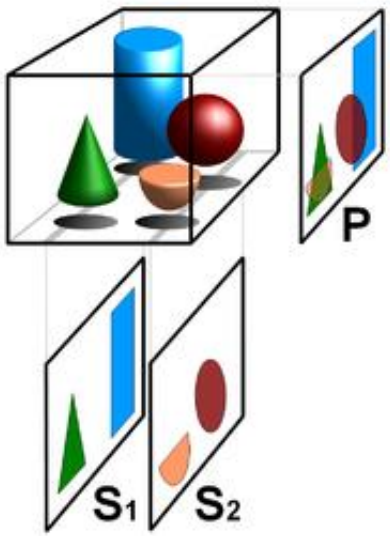


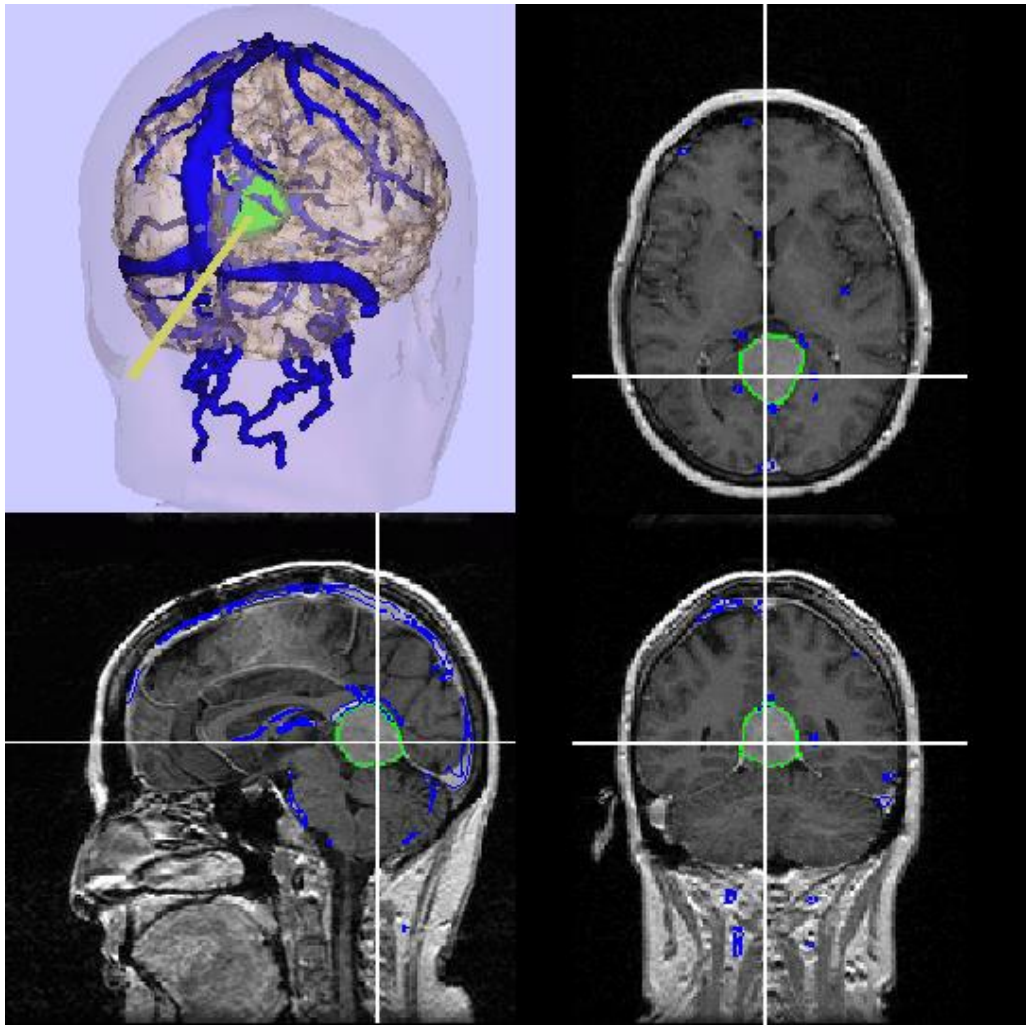
Reconstruction Methods

Carlos Oscar S. Sorzano

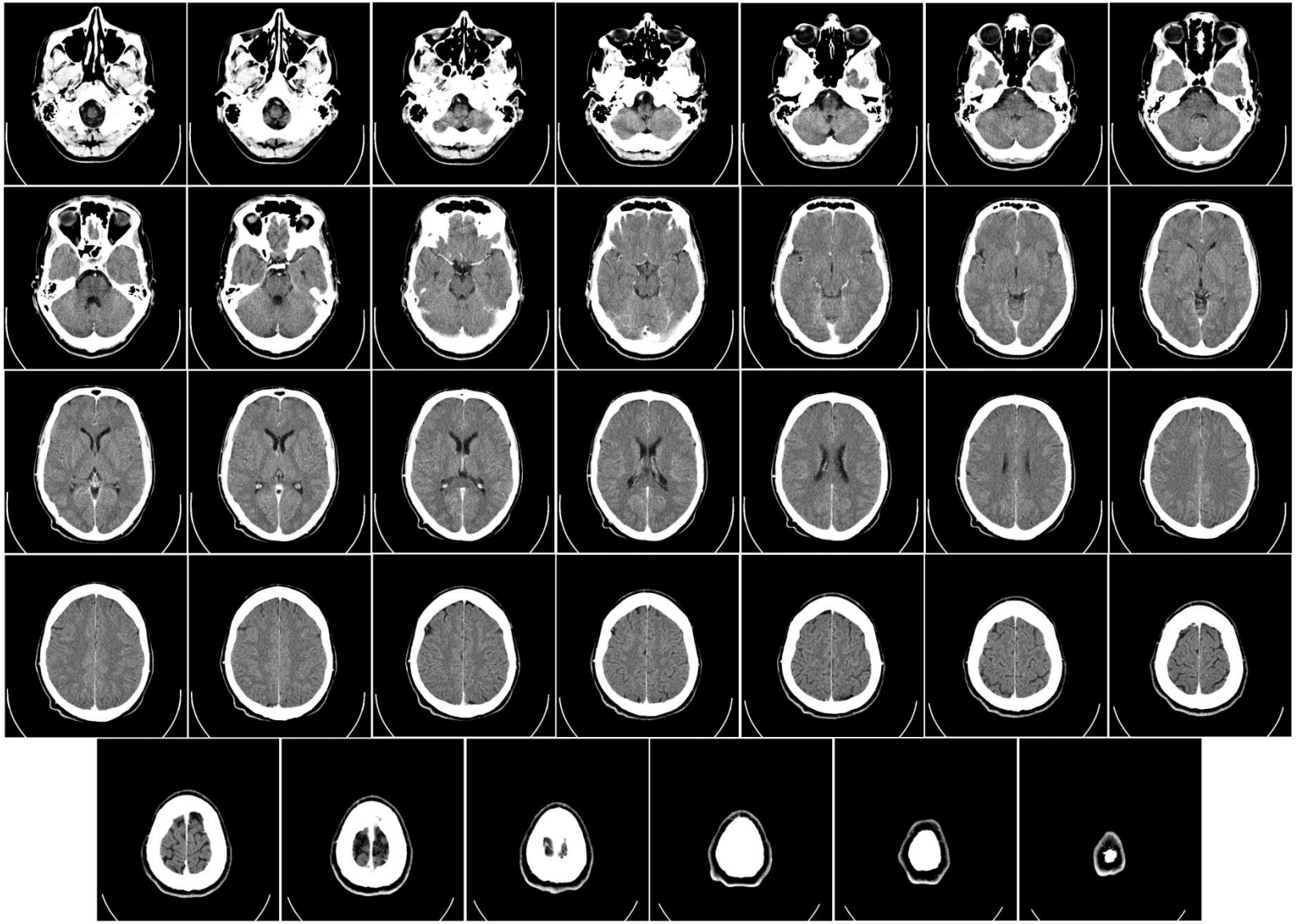
Instruct Image Processing Center

Tomography

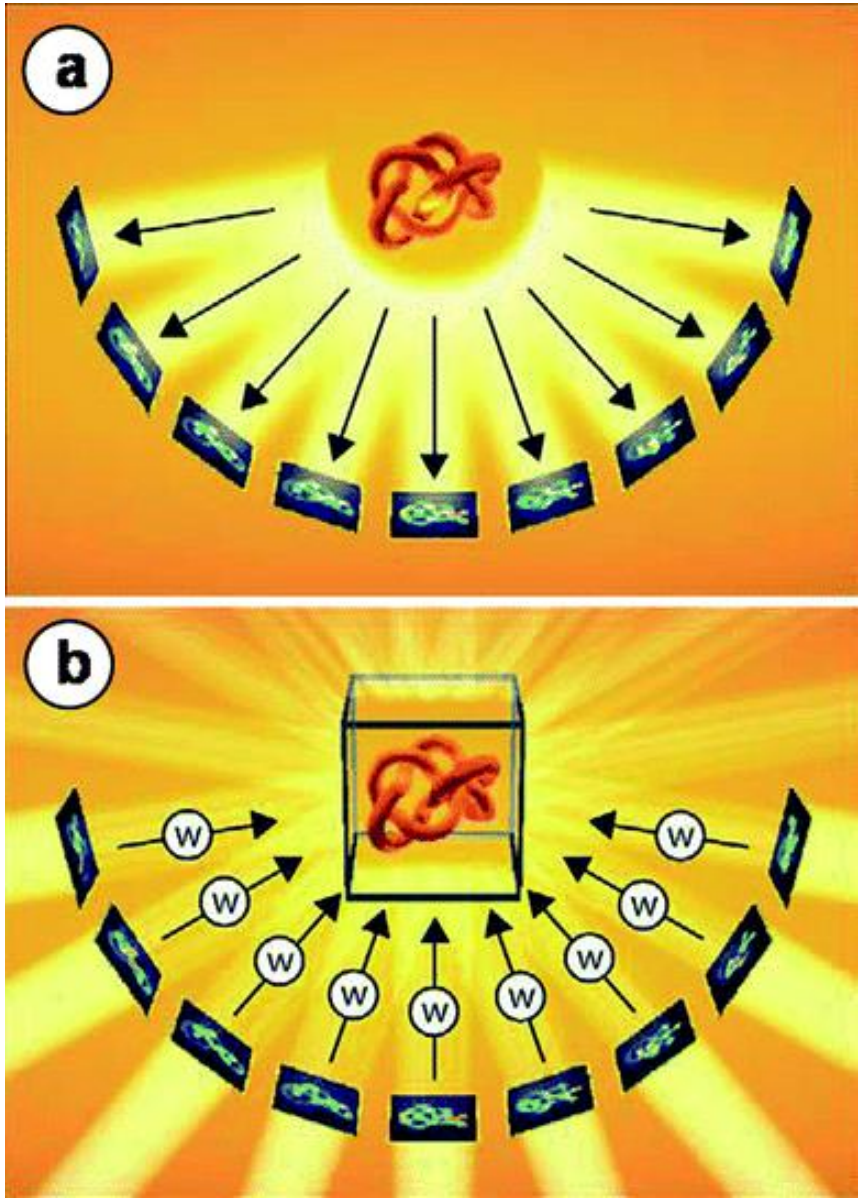




Tomography



Tomography



Weighted Back Projection

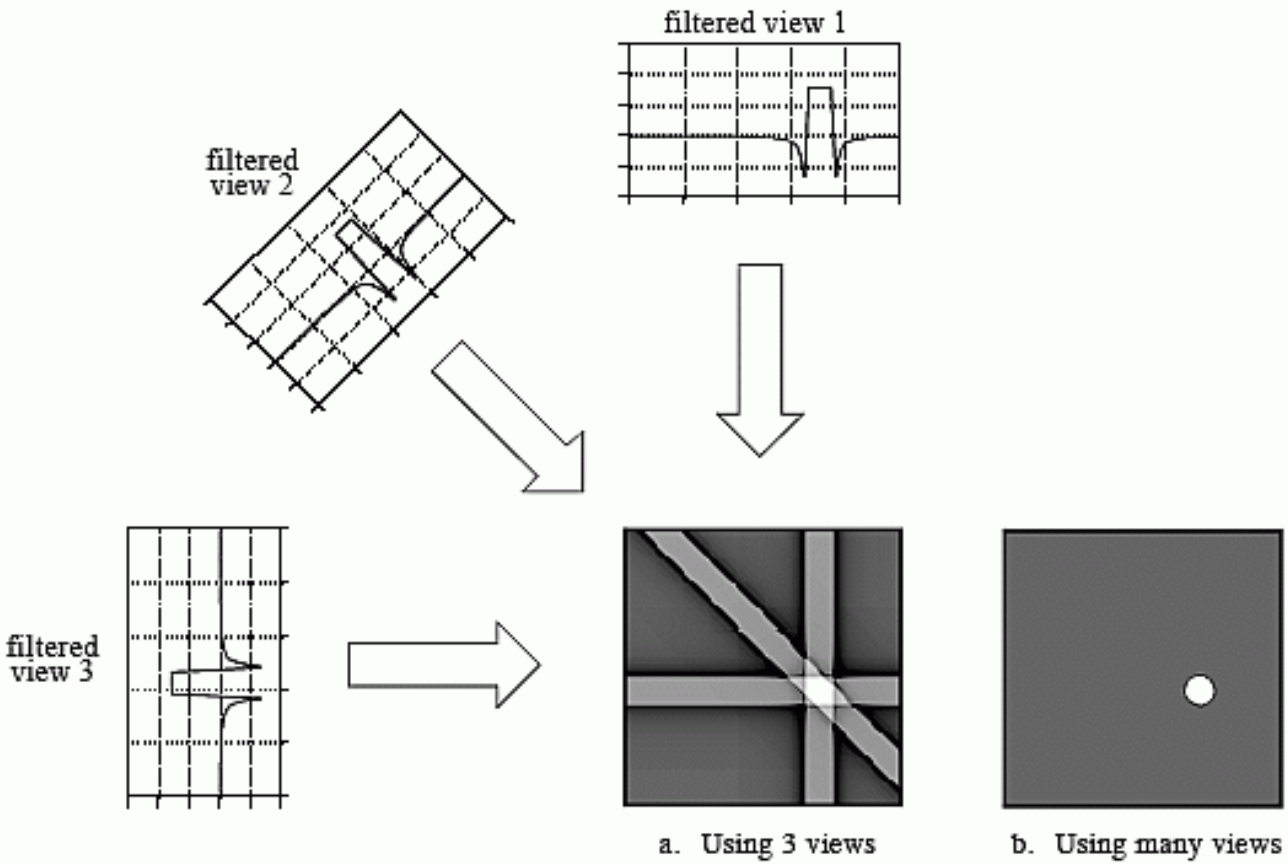
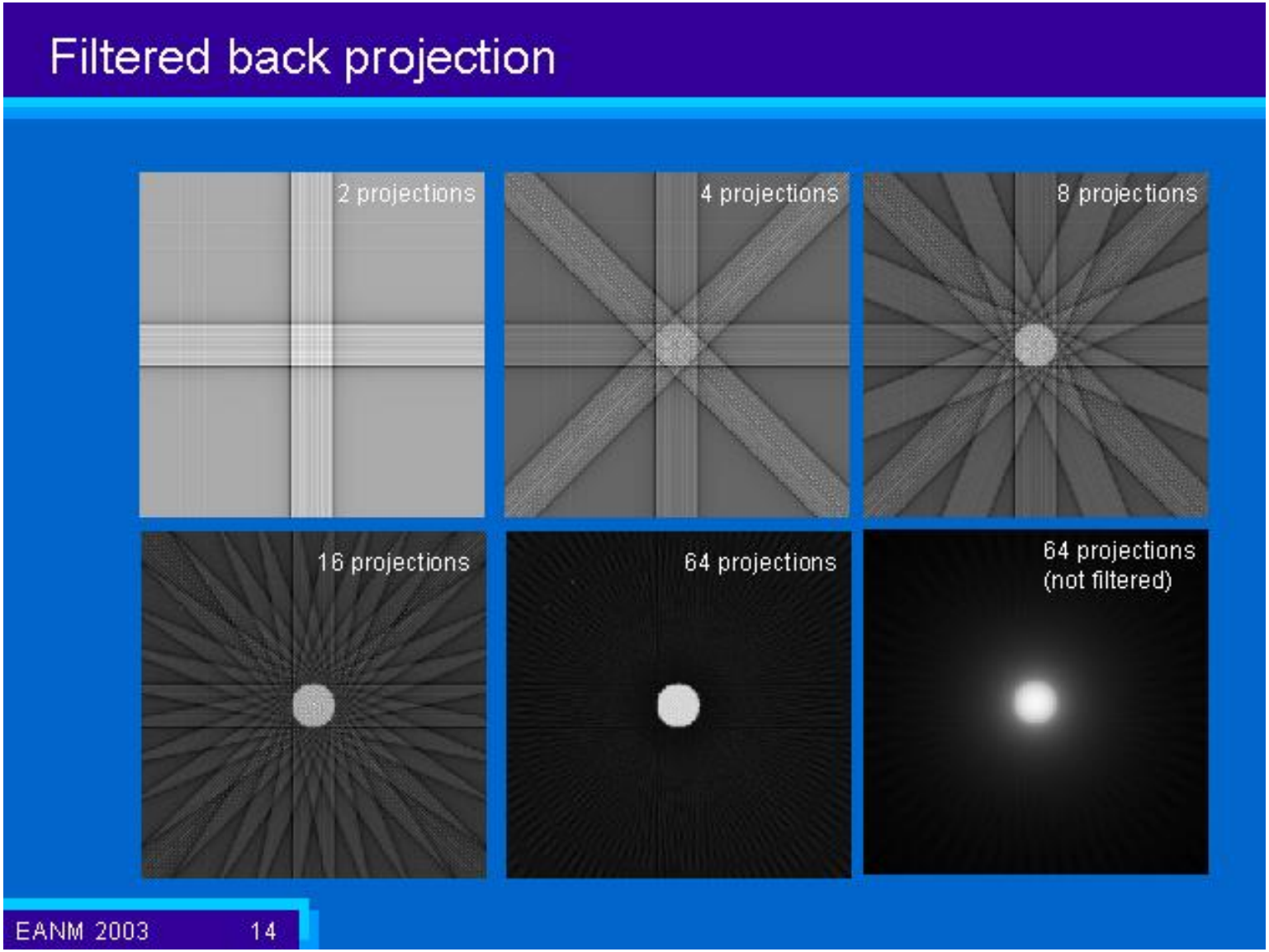
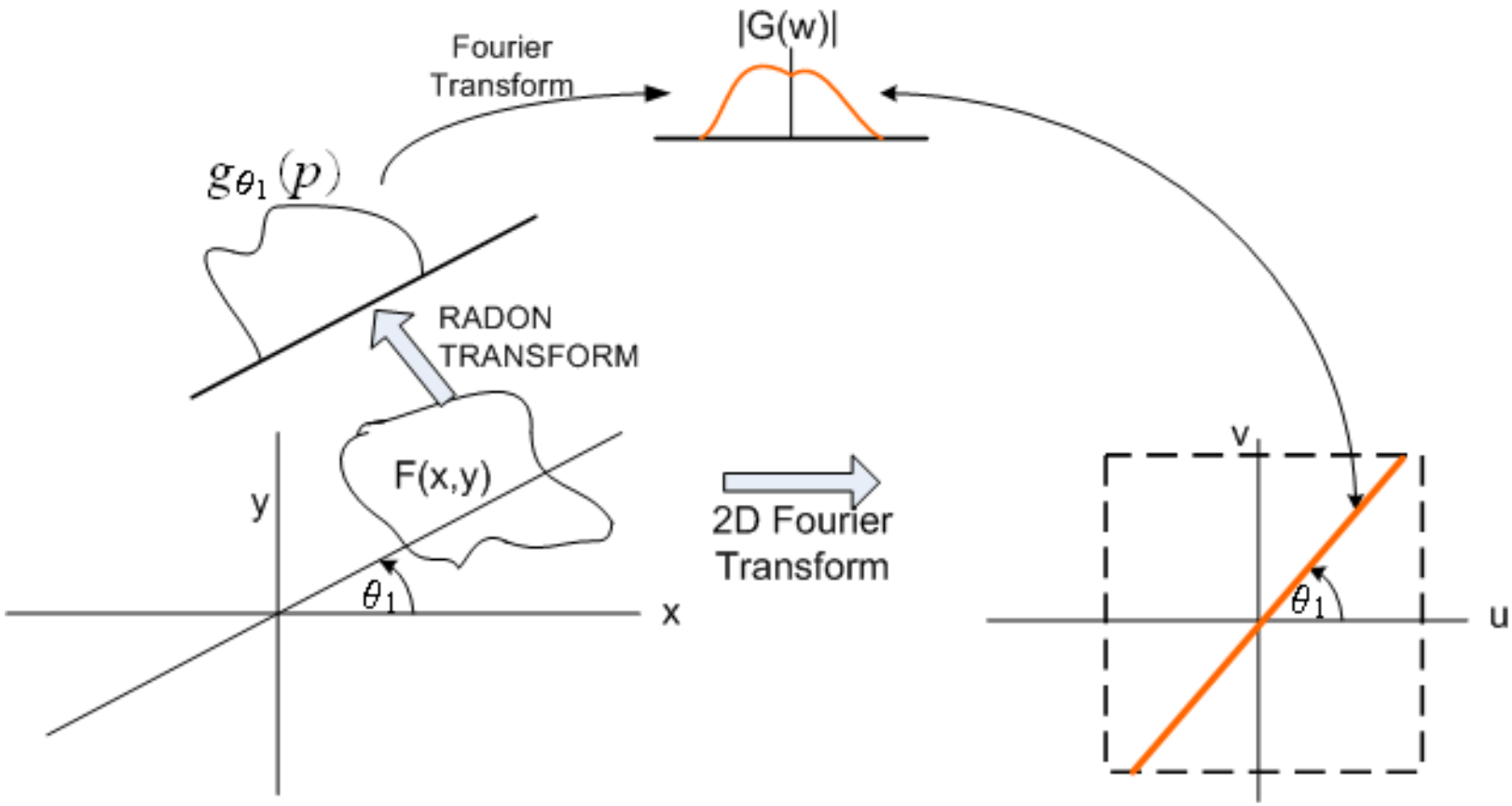


FIGURE 25-17
Filtered backprojection. Filtered backprojection reconstructs an image by filtering each view before backprojection. This removes the blurring seen in simple backprojection, and results in a mathematically exact reconstruction of the image. Filtered backprojection is the most commonly used algorithm for computed tomography systems.

Weighted Back Projection



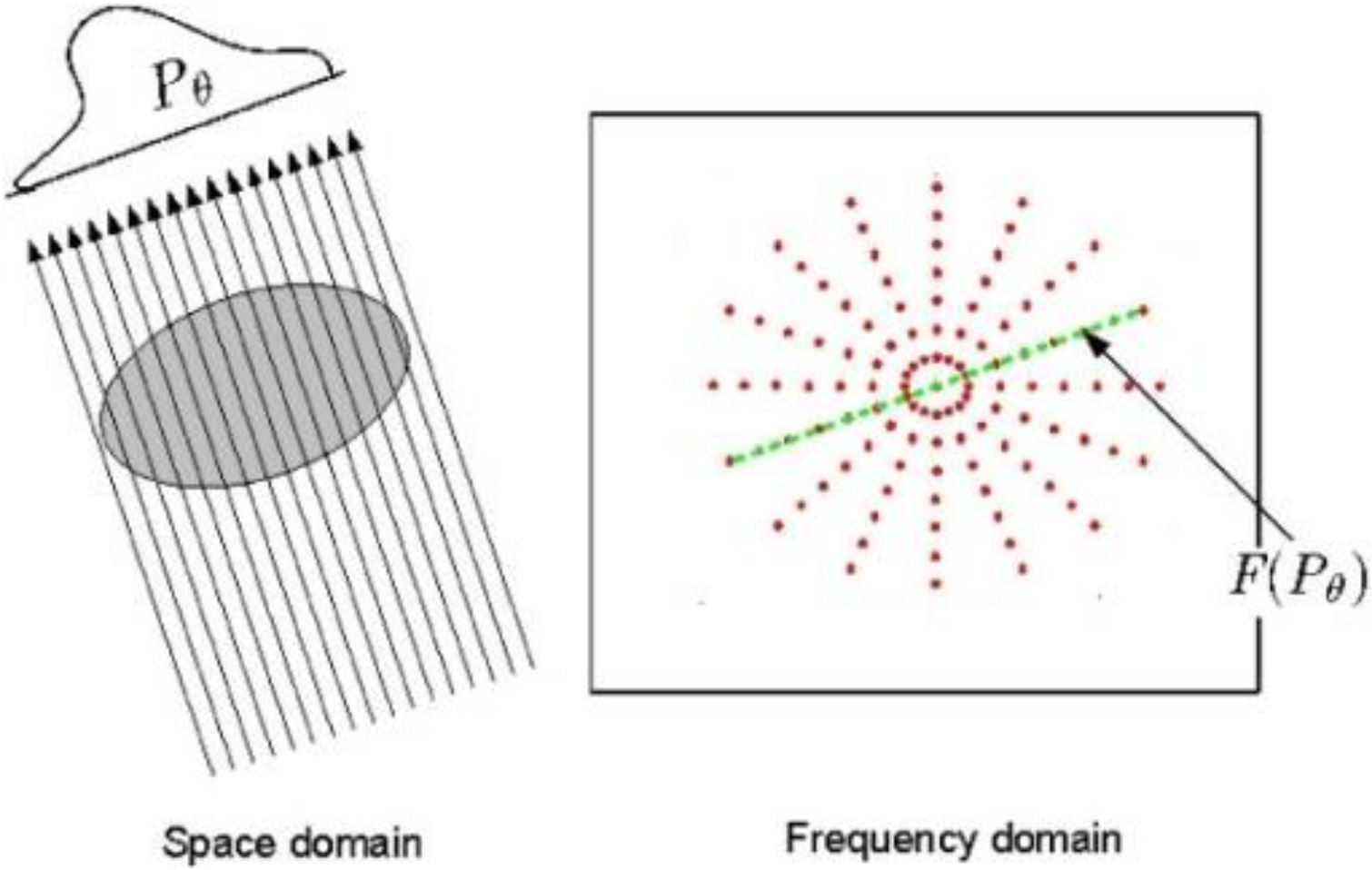
Central Slice Theorem



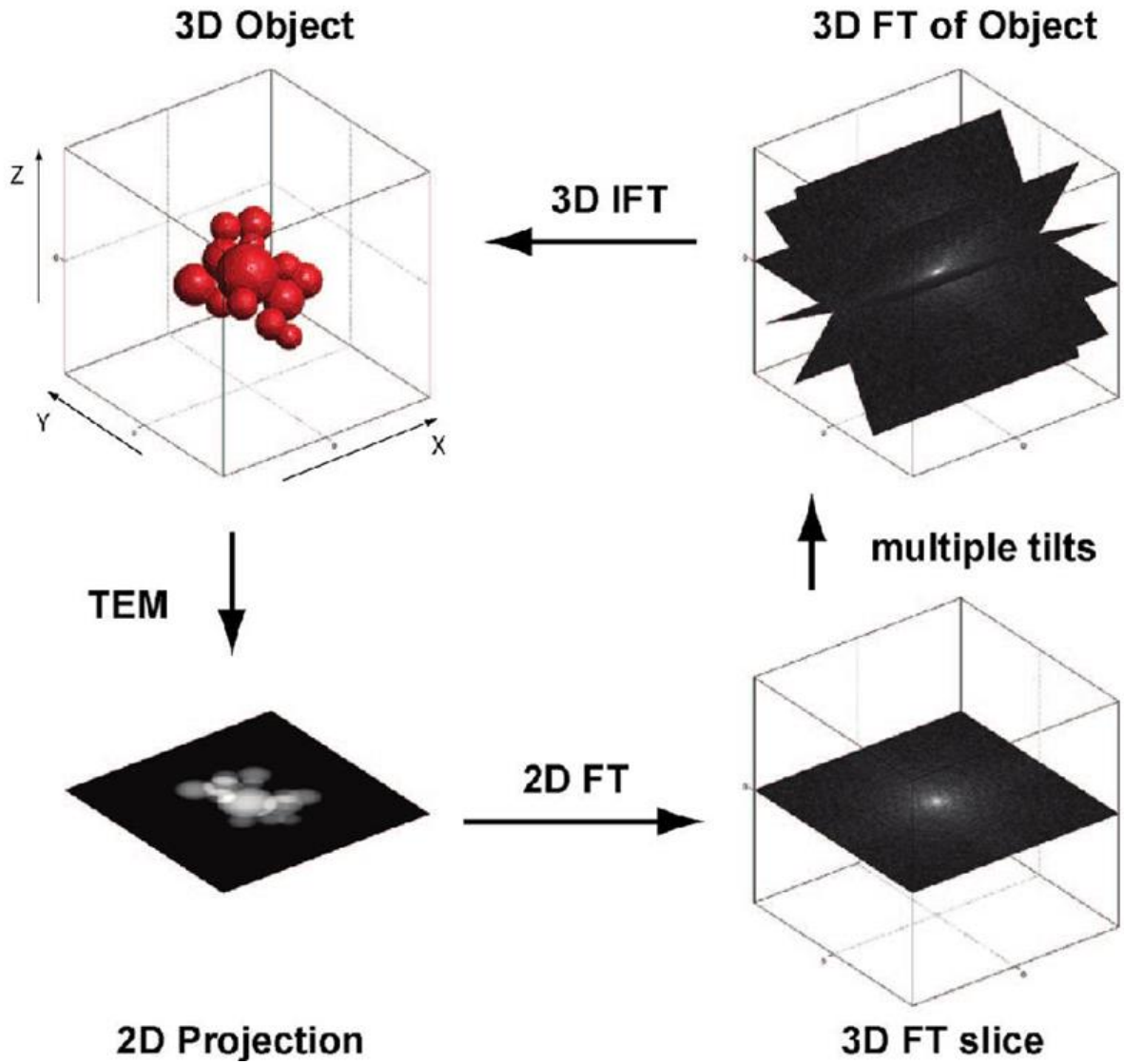
Nasser Abbasi
Nov 9.2008
Cst.vsd

Fourier Central Slice Theorem

Central Slice Theorem

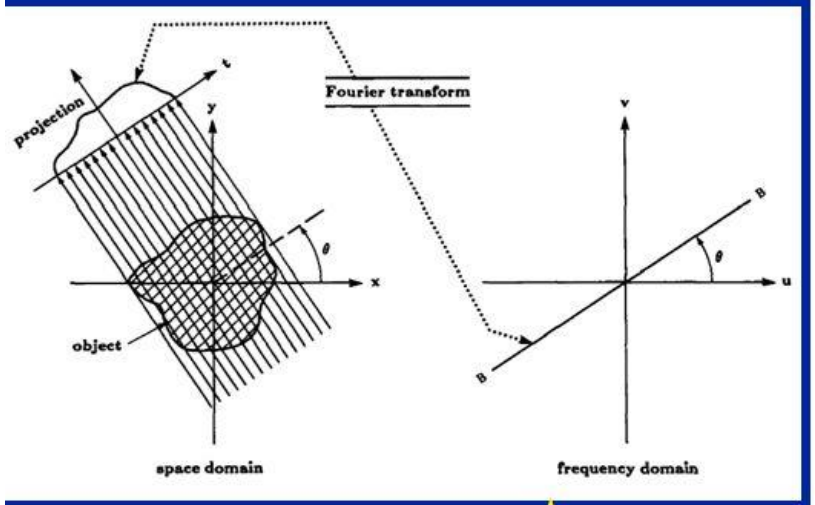


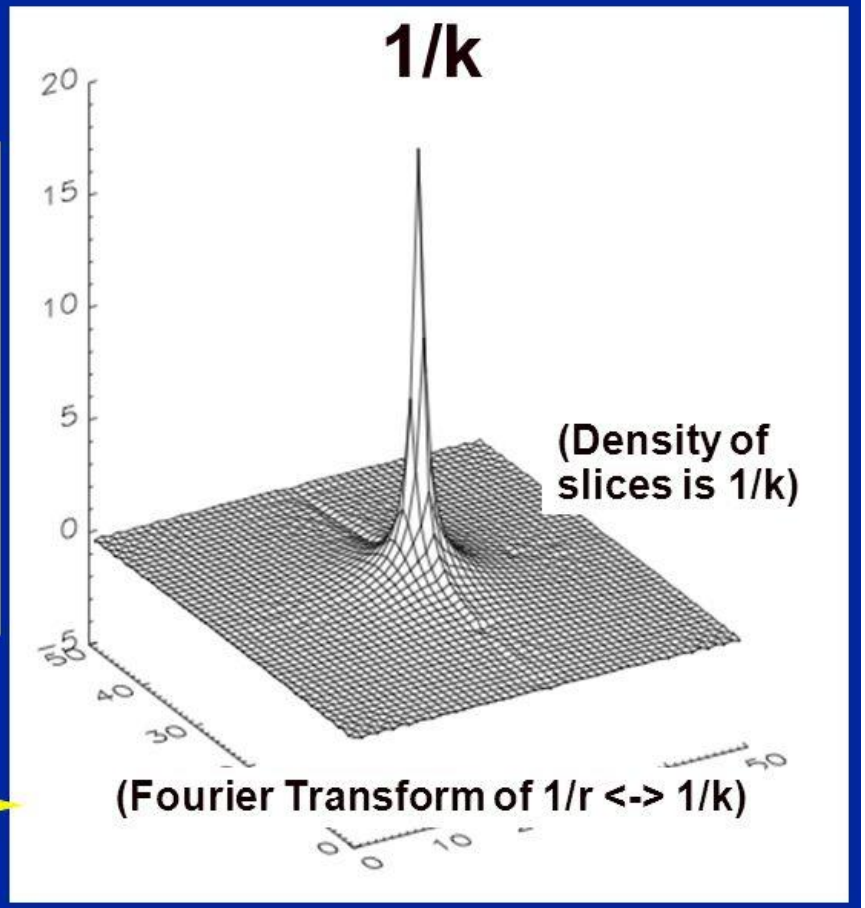
Central Slice Theorem



Weighted Back Projection

Projection -> Backprojection: k-space

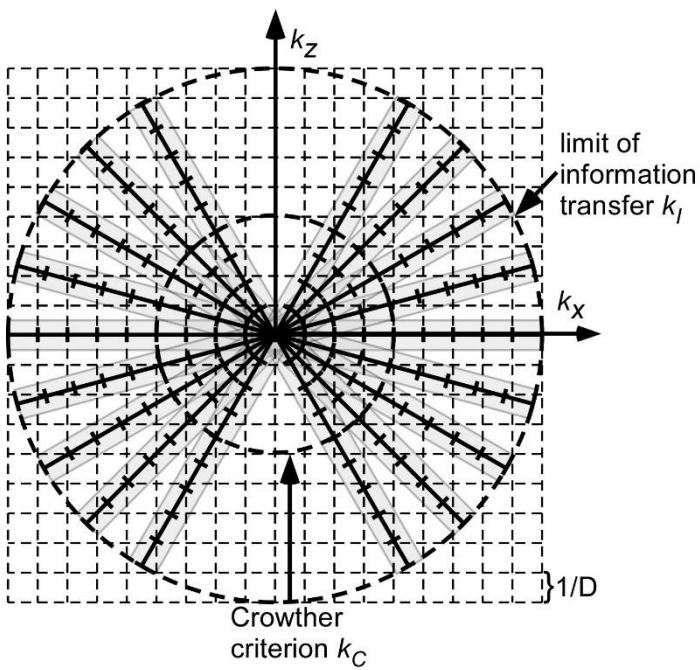


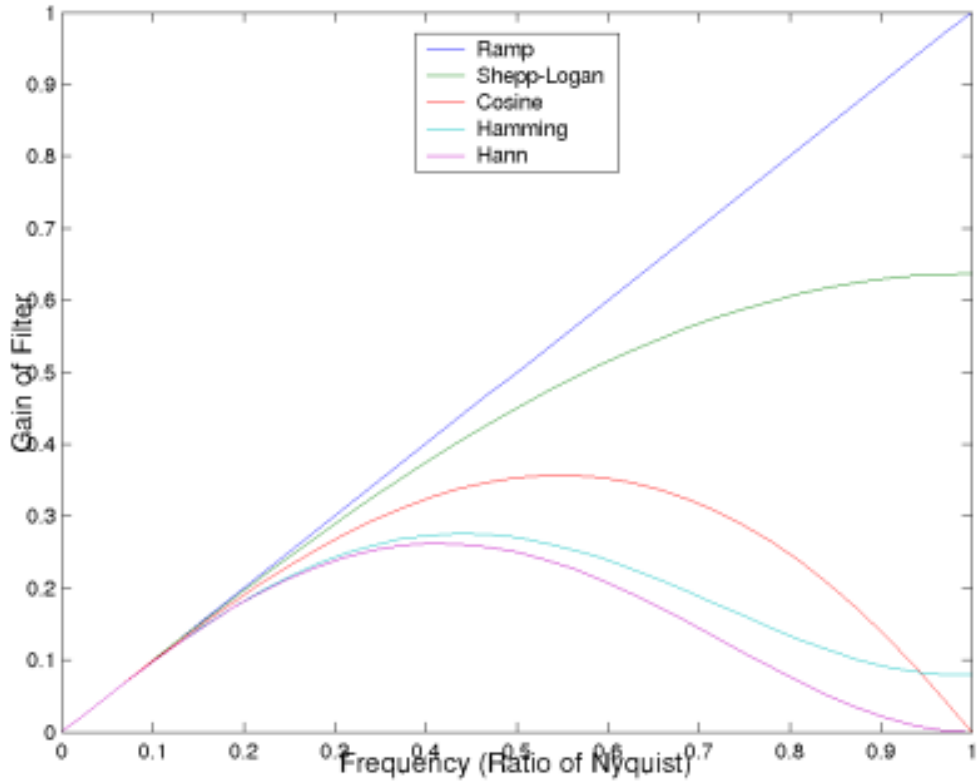


one projection

multiple projections

Weighted Back Projection



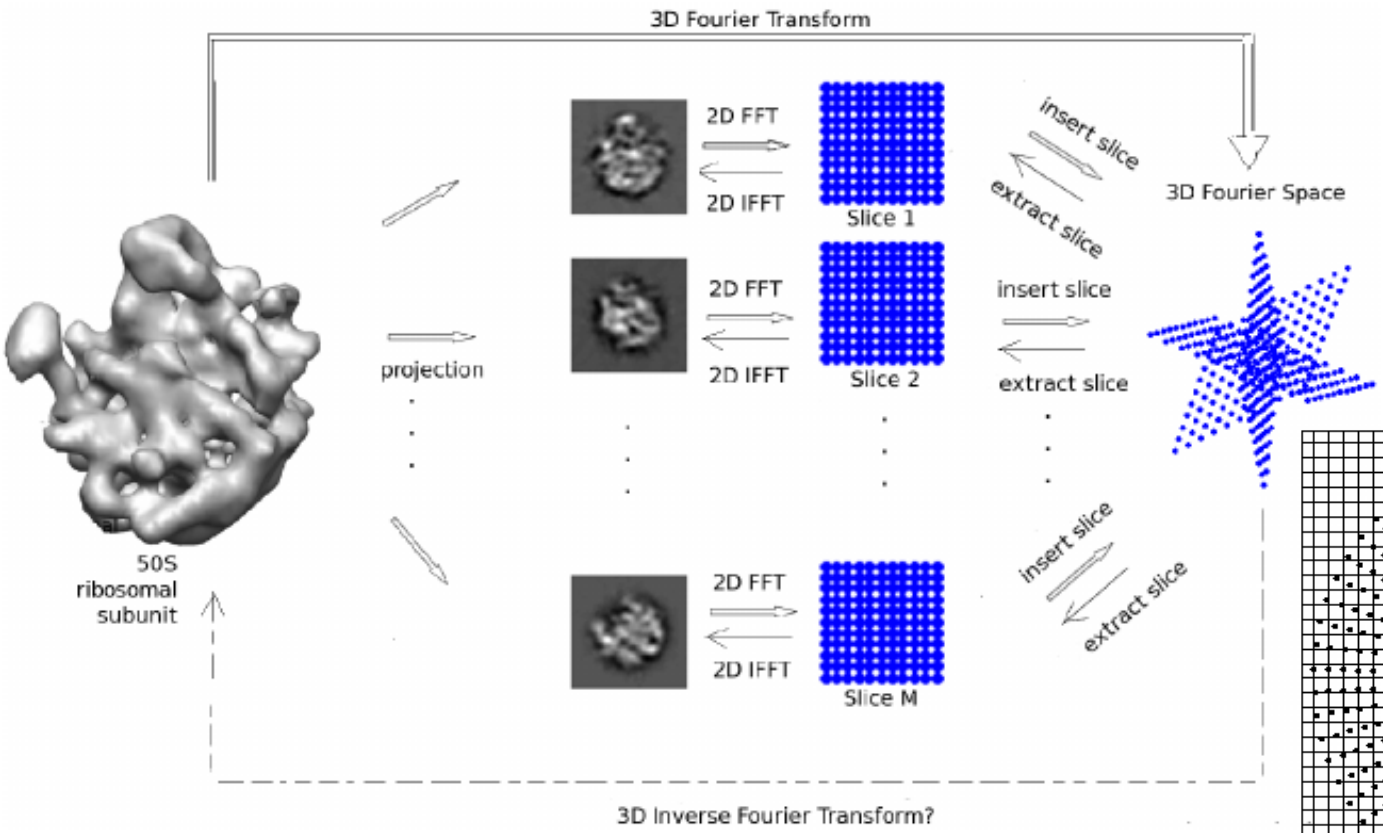


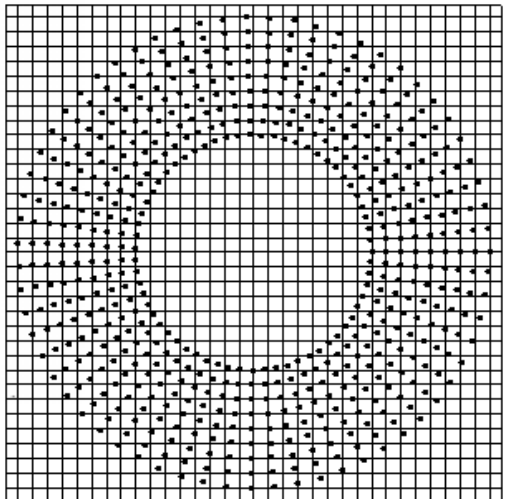
$$W_a(X, Y, Z) = \frac{1}{H(X, Y, Z)}$$

$$= \left\{ \sum_j 2a \operatorname{sinc}[2a\pi(X \sin\theta_j \cos\phi_j + Y \sin\theta_j \sin\phi_j + Z \cos\theta_j)] \right\}^{-1}$$

Radermacher, M. Frank, J. (Ed.) Weighted back-projection methods Electron Tomography, Plenum, 1992, 91-115

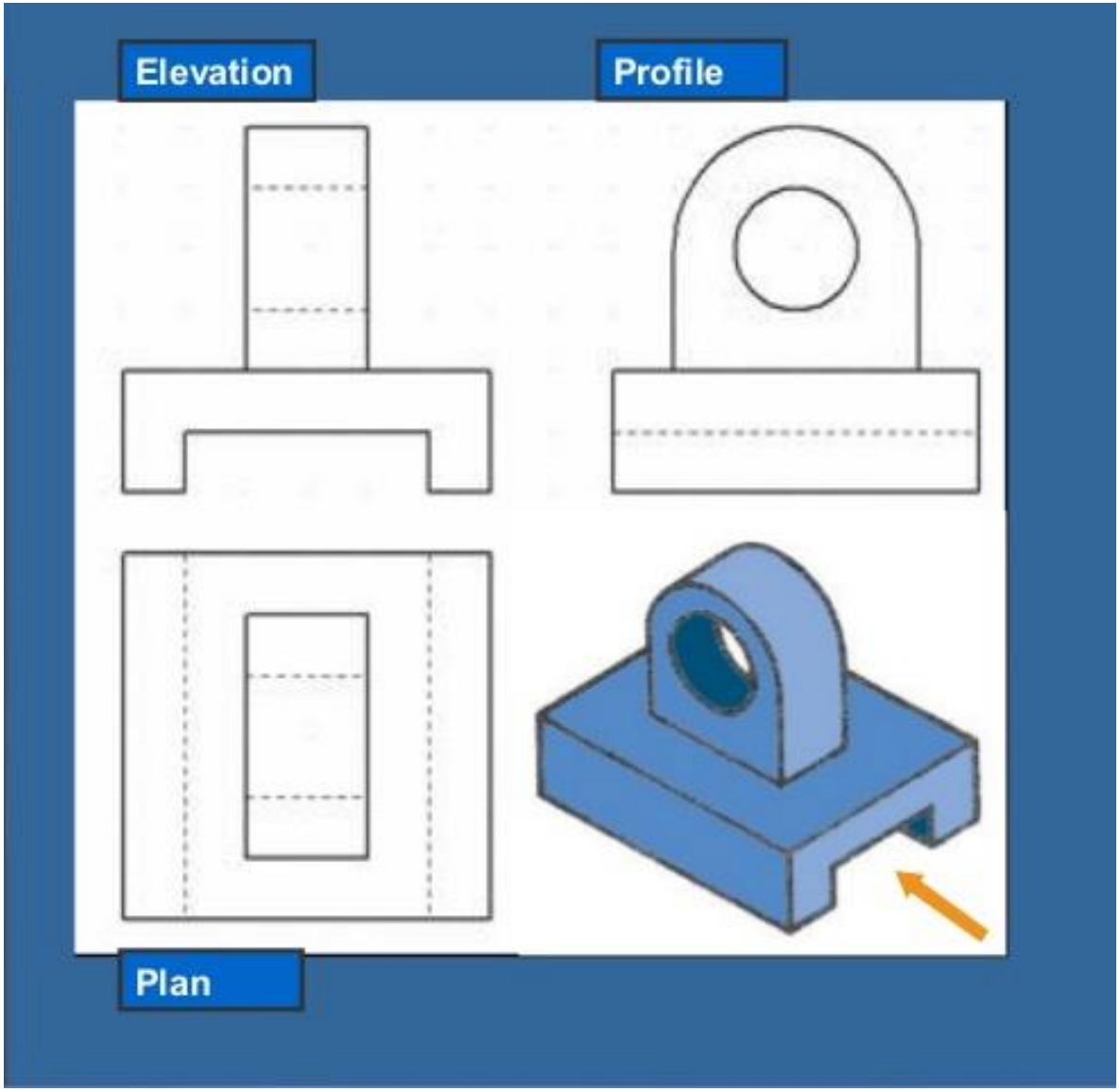
Fourier Reconstruction



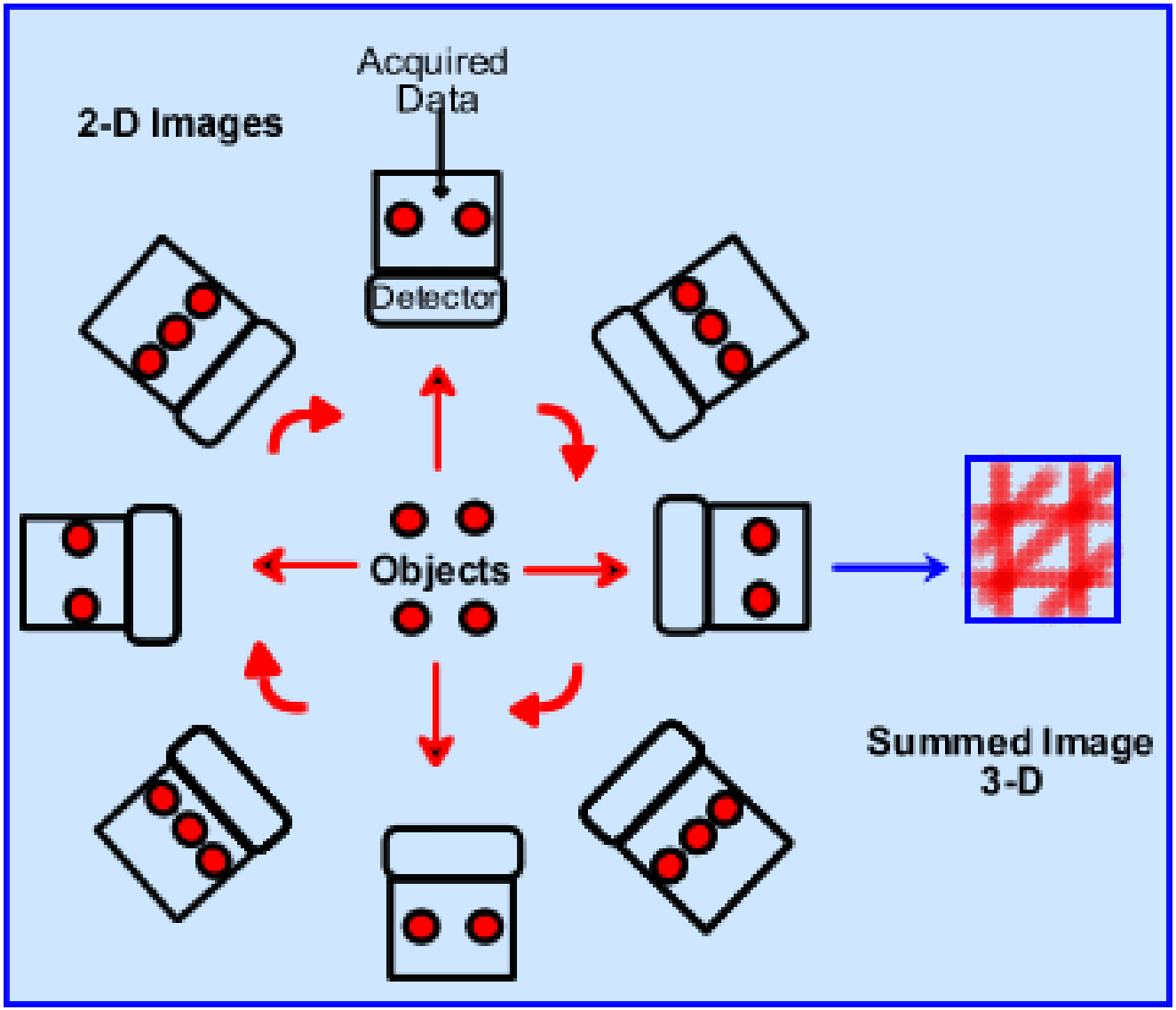


- Penczek, P.; Renka, R. & Schomberg, H. Gridding-based direct Fourier inversion of the three-dimensional ray transform J. Optical Society America A, 2004, 21, 499-509
- Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. J. Structural Biology, 2012, 180, 519-530
- Abrishami, V.; Bilbao-Castro, J. R.; Vargas, J.; Marabini, R.; Carazo, J. M. & Sorzano, C. O. S. A fast iterative convolution weighting approach for gridding-based direct Fourier three-dimensional reconstruction with correction for the contrast transfer function. Ultramicroscopy, 2015, 157, 79-87

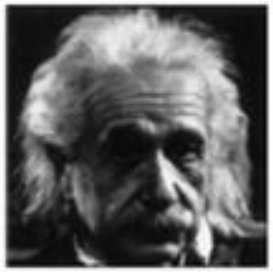
Tomography limitations



Tomography limitations

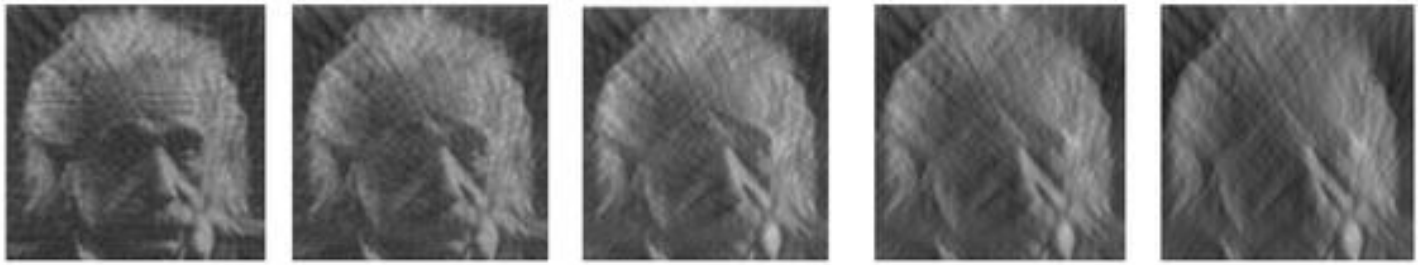


Tomography limitations



original image

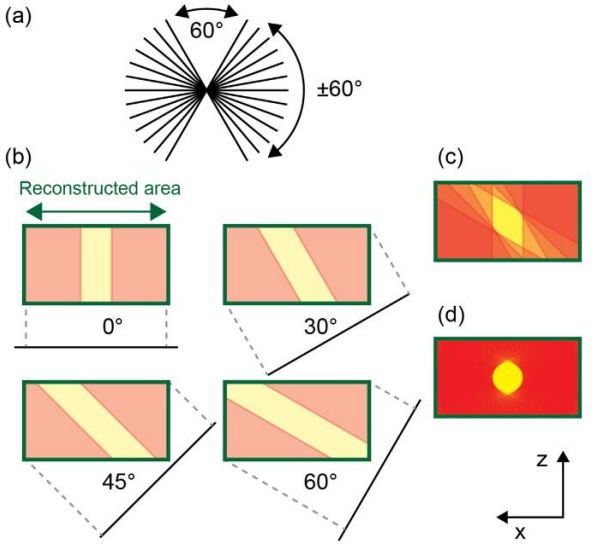
5 deg increment

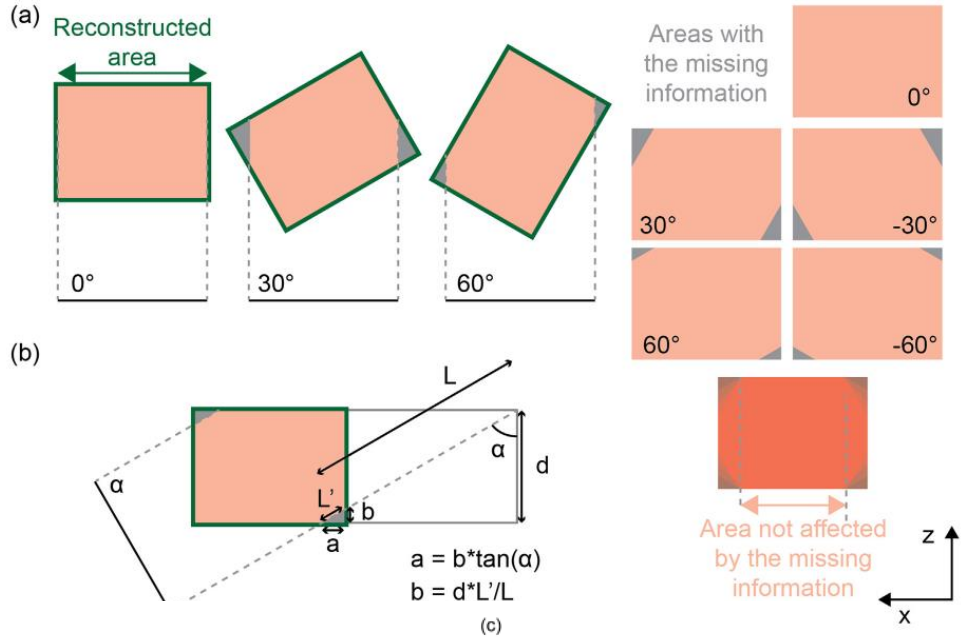


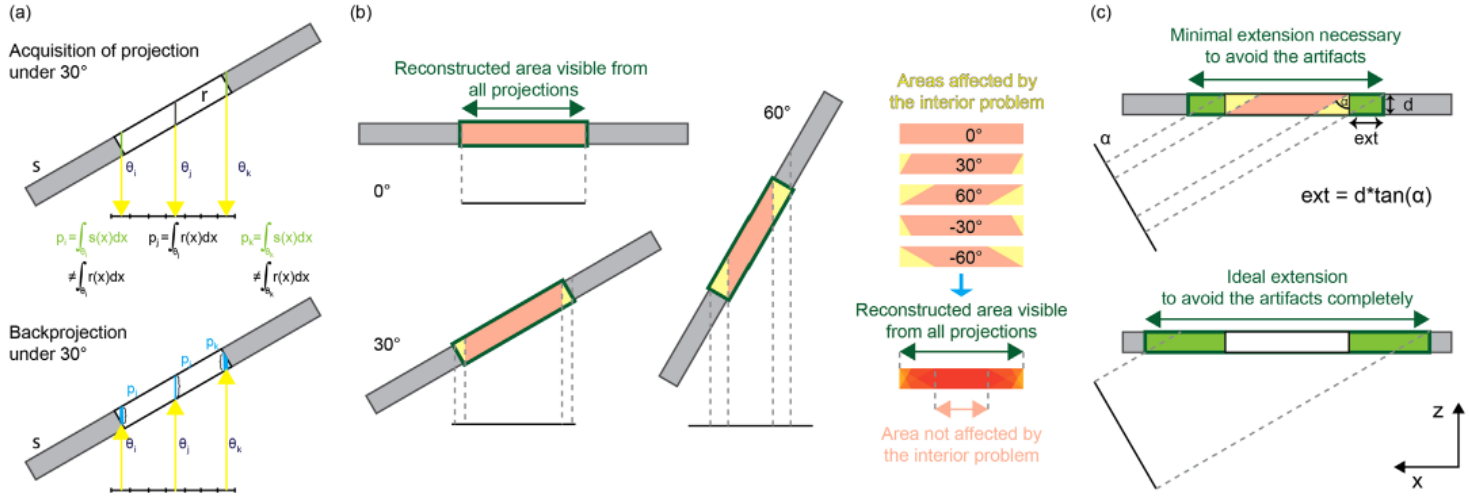
−90 – 90 deg −80 – 80 deg −70 – 70 deg −60 – 60 deg −50 – 50 deg

2 deg increment

Tomography limitations

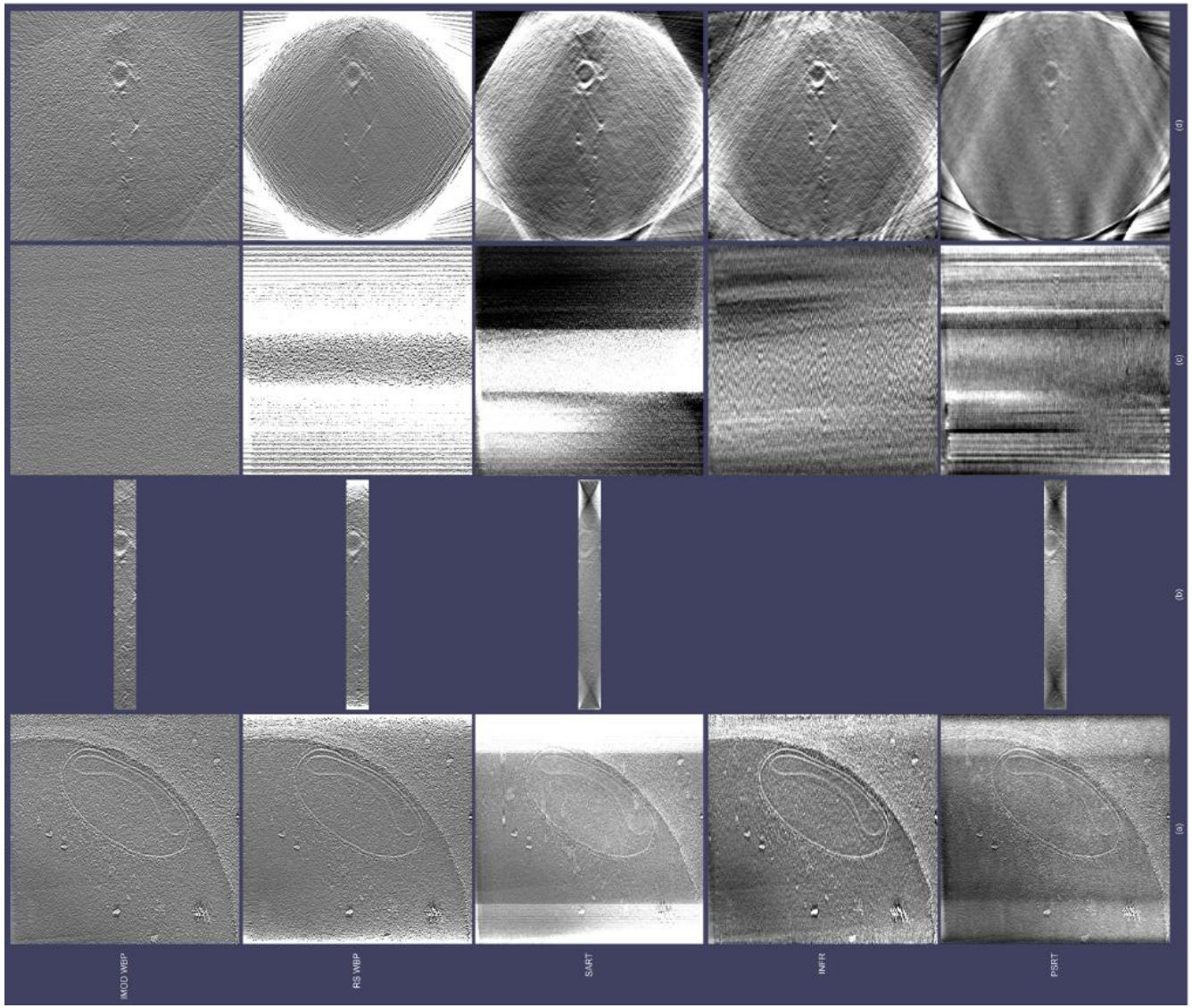






Turoňová, B.; Marsalek, L. & Slusallek, P. On geometric artifacts in cryo electron tomography. Ultramicroscopy, 2016 , 163 , 48-61

Tomography limitations



Turoňová, B.; Marsalek, L. & Slusallek, P. On geometric artifacts in cryo electron tomography. Ultramicroscopy, 2016 , 163 , 48-61

A little bit of algebra

Standard inner product: $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T \mathbf{v} \Rightarrow \|\mathbf{u}\|^2 = \langle \mathbf{u}, \mathbf{u} \rangle$

Noisy measurements: $\mathbf{f} = A\mathbf{x} + \mathbf{n}$

Least squares reconstruction:

$$\begin{aligned} \mathbf{x}^* &= \arg\min \|\mathbf{f} - A\mathbf{x}\|^2 \\ &= \arg\min (\mathbf{f} - A\mathbf{x})^T (\mathbf{f} - A\mathbf{x}) \\ &= \arg\min \mathbf{f}^T \mathbf{f} + \mathbf{x}^T A^T A\mathbf{x} - 2\mathbf{x}^T A^T \mathbf{f} \end{aligned}$$

First order condition:

$$\frac{\partial \|\mathbf{f} - A\mathbf{x}\|^2}{\partial \mathbf{x}} = 0 = 2A^T A\mathbf{x} - 2A^T \mathbf{f}$$

Normal equations:

$$\Rightarrow A^T A\mathbf{x} = A^T \mathbf{f}$$

For complex measurements:

$$A^H A\mathbf{x} = A^H \mathbf{f}$$

CSIC

instruct Integrating Biology

instruct image Processing Center

Some sizes

M images of size PxP
A volume of size PxPxP

$$\mathbf{f} = A\mathbf{x} + \mathbf{n}$$

$$\mathbf{f}, \mathbf{n} \in \mathbb{R}^{MP^2}$$

$$\mathbf{x} \in \mathbb{R}^{P^3}$$

$$A \in \mathbb{R}^{MP^2 \times P^3}$$

$$\begin{pmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \dots \\ \mathbf{f}_M \end{pmatrix} = \begin{pmatrix} A_1 \\ A_2 \\ \dots \\ A_M \end{pmatrix} \mathbf{x} + \begin{pmatrix} \mathbf{n}_1 \\ \mathbf{n}_2 \\ \dots \\ \mathbf{n}_M \end{pmatrix}$$

$$A_i \in \mathbb{R}^{P^2 \times P^3}$$

CSIC

instruct Integrating Biology

instruct image Processing Center

A little bit of algebra

Standard inner product: $\langle \mathbf{u}, \mathbf{v} \rangle_W = \langle \mathbf{u}, W\mathbf{v} \rangle = \langle W\mathbf{u}, \mathbf{v} \rangle \Rightarrow \|\mathbf{u}\|_W^2 = \langle \mathbf{u}, \mathbf{u} \rangle_W$

For any positive definite matrix W
In particular, our W is diagonal

Weighted Least squares

$$\begin{aligned}\mathbf{x}^* &= \arg\min \|\mathbf{f} - A\mathbf{x}\|_W^2 \\ &= \arg\min (\mathbf{f} - A\mathbf{x})^T W (\mathbf{f} - A\mathbf{x}) \\ &= \arg\min \mathbf{f}^T W \mathbf{f} + \mathbf{x}^T A^T W A \mathbf{x} - 2\mathbf{x}^T A^T W \mathbf{f}\end{aligned}$$

First order condition:

$$\frac{\partial \|\mathbf{f} - A\mathbf{x}\|_W^2}{\partial \mathbf{x}} = 0 = 2A^T W A \mathbf{x} - 2A^T W \mathbf{f}$$

Normal equations:

$$\Rightarrow A^T W A \mathbf{x} = A^T W \mathbf{f}$$

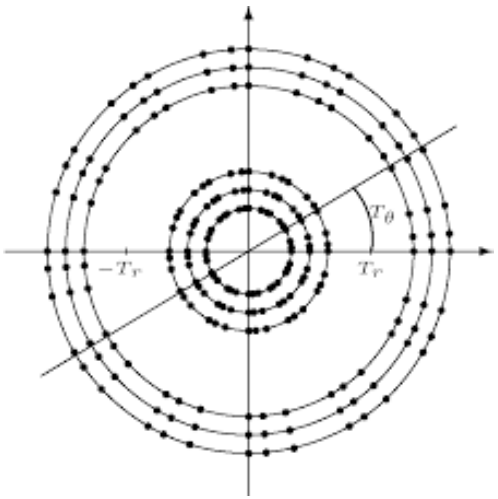
Any solution of the WLS is also a solution of the normal equations and viceversa

CSIC

instruct Integrating Biology

instruct image Processing Center

A little bit of algebra



INFR: Chen, Y. & Förster, F. Iterative reconstruction of cryo-electron tomograms using nonuniform fast Fourier transforms. *J Struct Biol,* **2014**, *185*, 309-316

$$W \in \mathbb{R}^{MP^2 \times MP^2}$$

The residual is W-orthogonal to the space spanned by A.

$$\forall \mathbf{y} \in \mathbb{R}^{P^3} \Rightarrow \langle A\mathbf{x} - \mathbf{f}, A\mathbf{y} \rangle_W = 0 \Leftrightarrow$$

$$\langle A\mathbf{x} - \mathbf{f}, WA\mathbf{y} \rangle = 0 \Leftrightarrow$$

$$\left\langle (WA)^T (A\mathbf{x} - \mathbf{f}), \mathbf{y} \right\rangle = 0 \Leftrightarrow$$

$$(WA)^T (A\mathbf{x} - \mathbf{f}) = \mathbf{0} \Leftrightarrow$$

$$A^T W A \mathbf{x} = A^T W \mathbf{f}$$

CSIC

instruct Integrating Biology

instruct image Processing Center

Direct solution

Normal equations: $A^T W A \mathbf{x} = A^T W \mathbf{f} \rightarrow \tilde{A}\mathbf{x} = \tilde{\mathbf{b}}$

Moore-Penrose pseudoinverse

$$\mathbf{x} = \underbrace{(A^T W A)^{-1}}_{P^3 \times P^3} \underbrace{A^T W \mathbf{f}}_{P^3 \times MP^2} = A_W^{\dagger} \mathbf{f}$$

This is a formidable inversion never performed in practice

Iterative methods

$$\tilde{A}\mathbf{x} = \tilde{\mathbf{b}}$$

Let us write $\tilde{A} = \tilde{M} - \tilde{N}$, then

$$(\tilde{M} - \tilde{N})\mathbf{x} = \tilde{\mathbf{b}}$$

$$\tilde{M}\mathbf{x}^{(k+1)} - \tilde{N}\mathbf{x}^{(k)} = \tilde{\mathbf{b}}$$

$$\mathbf{x}^{(k+1)} = \tilde{M}^{-1}\tilde{N}\mathbf{x}^{(k)} + \tilde{M}^{-1}\tilde{\mathbf{b}} = \mathbf{x}^{(k)} + \tilde{M}^{-1}(\tilde{\mathbf{b}} - \tilde{A}\mathbf{x}^{(k)})$$

<u>Jacobi</u> $\tilde{A} = \tilde{D} - (\tilde{L} + \tilde{U})$

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \tilde{D}^{-1}(\tilde{\mathbf{b}} - \tilde{A}\mathbf{x}^{(k)})$$

<u>Gauss-Seidel</u> $\tilde{A} = (\tilde{D} + \tilde{L}) - \tilde{U}$

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + (\tilde{D} + \tilde{L})^{-1}(\tilde{\mathbf{b}} - \tilde{A}\mathbf{x}^{(k)})$$

<u>Richardson</u> $\tilde{A} = I - (I - \tilde{A})$

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + (\tilde{\mathbf{b}} - \tilde{A}\mathbf{x}^{(k)}) = \tilde{\mathbf{b}} + (I - \tilde{A})\mathbf{x}^{(k)}$$

Iterative methods

Richardson

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + (\tilde{\mathbf{b}} - \tilde{A}\mathbf{x}^{(k)}) = \tilde{\mathbf{b}} + (I - \tilde{A})\mathbf{x}^{(k)}$$

$$\mathbf{x}^{(1)} = \tilde{\mathbf{b}}$$

$$\mathbf{x}^{(2)} = \tilde{\mathbf{b}} + (I - \tilde{A})\tilde{\mathbf{b}}$$

$$\mathbf{x}^{(3)} = \tilde{\mathbf{b}} + (I - \tilde{A})\tilde{\mathbf{b}} + (I - \tilde{A})^2\tilde{\mathbf{b}}$$

$$\ldots$$

$$\mathbf{x}^{(k)} = \left(\sum_{m=0}^{k-1} (I - \tilde{A})^m \right) \tilde{\mathbf{b}}$$

$$\sum_{m=0}^{\infty} (I - \tilde{A})^m = \tilde{A}^{-1}$$

$\lambda_i \in \square$

Converges iff $\forall i \; |1 - \lambda_i| < 1 \Rightarrow 0 < \lambda_i < 2$

CSIC

instruct Integrating Biology

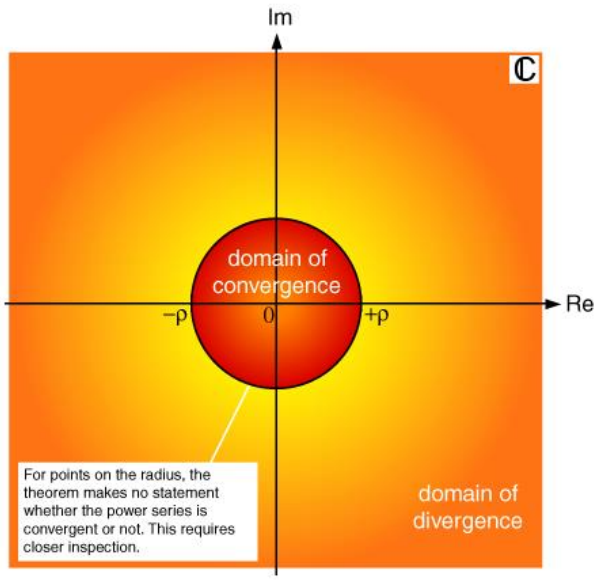
instruct image Processing Center

Iterative methods: Relaxation parameters

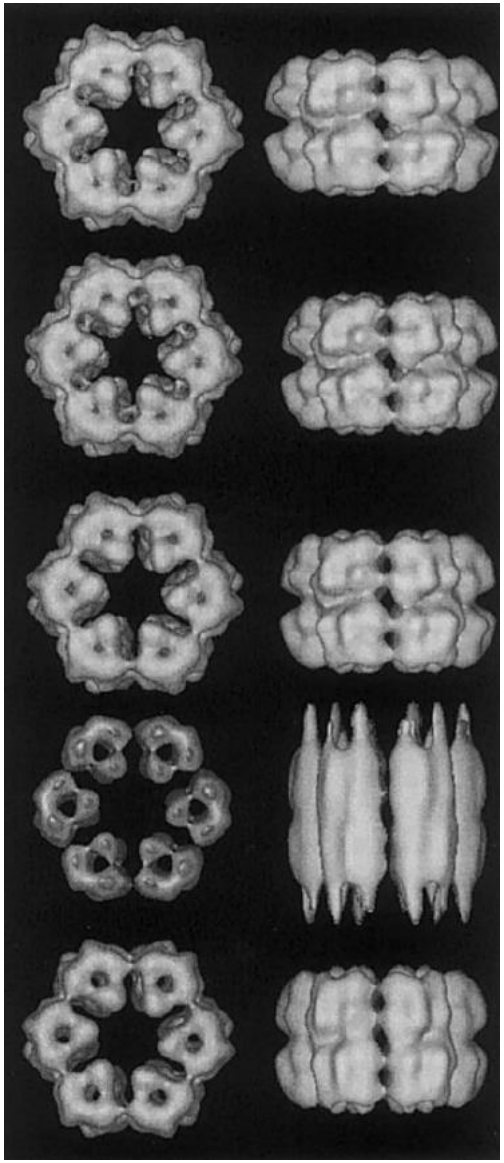
Richardson

$$\mathbf{x}^{(k+1)} = \tilde{\mathbf{b}} + \eta(I - \tilde{A})\mathbf{x}^{(k)}$$

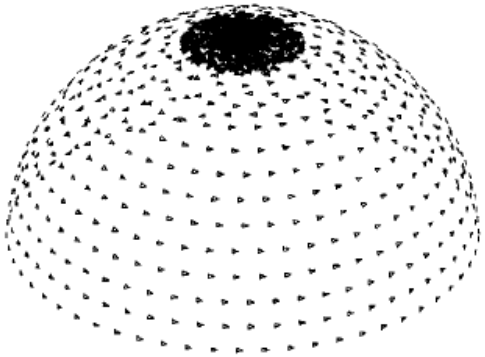
Relaxation factors help to increase the radius of convergence and the convergence speed



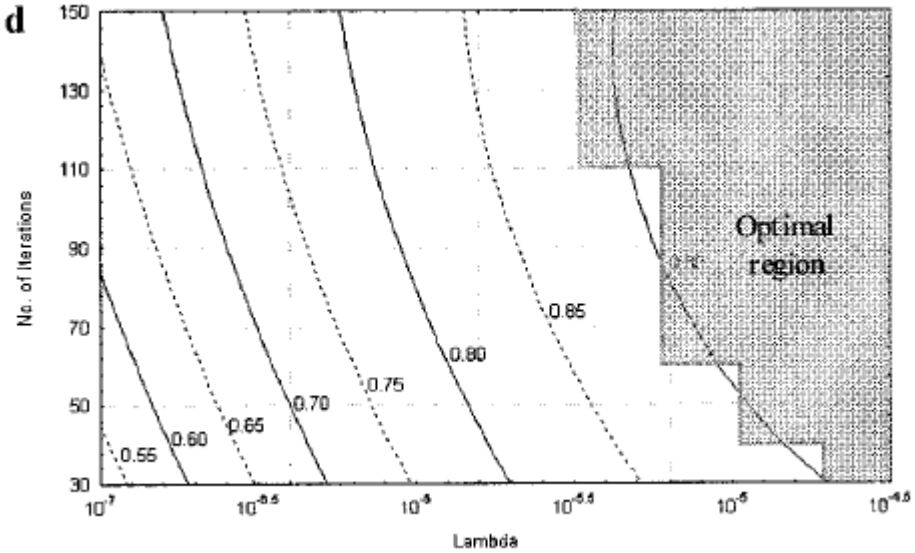
Choice of the relaxation parameter



ART optimal



SIRT optimal



WBP

SIRT non-optimal
“few” iterations

SIRT non-optimal
“many” iterations

C.O.S.Sorzano, R.Marabini, N.Boisset, E.Rietzel, R.Schröder, G.T.Herman, J.M.Carazo. The effect of overabundant projection directions on 3D reconstruction algorithms. Journal of Structural Biology, 133 (2,3), 108-118 (2001)

Iterative methods

$$A^T W A \mathbf{x} = A^T W \mathbf{f} \rightarrow \tilde{A}\mathbf{x} = \tilde{\mathbf{b}}$$

<u>SIRT is Jacobi</u>

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \eta \tilde{D}^{-1}(\tilde{\mathbf{b}} - \tilde{A}\mathbf{x}^{(k)})$$

$$A = \begin{pmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \cdots \\ \mathbf{a}_{MP^2}^T \end{pmatrix} \Rightarrow \begin{cases} \tilde{A} = \sum_{i=1}^{MP^2} w_i \mathbf{a}_i \mathbf{a}_i^T & \in \mathbb{R}^{P^3 \times P^3} \\ \tilde{\mathbf{b}} = \sum_{i=1}^{MP^2} w_i f_i \mathbf{a}_i & \in \mathbb{R}^{P^3} \end{cases}$$

$$A \in \mathbb{R}^{MP^2 \times P^3}$$

CSIC

instruct Integrating Biology

instruct image Processing Center

Iterative methods

$$A^T W A \mathbf{x} = A^T W \mathbf{f} \rightarrow \tilde{A}\mathbf{x} = \tilde{\mathbf{b}}$$

<u>SIRT is Jacobi</u>

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \eta \tilde{D}^{-1}(\tilde{\mathbf{b}} - \tilde{A}\mathbf{x}^{(k)})$$

$$\tilde{A} = \sum_{i=1}^{MP^2} w_i \mathbf{a}_i \mathbf{a}_i^T$$

$$\tilde{\mathbf{b}} = \sum_{i=1}^{MP^2} w_i f_i \mathbf{a}_i$$

$$A = \begin{pmatrix} \boldsymbol{\alpha}_1 & \boldsymbol{\alpha}_2 & \dots & \boldsymbol{\alpha}_{P^3} \end{pmatrix}$$

$$\Rightarrow \begin{cases} \tilde{d}_{jj} = \sum_{i=1}^{MP^2} w_i a_{ij}^2 = \left\| \boldsymbol{\alpha}_j \right\|_W^2 \\ (\tilde{\mathbf{b}} - \tilde{A}\mathbf{x}^{(k)})_j = \sum_{i=1}^{MP^2} \left(w_i f_i a_{ij} - \sum_{m=1}^{P^3} w_i a_{ij} a_{im} x_m^{(k)} \right) \end{cases}$$

$$= \left\langle \boldsymbol{\alpha}_j, \mathbf{f} \right\rangle_W - \left\langle \boldsymbol{\alpha}_j, A\mathbf{x}^{(k)} \right\rangle_W$$

$$= \left\langle \boldsymbol{\alpha}_j, \mathbf{f} - A\mathbf{x}^{(k)} \right\rangle_W$$

$$\boxed{x_j^{(k+1)} = x_j^{(k)} + \eta \frac{\left\langle \boldsymbol{\alpha}_j, \mathbf{f} - A\mathbf{x}^{(k)} \right\rangle_W}{\left\| \boldsymbol{\alpha}_j \right\|_W^2}}$$

CSIC

instruct Integrating Biology

instruct image Processing Center

Iterative methods

SIRT is Jacobi

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \eta \sum_{i=1}^{MP^2} \frac{\left\langle \mathbf{f} - A\mathbf{x}^{(k)}, \mathbf{a}_i \right\rangle_W}{\left\| \mathbf{a}_i \right\|_W^2} \mathbf{a}_i$$

$$= \mathbf{x}^{(k)} + \eta \tilde{D}^{-1} A^T W (\mathbf{f} - A\mathbf{x}^{(k)})$$

SIRT is related to gradient descent (Landweber)

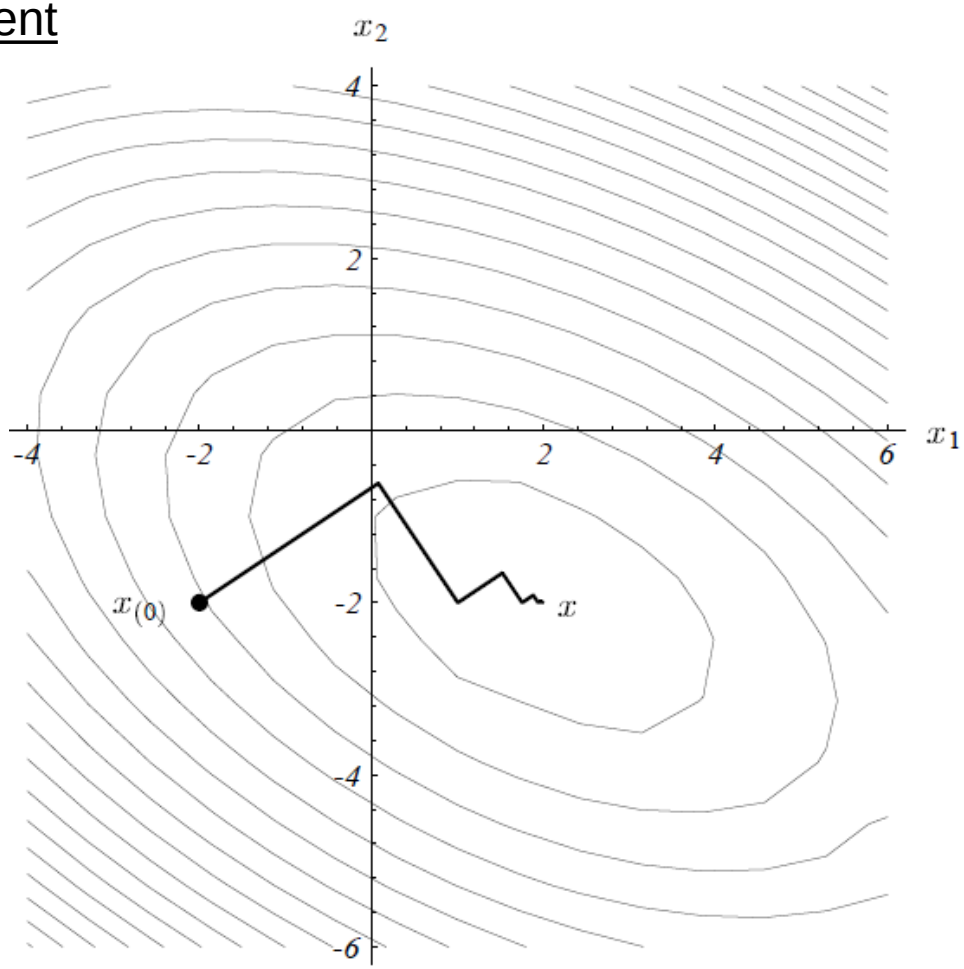
Weighted Least squares $\quad \mathbf{x}^* = \arg\min E(\mathbf{x}) = \left\| \mathbf{f} - A\mathbf{x} \right\|_W^2$

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - r_k \nabla E(\mathbf{x}^{(k)})$$

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + 2 r_k A^T W (\mathbf{f} - A\mathbf{x}^{(k)})$$

Iterative methods

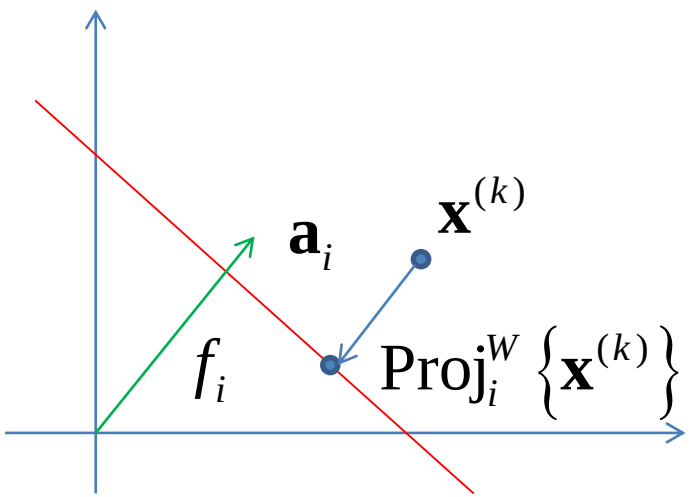
Gradient descent



Geometric interpretation

$$H_i = \left\{ \mathbf{x} \in \mathbb{R}^{P^3} \,\middle|\, \langle \mathbf{a}_i, \mathbf{x} \rangle_W = f_i \right\}$$

$$\mathrm{Proj}_i^W \left\{ \mathbf{x}^{(k)} \right\} = \arg\min_{\mathbf{x} \in H_i} \left\| \mathbf{x} - \mathbf{x}^{(k)} \right\|_W = \mathbf{x}^{(k)} + \frac{f_i - \langle \mathbf{x}^{(k)}, \mathbf{a}_i \rangle}{\left\| \mathbf{a}_i \right\|_W^2} W^{-1} \mathbf{a}_i$$



- ART
- SIRT
- SART
- Cimmino
- CAV
- BiCAV
- Diagonal weighting

Censor, Y. & Elfving, T. Block-iterative algorithms with diagonally scaled oblique projections for the linear feasibility problem. *SIAM J. Matrix Analysis and Applications,* **2002**

Iterative methods

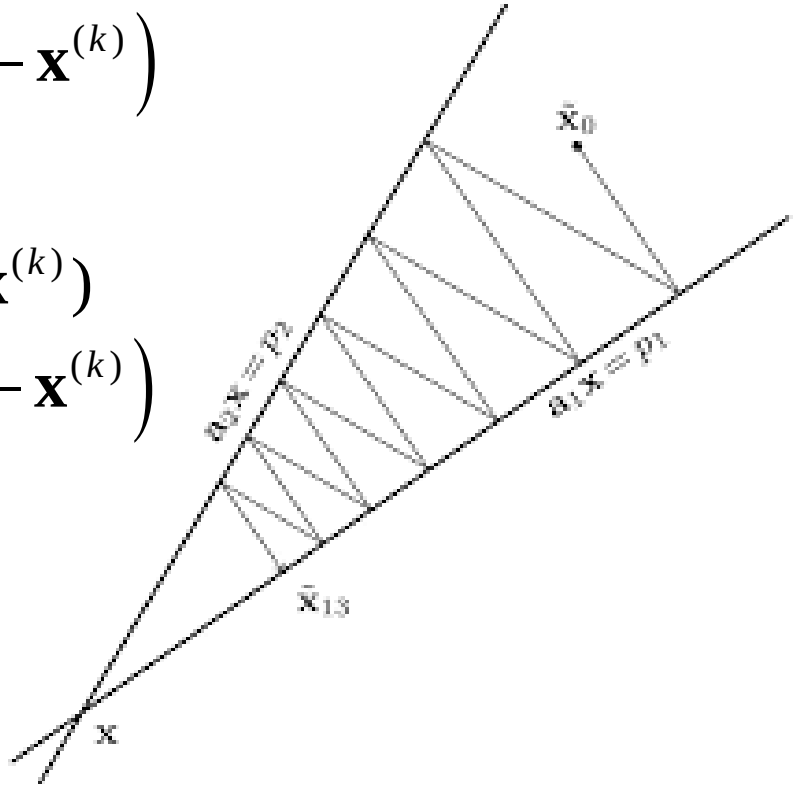
ART is Gauss-Seidel

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \eta\left(\mathrm{Proj}_{i(k)}\left\{\mathbf{x}^{(k)}\right\} - \mathbf{x}^{(k)}\right)$$

SIRT is an average of all hyperplane projections

$$\begin{aligned}\mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} + \eta \tilde{D}_{MP^2}^{-1} A^T W(\mathbf{f} - A\mathbf{x}^{(k)}) \\ &= \mathbf{x}^{(k)} + \eta \sum_{i=1}^{} \left(\mathrm{Proj}_i\left\{\mathbf{x}^{(k)}\right\} - \mathbf{x}^{(k)}\right)\end{aligned}$$

Block-ART (SART) is in between



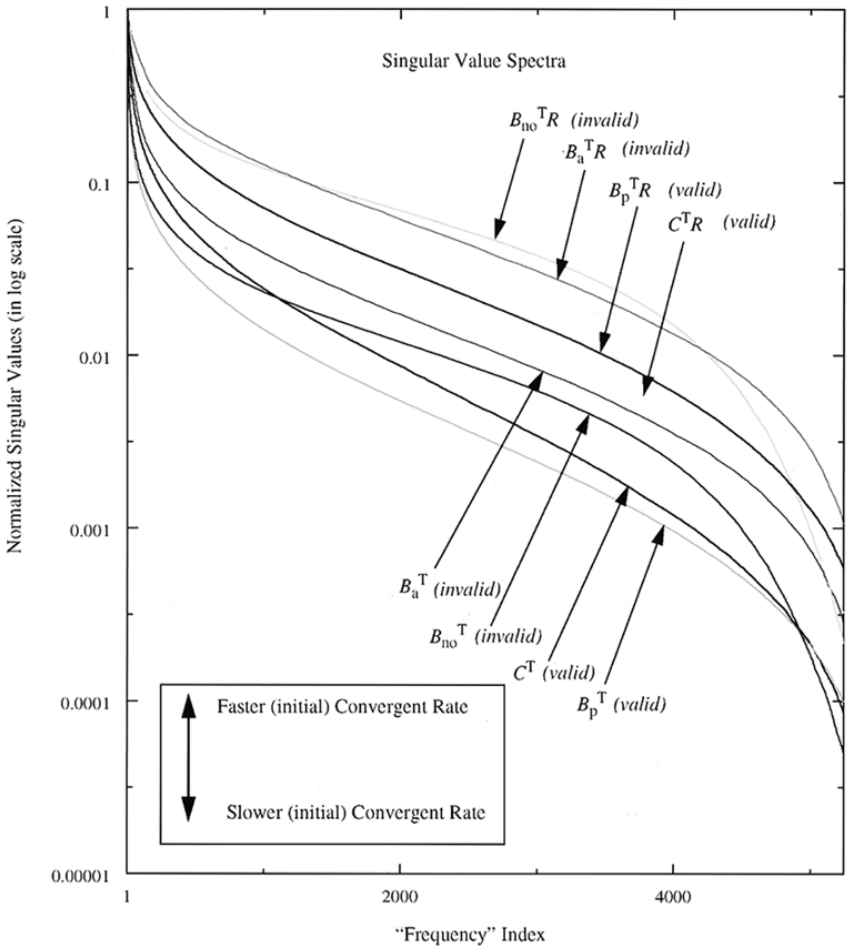
Unmatched projector/backprojector

Matched

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + 2r_k A^T W(\mathbf{f} - A\mathbf{x}^{(k)})$$

Unmatched

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + 2r_k B^T W(\mathbf{f} - A\mathbf{x}^{(k)})$$



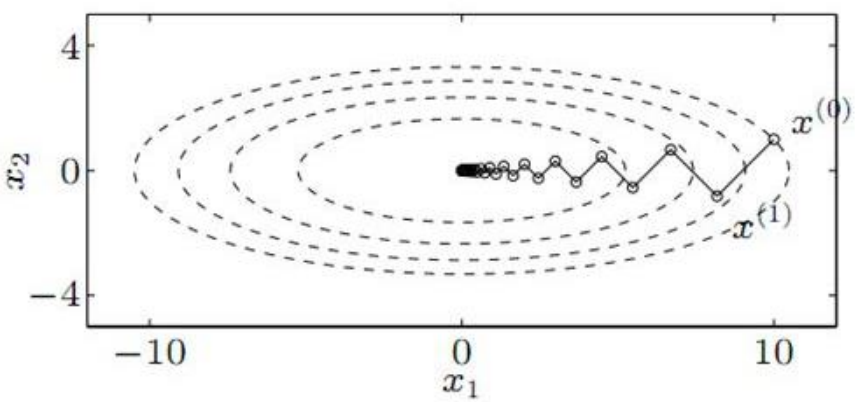
Zeng, G. L. & Gullberg, G. T. Unmatched projector/backprojector pairs in an iterative reconstruction algorithm. *IEEE Trans. Medical Imaging,* **2000**, *19*, 548-555

Speed of convergence and conditioning number

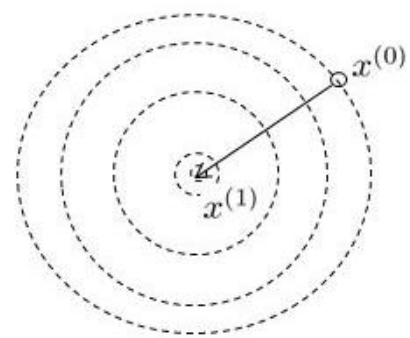
$$\mathbf{x}^* = \arg\min \|\mathbf{f} - A\mathbf{x}\|^2 \longrightarrow A^T A\mathbf{x} = A^T\mathbf{f}$$

$$\kappa(A) = \frac{\sigma_{\max}}{\sigma_{\min}} \longrightarrow \kappa(A^T A) = \kappa^2(A)$$

Singular value $\sigma_i(A) = Eig_i\{A^T A\}$



Condition number = 10



Condition number = 1

CSIC

instruct Integrating Biology

instruct image Processing Center

Conjugate gradient

Consider the system $\tilde{A}\mathbf{x} = \tilde{\mathbf{b}}$

Two vectors are conjugate with respect to $\tilde{A}$ iff $\langle \mathbf{u}, \mathbf{v} \rangle_{\tilde{A}} = 0$

We will construct the solution as $\mathbf{x}^{(k)} = \sum_{j=1}^{k} \frac{\langle \mathbf{p}_j, \tilde{\mathbf{b}} \rangle}{\|\mathbf{p}_j\|_{\tilde{A}}^2} \mathbf{p}_j$

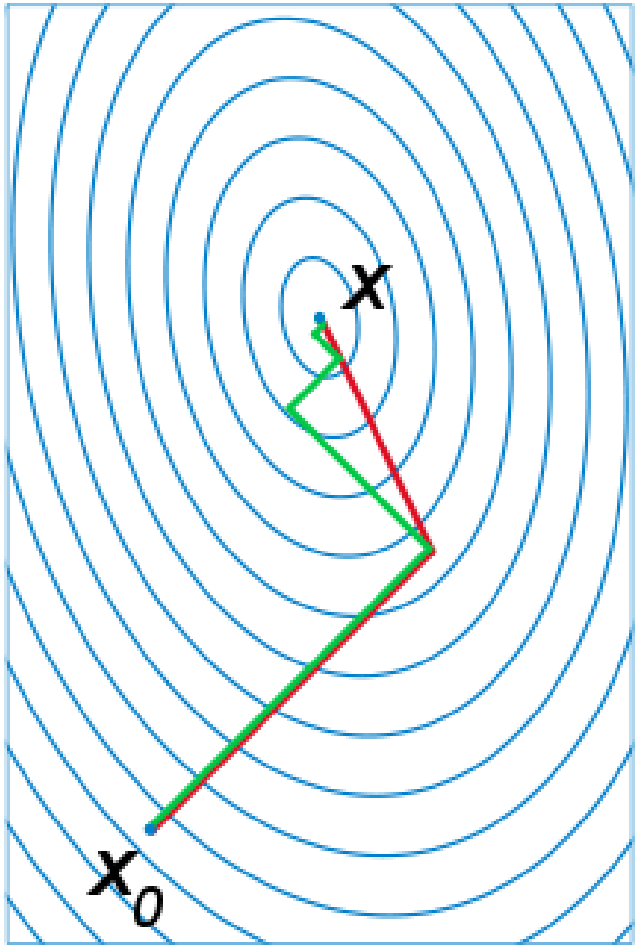
where the set $\{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{P^3}\}$ are conjugate directions with respect to $\tilde{A}$

This set is constructed iteratively as

$$\mathbf{p}_k = \left(\tilde{\mathbf{b}} - A\mathbf{x}^{(k)}\right) - \sum_{j=1}^{k-1} \frac{\langle \mathbf{p}_j, \tilde{\mathbf{b}} - A\mathbf{x}^{(k)} \rangle_{\tilde{A}}}{\|\mathbf{p}_j\|_{\tilde{A}}^2} \mathbf{p}_j$$

Conjugate gradient converges in at most P^3 steps.

Conjugate gradient



$r_0 = b - Ax_0;\quad p_0 = r_0$ initialization

FOR $k = 0, 1, \cdots,$ DO

$\alpha_k = \frac{r_k^T r_k}{p_k^T A p_k}$

$x_{k+1} = x_k + \alpha_k p_k$ update iterate

$r_{k+1} = r_k - \alpha_k A p_k$ update residual

$\beta_k = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$

$p_{k+1} = r_{k+1} + \beta_k p_k$ update direction vector

END FOR

INFR: Chen, Y. & Förster, F. Iterative reconstruction of cryo-electron tomograms using nonuniform fast Fourier transforms. *J Struct Biol,* **2014***, 185*, 309-316

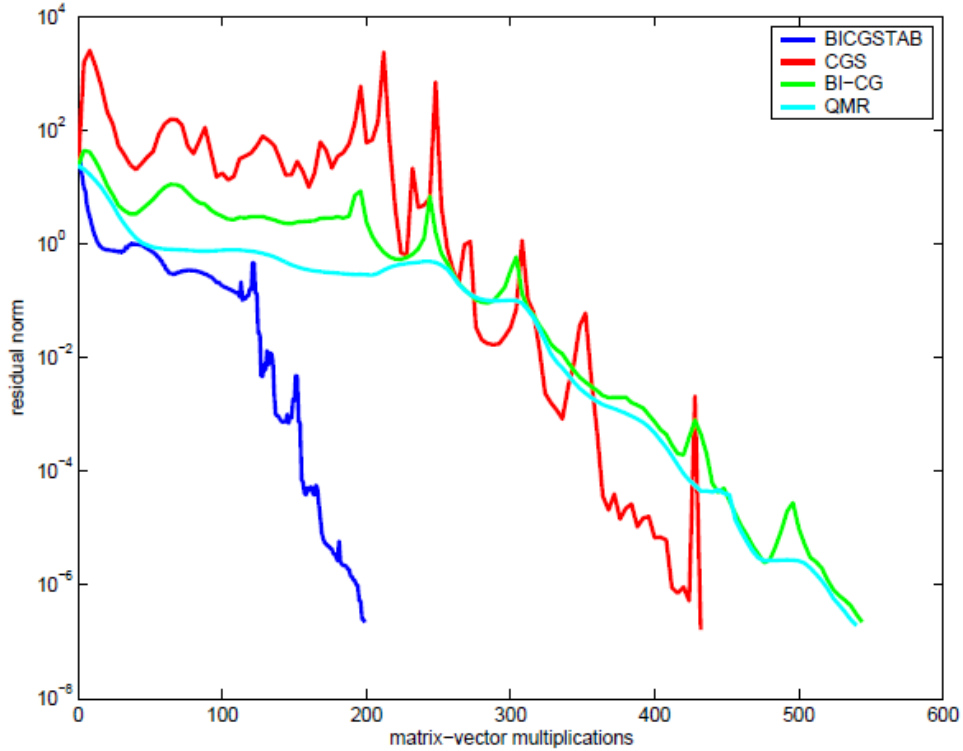
CSIC

instruct Integrating Biology

instruct image Processing Center

Conjugate gradient variants

- Conjugate residuals
- Biconjugate gradient
- Biconjugate gradient stabilizes
- Lanczos method
- Generalized Minimal Residuals (GMINRES)
- Bi-Lanczos
- Conjugate Gradient Squared
- Quasi Minimal Residuals



CSIC

instruct Integrating Biology

instruct image Processing Center

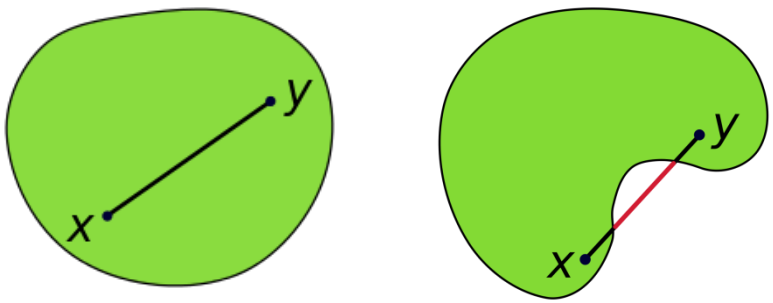
Constrained optimization

$$\mathbf{x}^* = \arg\min \|\mathbf{f} - A\mathbf{x}\|^2$$
$$s.t.\ \mathbf{x} \in C$$

being C is a convex set

$$\forall \mathbf{x}, \mathbf{y} \in C, \forall \alpha \in [0,1] \Rightarrow (1-\alpha)\mathbf{x} + \alpha\mathbf{y} \in C$$

- Non-negativity
- Symmetry
- Inside mask



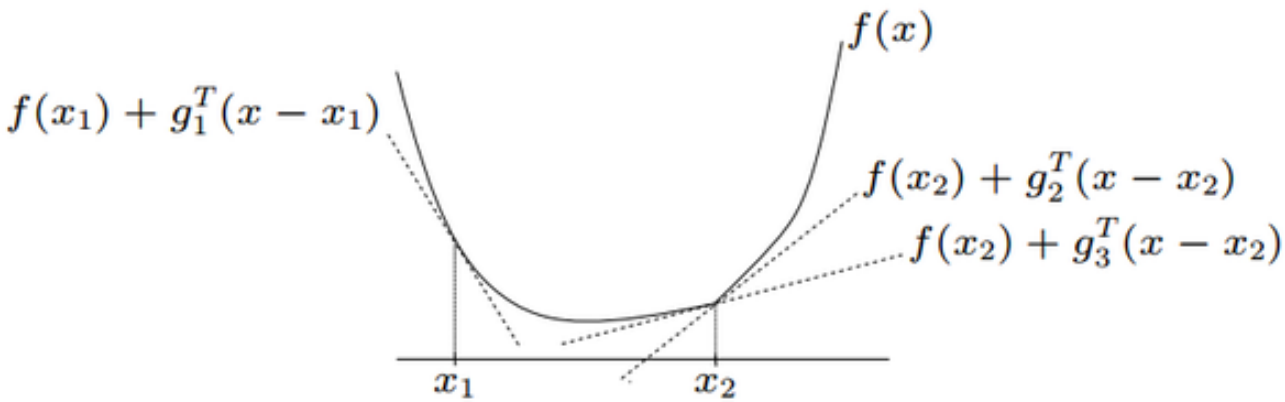
$$\mathbf{x}^{(k+1)} = \mathrm{Proj}_C \left\{ \mathbf{x}^{(k)} + \tilde{D}^{-1} A^T W (\mathbf{f} - A\mathbf{x}^{(k)}) \right\}$$

Subgradient descent

$$\mathbf{x}^* = \arg\min \phi(\mathbf{x})$$

Let Φ be a convex function. **g** is a subgradient of Φ at $\mathbf{x}_0$ iff

$$\phi(\mathbf{x}) - \phi(\mathbf{x}_0) \geq \langle \mathbf{g}, \mathbf{x} - \mathbf{x}_0 \rangle$$



The subgradient is a generalization of the gradient for non-differentiable convex functions

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - r_k \nabla \phi(\mathbf{x}^{(k)}) \longrightarrow \boxed{\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - r_k \mathbf{g}_k}$$

Projected subgradient and Superiorization

$$\mathbf{x}^* = \arg\min \phi(\mathbf{x})$$
$$s.t.\ \mathbf{x} \in C$$

<u>Projected subgradient</u>

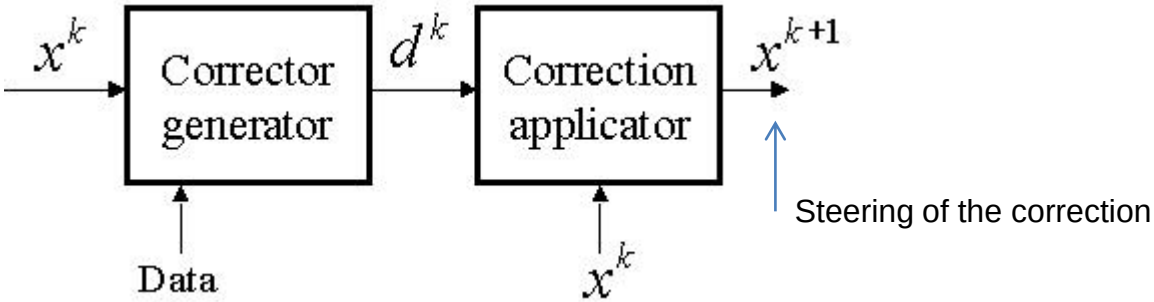
$$\mathbf{x}^{(k+1)} = \mathrm{Proj}_C \left\{ \mathbf{x}^{(k)} - r_k \mathbf{g}_k \right\}$$

<u>Superiorization</u>

set $n = 0$
set $y^{k,n} = y^k$
while $n < N$
  set $v^{k,n}$ to be a nonascending vector for ϕ at $y^{k,n}$
  set $loop = true$
  while $loop$
    set $\ell = \ell + 1$
    set $\beta_{k,n} = \eta_\ell$
    set $z = y^{k,n} + \beta_{k,n} v^{k,n}$
    if $\phi(z) \leq \phi(y^k)$ **then**
      set $n = n + 1$
      set $y^{k,n} = z$
      set $loop = false$
set $y^{k+1} = A_C(y^{k,N})$
set $k = k + 1$

Censor, Y.; Davidi, R.; Herman, G. T.; Schulte, R. W. & Tetruashvili, L. Projected subgradient minimization versus superiorization. *J. Optimization Theory and Applications*, **2014**, *160*, 730-747

Iterative perturbations



Censor, Y. & Matej, S. Binary steering of nonbinary iterative algorithms *Discrete tomography: Foundations, algorithms and applications, Birkhauser,* **1999**, 285-297

Algorithm 2 ***The Family-Constrained Algorithm (FCA)***

Initialization: *$x^0 \in R^n$ is arbitrary.*

Iterative step: *For every $k \geq 0$, given the current iterate x^k calculate the next iterate x^{k+1} by*

$$x^{k+1} = S_{k+1} Q(x^k). \tag{28}$$

Censor, Y.; Pantelimon, I. & Popa, C. Family constraining of iterative algorithms. *Numerical Algorithms, Springer,* **2014***, 66*, 323-338

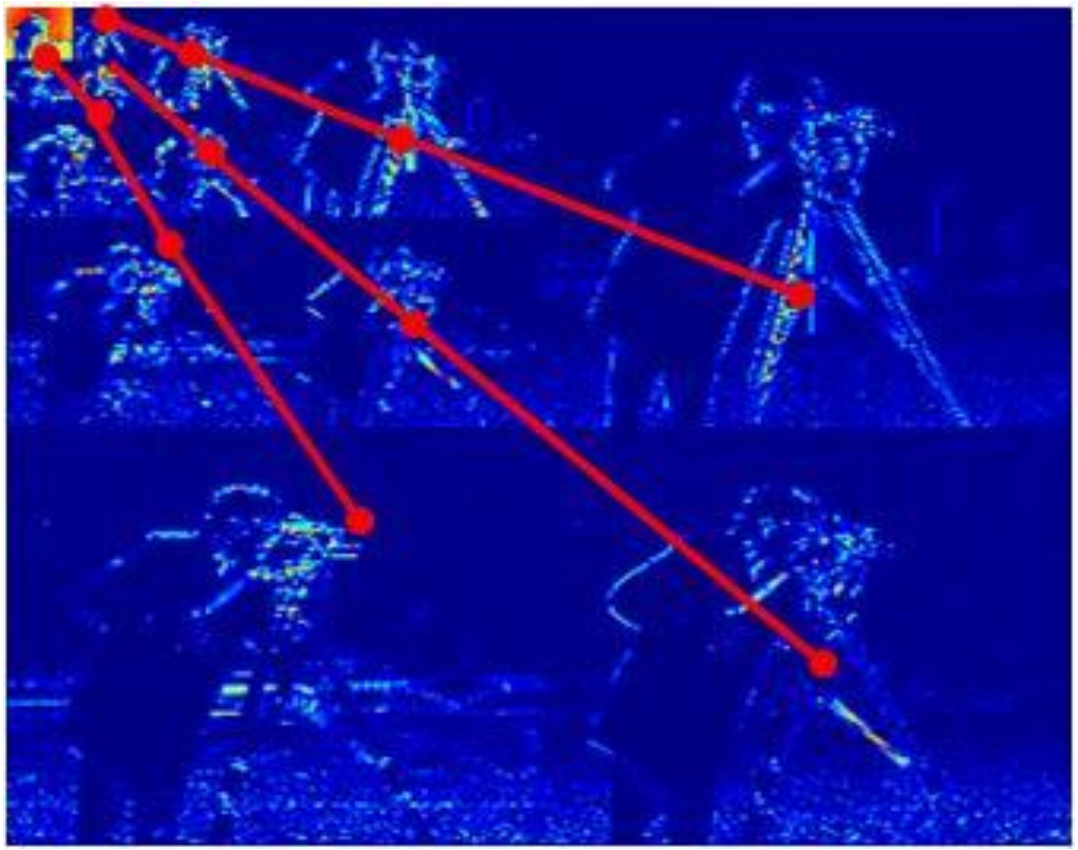
Censor, Y.; Davidi, R. & Herman, G. T. Perturbation resilience and superiorization of iterative algorithms. *Inverse Problems, IOP Publishing,* **2010***, 26*, 065008

CSIC

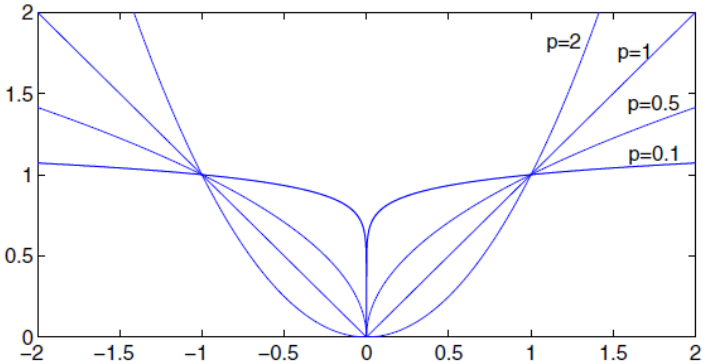
instruct Integrating Biology

instruct image Processing Center

Sparsity in image processing



$$\|\mathbf{x}\|_p = \left(\sum_i |x_i|^p \right)^{1/p}$$



$$\mathbf{x} = \arg\min \|\boldsymbol{\alpha}\|_0$$
$$s.t.\ \mathbf{x} = D\boldsymbol{\alpha}$$

Standard transformation (Wavelets, DCT, Fourier, …) or Ad hoc dictionary

NP hard

Bruckstein, A. M.; Donoho, D. L. & Elad, M. From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images. *SIAM Review,* **2009**, *51*, 34-81

Sparsity in image processing

$$(P_0^\delta): \quad \min_{\mathbf{x}} \ \|\mathbf{x}\|_0 \ \text{ subject to } \ \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2 \le \delta.$$

$$\min \|\mathbf{b} - A\mathbf{x}\|_2^2 \quad s.t. \|\mathbf{x}\|_0 \le N$$

- Matching pursuit
- Orthogonal matching pursuit

Orthogonal Matching Pursuit (OMP)

- Add an Orthogonal Projection to the residual calculation

1. set $I := \{\emptyset\},\ r := x,\ \gamma := 0$
2. while ($stopping\ test\ false$) do
3. $\quad k := \underset{k}{\operatorname{argmax}} |d_k^T r|$
4. $\quad I := (I, k)$
5. $\quad \gamma_I := (\mathbf{D}_I)^+ x$
6. $\quad r := x - \mathbf{D}_I \gamma_I$
7. end while

CSIC

instruct Integrating Biology

instruct image Processing Center

Sparsity in image processing

$$(P_1^\delta): \quad \min_{\mathbf{x}} \ \|\mathbf{W}\mathbf{x}\|_1 \ \text{ subject to } \ \|\mathbf{b}-\mathbf{A}\mathbf{x}\|_2 \le \delta,$$

$$\min \|\mathbf{b} - A\mathbf{x}\|_2^2 \quad s.t. \|\mathbf{x}\|_1 \le \delta$$

Ridge regression
(Tikhonov regularization)

$$\min \|\mathbf{b} - A\mathbf{x}\|_2^2 + \lambda \|\Gamma\mathbf{x}\|_2$$

- Lasso and its variants
- Iterative Reweighted Least Squares

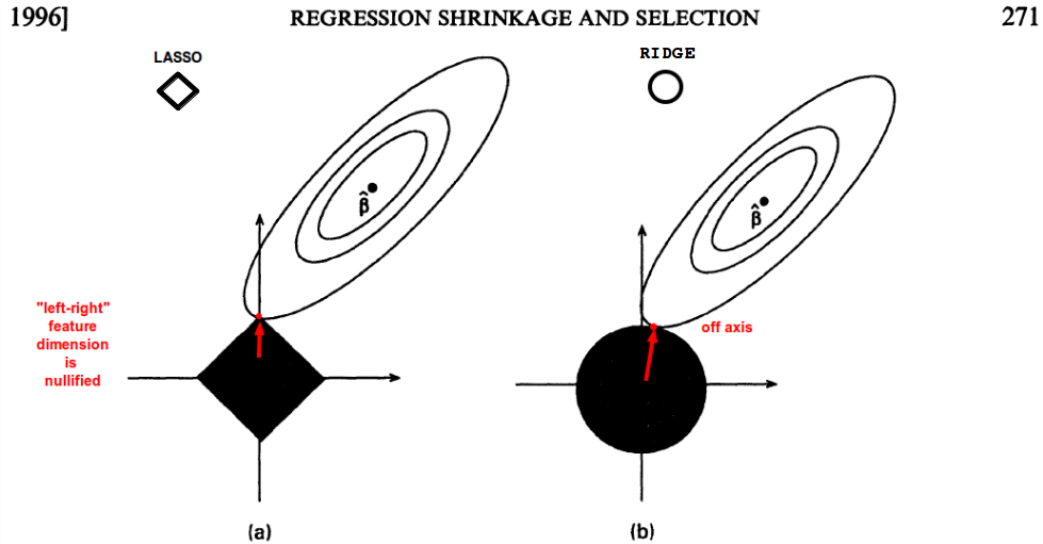


Fig. 2. Estimation picture for (a) the lasso and (b) ridge regression

CSIC
instruct Integrating Biology
instruct image Processing Center

Alternating Direction Method of Multipliers

- ADMM problem form (with f, g convex)

$$\begin{array}{ll} \text{minimize} & f(x) + g(z) \\ \text{subject to} & Ax + Bz = c \end{array}$$

  - two sets of variables, with separable objective

- $L_\rho(x, z, y) = f(x) + g(z) + y^T(Ax + Bz - c) + (\rho/2)\|Ax + Bz - c\|_2^2$

- ADMM:

$$\begin{array}{lcll} x^{k+1} & := & \mathrm{argmin}_x\, L_\rho(x, z^k, y^k) & \text{// } x\text{-minimization} \\ z^{k+1} & := & \mathrm{argmin}_z\, L_\rho(x^{k+1}, z, y^k) & \text{// } z\text{-minimization} \\ y^{k+1} & := & y^k + \rho(Ax^{k+1} + Bz^{k+1} - c) & \text{// dual update} \end{array}$$

Boyd, S.; Parikh, N.; Chu, E.; Peleato, B. & Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning, Now Publishers Inc.*, **2011**, *3*, 1-122

Sparsifying regularization

Classical reconstruction = linear algorithm

$$\mathbf{s}^\star = \operatorname{argmin} \underbrace{\|\mathbf{g} - \mathbf{H}\mathbf{s}\|_2^2}_{\text{data consistency}} + \underbrace{\lambda \mathcal{R}(\mathbf{s})}_{\text{regularization}}$$

- Quadratic regularization (Tikhonov)

  $\mathcal{R}(\mathbf{s}) = \|\mathbf{L}\mathbf{s}\|^2$

  Formal linear solution: $\mathbf{s} = (\mathbf{H}^T\mathbf{H} + \lambda \mathbf{L}^T\mathbf{L})^{-1}\mathbf{H}^T\mathbf{g} = \mathbf{R}_\lambda \cdot \mathbf{g}$

$\Updownarrow$ $\mathbf{L} = \mathbf{C}_s^{-1/2}$: Whitening filter

- Statistical formulation under Gaussian hypothesis

  Wiener (LMMSE) solution = Gauss MMSE = Gauss MAP

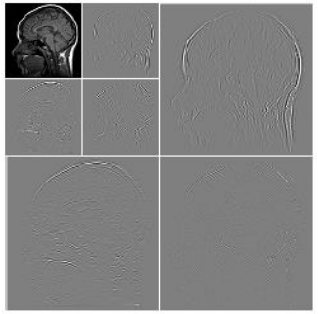
$$\mathbf{s}_{\mathrm{MAP}} = \arg\min_{\mathbf{s}} \underbrace{\frac{1}{\sigma^2}\|\mathbf{g} - \mathbf{H}\mathbf{s}\|_2^2}_{\text{Data Log likelihood}} + \underbrace{\|\mathbf{C}_s^{-1/2}\mathbf{s}\|_2^2}_{\text{Gaussian prior likelihood}}$$

Signal covariance: $\mathbf{C}_s = \mathbb{E}\{\mathbf{s} \cdot \mathbf{s}^T\}$

Sparsifying regularization

Current trend: non-linear algorithms (l_1 optimization)

$$\mathbf{s}^\star = \text{argmin} \underbrace{\|\mathbf{y} - \mathbf{H}\mathbf{s}\|_2^2}_{\text{data consistency}} + \underbrace{\lambda \mathcal{R}(\mathbf{s})}_{\text{regularization}}$$



- Wavelet-domain regularization

  Wavelet expansion: $\mathbf{s} = \mathbf{W}\mathbf{v}$ (typically, sparse)

  Wavelet-domain sparsity-constraint: $\mathcal{R}(\mathbf{s}) = \|\mathbf{v}\|_{\ell_1}$ with $\mathbf{v} = \mathbf{W}^{-1}\mathbf{s}$

  (Nowak et al., Daubechies et al. 2004)

- ℓ_1 regularization (Total variation=TV) (Rudin-Osher, 1992)

  $\mathcal{R}(\mathbf{s}) = \|\mathbf{L}\mathbf{s}\|_{\ell_1}$ with $\mathbf{L}$: gradient

  Aganj, I.; Bartesaghi, A.; Borgnia, M.; Liao, H. Y.; Sapiro, G. & Subramaniam, S. Regularization for Inverting the Radon Transform with Wedge Consideration. *Proc. Intl. Symposium of Biomedical Imaging (ISBI),* **2007**, 217-220

- Compressed sensing/sampling (Candes-Romberg-Tao; Donoho, 2006)

Sparsifying regularization

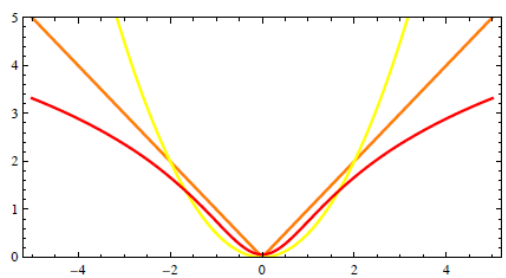
General form of MAP estimator

$$\mathbf{s}_{\mathrm{MAP}} = \operatorname{argmin}\left(\tfrac{1}{2}\,\|\mathbf{y}-\mathbf{H}\mathbf{s}\|_2^2 + \sigma^2 \textstyle\sum_n \Phi_U([\mathbf{L}\mathbf{s}]_n)\right)$$

Sparser

- Gaussian: $p_U(x) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-x^2/(2\sigma_0^2)}$ $\Rightarrow$ $\Phi_U(x) = \frac{1}{2\sigma_0^2}x^2 + C_1$
- Laplace: $p_U(x) = \frac{\lambda}{2} e^{-\lambda|x|}$ $\Rightarrow$ $\Phi_U(x) = \lambda|x| + C_2$
- Student: $p_U(x) = \dfrac{1}{B\left(r,\frac{1}{2}\right)}\left(\dfrac{1}{x^2+1}\right)^{r+\frac{1}{2}}$ $\Rightarrow$ $\Phi_U(x) = \left(r+\dfrac{1}{2}\right)\log(1+x^2) + C_3$

Potential: $\Phi_U(x) = -\log p_U(x)$

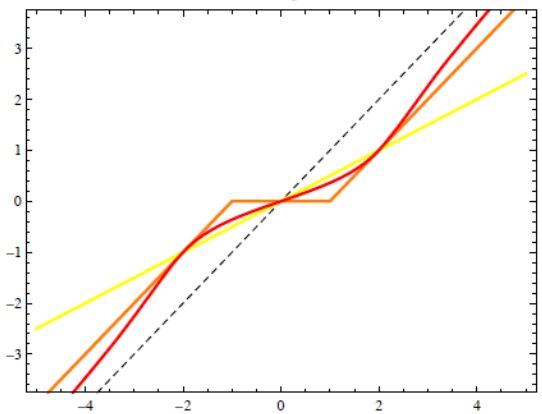


Sparsifying regularization

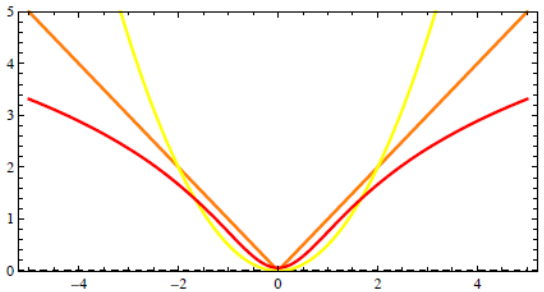
Proximal operator: pointwise denoiser

$$\mathrm{prox}_{\Phi_U}(y;\sigma^2) = \arg\min_{u\in\mathbb{R}} \frac{1}{2}|y-u|^2 + \sigma^2\Phi_U(u)$$

$$\tilde{u} = \mathrm{prox}_{\Phi_U}(y;1)$$



$$\sigma^2\Phi_U(u)$$



- linear attenuation — ℓ_2 minimization
- soft-threshold — ℓ_1 minimization
- shrinkage function — $\approx$ ℓ_p relaxation for $p \to 0$

Unser. Sparse modeling and the resolution of inverse problems in biomedical imaging. Plenary talk, IEEE Int. Symp. Biomedical Imaging (ISBI'15), 16-19 April, 2015, New York, USA

Journal Club

Journal of Structural Biology 195 (2016) 100–112

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

ICON: 3D reconstruction with 'missing-information' restoration in biological electron tomography

Yuchen Deng [a,d,1], Yu Chen [b,d,1], Yan Zhang [a], Shengliu Wang [a,d], Fa Zhang [b,*], Fei Sun [a,c,d,*]

[a] National Key Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
[b] Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China
[c] Center for Biological Imaging, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
[d] University of Chinese Academy of Sciences, Beijing, China

ARTICLE INFO

Article history:
Received 11 January 2016
Received in revised form 9 April 2016
Accepted 10 April 2016
Available online 11 April 2016

Keywords:
Compressed sensing
Electron tomography
Missing wedge
NUFFT
Sub-tomogram averaging

ABSTRACT

Electron tomography (ET) plays an important role in revealing biological structures, ranging from macromolecular to subcellular scale. Due to limited tilt angles, ET reconstruction always suffers from the 'missing wedge' artifacts, thus severely weakens the further biological interpretation. In this work, we developed an algorithm called Iterative Compressed-sensing Optimized Non-uniform fast Fourier transform reconstruction (ICON) based on the theory of compressed-sensing and the assumption of sparsity of biological specimens. ICON can significantly restore the missing information in comparison with other reconstruction algorithms. More importantly, we used the leave-one-out method to verify the validity of restored information for both simulated and experimental data. The significant improvement in sub-tomogram averaging by ICON indicates its great potential in the future application of high-resolution structural determination of macromolecules *in situ*.

ICON 3D

$$\text{argmin } ||Px||_{L_0}$$
$$\text{subject to}: \ ||A^h WAx - A^h Wf||_{L_2} < \varepsilon$$

Gradient descent →

$$H(x) = \begin{cases} 0, (\text{if } x < 0) \\ x, (\text{if } x \geqslant 0) \end{cases}$$

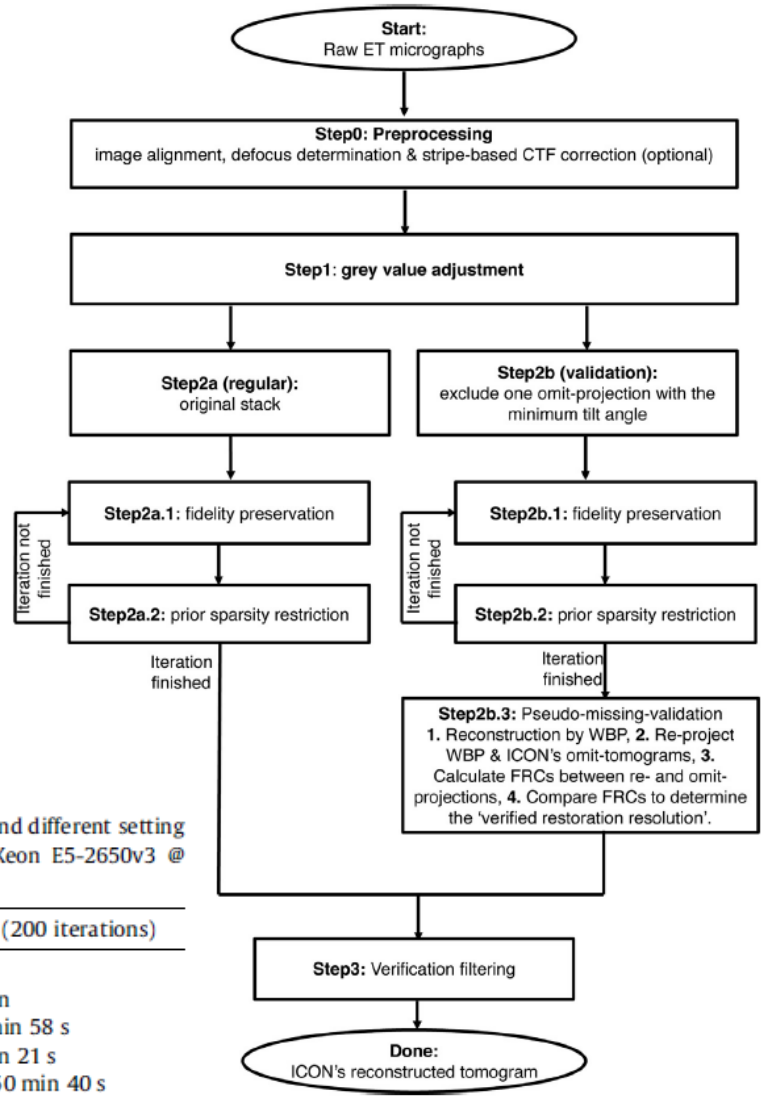
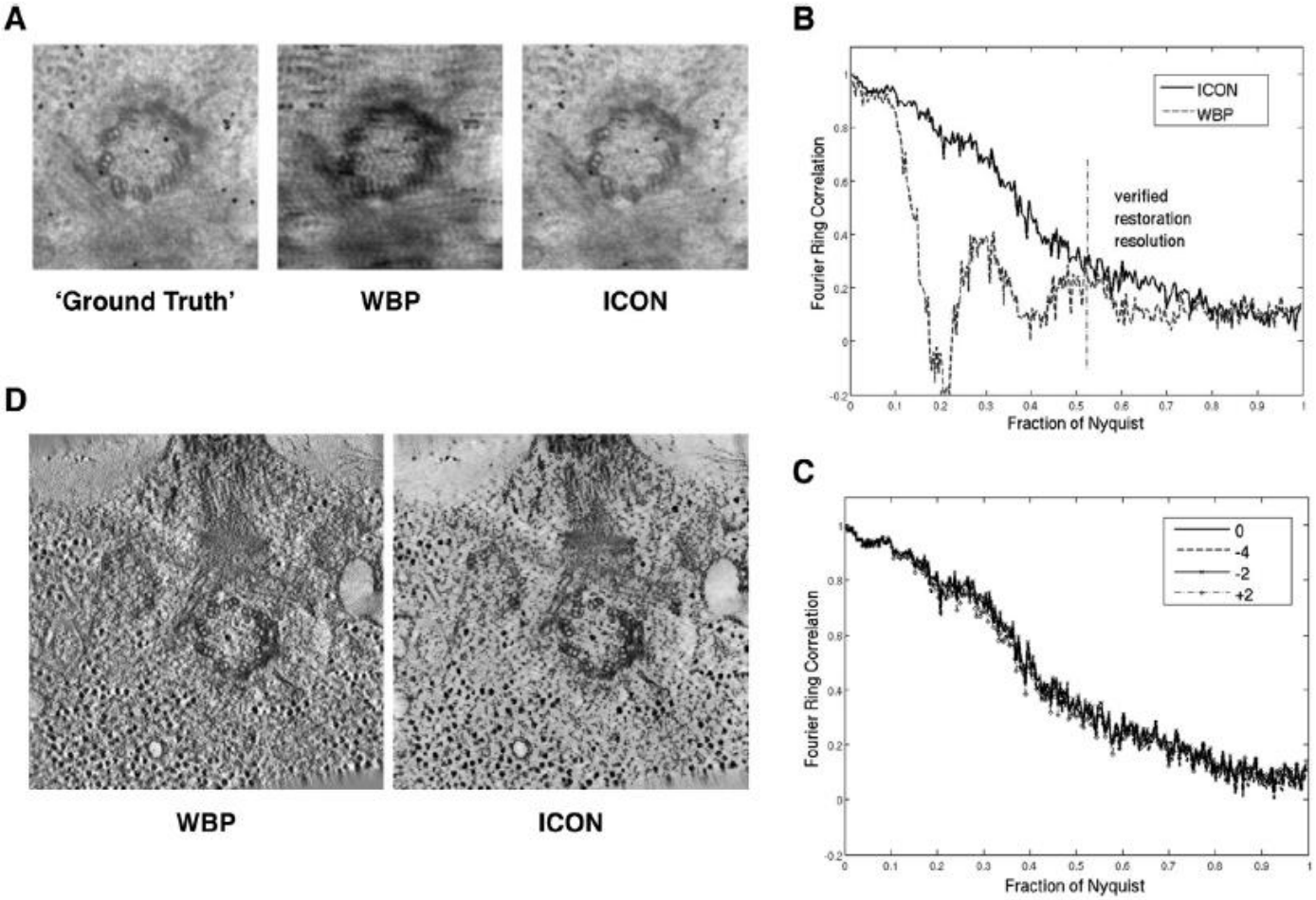


Table 1
Computational costs of ICON and SIRT for different sizes of data and different setting of CPU. One CPU represents one core of the processor Intel Xeon E5-2650v3 @ 2.3 GHz.

	ICON (200 iterations)	SIRT (200 iterations)
512^2 (one slice, one CPU)	20 s	22 s
2048^2 (one slice, one CPU)	4 min 56 s	8 min
4096^2 (one slice, one CPU)	18 min 32 s	36 min 58 s
512^2 (512 slices, 100 CPUs)	2 min 3 s	2 min 21 s
2048^2 (2048 slices, 100 CPUs)	1 h 42 min 24 s	2 h 50 min 40 s
4096^2 (4096 slices, 100 CPUs)	13 h 39 min 39 s	26 h 10 min 30 s

instruct image Processing Center

ICON 3D

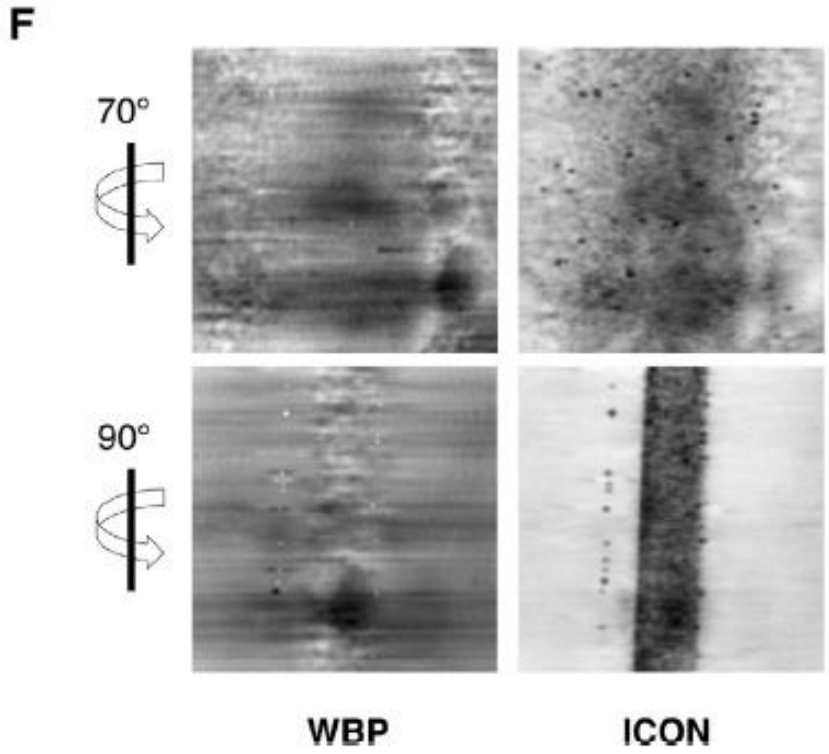


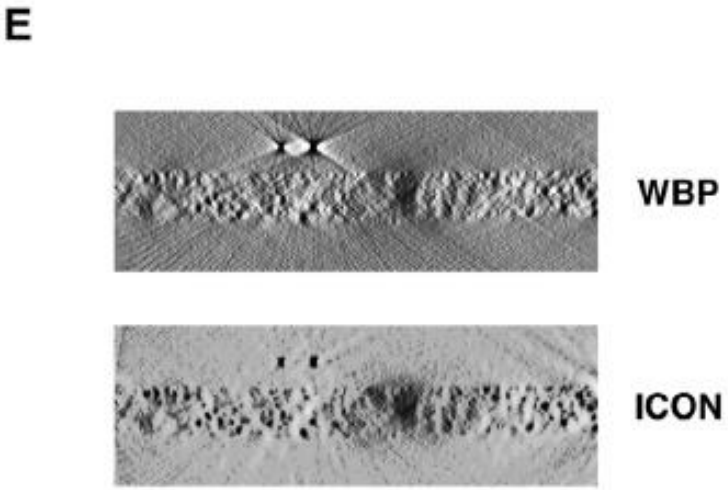
CSIC

instruct Integrating Biology

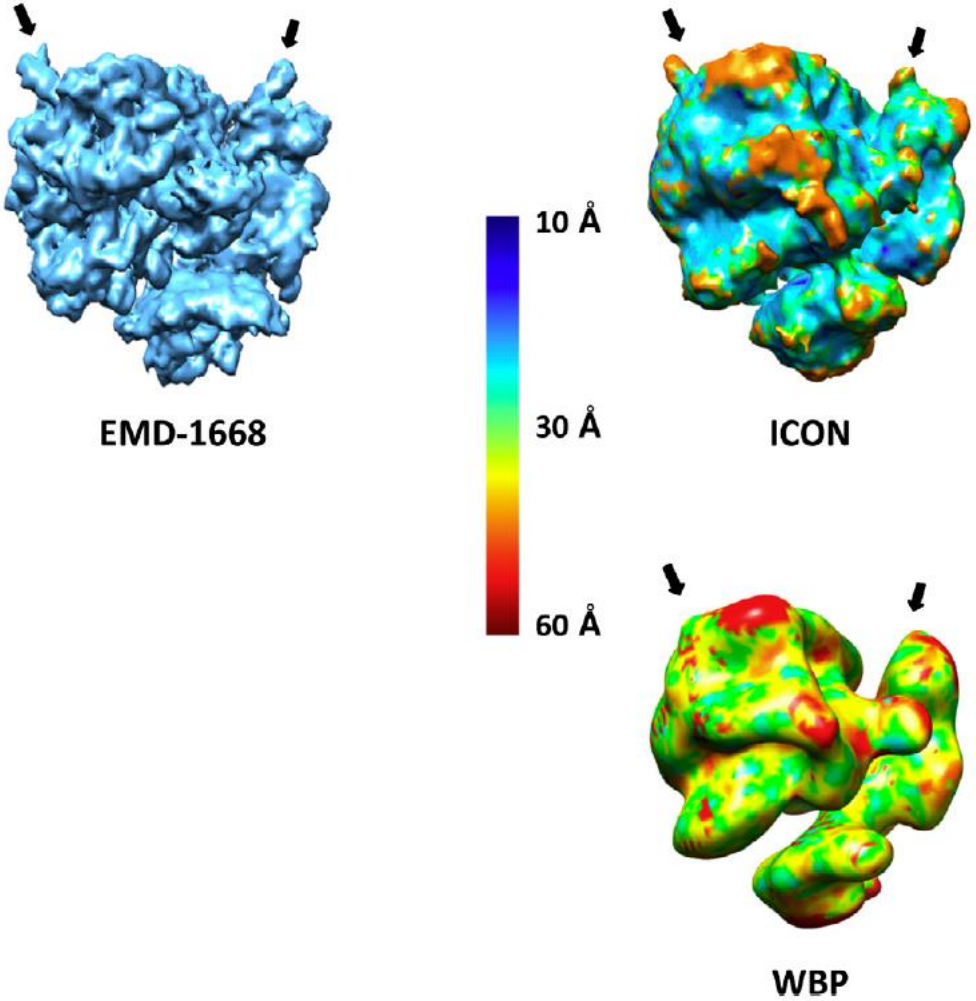
instruct image Processing Center

ICON 3D





ICON 3D



CSIC

instruct Integrating Biology

instruct image Processing Center

FIRT

Journal of Structural Biology 195 (2016) 49–61

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

FIRT: Filtered iterative reconstruction technique with information restoration

Yu Chen [a,c,1], Yan Zhang [b,1], Kai Zhang [b,c,*,1], Yuchen Deng [b,c], Shengliu Wang [b,c], Fa Zhang [a,*], Fei Sun [b,c,d,*]

[a] Key Lab of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China
[b] National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Beijing 100101, China
[c] University of Chinese Academy of Sciences, Beijing, China
[d] Center for Biological Imaging, Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Beijing 100101, China

ARTICLE INFO

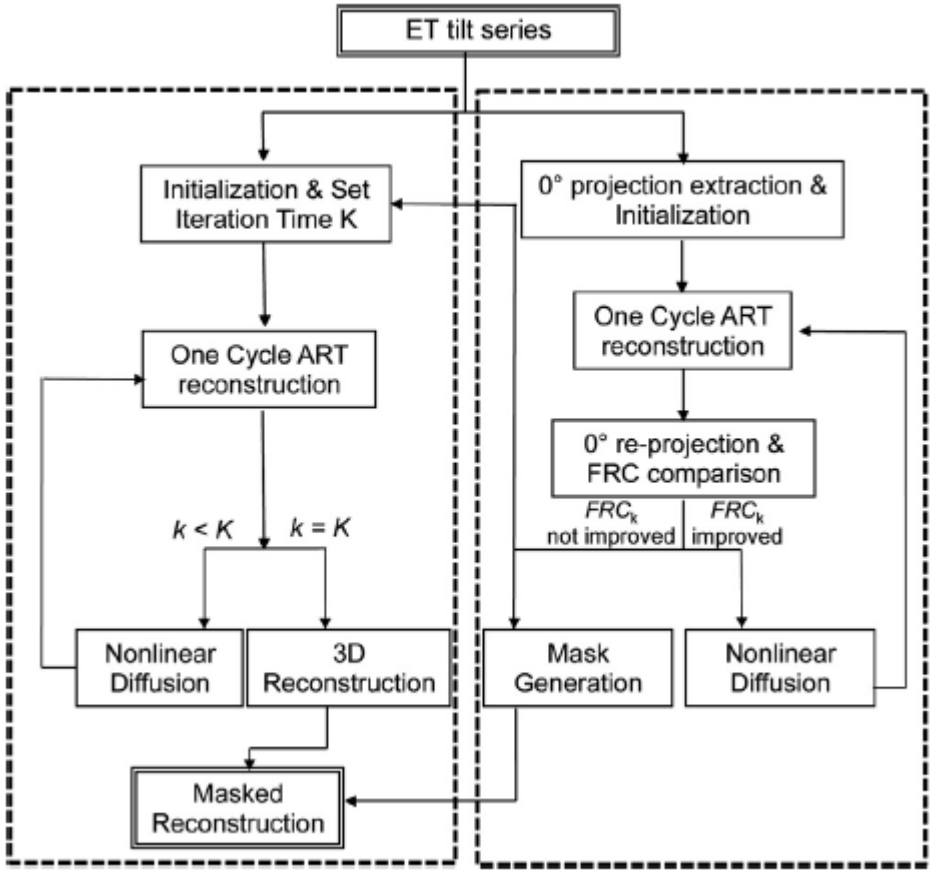
Article history:
Received 21 December 2015
Received in revised form 27 April 2016
Accepted 28 April 2016
Available online 29 April 2016

Keywords:
Algebra reconstruction technique
Electron tomography
Filtered iterative reconstruction technique
Information restoration
Nonlinear diffusion filter

ABSTRACT

Electron tomography (ET) combining subsequent sub-volume averaging has been becoming a unique way to study the in situ 3D structures of macromolecular complexes. However, information missing in electron tomography due to limited angular sampling is still the bottleneck in high-resolution electron tomography application. Here, based on the understanding of smooth nature of biological specimen, we present a new iterative image reconstruction algorithm, FIRT (filtered iterative reconstruction technique) for electron tomography by combining the algebra reconstruction technique (ART) and the nonlinear diffusion (ND) filter technique. Using both simulated and experimental data, in comparison to ART and weight back projection method, we proved that FIRT could generate a better reconstruction with reduced ray artifacts and significant improved correlation with the ground truth and partially restore the information at the non-sampled angular region, which was proved by investigating the 90° re-projection and by the cross-validation method. This new algorithm will be subsequently useful in the future for both cellular and molecular ET with better quality and improved structural details.

FIRT

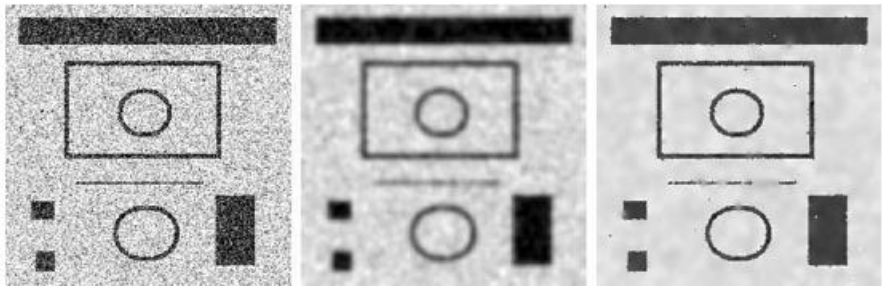


ART

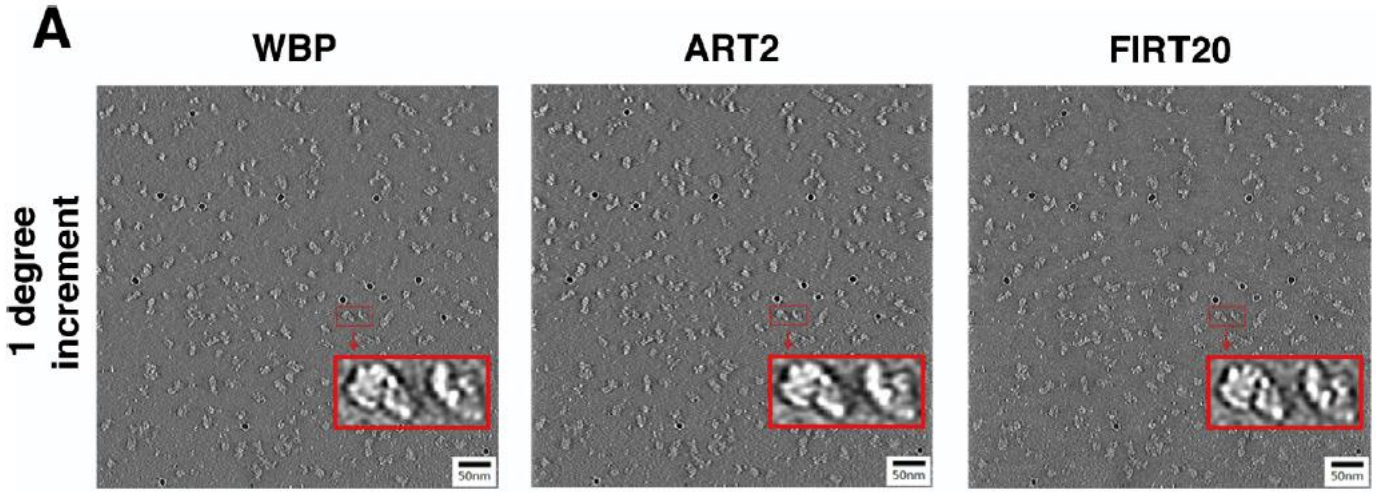
$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \eta\left(\mathrm{Proj}_{i(k)}\left\{\mathbf{x}^{(k)}\right\} - \mathbf{x}^{(k)}\right)$$

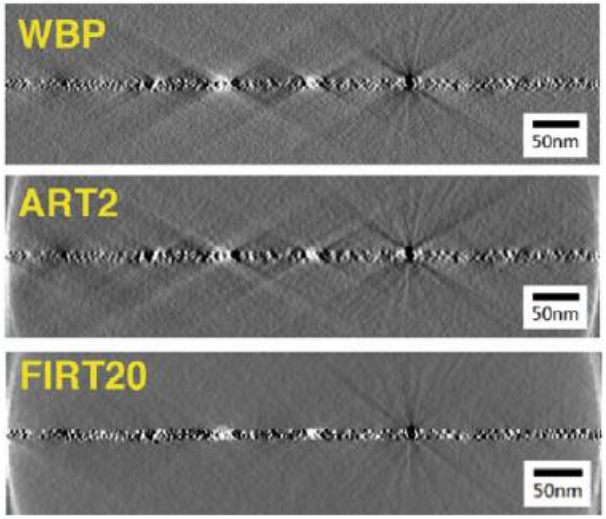
Non-linear diffusion

$$I^{t+1} = I^{t} + \tau \cdot \frac{\nabla I}{1 + \left(\frac{\nabla I}{K}\right)^{2}}$$



FIRT





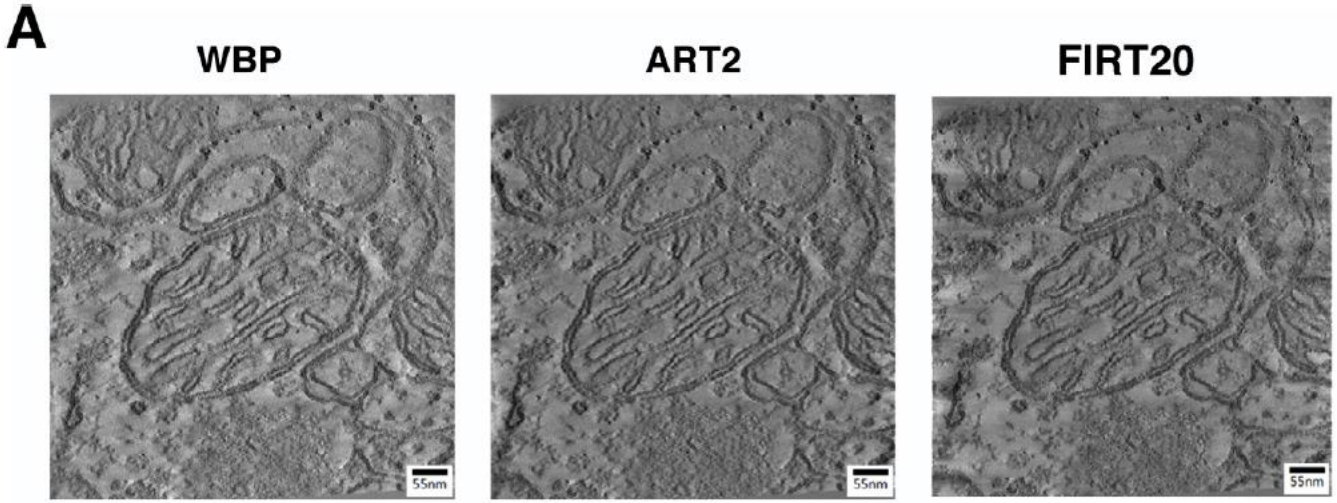
1 degree increment

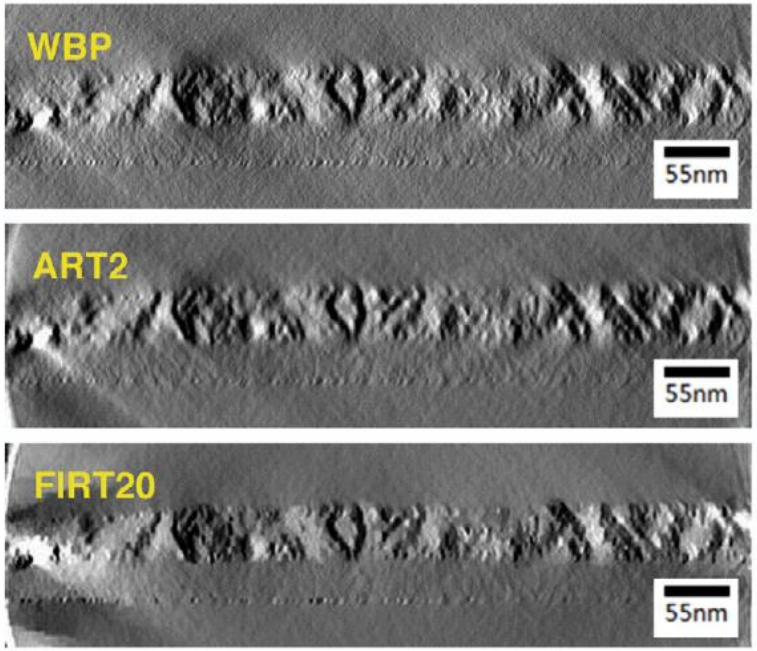
CSIC

instruct Integrating Biology

instruct image Processing Center

FIRT





CSIC

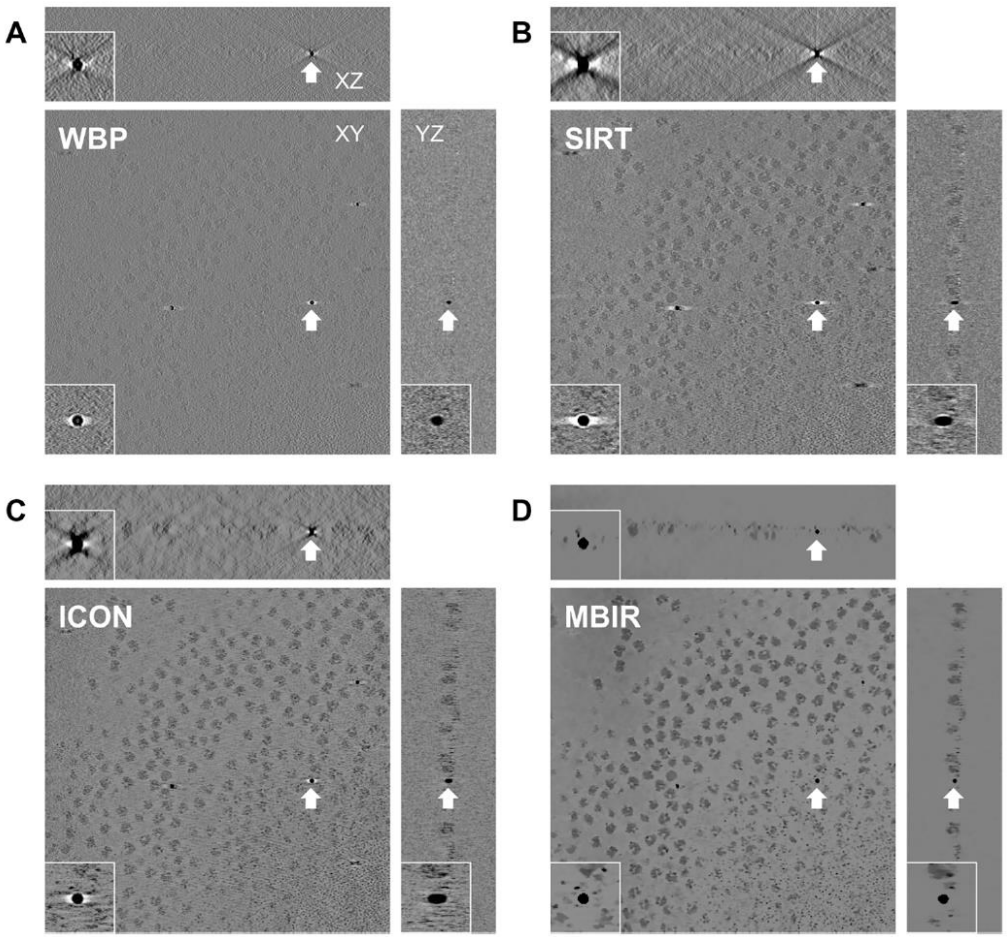
instruct Integrating Biology

BIO COMPUTING Unit

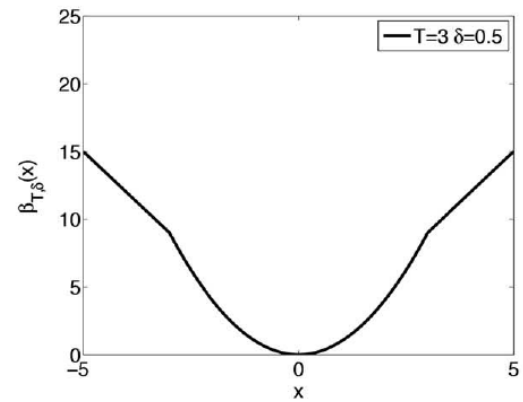
instruct image Processing Center

MBIR

$$(\hat{f}, \hat{\phi}) = \underset{f,\phi}{\operatorname{argmax}}\{\log p(f, \phi|g)\} = \underset{f,\phi}{\operatorname{argmin}}\{-\log p(g|f, \phi) - \log p(f)\}$$

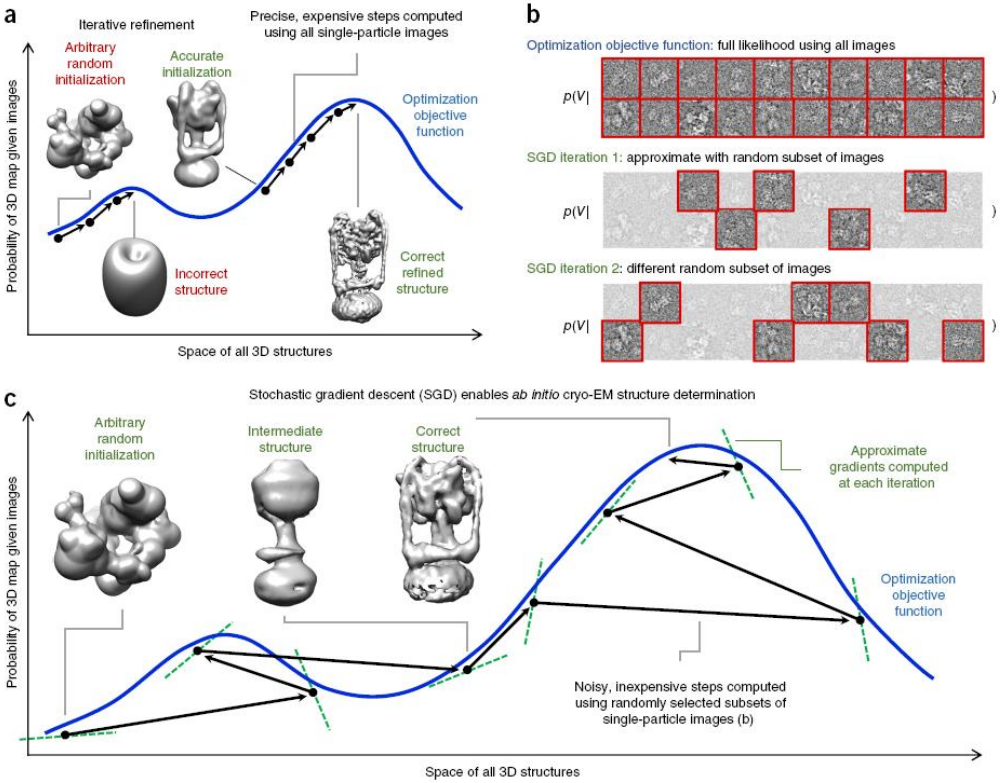


$$p(g|f, d, \sigma) = \frac{1}{Z}\exp\left\{-\frac{1}{2}\sum_{k=1}^{K}\sum_{i=1}^{M}\beta_{T,\delta}\left((g_{k,i} - A_{k,i,*}f - d_k)\frac{\sqrt{\Lambda_{k,ii}}}{\sigma}\right)\right\}$$



Yan, R.; Venkatakrishnan, S. V.; Liu, J.; Bouman, C. A.; Jiang, W. MBIR: A cryo-ET 3D reconstruction method that effectively minimizes missing wedge artifacts and restores missing information. Journal of structural biology, 2019, 206, 183-192
Venkatakrishnan, S. V.; Drummy, L. F.; Jackson, M.; De Graef, M.; Simmons, J. & Bouman, C. A. Model-Based Iterative Reconstruction for Bright-Field Electron Tomography. IEEE Transactions on Computational Imaging, 2015 , 1 , 1-15

Stochastic Gradient Descent



Punjani, A.; Brubaker, M. A. & Fleet, D. J. Building Proteins in a Day: Efficient 3D Molecular Structure Estimation with Electron Cryomicroscopy. IEEE transactions on pattern analysis and machine intelligence, 2017, 39, 706-718

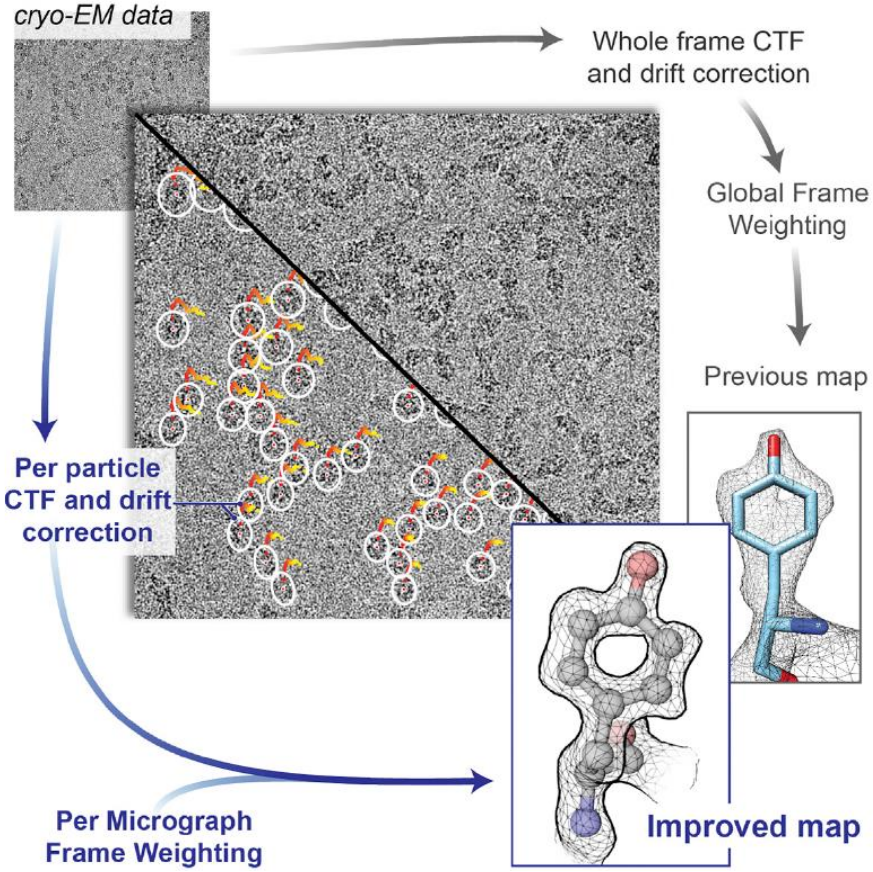
Punjani, A.; Rubinstein, J. L.; Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nature methods, 2017

CSIC

instruct Integrating Biology

instruct image Processing Center

Parameter refinement

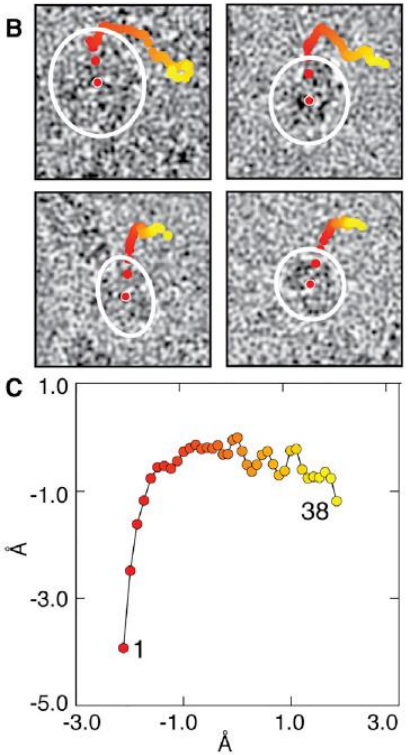


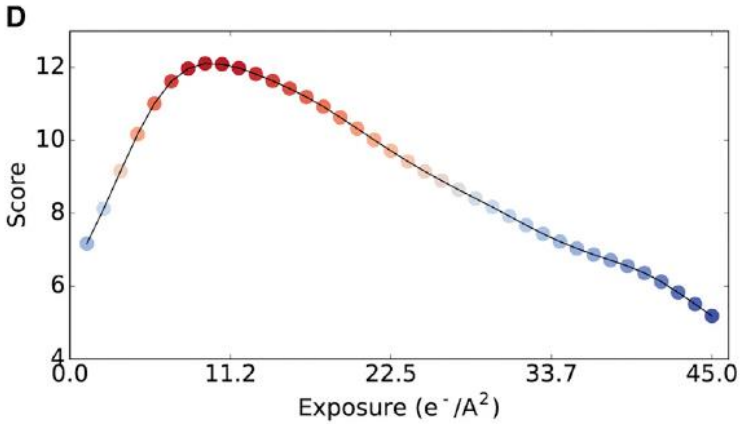
CSIC

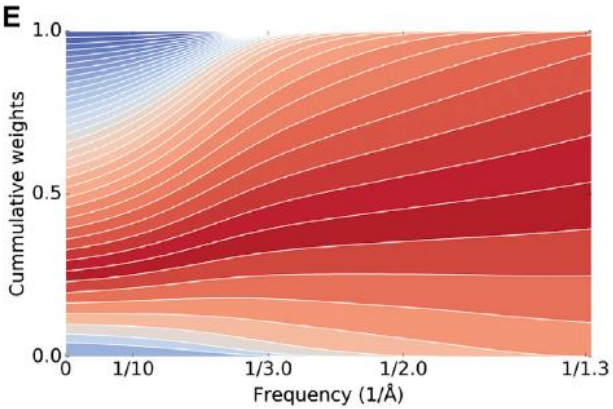
instruct
Integrating
Biology

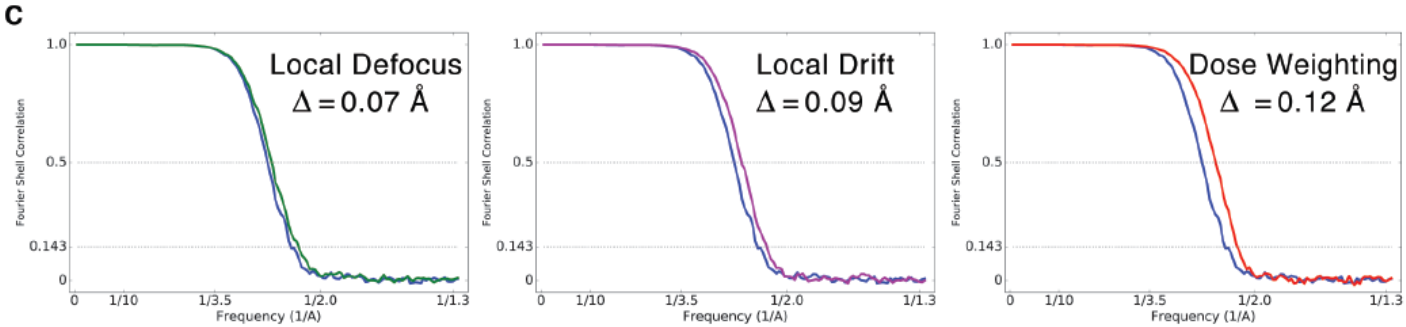
instruct
image
Processing
Center

Parameter refinement



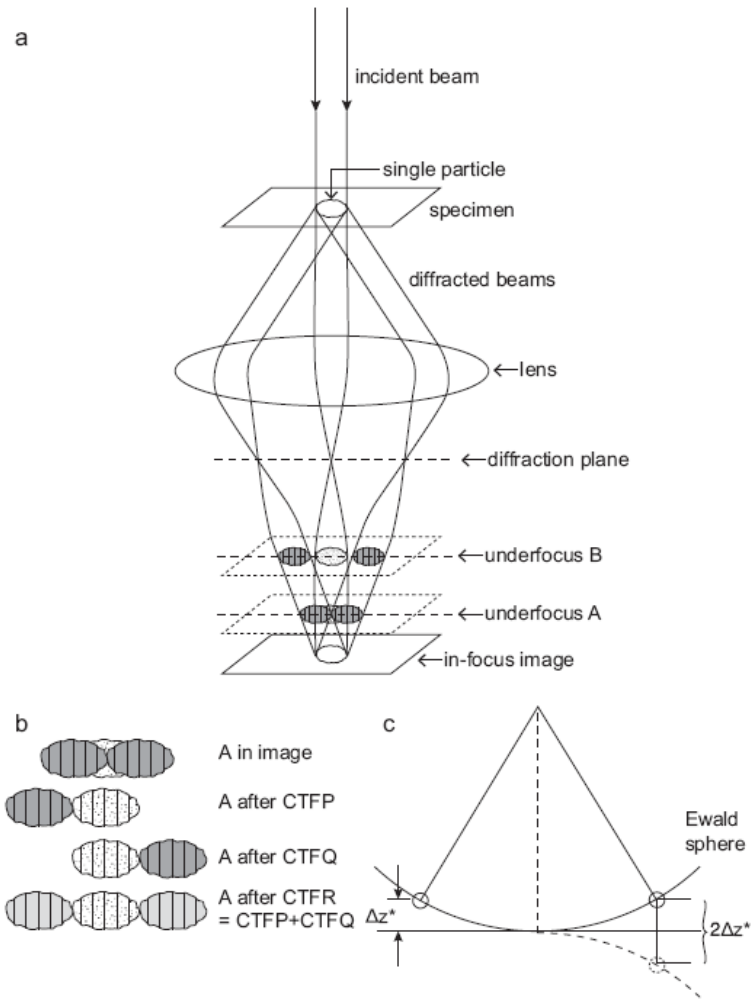


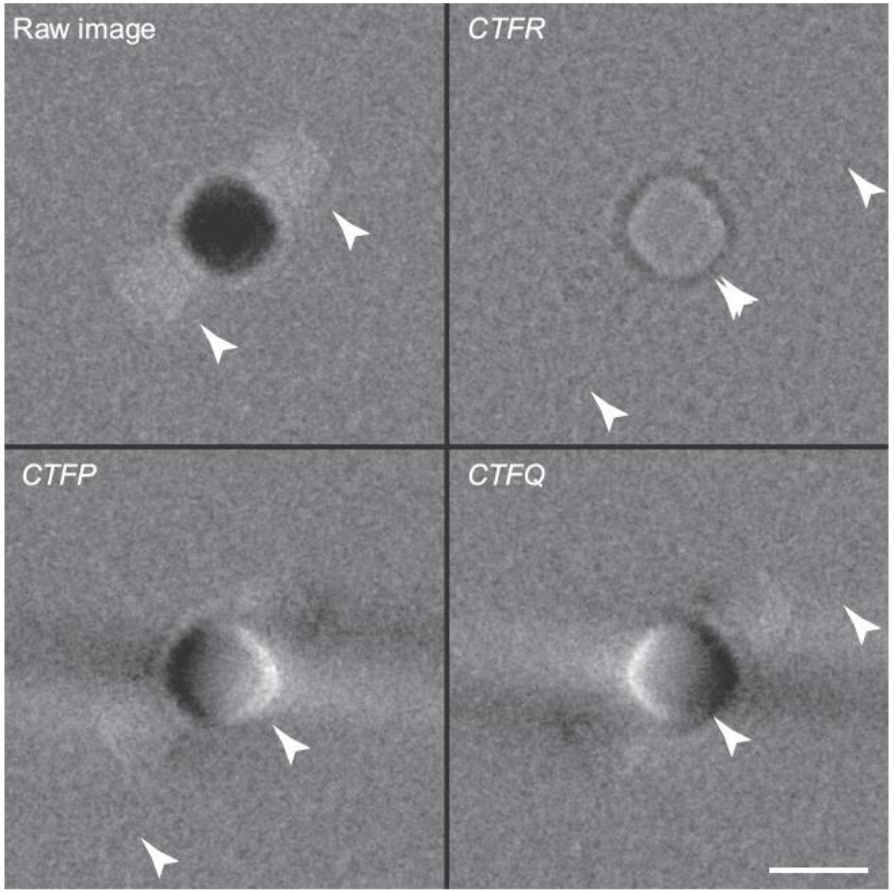




Bartesaghi, A.; Aguerrebere, C.; Falconieri, V.; Banerjee, S.; Earl, L. A.; Zhu, X.; Grigorieff, N.; Milne, J. L. S.; Sapiro, G.; Wu, X. & Subramaniam, S. Atomic Resolution Cryo-EM Structure of β-Galactosidase. Structure, 2018

Ewald sphere correction





Russo, C. J. & Henderson, R. Ewald sphere correction using a single side-band image processing algorithm. Ultramicroscopy, 2018, 187, 26-33

Zhu, D.; Wang, X.; Fang, Q.; Van Etten, J. L.; Rossmann, M. G.; Rao, Z. & Zhang, X. Pushing the resolution limit by correcting the Ewald sphere effect in single-particle Cryo-EM reconstructions. Nature communications, 2018 , 9 , 1552

Conclusions

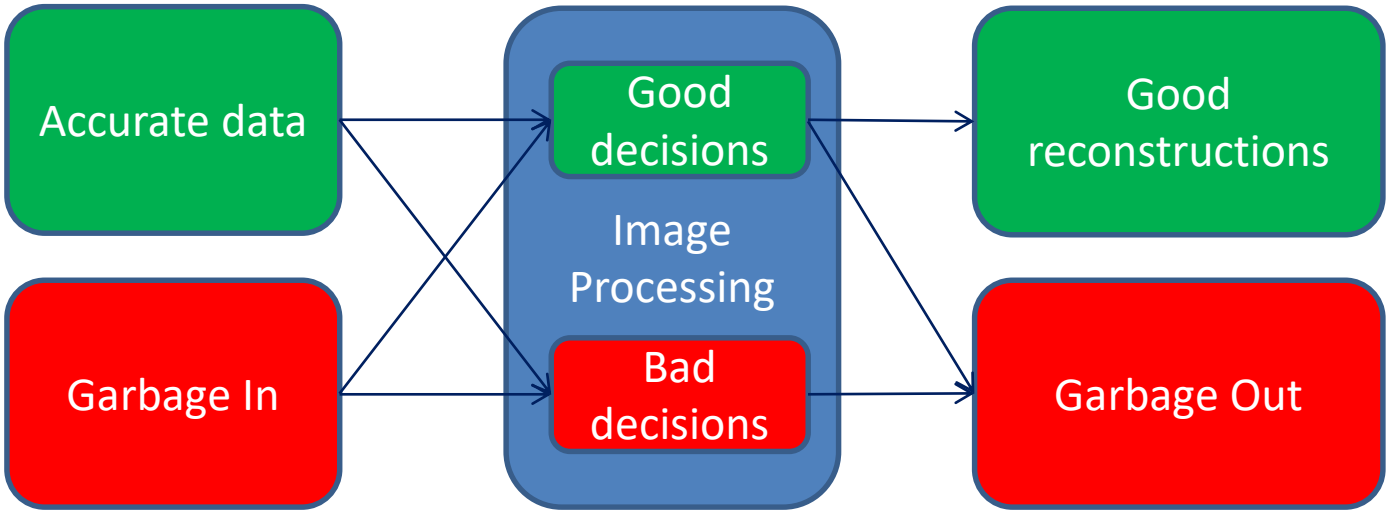
- WBP: “Easy”, traditional, not much used
- Fourier: Standard in SPA
- Iterative: Standard in Electron Tomography

Volume validation

Carlos Oscar S. Sorzano
Instruct Image Processing Center

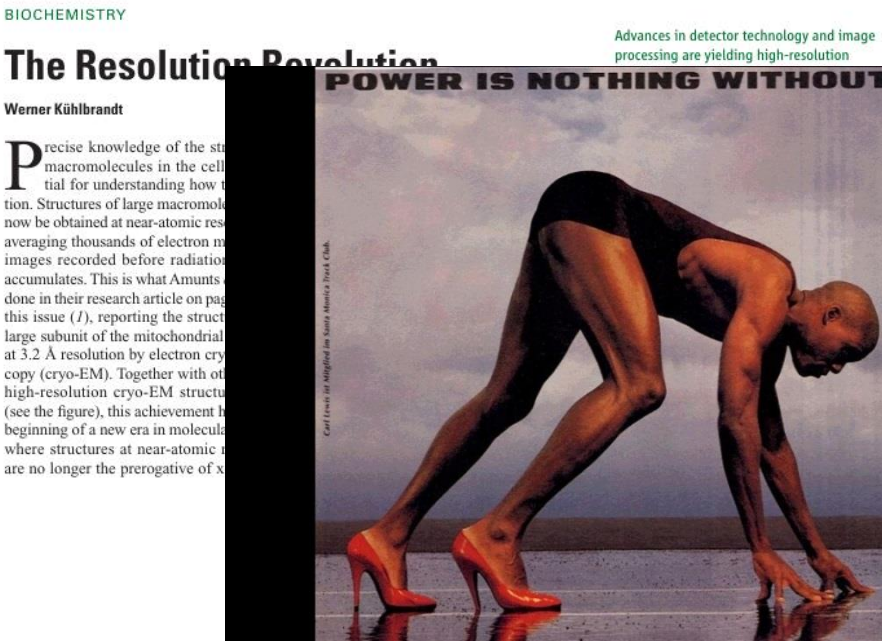
EM is a GIGO system

- The good thing about EM is that <u>it always gives a volume</u>
- The bad thing about EM is that <u>it always gives a volume</u>



EM is a GIGO system

- Do not go to the next step till you are sure of the current step

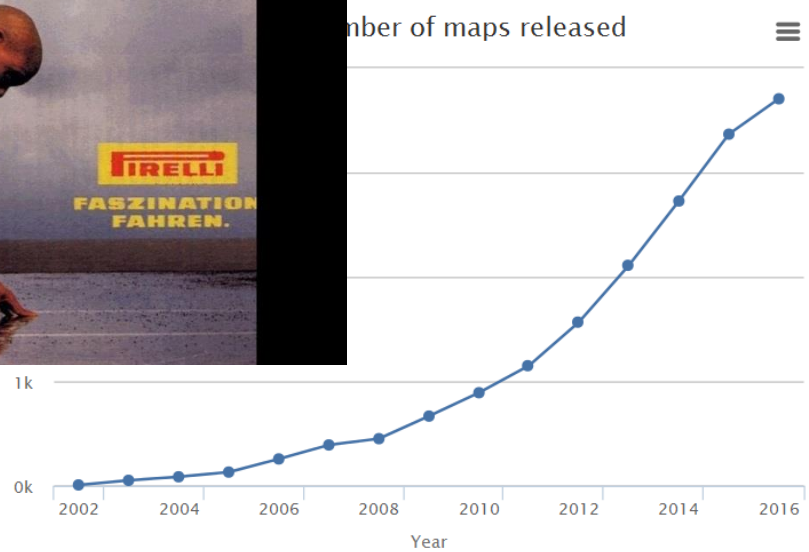

BIOCHEMISTRY

The Resolution Revolution

Advances in detector technology and image processing are yielding high-resolution

Werner Kühlbrandt

Precise knowledge of the str
macromolecules in the cell
tial for understanding how t
tion. Structures of large macromol
now be obtained at near-atomic res
averaging thousands of electron m
images recorded before radiatio
accumulates. This is what Amunts
done in their research article on pag
this issue (*1*), reporting the struct
large subunit of the mitochondrial
at 3.2 Å resolution by electron cry
copy (cryo-EM). Together with ot
high-resolution cryo-EM structu
(see the figure), this achievement h
beginning of a new era in molecula
where structures at near-atomic r
are no longer the prerogative of x



CSIC

instruct Integrating Biology

instruct image Processing Center

Controversial reconstructions

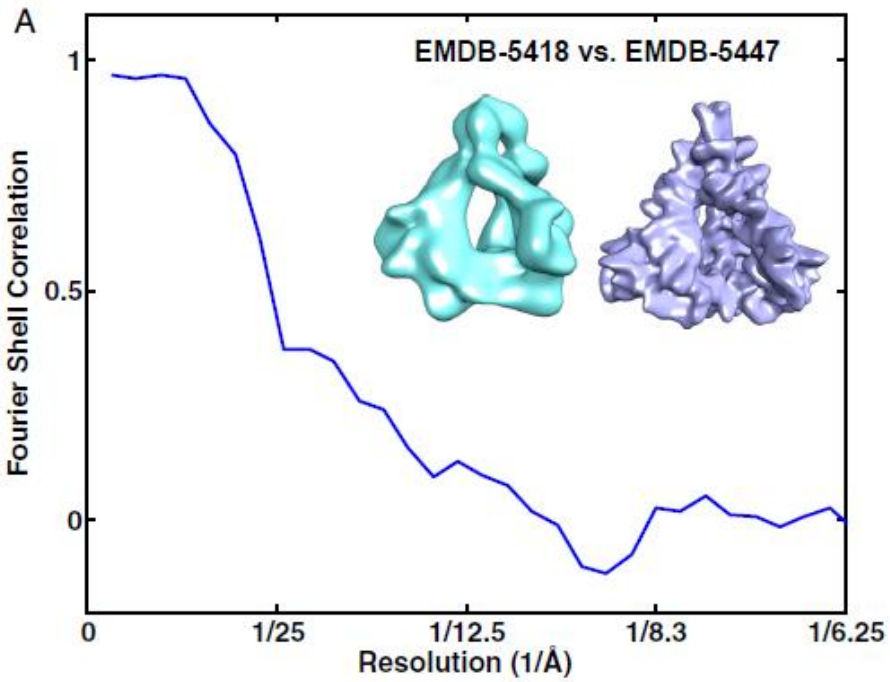
PNAS

LETTER

Structure of trimeric HIV-1 envelope glycoproteins

Mao et al. (1) published a paper describing a 3D structure of uncleaved, trimeric HIV-1 envelope glycoprotein (Env) at ~6-Å resolution, following a similar paper last year on the same structure at ~11-Å resolution (2).

uncleaved trimeric HIV-1 Env a
in the work of Mao et al. (1, 2)
ment with the structure of nati
trimeric HIV-1 Env presented i
(4) is incorrect. EMD-5019 repr



CSIC

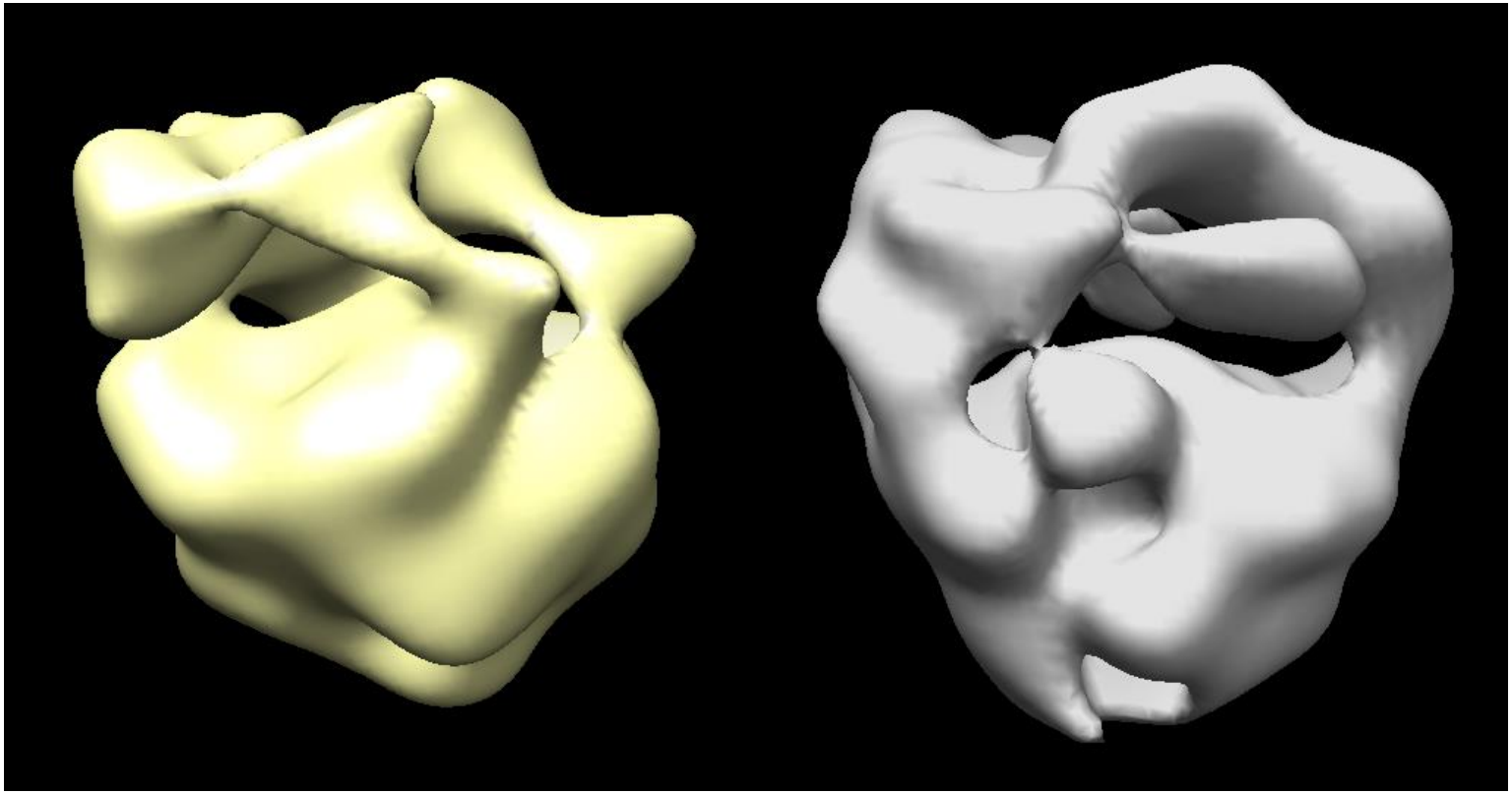
instruct Integrating Biology

instruct image Processing Center

Volume validation

- Necessary conditions:
  - Reprojection consistency
  - Angular distribution
  - Correlation histograms
  - Reproducibility
  - No overfitting
  - Better than noise
  - Homogeneity
- Resolution
  - Global, directional and local resolution
  - B-factor
- Consistency with data
  - Alignability
  - SSNR
- Absolute validation
  - Tilt pair validation
  - PDB fitting

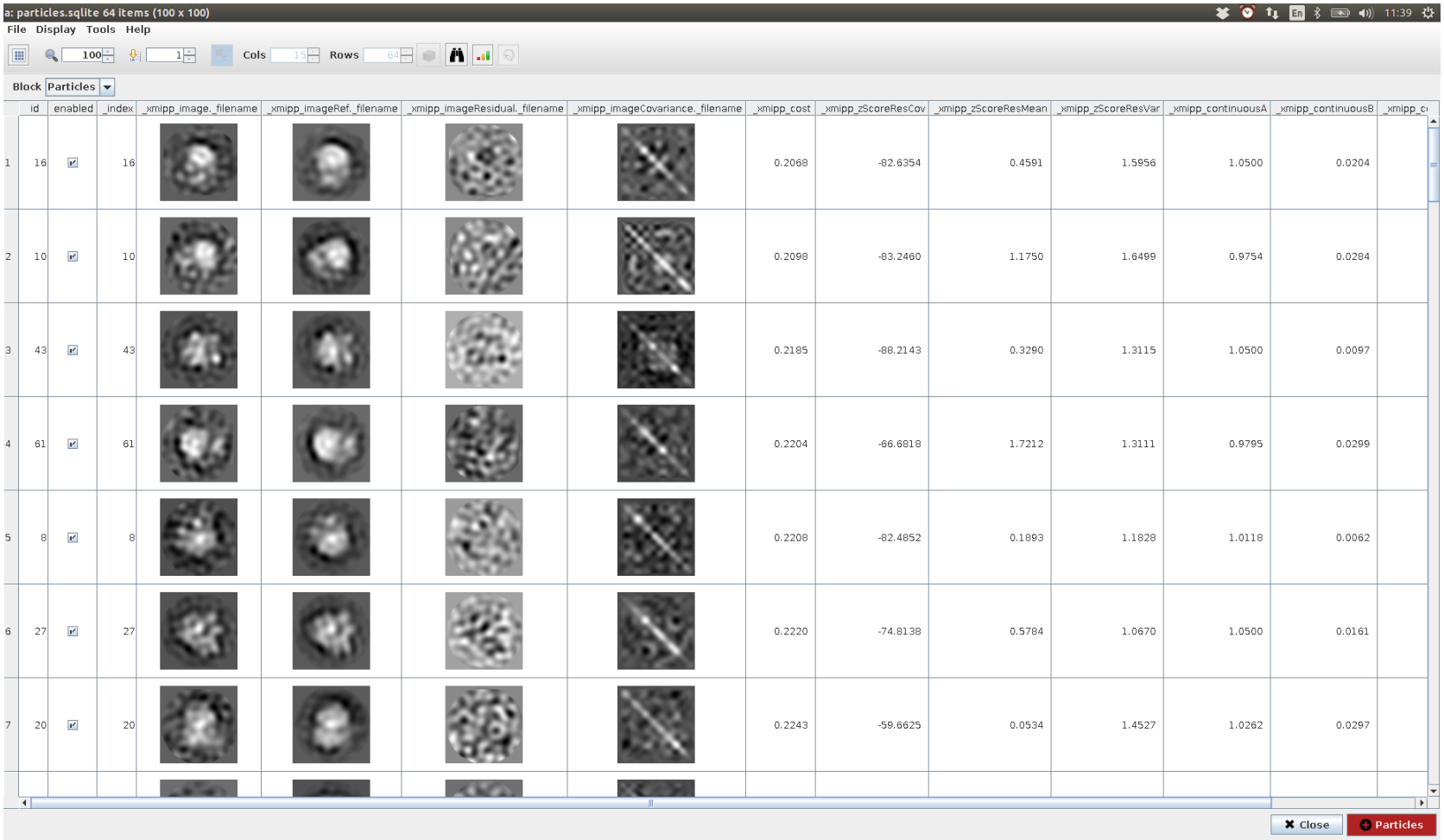
The initial volume is crucial



Eman | Significant

Scipion Menu: 3D Initial Volume

Reprojection consistency

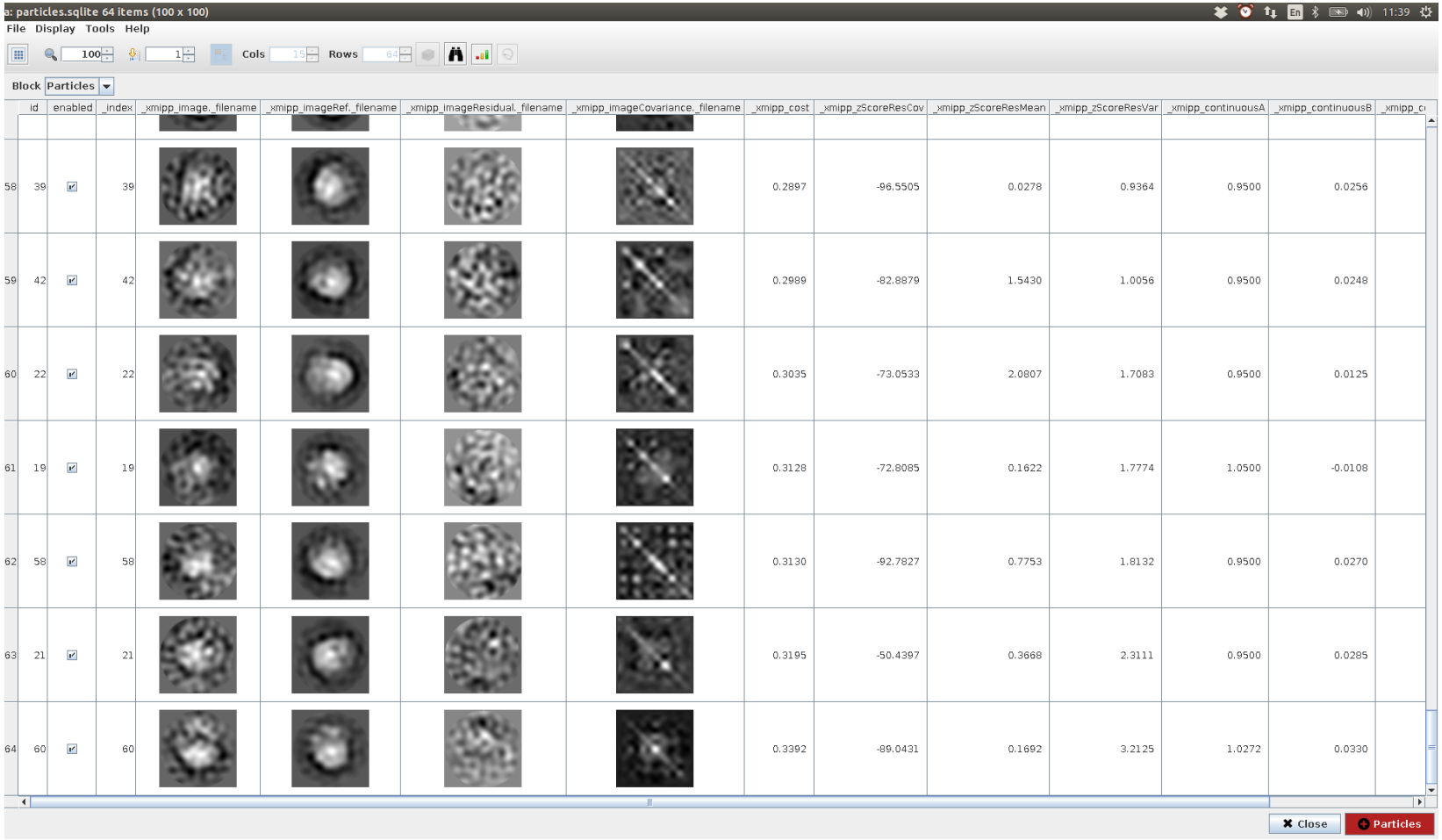


	id	enabled	_index	_xmipp_image._filename	_xmipp_imageRef._filename	_xmipp_imageResidual._filename	_xmipp_imageCovariance._filename	_xmipp_cost	_xmipp_zScoreResCov	_xmipp_zScoreResMean	_xmipp_zScoreResVar	_xmipp_continuousA	_xmipp_continuousB	_xmipp_c
1	16	☑	16					0.2068	-82.6354	0.4591	1.5956	1.0500	0.0204	
2	10	☑	10					0.2098	-83.2460	1.1750	1.6499	0.9754	0.0284	
3	43	☑	43					0.2185	-88.2143	0.3290	1.3115	1.0500	0.0097	
4	61	☑	61					0.2204	-66.6818	1.7212	1.3111	0.9795	0.0299	
5	8	☑	8					0.2208	-82.4852	0.1893	1.1828	1.0118	0.0062	
6	27	☑	27					0.2220	-74.8138	0.5784	1.0670	1.0500	0.0161	
7	20	☑	20					0.2243	-59.6625	0.0534	1.4527	1.0262	0.0297	

Best fitting particles

Scipion Prot: compare reprojections

Reprojection consistency

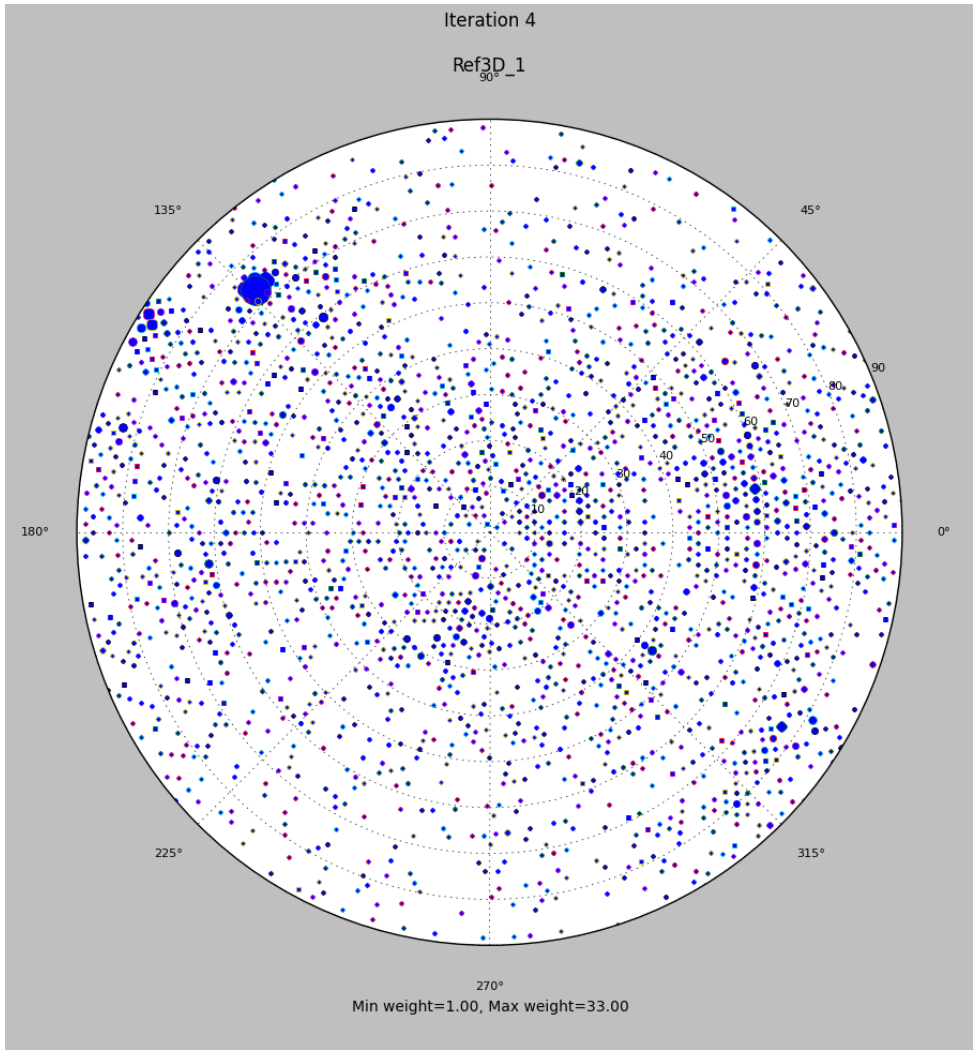


	id	enabled	_index	_xmipp_image._filename	_xmipp_imageRef._filename	_xmipp_imageResidual._filename	_xmipp_imageCovariance._filename	_xmipp_cost	_xmipp_zScoreResCov	_xmipp_zScoreResMean	_xmipp_zScoreResVar	_xmipp_continuousA	_xmipp_continuousB	_xmipp_c
58	39	☑	39					0.2897	-96.5505	0.0278	0.9364	0.9500	0.0256	
59	42	☑	42					0.2989	-82.8879	1.5430	1.0056	0.9500	0.0248	
60	22	☑	22					0.3035	-73.0533	2.0807	1.7083	0.9500	0.0125	
61	19	☑	19					0.3128	-72.8085	0.1622	1.7774	1.0500	-0.0108	
62	58	☑	58					0.3130	-92.7827	0.7753	1.8132	0.9500	0.0270	
63	21	☑	21					0.3195	-50.4397	0.3668	2.3111	0.9500	0.0285	
64	60	☑	60					0.3392	-89.0431	0.1692	3.2125	1.0272	0.0330	

Worst fitting particles

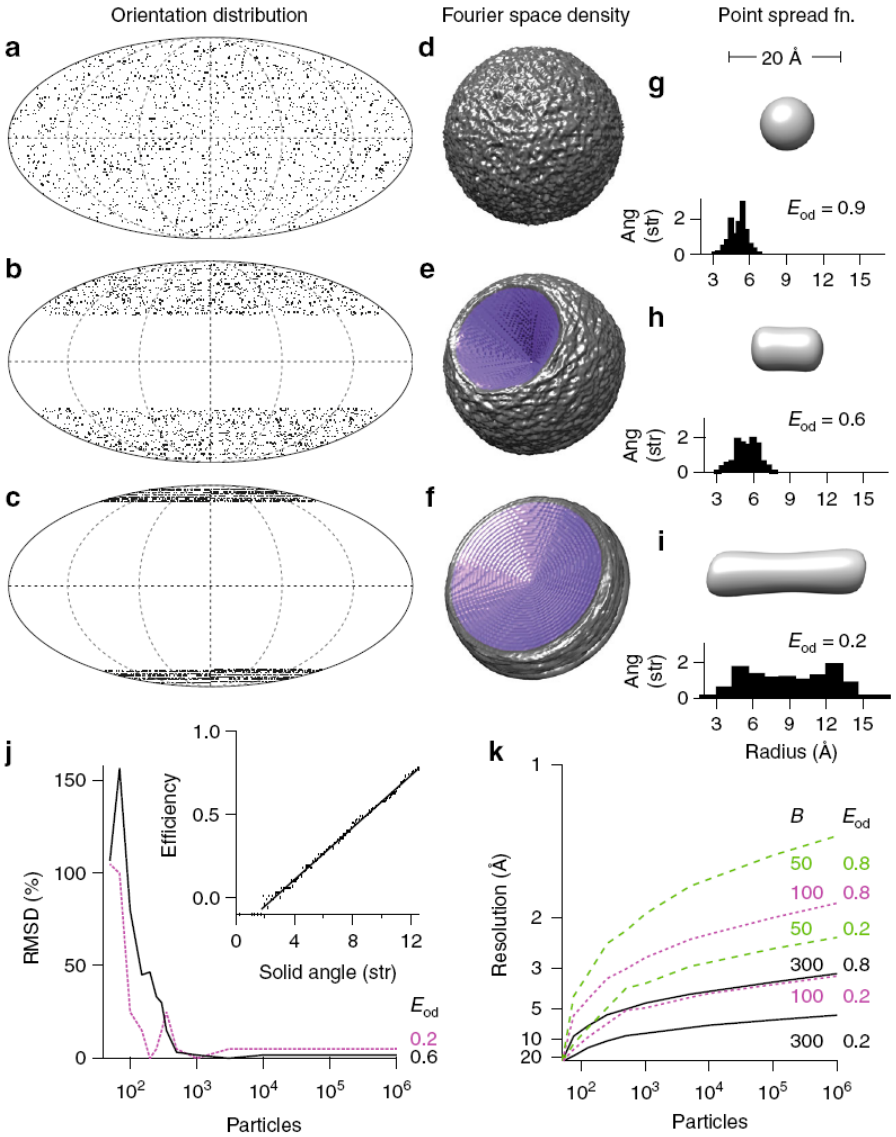
Scipion Prot: compare reprojections

Angular distribution



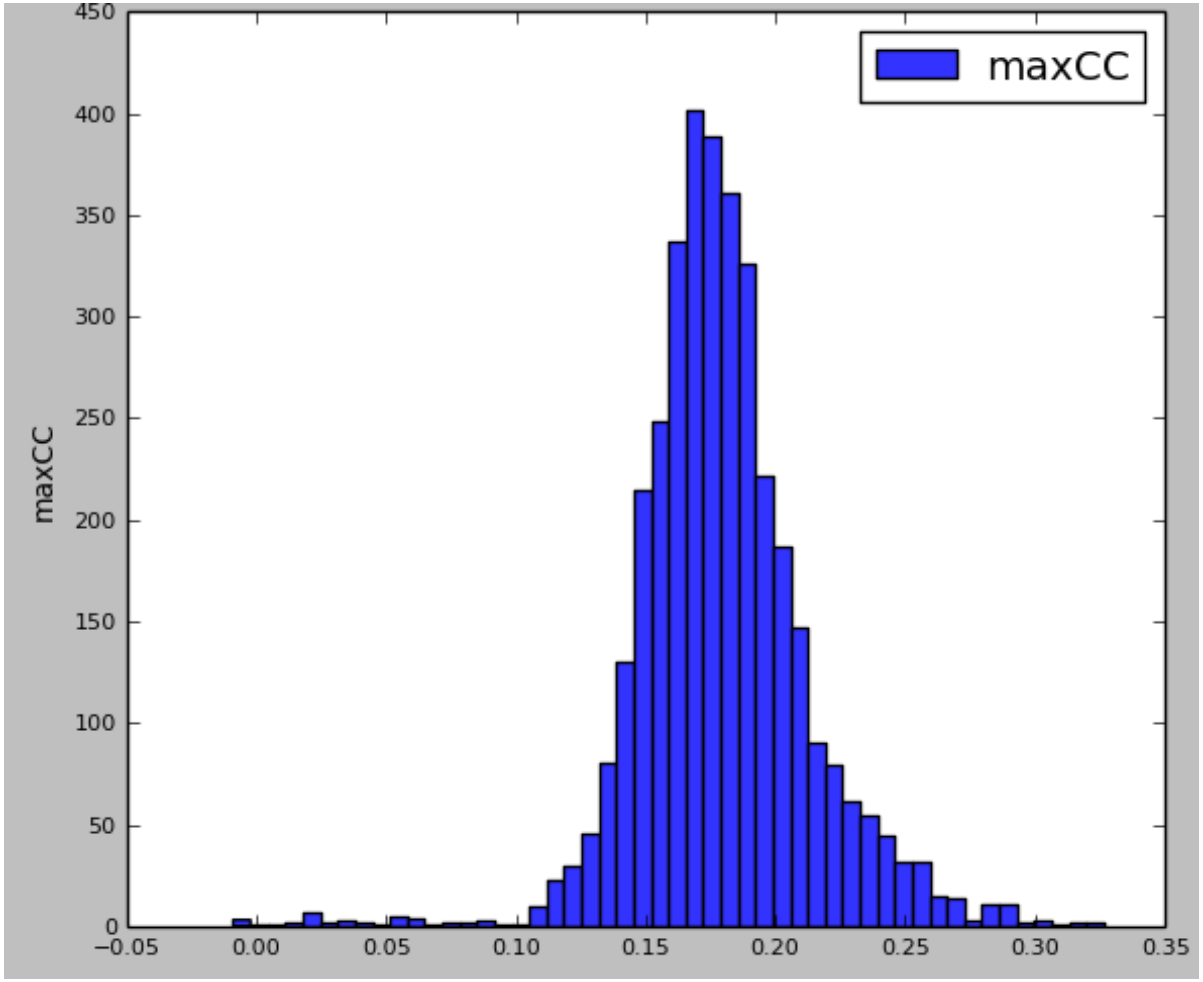
Scipion Prot: Projection Matching and most 3D Reconstruction

Angular distribution

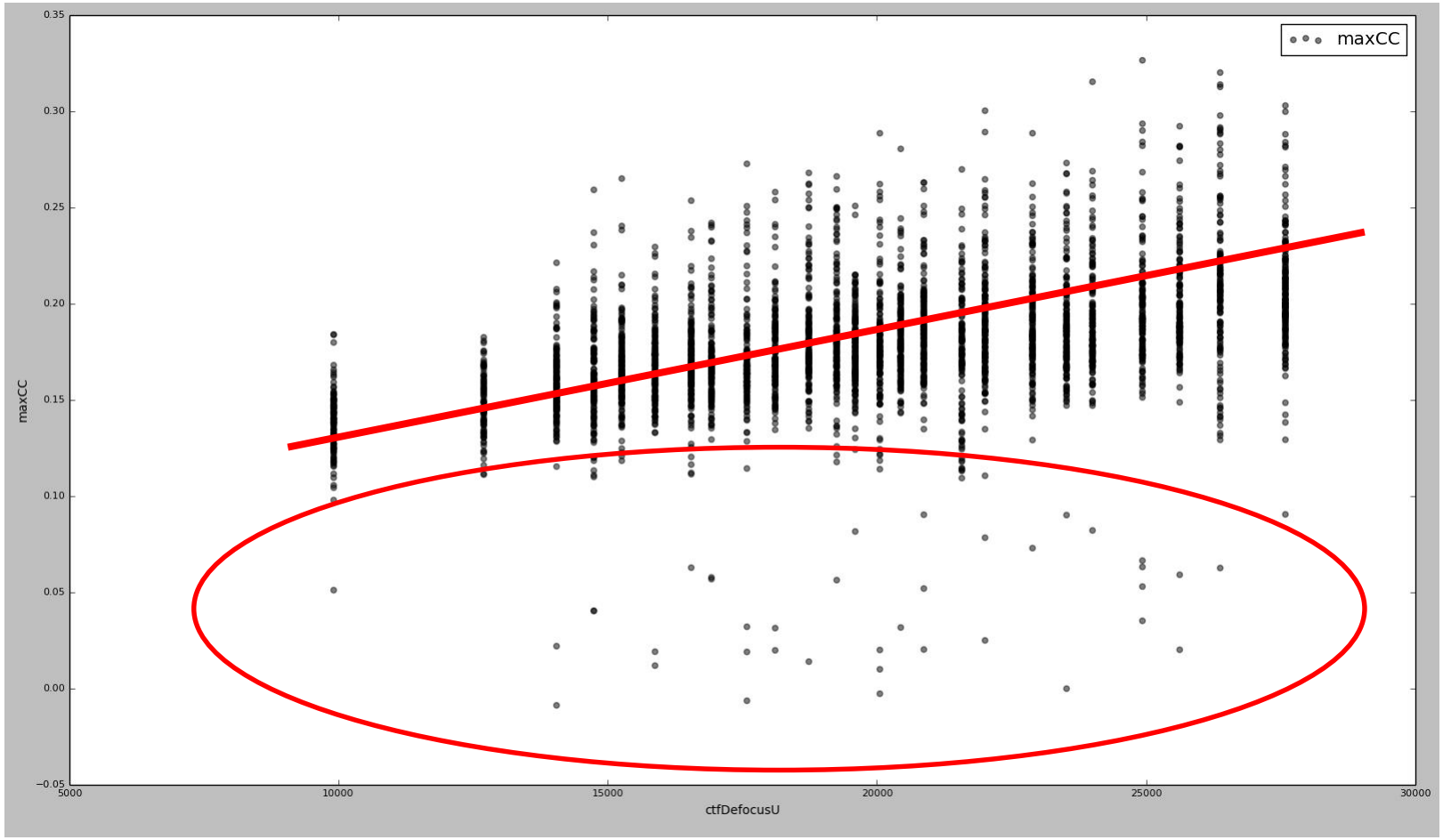


Naydenova, K. & Russo, C. J. Measuring the effects of particle orientation to improve the efficiency of electron cryomicroscopy. Nature communications, 2017, 8, 629

Correlation histograms



Correlation and defocus

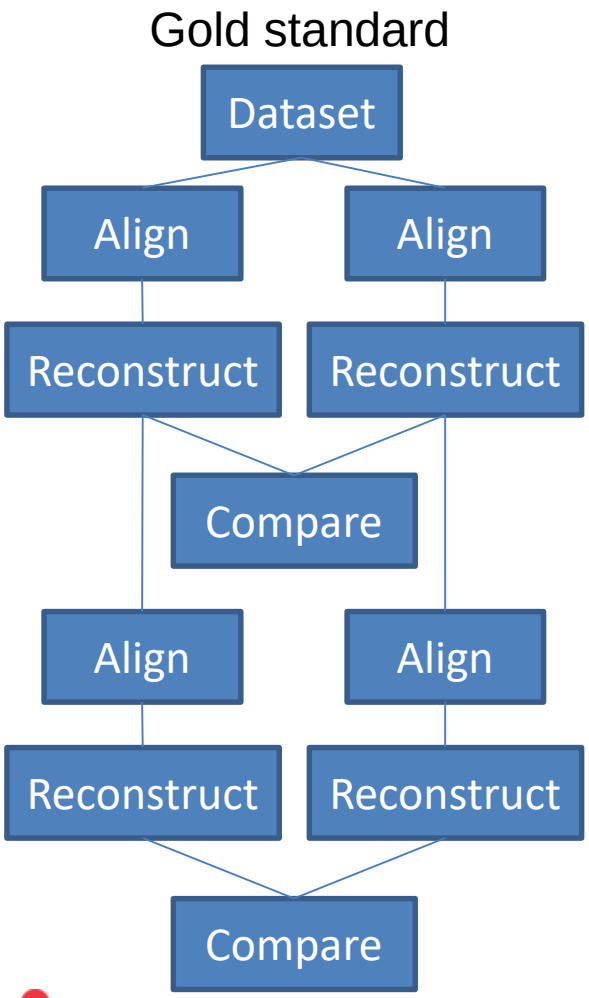


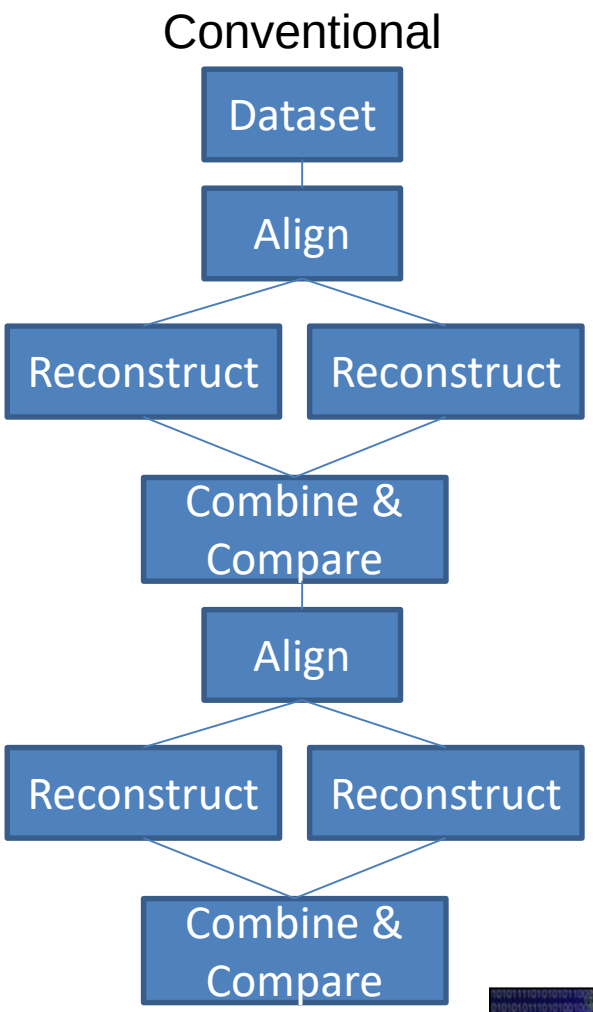
CSIC

instruct Integrating Biology

instruct image Processing Center

Reproducibility



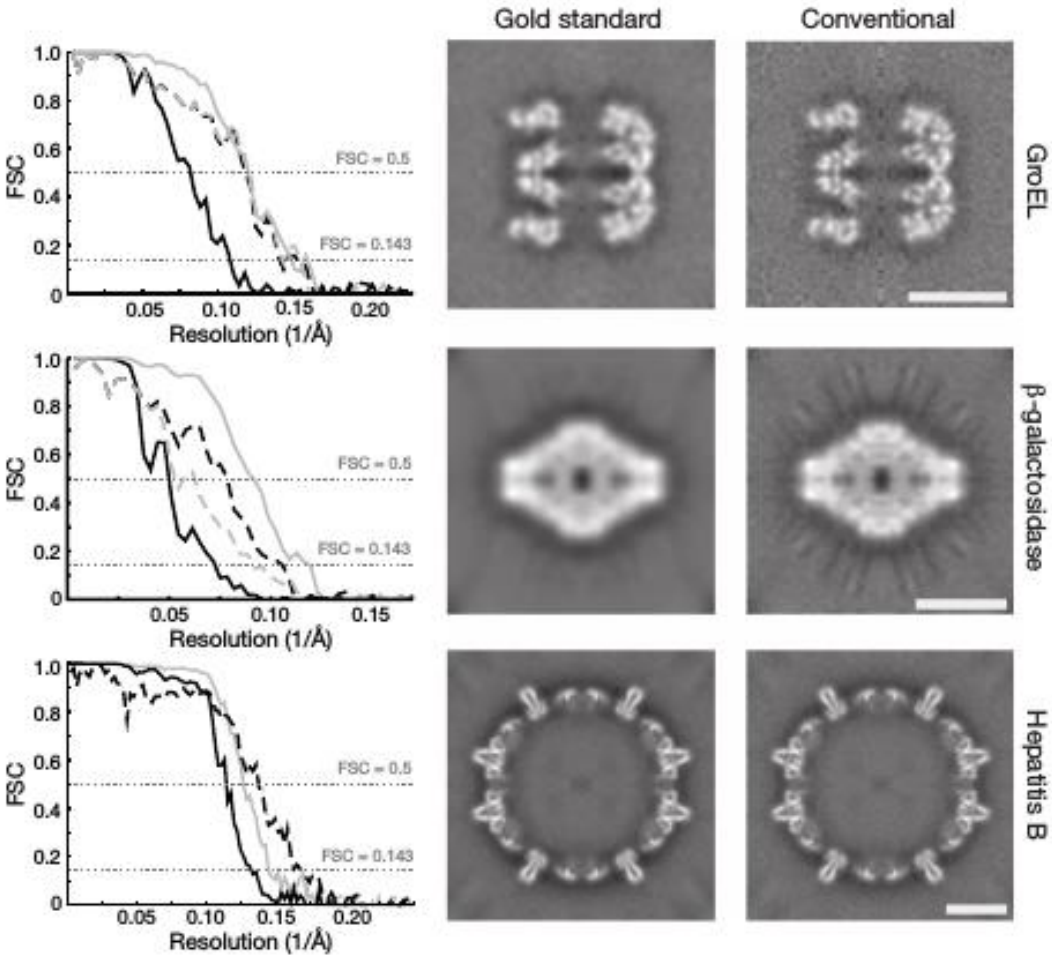


CSIC

instruct Integrating Biology

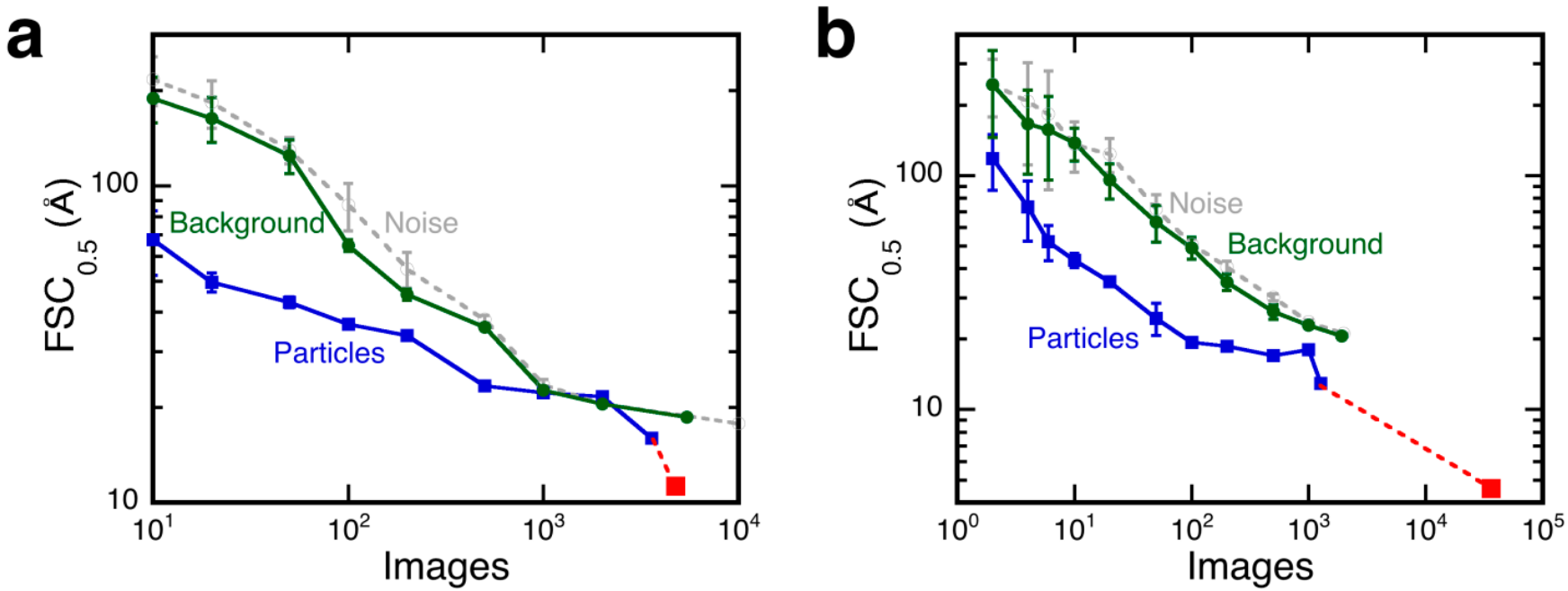
instruct image Processing Center

No overfitting



Scheres, S. H. W. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat Methods,* **2012***, 9*, 853-854

Better than noise

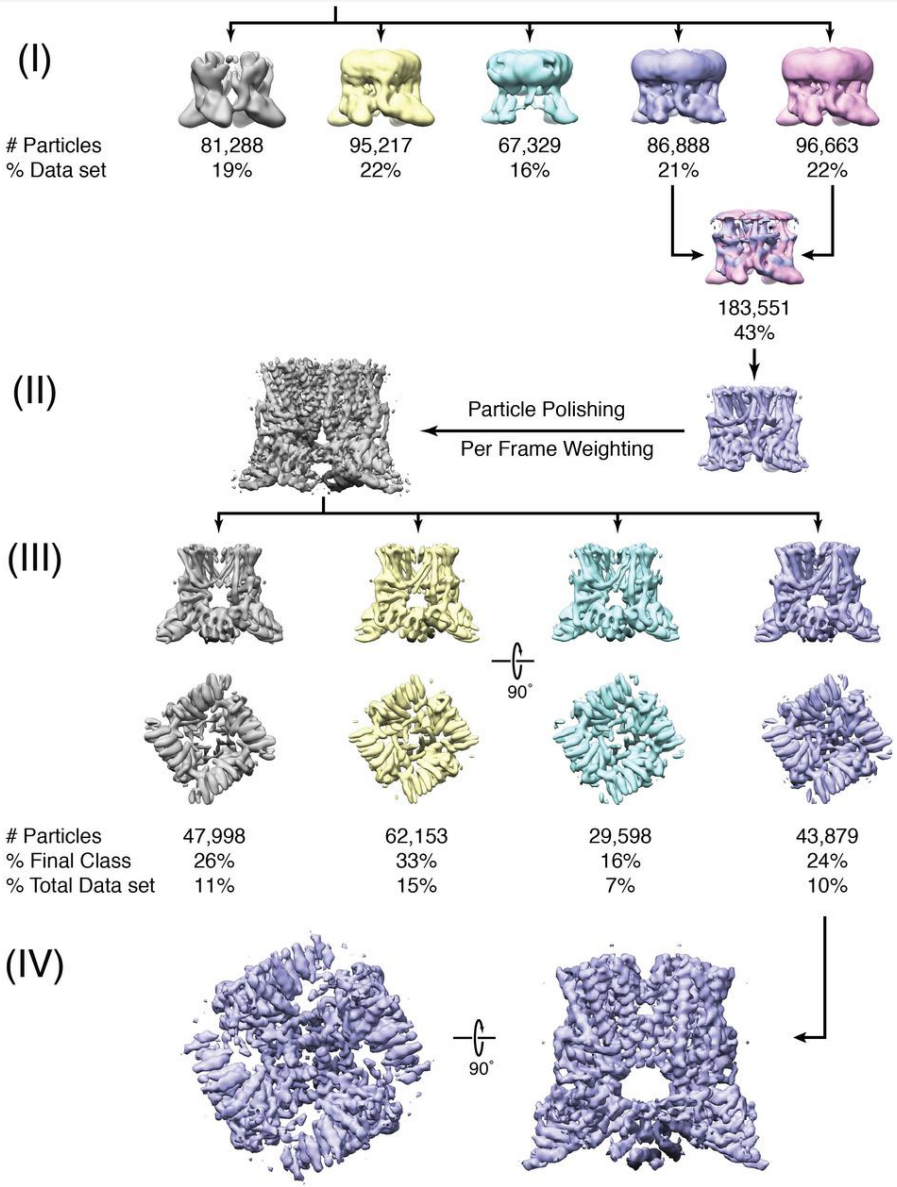


Scipion Prot: validate overfitting

Heymann, B. Validation of 3DEM Reconstructions: The phantom in the noise. *AIMS Biophysics,* **2015***, 2*, 21-35

Homogeneity

Zubcevic, L et al Cryo-electron microscopy structure of the TRPV2 ion channel. *Nat Struct Mol Biol,* **2016***, 23*, 180-186



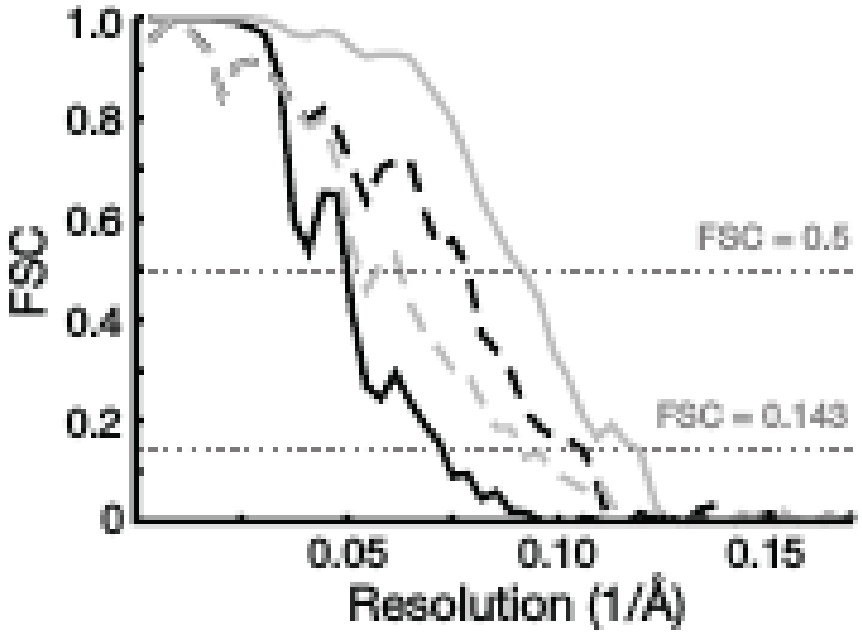
Scipion Menu: 3D Classification

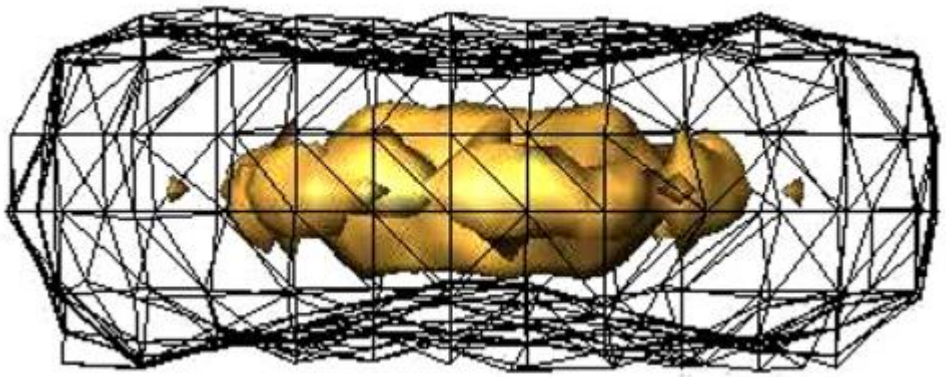
CSIC

instruct Integrating Biology

instruct image Processing Center

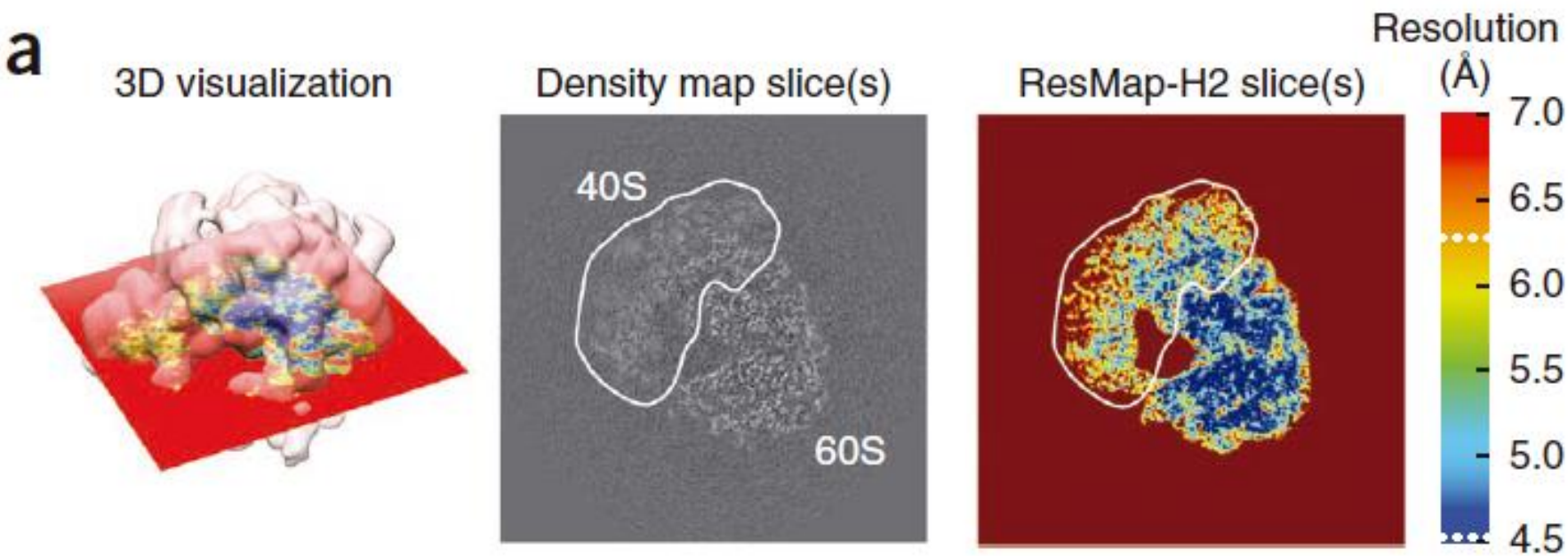
Global and directional resolution





Unser, M. et al. Spectral Signal-to-Noise Ratio and resolution assessment of 3D reconstructions
J. Structural Biology, **2005**, *149*, 243-255

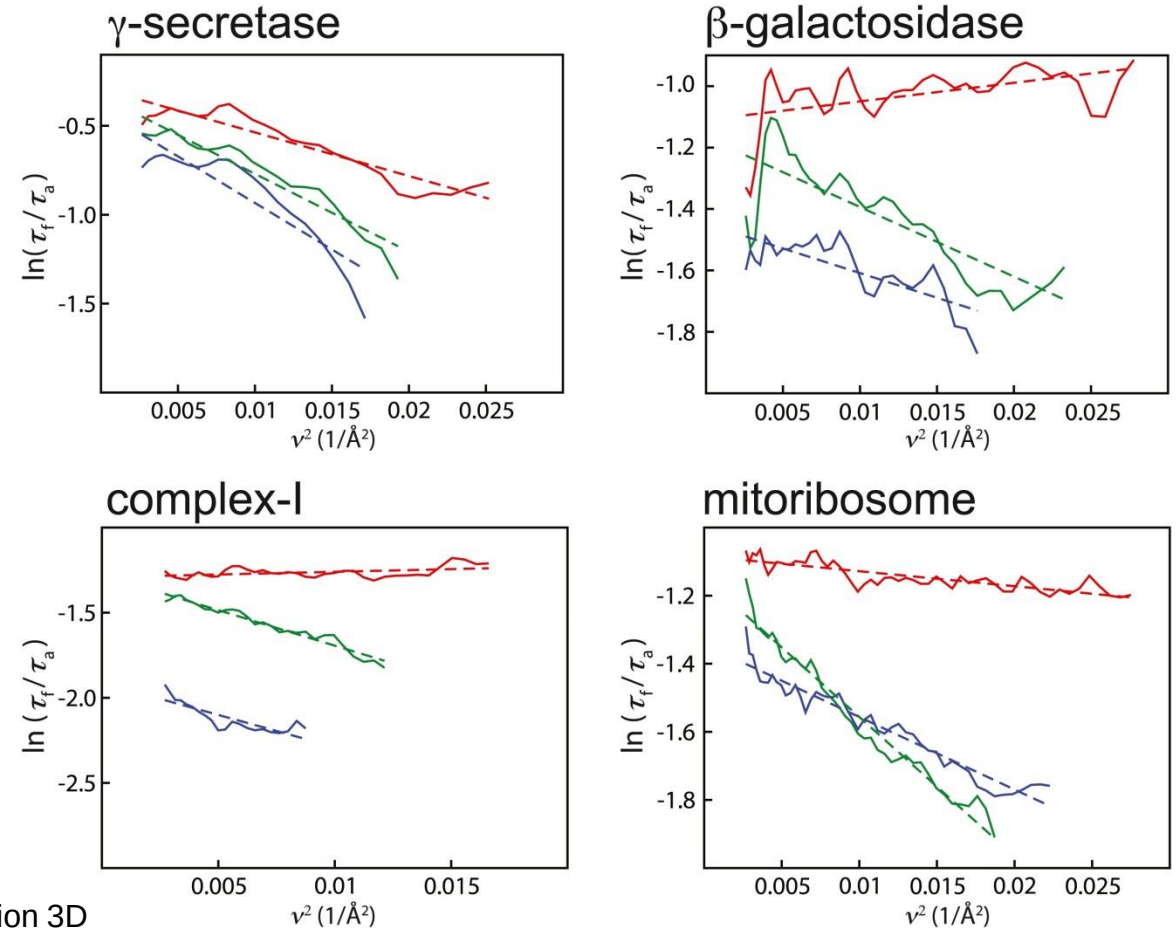
Local resolution



Scipion Prot: resmap

Kucukelbir, A.; Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat Methods*, **2014**, *11*, 63-65

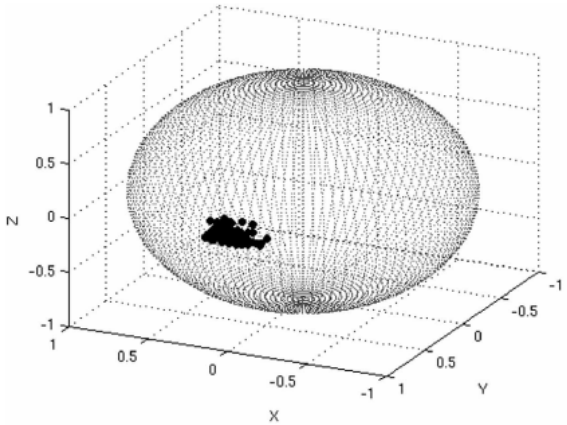
B-factor



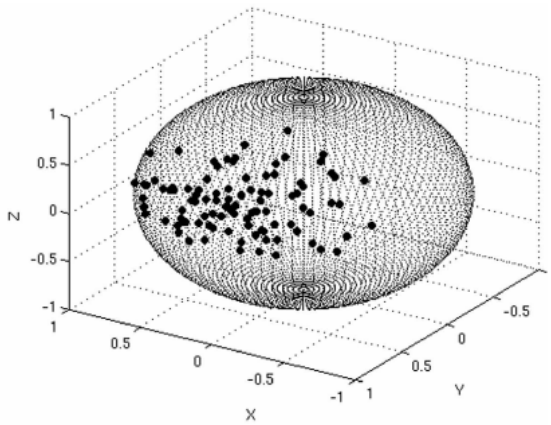
Scipion Prot: resolution 3D

Scheres, S. H. W. Beam-induced motion correction for sub-megadalton cryo-EM particles. *Elife,* **2014***, 3*, e03665

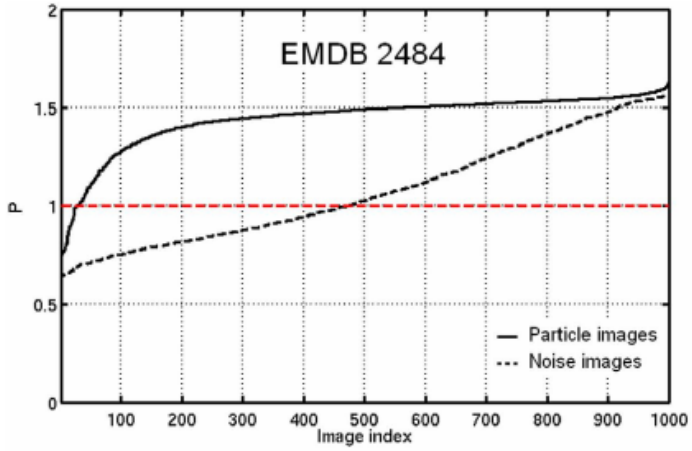
Alignability



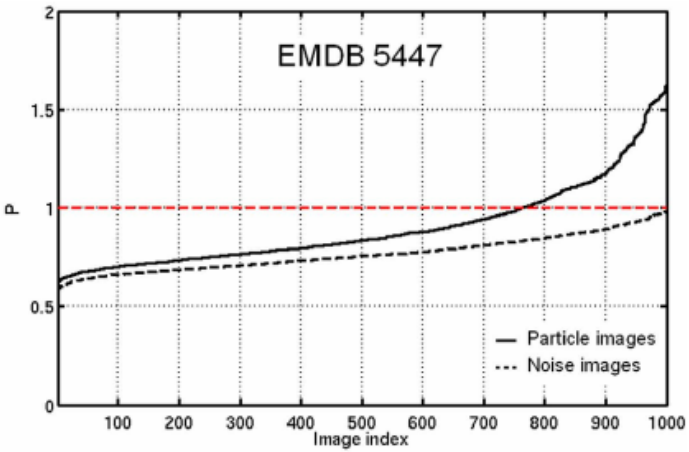
(a)



(b)



(a)



(b)

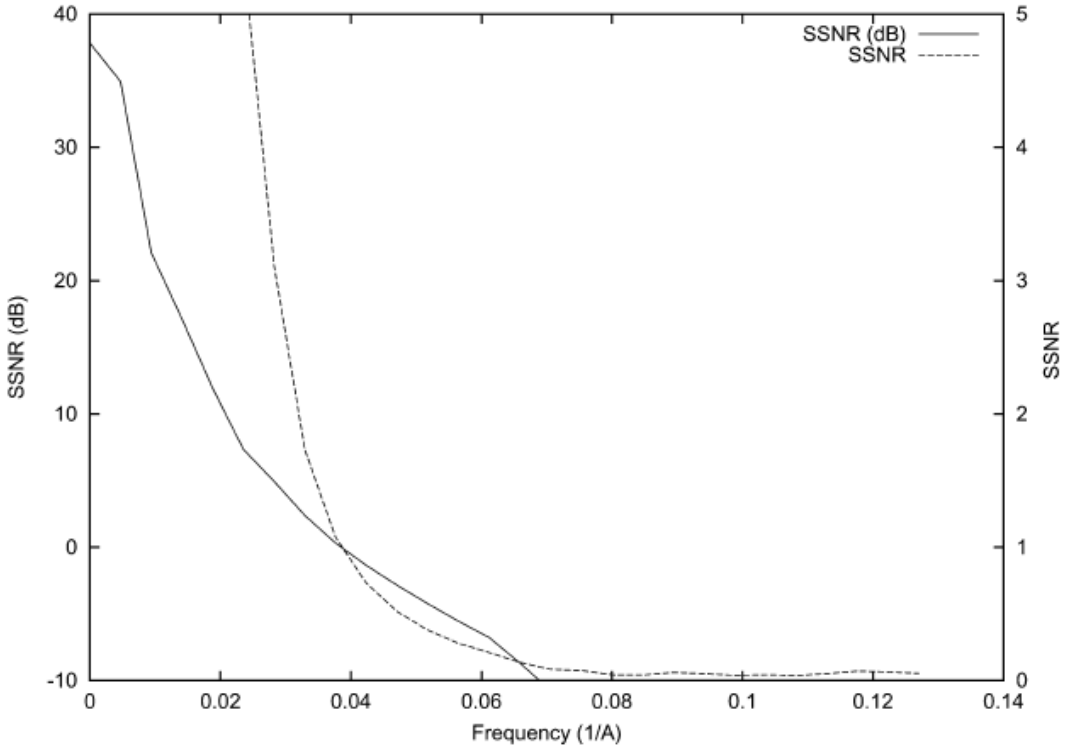
Scipion Prot: validate non-tilt

Vargas, J et al. Particle alignment reliability in single particle electron cryomicroscopy: a general approach. *Sci Rep,* **2016**, *6*, 21626

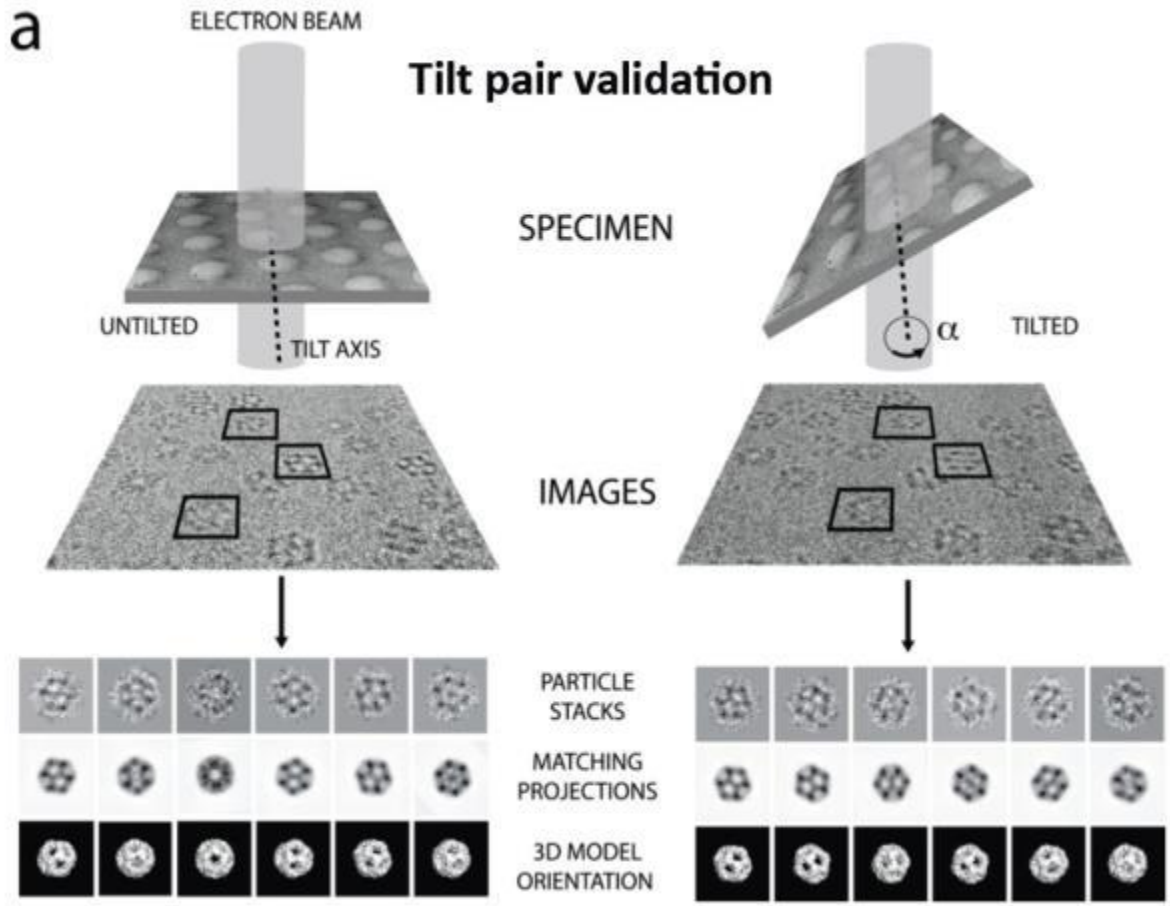
FSC vs SSNR

FSC(w)=Correlation between V_1 and V_2 bandpass filtered at frequency w

SSNR(w)=Power of V vs Power of noise at frequency w

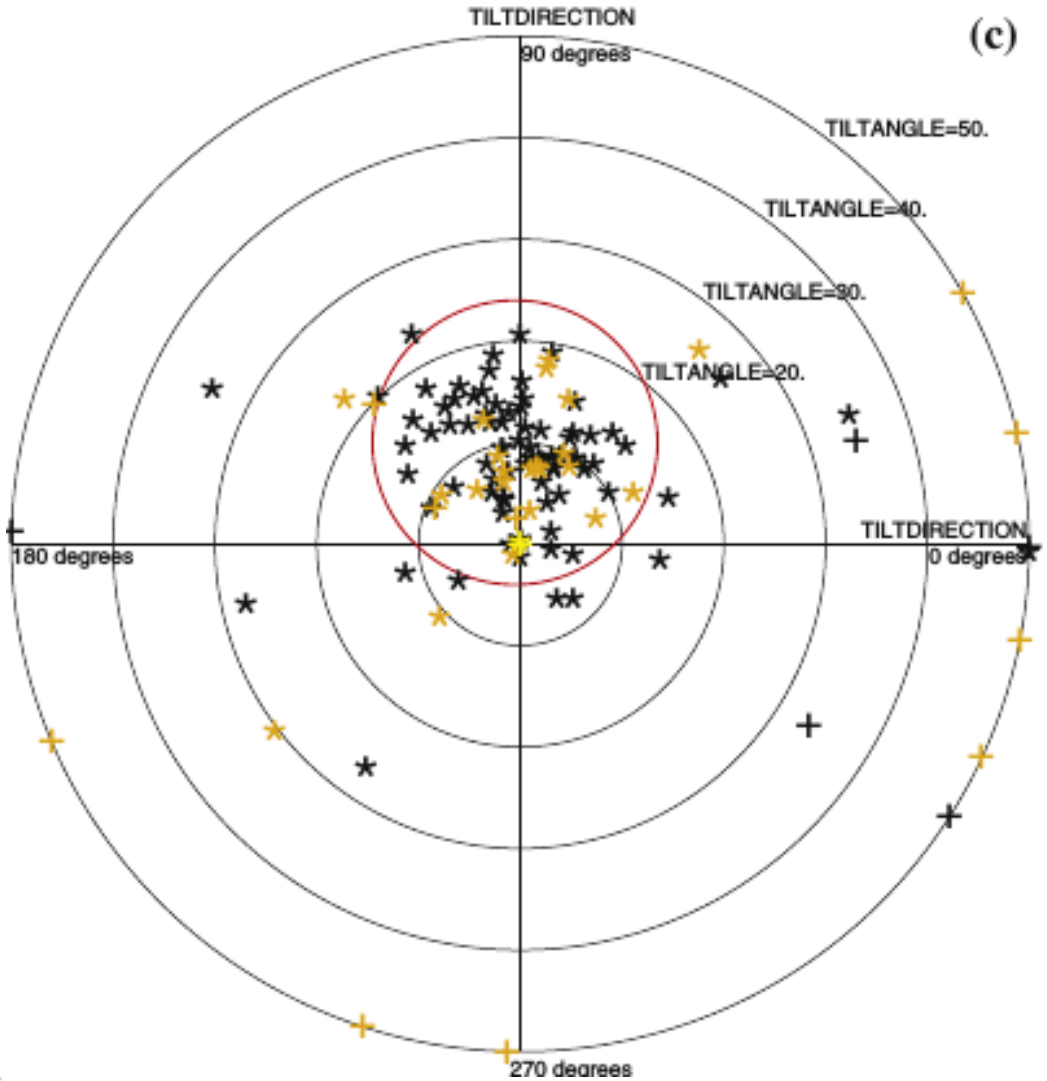


Tilt pair validation



Henderson, R. et al. Tilt-pair analysis of images from a range of different specimens in single-particle electron cryomicroscopy. *J Mol Biol,* **2011***, 413*, 1028-1046

Tilt pair validation



CSIC

instruct Integrating Biology

instruct image Processing Center

Tilt pair validation server

EMBL-EBI

Services | Research | Training | About us

Protein Data Bank in Europe

Bringing Structure to Biology

EM resources

Share | Feedback

EM Resources

- Home
- Statistics
- Validation
- EMDataBank
- EMPIAR
- Test data

EMDB

- Latest maps
- Latest headers
- Latest updates
- Search
- Browse
- FTP archive
- Deposit EM map/model
- EMDB data model

Tilt pair validation server

Welcome to the PDBe tilt pair validation server!
Tilt-pair validation analysis (Rosenthal and Henderson, 2003) can be used to assess the accuracy of initial angle assignment in single-particle processing. To perform this analysis you need to collect two corresponding sets of particle images - one untilted and the other tilted, then upload the stacks of images along with a 3D reconstruction based on the untilted images. This server is based on the Tilt-pair server developed at MRC National Institute for Medical Research (Wasilewski and Rosenthal, 2014), and we thank Sebastian Wasilewski and Peter Rosenthal for their help in developing and testing the current server.
You may upload map files in MRC or CCP4 format, and parameter files (containing Euler angles for individual particles) in Spider or Frealign format. We have some test data sets that you can use to try out the service here. We are still developing the server and appreciate your feedback!

Map (3D volume):	Browse...	No file selected.	
Untilted stack:	Browse...	No file selected.	
Orientation parameters for stack 1:	Browse...	No file selected.	Frealign
Tilted stack:	Browse...	No file selected.	
Pixel size (Å):			
Mask radius (pixels):			
Tilt search range (degrees):	20		
Resolution range (low to high; Å):	100		20
Email address:			
Job name:			
Perform CTF correction?			

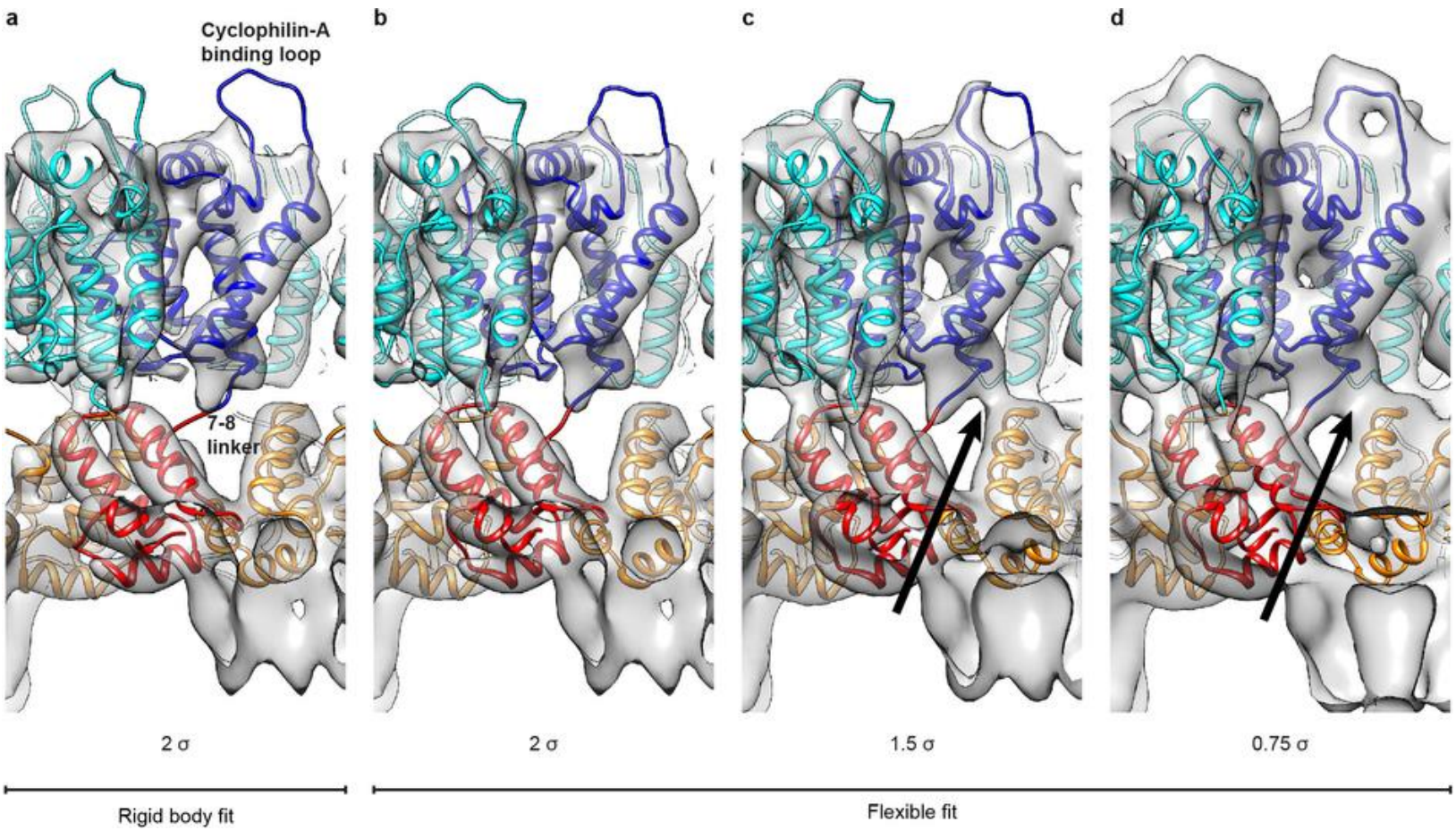
Compute

CSIC

instruct Integrating Biology

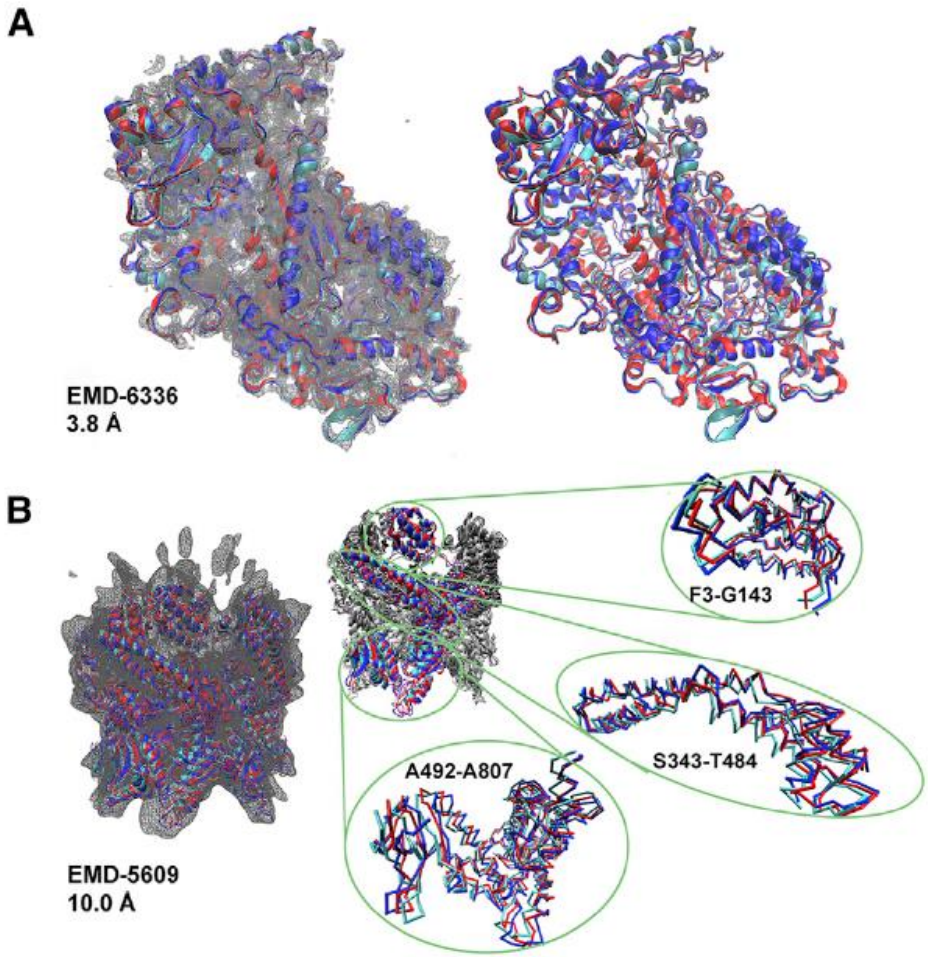
instruct image Processing Center

PDB Fitting



Schur, F. K. M. et al. Structure of the immature HIV-1 capsid in intact virus particles at 8.8 Å resolution. *Nature,* **2015***, 517*, 505-508

PDB Fitting Variability



- Monroe, L.; Terashi, G. & Kihara, D. Variability of Protein Structure Models from Electron Microscopy. Structure, 25:1-11, 2017
- Herzik, M. A.; Fraser, J. S. & Lander, G. C. A Multi-model Approach to Assessing Local and Global Cryo-EM Map Quality. Structure, 2019, 27, 344-358.e3

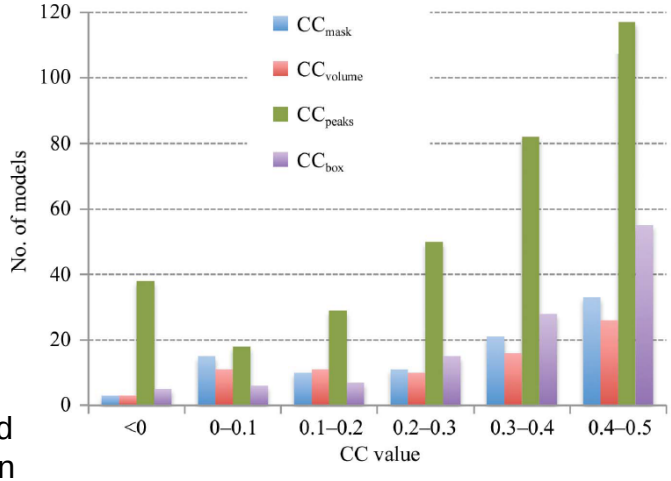
PDB Fitting Objective Quality

Table 2
Summary of map resolution estimates.

Metric	Objects used	Purpose	Values	Meaning, possible actions
d_{FSC}	Half-maps	Highest resolution at which the experimental data are confident	The higher the better	Resolution determined using half-maps method
d_{99}	Map	Resolution cutoff beyond which Fourier coefficients are negligibly small	$d_{99} \geq d_{FSC}$	Expected values
			$d_{99} < d_{FSC}$	Verify d_{FSC}; omit coefficients with $d_{99} \leq d < d_{FSC}$
			$d_{99} >> d_{FSC}$	Sharpen the map
d_{model}	Map and model	Resolution cutoff at which the model map is the most similar to the target map	$d_{model} \geq d_{FSC}$	Expected values
			$d_{model} < d_{FSC}$	Verify d_{FSC}; check ADP (too large?); validate map details
			$d_{model} >> d_{FSC}$	Sharpen the map
			$d_{model} << d_{99}$	Check ADP (too large?)
			$d_{model} >> d_{99}$	Check ADP (too small?); check the model
d_{FSC_model}	Map and model	Resolution cutoff up to which the model and map Fourier coefficients are similar	$d_{FSC\_model} \geq d_{FSC}$	Expected values
			$d_{FSC\_model} < d_{FSC}$	Verify d_{FSC}; omit coefficients with $d_{FSC\_model} \leq d < d_{FSC}$
			$d_{FSC\_model} \geq d_{FSC}$	Sharpen the map
			$d_{FSC\_model} >> d_{model}$	Omit coefficients with $d_{model} \leq d < d_{FSC\_model}$
			$d_{FSC\_model} << d_{model}$	Sharpen the map

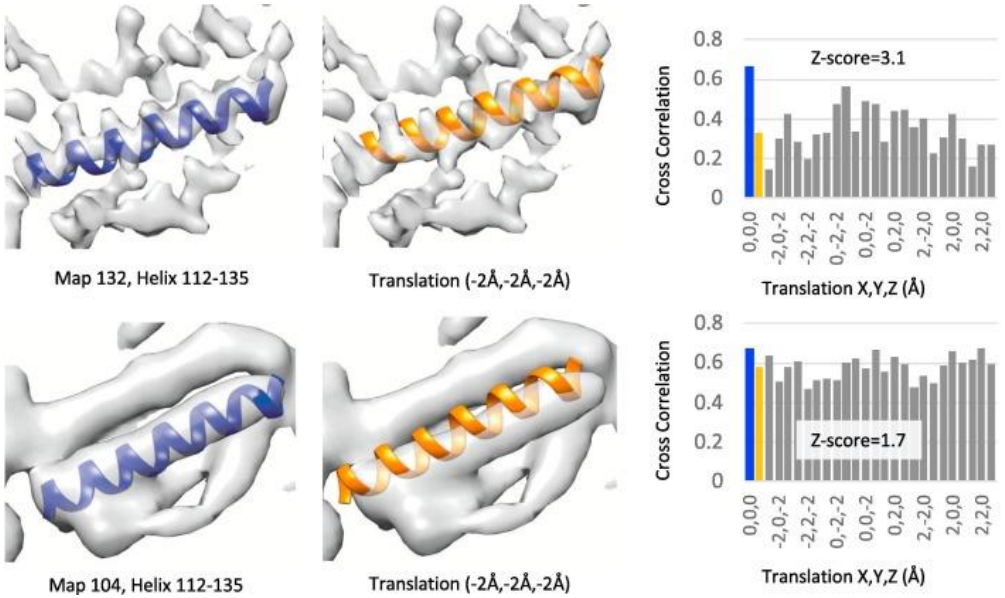
Table 3
Summary of map correlation coefficients used in this work.

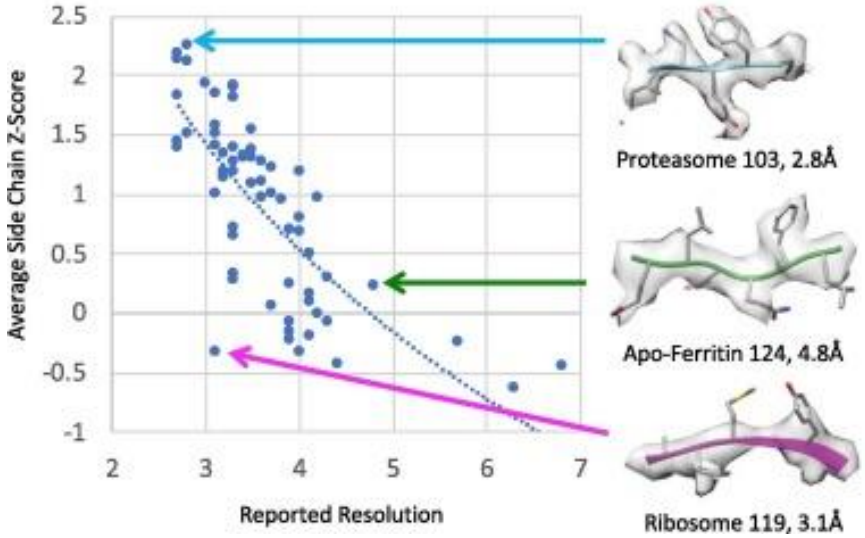
Metric	Region of the map used in calculation
CC_{box}	Whole map
CC_{mask}	Jiang & Brünger (1994) mask with a fixed radius
CC_{volume}	Mask of points with the highest values in the model map
CC_{peaks}	Mask of points with the highest values in the model and in the target maps
CC_{vr_mask}	Same as CC_{mask} but atomic radii are variable and function of resolution, atom type and ADP



Afonine, P. V.; Klaholz, B. P.; Moriarty, N. W.; Poon, B. K.; Sobolev, O. V.; Terwilliger, T. C.; Adams, P. D. & Urzhumtsev, A. New tools for the analysis and validation of cryo-EM maps and atomic models. Acta crystallographica. Section D, Structural biology, 2018 , 74 , 814-840

PDB Fitting Objective Quality

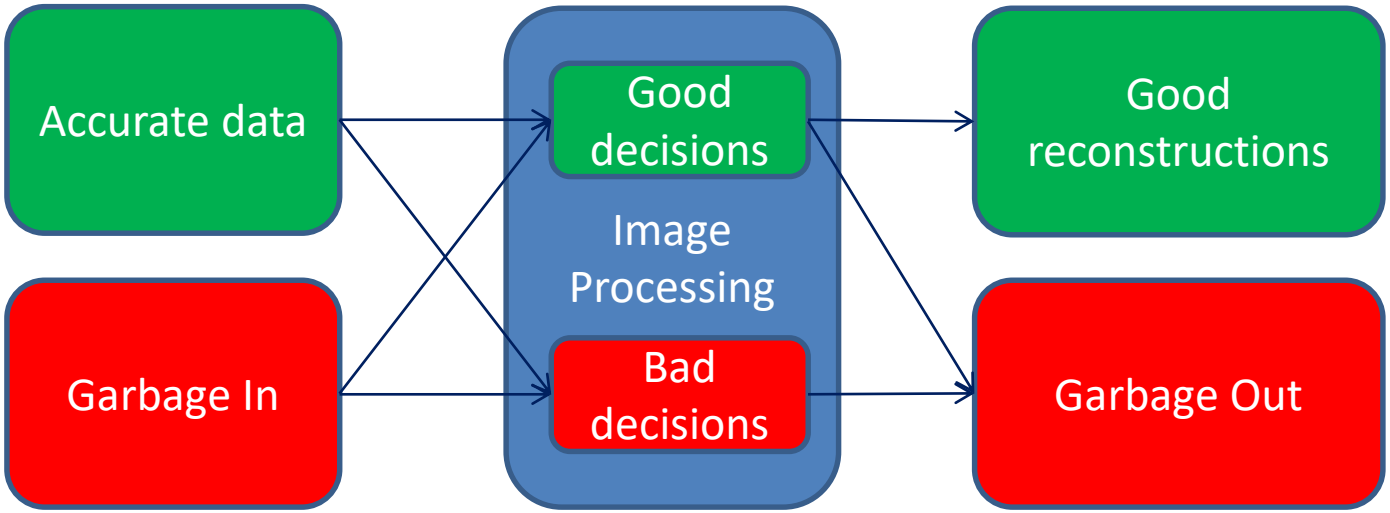




Pintilie, G. and Chiu, W. Assessment of structural features in Cryo-EM density maps using SSE and side chain Z-scores. Journal of structural biology, 2018, 204, 564-571

EM is a GIGO system

- The good thing about EM is that <u>it always gives a volume</u>
- The bad thing about EM is that <u>it always gives a volume</u>



3D Heterogeneity

Carlos Oscar S. Sorzano
Instruct Image Processing Center

CSIC

instruct Integrating Biology

instruct image Processing Center

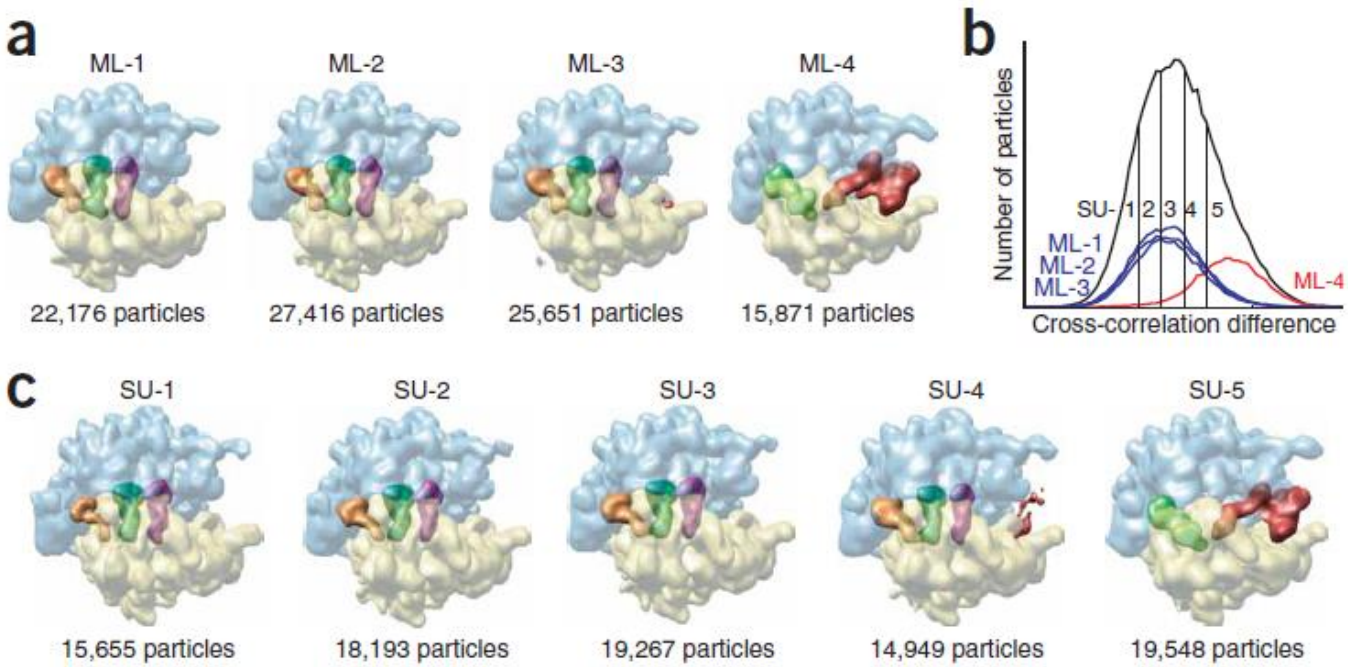
Maximum Likelihood (data fidelity)

Class Representative

$$L(\Theta) = \sum_{i=1}^{I} \ln \sum_{k=1}^{K} \int_{\varphi} P(X_i|k, \varphi, \Theta) P(k, \varphi|\Theta) d\varphi$$

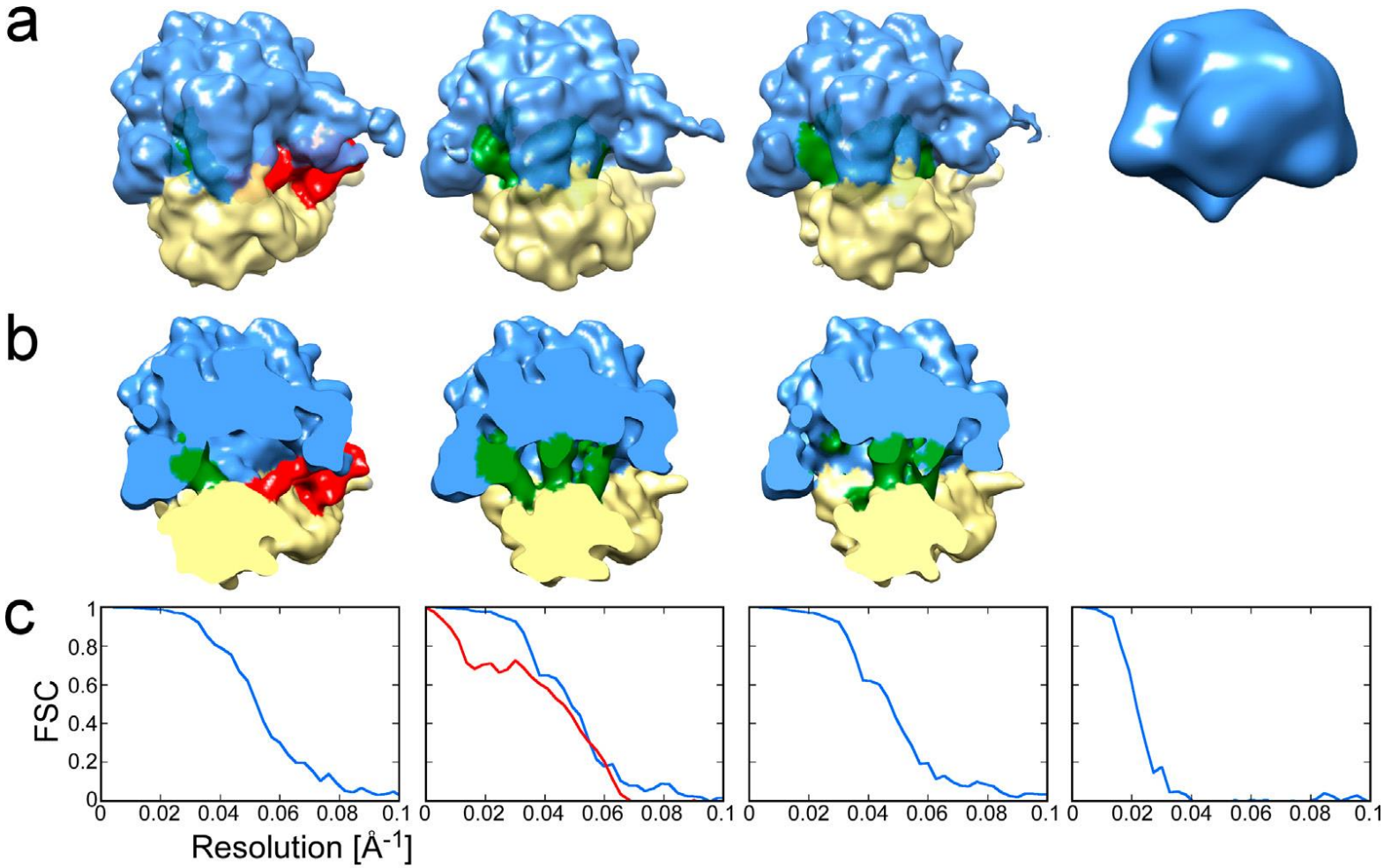
Images

Class & Orientation



Scheres, S. H. W.; Gao, H.; Valle, M.; Herman, G. T.; Eggermont, P. P. B.; Frank, J. & Carazo, J. M. Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. Nature Methods, 2007, 4, 27-29

Maximum Likelihood (data fidelity)

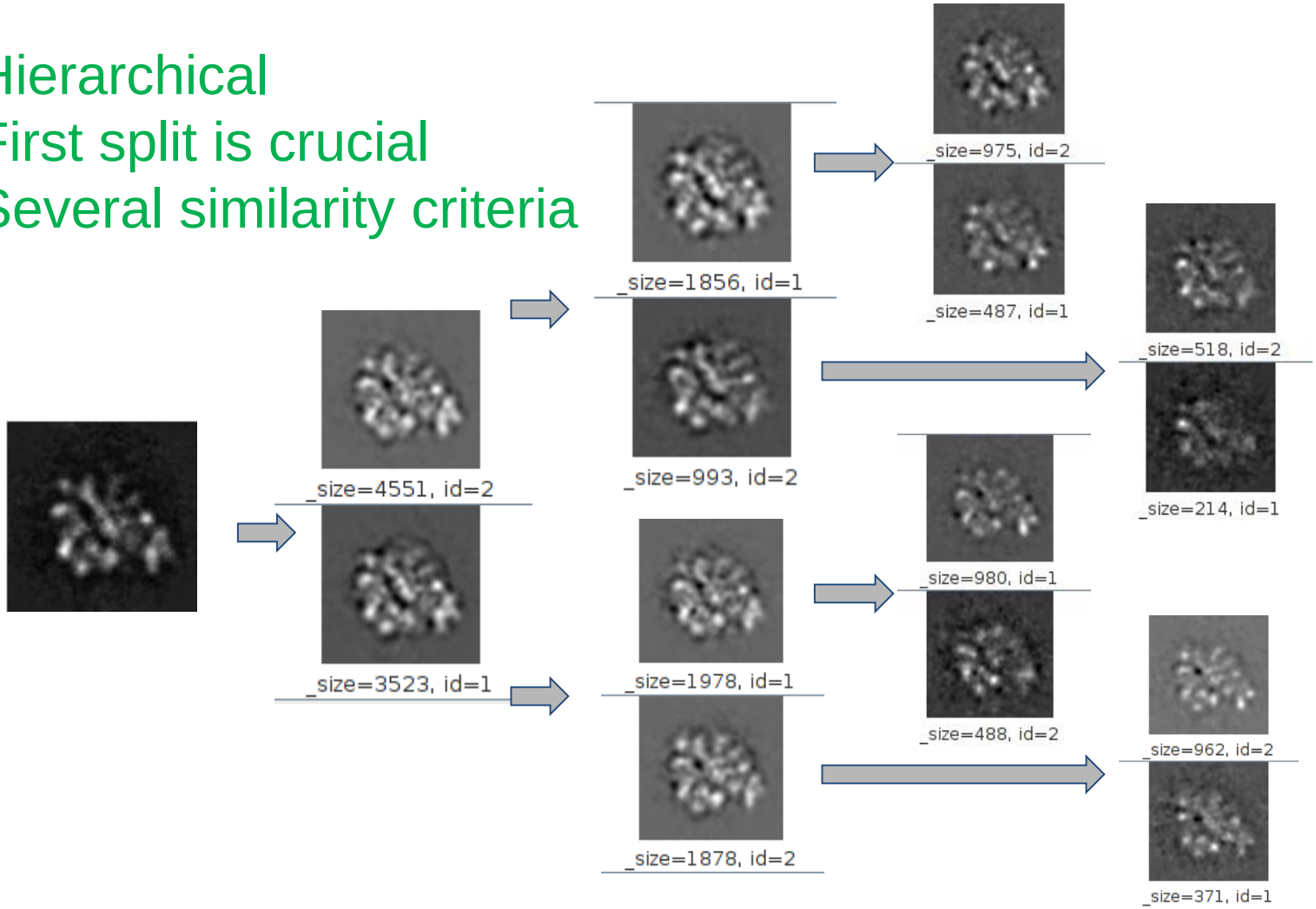


43% (17 Å)	23% (18 Å)	24% (18 Å)	10% (40 Å)

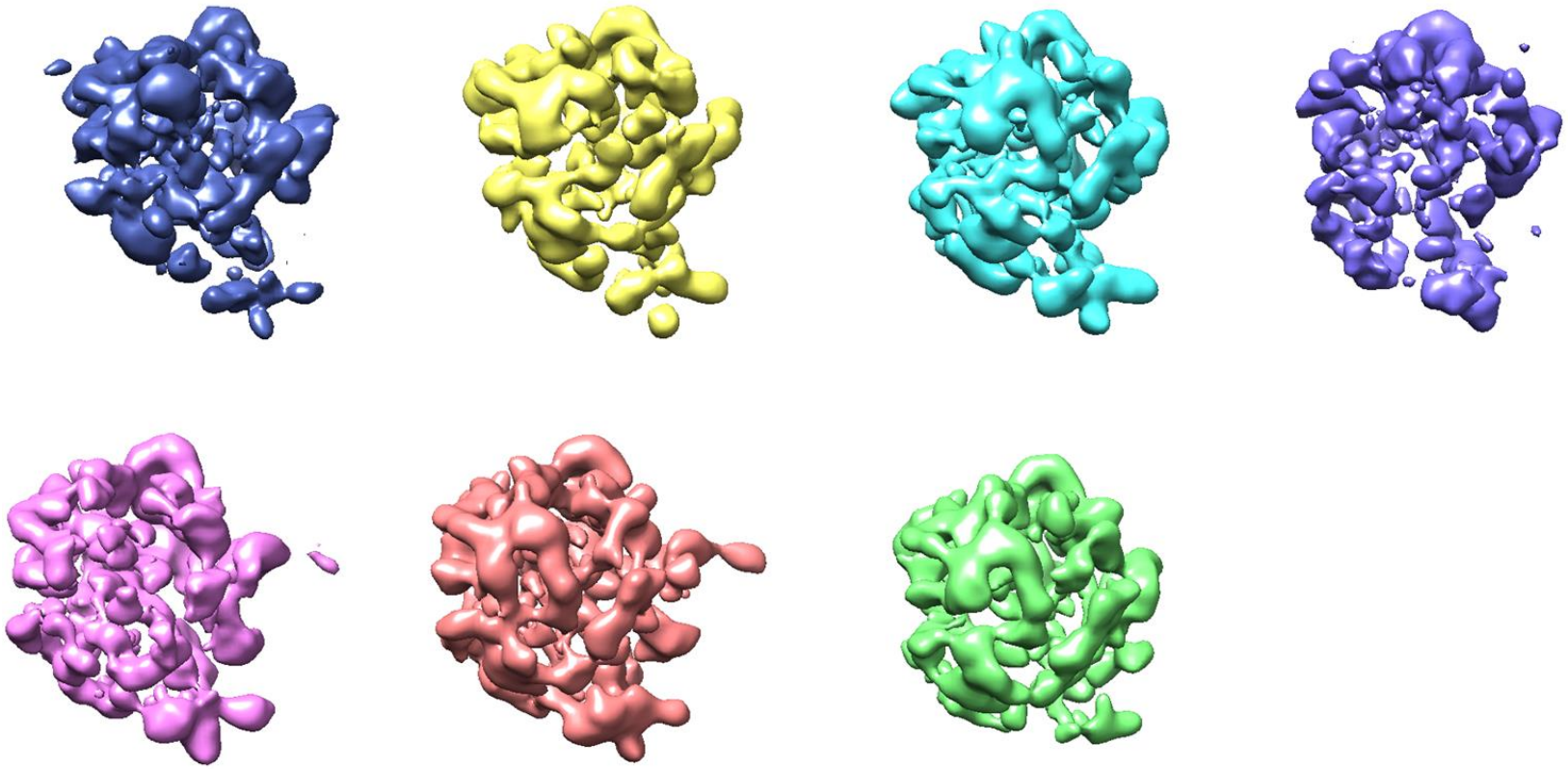
Lyumkis, D.; Brilot, A. F.; Theobald, D. L. & Grigorieff, N. Likelihood-based classification of cryo-EM images using FREALIGN. J. Structural Biology, 2013, 183, 377-388

CL3D

- Hierarchical
- First split is crucial
- Several similarity criteria



CL3D



Bayesian (a priori)

$$P(\Theta|X,Y) \propto P(X|\Theta,Y)P(\Theta|Y)$$

$$c_{MAP} = \underset{c \in C}{\operatorname{argmax}}\, P(c|d)$$

MAP is "maximum a posteriori" = most likely class

$$= \underset{c \in C}{\operatorname{argmax}} \frac{P(d|c)P(c)}{P(d)}$$

Bayes Rule

$$= \underset{c \in C}{\operatorname{argmax}} \underbrace{P(d|c)}_{\text{Data fidelity}}\underbrace{P(c)}_{\text{Prior}}$$

Dropping the denominator

Scheres, S. H. W. A Bayesian view on cryo-EM structure determination. J. Molecular Biology, 2012, 415, 406-418

Bayesian (a priori)

Calculation of the prior relies on the assumption of smoothness in the reconstruction. Smoothness is encoded in the assumption that all Fourier components V_l are independent and Gaussian distributed with zero mean and unknown variance τ_l^2, so that:

$$P(\Theta|Y) = \prod_{l=1}^{L} \frac{1}{2\pi\tau_l^2} \exp\left(\frac{|V_l|^2}{-2\tau_l^2}\right) \qquad (8)$$

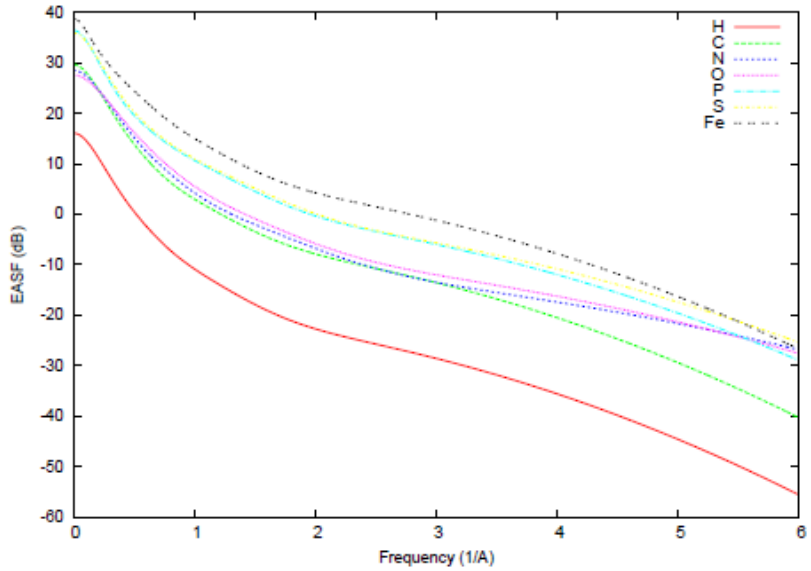
$$V_l^{(n+1)} = \frac{\sum_{i=1}^{N} \int_{\phi} \Gamma_{i\phi}^{(n)} \sum_{j=1}^{J} \mathbf{P}^{\phi_{lj}^T} \frac{\mathrm{CTF}_{ij} X_{ij}}{\sigma_{ij}^{2(n)}} d\phi}{\sum_{i=1}^{N} \int_{\phi} \Gamma_{i\phi}^{(n)} \sum_{j=1}^{J} \mathbf{P}^{\phi_{lj}^T} \frac{\mathrm{CTF}_{ij}^2}{\sigma_{ij}^{2(n)}} d\phi + \frac{1}{\tau_l^{2(n)}}}$$

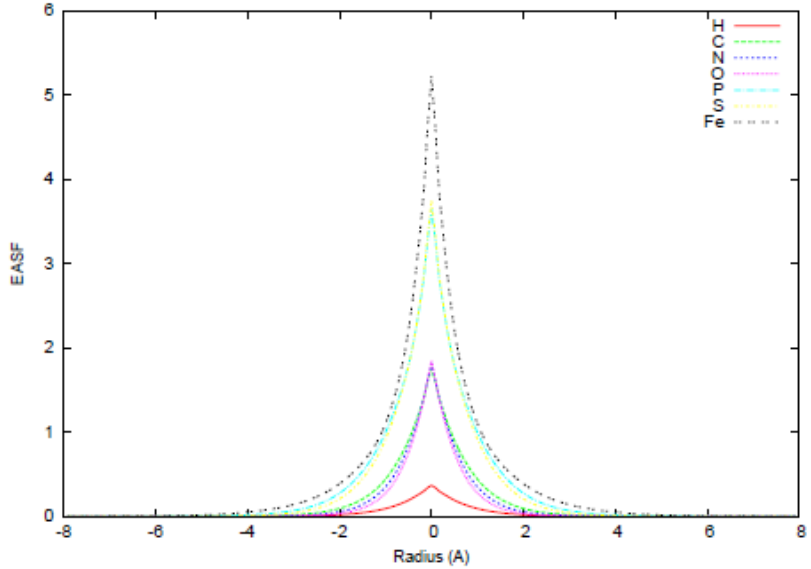
τ ↓→V↓

estimates for τ_l^2 were multiplied by a constant, $T=4$, in an attempt to account for the correlations between Fourier components in the signal. As expected, values of T close to 1 were observed to yield reconstructions with suboptimal resolutions, whereas for values larger than four, noticeable amounts of overfitting were observed (results not shown). One

Scheres, S. H. W. A Bayesian view on cryo-EM structure determination. J. Molecular Biology, 2012, 415, 406-418

Bayesian (a priori)



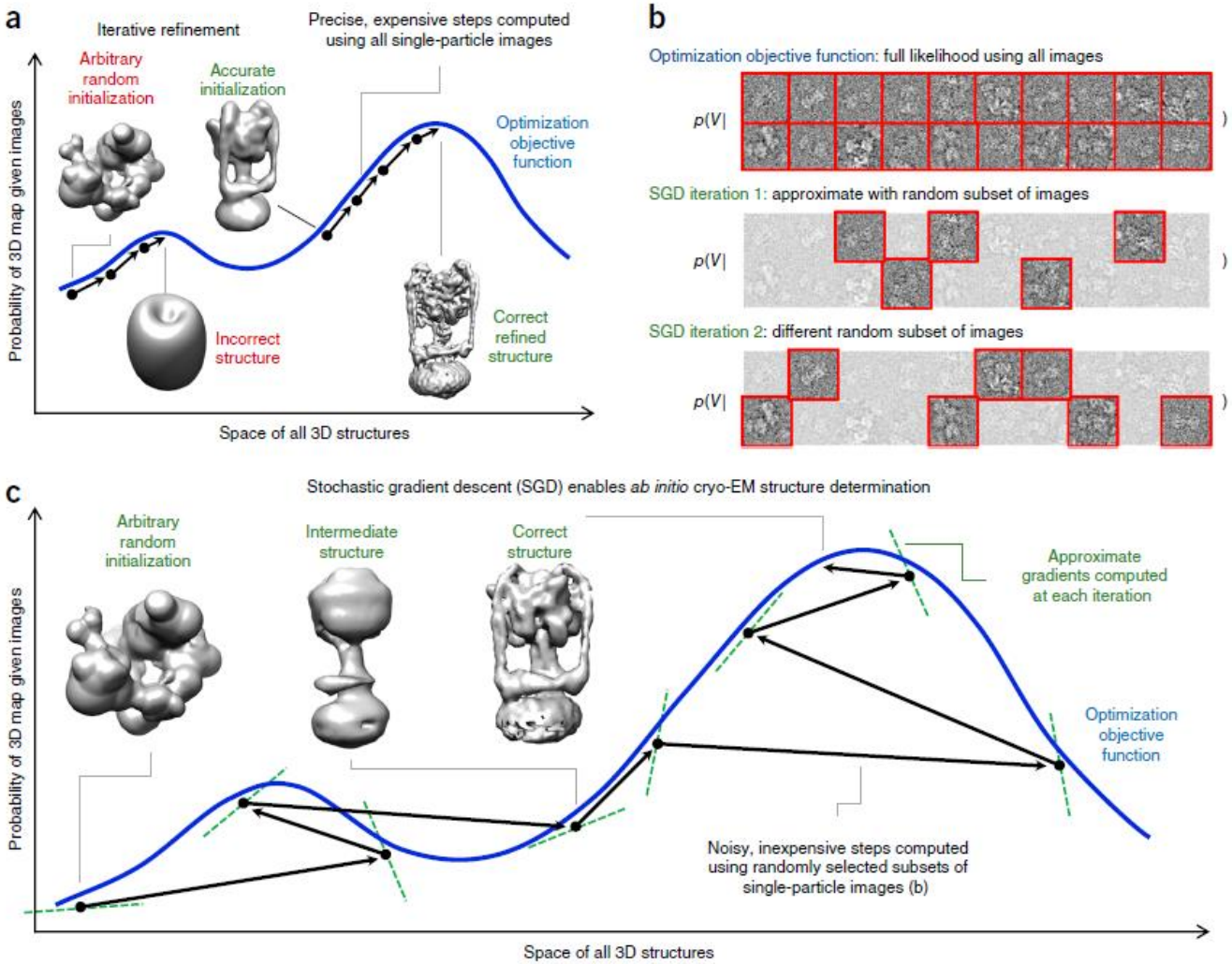


C.O.S. Sorzano, J. Vargas, J. Otón, V. Abrishami, J.M. de la Rosa-Trevín, S. del Riego, A. Fernández-Alderete, C. Martínez-Rey, R. Marabini, J.M. Carazo. Fast and accurate conversion of atomic models into electron density maps. AIMS Biophysics, 2: 8-20 (2015)

500 PDBs

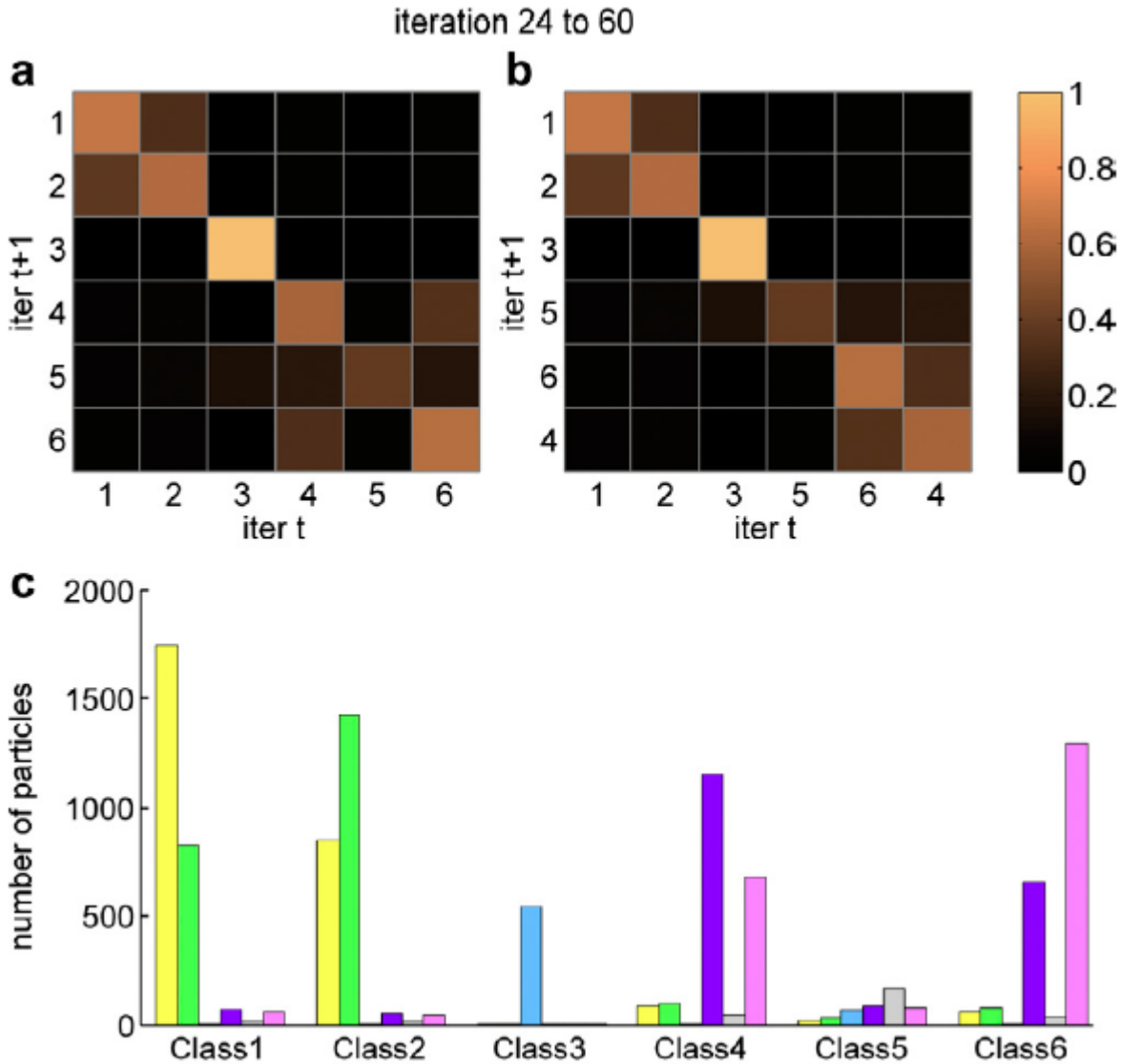
- 83% of Fourier coefficients are dependent on other coefficients
- 86% of Fourier coefficients are not Gaussian
- 100% of Fourier coefficients are not zero-mean
- Real and imaginary parts do not have the same variance
- Real and imaginary parts are correlated.

Bayesian (a priori)



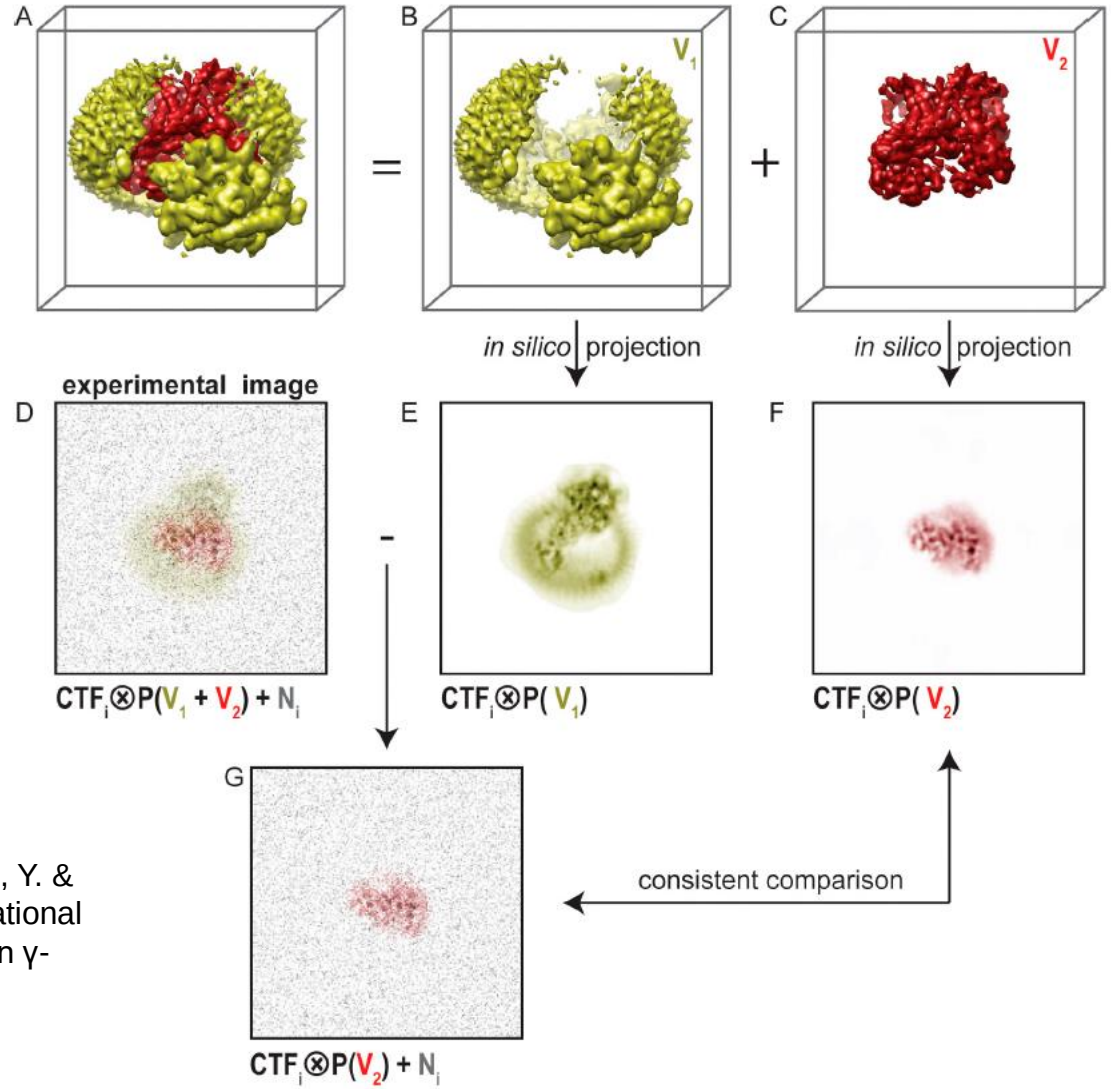
Punjani, A.; Rubinstein, J. L.; Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nature methods, 2017

Class stability



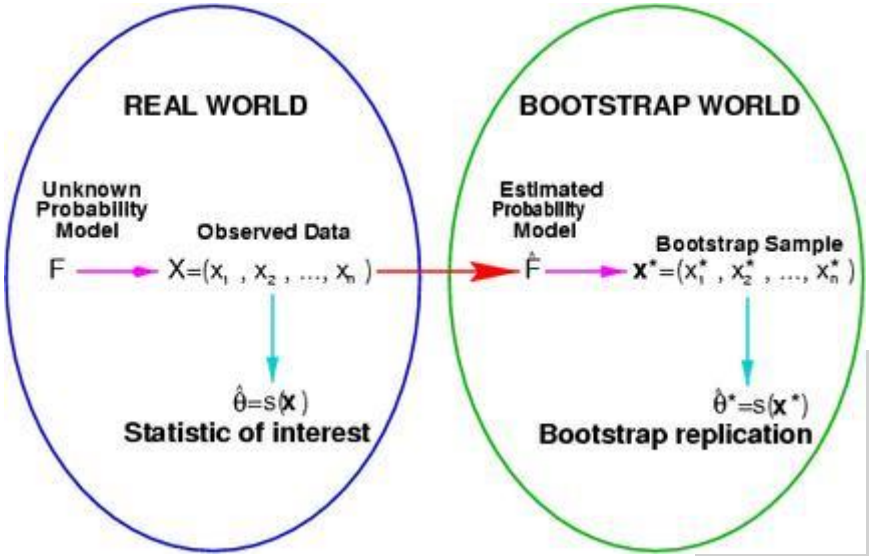
Chen, B.; Shen, B. & Frank, J. Particle migration analysis in iterative classification of cryo-EM single-particle data. J Struct Biol, 2014, 188, 267-273

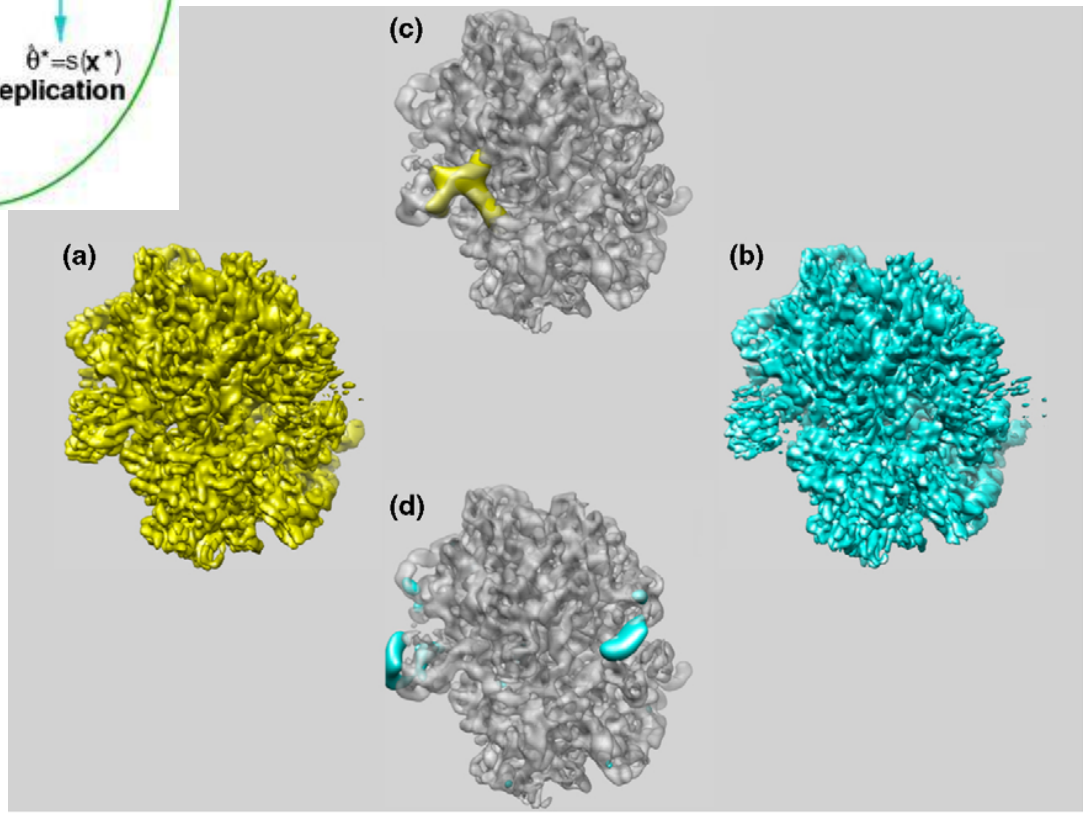
Focused Classification



Bai, X.-C.; Rajendra, E.; Yang, G.; Shi, Y. & Scheres, S. H. Sampling the conformational space of the catalytic subunit of human γ-secretase. Elife, 2015, 4

Bootstrapping





Penczek, P. A.; Yang, C.; Frank, J. & Spahn, C. M. Estimation of variance in single-particle reconstruction using the bootstrap technique J. Structural Biology, 2006, 154, 168-183

Covariance

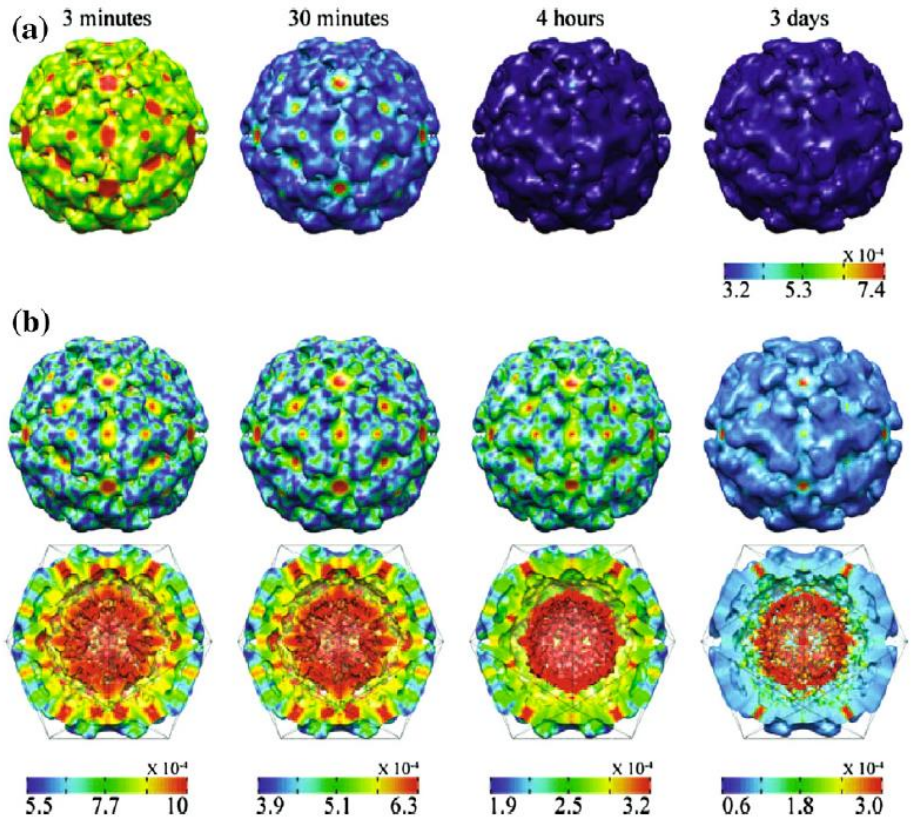
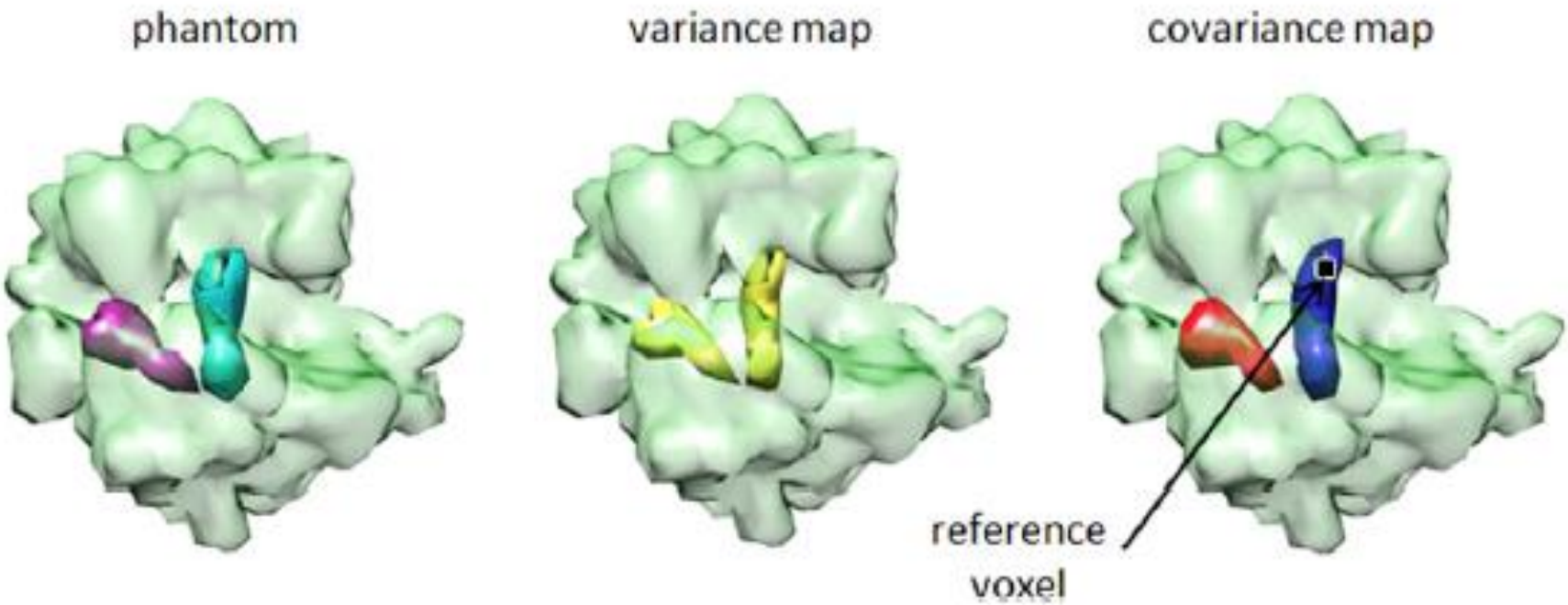


Fig. 2. The four time-resolved reconstructions. Panel A: surface of each of the four reconstructions colored by the square root of the variance map (i.e., the standard deviation map) and displayed using the VIPERdb convention. The same color map is used in all images. Panel B: the surface and a cross section perpendicular to a 2-fold axis of each of the four reconstructions colored by the standard deviation map. The surface and cross section visualizations at a particular time point share the same color map. Different color maps are used at different time points. Visualization by UCSF Chimera (Pettersen et al., 2004).

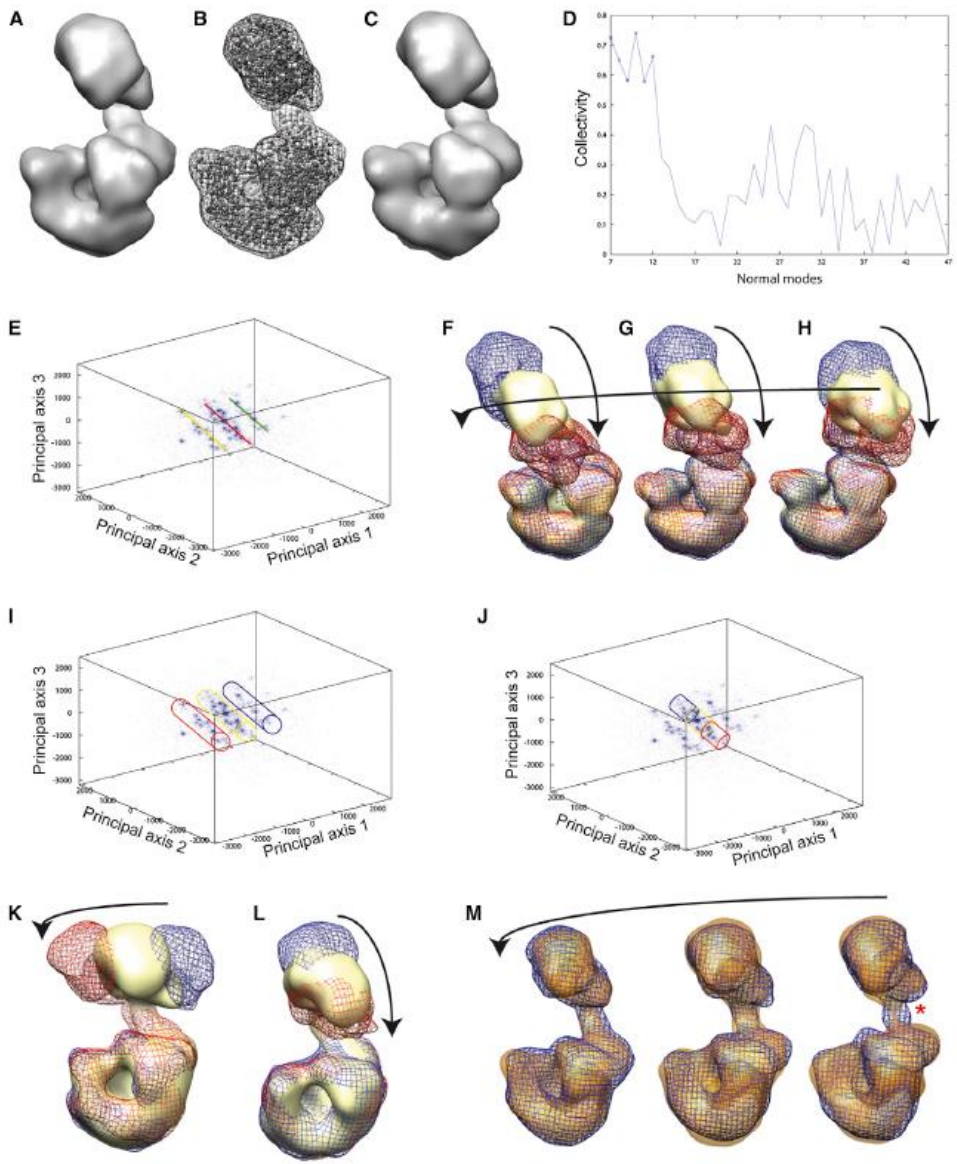
- Zheng, Y.; Wang, Q. & Doerschuk, P. C. Three-dimensional reconstruction of the statistics of heterogeneous objects from a collection of one projection image of each object. Journal of the Optical Society of America. A, Optics, image science, and vision, 2012, 29, 959-970
- Wang, Q.; Matsui, T.; Domitrovic, T.; Zheng, Y.; Doerschuk, P. C. & Johnson, J. E. Dynamics in cryo EM reconstructions visualized with maximum-likelihood derived variance maps. J Struct Biol, 2013, 181, 195-206

Covariance



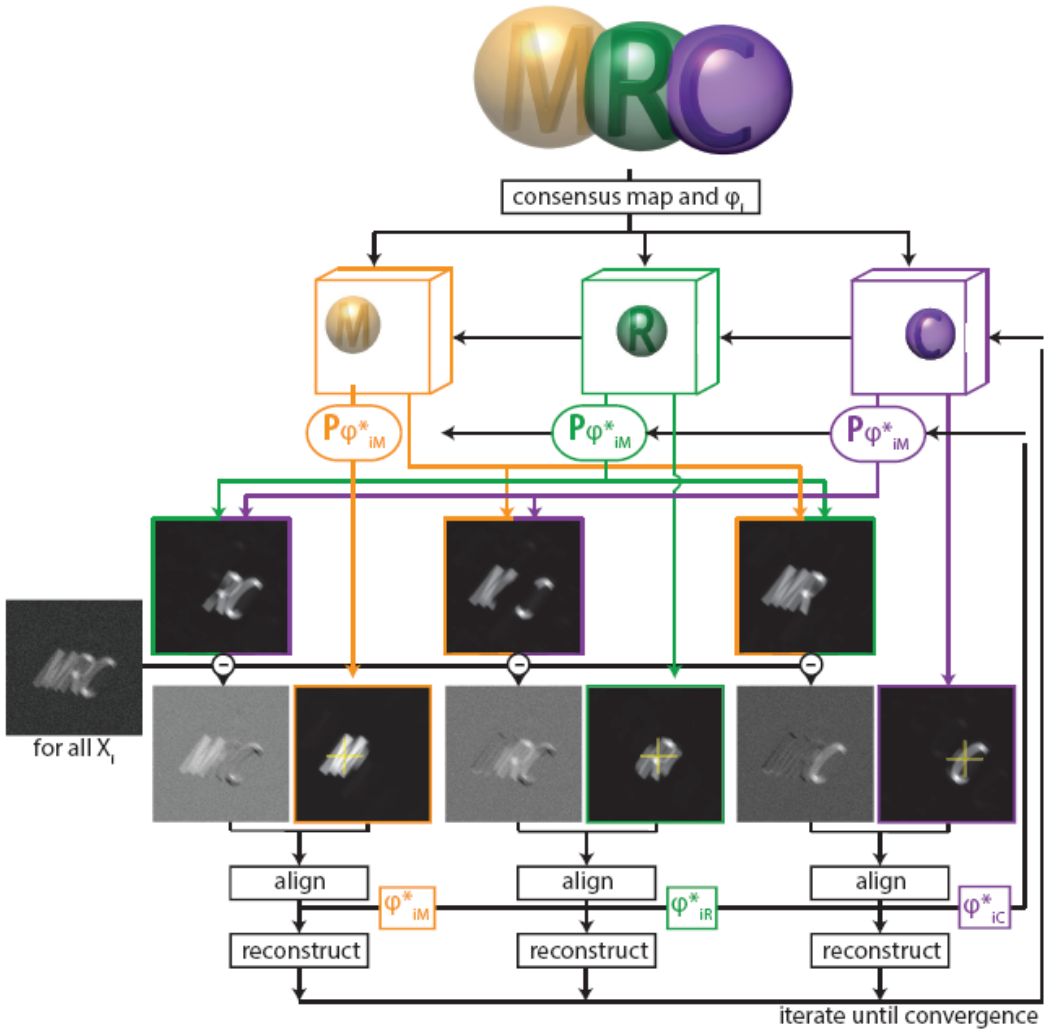
- Andén, J.; Katsevich, E. & Singer, A. Covariance estimation using conjugate gradient for 3D classification in Cryo-EM Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on, 2015, 200-204
- Katsevich, E.; Katsevich, A. & Singer, A. Covariance Matrix Estimation for the Cryo-EM Heterogeneity Problem. SIAM journal on imaging sciences, 2015, 8, 126-185
- Liao, H. Y.; Hashem, Y. & Frank, J. Efficient estimation of three-dimensional covariance and its application in the analysis of heterogeneous samples in cryo-electron microscopy. Structure, 2015, 23, 1129-1137
- Tagare, H. D.; Kucukelbir, A.; Sigworth, F. J.; Wang, H. & Rao, M. Directly reconstructing principal components of heterogeneous particles from cryo-EM images. J Struct Biol, 2015, 191, 245-262
- Andén, J. & Singer, A. Structural Variability from Noisy Tomographic Projections. SIAM Journal on Imaging Sciences, 11(2), pp. 1441-1492, 2018.
- Zhang, C.; Cantara, W.; Jeon, Y.; Musier-Forsyth, K.; Grigorieff, N. & Lyumkis, D. Analysis of discrete local variability and structural covariance in macromolecular assemblies using Cryo-EM and focused classification. Ultramicroscopy, 2019 , 203 , 170-180

Continuous heterogeneity



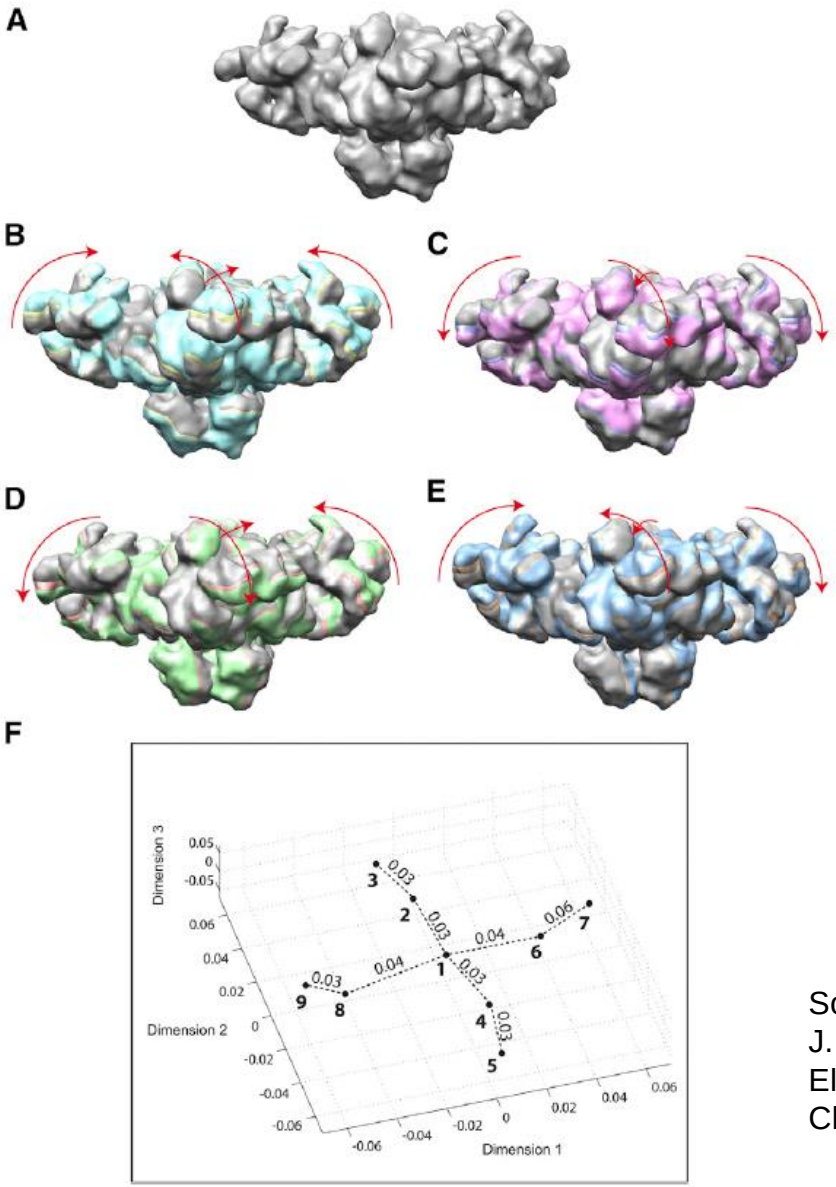
Jin, Q.; Sorzano, C. O. S.; de la Rosa-Trevín, J. M.; Bilbao-Castro, J. R.; Núñez-Ramírez, R.; Llorca, O.; Tama, F. & Jonic;, S. Iterative Elastic 3D-to-2D Alignment Method Using Normal Modes for Studying Structural Dynamics of Large Macromolecular Complexes. Structure, 2014, 22, 496-506

Multibody refinement



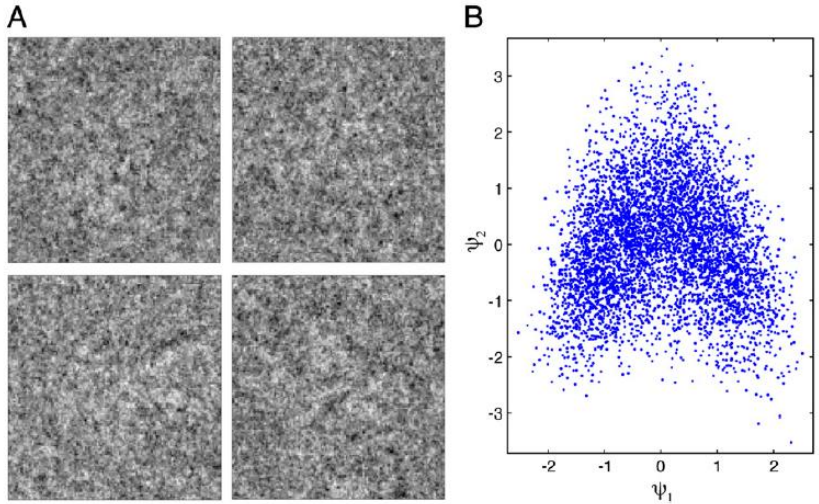
Nakane, T.; Kimanius, D.; Lindahl, E. & Scheres, S. H. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. eLife, 2018, 7

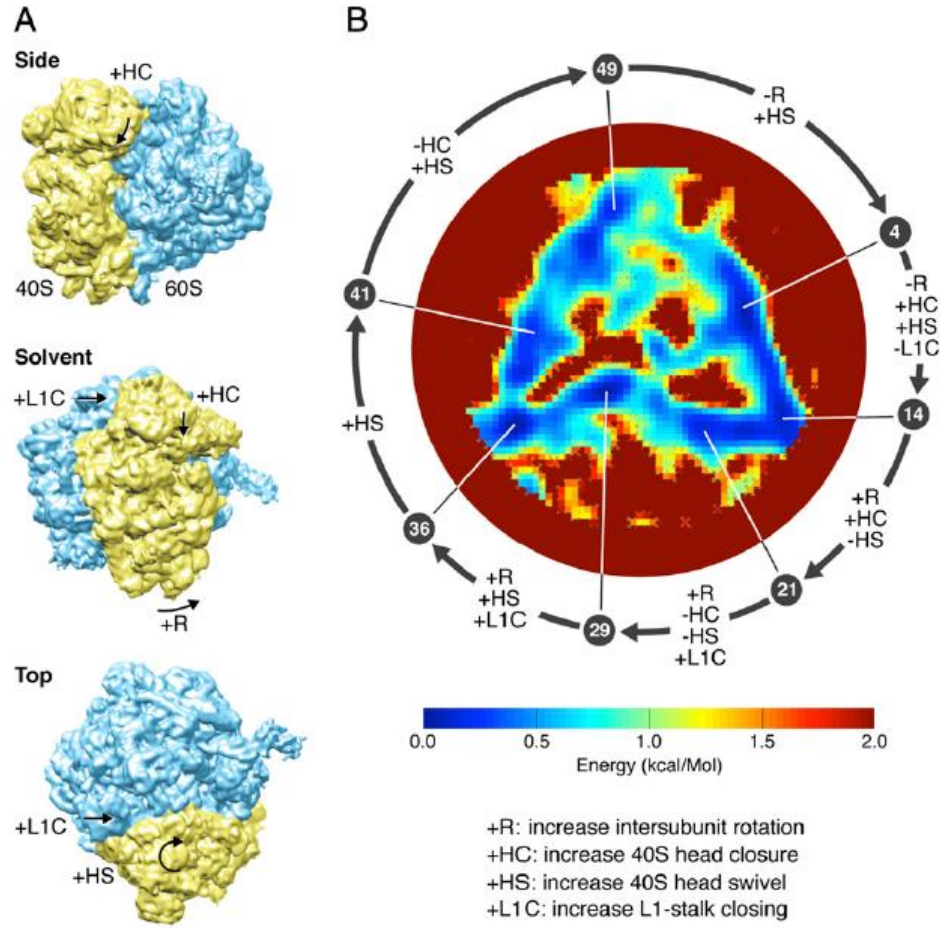
Continuous heterogeneity



Sorzano, C. O. S.; Alvarez-Cabrera, A. L.; Kazemi, M.; Carazo, J. M. & Jonić, S. StructMap: Elastic Distance Analysis of Electron Microscopy Maps for Studying Conformational Changes. Biophys J, 2016, 110, 1753-1765

Continuous heterogeneity

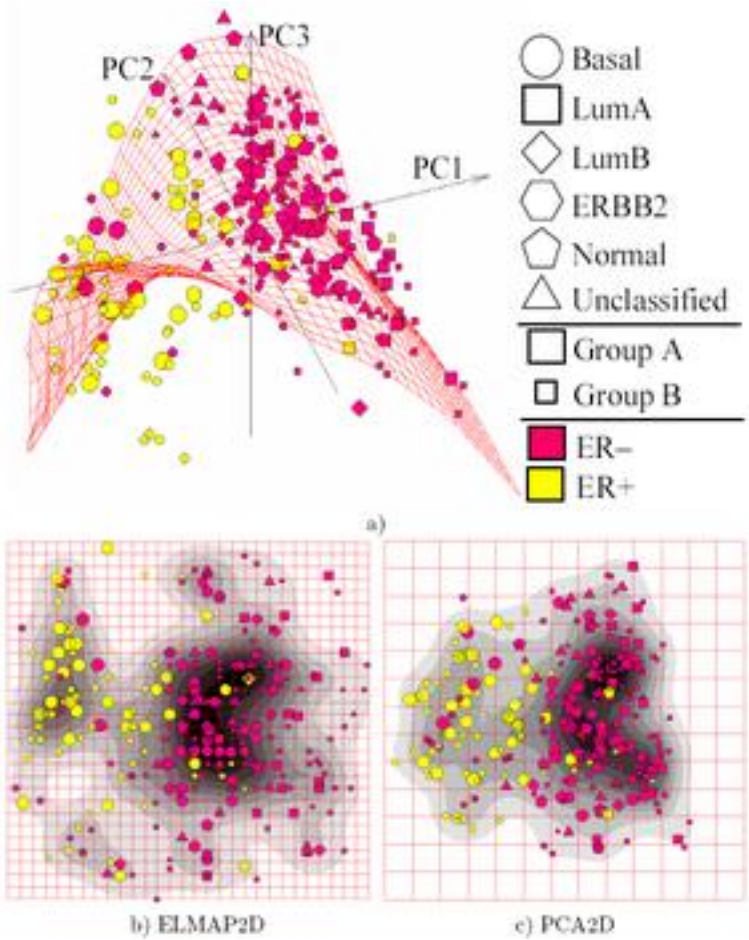


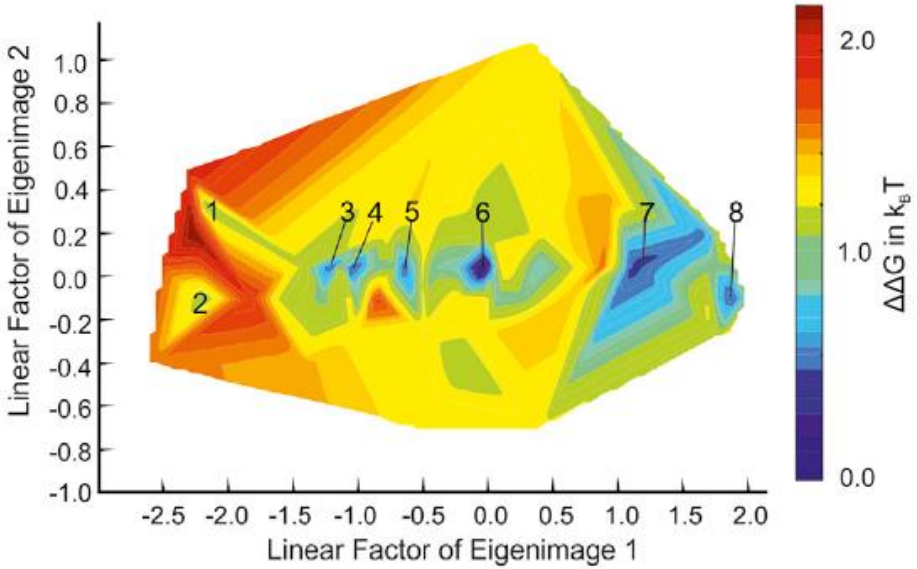


Dashti, A.; Schwander, P.; Langlois, R.; Fung, R.; Li, W.; Hosseinizadeh, A.; Liao, H. Y.; Pallesen, J.; Sharma, G.; Stupina, V. A.; Simon, A. E.; Dinman, J. D.; Frank, J. & Ourmazd, A. Trajectories of the ribosome as a Brownian nanomachine. Proc. Natl. Acad. Sci. USA, 2014, 111, 17492-17497

Continuous heterogeneity

Random subsets of 200.000 into 20 3D classes with Relion and PCA

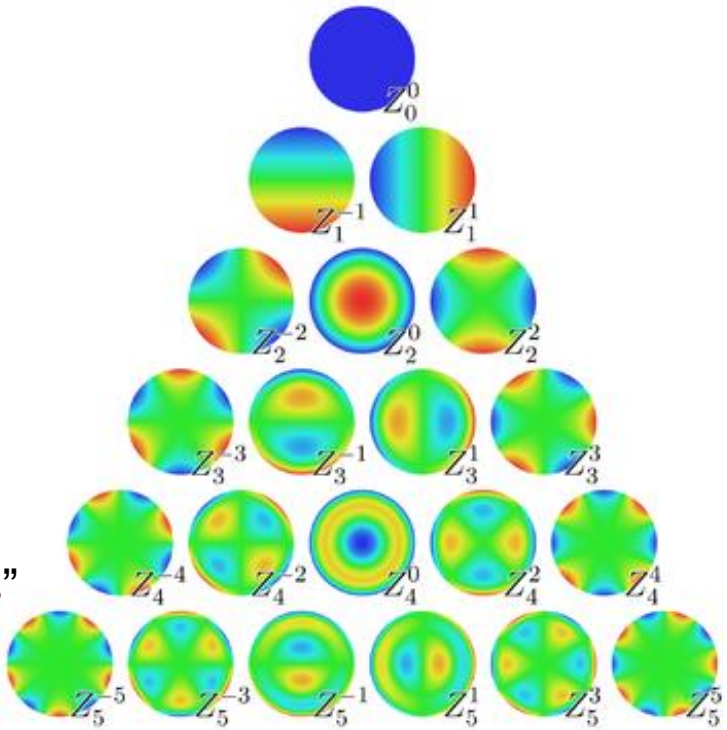




Haselbach, D.; Komarov, I.; Agafonov, D. E.; Hartmuth, K.; Graf, B.; Dybkov, O.; Urlaub, H.; Kastner, B.; Lührmann, R. & Stark, H. Structure and Conformational Dynamics of the Human Spliceosomal Bact Complex. Cell, 2018, 172, 454-464.e11

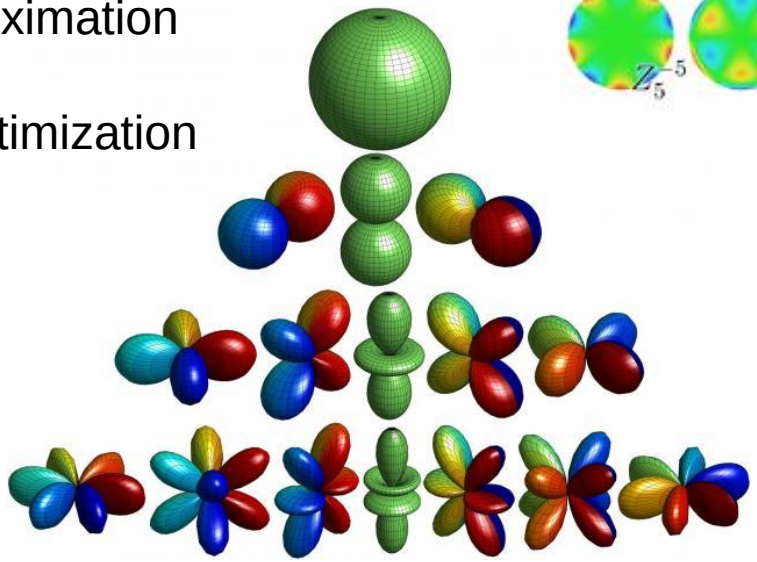
Continuous heterogeneity

$$V_2(\boldsymbol{r}) = V_1\left(\boldsymbol{r} + \begin{pmatrix} \sum_{lnm} c^x_{lnm} Z^m_l\left(\frac{r}{R_{lnm}}\right) Y^n_l(\theta,\varphi) \\ \sum_{lnm} c^y_{lnm} Z^m_l\left(\frac{r}{R_{lnm}}\right) Y^n_l(\theta,\varphi) \\ \sum_{lnm} c^z_{lnm} Z^m_l\left(\frac{r}{R_{lnm}}\right) Y^n_l(\theta,\varphi) \end{pmatrix}\right)$$

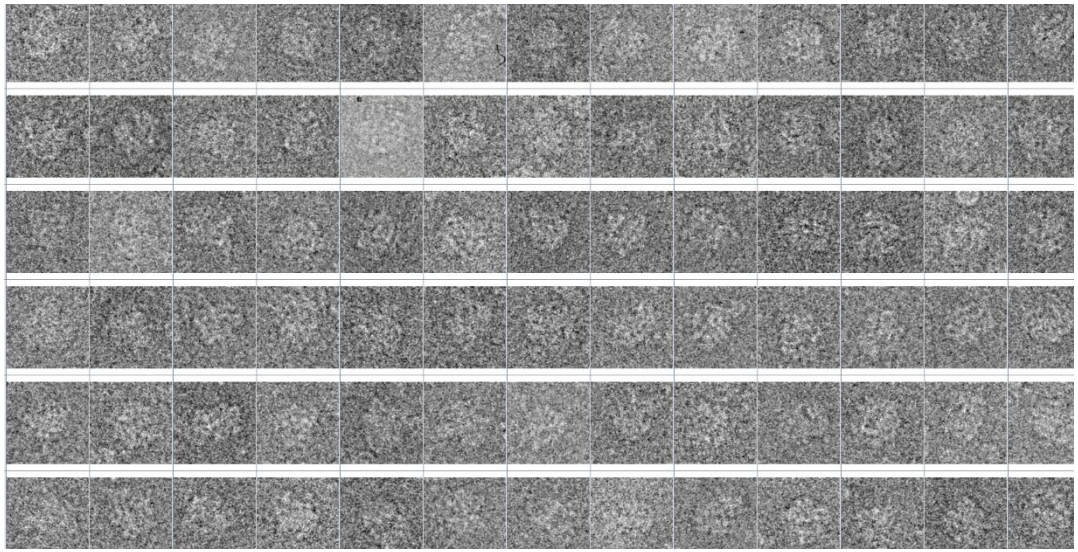


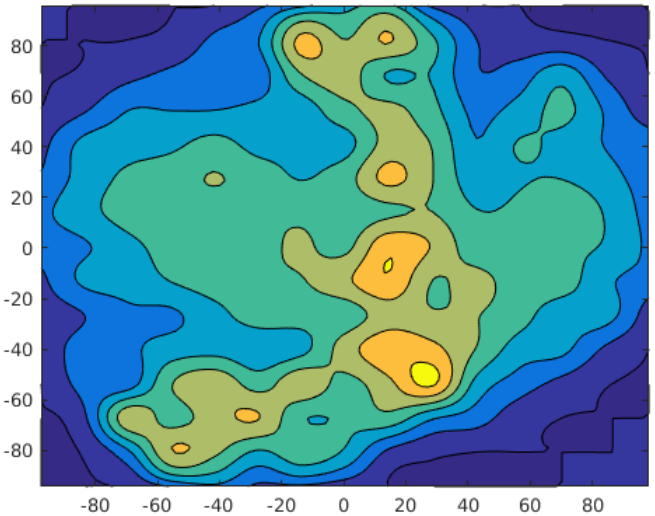
Zernike polynomials and Spherical Harmonics

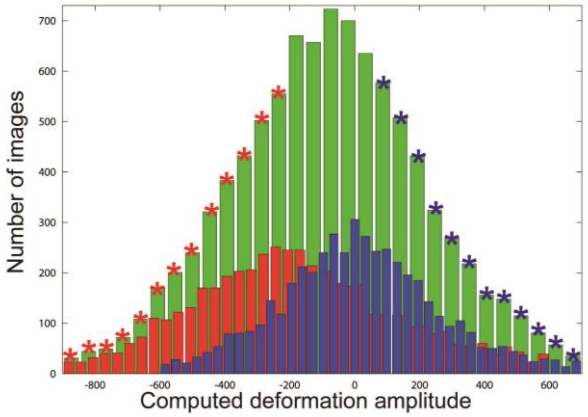
- No need to select the “interesting normal modes”
- Parameter-free approximation
- Multiresolution
- Almost orthogonal optimization

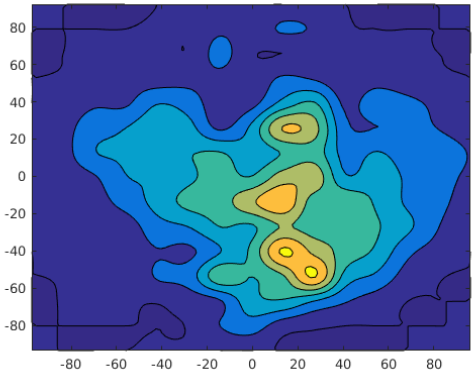


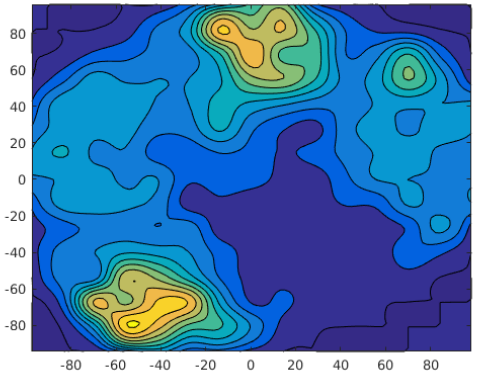
Continuous heterogeneity



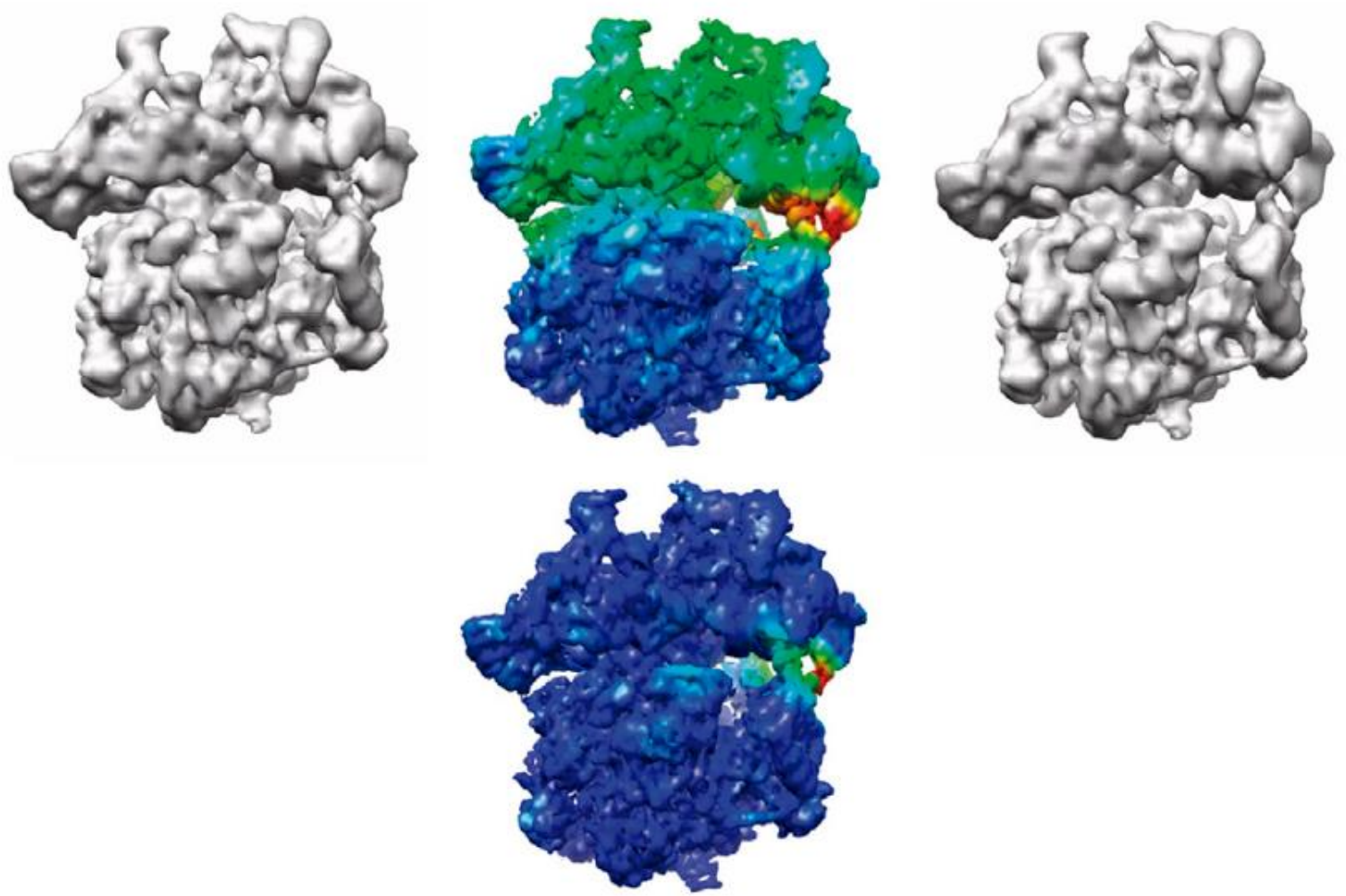








Continuous heterogeneity



Sorzano, C. O. S.; Martín-Ramos, A.; Prieto, F.; Melero, R.; Martín-Benito, J.; Jonic, S.; Navas-Calvente, J.; Vargas, J.; Otón, J.; Abrishami, V.; de la Rosa-Trevín, J. M.; Gómez-Blanco, J.; Vilas, J. L.; Marabini, R. & Carazo, J. M. Local analysis of strains and rotations for macromolecular electron microscopy maps. Journal of structural biology, 2016, 195, 123-128

Conclusions

- Heterogeneity is currently one of the most limiting factors
- Priors are useful if they are true
- Class stability?
- Number of classes?
- Focused classification
- Variance/Covariance matrix
- Continuous heterogeneity

Understanding EM resolution

Carlos Oscar S. Sorzano

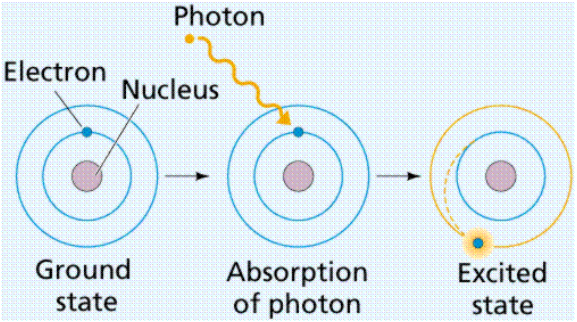
Instruct Image Processing Center

CSIC

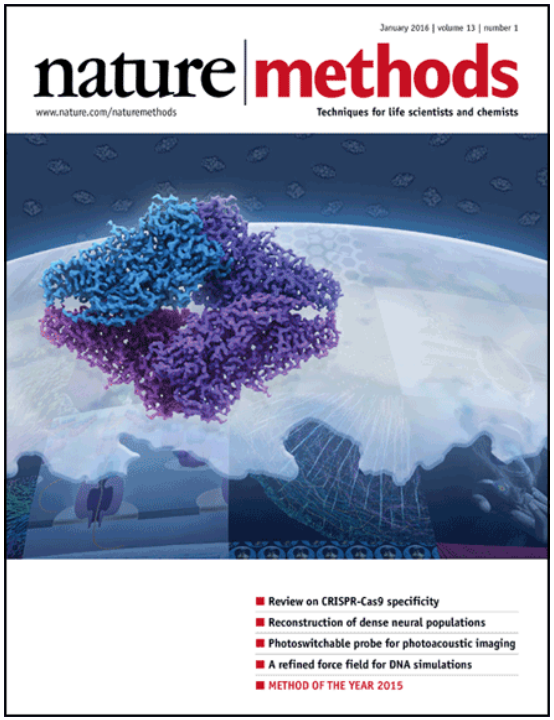
instruct Integrating Biology

instruct image Processing Center

State of the field



1. Bai, X. C.; McMullan, G. & Scheres, S. H. W. **How cryo-EM is revolutionizing structural biology**. Trends in Biochemical Sciences, 2014, 40, 49-57
2. Kühlbrandt, W. **The resolution revolution**. Science, 2014, 343, 1443-1444
3. Carazo, J. M.; **Sorzano, C. O. S**.; Otón, J.; Marabini, R. & Vargas, J. Three-dimensional reconstruction methods in Single Particle Analysis from transmission electron microscopy data. Arch Biochem Biophys, 2015, 581, 39-48
4. Cheng, Y.; Grigorieff, N.; Penczek, P. A. & Walz, T. A primer to single-particle cryo-electron microscopy. Cell, 2015, 161, 438-449
5. Cheng, Y. **Single-Particle Cryo-EM at Crystallographic Resolution**. Cell, 2015, 161, 450-457
6. Elmlund, D. & Elmlund, H. Cryogenic electron microscopy and single-particle analysis. Annu Rev Biochem, 2015, 84, 499-517
7. Henderson, R. Overview and future of single particle electron cryomicroscopy. Arch Biochem Biophys, 2015, 581, 19-24
8. Nogales, E. & Scheres, S. H. W. **Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity**. Mol Cell, 2015, 58, 677-689
9. Schröder, R. R. Advances in electron microscopy: A qualitative view of instrumentation development for macromolecuar imaging and tomography. Archives of Biochemistry and Biophysics, 2015, 581, 25-38
10. Carroni, M. & Saibil, H. R. Cryo electron microscopy to determine the structure of macromolecular complexes. Methods, 2016, 95, 78-85
11. Eisenstein, M. **The field that came in from the cold**. Nature Methods, 2016, 13, 19-22
12. Glaeser, R. M. **How good can cryo-EM become?** Nat Methods, 2016, 13, 28-32
13. Wu, S.; Armache, J.-P. & Cheng, Y. Single-particle cryo-EM data acquisition by using direct electron detection camera. Microscopy, 2016
14. Nogales, E. **The development of cryo-EM into a mainstream structural biology technique.** Nature Methods, 2016, 13, 24-27
15. Subramaniam, S.; Kühlbrandt, W. & Henderson, R. **CryoEM at IUCrJ: a new era**. IUCrJ, 2016, 3, 3-7



High-throughput

EMBL-EBI

Protein Data Bank in Europe

Bringing Structure to Biology

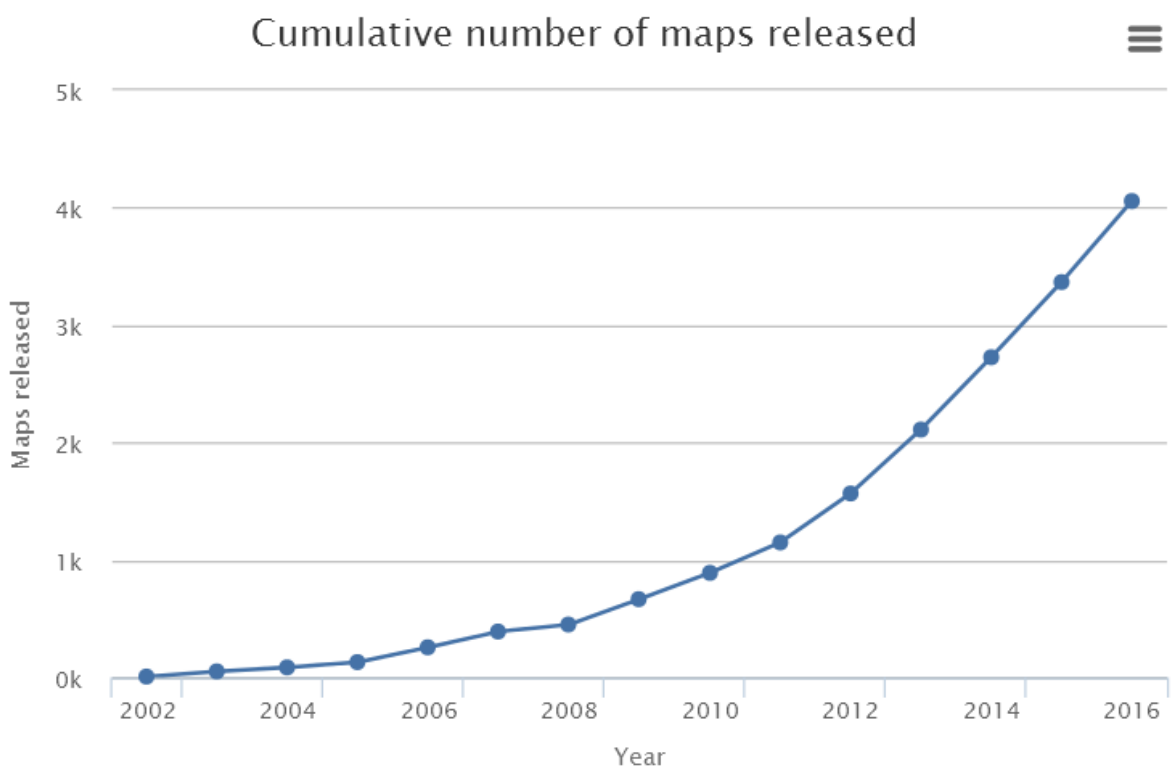
EMDB statistics

EM Resources

- Home
- Statistics
- Validation
- EMDataBank
- EMPIAR
- Test data

EMDB

- Latest maps
- Latest headers
- Latest updates
- Search
- Browse
- FTP archive
- Deposit EM map/model
- EMDB data model



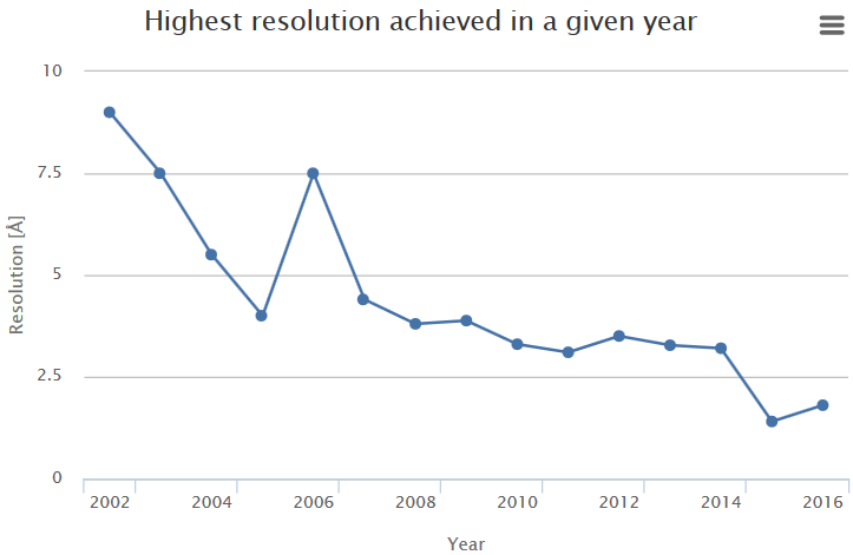
High-resolution

EMBL-EBI | Services | Research | Training | About us

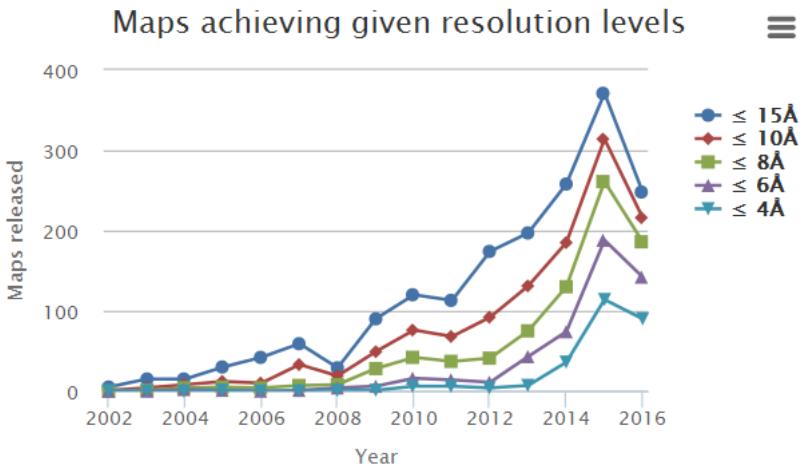
Protein Data Bank in Europe
Bringing Structure to Biology

EMDB statistics

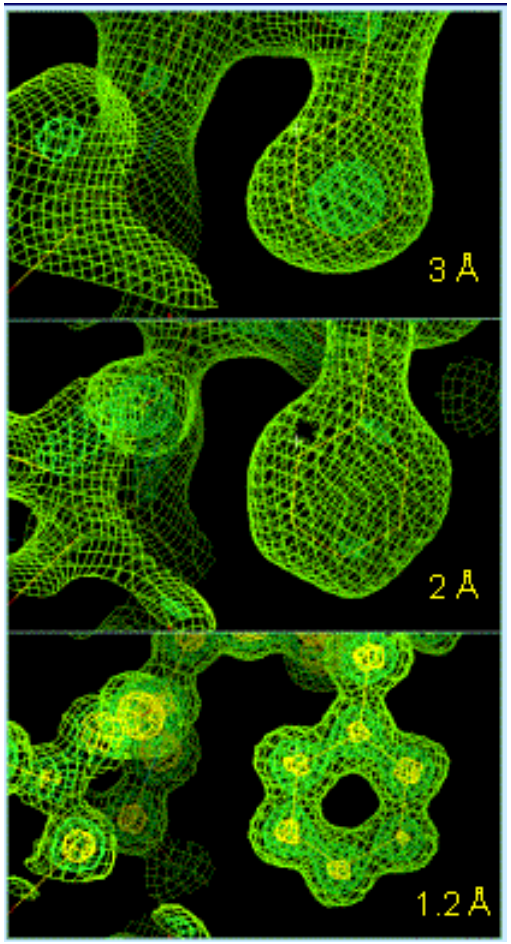
Trend of highest resolutions achieved annually for released maps



Number of released maps achieving given resolution levels

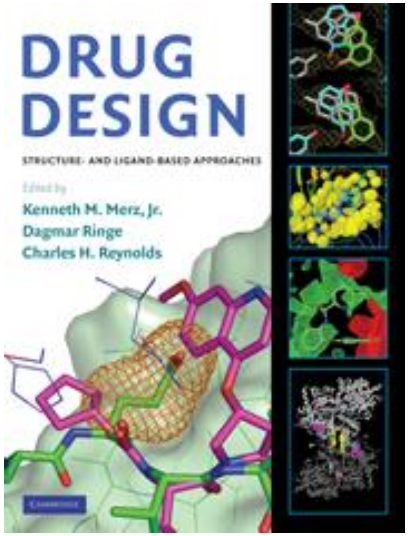


High-resolution



Experiment	Typical Cost per compound ($)
Computer Modeling	10
Biochemical assay	400
Cell culture assay	4,000
Rat acute toxicity	12,000
Protein crystal structure	100,000
Animal efficacy trial	300,000
Rat 2-year chronic oral toxicity	800,000
Human clinical trial	500,000,000

Young, D. C. Computational drug design John Wiley & Sons, 2009



High-resolution, High-throughput EM

Please cite this article in press as: Merk et al., Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery, Cell (2016), http://dx.doi.org/10.1016/j.cell.2016.05.040

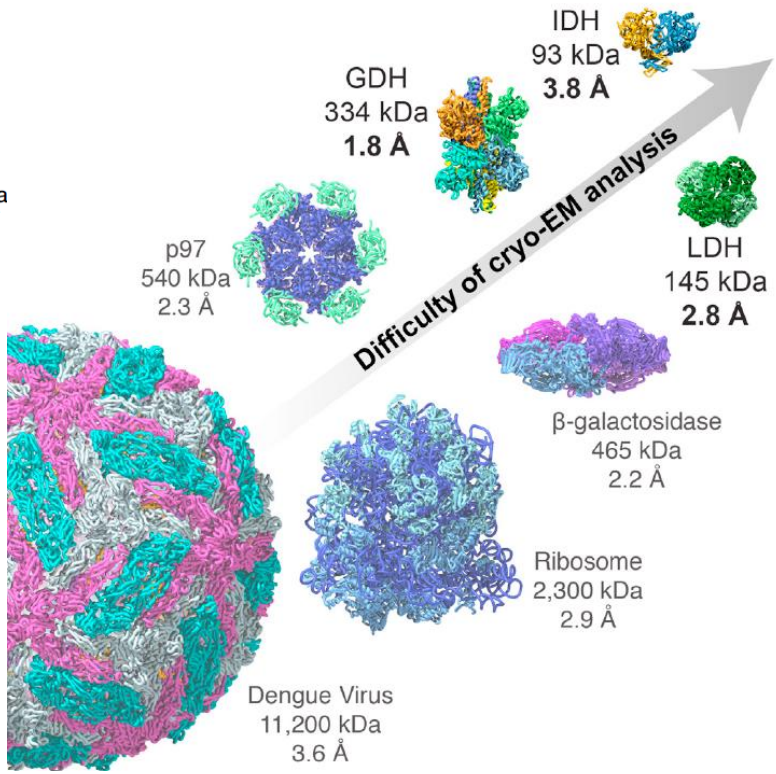
Article

Cell

Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery

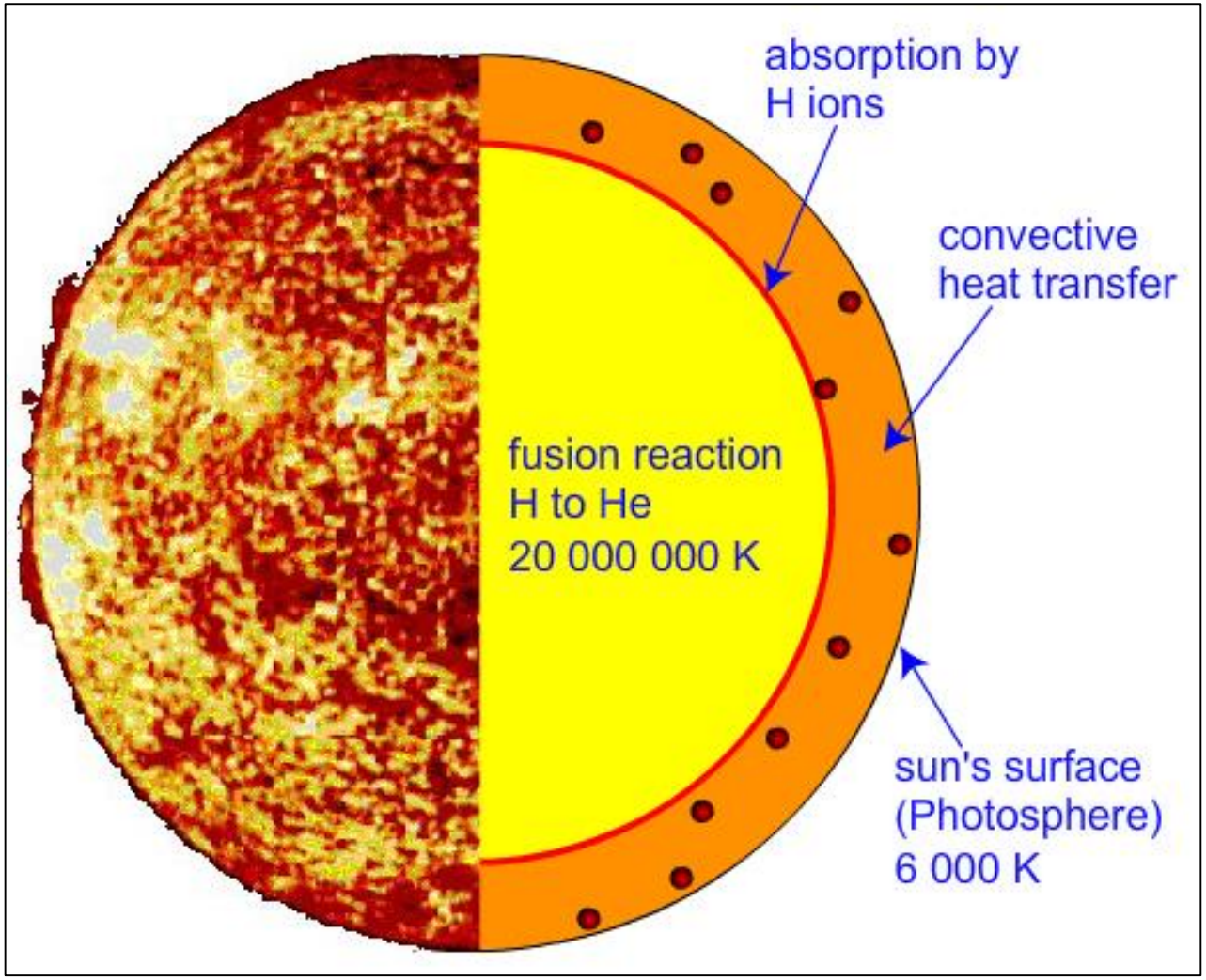
Alan Merk,[1,3] Alberto Bartesaghi,[1,3] Soojay Banerjee,[1,3] Veronica Falconieri,[1] Prashant Rao,[1] Mindy I. Da Rajan Pragani,[2] Matthew B. Boxer,[2] Lesley A. Earl,[1] Jacqueline L.S. Milne,[1] and Sriram Subramaniam[1,*]

drogenase (145 kDa) and glutamate dehydrogenase (334 kDa), respectively. With these results, two perceived barriers in single-particle cryo-EM are overcome: (1) crossing 2 Å resolution and (2) obtaining structures of proteins with sizes < 100 kDa, demonstrating that cryo-EM can be used to investigate a broad spectrum of drug-target interactions and dynamic conformational states.



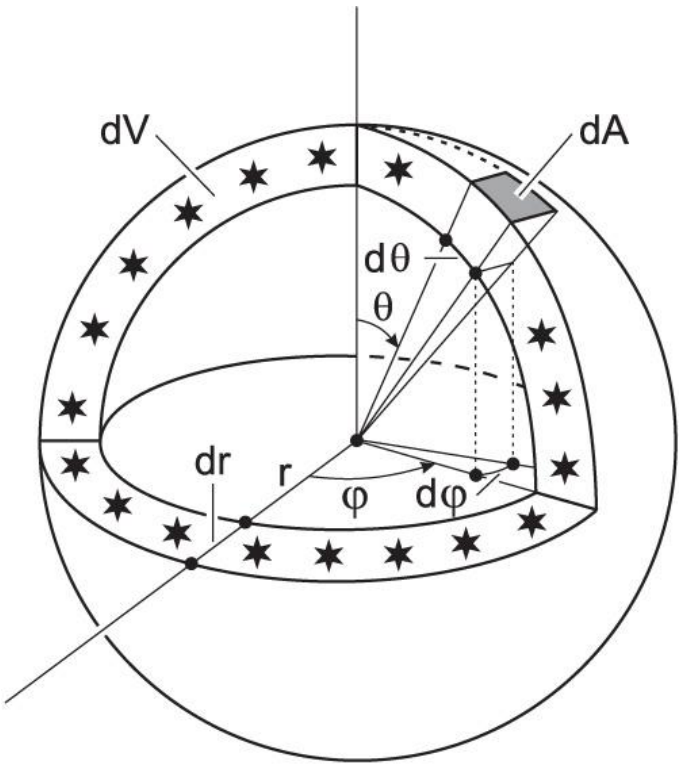
Merk, A et al, Cell, 2016, *165*, 1698-1707

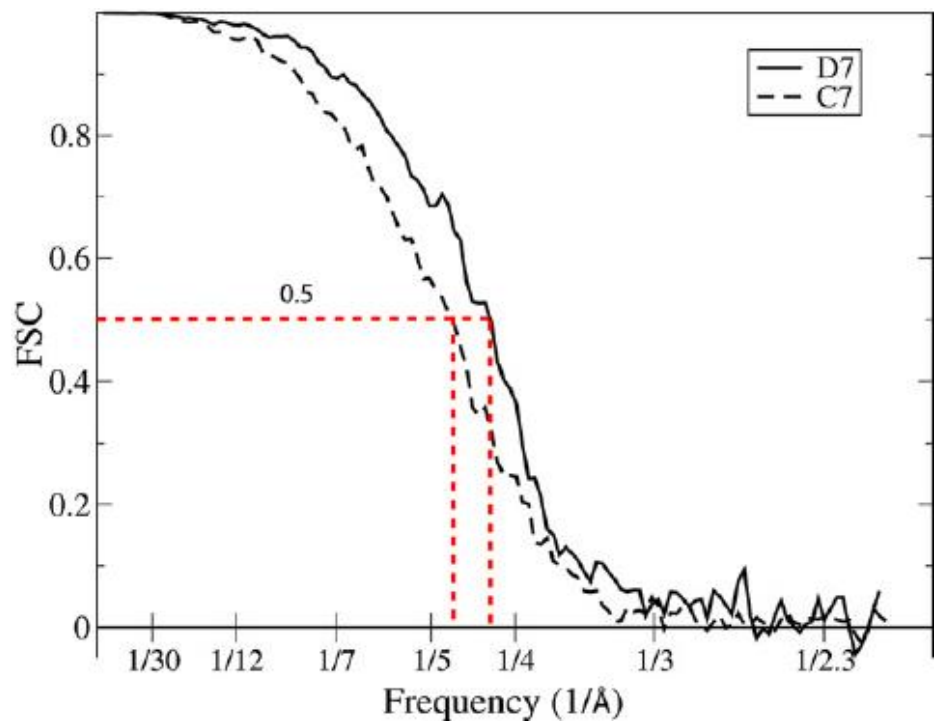
Let’s try to understand this number



Fourier Shell Correlation

$$FSC(R,\Delta R) = \frac{\sum_{\mathbf{R}\in\mathfrak{S}(R,\Delta R)} F_1(\mathbf{R})F_2^*(\mathbf{R})}{\sqrt{\left(\sum_{\mathbf{R}\in\mathfrak{S}(R,\Delta R)} |F_1(\mathbf{R})|^2\right)\left(\sum_{\mathbf{R}\in\mathfrak{S}(R,\Delta R)} |F_2(\mathbf{R})|^2\right)}},$$





Fourier Shell Correlation

FSC

$$FSC(R,\Delta R) = \frac{\sum_{\mathbf{R}\in\mathfrak{S}(R,\Delta R)} F_1(\mathbf{R})F_2^*(\mathbf{R})}{\sqrt{\left(\sum_{\mathbf{R}\in\mathfrak{S}(R,\Delta R)} |F_1(\mathbf{R})|^2\right)\left(\sum_{\mathbf{R}\in\mathfrak{S}(R,\Delta R)} |F_2(\mathbf{R})|^2\right)}},$$

Cross correlation in real-space

$$(f\star g)(\tau) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f^*(t)\, g(t+\tau)\, dt,$$

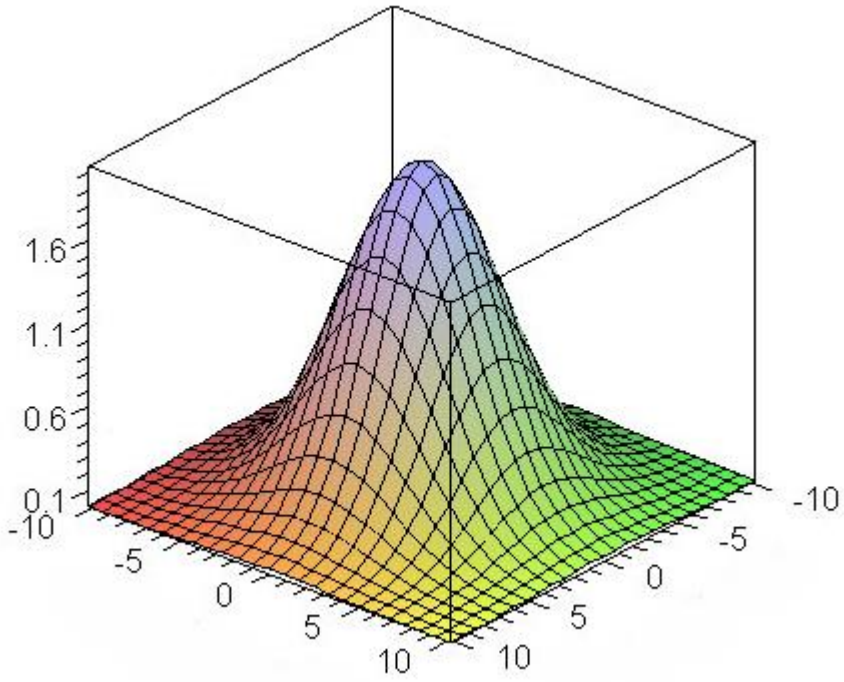
Normalized cross correlation in Fourier-space

$$NCC_{f_1,f_2}(\mathbf{0}) = \frac{\sum_{\mathbf{R}\in\left[-\frac{N}{2},\frac{N}{2}\right]^3} F_1(\mathbf{R})F_2^*(\mathbf{R})}{\sqrt{\left(\sum_{\mathbf{R}\in\left[-\frac{N}{2},\frac{N}{2}\right]^3} |F_1(\mathbf{R})|^2\right)\left(\sum_{\mathbf{R}\in\left[-\frac{N}{2},\frac{N}{2}\right]^3} |F_2(\mathbf{R})|^2\right)}}$$

FSC=Normalized cross correlation with zero-lag of two volumes after band-pass filtering them

Fourier Shell Correlation

$$FSC(R,\Delta R) = \frac{\sum_{\mathbf{R}\in\mathfrak{S}(R,\Delta R)} F_1(\mathbf{R})F_2^*(\mathbf{R})}{\sqrt{\left(\sum_{\mathbf{R}\in\mathfrak{S}(R,\Delta R)} |F_1(\mathbf{R})|^2\right)\left(\sum_{\mathbf{R}\in\mathfrak{S}(R,\Delta R)} |F_2(\mathbf{R})|^2\right)}},$$



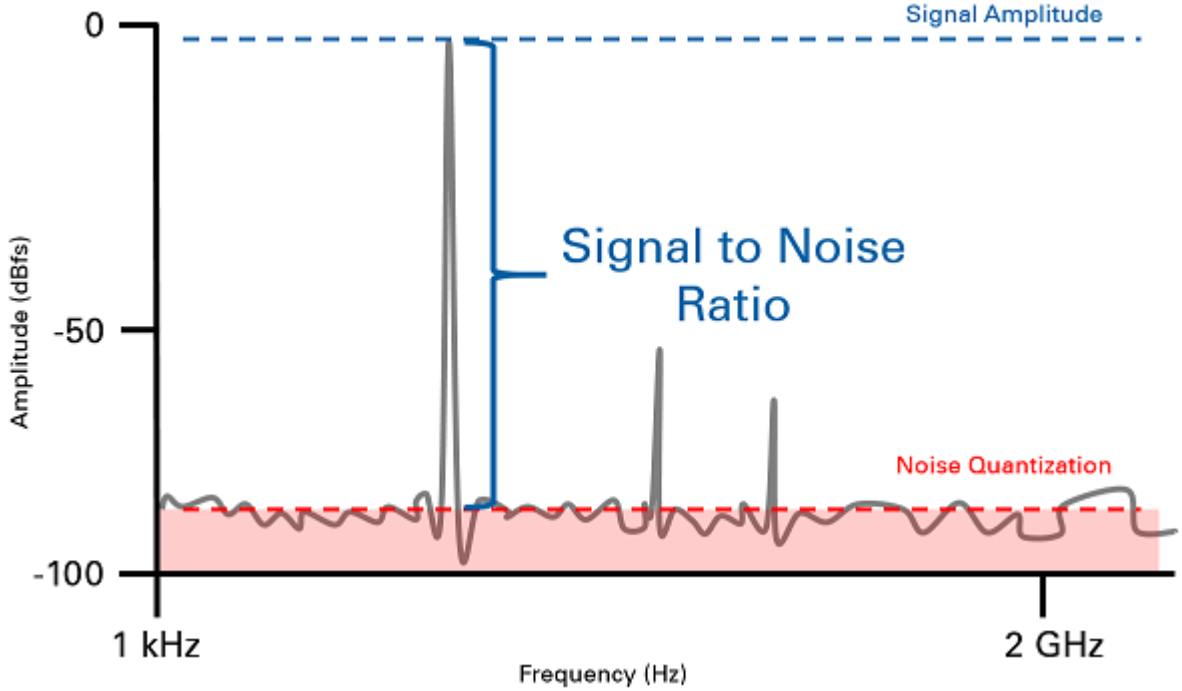
The FSC is invariant to an isotropic, non-vanishing filter

FSC and SNR

SNR

. . . there was ~~light~~

Spectral Signal-to-Noise Ratio



$$\mathrm{SNR} = \frac{P_{\mathrm{signal}}}{P_{\mathrm{noise}}} = \left(\frac{A_{\mathrm{signal}}}{A_{\mathrm{noise}}}\right)^2,$$

Spectral Signal-to-Noise Ratio

$$F_k(\mathbf{R}) = X(\mathbf{R}) + N_k(\mathbf{R}),$$

Multiple measurements of the same Fourier coefficient

$$\widehat{X}(\mathbf{R}) = \frac{1}{K}\sum_{k=1}^{K} F_k(\mathbf{R}) = X(\mathbf{R}) + \frac{1}{K}\sum_{k=1}^{K} N_k(\mathbf{R})$$

Estimate of the signal

$$\widehat{S}(R,\Delta R) = \frac{1}{|\mathfrak{S}(R,\Delta R)|}\sum_{\mathbf{R}\in\mathfrak{S}(R,\Delta R)} |\widehat{X}(\mathbf{R})|^2$$

Estimate of the signal power

$$\widehat{N}(R,\Delta R) = \frac{1}{|\mathfrak{S}(R,\Delta R)|}\sum_{\mathbf{R}\in\mathfrak{S}(R,\Delta R)} \left(\frac{1}{K-1}\sum_{k=1}^{K} |F_k(\mathbf{R}) - \widehat{X}(\mathbf{R})|^2\right)$$

Estimate of the noise

Spectral Signal-to-Noise Ratio

In the absence of signal

$$\widehat{S}(R,\Delta R) = \frac{1}{|\mathfrak{S}(R,\Delta R)|} \sum_{\mathbf{R}\in\mathfrak{S}(R,\Delta R)} \left|\frac{1}{K}\sum_{k=1}^{K} N_k(\mathbf{R})\right|^2$$

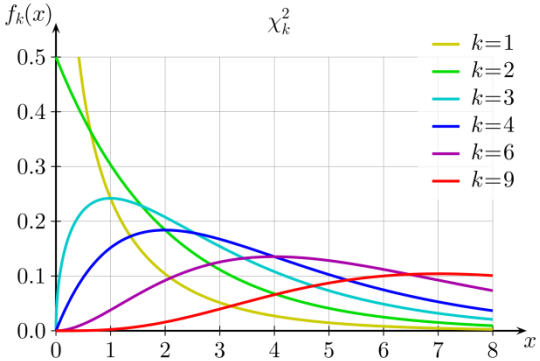
$$\chi_1^2 = |\mathfrak{S}(R,\Delta R)| \frac{\widehat{S}(R,\Delta R)}{\frac{\sigma_R^2}{2K}}$$

Chi2 with $|\mathfrak{S}(R,\Delta R)|$ dof

$$\widehat{N}(R,\Delta R) = \frac{1}{|\mathfrak{S}(R,\Delta R)|} \sum_{\mathbf{R}\in\mathfrak{S}(R,\Delta R)} \left(\frac{1}{K-1} \sum_{k=1}^{K} \left|\frac{K-1}{K} N_k(\mathbf{R}) - \frac{1}{K} \sum_{\substack{k'=1 \\ k'\neq k}}^{K} N_{k'}(\mathbf{R})\right|^2 \right)$$

$$\chi_2^2 = |\mathfrak{S}(R,\Delta R)| \frac{\widehat{N}(R,\Delta R)}{\frac{\sigma_R^2}{2(K-1)}}$$

Chi2 with $K|\mathfrak{S}(R,\Delta R)|$ dof

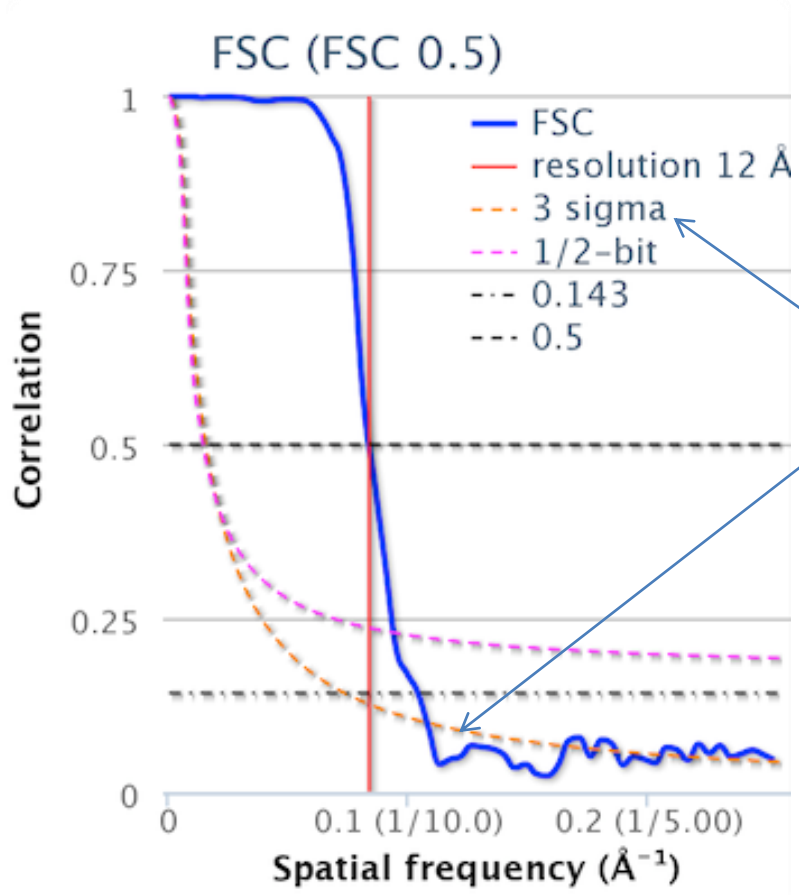


Spectral Signal-to-Noise Ratio

$$F(R,\Delta R) = \frac{\frac{\chi_1^2}{|\mathfrak{S}(R,\Delta R)|}}{\frac{\chi_2^2}{K|\mathfrak{S}(R,\Delta R)|}}$$

F-Snedecor $|\mathfrak{S}(R,\Delta R)|$, $K|\mathfrak{S}(R,\Delta R)|$ dof

$$\begin{aligned} \mathrm{E}\{F(R,\Delta R)\} &= \frac{d_2}{d_2-2} = \frac{K|\mathfrak{S}(R,\Delta R)|}{K|\mathfrak{S}(R,\Delta R)|-2} \\ \mathrm{Var}\{F(R,\Delta R)\} &= \frac{2d_2^2(d_1+d_2-2)}{d_1(d_2-2)^2(d_2-4)} = \frac{2K^2|\mathfrak{S}(R,\Delta R)|((K+1)|\mathfrak{S}(R,\Delta R)|-2)}{(K|\mathfrak{S}(R,\Delta R)|-2)^2(K|\mathfrak{S}(R,\Delta R)|-4)} \end{aligned}$$



1. **This curve is obtained only after a Gaussian approximation of the F-Snedecor.**
2. **There is not a single curve, but a whole distribution.**
3. **3-sigma is a "proxy" for hypothesis testing**

Spectral Signal-to-Noise Ratio

In the presence of signal

$$\chi_1^2 = |\mathfrak{S}(R,\Delta R)| \frac{\widehat{S}(R,\Delta R)}{\frac{\sigma_R^2}{2K}}$$

Non-central Chi2 with $|\mathfrak{S}(R,\Delta R)|$ dof

$$\lambda = K|\mathfrak{S}(R,\Delta R)| \cdot SSNR(R,\Delta R).$$

$$\begin{aligned} \mathrm{E}\{F(R,\Delta R)\} &= \frac{d_2(d_1+\lambda)}{d_1(d_2-2)} \approx 1 + K \cdot SSNR(R,\Delta R) \\ \mathrm{Var}\{F(R,\Delta R)\} &= 2\frac{(d_1+\lambda)^2+(d_1+2\lambda)(d_2-2)}{(d_2-2)^2(d_2-4)}\frac{d_2^2}{d_1^2} \\ &\approx 2\frac{K^2 SSNR^2(R,\Delta R)+2K(K+1)SSNR(R,\Delta R)+(K+1)}{K|\mathfrak{S}(R,\Delta R)|} \end{aligned}$$

True image SSNR before averaging

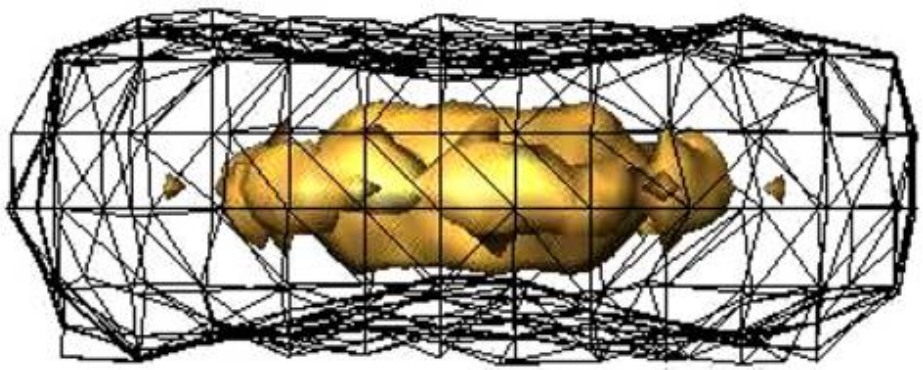
$$\boxed{\widehat{SSNR}(R,\Delta R) = \begin{cases} F(R,\Delta R) - 1 & F(R,\Delta R) \geq 1 \\ 0 & F(R,\Delta R) < 1 \end{cases}}$$

Estimate of the Volume SSNR after averaging

Spectral Signal-to-Noise Ratio

$$\widehat{S}(R,\Delta R) = \frac{1}{|\mathfrak{S}(R,\Delta R)|} \sum_{\mathbf{R}\in\mathfrak{S}(R,\Delta R)} \left|\frac{1}{K}\sum_{k=1}^{K} N_k(\mathbf{R})\right|^2$$

$$\widehat{N}(R,\Delta R) = \frac{1}{|\mathfrak{S}(R,\Delta R)|} \sum_{\mathbf{R}\in\mathfrak{S}(R,\Delta R)} \left(\frac{1}{K-1}\sum_{k=1}^{K} |F_k(\mathbf{R}) - \widehat{X}(\mathbf{R})|^2\right)$$



Unser, M.; Sorzano, C. O. S.; Thévenaz, P.; Jonic, S.; El-Bez, C.; De Carlo, S.; Conway, J. & Trus, B. L. Spectral Signal-to-Noise Ratio and resolution assessment of 3D reconstructions. *J. Structural Biology,* **2005***, 149*, 243-255

SSNR ↔ FSC

$$SSNR = \frac{FSC}{1 - FSC}$$

1974:

On Estimating Signal-to-Noise Ratio Using the Sample Correlation Coefficient

N. J. BERSHAD AND A. J. ROCKMORE

Abstract—Two versions of a Gaussian signal, corrupted by independent Gaussian noises, are available for estimating the signal-to-noise power ratio. This correspondence demonstrates that the sample correlation coefficient can be used to construct an unbiased estimator that is asymptotically efficient.

$$x(t) = s(t) + n_1(t)$$

$$y(t) = s(t) + n_2(t)$$

1. **x and y** are stationary, white, Gaussian, zero-mean random processes
2. Equally distributed noise
3. Independent of each other
4. Independent of the signal

~~Stationary~~:
Macromolecule!! (local mean, variance)

~~White~~:
Macromolecule, CTF, missing directions!! A voxel is not independent from its neighbors! (local autocorrelation)

~~Zero-mean~~:
The background is zero-mean, but not the macromolecule

~~Not studied for Fourier space~~
Bershad and Rockmore derivation was performed in real-space

$$SNR(\mathbf{R}) = \frac{NCC_{x,y}(\mathbf{R})}{1 - NCC_{x,y}(\mathbf{R})}$$

Fourier coefficient

$$SNR(R) = \frac{NCC_{x,y}(R)}{1 - NCC_{x,y}(R)}$$

Fourier shell

And still it is approximately true!!

SSNR ↔ FSC, Kσ-threshold family

Saxton and Baumeister 1982

$$NCC_{f_1,f_2} = \frac{\sum_{\mathbf{r}\in\left[-\frac{N}{2},\frac{N}{2}\right]^3} f_1(\mathbf{r})f_2(\mathbf{r})}{\sqrt{\sum_{\mathbf{r}\in\left[-\frac{N}{2},\frac{N}{2}\right]^3} f_1^2(\mathbf{r})}\sqrt{\sum_{\mathbf{r}\in\left[-\frac{N}{2},\frac{N}{2}\right]^3} f_2^2(\mathbf{r})}} = \frac{c}{\sqrt{d}}$$

$$\approx \frac{\mathrm{E}\{c\}}{\sqrt{\mathrm{E}\{d\}}} + \frac{1}{\sqrt{\mathrm{E}\{d\}}}\Delta c - \frac{\mathrm{E}\{c\}}{2\left(\mathrm{E}\{d\}\right)^{\frac{3}{2}}}\Delta d$$

Taylor expansion (Error in Saxton and Baumeister 1982)

Taylor expansion of order 0!!!

$$\mathrm{E}\{NCC_{f_1,f_2}\} \approx \frac{\mathrm{E}\{c\}}{\sqrt{\mathrm{E}\{d\}}} = \frac{(N+1)^3\sigma_s^2}{\sqrt{(N+1)^6\sigma^4(1+SNR)^2}} = \frac{SNR}{1+SNR}$$

$$SNR = \frac{NCC}{1-NCC}$$

SSNR ↔ FSC

Taylor expansion of order 1!!!

$$\begin{aligned}
\mathrm{Var}\left\{NCC_{f_1,f_2}\right\} &\approx \left(\frac{1}{\sqrt{\mathrm{E}\{d\}}}\right)^2 \mathrm{Var}\left\{\Delta c\right\} + \left(\frac{\mathrm{E}\{c\}}{2(\mathrm{E}\{d\})^{\frac{3}{2}}}\right)^2 \mathrm{Var}\left\{\Delta d\right\} = \\
&= \frac{(N+1)^3\sigma^4(2SNR+1)}{(N+1)^6\sigma^4(1+SNR)^2} + \\
&\quad \left(\frac{(N+1)^3\sigma_s^2}{2((N+1)^6\sigma^4(1+SNR)^2)^{\frac{3}{2}}}\right)^2 4(N+1)^9\sigma^8(1+2SNR)(1+SNR)^2 \\
&= \frac{1+2SNR}{(N+1)^3(1+SNR)^2}\left(1+\frac{SNR^2}{(1+SNR)^2}\right)
\end{aligned}$$

SNR=0

$$\mathrm{Var}\left\{NCC_{f_1,f_2}\right\} \approx \frac{1}{(N+1)^3}$$

Resolution of the map when FSC drops below a multiple of Var

- 2 (Saxton and Baumeister, 1982; van Heel and Stoffler-Meilicke, 1985) 1-α=97.7%
- 3 (Orlova et al, 1997) 1-α=99.9%
- 5 (Radermacher, 1988) 1-α=99.99997%

$$NCC_{n_1,n_2,\alpha} = \tanh\left(\frac{z_{1-\alpha}}{\sqrt{(N+1)^3-3}}\right) \approx \frac{z_{1-\alpha}}{(N+1)^{\frac{3}{2}}}$$

SSNR ↔ FSC

Analysis for the FSC (not the NCC)

$$\mathbf{F}_1 = \mathbf{X} + \mathbf{N}_1$$

$$\mathbf{F}_2 = \mathbf{X} + \mathbf{N}_2$$

$$\tilde{\mathbf{F}}_i = C\mathbf{F}_i$$

Multivariate Gaussian

mean $C\mathbf{X}$

Cov.Matrix $\sigma^2 CC^t = \sigma^2 C^2 = \sigma^2 C$

Even filter

$$FSC(R, \Delta R) = \frac{\sum_{\mathbf{R}\in\mathfrak{S}(R,\Delta R)} F_1(\mathbf{R})F_2^*(\mathbf{R})}{\sqrt{\left(\sum_{\mathbf{R}\in\mathfrak{S}(R,\Delta R)} |F_1(\mathbf{R})|^2\right)\left(\sum_{\mathbf{R}\in\mathfrak{S}(R,\Delta R)} |F_2(\mathbf{R})|^2\right)}} = \frac{c}{\sqrt{d}}$$

$$c = \tilde{\mathbf{F}}_1^t \tilde{\mathbf{F}}_2$$

$$d = \tilde{\mathbf{F}}_1^t \tilde{\mathbf{F}}_1 \tilde{\mathbf{F}}_2^t \tilde{\mathbf{F}}_2$$

SSNR ↔ FSC

$$
\begin{aligned}
\mathrm{E}\{c\} &= \mathrm{E}\left\{\tilde{\mathbf{F}}_1^t \tilde{\mathbf{F}}_2\right\} = \mathbf{X}^t C^2 \mathbf{X} = \|C\mathbf{X}\|^2 \\
\mathrm{Var}\{c\} &= \sigma^4 \|C\|^2 (2SNR_C + 1) \\
\mathrm{E}\{d\} &= \sigma^4 \|C\|^4 (1 + SNR_C)^2 \\
\mathrm{Var}\{d\} &= 4\sigma^8 \|C\|^6 (1 + SNR_C)^2 (1 + 2SNR_C) \\
\mathrm{E}\{FSC\} &= \mathrm{E}\{NCC_{f_1,f_2}\} = \mathrm{E}\left\{\frac{c}{\sqrt{d}}\right\} \approx \frac{\mathrm{E}\{c\}}{\sqrt{\mathrm{E}\{d\}}} \\
&= \frac{\|C\mathbf{X}\|^2}{\sqrt{\sigma^4 \|C\|^4 (1+SNR_C)^2}} = \frac{\|C\mathbf{X}\|^2}{\sigma^2 \|C\|^2 (1+SNR_C)} = \frac{SNR_C}{1+SNR_C}
\end{aligned}
$$

SSNR after applying the shell filter to the signal AND noise

$$
SNR_C = \frac{\|C\mathbf{X}\|^2}{\sigma^2 \|C\|^2}
$$

Euclidean norm

Frobenius norm

$$
SNR_C = \frac{FSC}{1 - FSC}
$$

It is true for SNR_C, and after a Taylor approximation of order 0

SSNR ↔ FSC

$$\mathrm{Var}\{FSC\} \approx \frac{1+2\frac{\mathrm{E}\{FSC\}}{1-\mathrm{E}\{FSC\}}}{\|C\|^2\left(1+\frac{\mathrm{E}\{FSC\}}{1-\mathrm{E}\{FSC\}}\right)^2}\left(1+\frac{\left(\frac{\mathrm{E}\{FSC\}}{1-\mathrm{E}\{FSC\}}\right)^2}{\left(1+\frac{\mathrm{E}\{FSC\}}{1-\mathrm{E}\{FSC\}}\right)^2}\right) = \frac{1-\mathrm{E}^4\{FSC\}}{\|C\|^2}$$

Taylor order 0

Instead of

$$\mathrm{Var}\{NCC_{f_1,f_2}\} \approx \frac{1}{(N+1)^3}$$

We do not assume stationarity, zero mean or whiteness

Variance in the absence of signal depends on the shell shape

SSNR ↔ FSC

FSC=0.5 $\Rightarrow SNR_C$=1

$$\mathrm{E}\{FSC\} \approx \frac{SNR_C}{1+SNR_C}$$

The approximations come after an order 0 Taylor approximation, and they are true on average.

FSC=0.3 $\Rightarrow SNR_C^{(full)} = 1$

$$\mathrm{E}\{FSC(R,\Delta R)\} \approx \frac{SNR_C^{(half)}}{1+SNR_C^{(half)}} = \frac{\frac{SNR_C^{(full)}}{2}}{1+\frac{SNR_C^{(full)}}{2}} = \frac{SNR_C^{(full)}}{2+SNR_C^{(full)}}$$

$$\mathrm{E}\left\{FSC^{(full)}(R,\Delta R)\right\} \approx \frac{2\mathrm{E}\{FSC(R,\Delta R)\}}{1+\mathrm{E}\{FSC(R,\Delta R)\}}$$

FSC=0.143=1/7 $\Rightarrow FSC_{ref}$=0.5

$$f_1(\mathbf{r}) = x(\mathbf{r}) + n_1(\mathbf{r}) \text{ and } f_2(\mathbf{r}) = x(\mathbf{r})$$

$$\mathrm{E}\{FSC_{ref}(R,\Delta R)\} \approx \sqrt{\frac{2SNR_C}{2SNR_C+1}}$$

$$\mathrm{E}\{FSC_{ref}(R,\Delta R)\} \approx \sqrt{\frac{2\mathrm{E}\{FSC\}}{1+\mathrm{E}\{FSC\}}}$$

Rosenthal, P. B. & Henderson, R. *J. Molecular Biology*, **2003**, *333*, 721-745

SSNR ↔ FSC

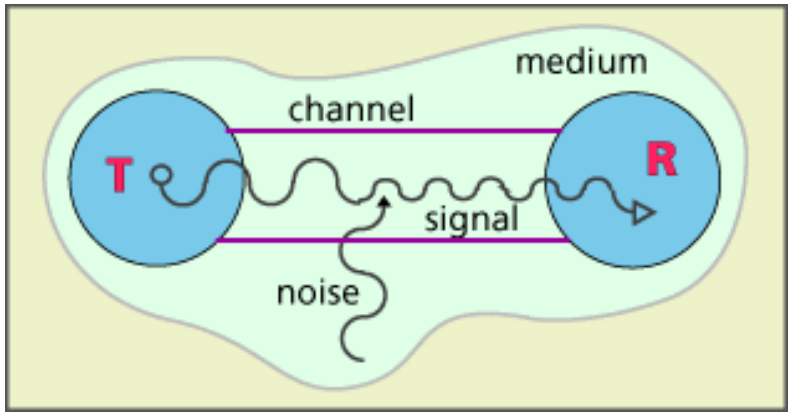
<u>0.5 bits</u>

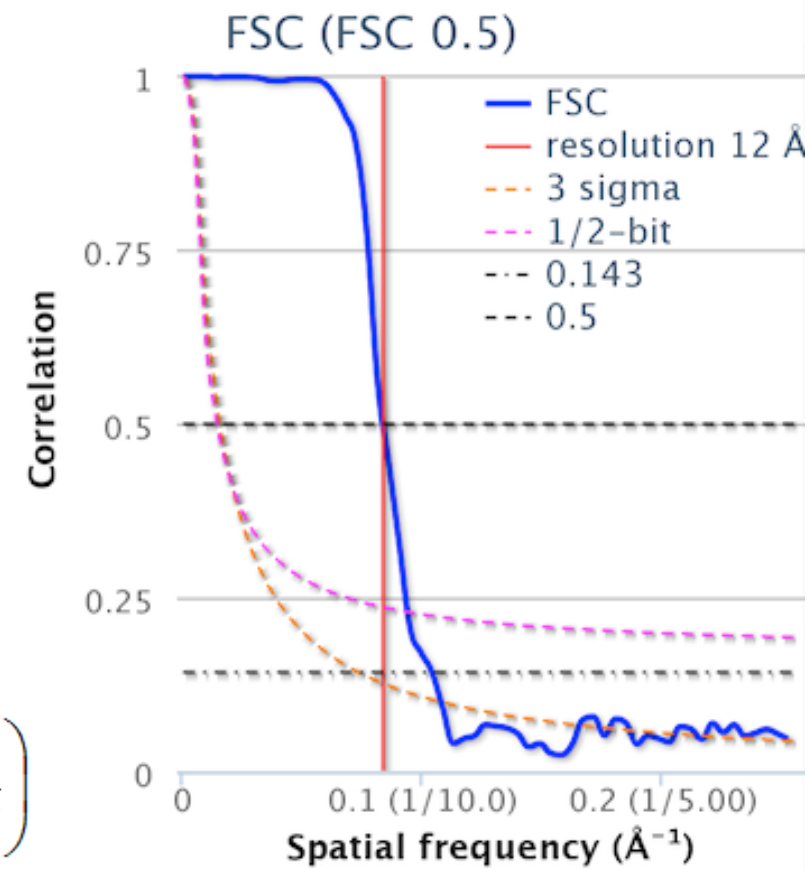
Shannon's channel capacity

van Heel, M. & Schatz, M.
J. Structural Biology, **2005**, *151*, 250-262

$$C = \log_2(1 + SNR)$$

$$C = 0.5 \Rightarrow SNR = 0.4142$$





~~There is no channel~~
~~Real statistics~~

$$C = B\log_2\left(1 + \frac{S}{N}\right)$$

SSNR ↔ FSC

Relationship between SSNR and FSC

$$\mathrm{E}\left\{FSC(R,\Delta R)\right\} \approx \frac{SNR_C}{1+SNR_C} = \frac{\mathrm{E}\left\{\widehat{SSNR}(R,\Delta R)\right\}}{2K+\mathrm{E}\left\{\widehat{SSNR}(R,\Delta R)\right\}}$$

Depends on the number of images, the frequency and the orientation distribution

Penczek, P. *J. Structural Biology,* **2002**, *138*, 34-46

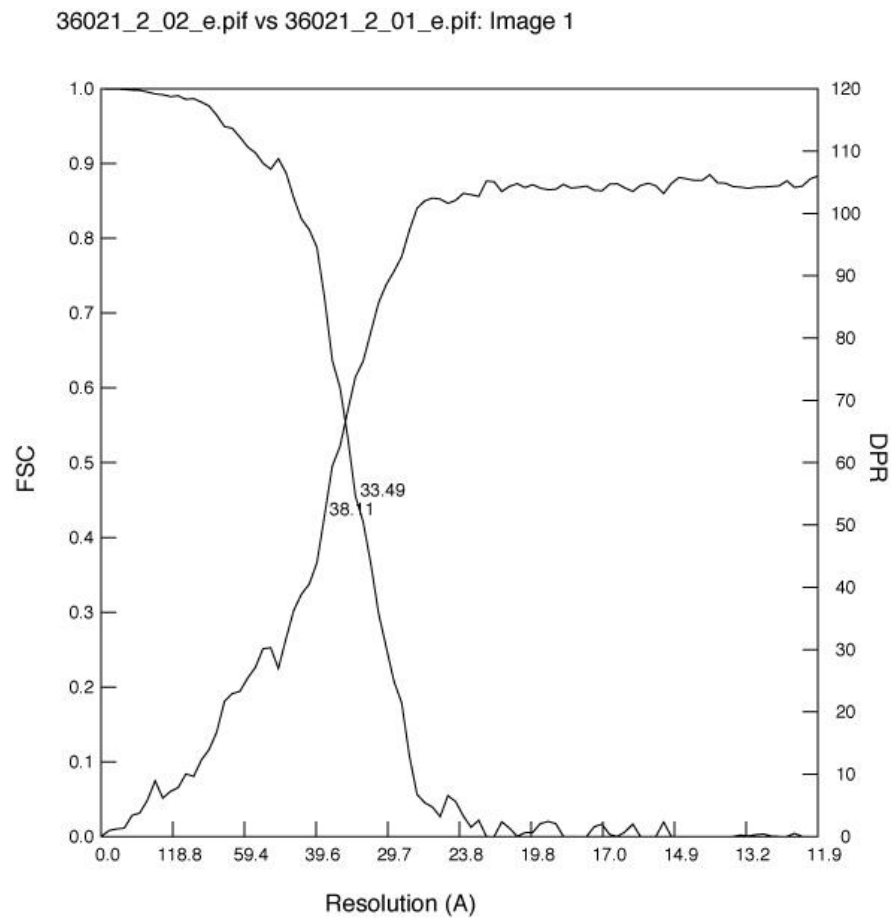
$$\mathrm{E}\left\{FSC(R,\Delta R)\right\} = \frac{\mathrm{E}\left\{\widehat{SSNR}(R,\Delta R)\right\}}{1+\mathrm{E}\left\{\widehat{SSNR}(R,\Delta R)\right\}}$$

There was a mistake when going from 2D to 3D.

Differential Phase Residual

$$DPR(R,\Delta R) = \sqrt{\frac{\sum_{\mathbf{R}\in\mathfrak{S}(R,\Delta R)} (|F_1(\mathbf{R})|+|F_2(\mathbf{R})|)(\measuredangle\{F_1(\mathbf{R}),F_2(\mathbf{R})\})^2}{\sum_{\mathbf{R}\in\mathfrak{S}(R,\Delta R)} (|F_1(\mathbf{R})|+|F_2(\mathbf{R})|)}}$$

$$= \sqrt{\sum_{\mathbf{R}\in\mathfrak{S}(R,\Delta R)} w(\mathbf{R})\,(\measuredangle\{F_1(\mathbf{R}),F_2(\mathbf{R})\})^2},$$



Differential Phase Residual

$$\measuredangle\{F_1(\mathbf{R})\} = \tan^{-1}\frac{\mathrm{Im}\{X(\mathbf{R})\} + \mathrm{Im}\{N_i(\mathbf{R})\}}{\mathrm{Re}\{X(\mathbf{R})\} + \mathrm{Re}\{N_i(\mathbf{R})\}}$$

$$\Delta\theta_i(\mathbf{R}) = \measuredangle\{F_i(\mathbf{R})\} - \measuredangle\{X(\mathbf{R})\}$$

<u>Sinusoidal corrupted by narrowband Gaussian noise</u>

$$p_{\Delta\theta_i(\mathbf{R})}(\Delta\theta_i(\mathbf{R})) = \frac{1}{2\pi}e^{-\frac{|X(\mathbf{R})|^2}{\sigma^2}}\left(1 + 2\frac{|X(\mathbf{R})|}{\sigma}\sqrt{\pi}\cos(\Delta\theta_i(\mathbf{R}))\, e^{\frac{|X(\mathbf{R})|^2\cos^2(\Delta\theta_i(\mathbf{R}))}{\sigma^2}} Q\left(\frac{\sqrt{2}|X(\mathbf{R})|\cos(\Delta\theta_i(\mathbf{R}))}{\sigma}\right)\right)$$

<u>Particular case: No signal</u>

$$p_{\Delta\theta_i(\mathbf{R})}(\Delta\theta_i(\mathbf{R})) = \frac{1}{2\pi}$$

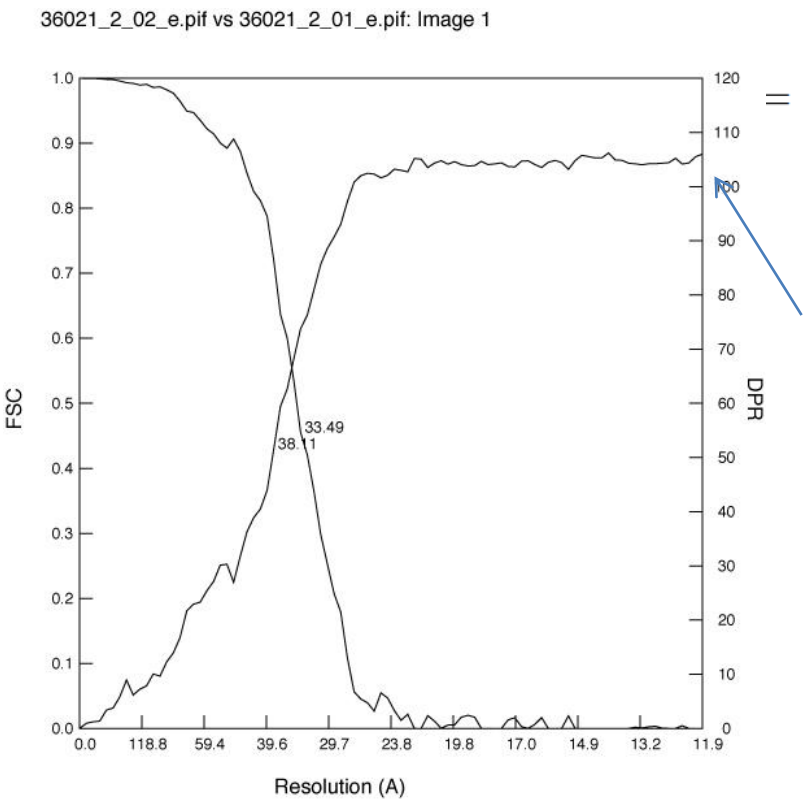
<u>Particular case: Strong signal</u>

$$p_{\Delta\theta_i(\mathbf{R})}(\Delta\theta_i(\mathbf{R})) \approx \frac{1}{\sqrt{\pi}}\frac{|X(\mathbf{R})|}{\sigma}e^{-\frac{|X(\mathbf{R})|^2}{\sigma^2}(\Delta\theta_i(\mathbf{R}))^2}$$

Differential Phase Residual

In the case of SNR=0

$$
\begin{aligned}
\mathrm{E}\left\{DPR(R,\Delta R)\right\} &= \sqrt{\frac{|\mathfrak{S}(R,\Delta R)|\frac{\sigma\pi^2\sqrt{\pi}}{3}}{|\mathfrak{S}(R,\Delta R)|\sigma\sqrt{\pi}}} = \frac{\pi}{\sqrt{3}} \\
\mathrm{Var}\left\{DPR(R,\Delta R)\right\} &= \frac{1}{4}\frac{1}{|\mathfrak{S}(R,\Delta R)|\frac{\sigma\pi^2\sqrt{\pi}}{3}|\mathfrak{S}(R,\Delta R)|\sigma\sqrt{\pi}}|\mathfrak{S}(R,\Delta R)|\sigma^2\pi^4\left(\frac{2}{5}-\frac{\pi}{90}\right) \\
&\quad + \frac{1}{4}\frac{|\mathfrak{S}(R,\Delta R)|\frac{\sigma\pi^2\sqrt{\pi}}{3}}{\left(|\mathfrak{S}(R,\Delta R)|\sigma\sqrt{\pi}\right)^3}|\mathfrak{S}(R,\Delta R)|2\sigma^2\left(1-\frac{\pi}{4}\right) \\
&= \frac{\pi}{20|\mathfrak{S}(R,\Delta R)|}\left(\frac{28}{3}-\pi\right).
\end{aligned}
$$



Taylor approximation of order 1

103º and not 90º

Unser, M.; et al. *Ultramicroscopy,* **1987**, *23*, 39-52

Differential Phase Residual

In the case of large SNR

$$\mathrm{E}\{DPR(R,\Delta R)\} = \frac{1}{\sqrt{SNR(R)}}$$

$$\mathrm{Var}\{DPR(R,\Delta R)\} = \frac{8+8SNR(R)-\frac{3\pi}{4}L^2_{\frac{1}{2}}(-SNR(R))}{4\pi|\mathfrak{S}(R,\Delta R)|SNR(R)L^2_{\frac{1}{2}}(-SNR(R))}$$

Laguerre polynomial of order 1/2

$E\{DPR(R,\Delta R)\} = 45° \Rightarrow SNR(R) = 1.6$ Frank, J. 2006

This is not large SNR

Chebyshev's inequality

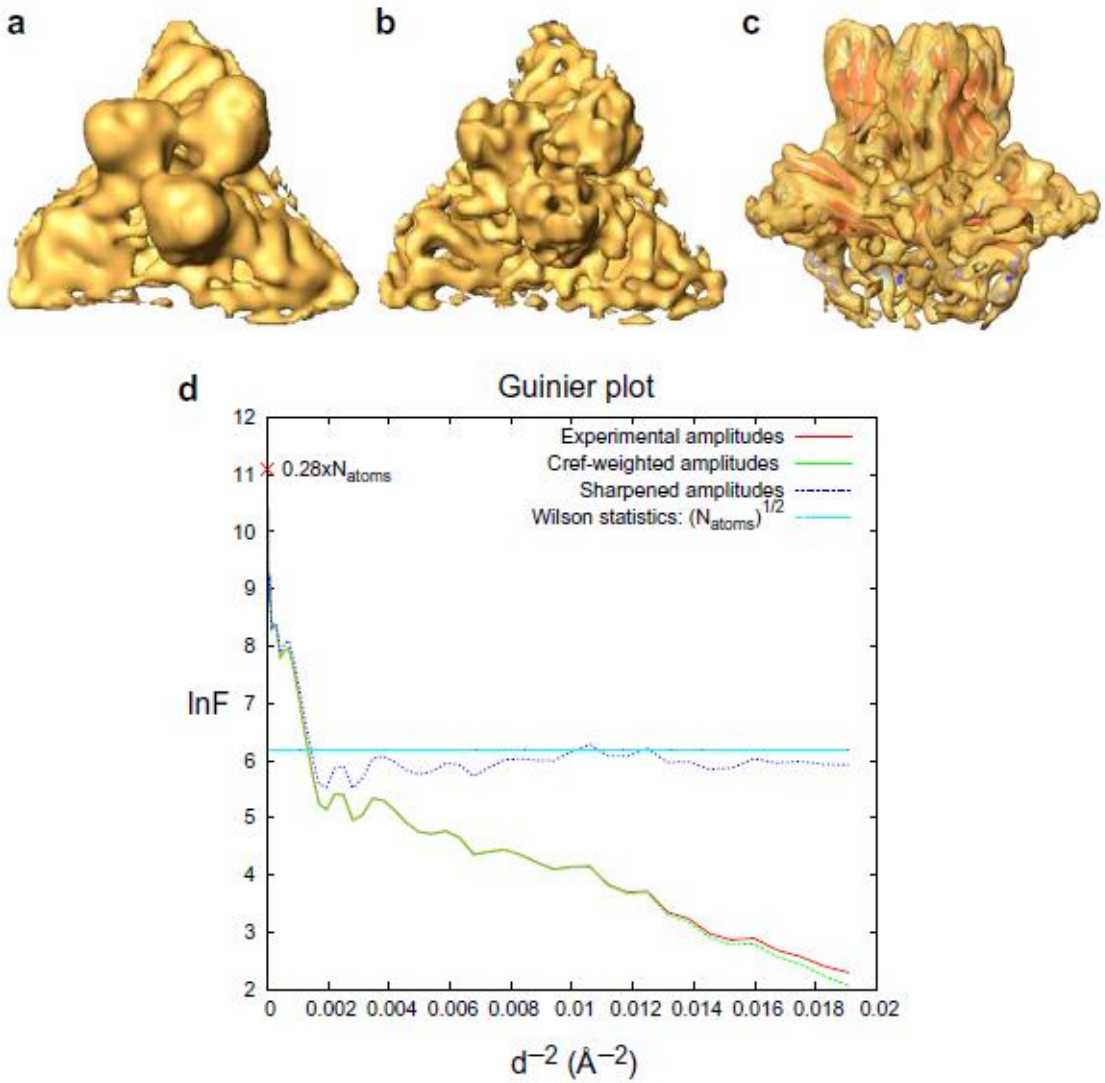
$$\Pr\left\{|DPR(R,\Delta R) - \mathrm{E}\{DPR(R,\Delta R)\}| \geq \sqrt{\frac{\mathrm{Var}\{DPR(R,\Delta R)\}}{\alpha}}\right\} \leq \alpha$$

$$DPR(R,\Delta R) \geq \frac{\pi}{\sqrt{3}} - \sqrt{\frac{\pi}{20|\mathfrak{S}(R,\Delta R)|\alpha}\left(\frac{28}{3}-\pi\right)}$$

$FSC = 2\sigma \Rightarrow DPR(R) = 85°$

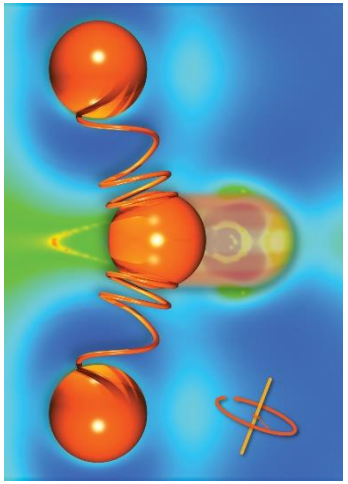
Radermacher, M. *J. Electron Microscopy Technique*, **1988**, *9*, 359-394
de la Fraga, L. G. *Ultramicroscopy*, **1995**, *60*, 385-391

B-factor correction



Fernández, J. J. *J. Structural Biology,* **2008***, 164*, 170-175

B-factor due to thermal vibrations



$$f(\mathbf{r}) = \sum_{j=1}^{N_{\text{atoms}}} w_j \delta(\mathbf{r} - \mathbf{r}_j)$$

Infinitely thin atoms

$$F(\mathbf{R}) = \sum_{j=1}^{N_{\text{atoms}}} w_j e^{-i2\pi\langle \mathbf{R}, \mathbf{r}_j \rangle} = \sum_{j=1}^{N_{\text{atoms}}} w_j \cos(2\pi \langle \mathbf{R}, \mathbf{r}_j \rangle) - i \sum_{j=1}^{N_{\text{atoms}}} w_j \sin(2\pi \langle \mathbf{R}, \mathbf{r}_j \rangle)$$

Uniform distribution

<u>Wilson statistics</u>

Atom location follows uniform distribution within a box

$$\mathrm{E}\{\mathrm{Re}\{F(\mathbf{R})\}\} = 0$$

$$\mathrm{Var}\{\mathrm{Re}\{F(\mathbf{R})\}\} = \frac{1}{2} \sum_{j=1}^{N_{\text{atoms}}} w_i^2$$

σ^2_{atoms}

$$\mathrm{E}\{|F(\mathbf{R})|^2\} = 2\frac{1}{2}\sigma^2_{\text{atoms}}$$

$$\mathrm{Var}\{|F(\mathbf{R})|^2\} = 2 \cdot 2 \left(\frac{\sigma^2_{\text{atoms}}}{2}\right)^2 = \sigma^4_{\text{atoms}}$$

B-factor due to thermal vibrations

$$F_{\text{thermal}}(\mathbf{R}) = \sum_{j=1}^{N_{\text{atoms}}} w_j e^{-i2\pi\langle \mathbf{R}, \mathbf{r}_j + \Delta\mathbf{r}_j\rangle}$$

Thermal vibration

Taylor expansion of order 2

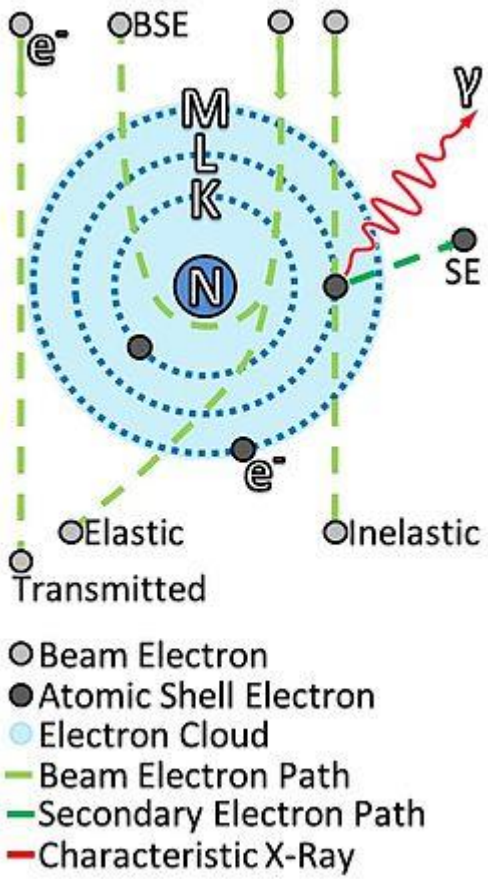
$$F_{\text{thermal}}(\mathbf{R}) \approx \sum_{j=1}^{N_{\text{atoms}}} w_j e^{-i2\pi\langle \mathbf{R}, \mathbf{r}_j\rangle} \left(1 - i2\pi\,\langle \mathbf{R}, \Delta\mathbf{r}_j\rangle + \frac{(-i2\pi\langle \mathbf{R}, \Delta\mathbf{r}_j\rangle)^2}{2}\right)$$

$$\begin{aligned}
\mathrm{E}_{\Delta\mathbf{r}_j}\left\{F_{\text{thermal}}(\mathbf{R})\right\} &= \sum_{j=1}^{N_{\text{atoms}}} w_j e^{-i2\pi\langle \mathbf{R}, \mathbf{r}_j\rangle} \left(1 - 2\pi^2 \mathrm{E}_{\Delta\mathbf{r}_j}\left\{\langle \mathbf{R}, \Delta\mathbf{r}_j\rangle^2\right\}\right) \\
&= \sum_{j=1}^{N_{\text{atoms}}} w_j e^{-i2\pi\langle \mathbf{R}, \mathbf{r}_j\rangle} \left(1 - 2\pi^2 \sigma^2_{\Delta_x} |\mathbf{R}|^2\right) \\
&\approx \sum_{j=1}^{N_{\text{atoms}}} w_j e^{-i2\pi\langle \mathbf{R}, \mathbf{r}_j\rangle} e^{-2\pi^2 \sigma^2_{\Delta_x} |\mathbf{R}|^2} \\
&= F(\mathbf{R}) e^{-2\pi^2 \sigma^2_{\Delta_x} |\mathbf{R}|^2} = F(\mathbf{R}) e^{-B|\mathbf{R}|^2}.
\end{aligned}$$

B-factor

This is true on average, after a number of approximations and it is dealing with distributions with a wild variance

B-factor due to electron scattering



Under Born approximation (the scattered wave from a single scatterer is not affected by the presence of other scatterers)

$$I(\mathbf{R}) = |F(\mathbf{R})|^2 = \iint f(\mathbf{r}_1) f(\mathbf{r}_2) \cos(\langle \mathbf{R}, \mathbf{r}_1 - \mathbf{r}_2 \rangle) d\mathbf{r}_1 d\mathbf{r}_2$$

Radial average → Debye scattering equation

$$\overline{|F(R)|^2} = \iint f(\mathbf{r}_1) f(\mathbf{r}_2) \frac{\sin(R\|\mathbf{r}_1 - \mathbf{r}_2\|)}{R\|\mathbf{r}_1 - \mathbf{r}_2\|} d\mathbf{r}_1 d\mathbf{r}_2$$

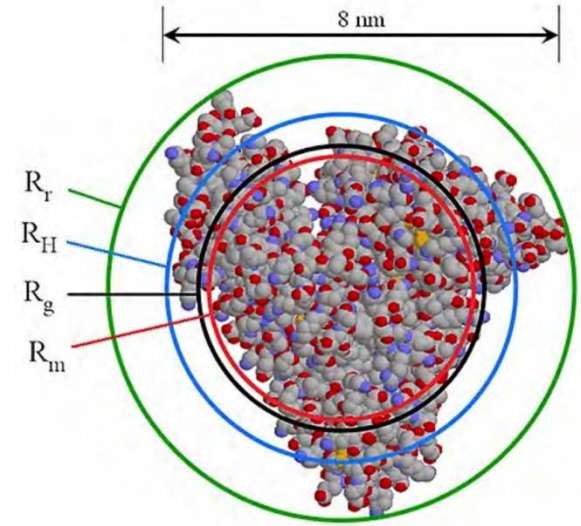
$$= \int_0^D r^2 \Phi_f(r) \frac{\sin(Rr)}{Rr} dr$$

$$\Phi_f(r) = \int_0^{2\pi} \int_0^{\pi} \int f(\mathbf{r}_1) f(\mathbf{r}_1 + \mathbf{r}) \sin(\alpha) d\mathbf{r}_1 d\alpha d\beta$$

Radial average of the autocorrelation function

B-factor due to electron scattering

Radius of rotation, hydrodynamic radius, radius of gyration, mass radius



Guinier approximation

$$\frac{\sin(Rr)}{Rr} = 1 - \frac{R^2 r^2}{6} + \frac{R^4 r^4}{100} - \ldots$$

$$\begin{aligned}
\overline{|F(R)|^2} &\approx \int_0^D r^2 \Phi_f(r) dr - R^2 \int_0^D \frac{r^4}{6} \Phi_f(r) dr \\
&= \int_0^D r^2 \Phi_f(r) dr \left(1 - R^2 \frac{\int_0^D \frac{r^4}{6} \Phi_f(r) dr}{\int_0^D r^2 \Phi_f(r) dr} \right) \\
&= F_0^2 \left(1 - \frac{R_g^2}{3} R^2 \right)
\end{aligned}$$

Radius of gyration

For infinitely small atoms

Same as Wilson

$$F_0^2 = \overline{|F(0)|^2} = \sum_{i=1}^{N_{\text{atoms}}} \sum_{j=1}^{N_{\text{atoms}}} w_i w_j \delta(\|\mathbf{r}_i - \mathbf{r}_j\|) = \sum_{j=1}^{N_{\text{atoms}}} w_j^2$$

If all atoms have equal weight, it Is the number of atoms

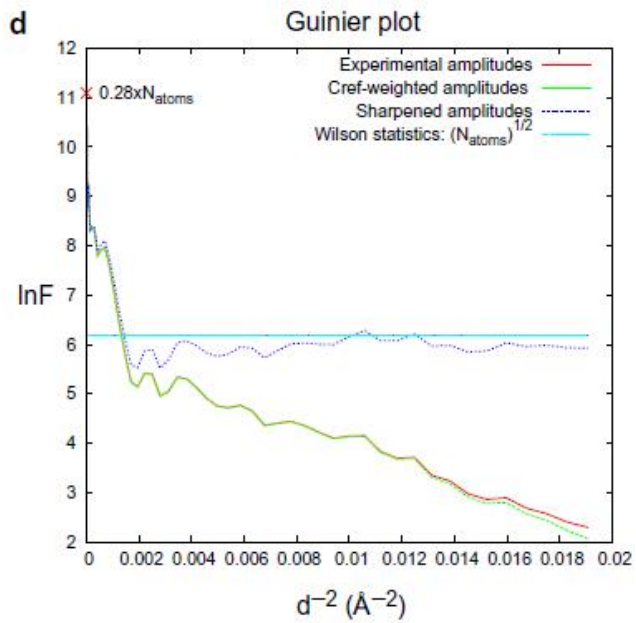
Rosenthal, P. B. *J. Molecular Biology,* **2003**, *333*, 721-745

B-factor due to electron scattering

Guinier approximation

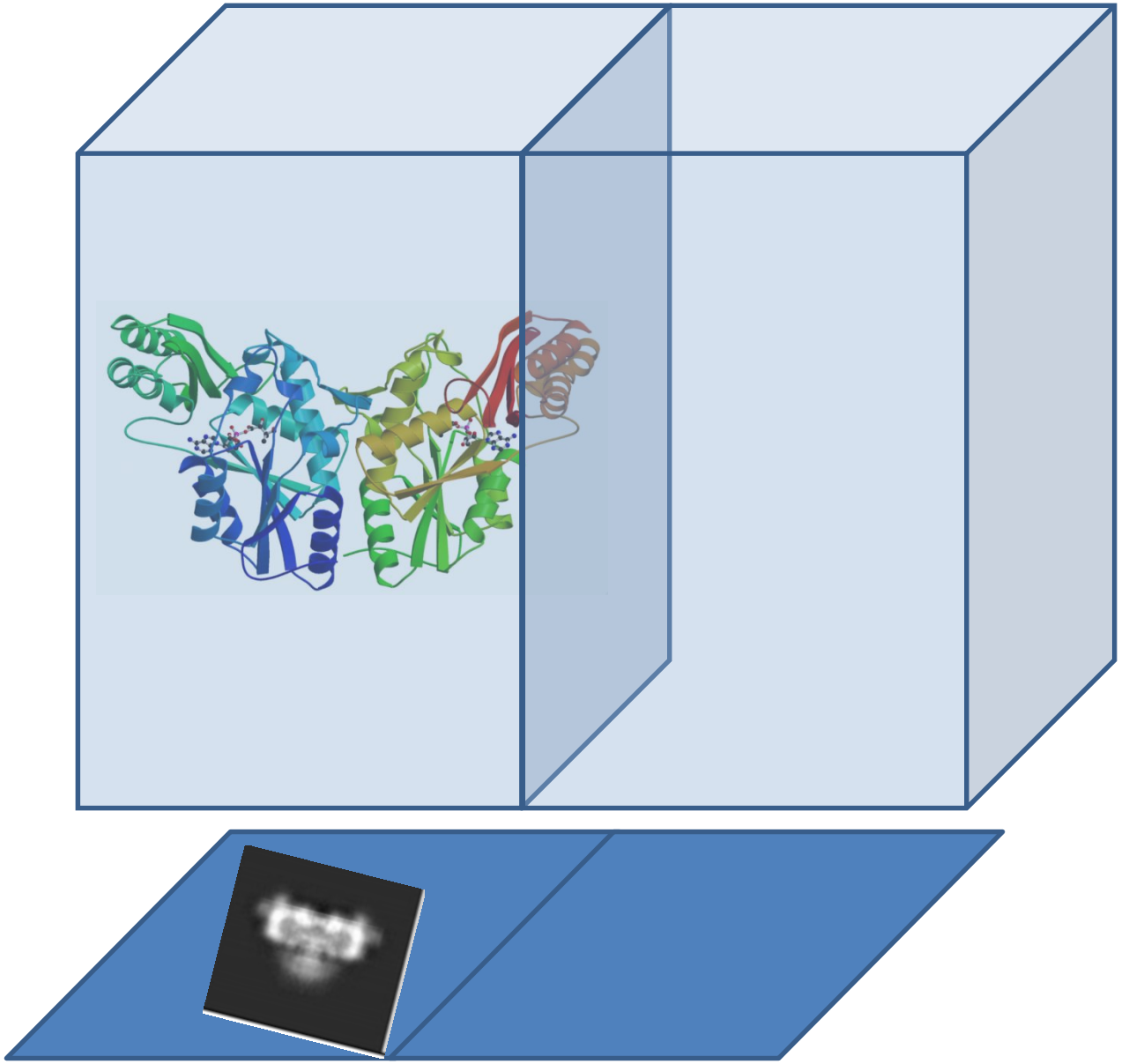
$$\overline{|F(R)|^2} \approx F_0^2\left(1-\frac{R_g^2}{3}R^2\right) \approx F_0^2 e^{-\frac{R_g^2}{3}R^2}$$

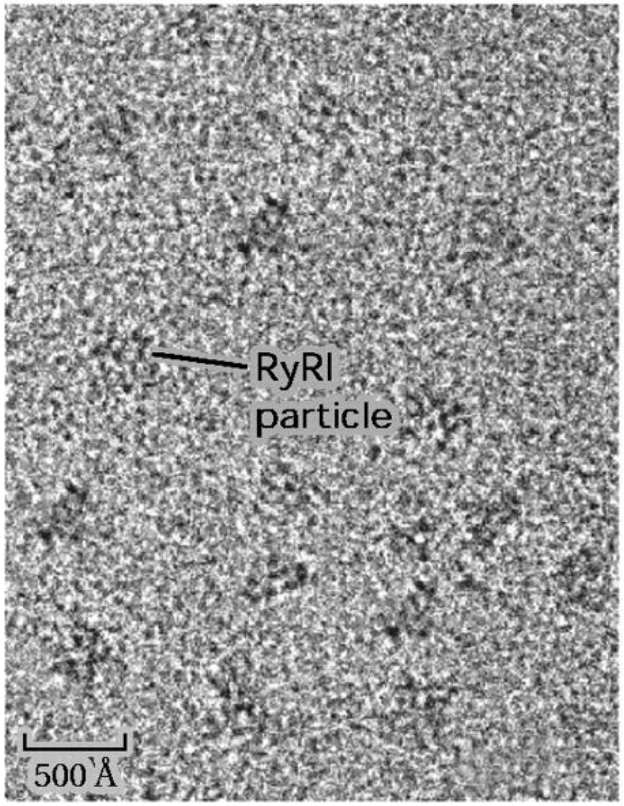
$$\log \overline{|F(R)|^2} = \log(F_0^2) - \frac{R_g^2}{3}R^2$$



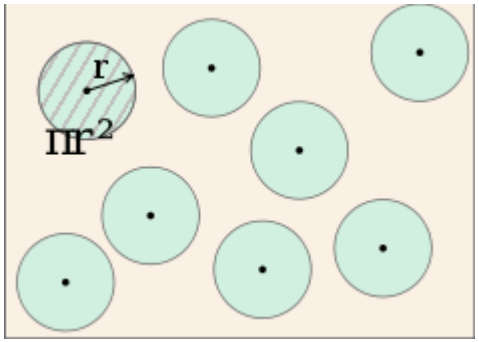
The curve is flat only for infinitely small atoms, but not for a macromolecule

SNR (Dose) for detecting a particle in ice





SNR (Dose) for detecting a particle in ice



$$\sigma_Z = \frac{1.4 \cdot 10^{-4} Z^{3/2}}{\beta^2}\left(1 - \frac{0.26Z}{137\beta}\right) [\text{Å}^2/\text{atom}]$$

$$\beta = \sqrt{1 - \left(\frac{m_e c^2}{q_e V_0 + m_e c^2}\right)^2}$$

e.g., carbon atom cross section at 200kV is 41.9pm^2

Between 2Å and 5Å, the fraction of elastically scattered electrons that reach a resolution R is 1/R

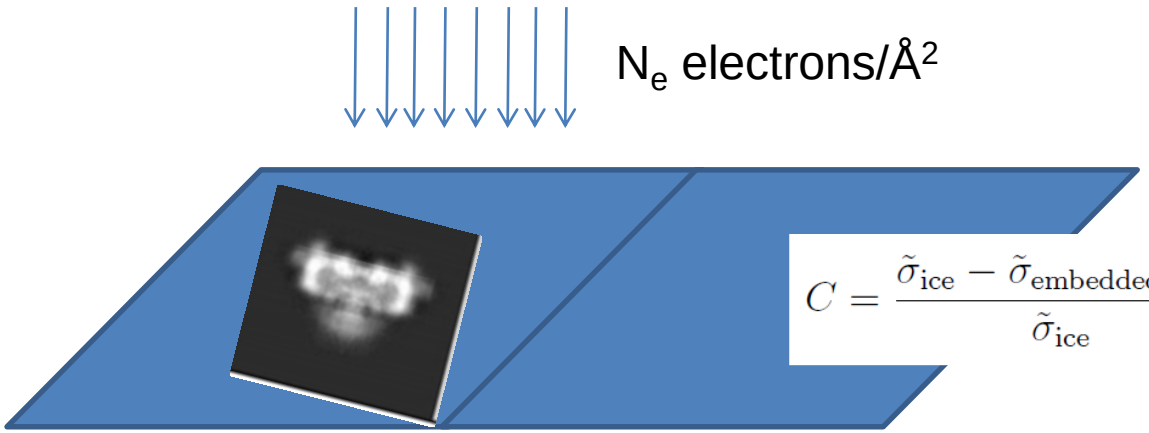
Langmore, J. P. *Ultramicroscopy,* **1992**, *46*, 349-373

$$\tilde{\sigma}_{\text{prot}} = D\rho_{\text{prot}}\frac{1}{MW_{\text{C}}}R_{max}\sigma_{Z=6}$$

$$\tilde{\sigma}_{\text{ice}} = H\rho_{\text{ice}}\frac{1}{MW_{\text{O}}}R_{max}\sigma_{Z=8}$$

$$\tilde{\sigma}_{\text{embedded.prot}} = D\rho_{\text{prot}}\frac{1}{MW_{\text{C}}}R_{max}\sigma_{Z=6} + (H-D)\rho_{\text{ice}}\frac{1}{MW_{\text{O}}}R_{max}\sigma_{Z=8}$$

SNR (Dose) for detecting a particle in ice



$$C = \frac{\tilde{\sigma}_{\text{ice}} - \tilde{\sigma}_{\text{embedded.prot}}}{\tilde{\sigma}_{\text{ice}}} = \frac{D}{H}\left(1 - \frac{\rho_{\text{prot}}}{\rho_{\text{ice}}}\frac{MW_{\text{O}}}{MW_{\text{C}}}\frac{\sigma_{Z=6}}{\sigma_{Z=8}}\right)$$

$$C = 0.0047\frac{D}{H}$$ (200kV)

$$\lambda_{\text{embedded.prot}} = N_e \tilde{\sigma}_{\text{embedded.prot}} D^2$$

$$\lambda_{\text{ice}} = N_e \tilde{\sigma}_{\text{ice}} D^2$$

The number of scattered electrons follow a Poisson

We can detect a particle with confidence level 1-α and statistical power β if

$$\lambda_{\text{embedded.prot}} + z_{1-\alpha}\sqrt{\lambda_{\text{embedded.prot}}} < \lambda_{\text{ice}} - z_{1-\beta}\sqrt{\lambda_{\text{ice}}}$$

$$N_e > \left(\frac{z_{1-\alpha}\sqrt{\tilde{\sigma}_{\text{embedded.prot}}} + z_{1-\beta}\sqrt{\tilde{\sigma}_{\text{ice}}}}{D(\tilde{\sigma}_{\text{ice}} - \tilde{\sigma}_{\text{embedded.prot}})}\right)^2 = \frac{\left(z_{1-\alpha}\sqrt{\frac{\tilde{\sigma}_{\text{embedded.prot}}}{\tilde{\sigma}_{\text{ice}}}} + z_{1-\beta}\right)^2}{\tilde{\sigma}_{\text{ice}}}\frac{1}{D^2C^2}$$

This factor is set to 25

This analysis is for 2D detection, not for 3D resolution

Glaeser, R. M. *J. Structural Biology*, **1999**, *128*, 3-14

Number of particles for a given SNR

$$\tilde{\sigma} = \int |F(\mathbf{R})|^2 d\mathbf{R} = 2\pi \int_0^{R_{max}} |F(R)|^2 R dR = 2\pi F_0^2$$

Radially symmetric

Henderson, R. *J. Molecular Biology*, **1995**, *247*, 726-738

$$\tilde{\sigma} = 2\pi F_0^2 R_{max} \left(1 - \frac{R_g^2}{9} R_{max}^3\right)$$

Taylor approximation of order 2

$$|F_{\text{prot}}(R_{max})|^2 = \frac{D\rho_{\text{prot}}\sigma_{Z=6}\left(1 - \frac{R_{g,prot}^2}{3} R_{max}^2\right)}{2\pi MW_C\left(1 - \frac{R_{g,prot}^2}{9} R_{max}^3\right)}$$

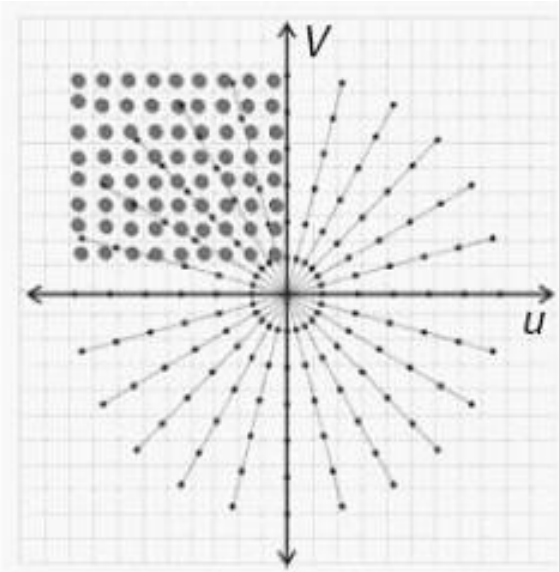
$$|F_{\text{surr.ice}}(R_{max})|^2 = \frac{(H-D)\rho_{\text{ice}}\sigma_{Z=8}\left(1 - \frac{R_{g,ice}^2}{3} R_{max}^2\right)}{2\pi MW_O\left(1 - \frac{R_{g,ice}^2}{9} R_{max}^3\right)}$$

Number of particles for a given SNR

$$SNR = \frac{|F_{\text{prot}}(R_{max})|^2}{\frac{1}{N_{\text{meas}}}|F_{\text{surr.ice}}(R_{max})|^2}$$

$$\begin{aligned} N_{\text{meas}} &> SNR_0 \frac{|F_{\text{surr.ice}}(R_{max})|^2}{|F_{\text{prot}}(R_{max})|^2} \\ &= SNR_0 \left(\frac{H-D}{D}\right) \left(\frac{\rho_{\text{prot}} MW_C \sigma_{Z=8}}{\rho_{\text{ice}} MW_O \sigma_{Z=6}}\right) \left(\frac{3-R_{g,ice}^2 R_{max}^2}{3-R_{g,prot}^2 R_{max}^2}\right) \left(\frac{9-R_{g,prot}^2 R_{max}^3}{9-R_{g,ice}^2 R_{max}^3}\right) \end{aligned}$$

Number of times a Fourier coefficient needs to be measured to achieve a given SNR

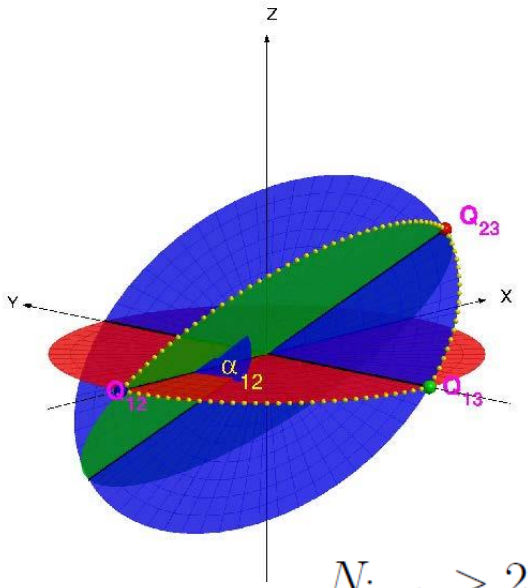


$$\frac{\pi}{N_{imgs}} R_{max} < \frac{1}{D} \Rightarrow N_{imgs} > \pi D R_{max}$$

Bracewell, R.*Proc. IRE,* **1958**, *46*, 97
Bracewell, R. *The Astrophysical J.,* **1967**, *150*, 427-434
Henderson, R. *J. Molecular Biology,* **1995**, *247*, 726-738

Valid only for single-tilt axis geometry

Number of particles for a given SNR



$$\rho_{\text{total samples},R} = \frac{D^2 N_{\text{imgs}}}{2R}$$

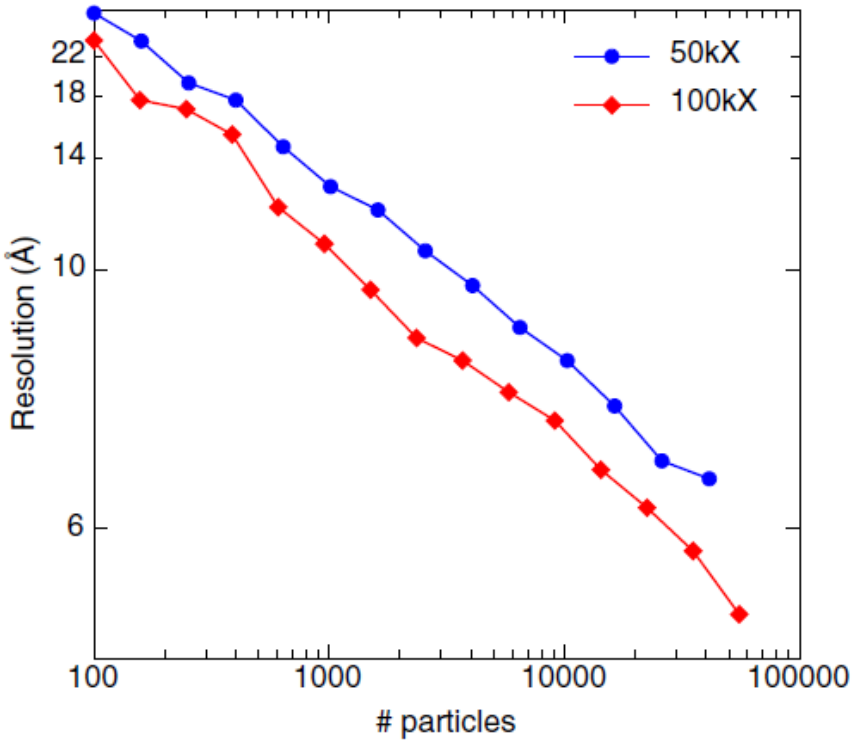
Surface density of samples at a frequency R when the samples are separated by 1/D

$$N_{\text{imgs}} > 2SNR_0 \left(\frac{H-D}{D}\right) \left(\frac{\rho_{\text{prot}} MW_C \sigma_{Z=8}}{\rho_{\text{ice}} MW_O \sigma_{Z=6}}\right) R_{max}(3 + R^2_{g,prot} R^2_{max})$$

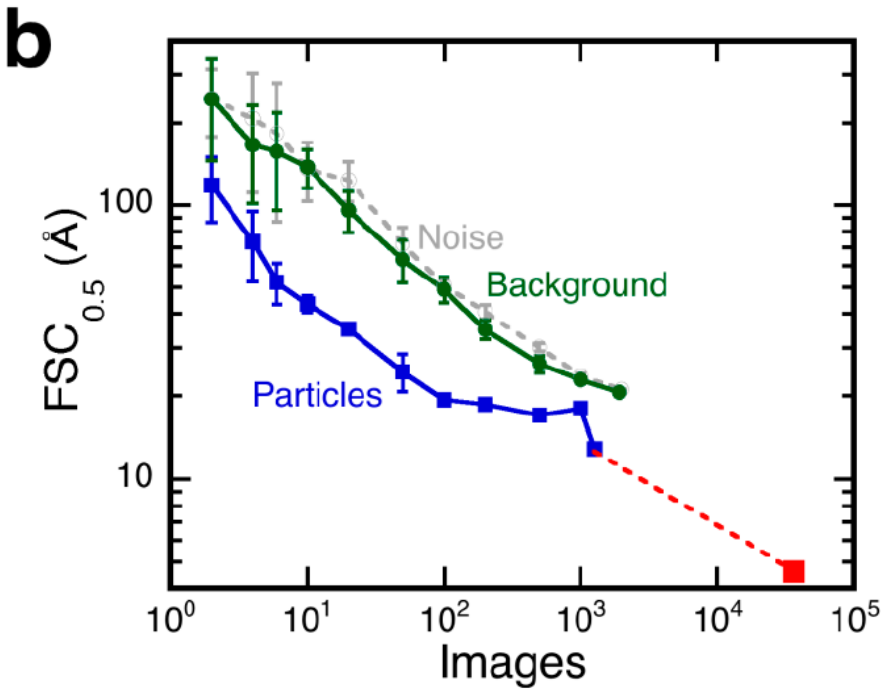
Depends on:

- Ice thickness
- Molecular weight
- Molecular shape
- Microscope voltage
- Target resolution (**more than linearly!!**)

Number of particles for a given SNR



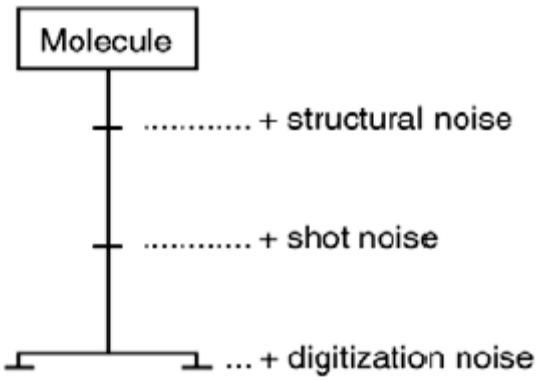
Stagg, S. M. *J Struct Biol,* **2008**, *163*, 29-39



Heymann, B. *AIMS Biophysics,* **2015**, *2*, 21-35

$$N_{\text{imgs}} = 10^{\frac{R_{max} - R_0(M,V)}{c}}$$

How to measure the SNR?



structure

microscope (3D/2D/3D)

camera

Baxter, W. T. *J. Structural Biology,* **2009**, *166*, 126-132

$$1 + \frac{1}{SNR_{comp}} = \left(1 + \frac{1}{SNR_1}\right)\left(1 + \frac{1}{SNR_2}\right)\left(1 + \frac{1}{SNR_3}\right)$$

Frank, J. *Nature,* **1975**, *256*, 376-379

$$G_1 G_2 G_3 + \frac{1}{SNR_{comp}} = \left(G_1 + \frac{1}{SNR_1}\right)\left(G_2 + \frac{1}{SNR_2}\right)\left(G_3 + \frac{1}{SNR_3}\right)$$

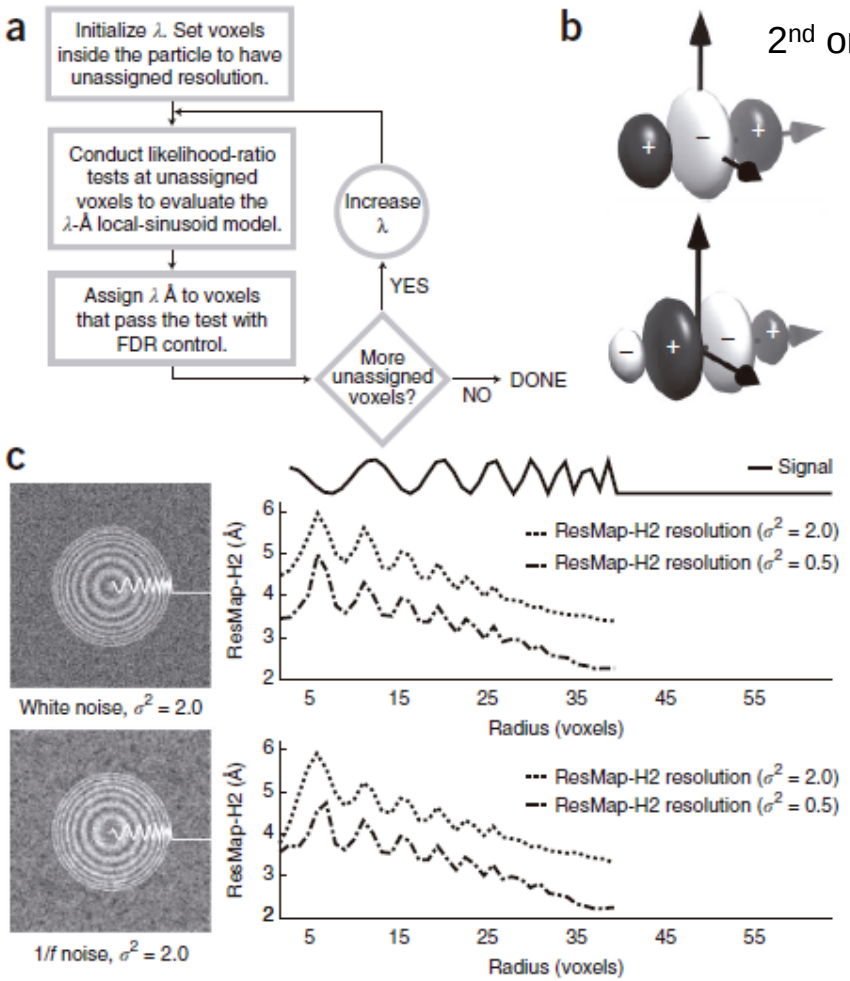
There is energy loss caused by the particle that depends on frequency

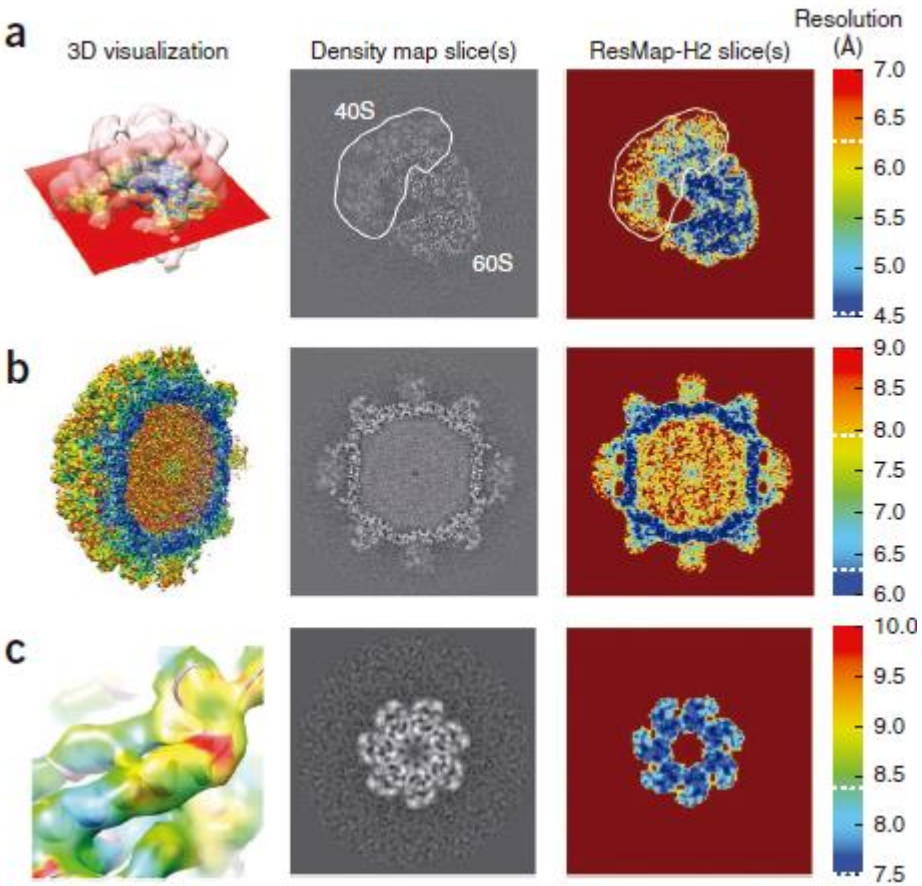
CTF and angular distribution are not considered

DQE is not 1 at all frequencies

Local Resolution

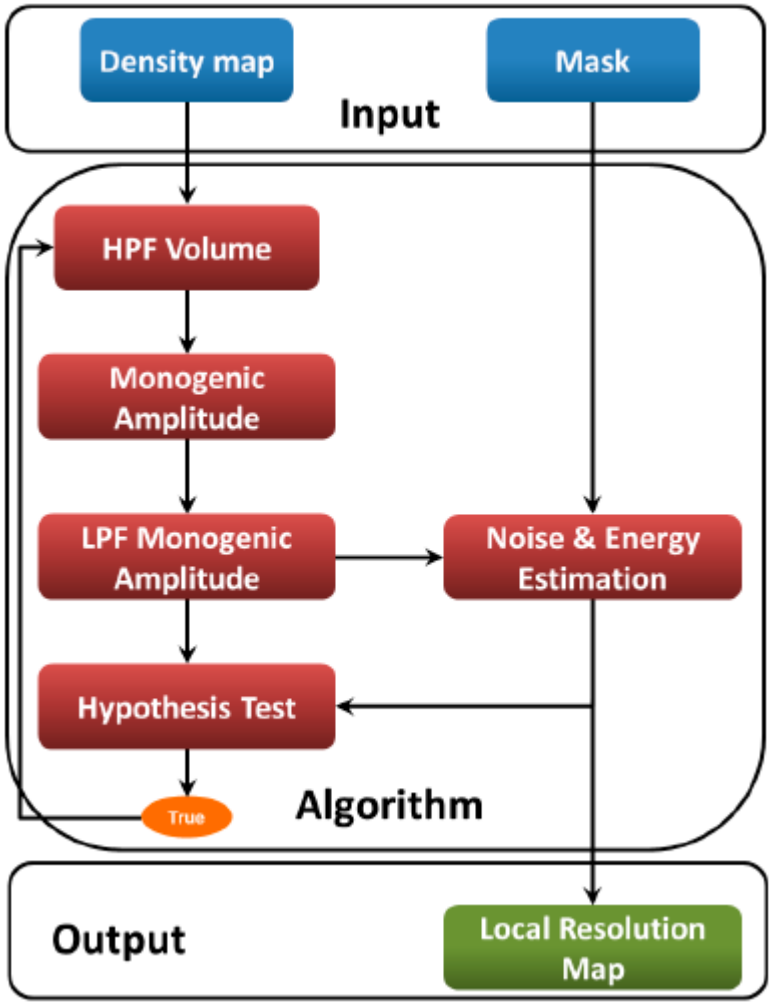
2nd order Hermite polynomials x Gaussian

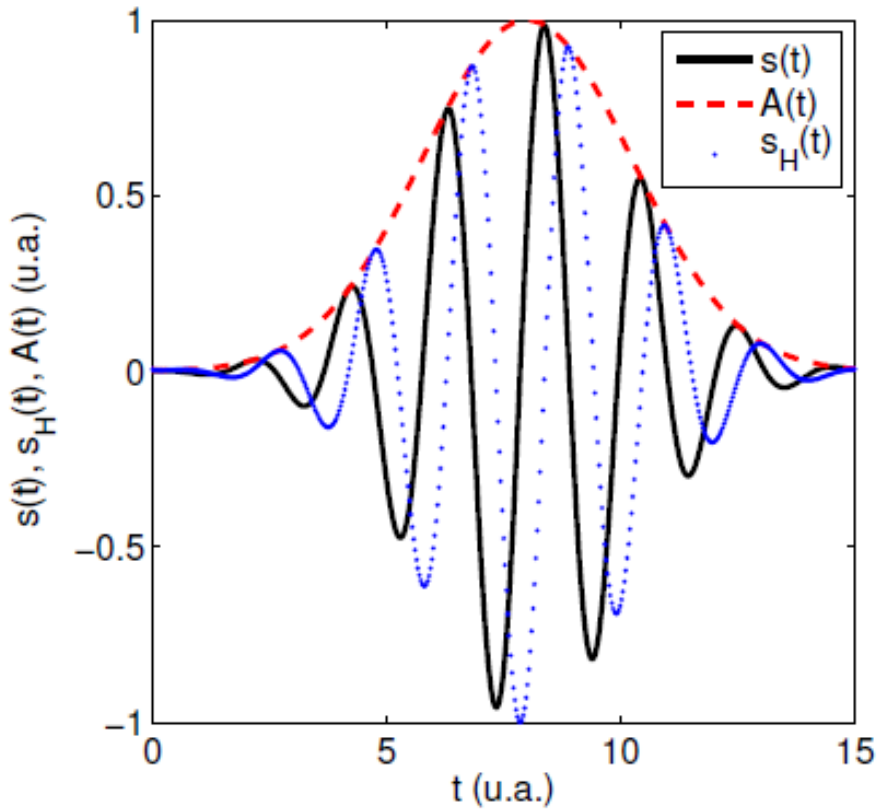




Kucukelbir, A.; Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. Nature Methods, 2014, 11, 63-65

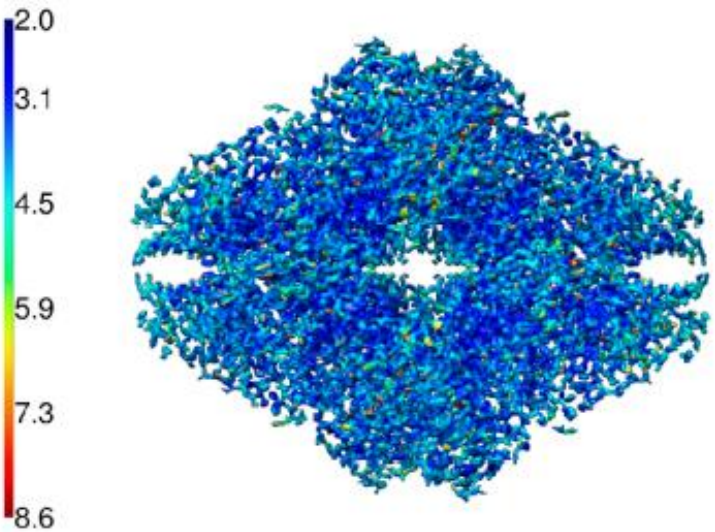
Local Resolution

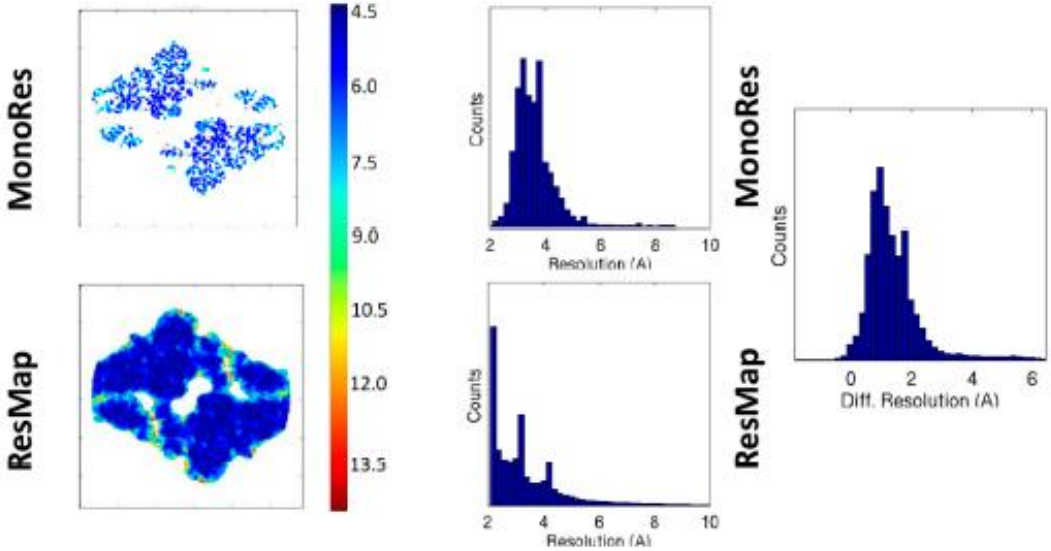




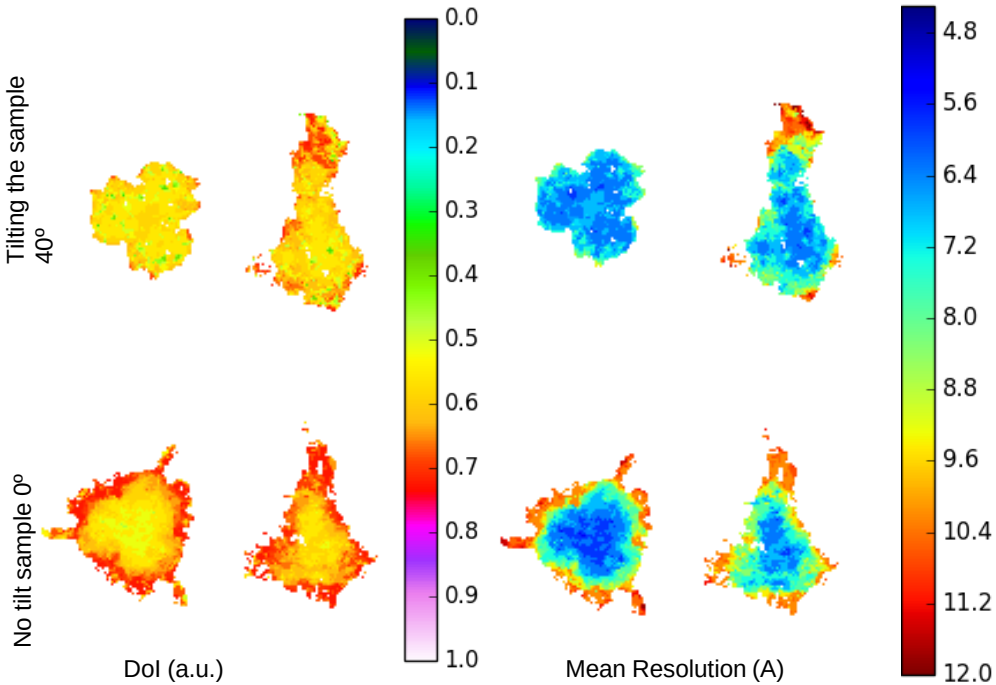
Jose Luis Vilas, Josué Gómez-Blanco, Pablo Conesa, Roberto Melero, José Miguel de la Rosa-Trevín, Joaquin Otón, Jesús Cuenca, Roberto Marabini, José María Carazo, Javier Vargas, Carlos Oscar S.Sorzano. MonoRes: Automatic and Accurate Estimation of Local Resolution for Electron Microscopy Maps. Volume 26, Issue 2, 6 February 2018, Pages 337-344.e4

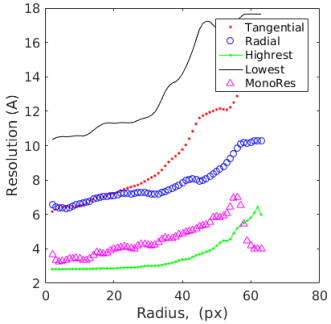
Local Resolution

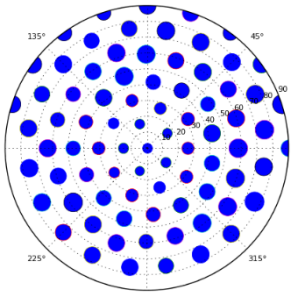


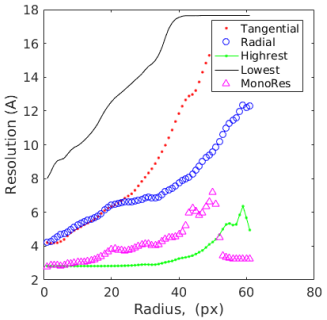


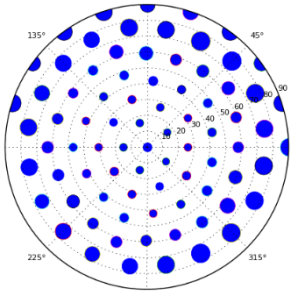
MonoDir











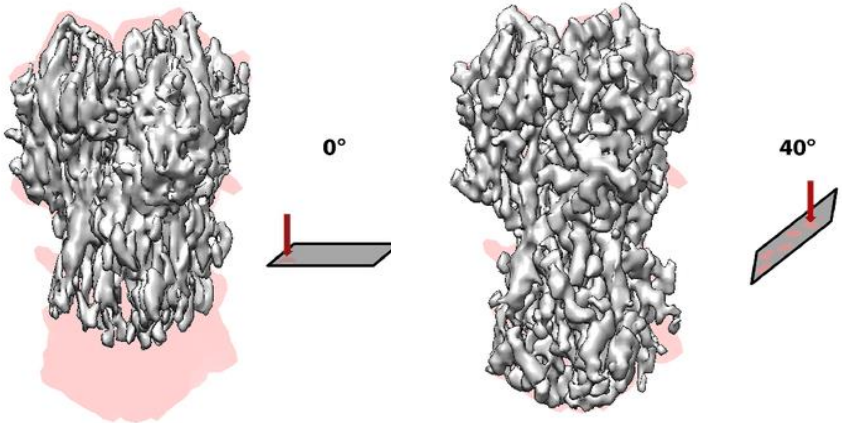
nature methods

Brief Communication | Published: 03 July 2017

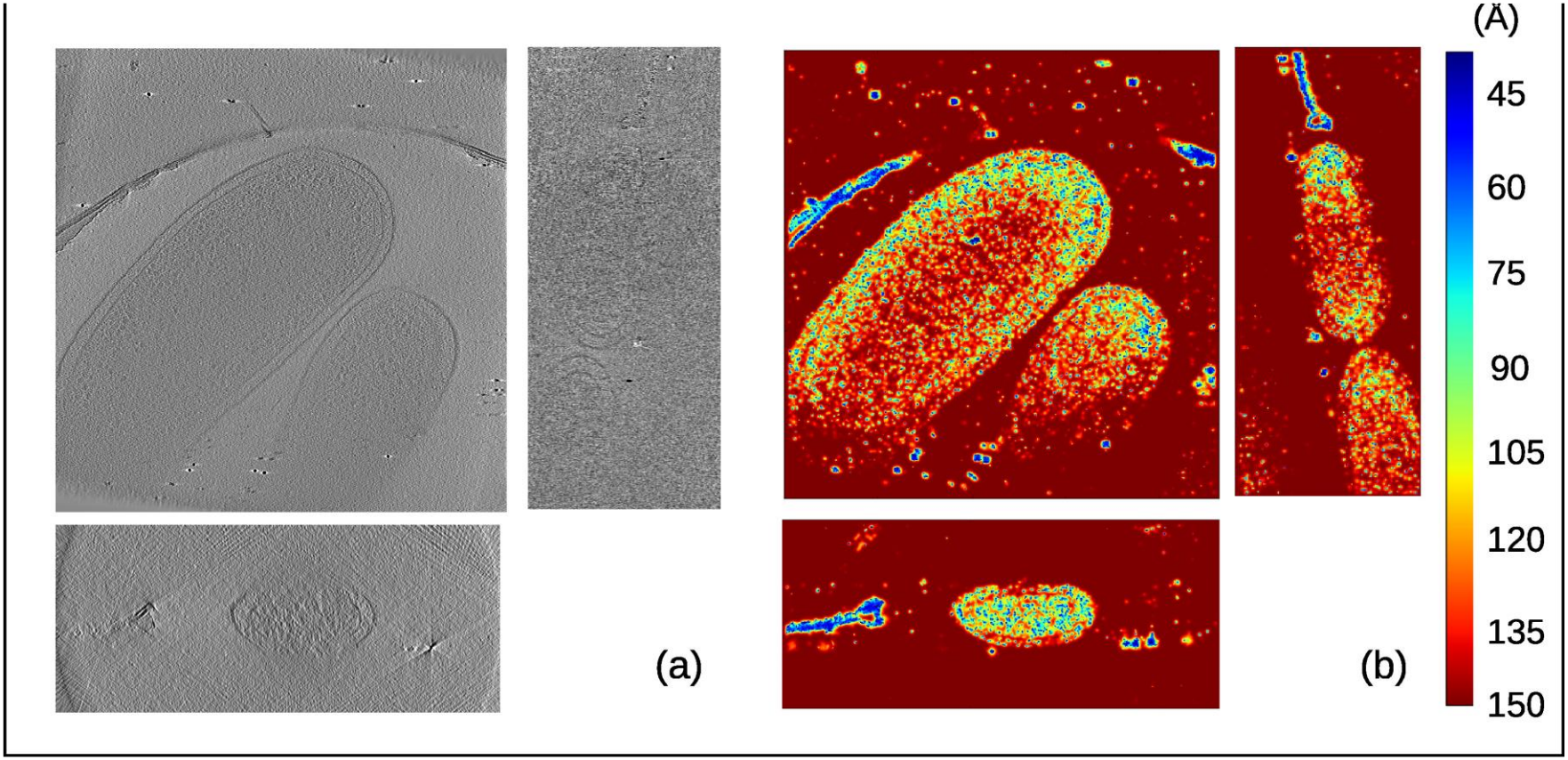
Addressing preferred specimen orientation in single-particle cryo-EM through tilting

Yong Zi Tan, Philip R Baldwin, Joseph H Davis, James R Williamson, Clinton S Potter, Bridget Carragher & Dmitry Lyumkis

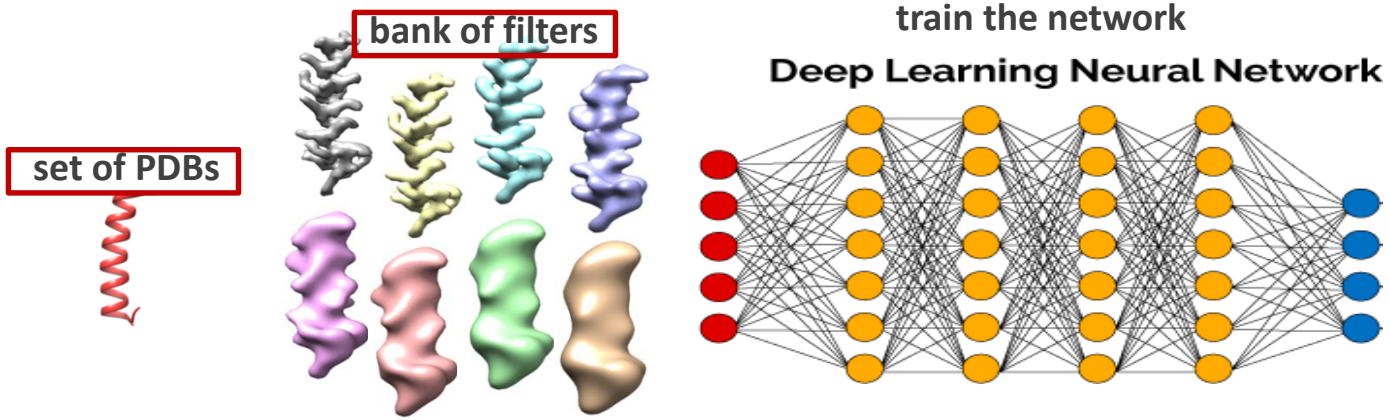
Nature Methods **14**, 793–796 (2017) | Download Citation

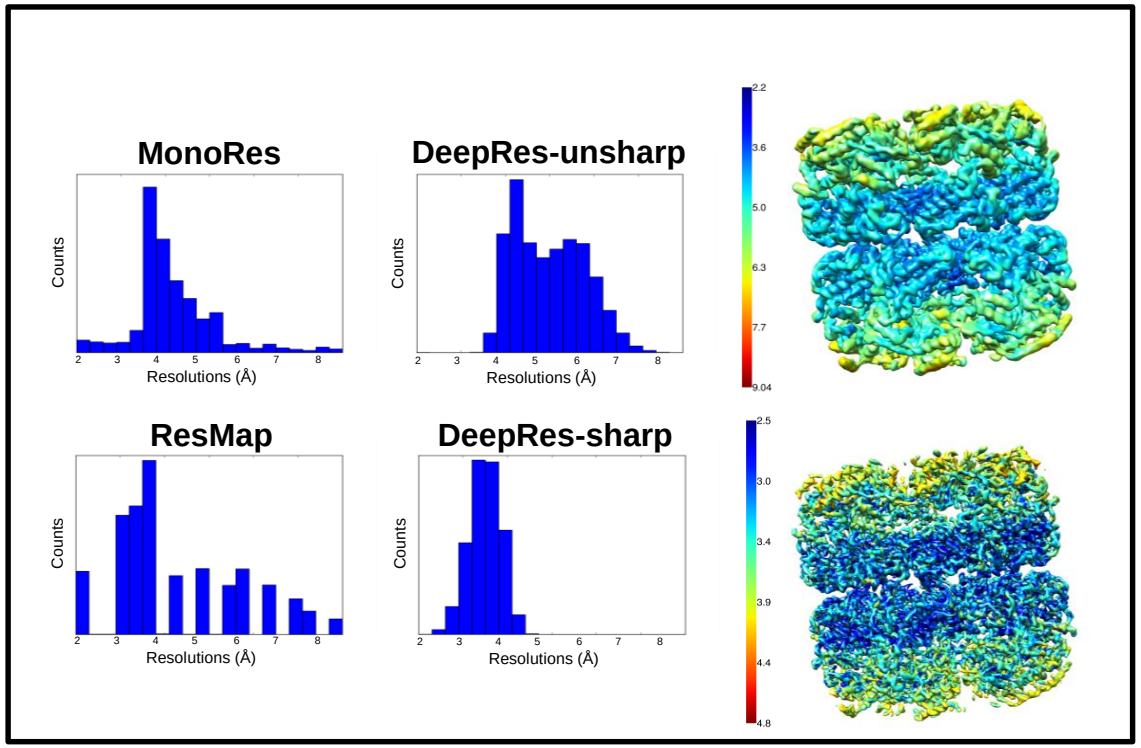


MonoTomo

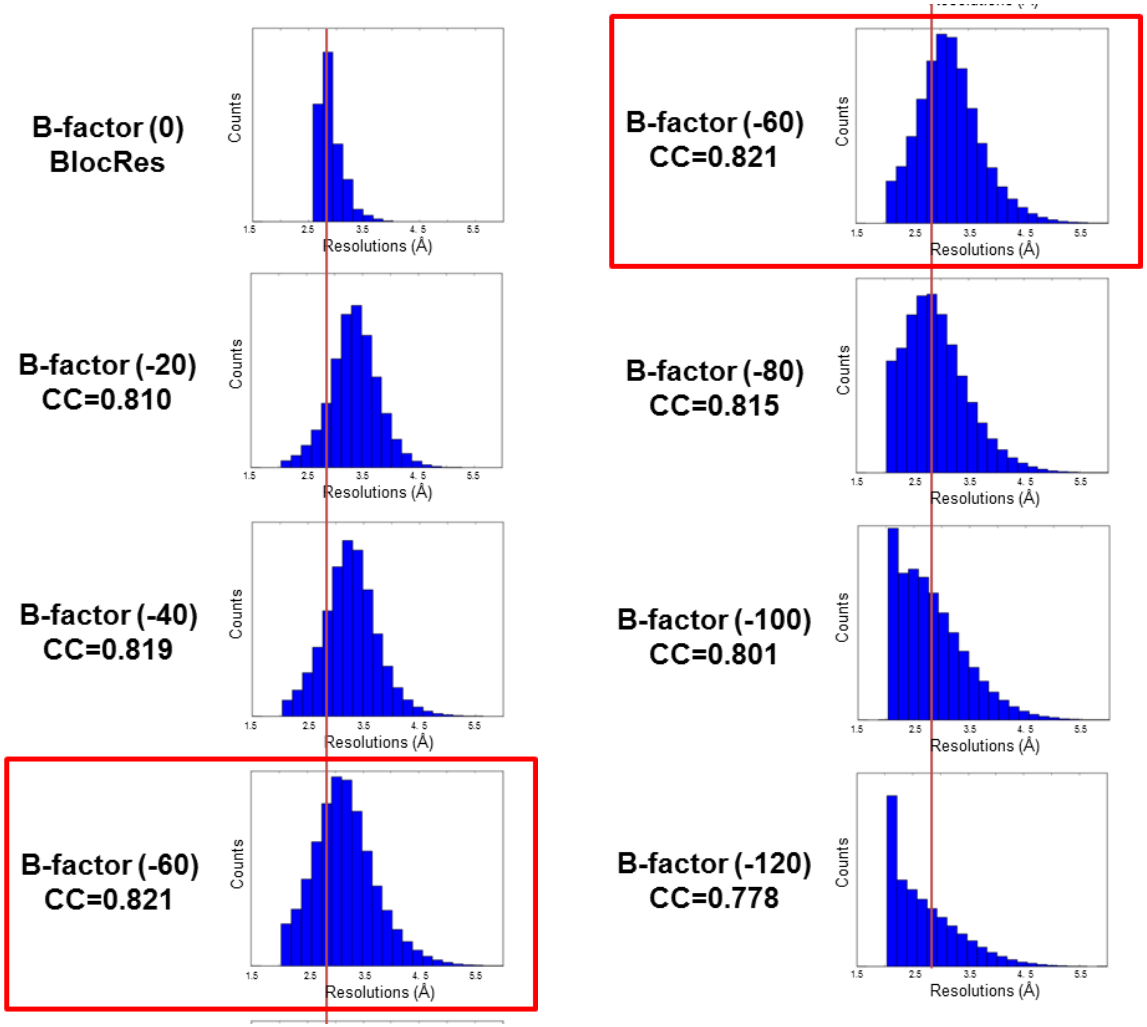


DeepRes

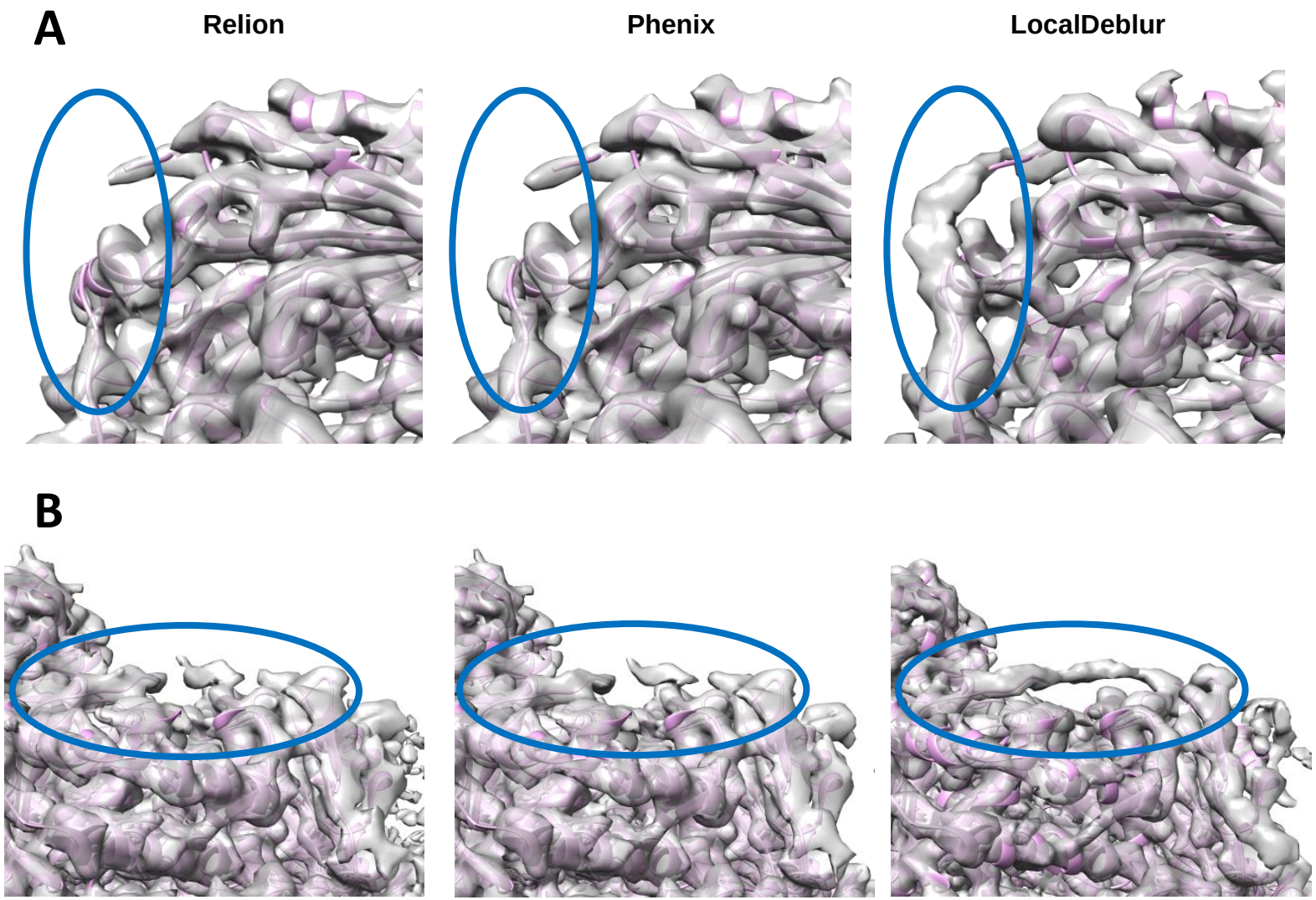

bank of filters
set of PDBs
train the network
Deep Learning Neural Network


MonoRes
DeepRes-unsharp
ResMap
DeepRes-sharp
Counts
Resolutions (Å)

DeepRes


B-factor (0)
BlocRes
B-factor (-20)
CC=0.810
B-factor (-40)
CC=0.819
B-factor (-60)
CC=0.821
B-factor (-60)
CC=0.821
B-factor (-80)
CC=0.815
B-factor (-100)
CC=0.801
B-factor (-120)
CC=0.778
Counts
Resolutions (Å)
1.5
2.5
3.5
4.5
5.5

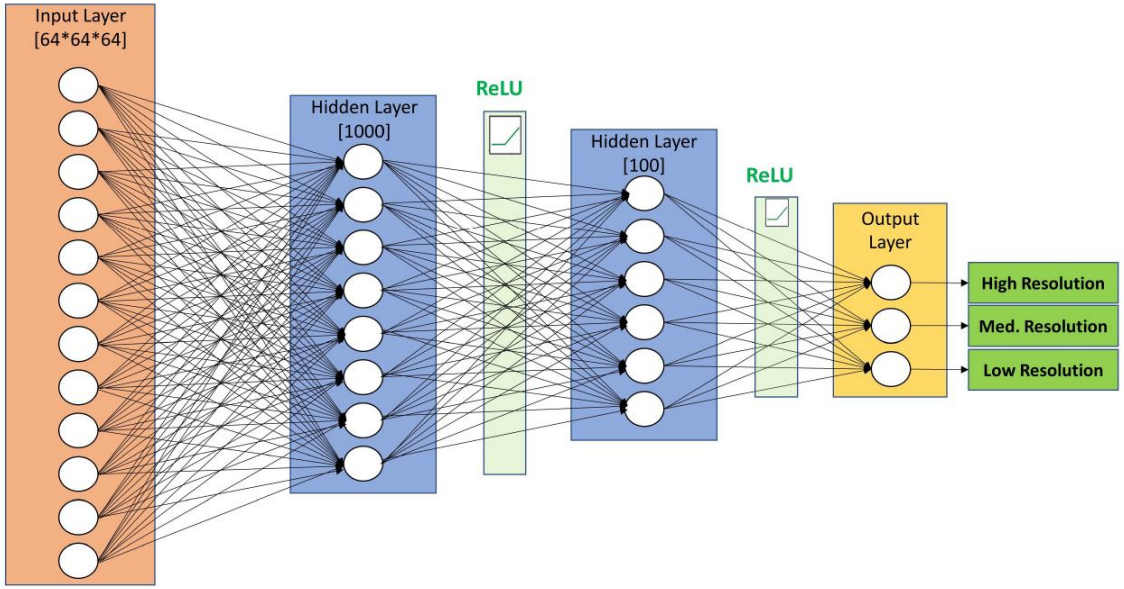
Local Deblur

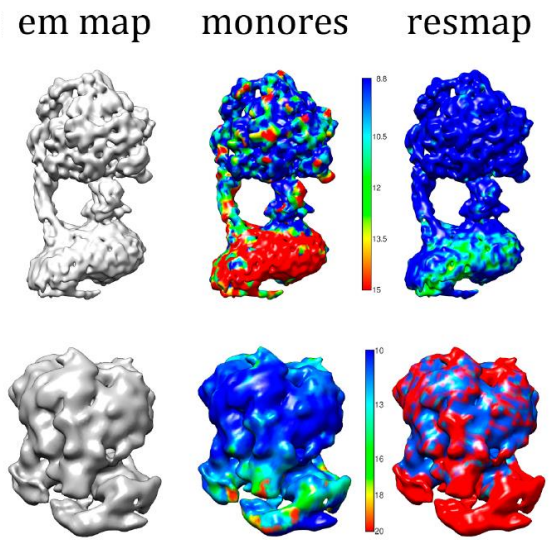


Qualitative Resolution

Table 4. CNN confusion matrix.

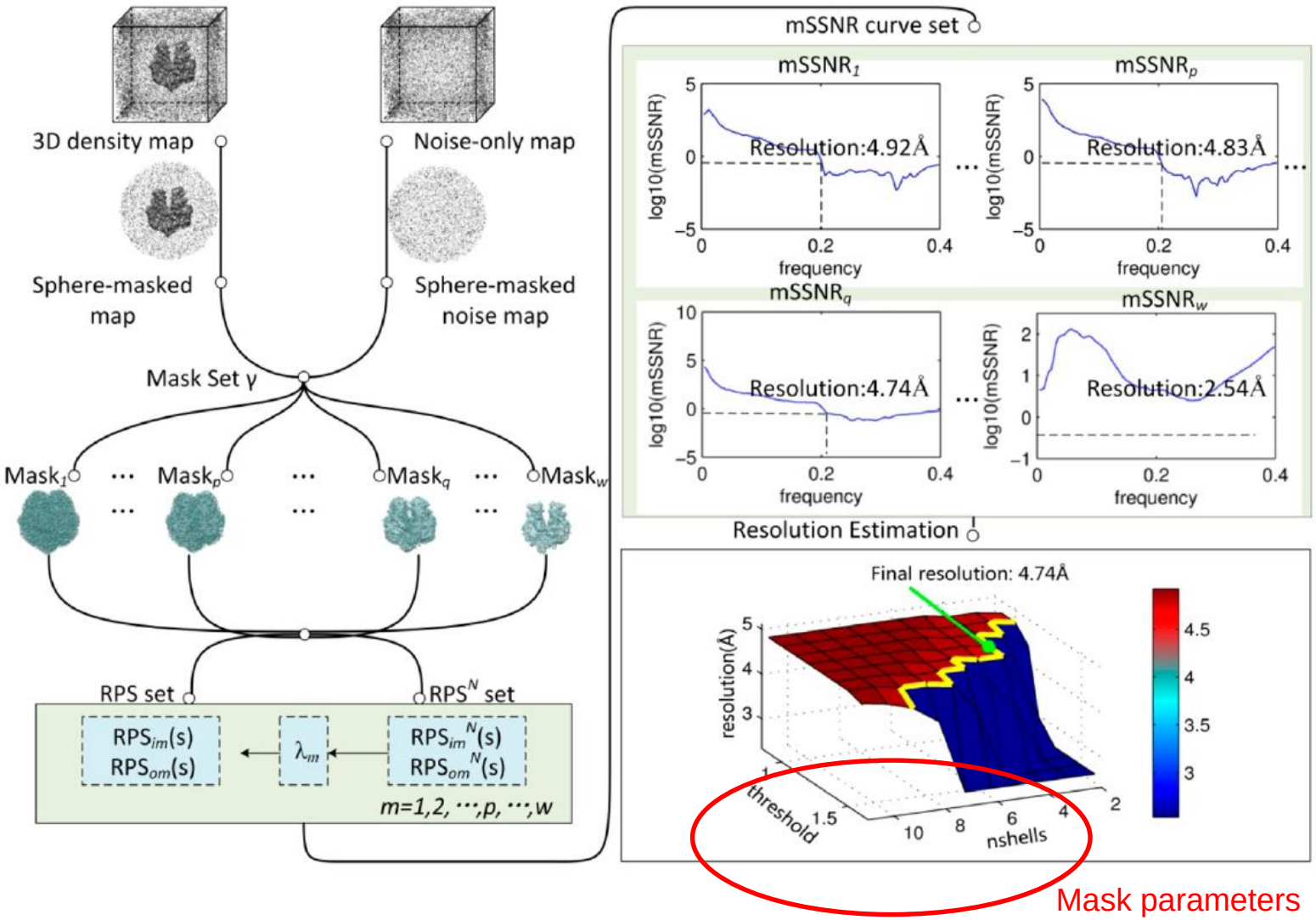
3D CNN Confusion Matrix		Predicted Resolution		
		High	Medium	Low
Published Resolution	High	10	0	0
	Medium	5	4	1
	Low	3	4	3





Avramov, T. K.; Vyenielo, D.; Gomez-Blanco, J.; Adinarayanan, S.; Vargas, J. & Si, D. Deep Learning for Validating and Estimating Resolution of Cryo-Electron Microscopy Density Maps. Molecules, 2019 , 24

Dependence of FSC on mask



Yang, Y.-J.; Wang, S.; Zhang, B.; Shen, H.-B. Resolution Measurement from a Single Reconstructed Cryo-EM Density Map with Multiscale Spectral Analysis. Journal of chemical information and modeling, 2018, 58, 1303-1311

Conclusions

Conclusions

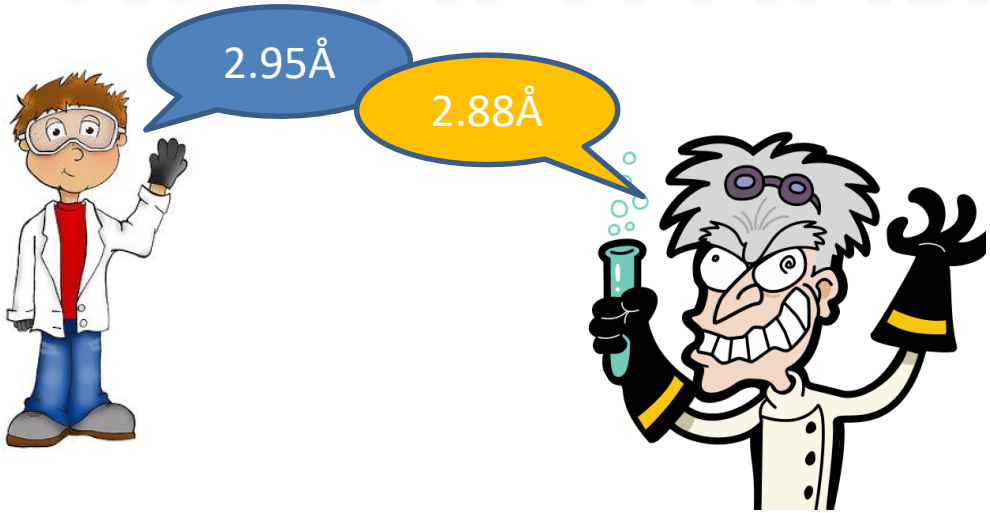
1

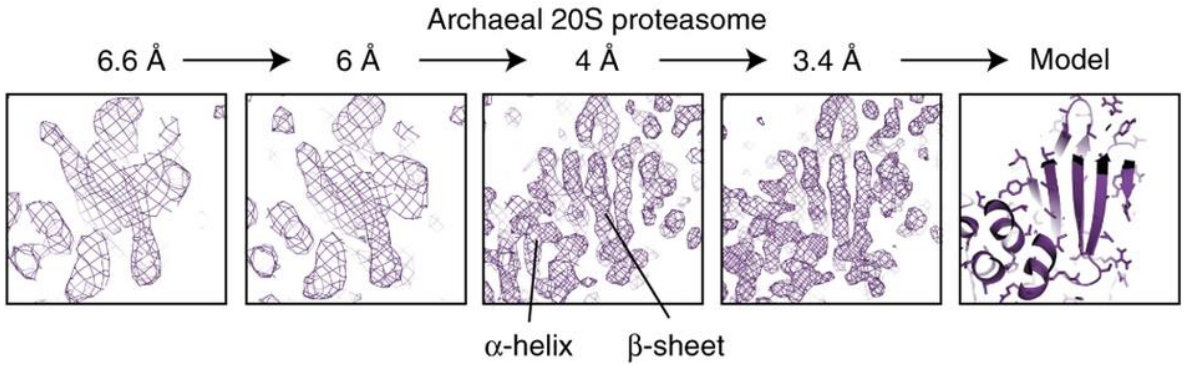
‘There is nothing more practical than a good theory’.

*Phrase attributed to Kurt Lewin, German-American psychologist, known as one of the modern pioneers of social, organizational, and applied psychology.

Conclusions

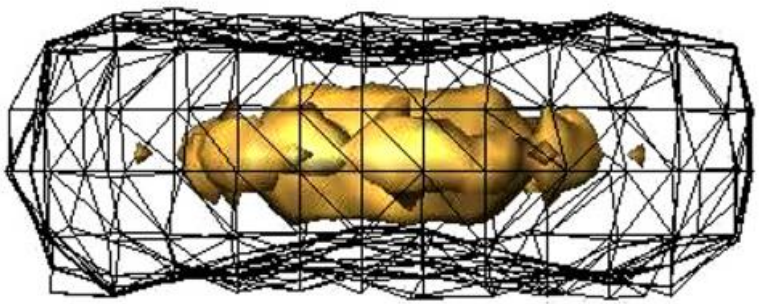
2 **RESOLUTION**

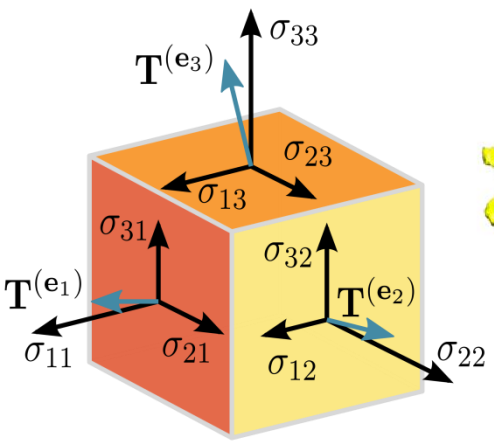


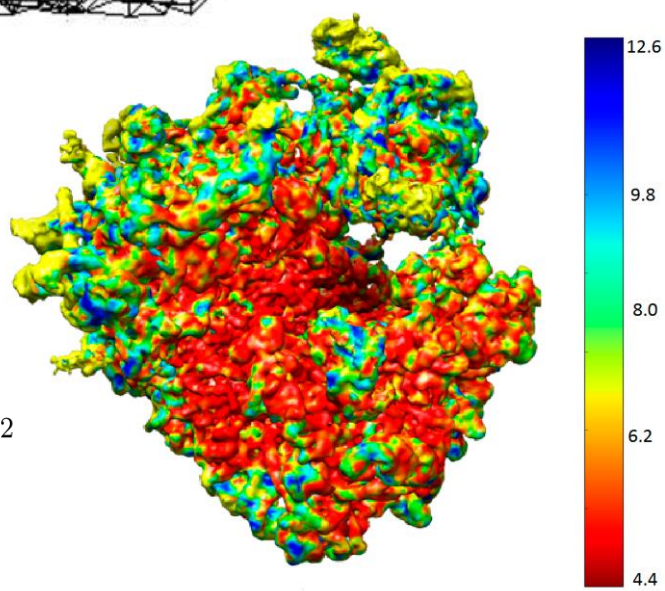


Conclusions

2 **RESOLUTION**







Conclusions

3

Don’t trust everything that is written in papers (even mine).

- The wide-spread use of Gaussian in Science is attributed to the fact that mathematicians believed it was a physical principle and physicists believed it was a mathematical principle.

instruct
image
Processing
Center

"**Theory** is when you know everything but nothing works.

Practice is when everything works but no one knows why.

In our lab, theory and practice are combined: **nothing works and no one knows why**."

Mathematics is the language of Nature