Processing

09:00 Sjors Scheres: Intro and new approaches

10:00 Coffee Break

10:30 Niko Grigorieff: New challenges

11:15 Steve Ludtke: Deep learning methods

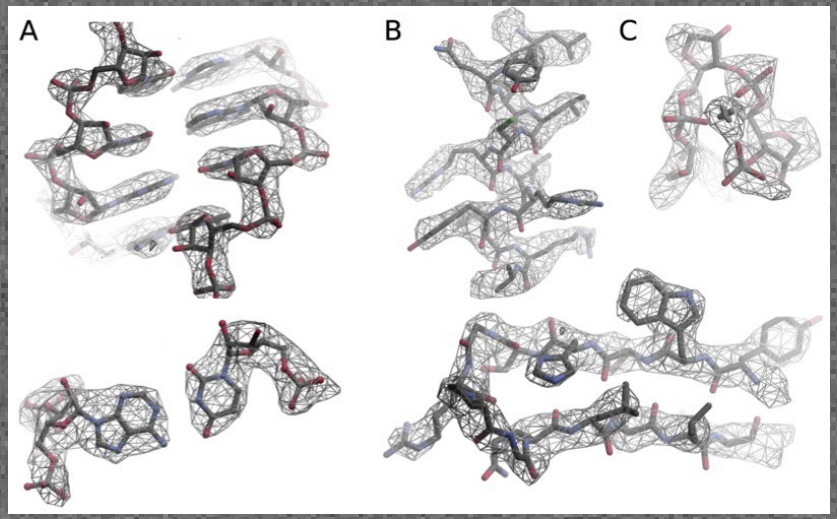
12:00 Lunch

13:00 Marcus Brubaker: Bayesian methods

13:30 Michael Cianfrocco: Cloud computing

14:00 Panel discussion (Chair John Rubinstein)

Processing:
Introduction and new approaches



Sjors H.W. Scheres

NRAMM cryo-EM workshop,
NYSBC, 1 November 2017

MRC | Laboratory of Molecular Biology

Introduction and new approaches

A comprehensive overvi[illegible] in the last few years that [illegible] resolution.

Lots of hard work in early image processing developments
(Joachim, Marin, Michael, Pawel, …)

Topics to be covered include:

- 3D reconstruction
- image restoration techniques
- how to deal with heterogeneous populations.

- What are the hot topics in processing?
- What are the major mathematical approaches and available software?

Introduction and new approaches

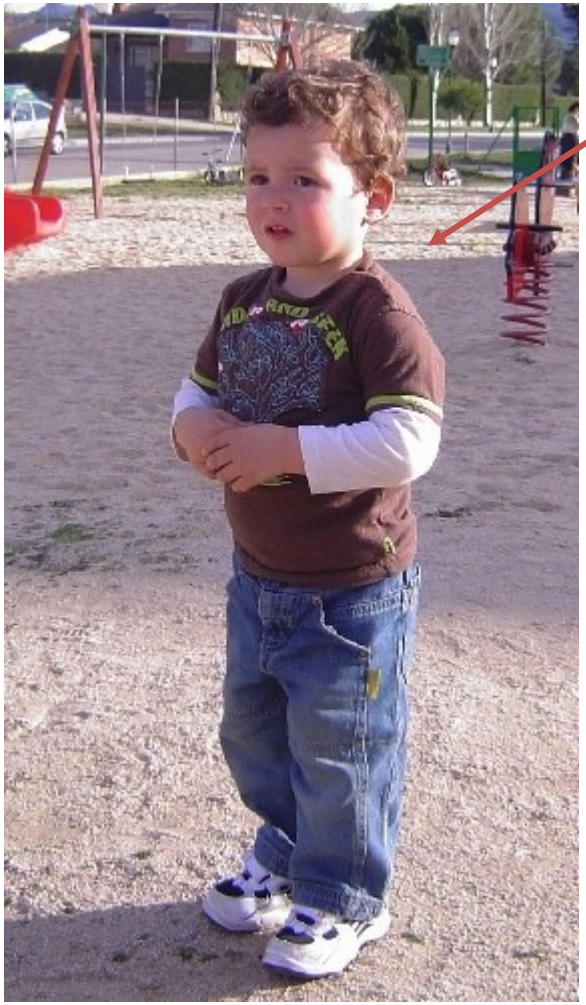
A comprehensive overview of the major advances that have taken place in the last few years that have enabled maps to achieve "atomic" resolution.

Topics to be covered include:

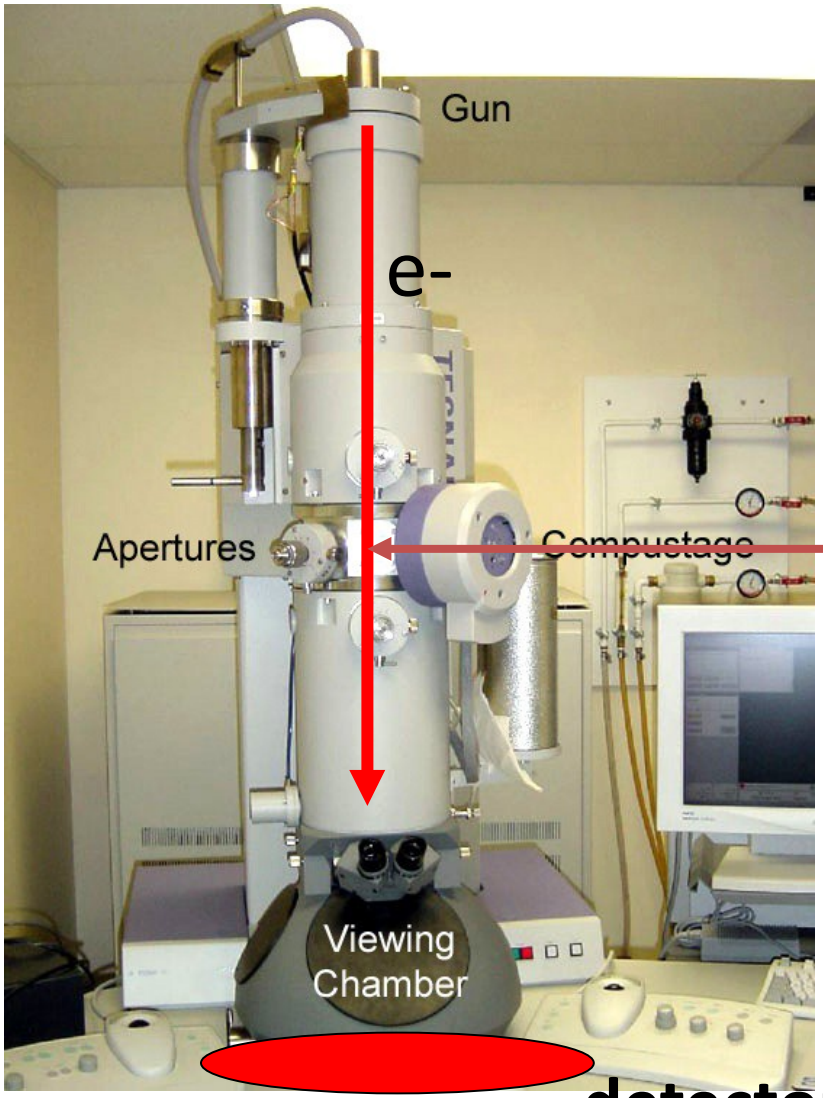
- 3D reconstruction
- image restoration techniques
- how to deal with heterogeneous populations.

- What are the hot topics in processing?
- What are the major mathematical approaches and available software?

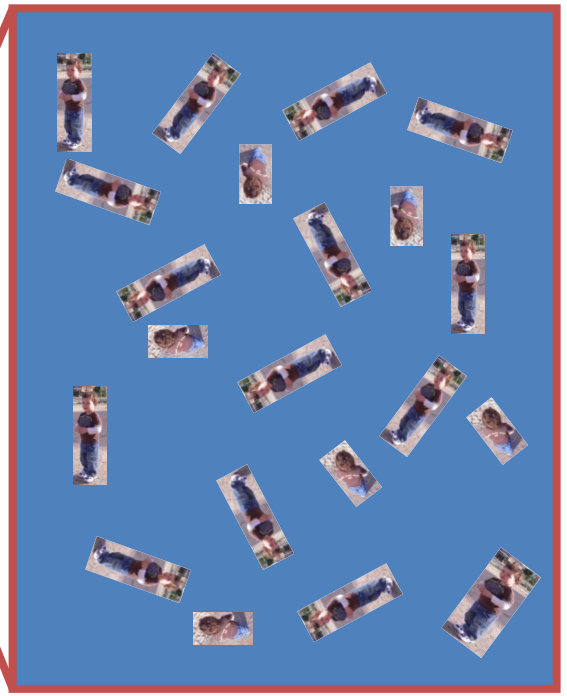
An example “protein”



Experimental setup



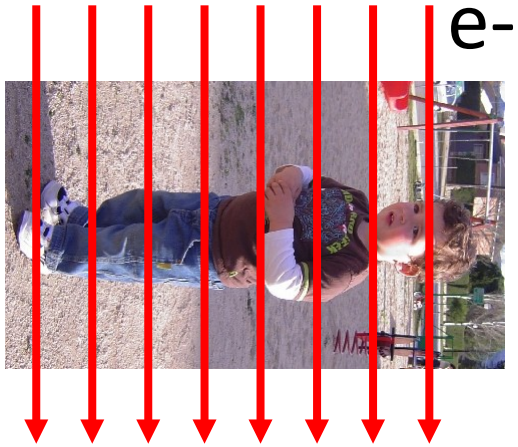
sample



detector

Electron microscopy imaging

3D object



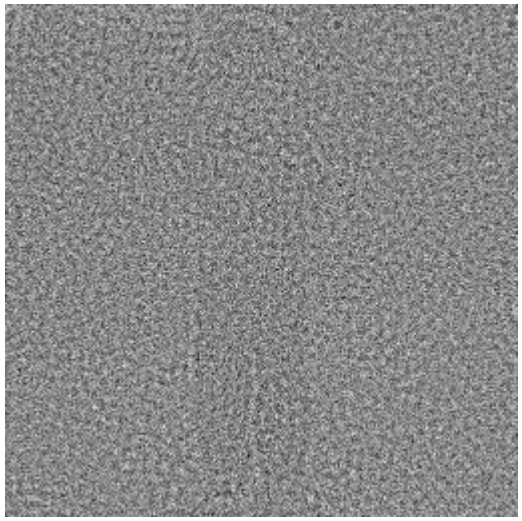
2D projection

We collect data in 2D,

but we want 3D info!

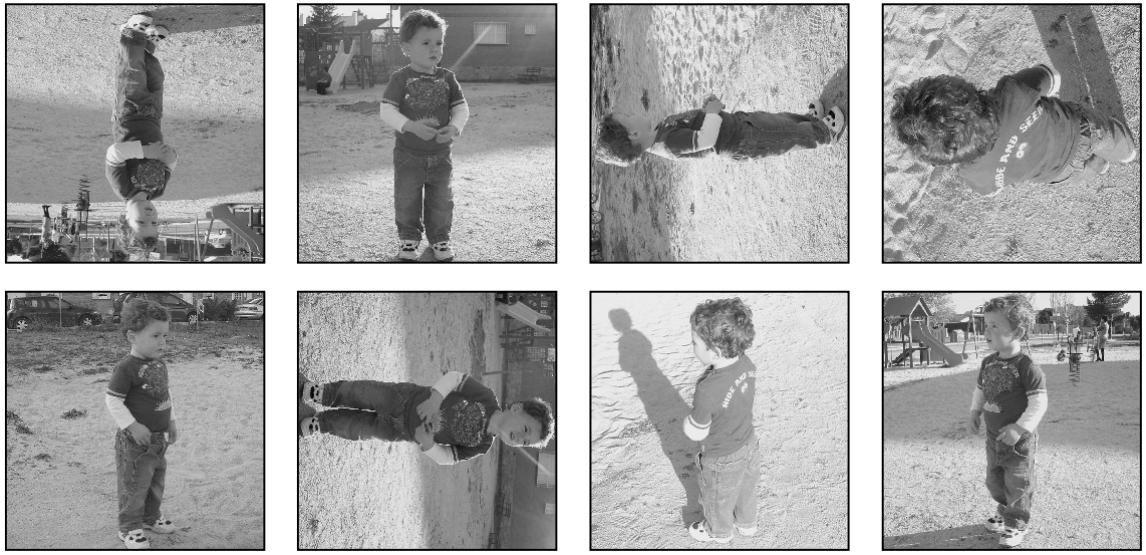
Further inconveniences

- Defocussing & microscope imperfections introduce artefacts
- Low dose: large amounts of noise



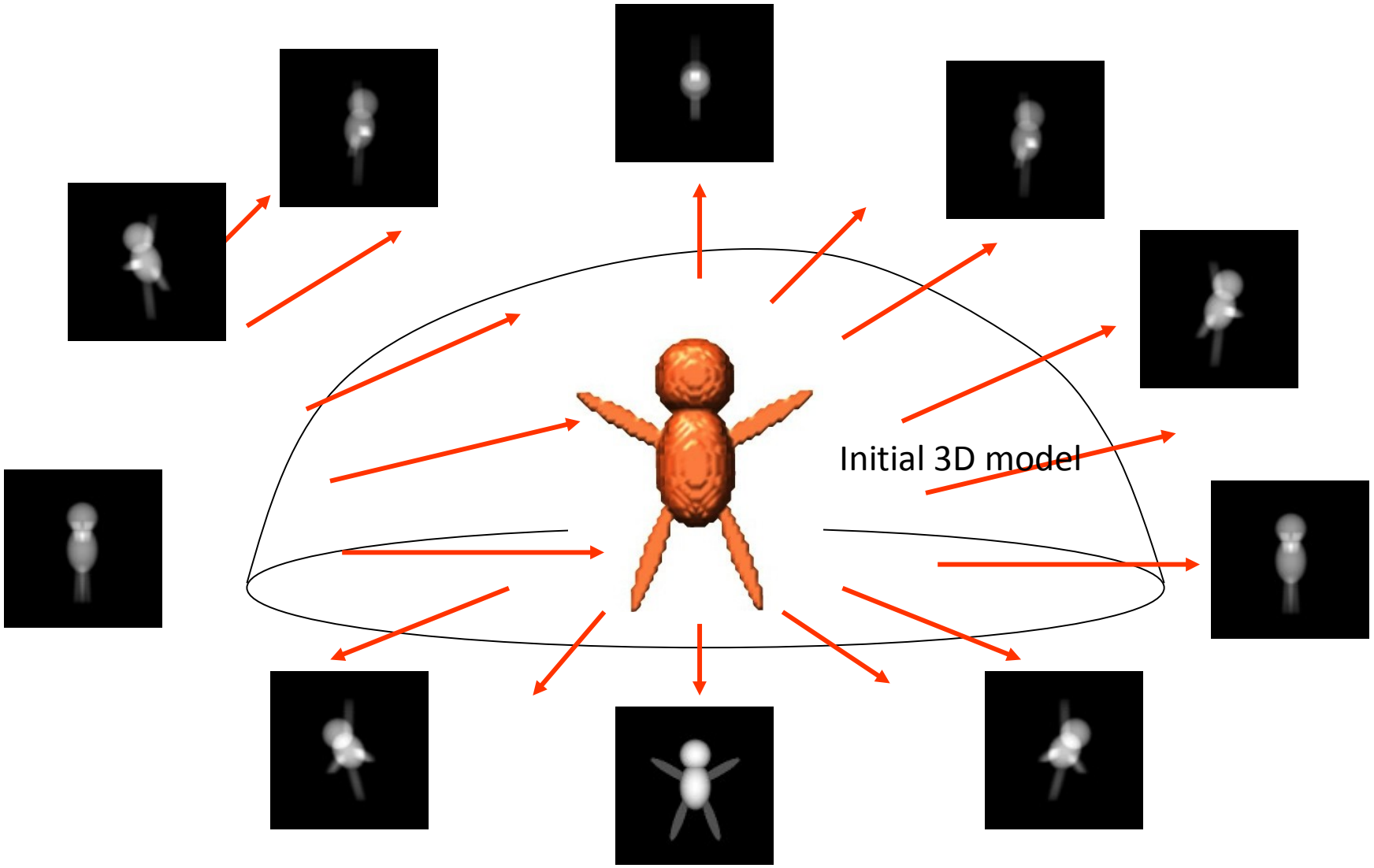
Single particle analysis

- Embedded in ice: many unknown orientations

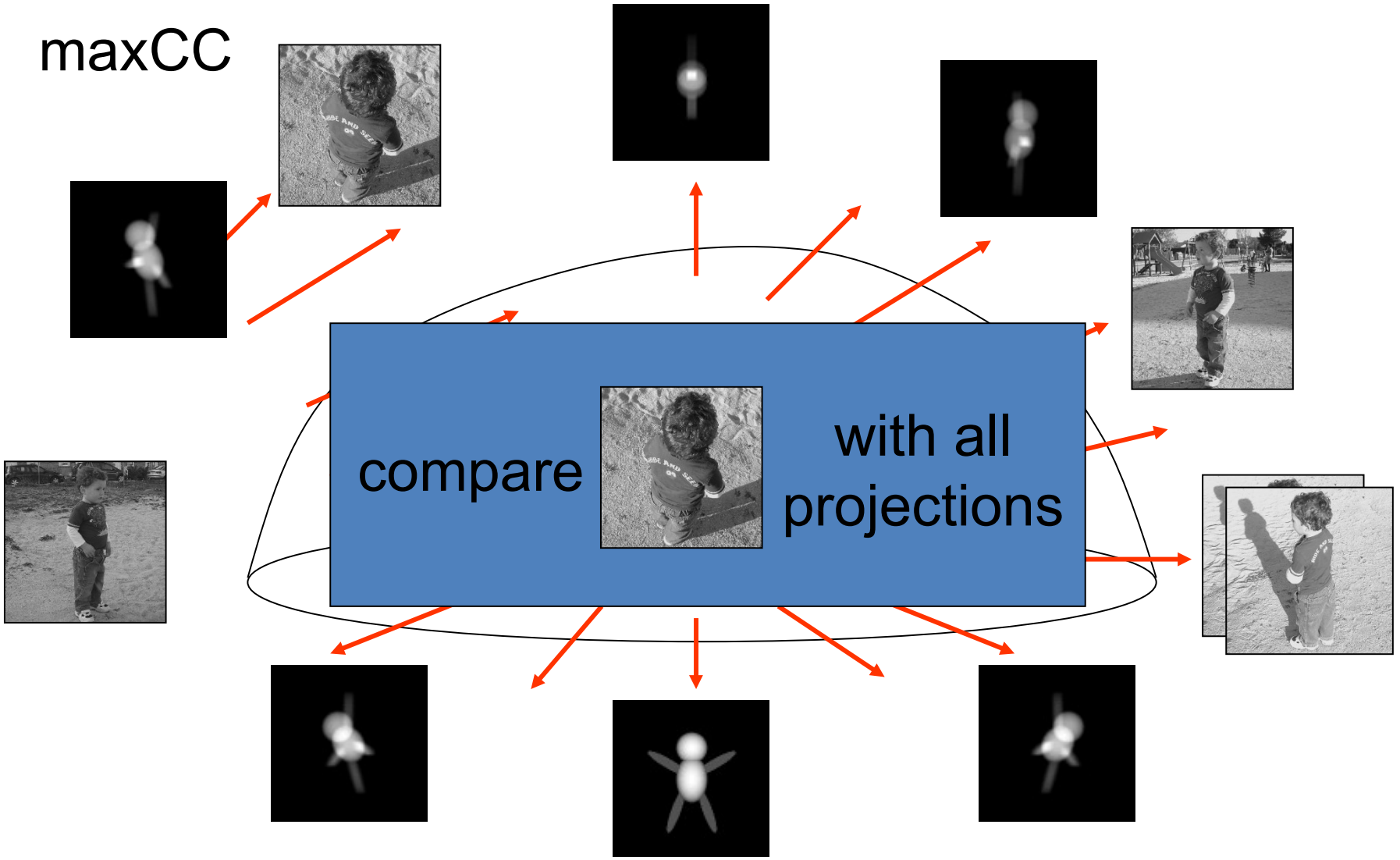


- Combine all 2D projections into a 3D reconstruction

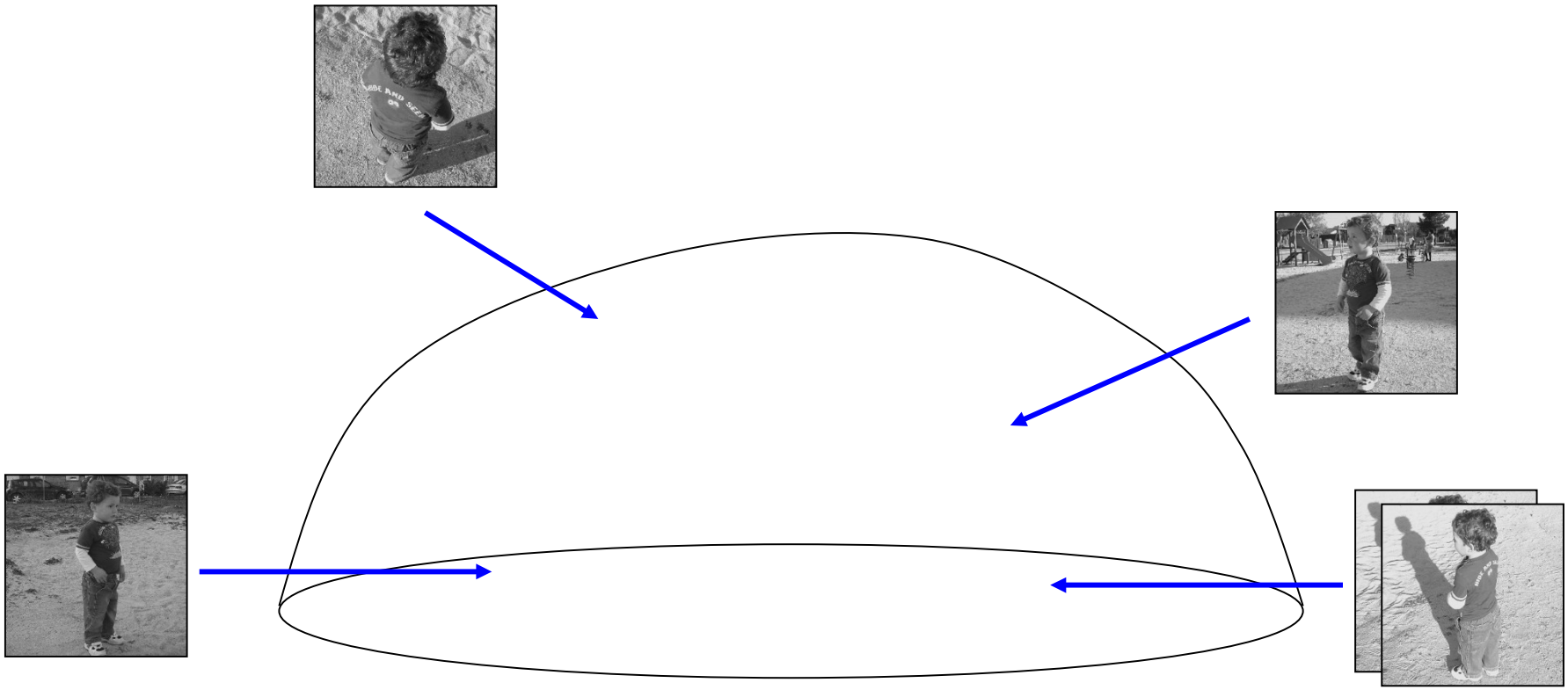
Projection matching



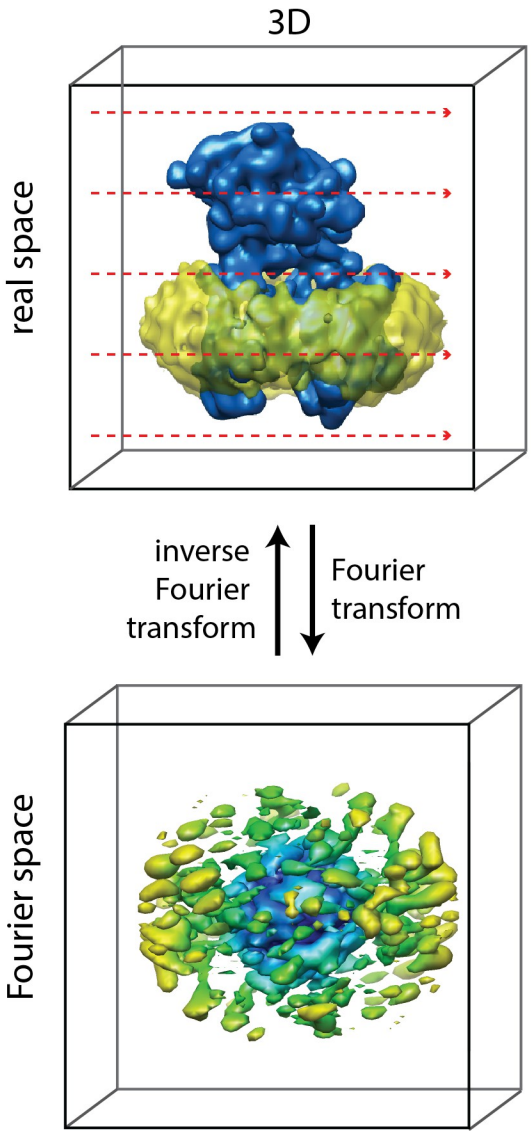
Projection matching



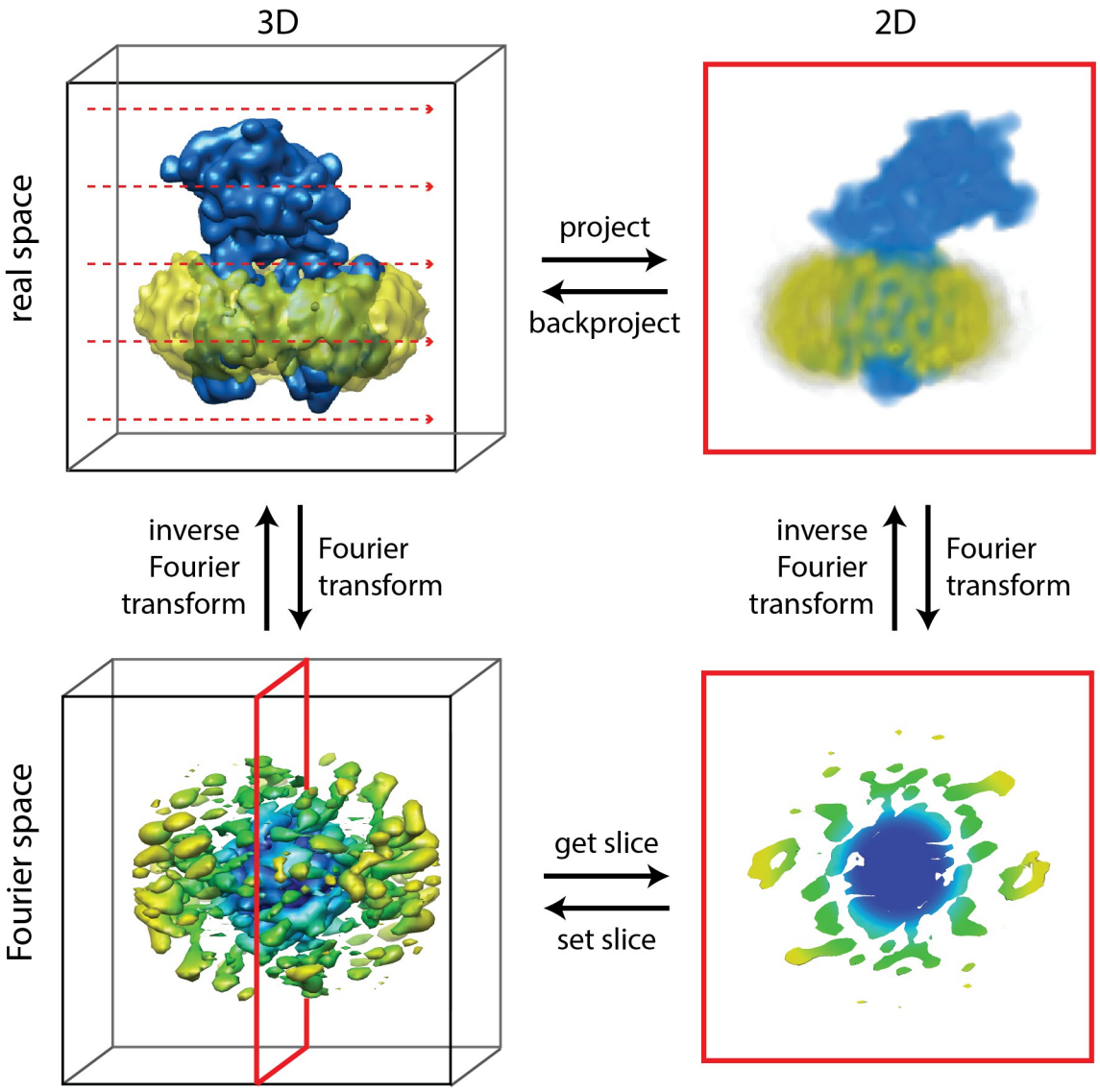
3D reconstruction



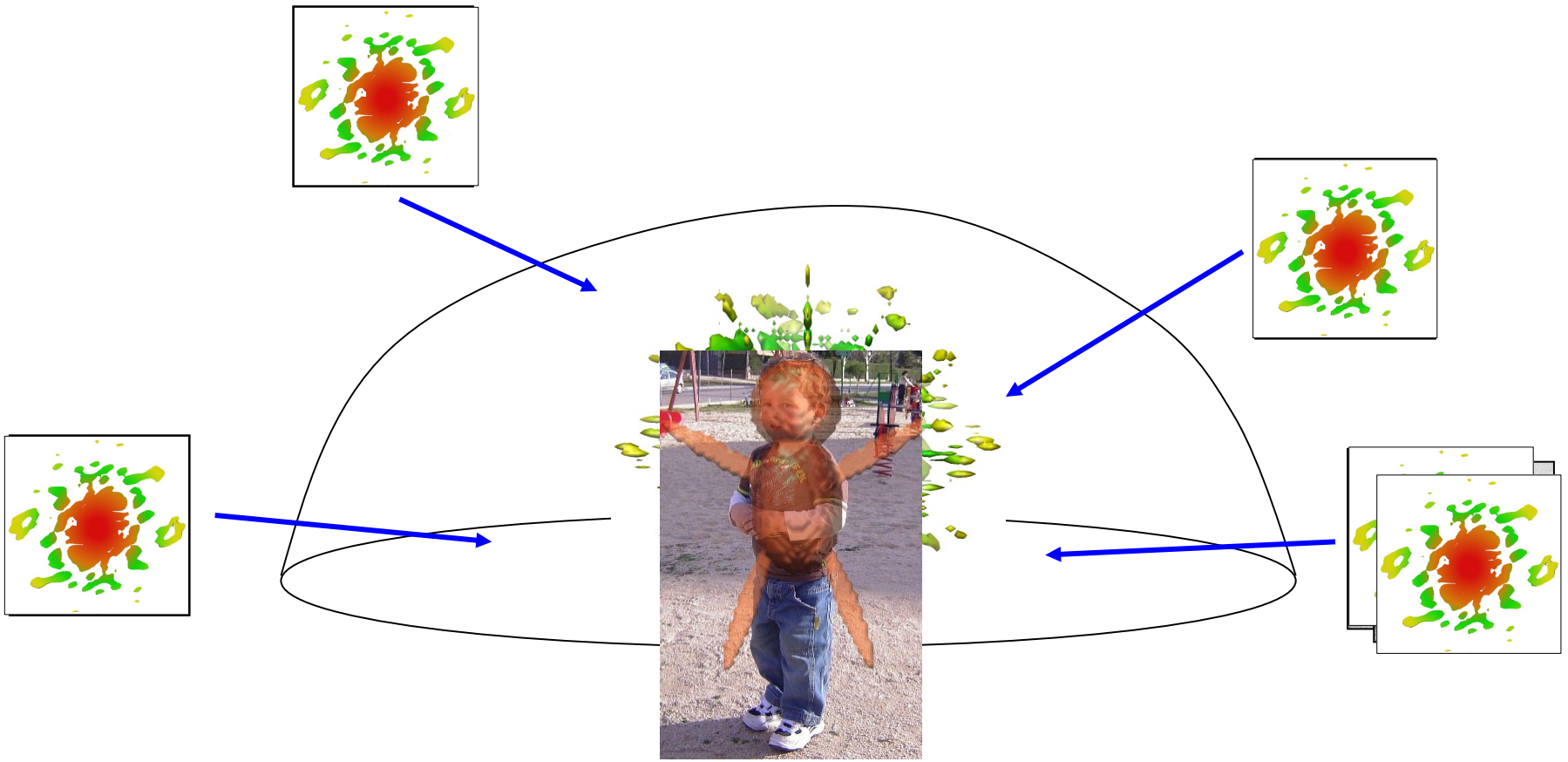
Projection slice theorem



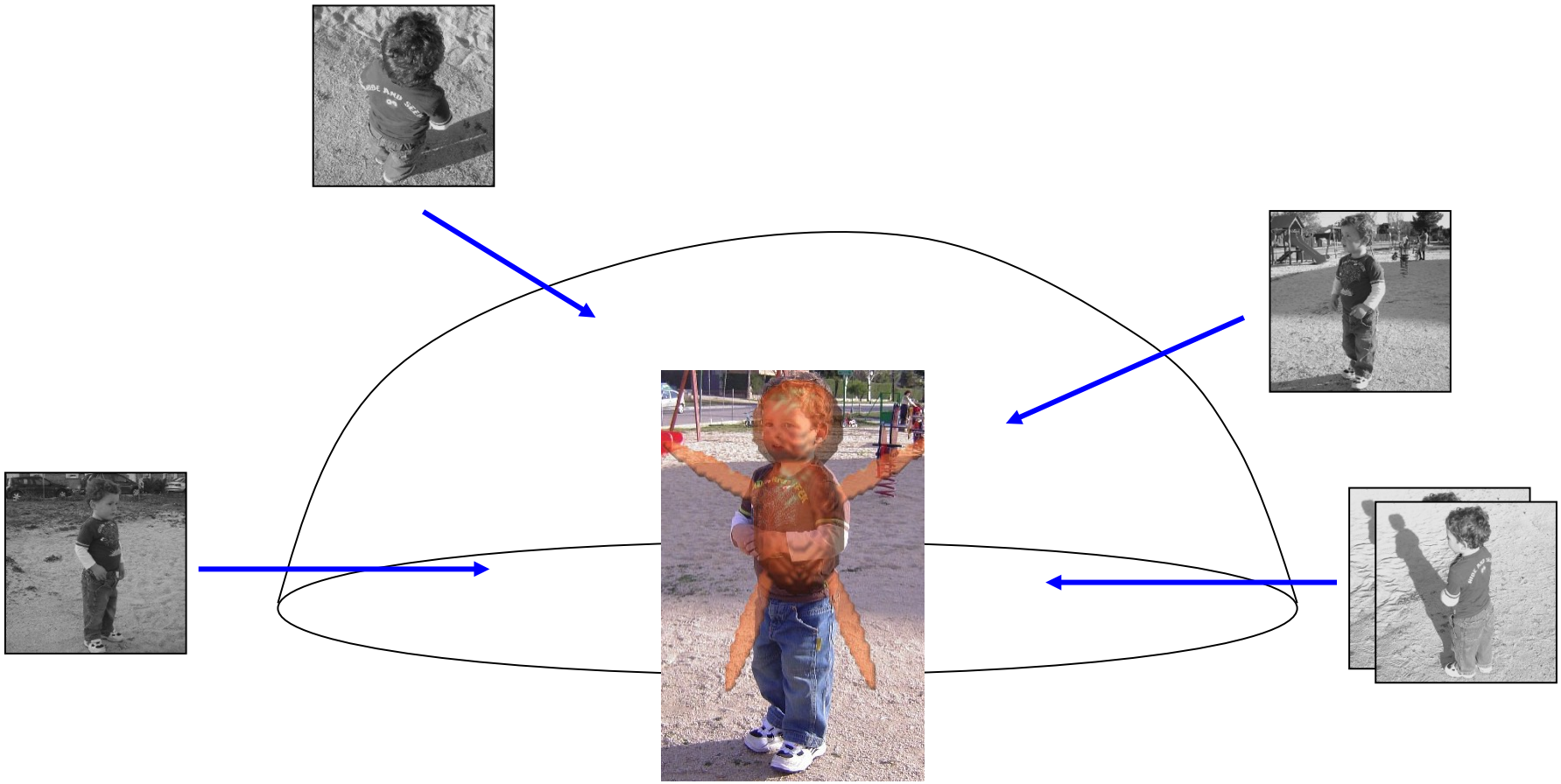
Projection slice theorem



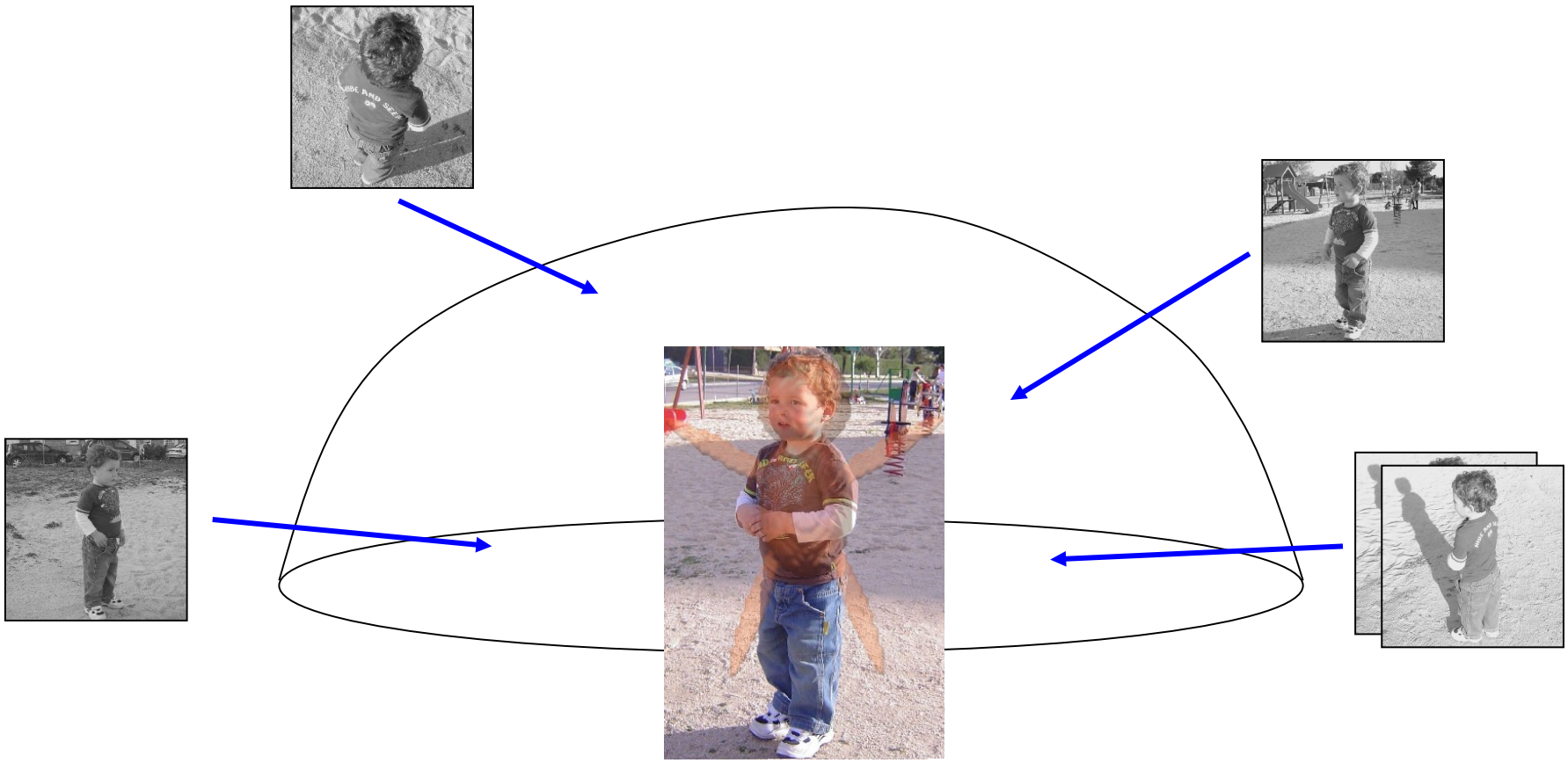
Iterative refinement



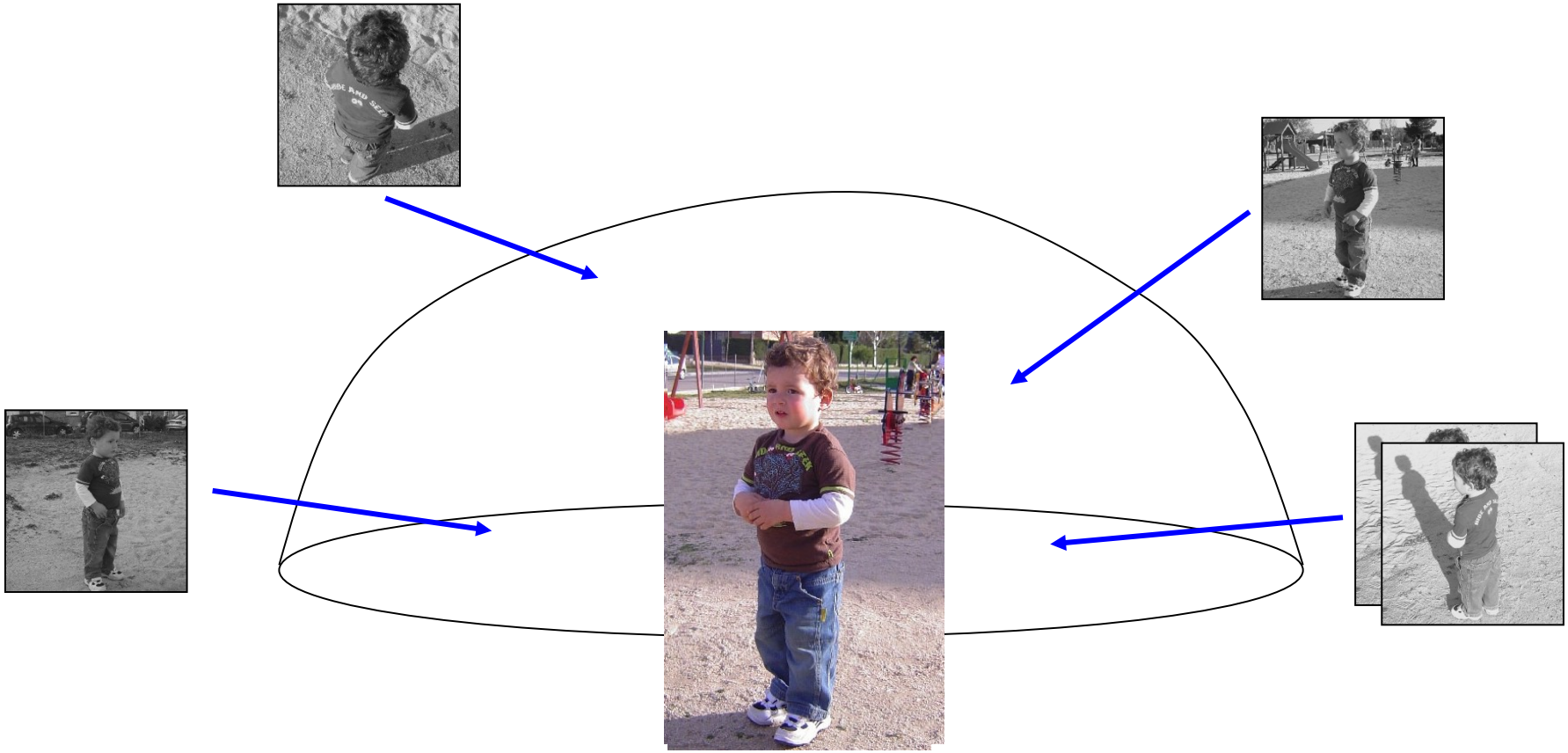
3D reconstruction



Iterative refinement



Iterative refinement



Introduction and new approaches

A comprehensive overview of the major advances that have taken place in the last few years that have enabled maps to achieve "atomic" resolution.

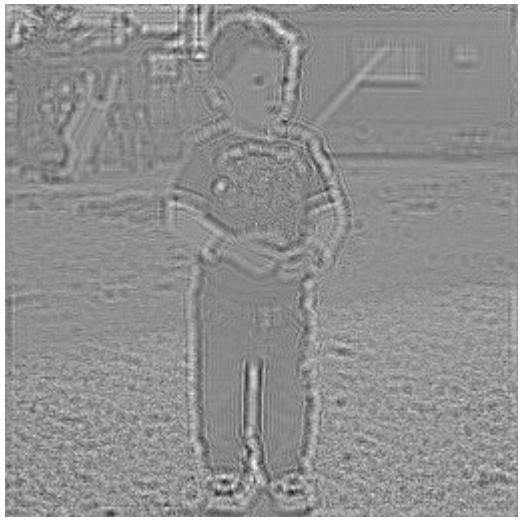
Topics to be covered include:

- 3D reconstruction
- image restoration techniques
- how to deal with heterogeneous populations.

- What are the hot topics in processing?
- What are the major mathematical approaches and available software?

Further inconveniences

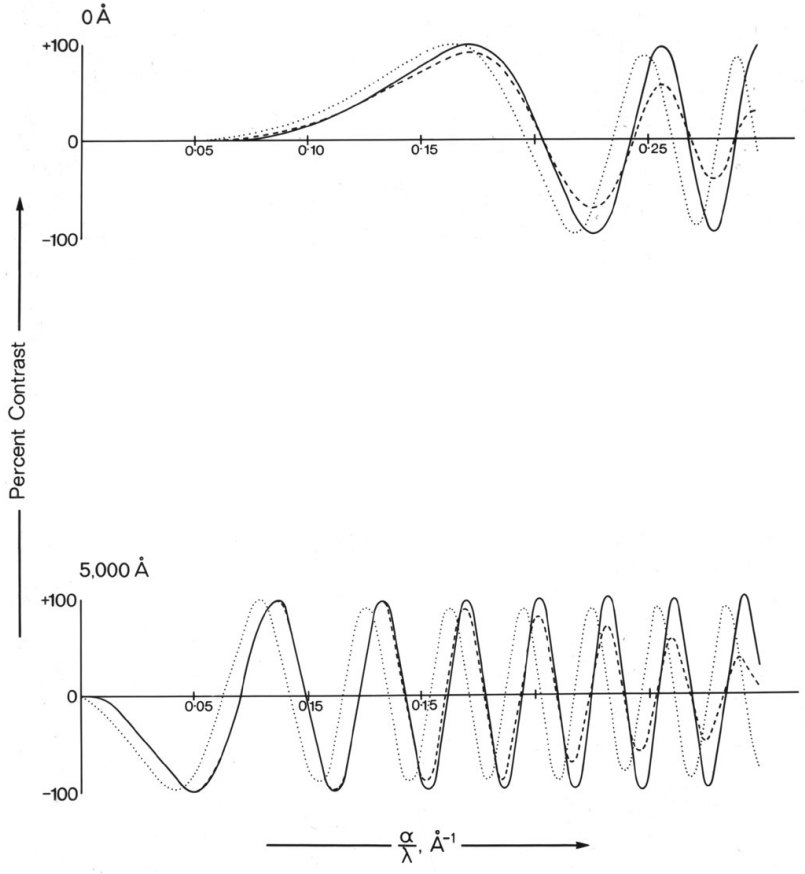
- Defocussing & microscope imperfections introduce artefacts



Phil. Trans. Roy. Soc. Lond. B. **261**, 105–118 (1971) [105]
Printed in Great Britain

Measurement and compensation of defocusing and aberrations by Fourier processing of electron micrographs

BY H. P. ERICKSON AND A. KLUG, F.R.S.
Medical Research Council Laboratory of Molecular Biology, Cambridge



Data model

- **Real-space**

$$X_i = \mathrm{CTF}_i \otimes \mathbf{P}_\varphi V_k + N_i$$

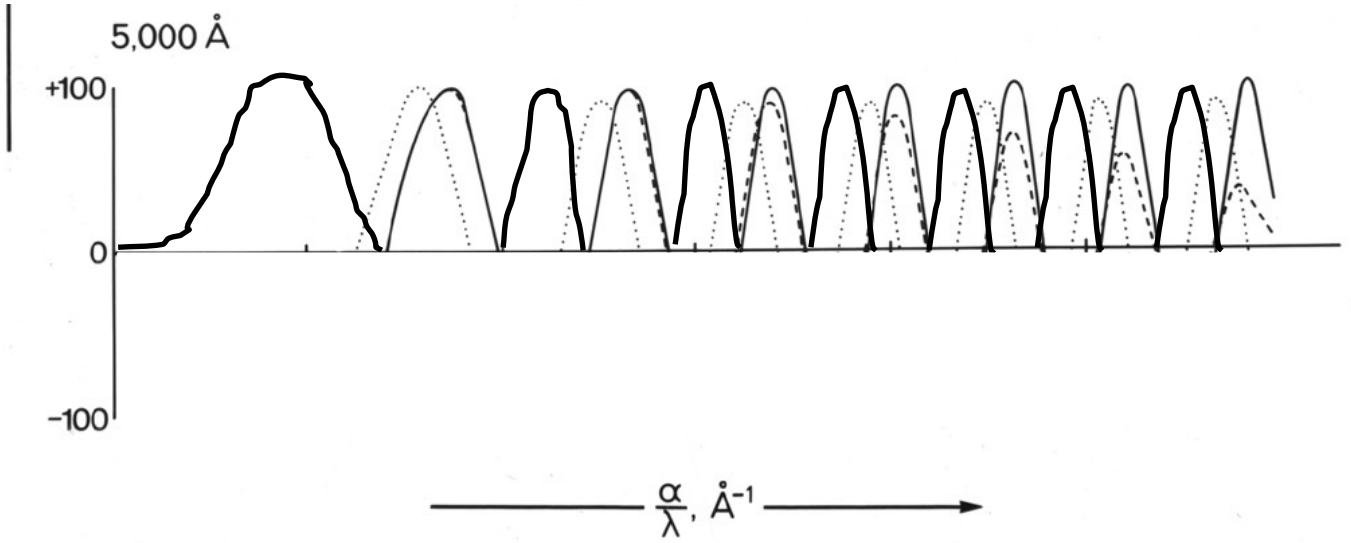
- Convolute w/ CTF
- $\mathbf{P}\phi$ implements integrals

- **Fourier space**

$$X_i = \mathrm{CTF}_i \mathbf{P}_\varphi V_k + N_i$$

- Multiply w/ CTF
- $\mathbf{P}\phi$ takes a slice

Phase flipping



- Easy to do
- Reasonably effective
- Problems in classification?

(3D) Wiener filter

$$V = \frac{\sum_{i=1}^{N} \mathbf{P}_{\varphi}^{\mathrm{T}} \frac{\mathrm{CTF}_i}{\sigma_i^2} X_i}{\sum_{i=1}^{N} \mathbf{P}_{\varphi}^{\mathrm{T}} \frac{\mathrm{CTF}_i^2}{\sigma_i^2} + \frac{1}{\tau^2}}$$

Optimal linear filter

- $\sigma 2$: noise power
- $\tau 2$: signal power

- Low-pass filters & corrects for CTF
- $\tau 2/\sigma 2$ is often approximated as a constant
  => low-pass filter effect is lost
- You cannot pre-Wiener filter your data!

Introduction and new approaches

A comprehensive overview of the major advances that have taken place in the last few years that have enabled maps to achieve "atomic" resolution.

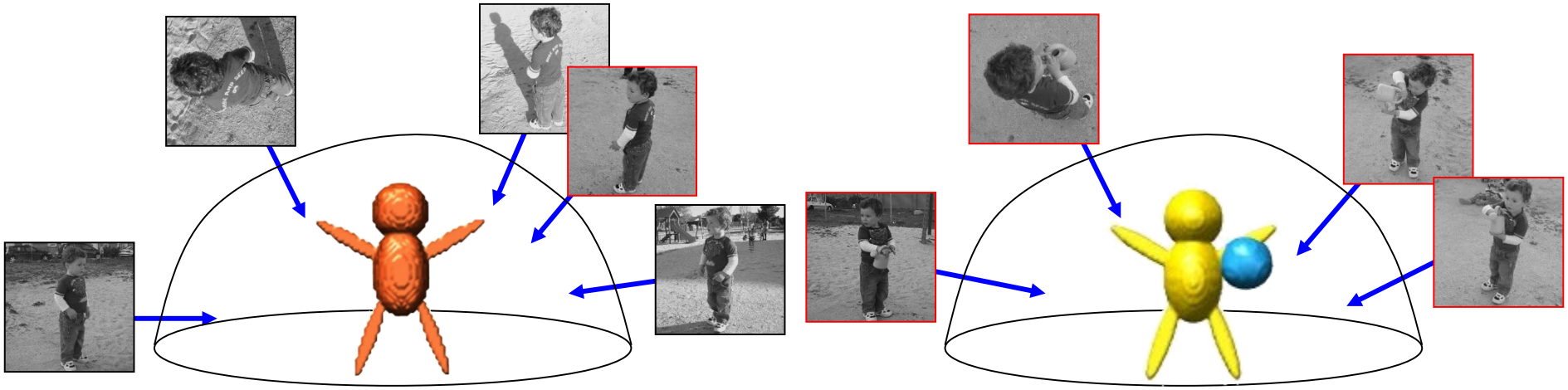
Topics to be covered include:

- 3D reconstruction
- image restoration techniques
- how to deal with heterogeneous populations.

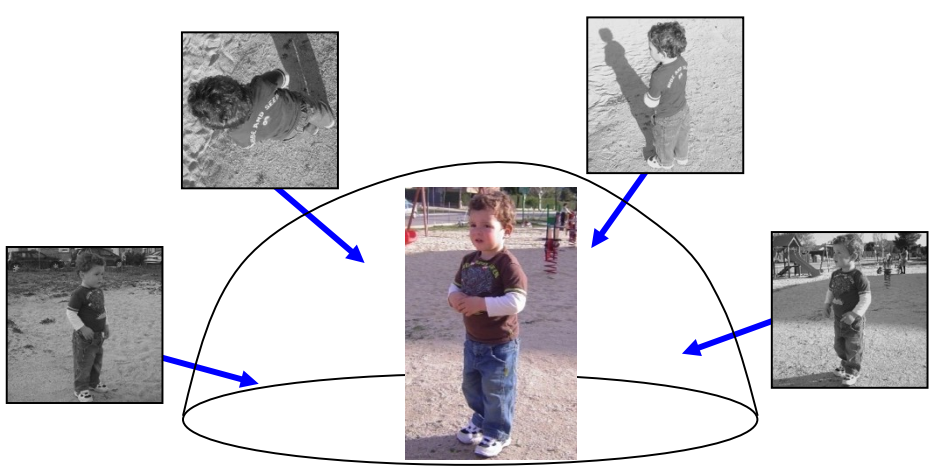
- What are the hot topics in processing?
- What are the major mathematical approaches and available software?

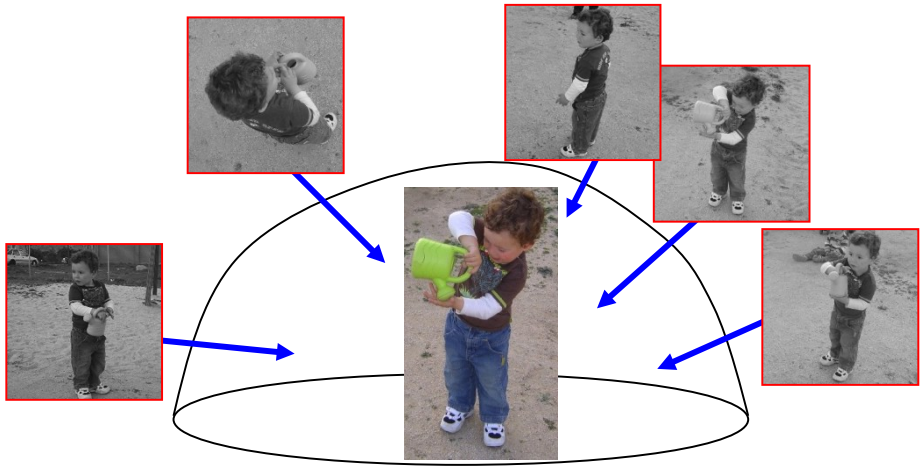
Structural heterogeneity

Multi-reference refinement



Multi-reference refinement





Introduction and new approaches

A comprehensive overview of the major advances that have taken place in the last few years that have enabled maps to achieve "atomic" resolution.

Topics to be covered include:

- 3D reconstruction
- image restoration techniques
- how to deal with heterogeneous populations.

- What are the hot topics in processing?
- What are the major mathematical approaches and available software?

Hot topics?

- Beam-induced motion correction
- Robust initial model generation
- 3D classification
- Computational costs

Introduction and new approaches

A comprehensive overview of the major advances that have taken place in the last few years that have enabled maps to achieve "atomic" resolution.

Topics to be covered include:

- 3D reconstruction
- image restoration techniques
- how to deal with heterogeneous populations.

- What are the hot topics in processing?
- What are the major mathematical approaches and available software?

Hot topics?

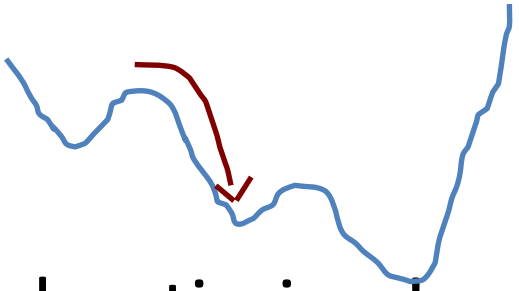
Discussed by John yesterday

- Beam-induced motion correction
- Robust initial model generation
- 3D classification
- Computational costs

Hot topics?

- Beam-induced motion correction
- Robust initial model generation
- 3D classification
- Computational costs

Initial model



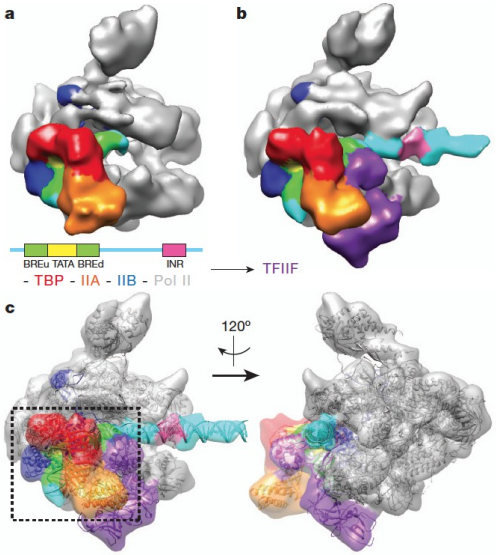
- Expectation-Maximisation is a local optimizer!
  - Gets stuck in nearest (local) minimum
- Bad model in -> bad model out!!!
  - Much less of a problem with high-resolution data
- Stochastic methods may reach global minimum
  - Stochastic Hill Climbing (Hans Elmlund: SIMPLE, SPARX)

Failures…

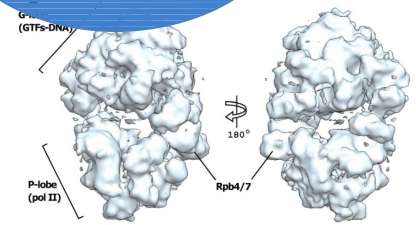
- Get stuck with a wrong initial model

Human RNA polymerase II PIC

He et al & Nogales, Nature (2013)



As resolutions have improved, this has become ever less of a problem.

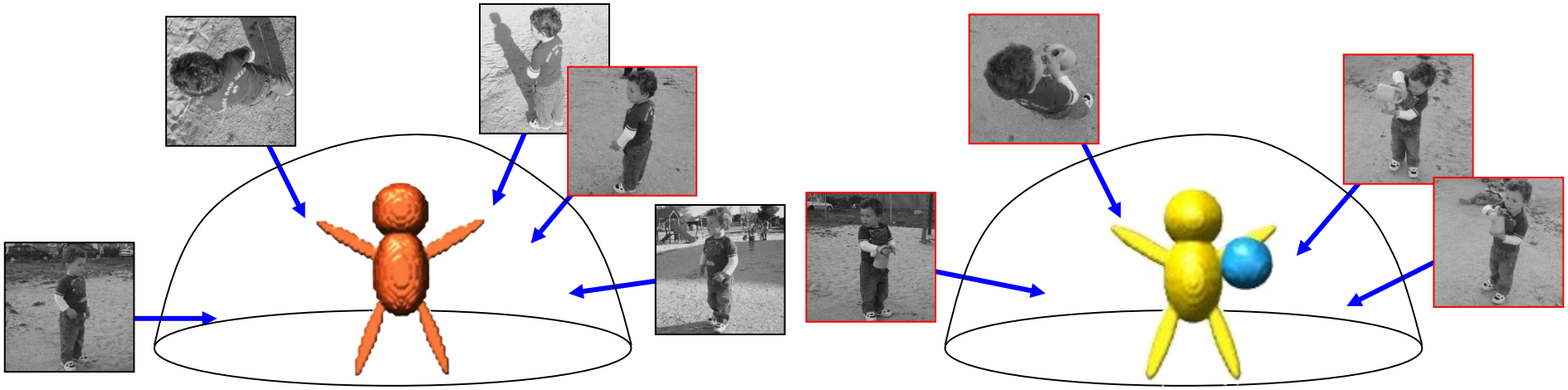


Validation session tomorrow!

Hot topics?

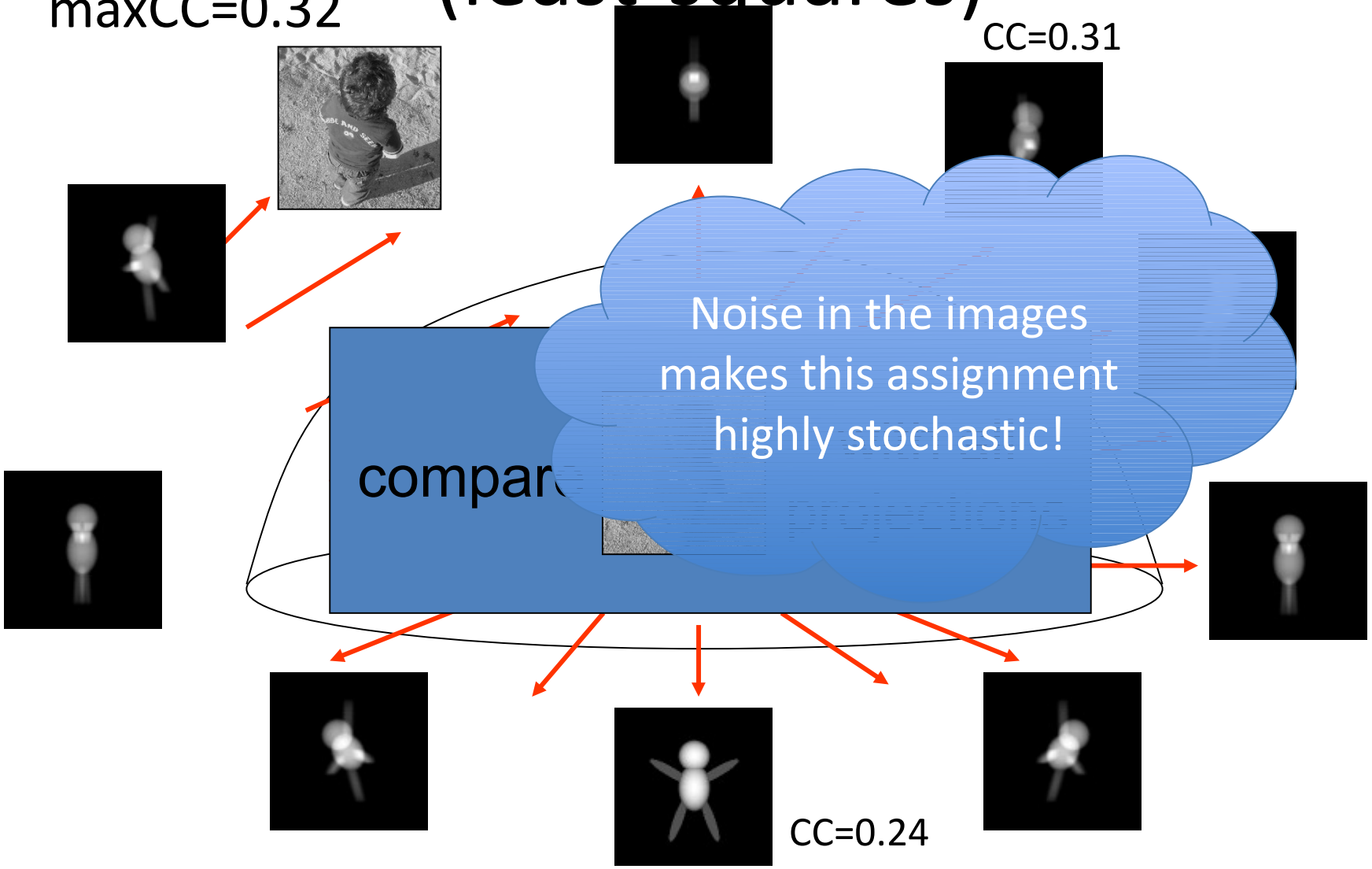
- Beam-induced motion correction
- Robust initial model generation
- 3D classification
- Computational costs

Supervised classification

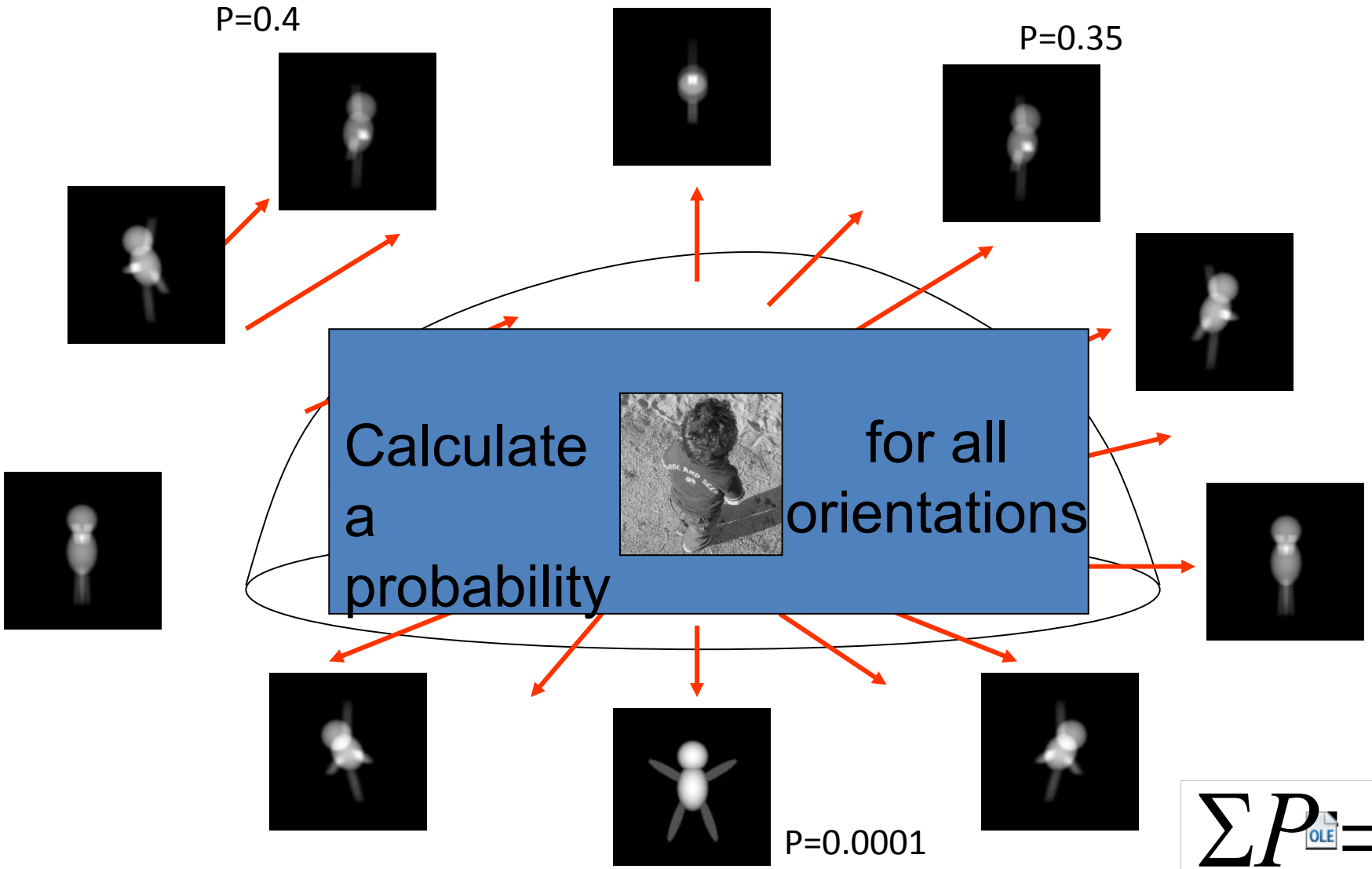


You kind-of need to know the answer already….

Maximum cross-correlation (least-squares)

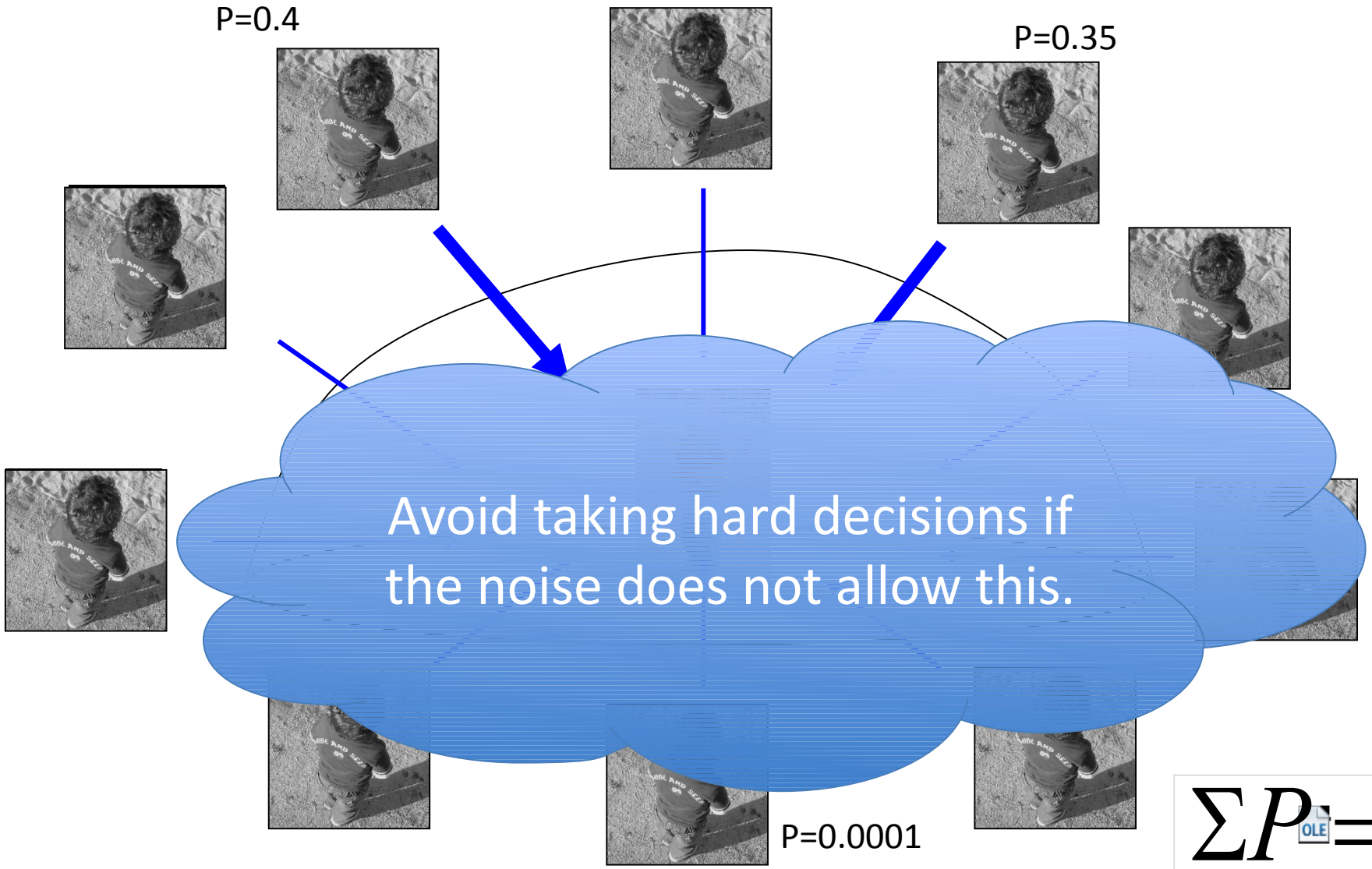


Maximum likelihood



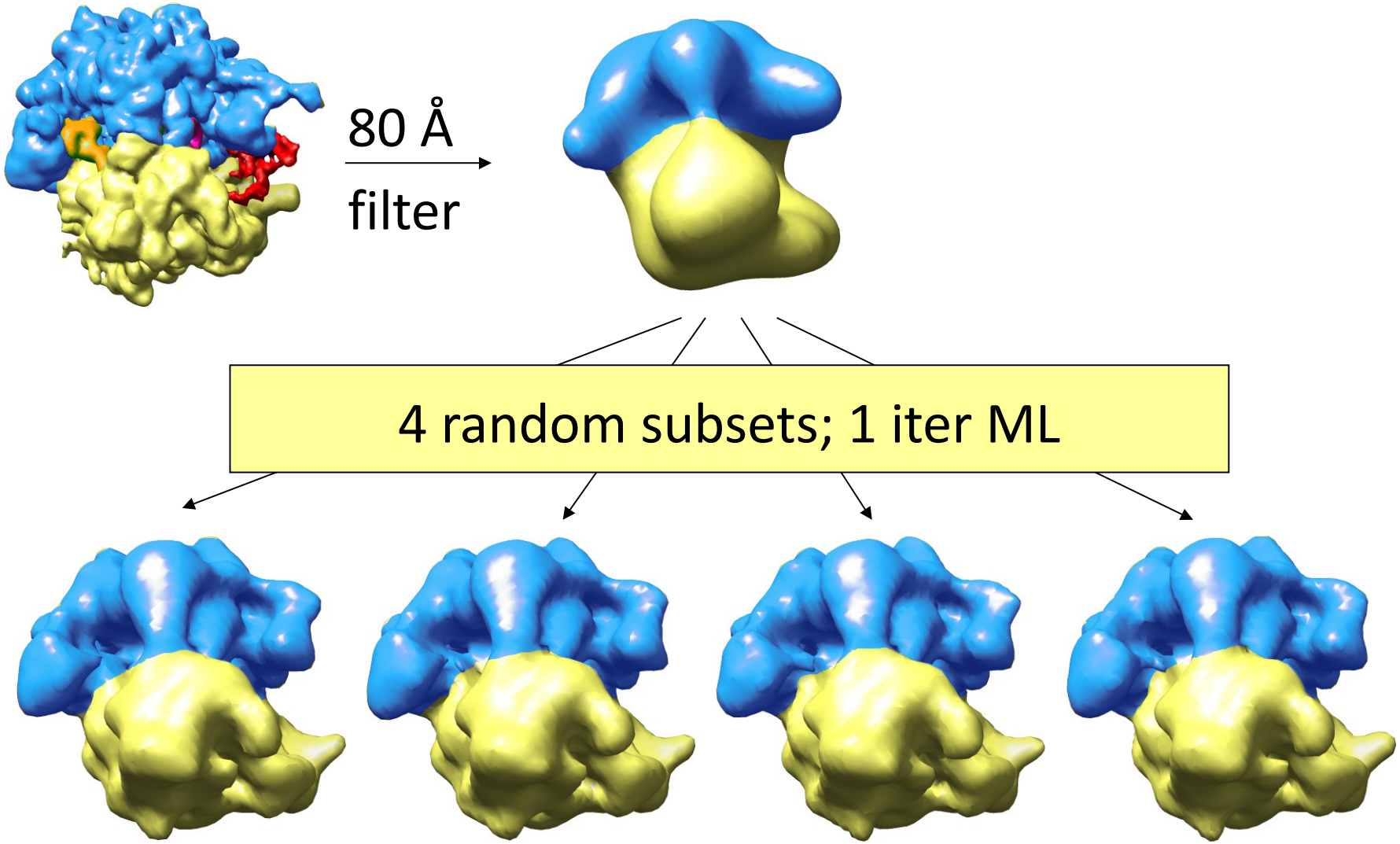
$$\sum P = 1$$

Maximum likelihood



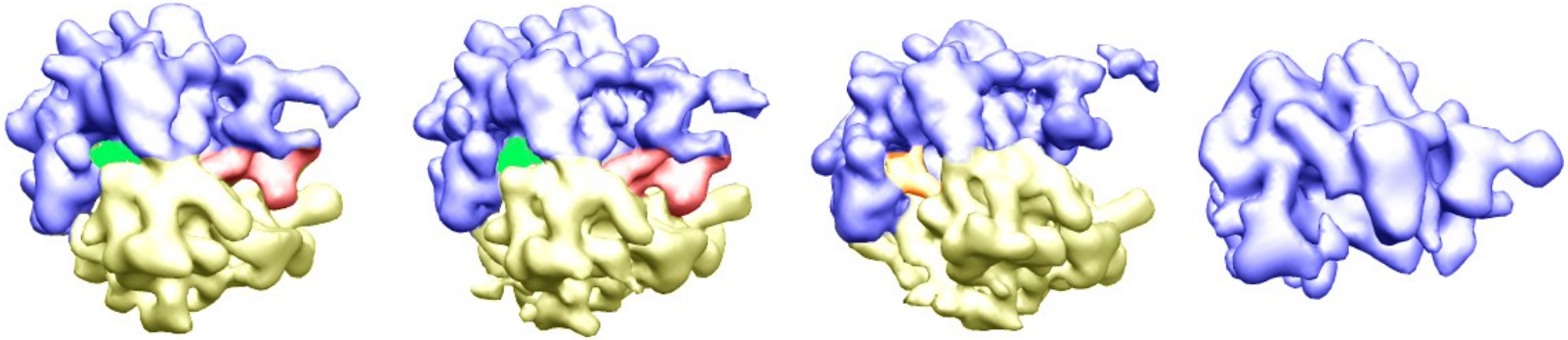
$$\sum P_i = 1$$

Seed generation



Classify structural variability

- Standard data set from the Frank lab
  - 10,000 70S ribosomes (50% +EFG; 50% -EFG)
  - MAP-refinement K=4



24%	28%	42%	6%
26Å	19Å	19Å	30Å

Maximum-likelihood approaches

- Marginalize over orientations & classes
  - Probability-weighted assignments
- First described by Fred Sigworth (JSB-1998)
  - For 2D-alignment, single-reference
  - Real-space data model (white-noise model)
  - Matlab scripts
- Then extended for 2D & 3D classification

Regularised likelihood approach

(2012)

- Data model in Fourier-space
  - Colored (correlated) noise
  - CTF-correction
- Marginalize over orientations & classes
  - Probability-weighted assignments
- Regularization term

Hot topics?

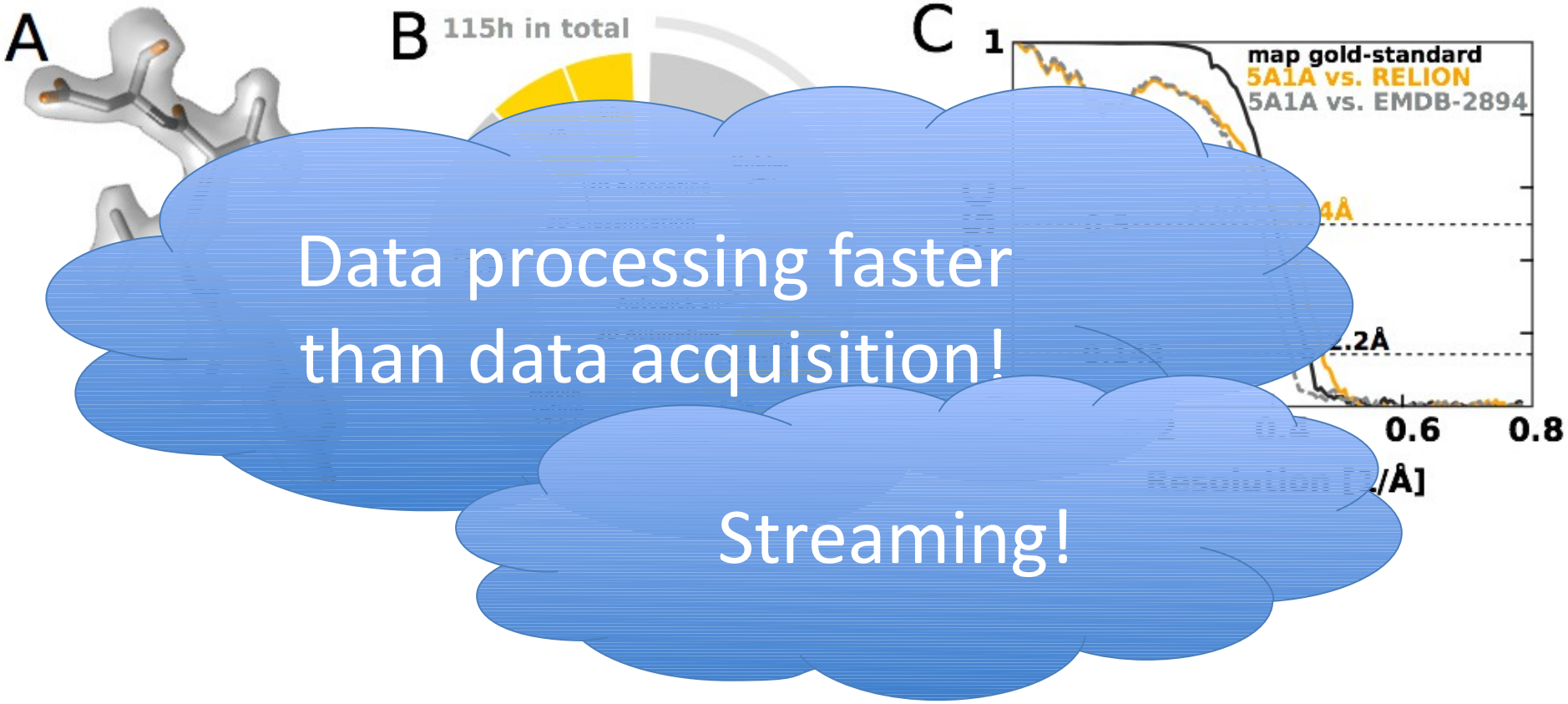
- Beam-induced motion correction
- Robust initial model generation
- 3D classification
- Computational costs

Mike's talk

Reducing computational costs

- Local searches of orientations
  - Formalised by branch-and-bound in cryoSPARC (Marcus)
- GPU-implementations
  - MotionCor(r,2), Gctf, EMAN, RELION, cryoSPARC
- Faster CPU-implementations
  - FREALIGN, CTFFIND4, RELION (v3?)

Desktop-based structure determination



Hot topics?

- Beam-induced motion correction
- Robust initial model generation
- 3D classification
- Computational costs

Hot topics?

- Beam-induced motion correction
- Robust initial model generation
- 3D classification
- Computational costs

Introduction and new approaches

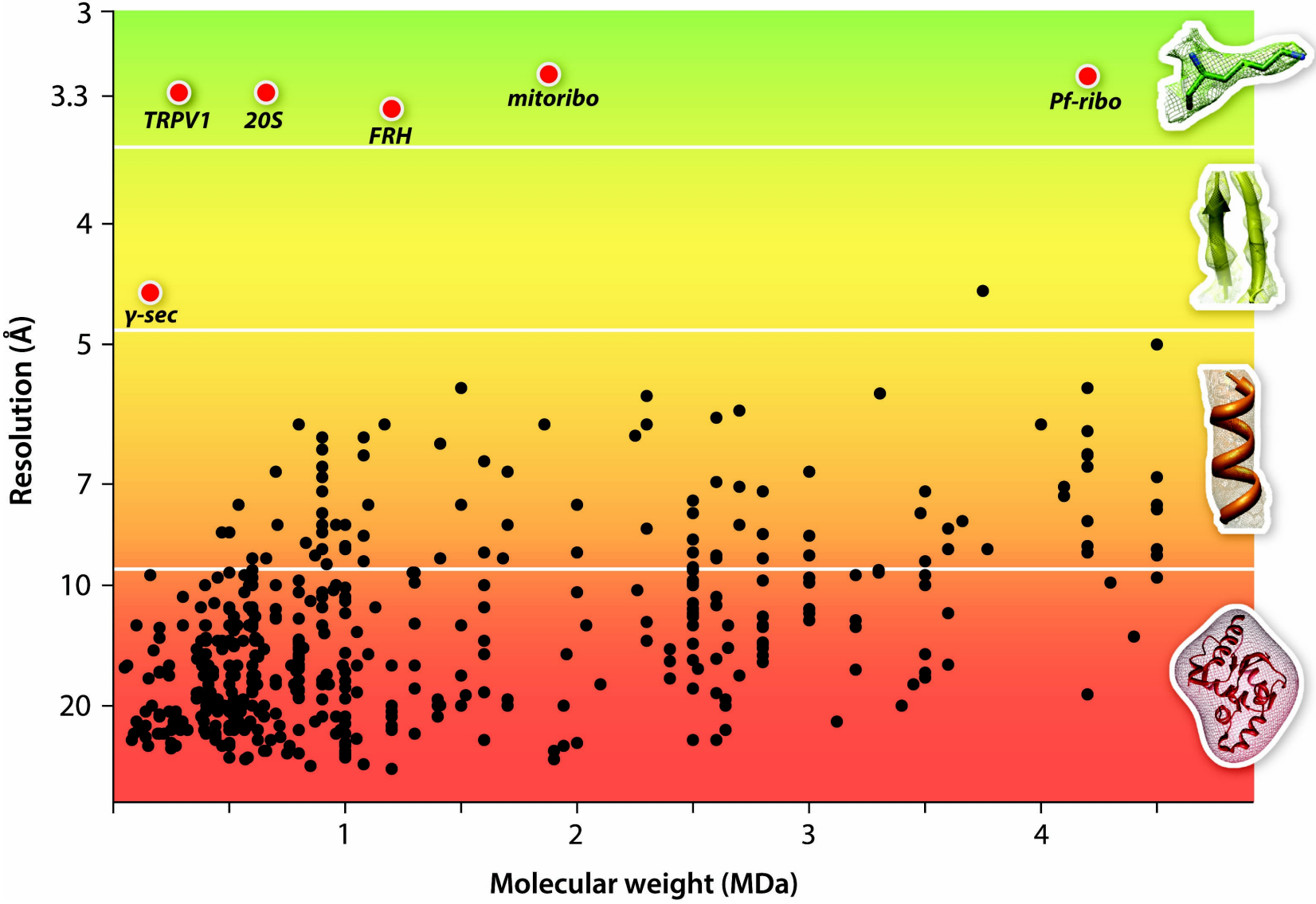
A comprehensive overview of the major advances that have taken place in the last few years that have enabled maps to achieve "atomic" resolution.

Topics to be covered include:

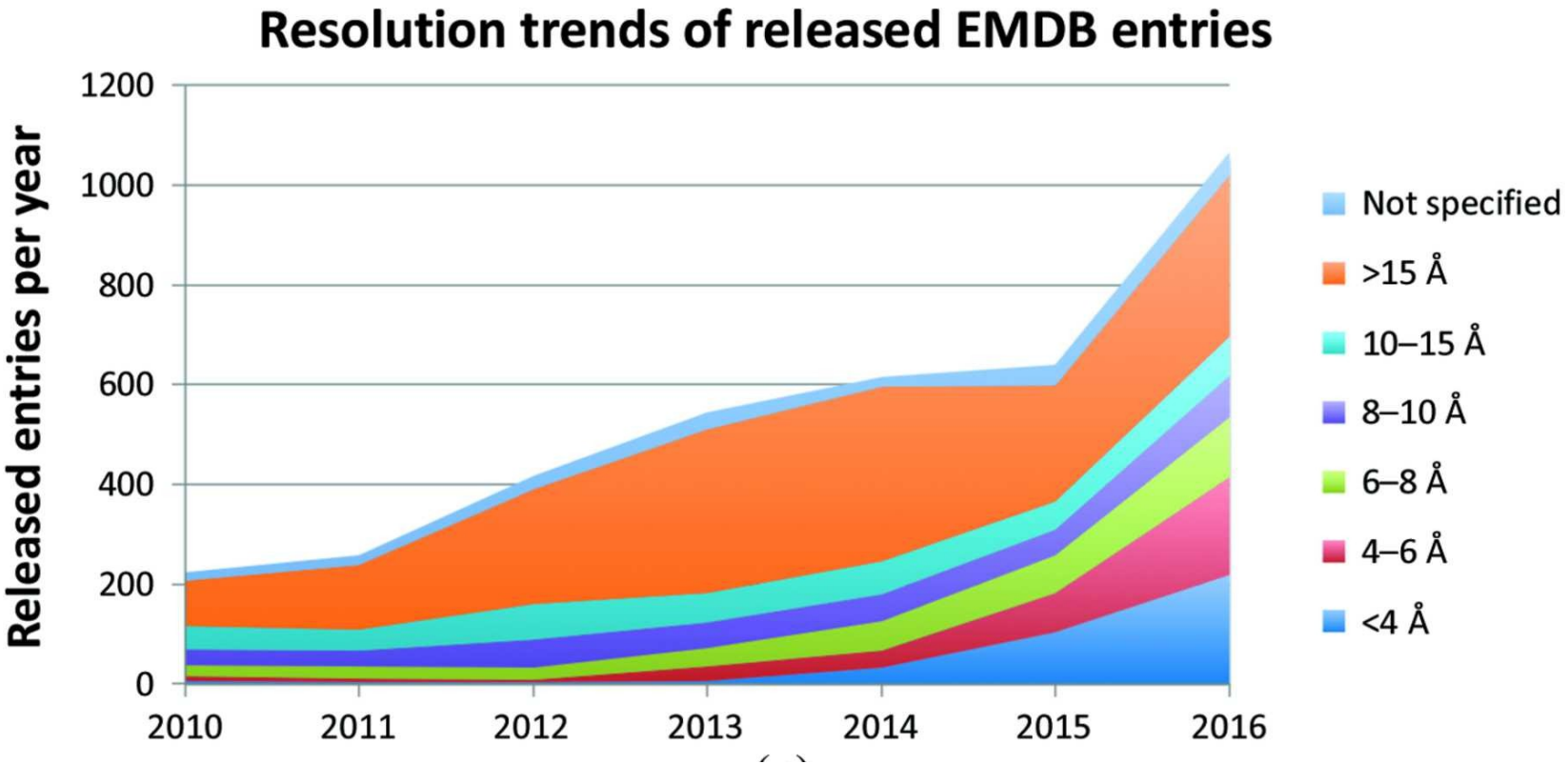
- 3D reconstruction
- image restoration techniques
- how to deal with heterogeneous populations.

- What are the hot topics in processing?
- What are the major mathematical approaches and available software?

Success Stories (2014)

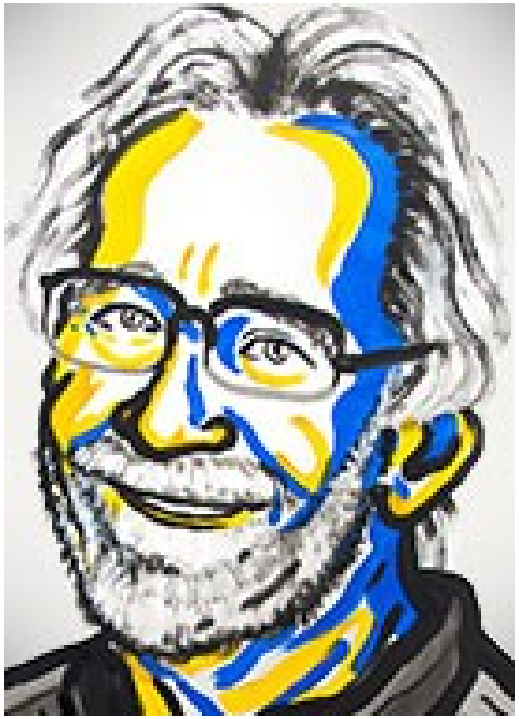


Success stories (2017)



Patwardhan, IUCr J, 2016

Success stories (2017)



Jacques Dubochet

Joachim Frank

Richard Henderson

Introduction and new

A comprehensive review of the
in the last
resolu

You never hear about these……..

We have them very often! Mostly related to sample or grid preparation….

Topics

- 3D reconstruction
- image restoration techniques
- how to deal with heterogeneous populations.

- What are the hot topics in processing?
- What are the major mathematical approac
software?

We don't like: negative stain & cross-linking

Introduction and new approaches

A comprehensive overview of the major advances that have taken place in the last few years that have enabled maps to achieve "atomic" resolution.

Topics to be covered include:

- 3D reconstruction
- image restoration techniques
- how to deal with heterogeneous populations.

- What are the hot topics in processing?
- What are the major mathematical approaches and available software?

Challenges in processing

Marcus' talk

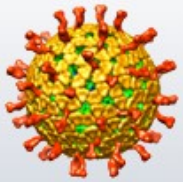
- Ever higher resolutions
  - Beam-tilt, Ewald sphere, precise CTF-estimation, (anisotropic) magnification, beam-induced motion correction

Discussed by Rado yesterday

- Smaller complexes
  - Phase plates

Niko's talk

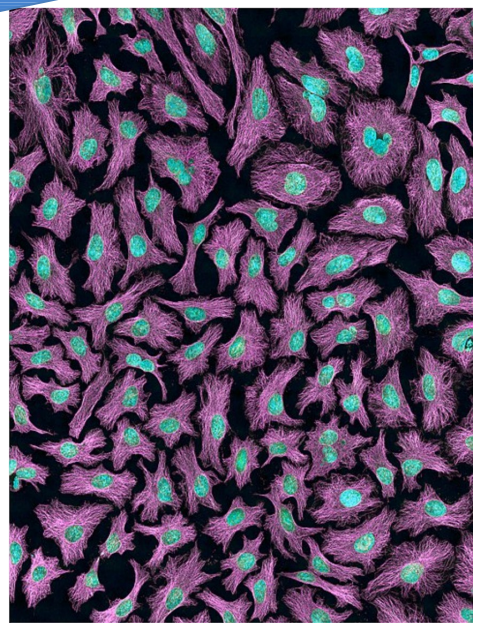
- Structural heterogeneity

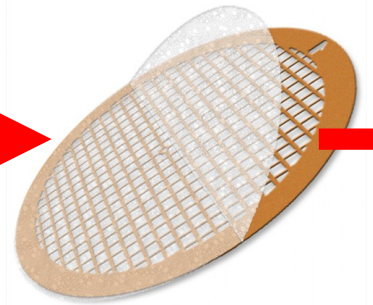


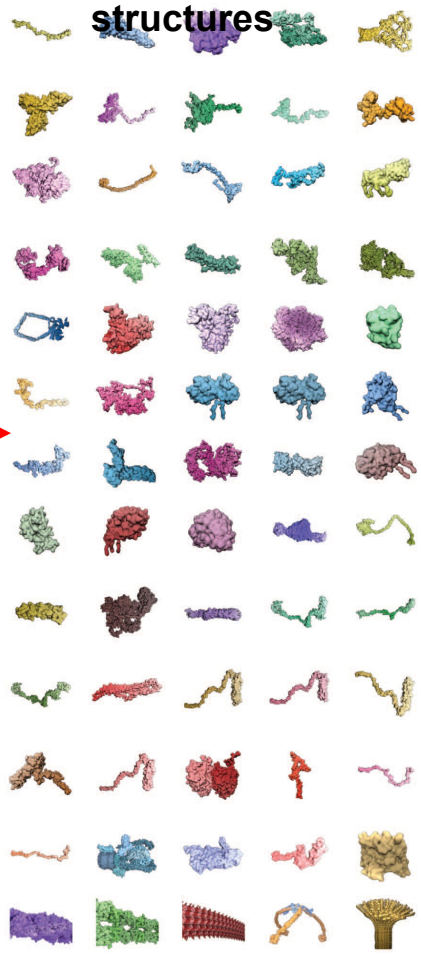
Niko's talk!

EM grid

3D structures



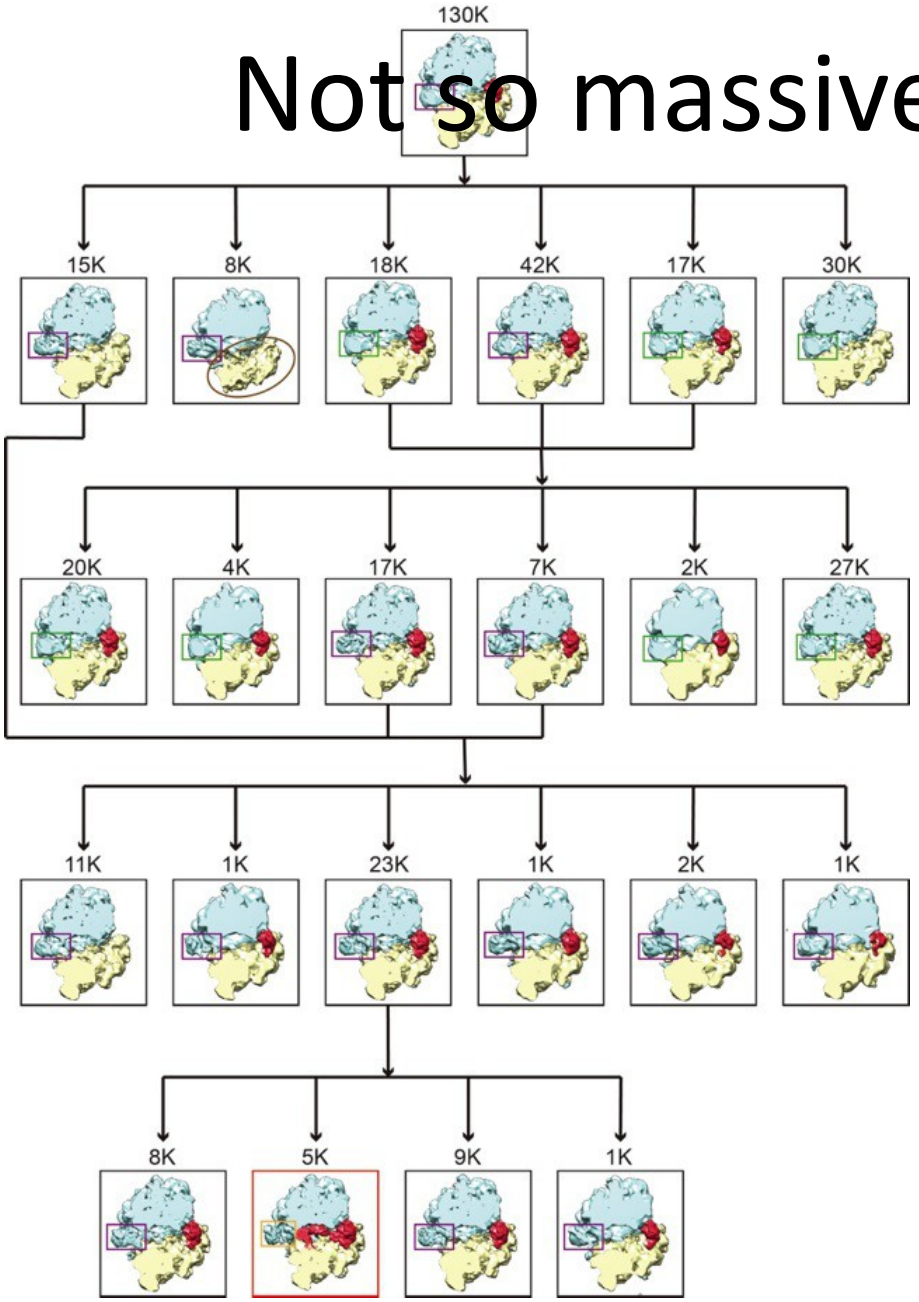


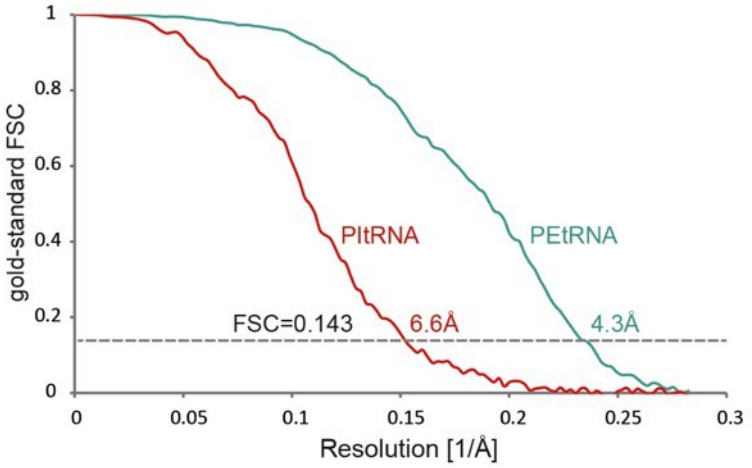


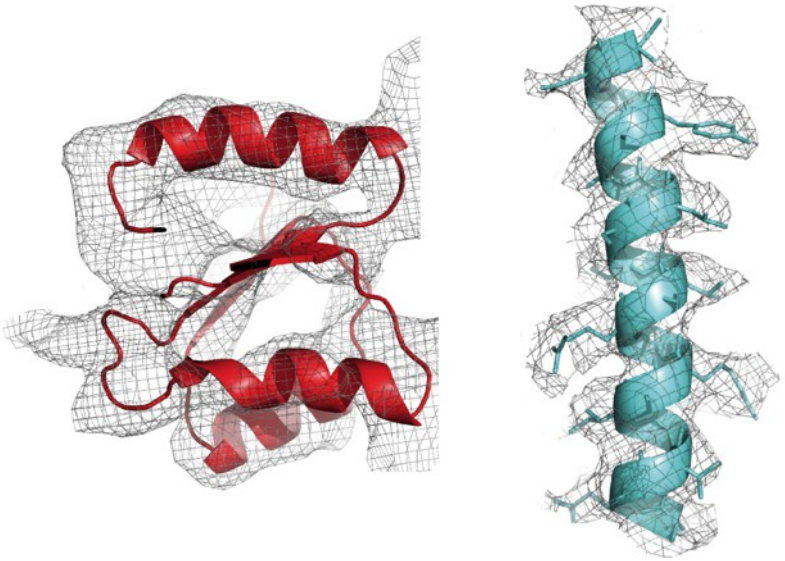
What are the challenges?

commons.wikimedia.org amazon.com emresolutions.com Wilhelm et al. 2014

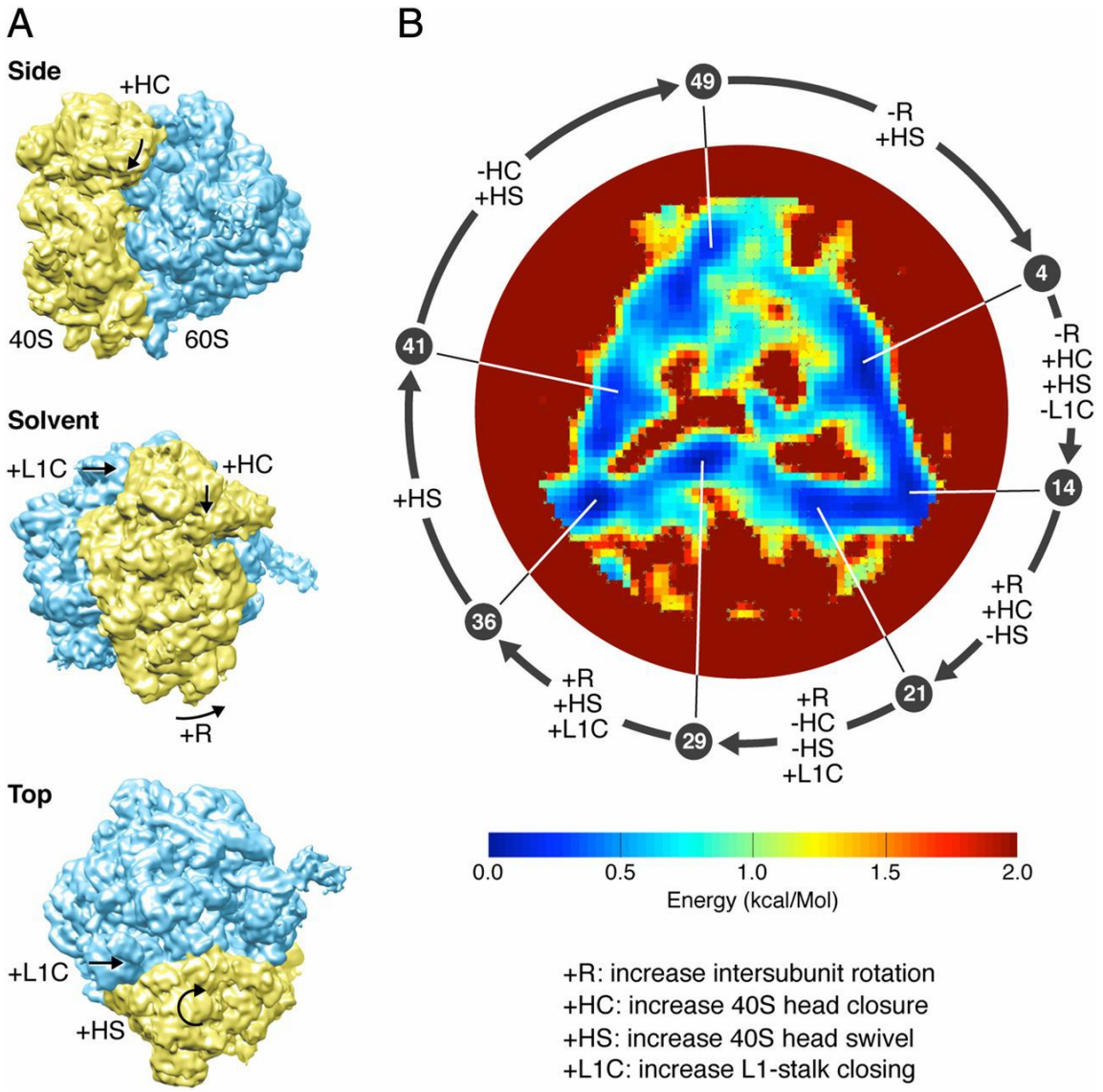
Not so massive classification







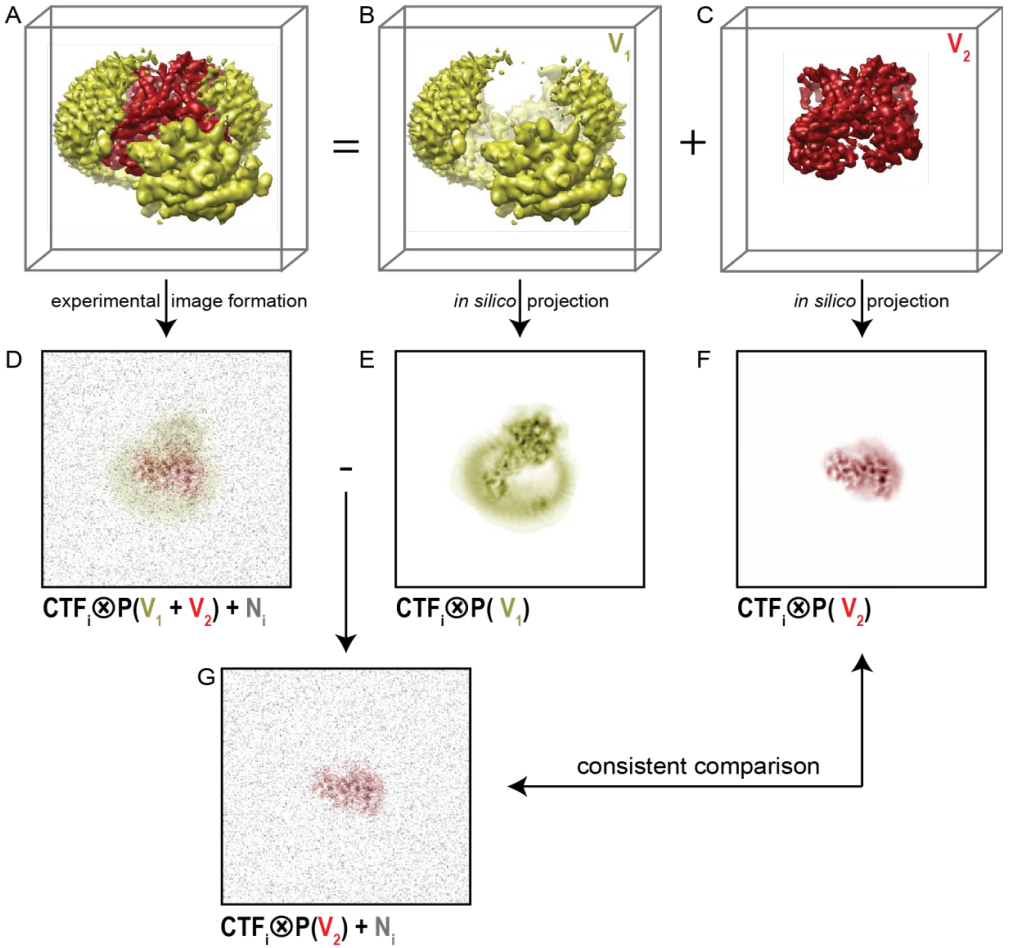
Manifold embedding

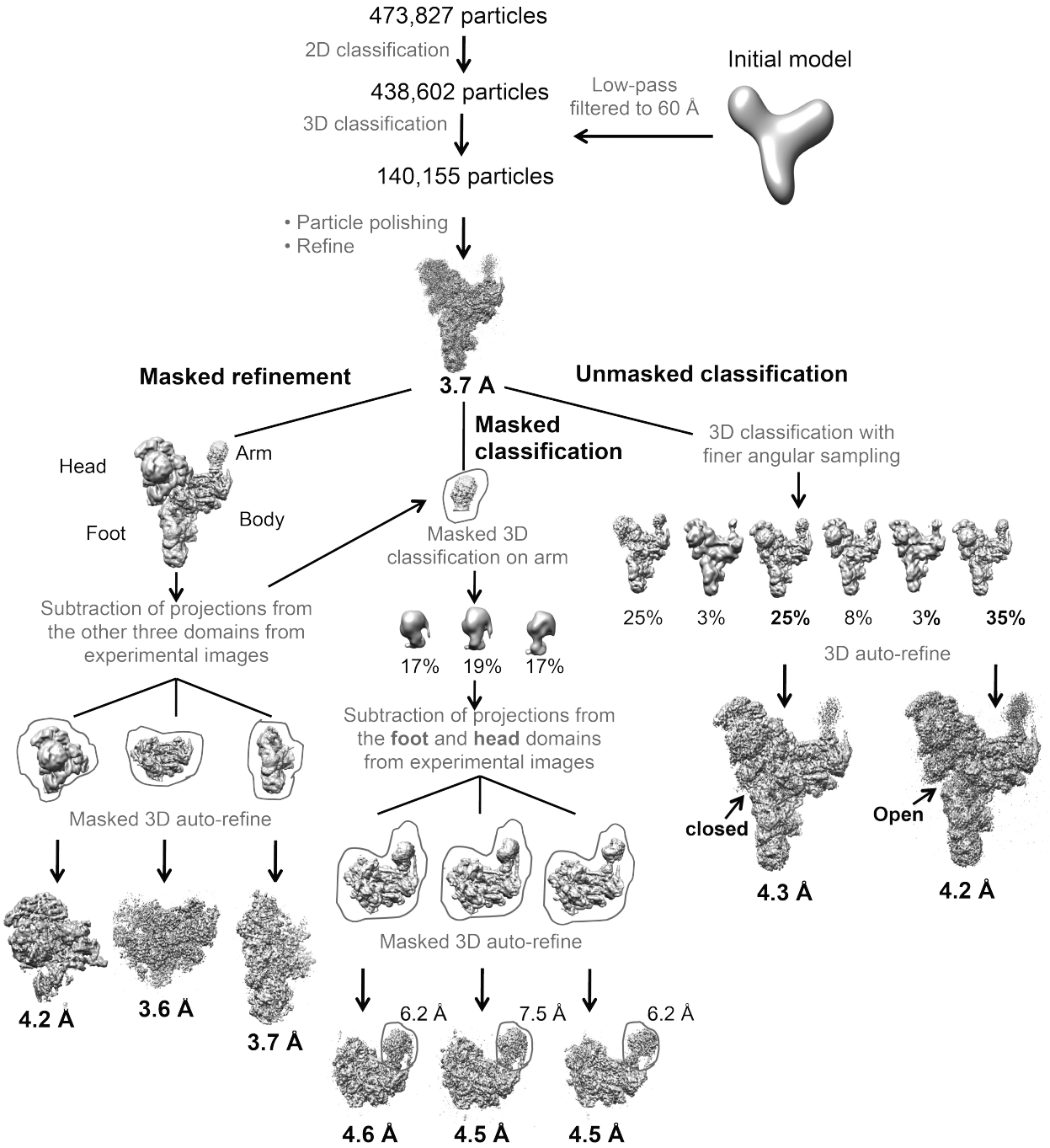


Ali Dashti et al. PNAS 2014;111:17492-17497

Continuous heterogeneity

- Focused refinement (partial signal subtraction)
  - Juha Huiskonen
  - Hongwei Wang
  - Ourselves



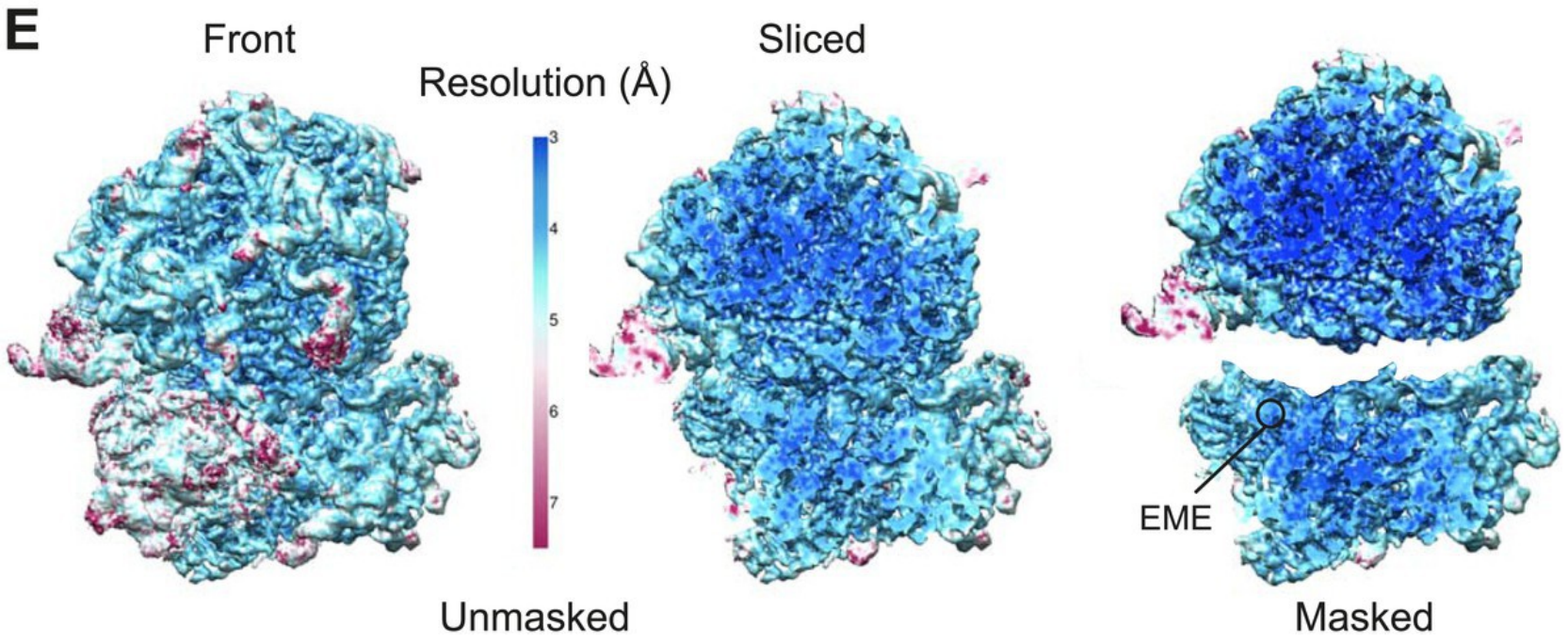


Kelly Nguyen, Kiyoshi Nagai, tri-snRNP spliceosome 2015-2016

Multi-body refinement

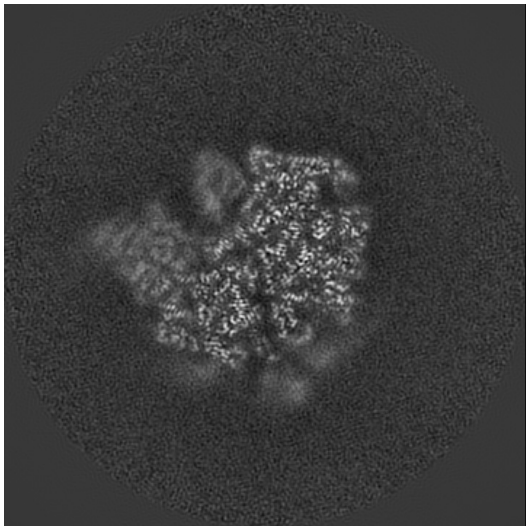
- Divide complex in user-defined bodies
  - Assume each moves as a rigid body…
  - Provide (possibly overlapping) soft masks
- Within each E-M iteration:
  - Focused refinement for each body
  - Update orientations for all bodies continuously

P. falciparum Ribosome

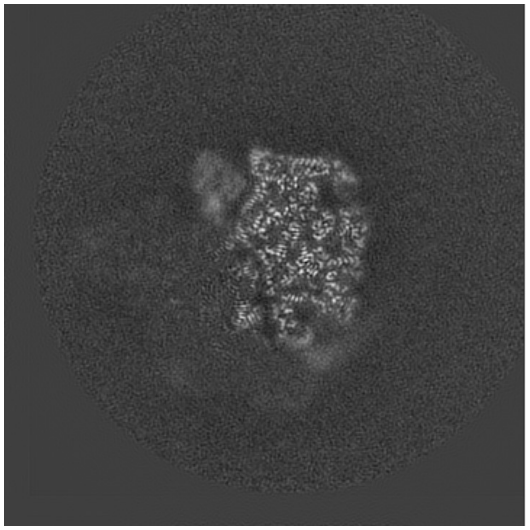


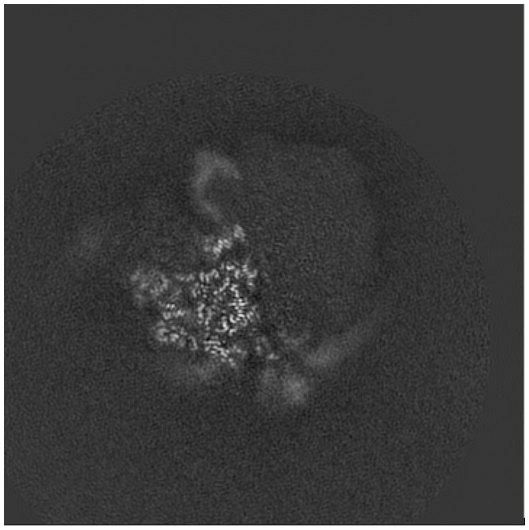
Wong et al, eLife 2014

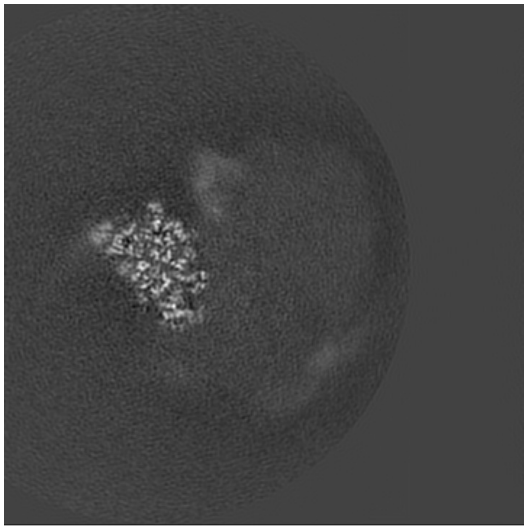
Consensus refinement



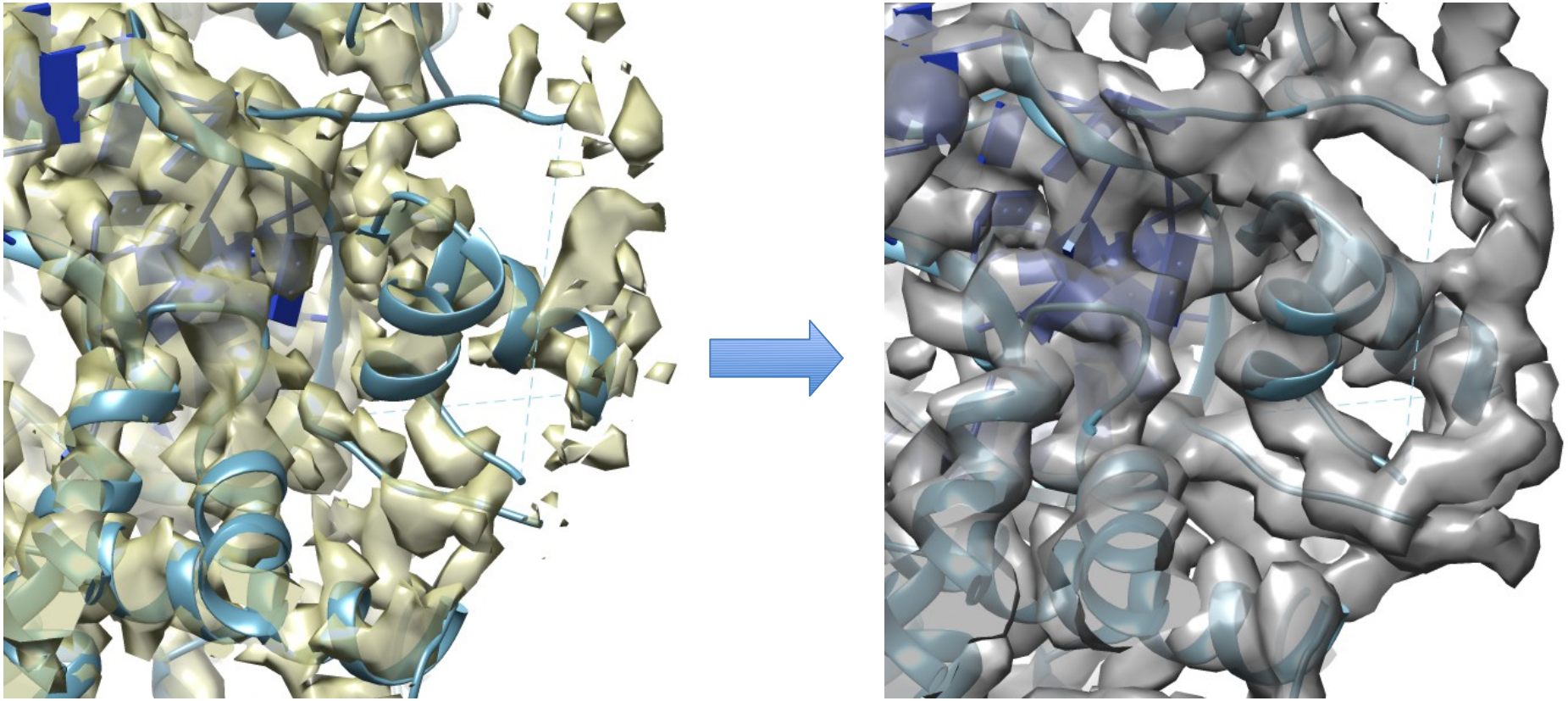
3-body ribosome refinement



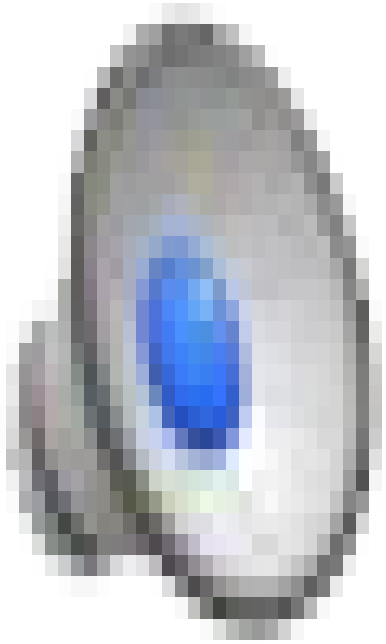




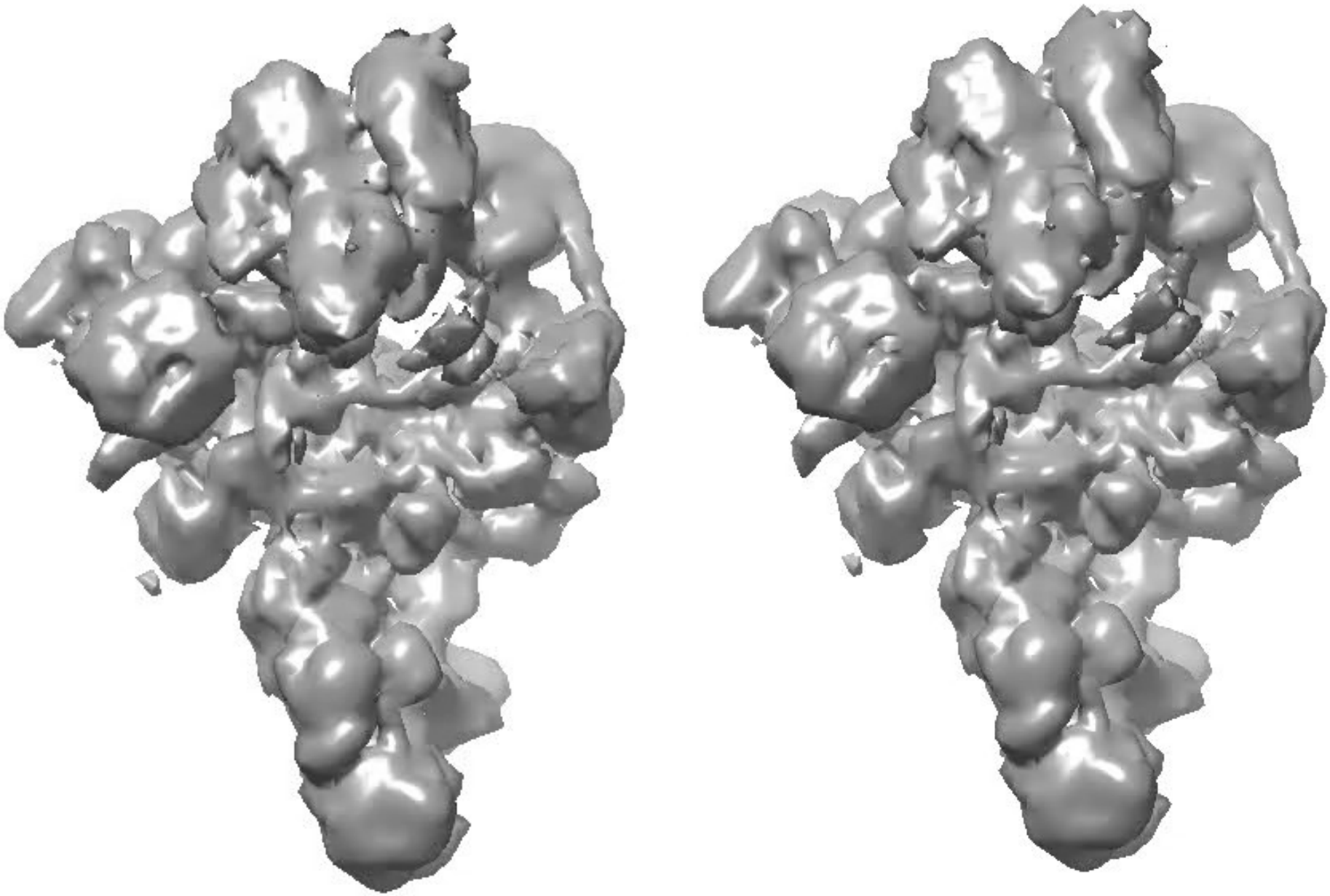
Improved head density



PCA on body orientations



Spliceosomal B-complex



Introduction and new approaches

A comprehensive overview of the major advances that have taken place in the last few years that have enabled maps to achieve "atomic" resolution.

Topics to be covered include:

- 3D reconstruction
- image restoration techniques
- how to deal with heterogeneous [illegible]

Yes, please! Many outsiders coming into the field. Not only biologists, also computational scientists and mathematicians!

- What are the hot topics in processing?

Steve's talk

- What are the major mathematical approaches and available software?

Introduction and new approaches

A comprehensive overview of the major advances that have taken place in the last few years that have enabled maps to achieve "atomic" resolution.

Topics to be covered include:

- 3D reconstruction
- image restoration techniques
- how to deal with heterogeneous populations.

- What are the hot topics in processing?
- What are the major mathematical approaches and available software?

Mistakes to avoid (2014)

New software:
SPHIRE, CisTEM,
cryoSPARC, *etc*!

Detectors: K2, Falcon, VIPS, …

Software: SPIDER, IMAGIC, EMAN, SPARX, XMIPP, BSOFT, FREALIGN, RELION, APPION, …

- Open-source
- SPIDER
- EMAN
- SPARX
- XMIPP
- RELION
- FREALIGN/CTFFIND/ UNBLUR

Who will build on closed-source software?

Conclusions (2014=2017)

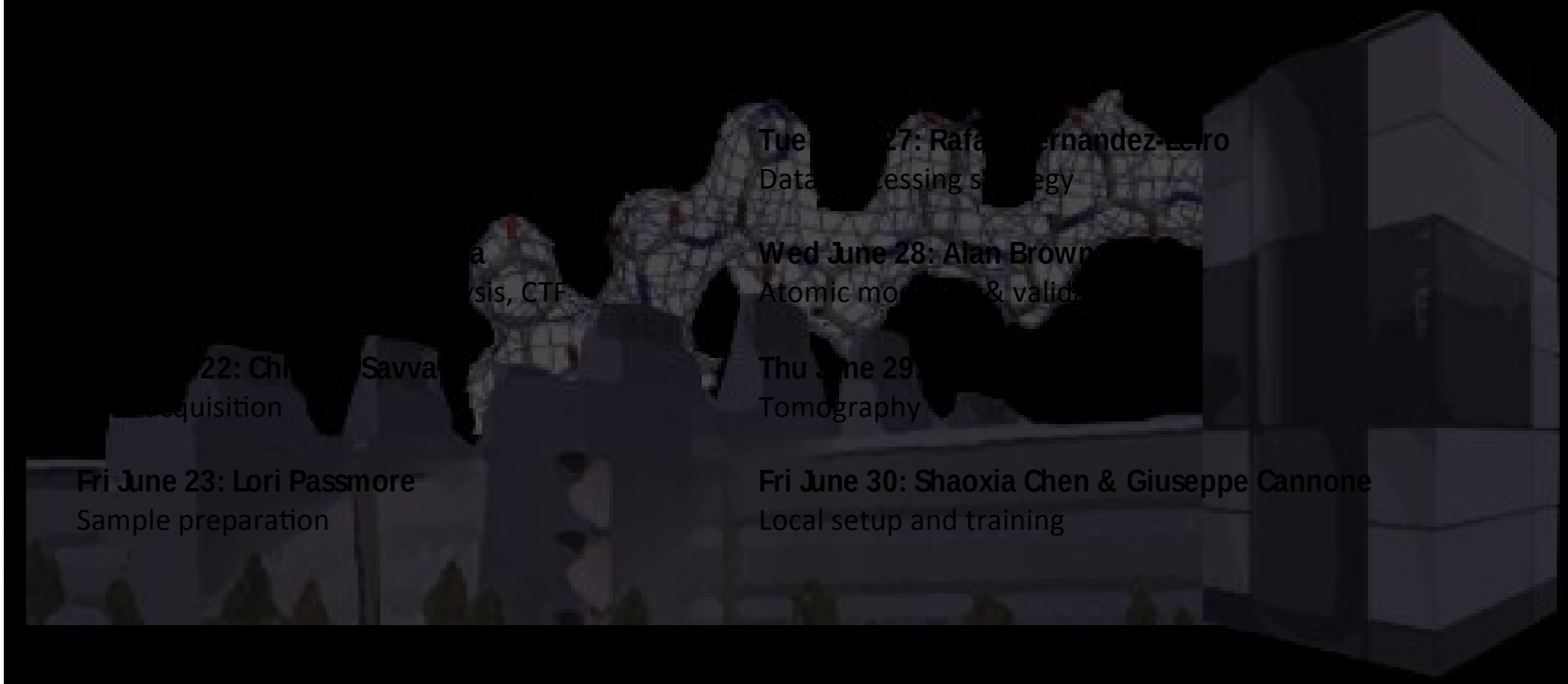
- Image pr
  field

- As has jus
  contin

Michel Goedert and myself
are looking for post-docs with experience in
cryo-EM for studying amyloids in
neurodegenerative disease

Making good samples already was crucial, but will be ever more important!

LMB cryo-EM course 2017



Tue [illegible]27: Rafa[illegible]ernandez-Leiro
Data [illegible]cessing s[illegible]egy

Wed June 28: Alan Brown[illegible]
Atomic mo[illegible]& valid[illegible]

Thu [illegible]ne 29[illegible]
Tomography

Fri June 30: Shaoxia Chen & Giuseppe Cannone
Local setup and training

[illegible]22: Chi[illegible] Savva[illegible]
[illegible]equisition

Fri June 23: Lori Passmore
Sample preparation

[illegible]sis, CTF[illegible]

ftp[illegible]e

Processing

09:00 Sjors Scheres: Intro and new approaches

10:00 Coffee Break

10:30 Niko Grigorieff: New challenges

11:15 Steve Ludtke: Deep learning methods

12:00 Lunch

13:00 Marcus Brubaker: Bayesian methods

13:30 Michael Cianfrocco: Cloud computing

14:00 Panel discussion (Chair John Rubinstein)