Workshop on Management of Large CryoEM Facilities

PROCESSING...

Edward T Eng | Anchi Cheng February 6, 2017



MICROSCOPY CENTER

New York Structural Biology Center



Processing at SEMC







Infrastructure

Efficiency



Balance



Strategy







Top requests for processing





2017 data generation at SEMC **EM** instrumentation Cameras FEI Titan Krios#1 / #2 / #3 Falcon3 x3 **7** direct detectors Since 2015: K2 x3 1,548,411 total images on 4 TEMs FEI Tecnai F20 (1.1 mil images **DE20** since 2016) **TVIPS 4K CMOS FEI Tecnai Biotwin** For 2016: **TVIPS 4K CMOS** 4 CMOS/CCDs 170,933 Krios1 **JEOL 1230** on 3 TEMs + direct detector movies Gatan US4000 CCD (exposure images) 1 SEM **FEI Helios 650 Quorum cryostage**

Data acquisition and processing platforms

EM instrumentation

FEI Titan Krios#1 / #2 / #3 Falcon3 x3 K2 x3

> FEI Tecnai F20 DE20 TVIPS 4K CMOS

FEI Tecnai Biotwin TVIPS 4K CMOS

JEOL 1230 Gatan US4000 CCD

FEI Helios 650 Quorum cryostage









SEMC computational support

- HPC Server and storage (DDN):
- 2 x 42U rack enclosures
- DDN GRIDScaler GS7K appliance with 1.1PB GPFS paralegal file system
- 420TB DDN WOS object storage for archival
- 1056 x CPU cores. 44 x SuperMicro nodes each with 24 x CPU cores and 256GB RAM
- 4 x GPU nodes each with one GPU and 128GB RAM. One GPU server with 8 x GPUs and 512GB RAM and 2 x GPU servers each with 4 x GPUs and 512GB RAM.
- 4 buffer servers each with 51TB local storage, 2 x GPUs, 128GB RAM and 10G Fiber Network cards.
- 5 x 36 QSFP port 56Gb/s FDR InfiniBand switches.
- Bright Cluster Manager
- Basic Onsite Support; 7x24 remote support



Additional computational resources Web resources **On site resources Cluster resources** EM session user Data processing user SEMC Leginon / support Amira emgweb.nysbc.org Goby cluster workstations (9) workstations (2) emg.nysbc.org/ **GPU** workstations Control room SEMC redmine (5) workstations (3) **Guppy cluster** SEMC Data transfer semc.nysbc.org cluster computers (2)





PDB to Model EMDB to Model Upload model 1 available

Example: Single-particle workflow During EM session After EM session Appion Data Processing Home institution **SEMC** computing computing/ Project: SEMC - Glutamate Dehydrogenase (C Session: 17jan26c - Krios2 gdh phase pla Image Path: /gpfs/leginon/rice/17jan26c/rawdata Micrograph/ 2D classification Hide | Expand | Contract Particle sorting Object Selection : 1 CTF Estimation Stacks : 3 Particle Alignment Initial 2D Ab Initio Reconstruction **3D** classification classification Refine Reconstruction Helical Processing Tomography Direct Detector Tools Import tools Initial model **3D** refinement Upload particles generation Upload template 23 available Upload CTF Upload stack Upload reconstruction

3D refinement

Model building

Web integrated processing



During EM session

Appion Data Processing

Project: SEMC - Glutamate Dehydrogenase (C Session: 17jan26c - Krios2 gdh phase plat Image Path: /gpfs/leginon/rice/17jan26c/rawdata

 > Object Selection : 1 > CTF Estimation > Stacks : 3 > Particle Alignment > Ab Initio Reconstruction > Refine Reconstruction > Helical Processing > Tomography > Direct Detector Tools > Import tools Upload particles Upload template 23 available Upload stack Upload stack Upload reconstruction PDB to Model EMDB to Model Upload model 	Hide Expand Contract	
 CTF Estimation Stacks : 3 Particle Alignment Ab Initio Reconstruction Refine Reconstruction Helical Processing Tomography Direct Detector Tools Import tools Upload particles Upload template 23 available Upload CTF Upload stack Upload reconstruction PDB to Model EMDB to Model Upload model 	Object Selection : 1	
 Stacks : 3 Particle Alignment Ab Initio Reconstruction Refine Reconstruction Helical Processing Tomography Direct Detector Tools Import tools Upload particles Upload template 23 available Upload Stack Upload stack Upload reconstruction PDB to Model EMDB to Model Upload model 	CTF Estimation	
 Particle Alignment Ab Initio Reconstruction Refine Reconstruction Helical Processing Tomography Direct Detector Tools Import tools Upload particles Upload template 23 available Upload CTF Upload stack Upload reconstruction PDB to Model EMDB to Model Upload model 	Stacks : 3	
 Ab Initio Reconstruction Refine Reconstruction Helical Processing Tomography Direct Detector Tools Import tools Upload particles Upload template 23 available Upload CTF Upload reconstruction PD to Model EMDB to Model Upload model 	Particle Alignment	
 Refine Reconstruction Helical Processing Tomography Direct Detector Tools Import tools Upload particles Upload template 23 available Upload CTF Upload stack Upload reconstruction PDB to Model EMDB to Model Upload model 	Ab Initio Reconstruction	
 Helical Processing Tomography Direct Detector Tools Import tools Upload particles Upload template 23 available Upload CTF Upload stack Upload reconstruction PDB to Model EMDB to Model Upload model 	Refine Reconstruction	
 Tomography Direct Detector Tools Import tools Upload particles Upload template 23 available Upload CTF Upload stack Upload reconstruction PDB to Model EMDB to Model Upload model 	Helical Processing	
Direct Detector Tools Import tools Upload particles Upload template 23 available Upload CTF Upload stack Upload reconstruction PDB to Model EMDB to Model Upload model	▶ Tomography	
Vipload particles Upload template 23 available Upload CTF Upload stack Upload reconstruction PDB to Model EMDB to Model Upload model	Direct Detector Tools	
Upload particles Upload template 23 available Upload CTF Upload stack Upload reconstruction PDB to Model EMDB to Model Upload model	Import tools	
Upload template 23 available Upload CTF Upload stack Upload reconstruction PDB to Model EMDB to Model Upload model	the first state of the second state of	
23 available Upload CTF Upload stack Upload reconstruction PDB to Model EMDB to Model Upload model	Upload particles	
Upload CTF Upload stack Upload reconstruction PDB to Model EMDB to Model Upload model	Upload template	
Upload stack Upload reconstruction PDB to Model EMDB to Model Upload model	Upload template 23 available	
Upload reconstruction PDB to Model EMDB to Model Upload model	Upload particles Upload template 23 available Upload CTF	
PDB to Model EMDB to Model Upload model	Upload particles Upload template 23 available Upload CTF Upload stack	
EMDB to Model Upload model	Upload particles Upload template 23 available Upload CTF Upload stack Upload reconstruction	
Upload model	Upload particles Upload template 23 available Upload CTF Upload stack Upload reconstruction PDB to Model	
	Upload particles Upload template 23 available Upload CTF Upload stack Upload reconstruction PDB to Model EMDB to Model	
1 available	Upload particles Upload template 23 available Upload CTF Upload stack Upload reconstruction PDB to Model EMDB to Model Upload model	

Leginon / Appion web interface



Packages offered



After EM session

CTF Estimation Selection Page

Project: SEMC - Glutamate Dehydrogenase (GDH) (256) Session: 16dec06e - 1.68 mg/mL GDH + 0.001% DDM (1.3/1.4 UltrAu 300mesh) Image Path: (gsfolegion/til/16dec06e/awadata

Hide | Expand | Contract Colject Selection : 3 Select Particle Picker...

Repeat an image loop run

Repeat an image loop run

4 complete

V Stacks : 9

Stack creation

11 complete

CTF Estimation

CTF Estimation Procedures

GCTF v1.06

robust astigmatism estimate.

ACE 2

Appion using this program

Estimate the CTF... 5 complete 1 running Transfer results to another preset

CTFFIND v4

Appion

CTFFIND uses a robust grid search algorithm to find the optimal CTF parameters. Please see the Grigorieff lab website @ for more information.

This is a GPU accelerated program for real-time CTF determination, refinement,

evaluation and correction. Please see the Dynein lab website of for more information.

ACE 2 is an unpublished re-implementation of ACE1, but written in objective-C ACE2

make several improvements over ACE1 including a several speed enhancements and a

Note: It was designed around FEI Tecnai FEG data and other have reported problems

7

more stack tools

Particle Alignment

- Select Particle Alignment... 2 complete
- Run Feature Analysis...

3 complete

- Run Particle Clustering... 4 complete
- Run MaskitOn

Template Stacks

Ab Initio Reconstruction
 OptiMod Common Lines



SIMPLE Common Lines

Refine Reconstruction



Appion web interface



Balance

Who does the processing workflow?



After EM session



Appion web interface

Cluster Stack List		
Project: SEMC - Glutamate Dehydrogenase (GDH) (256) session: 16dec06e - 1.68 mg/mL GDH + 0.001% DDM (1.3/1.4 UltrAu 300mesh) Image Path: /gel/signor/lifut/66ce06/avada		
Hide Expand Contract		
Select Particle Picker	Cluster Stack List	
4 complete		
Repeat an image loop run	Show Composite Page	
 CTF Estimation 		
Estimate the CTF	Clustering Info: kerden3 (ID: 13) with 1 clusters	
5 complete	Turne Verlag KarDag 2014	
1 running	Type: Xmipp Kerben SOM	
Transfer results to another preset	- <u>b, b, 9 7 9 7 -</u>	
Repeat an image loop run		
▼ Stacks : 9		
Stack creation		
11 complete		
more stack tools		
Particle Alignment	 View montage of self-organizing map 	
Select Particle Alignment	 View montage as a stack for further processing (ID 39) 	
2 complete		
Run Feature Analysis		
3 complete	Clustering Info: coran1 (ID: 12) with 2 clusters	
Run Particle Clustering		
4 complete	Type: SPIDER Coran	
Run MaskitOn	Factor list:	
Template Stacks	- 160 Close Averages (variance) (ID 41)	
Ab Initio Reconstruction	80 Class Averages [variance] (ID 41)	
OptiMod Common Lines		
EMAN Common Lines		
SIMPLE Common Lines	Clustering Info: kerden1 (ID: 11) with 1 clusters	
Refine Reconstruction		
Run Single-Model Refinement	Type: Xmipp KerDen SOM	
1 jobs ready to run		
Run Multi-Model Refinement		
Helical Processing		
Helical Image Processing		
Direct Detector Tools		
	that had had had had her P-R P-R	

Balance

Who does the processing workflow?



After EM session



Appion web interface for TEM use





Too busy to improve?

audiencestack.com



Processing at SEMC





Infrastructure

Community

What is our user community asking for?

What is our infrastructure?

Efficiency How does processing lead

data collection?

to more efficient user effort?

What is the balance between staff and

Balance

Strategy How do we evolve and prepare for future advances?



PROCESSING (cont)

before scaling up and support issues

...Anchi





Anchi Cheng Sargis Dallakyan Res. Staff Scientist Res. Programmer

Carl Negro Res. Programmer



SEM

2015-2016 computation support

- Guppy (8 nodes with 4-8 CPU cores)
- 1 Database and 1 web server
- 3 gpu workstations
- Individually mounted storage 300 TB
- GPFS storage 0.5 PB
- DE20 dedicated processing server (24 Processors)
- 1056 x CPU cores. 44 x SuperMicro nodes each with 24 x CPU cores and 256GB RAM
- 4 x GPU nodes each with one GPU and 128GB RAM.



Typical Problems

During EM session

- Frame alignment from the last user is not finished
- Web service image loading can not keep up with multiple user viewing.
- Database query lock up.
- Computation resource upgrade in progress



- Inefficient request of resource
 - Ask more than needed
 - Ask less than required
- Small one processor jobs compete with MPI processes.
- Users do not (know how to) clean up bad processing results.
- New development breaks the pipeline.



Simons Electron Microscopy Center





Anchi Cheng Sargis Dallakyan Res. Staff Scientist Res. Programmer



Yong Zi Tan Grad. Student

Bill Rice

EM Manager

Venkat Dandey

Alex Noble Post Doc.



Ed Eng Staff Scientist



Micah Rapp

Ashleigh Raczkowski Kelsey Jordan Senior Technician Technician



Laura Kim Research Associate





Clint Potter

Director

Bridget Carragher Director

Staff Scientist



Priyamvada Acharya Giovanna Scapin Julia Brasch Embedded Post Doc. Embedded Scientist Embedded Post Doc.





Crystal Premo Administrator



National Resource for Automated Molecular Microscopy http://nramm.nysbc.org





Carl Negro Res. Programmer

Phil Baldwin Alex Wei Technician