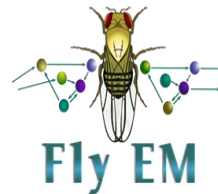


Scaling Up the Connectome *in the Adult Fly Brain*



Stephen Plaza



<https://www.janelia.org/project-team/fly-em>

@janelia-flyem



@stephenplaza

@janeliaflyem



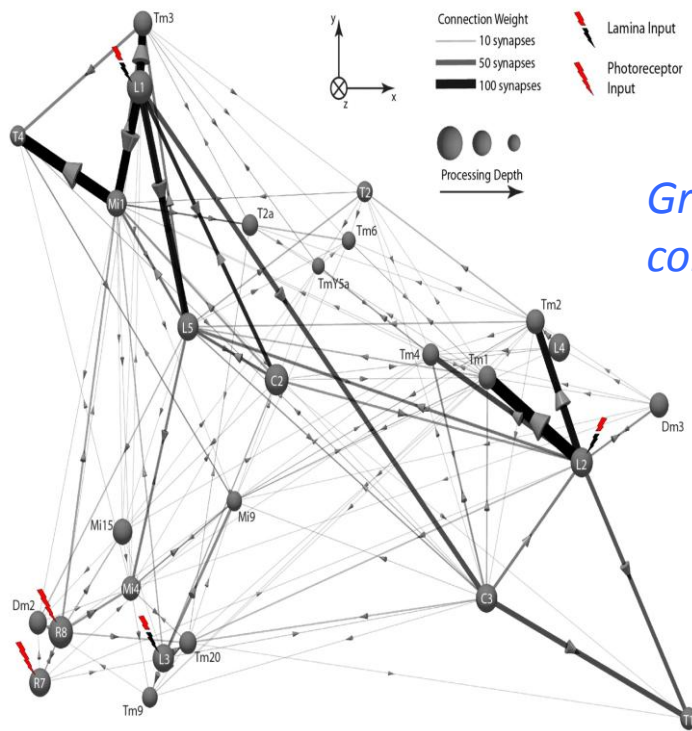
Outline

- EM connectomics in the Fly Brain
- Challenges (bottlenecks) in generating a connectome
- Segmenting and evaluating large data well at scale
- Collaborative segmentation-based tracing

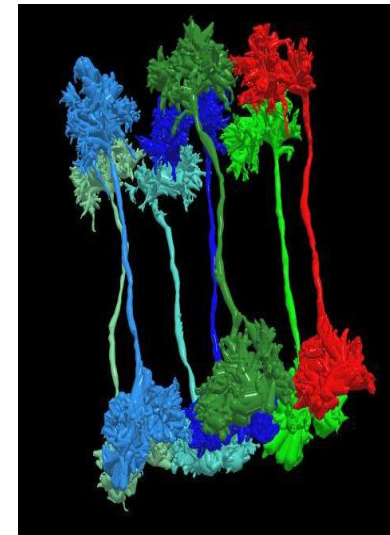
What is a Connectome?

- A list of neurons/nodes and how they are connected
- Does not necessarily indicate synapse sign, strength, delay, and other dynamics
- Scalability is a challenge (*main focus of talk*)

Detailed morphology →
electrical simulation

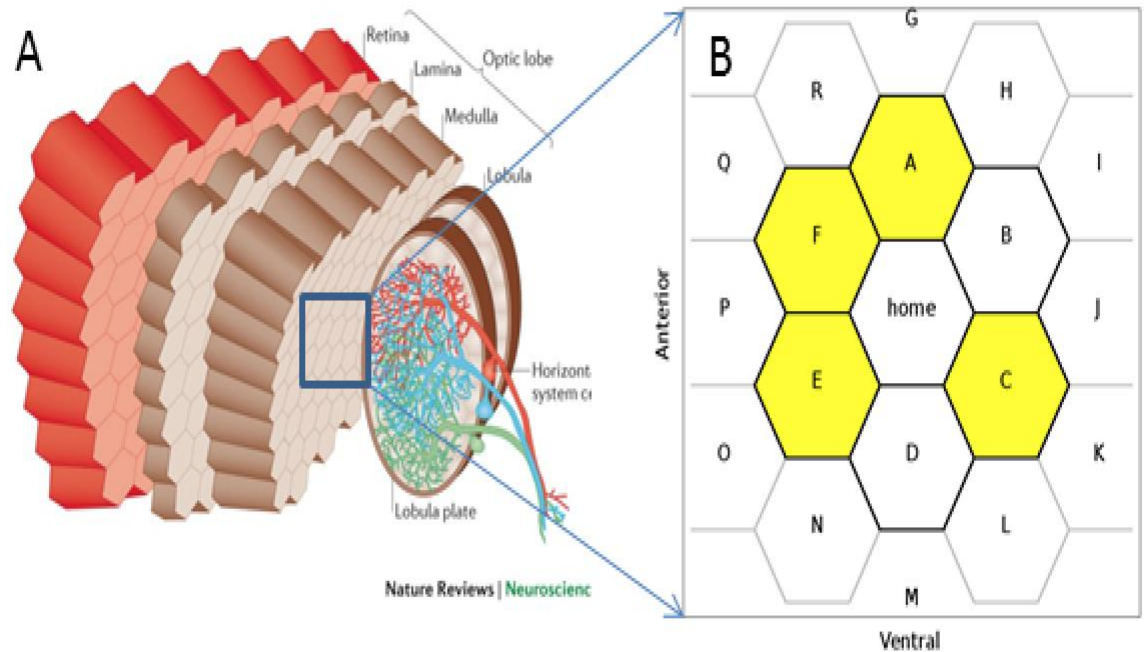
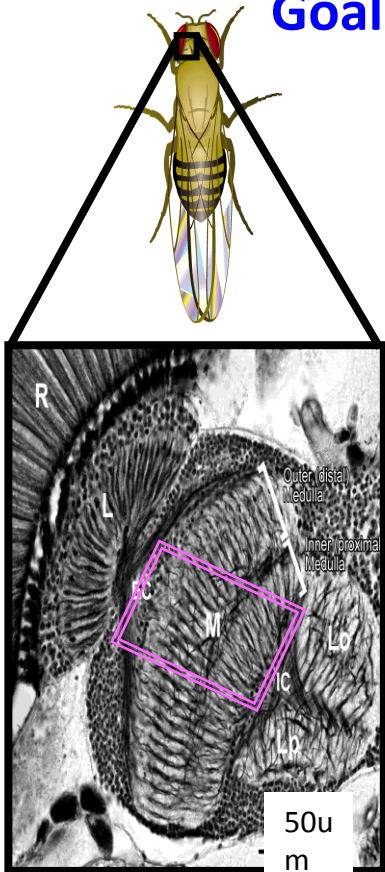


Graph with
connection strengths

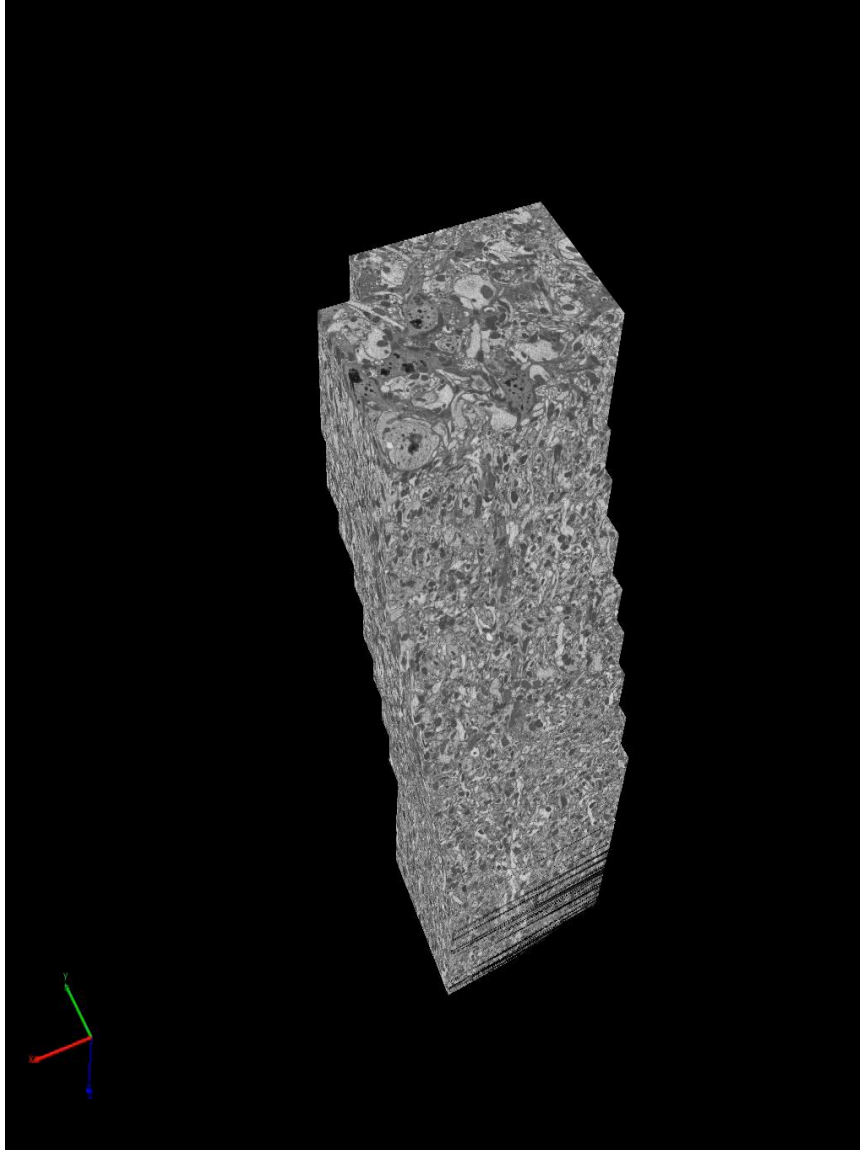


Example: Connectome in the Fly Optic Lobe

Goals: motion detection circuitry, wiring stereotypy in *medulla*



Optic Lobe Reconstruction (medulla)



Annotated 53,401 Tbars
and 315,421 PSDs (more
complete than previous
medulla reconstruction:
10,093 Tbars and 38,465
PSDs)

>3x faster than previous
reconstruction

~842 reconstructed cells

*Video courtesy
of Ting Zhao*

Outline

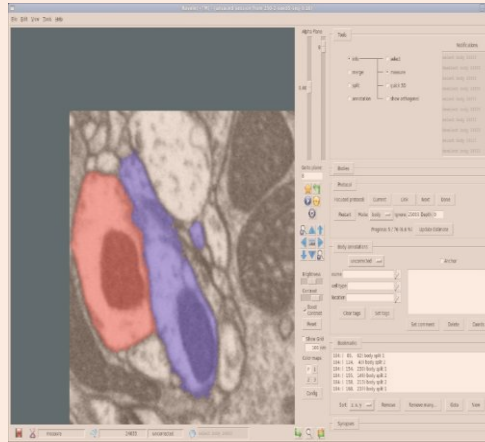
- EM connectomics in the Fly Brain
- Challenges (bottlenecks) in generating a connectome
- Segmenting and evaluating large data well at scale
- Collaborative segmentation-based tracing

Bottlenecks in Generating Connectomes



Imaging Challenges

- Years to image something like a mouse brain (even with latest advances)
- Fly brain is already 100 TB of data



Proofreading Dataset

- Extensive manual component
- Worse than imaging (e.g., 1 week of imaging → 1 year of proofreading)



Analysis Challenges

- ?? Analysis time
- Analogous to 'genome' – what to do with the data

Challenges to Speedup Proofreading

Segmentation (better segmentation → less work)

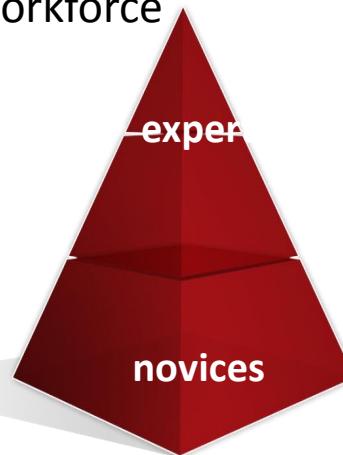
- Ambiguous parts of dataset
- Lack of contextual awareness (currently)

Proofreading is difficulty (scaling-out manual effort)



*more
experience
required*

Available Tracing
Workforce

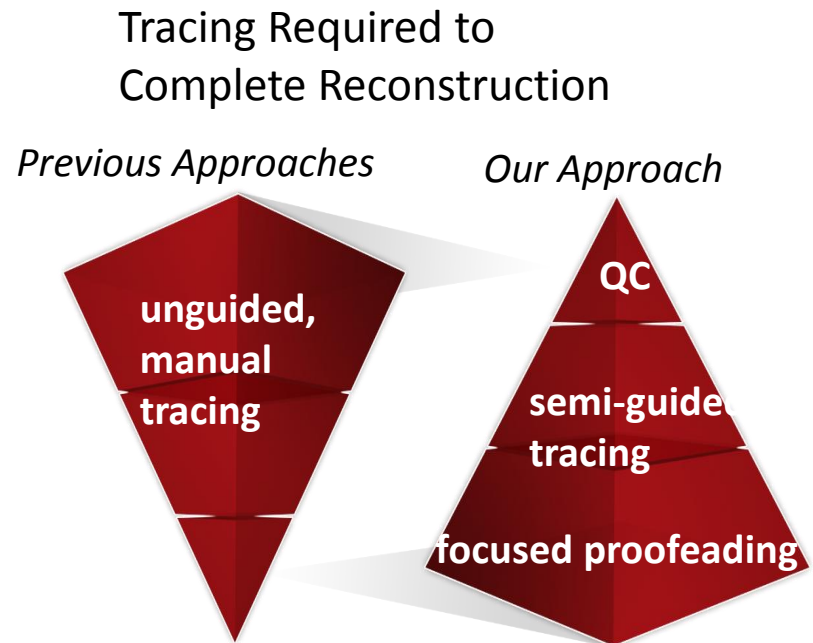


Tracing Required to
Complete
Reconstruction



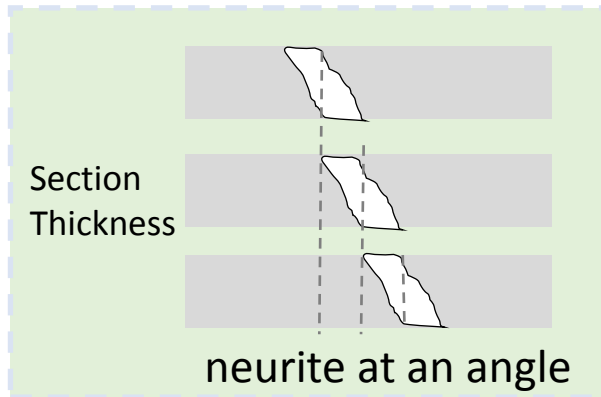
Our Solutions

- Use focused ion beam to produce high-resolution image → improved segmentation
- Automated synapse annotation
- Machine-guided proofreading



Imaging: Focused Ion Beam (FIB)

Section cutting → thick sections →
difficult to align/interpret



Use FIB instead



Polish with Ga Ion Beam (~2-10 nm)



Image with scanning EM

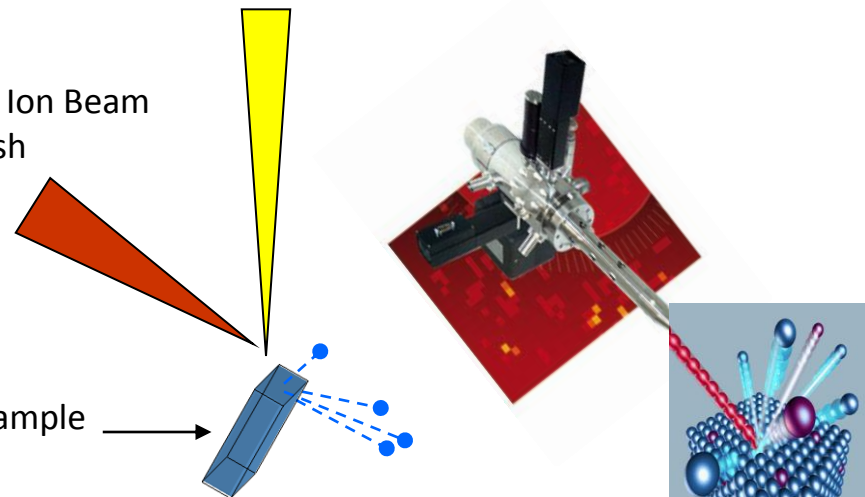
FIB Problems

1. Field-of-view
2. Speed
(but still faster than proofreading)

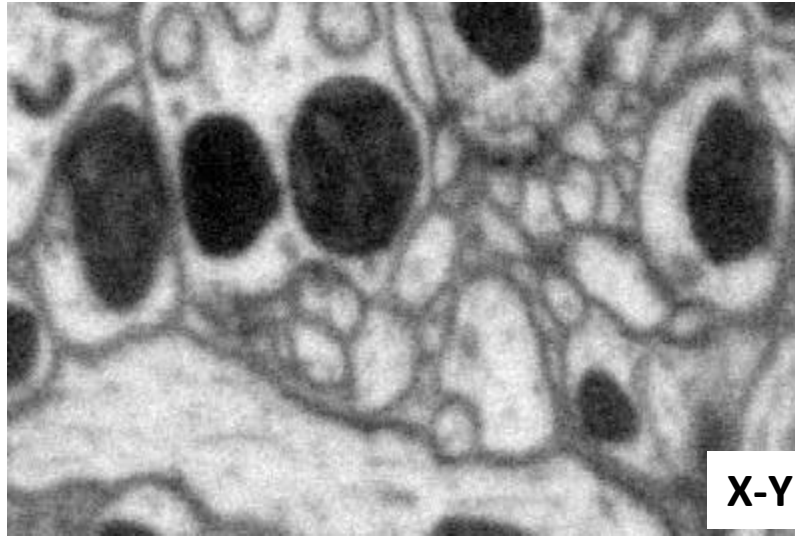
Focused Ion Beam
cut/polish

Sample

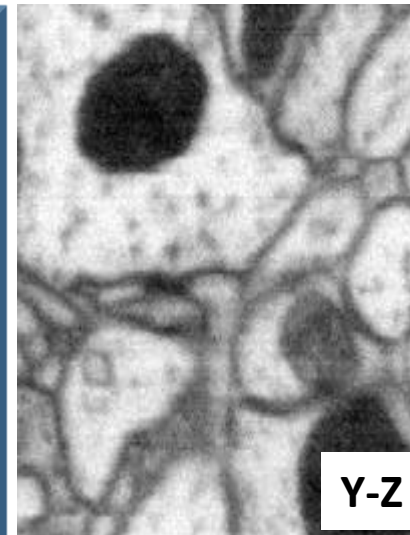
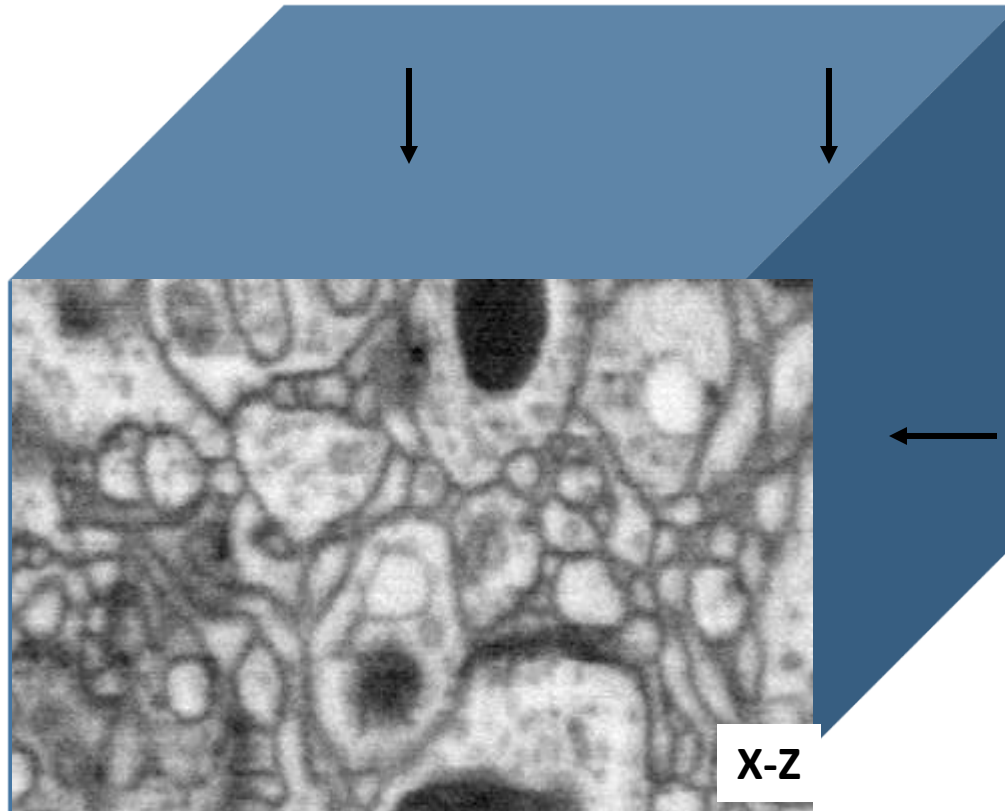
Scanning Electron Beam
Image



FIB-SEM

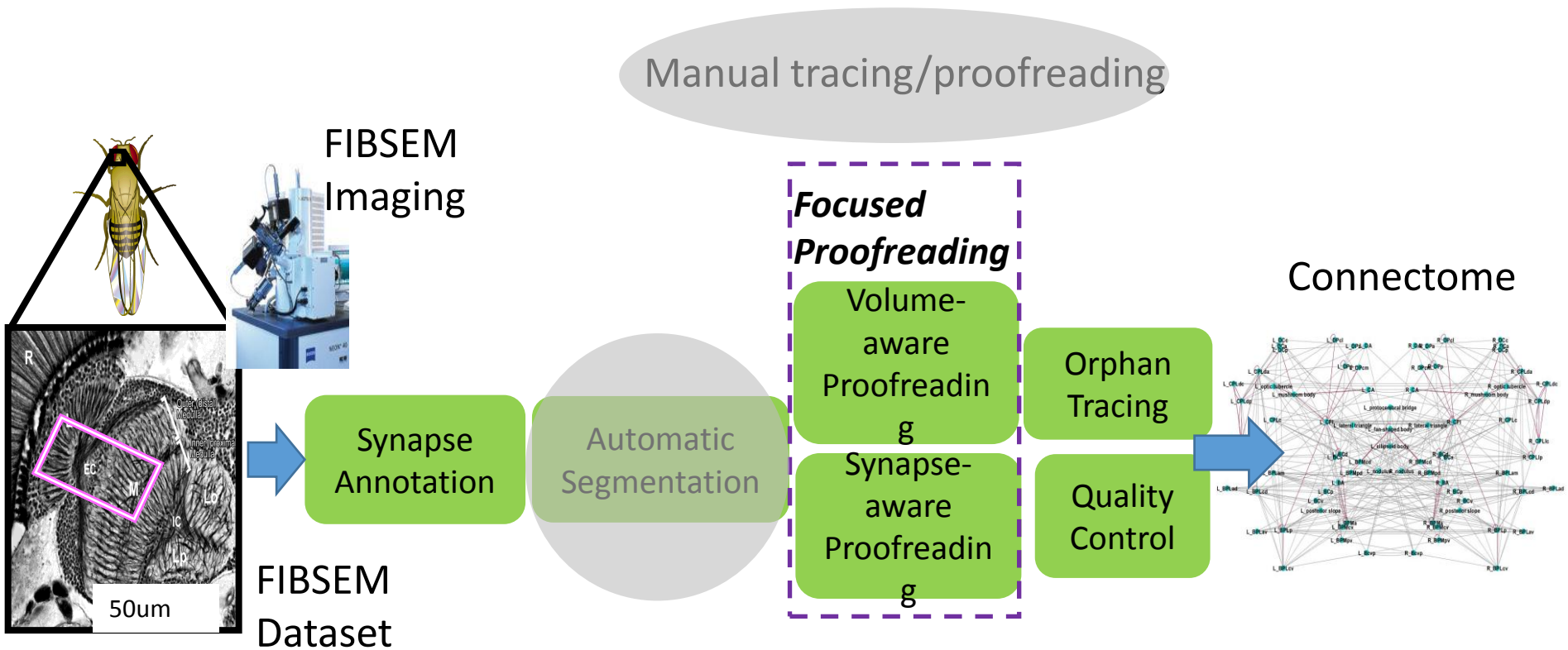


FIB-SEM
Isotropic
voxel



Credit:
Harald Hess Lab
(Shan Xu)

Our Reconstruction Pipeline



Outline

- EM connectomics in the Fly Brain
- Challenges (bottlenecks) in generating a connectome
- Segmenting and evaluating large data well at scale
- Collaborative segmentation-based tracing

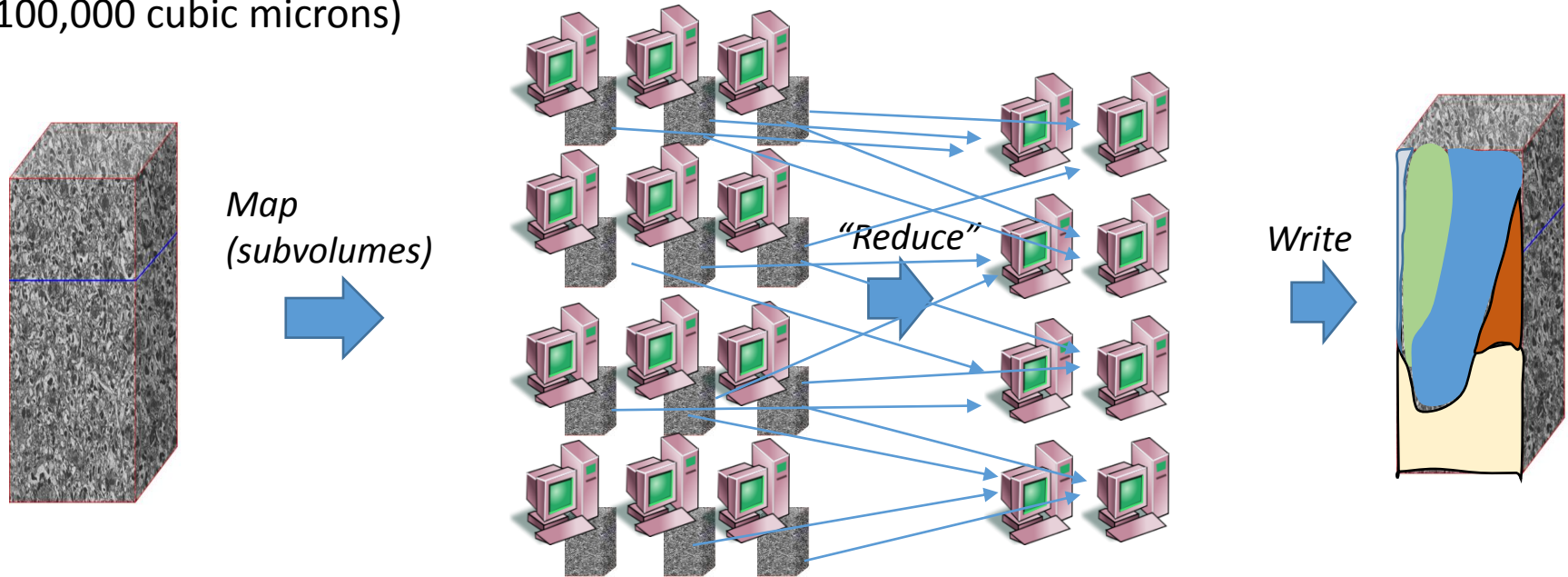
Problem: Segmenting Large Datasets Well

Dataset
(e.g., >200 GB-2TB
>100,000 cubic microns)

Boundary prediction,
watershed, agglomeration

Stitch local volumes
(consistent labeling)

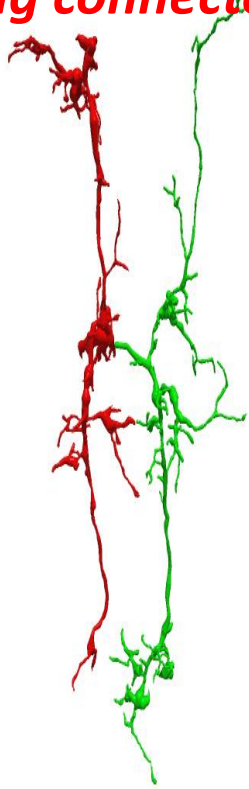
Commit
segmentation



- Mostly local computation
- Pretty scalable (not compute limited currently)
- Long range segmentation sometimes error

Global Segmentation: Susceptibility to Small Inaccuracies

***Segmentation: accuracy is critical
(small errors can cause
big connectome changes)***

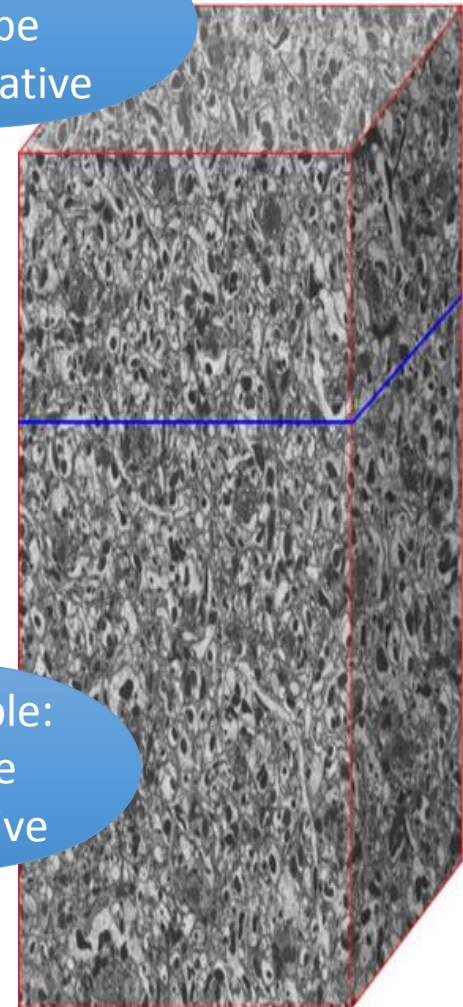


I'm difficult:
let's be
conservative

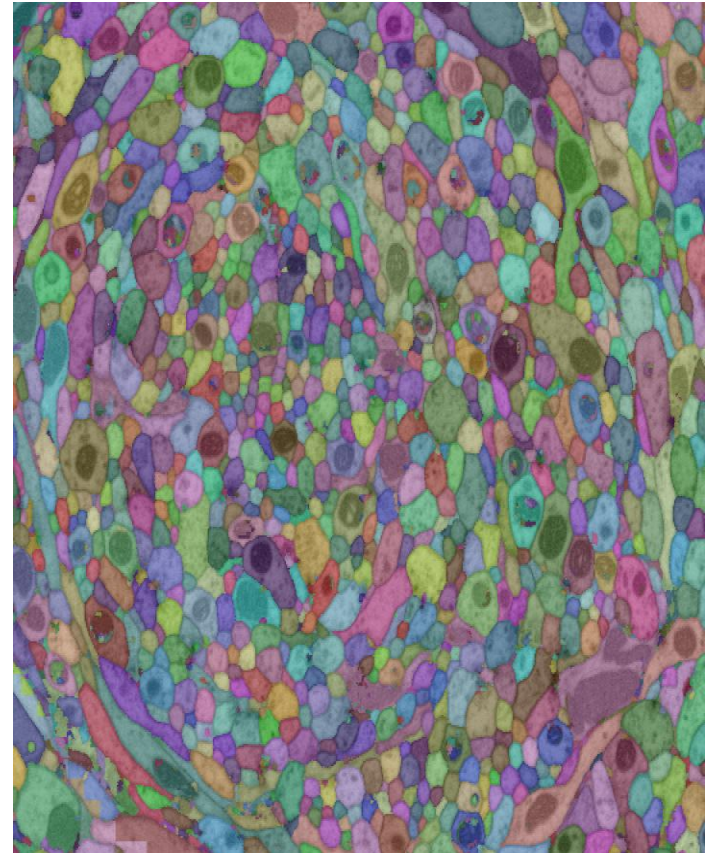
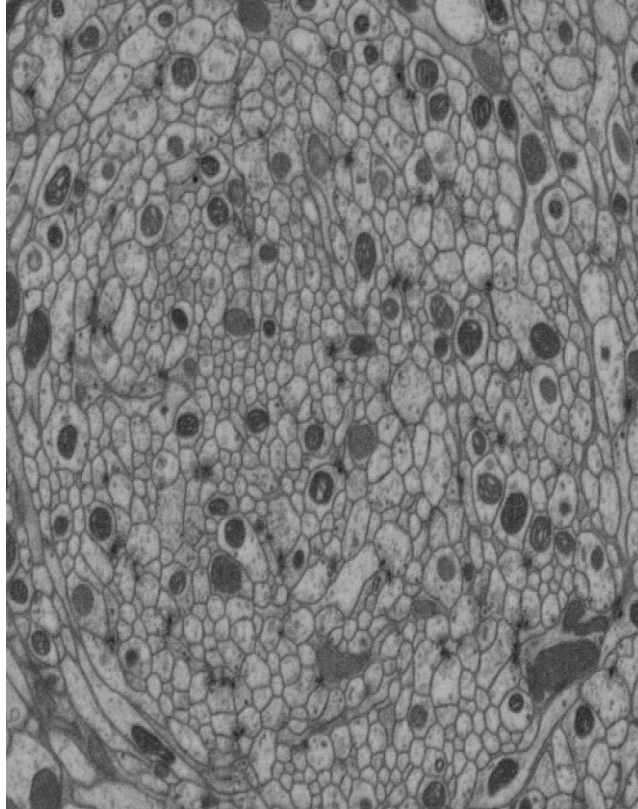


***Different regions ->
different classifiers***

I'm simple:
let's be
aggressive

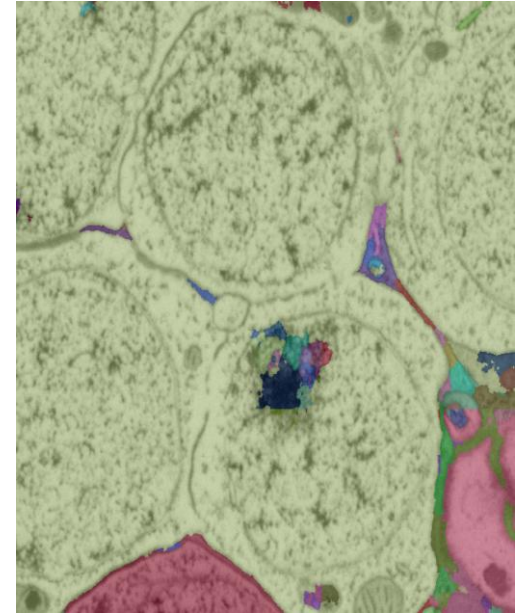
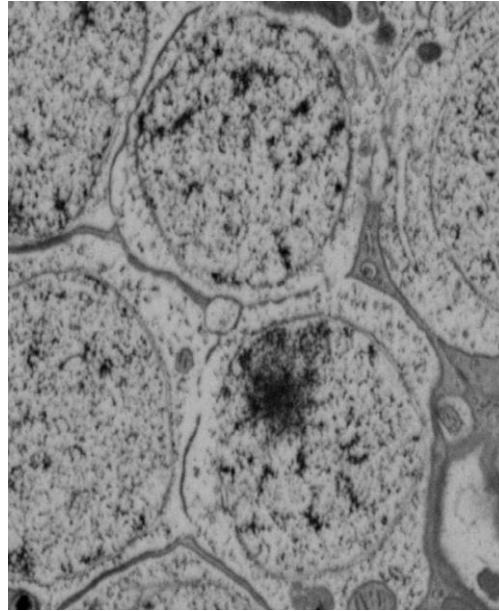


Examples: The Good

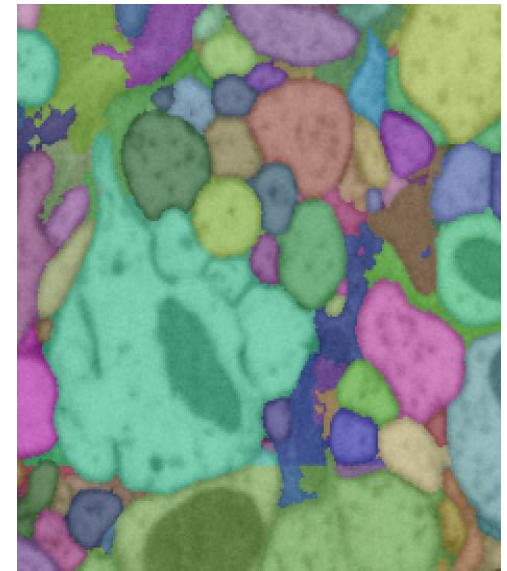
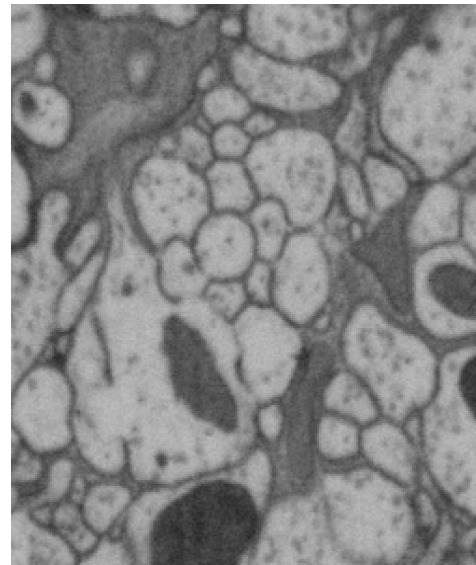


Segmentation: The Bad

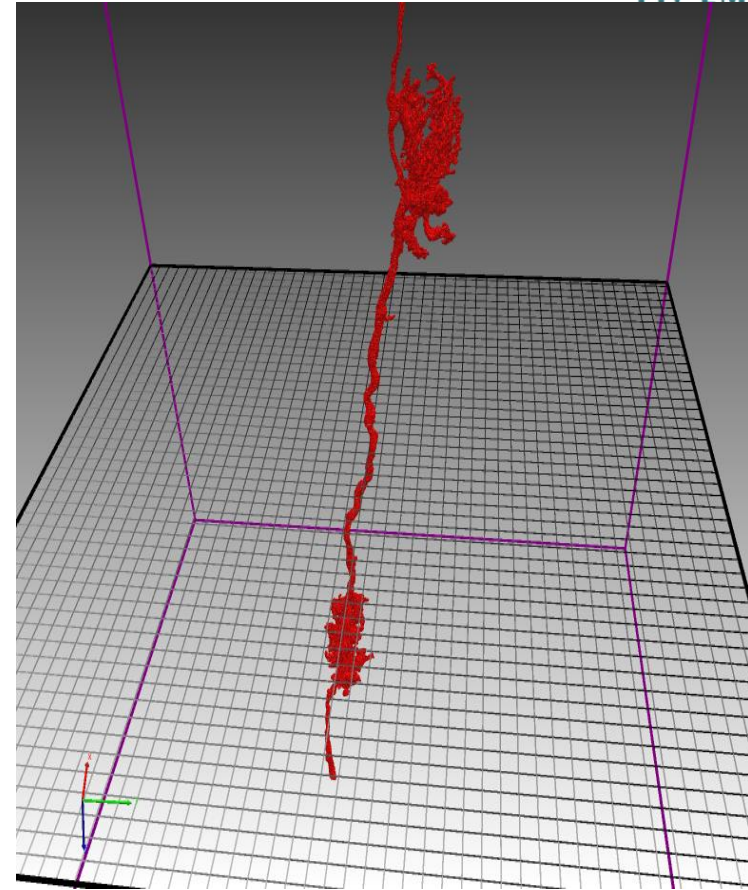
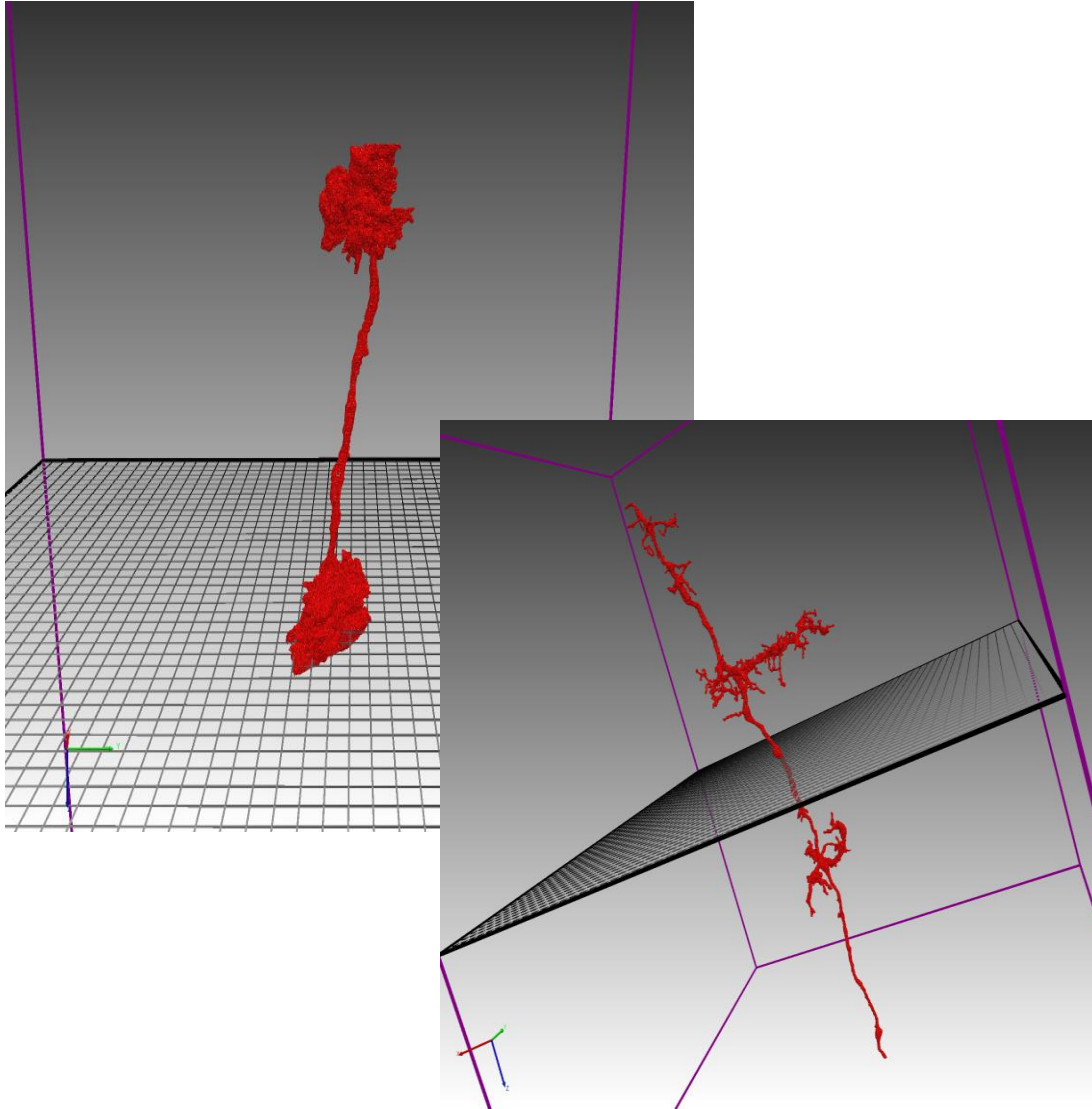
Poor classifier generalizability
(soma wasn't considered)



Artifacts (e.g., membrane holes)



Careful Stitching => Better Automatic Reconstruction



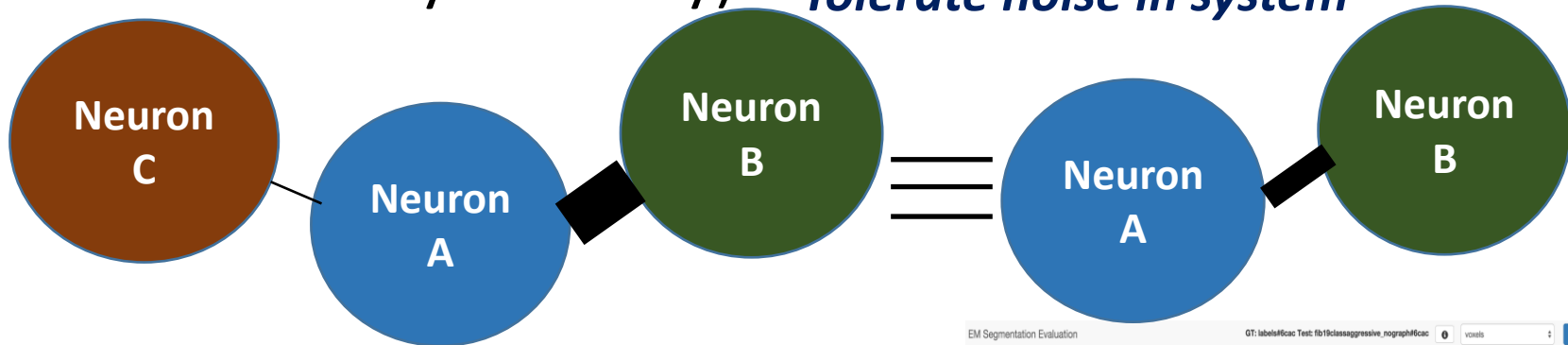
- How do we know that the segmentation is good?
- Typically compare two segmentations (one often 'ground-truth') to achieve some similarity score
- How to facilitate comparisons on very large datasets? (where does such ground truth come from? FlyEM)

Solutions/Advances

- Spark-based system to assess segmentation on large dataset

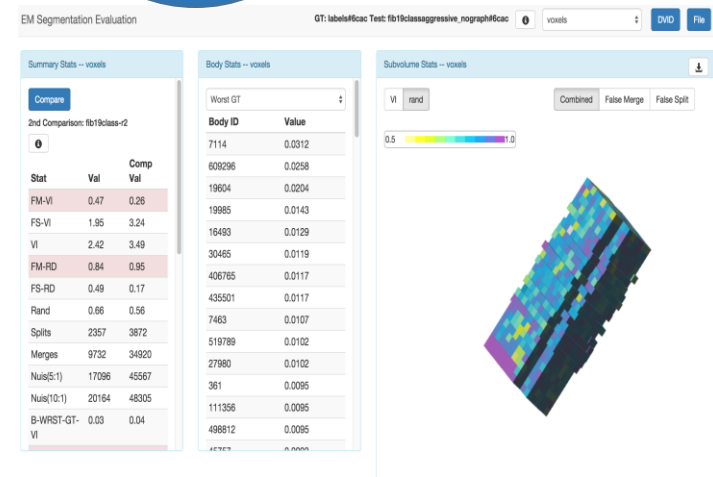
<https://github.com/janelia-flyem/DVIDSparkServices>

- Generate several metrics (e.g., edit distance, connectivity accuracy) *Tolerate noise in system*



- Web application to conveniently analyze results

<https://github.com/janelia-flyem/SegmentationEvaluationConsole>

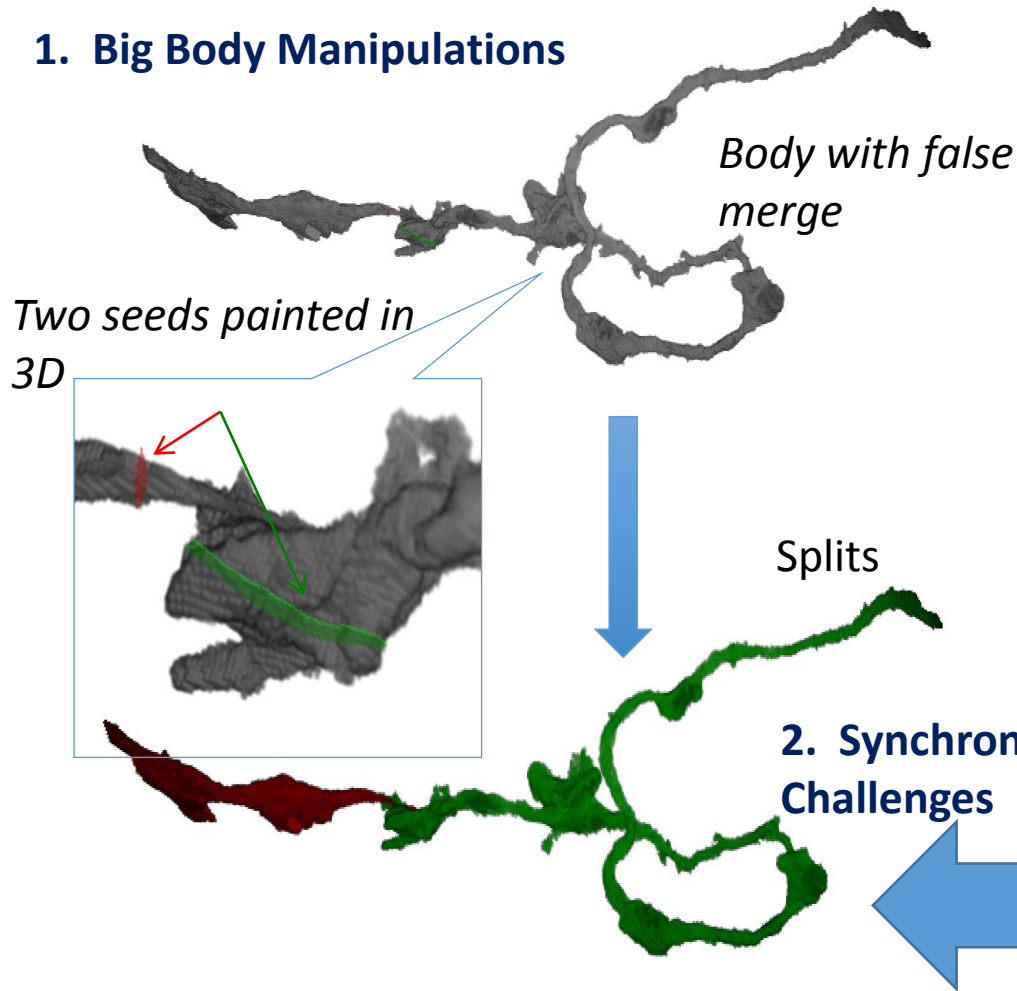


Outline

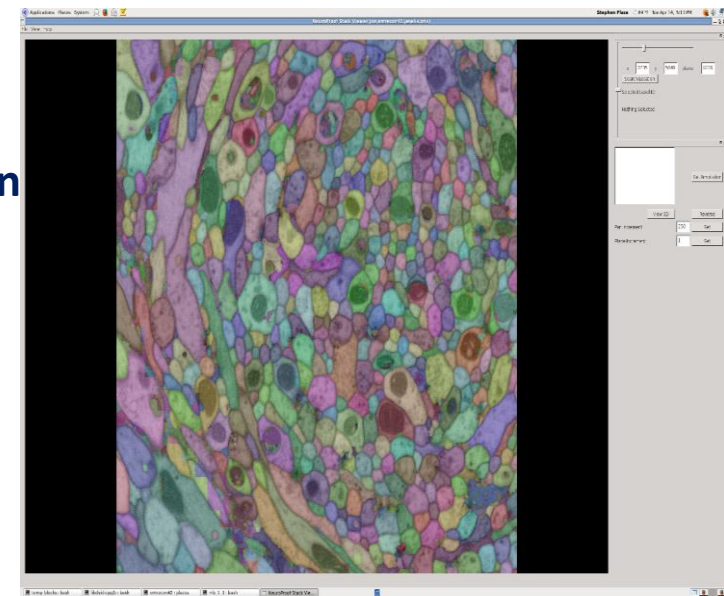
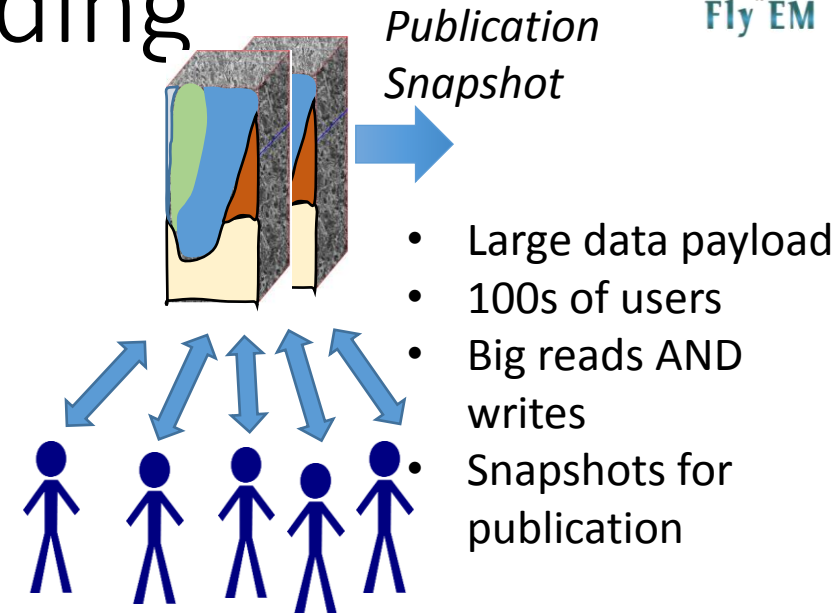
- EM connectomics in the Fly Brain
- Challenges (bottlenecks) in generating a connectome
- Segmenting and evaluating large data well at scale
- Collaborative segmentation-based tracing

Problem: Large-scale Interactive Proofreading

1. Big Body Manipulations



2. Synchronization Challenges

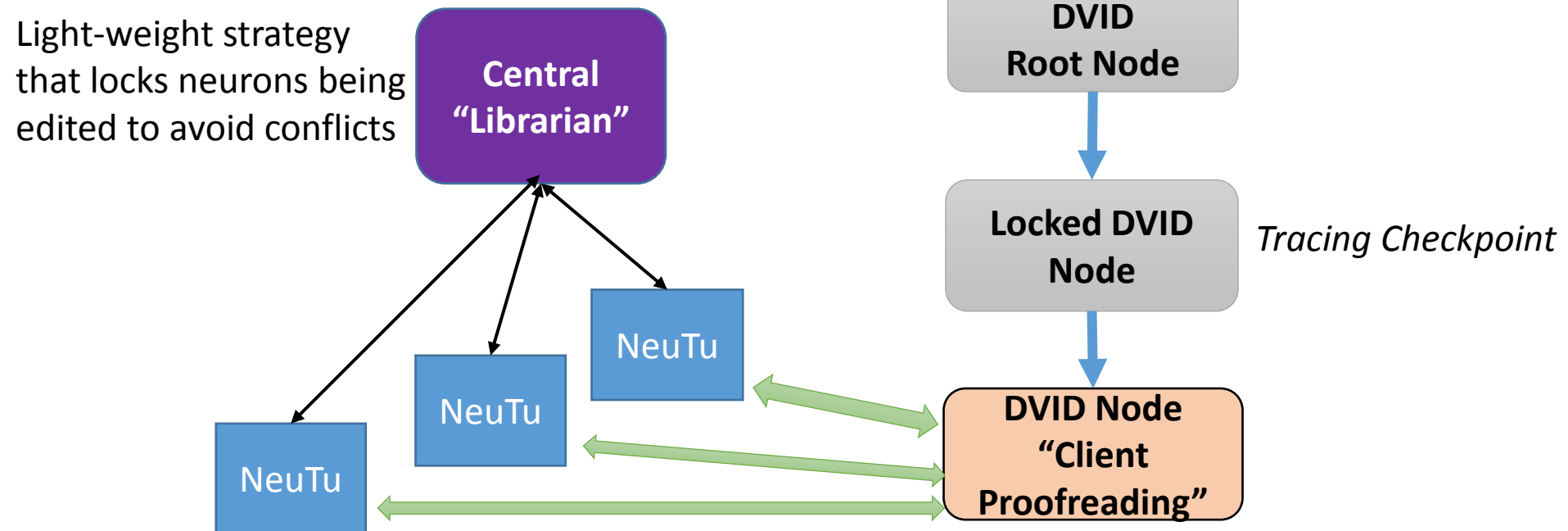


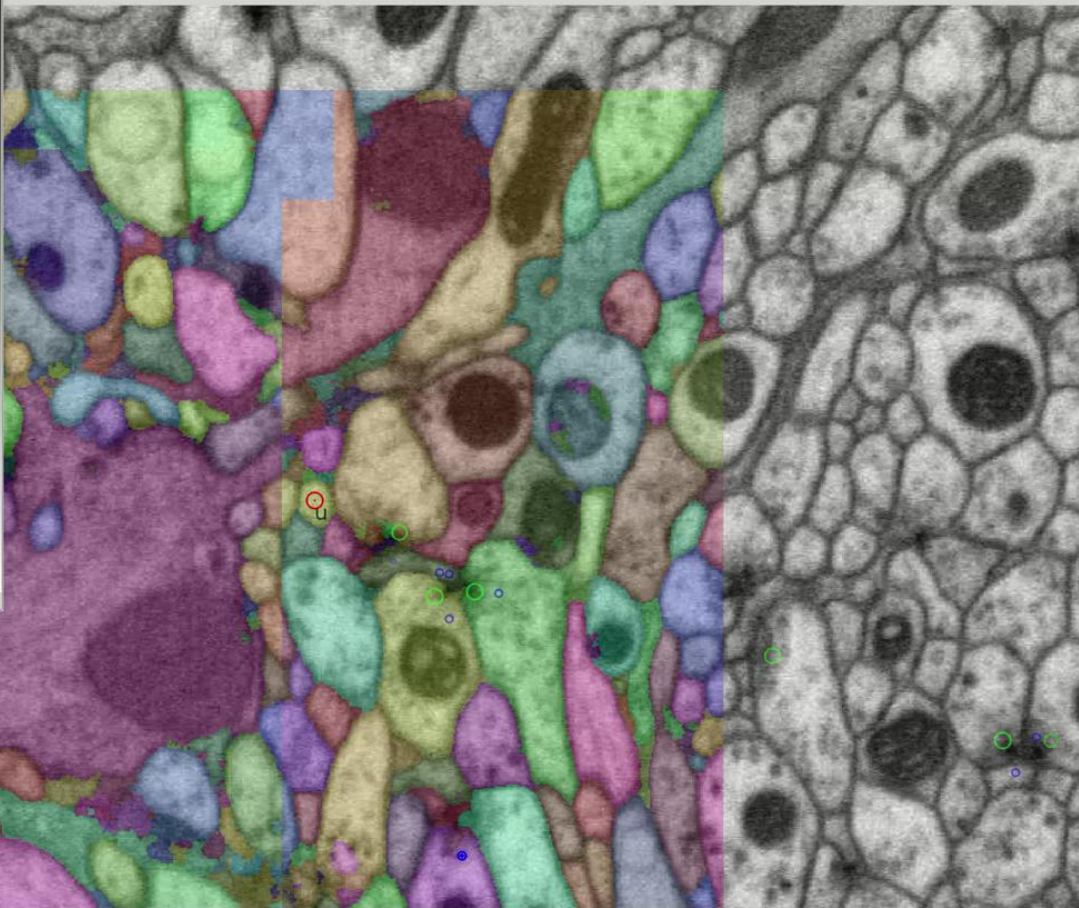
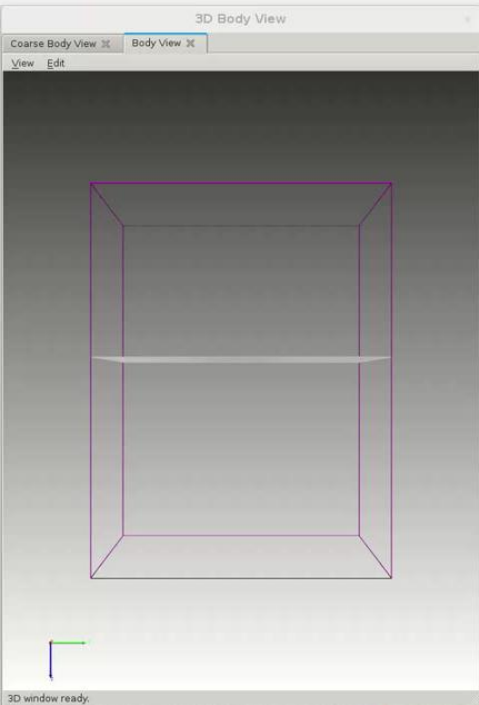
Solution: Segmentation-driven Proofreading using DVID

- NeuTu: Proofreading client that facilitates fast merge and split operations
- Collaborative tracing (changes viewable by other tracers)
- Integration with DVID



NeuTu: Ting Zhao





Select Database Menu

```
{
  "name": "MB_Practice",
  "address": "emdata1.int.janelia.org",
  "comment": "MB seg for practice(branched from 66ba)",
  "port": 8500,
  "background": 255,
  "uid": "eafc",
  "body_label": "bodies3",
  "label_block": "labels3"
}
```

Coarse Body View Body View Launch Split

Segmentation Size: + Full View

Merge Upload Result

Assigned Bookmark User Bookmark

	type	X	Y	Z	Bod
1	Other	3284	5531	6432	164401
2	Other	3305	6096	5566	156369
3	Other	4274	7737	3194	698648
4	Other	4658	6985	3904	673671
5	Other	4330	3375	9868	149827
6	Other	4720	6986	4868	505123
7	Other	4846	6970	4905	102939
8	Split	4431	6456	1721	229250
9	Merge	4422	6437	1721	229250
10	Other	3265	5537	6485	142247
11	Other	4668	6971	4841	346030
12	Other	5306	6418	5146	200004
13	Other	4133	4838	6594	122446
14	Other	3594	6665	5215	535284
15	Other	3302	5743	6073	535284
16	Other	4790	6936	4862	876448
17	Other	4989	6763	4897	200004
18	Other	5105	5413	7024	918181
19	Other	3882	6892	5280	950521
20	Other	3883	5746	8233	128180
21	Other	4526	6488	6862	121910
22	Other	3659	3917	8701	136413
23	Split	2921	7328	2465	982128
24	Other	4928	6479	4847	0
25	Other	4838	6762	3740	759860
26	Other	4407	4901	6516	351301
27	Other	3551	4933	6576	0

Message

3513984 selected.
No body selected.
3513984 selected.
Syncing 3D Body view ...
3D Body view updated.
No body selected.
Syncing 3D Body view ...
3D Body view updated.

Closing Thoughts

- Connectomics is a burgeoning field with great promise
- Scaling-up connectomics (*i.e., speeding up proofreading*) is a great challenge
(can computation come to rescue?)
- Image segmentation needs to go from small-scale to large scale => new challenges
- Evaluation metrics must reflect the domain
- DVID + NeuTu enables a collaborative segmentation-driven proofreading solution

Acknowledgements

Fly EM Project

Team

Shinya Takemura
Bill Katz
Arjun Bharioke
Kazunori Shinomiya
Don Olbris
Lowell Umayam
Jody Clements
Charlotte Weaver
Rob Svirskas
Ting Zhao
Lou Scheffer
Gary Huang
Harald Hess (Hess Lab)
Toufiq Parag
Zhiyuan Lu
Shan Xu (Hess Lab)
Kenneth Hayworth
(Hess Lab)
Rick Fetter
Stuart Berg
Pat Rivlin

Fly EM Proofreader

Team

Satoko Takemura
Chris Sigmund
Lei-Ann Chang
Julie Tran
Omotara Ogundeyi
Roxanne Aniceto
Shirely Lauchie
Chris Ordish
Aya Shinomiya

Collaborating Labs

Ian Meinertzhagen (Dalhousie/Janelia
Senior Fellow)
Jane Anne Horne (Dalhousie)

Fred Hamprecht (Heidelberg)

Dmitri Chklovskii (Simons Foundation)