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
Introduction and new approaches

A comprehensive overview in the last few years that has 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?
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
Electron microscopy imaging

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

Initial 3D model
Projection matching

maxCC

compare with all projections
Projection slice theorem
Projection slice theorem
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
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**

\[ X_i = \text{CTF}_i \ast P_\phi V_k + N_i \]

- Convoluted w/ CTF
- \( P_\phi \) implements integrals

- **Fourier space**

\[ X_i = \text{CTF}_i P_\phi V_k + N_i \]

- Multiply w/ CTF
- \( P_\phi \) takes a slice
Phase flipping

- Easy to do
- Reasonably effective
- Problems in classification?
(3D) Wiener filter

\[ V = \frac{\sum_{i=1}^{N} P_{\varphi}^{T} \frac{CTF_i}{\sigma_i^2} X_i}{\sum_{i=1}^{N} P_{\varphi}^{T} \frac{CTF_i^2}{\sigma_i^2} + \frac{1}{\tau^2}} \]

- \( \sigma^2 \): 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?

• Beam-induced motion correction

• Robust initial model generation

• 3D classification

• Computational costs

Discussed by John yesterday
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

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)

maxCC=0.32

Noise in the images makes this assignment highly stochastic!

CC=0.31

CC=0.24
Maximum likelihood

Calculate a probability for all orientations

P=0.4
P=0.35
P=0.0001

ΣP_{OLE}=1
Maximum likelihood

Avoid taking hard decisions if the noise does not allow this.

$P = 0.4$

$P = 0.35$

$P = 0.0001$

$\sum_{P} = 1$
Seed generation

80 Å filter

4 random subsets; 1 iter ML
Classify structural variability

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

<table>
<thead>
<tr>
<th></th>
<th>24%</th>
<th>28%</th>
<th>42%</th>
<th>6%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Å</td>
<td>26Å</td>
<td>19Å</td>
<td>19Å</td>
<td>30Å</td>
</tr>
</tbody>
</table>
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
  - Penalize high-frequency components
  - Elegant derivation of 3D Wiener filter
  - Iteratively learn power of signal and noise from the data
  - No user-expertise required to optimally filter data/map
  - Objectivity

- RELION (from 2012)
- cryoSPARC (Marcus, from 2015)
Hot topics?

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

Data processing faster than data acquisition!

Streaming!
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)

Resolution trends of released EMDB entries

- Not specified
- >15 Å
- 10–15 Å
- 8–10 Å
- 6–8 Å
- 4–6 Å
- <4 Å

Patwardhan, IUCr J, 2016
Success stories (2017)

Jacques Dubochet

Joachim Frank

Richard Henderson
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?
- What are the success stories and the failures?
- Where are the greatest challenges right now and how are we approaching these?
- Do we need completely new algorithms or just incremental improvements on the current ones?

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

You never hear about these.......

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

- Ever higher resolutions
  - Beam-tilt, Ewald sphere, precise CTF-estimation, (anisotropic) magnification, beam-induced motion correction
- Smaller complexes
  - Phase plates
- Structural heterogeneity

Discuss by Rado yesterday

Marcus’ talk

Niko’s talk
What are the challenges?
Not so massive classification
Manifold embedding

A

Side

+HC

40S |

+L1C

60S

Solvent

+HC

+R

Top

+L1C

+HS

B

-R

+HS

+HC

-R

+HS

-L1C

49

4

36

29

21

41

Energy (kcal/Mol)

+R: increase intersubunit rotation
+HC: increase 40S head closure
+HS: increase 40S head swivel
+L1C: increase L1-stalk closing

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

- Fully automated
  - Continuation of 'consensus' Refine3D run
P. falciparum Ribosome

Wong et al, eLife 2014
Consensus refinement

3-body ribosome refinement
Improved head density
PCA on body orientations
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
- What are the hot topics in processing?
- What are the major mathematical approaches and available software?

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

Steve’s talk
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)

Detectors: FEI, Jeol, Zeiss, ...
Detectors: K2, Falcon, DE, TVIPS, ...

Software: SPIDER, IMAGIC, EMAN, SPARX, ...
Software: SPHER, CisTEM, cryoSPARC, etc!

New software: SPHER, CisTEM, cryoSPARC, etc!
Who will build on closed-source software?
Conclusions (2014=2017)

- Image processing will continue to drive this field forward.
- As has just happened: new hardware will also continue to have huge impacts.
- Making good samples already was crucial, but will be ever more important!

Michel Goedert and myself are looking for post-docs with experience in cryo-EM for studying amyloids in neurodegenerative disease.
LMB cryo-EM course 2017

Mon June 19: Local Setup

Tue June 20: Javier De La Cruz & Joel Houston
Data processing strategy

Wed June 21: John Hough
Image analysis, CTF

Fri June 22: Chiara Savalli
Acquisition

Fri June 23: Lori Passmore
Sample preparation

Tue June 27: Rafael Fernandez-Ejido
Data processing strategy

Wed June 28: Alan Brown
Atomic models & validation

Thu June 29: Annette Hoppe
Tomography

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