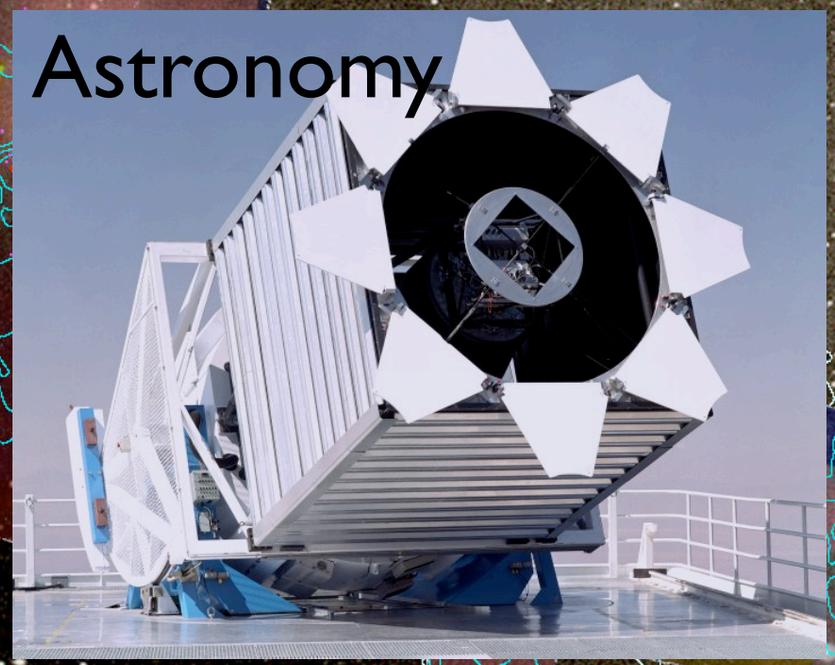
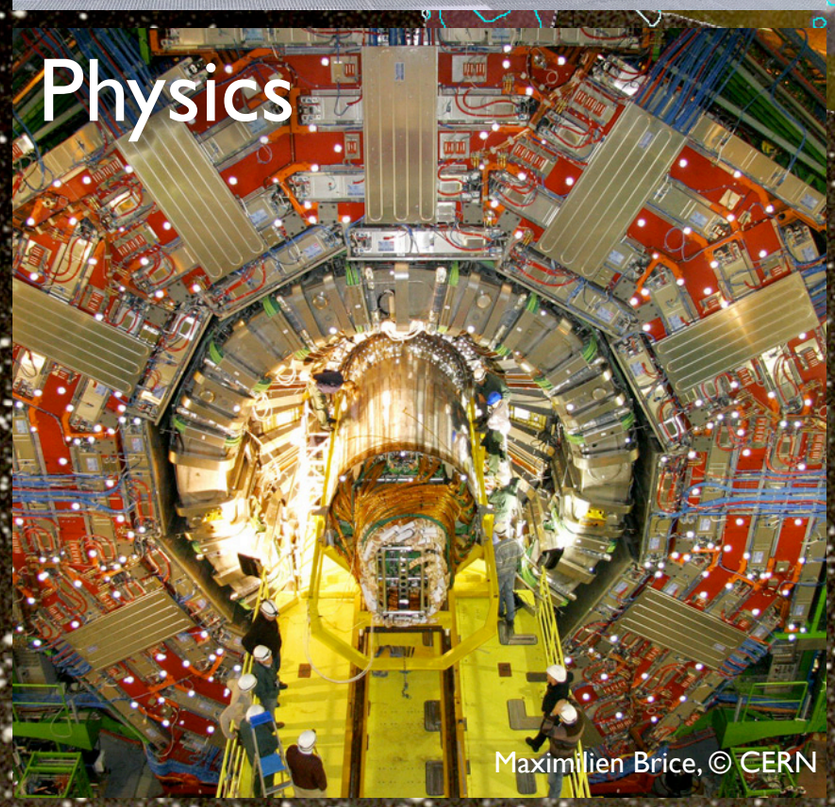
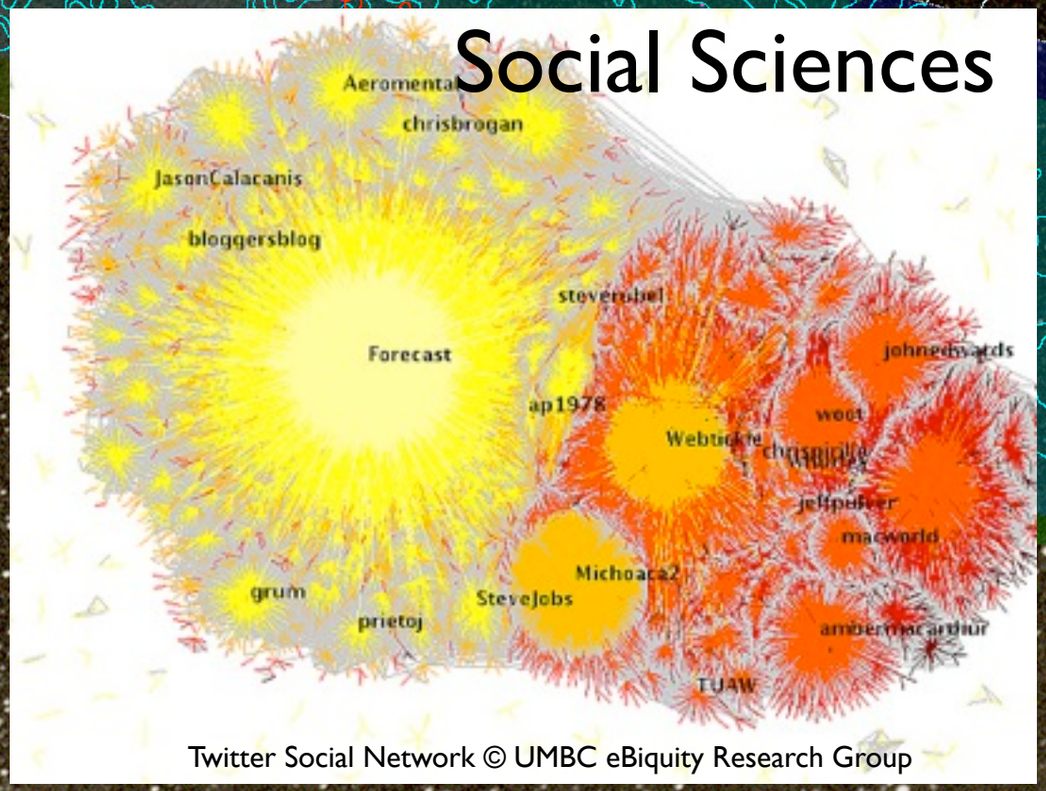
High-Throughput Science

Hanspeter Pfister, Harvard University
pfister@seas.harvard.edu

Scientific Data Explosion







Twitter Social Network © UMBC eBiquity Research Group

Computing Revolution

- Commodity energy-efficient supercomputers

4 Teraflops @ 1400 W

Challenge

- Limited by communication, not processor performance [Jim Grey, Computing Economics]

the INQUIRER
News, reviews, facts and friction

Search

Saturday, 12 September 2009 | INQ Mobile | RSS

Pigeon protocol offers faster data delivery

The pigeon is mightier than ADSL

By **David Neal**

Thursday, 10 September 2009, 14:15

AN UNORTHODOX RACE in South Africa has revealed that it is faster to transmit 4GB of data via carrier pigeon than it is to send it over the country's main ADSL services.

Traditional

Input

Modeling
Simulation
Analysis

Traditional

Input

Modeling
Simulation
Analysis

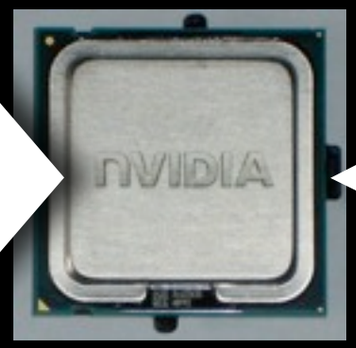
Data Reduction

Data
Input
Reduction

Modeling
Simulation
Analysis

High-Throughput

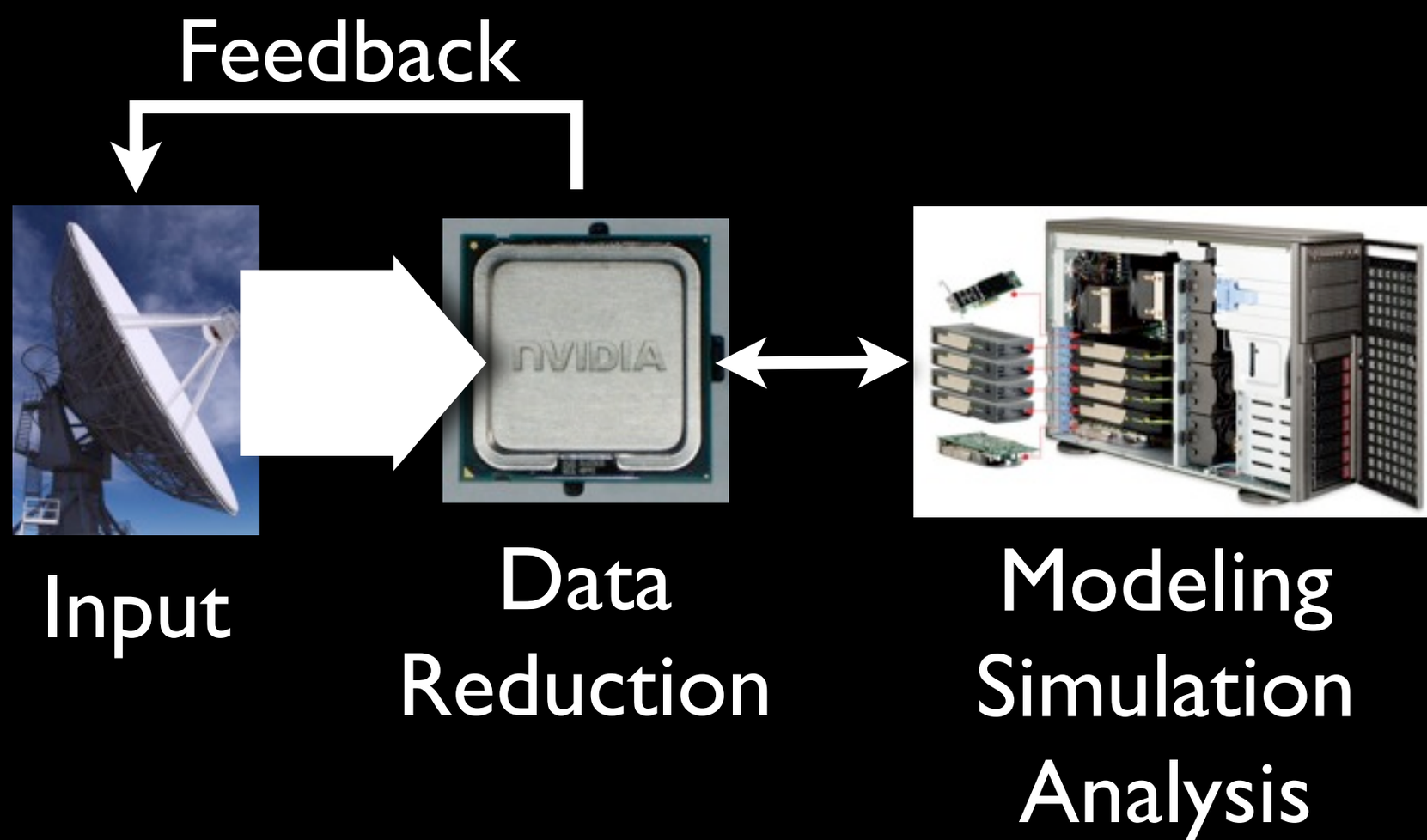
Input



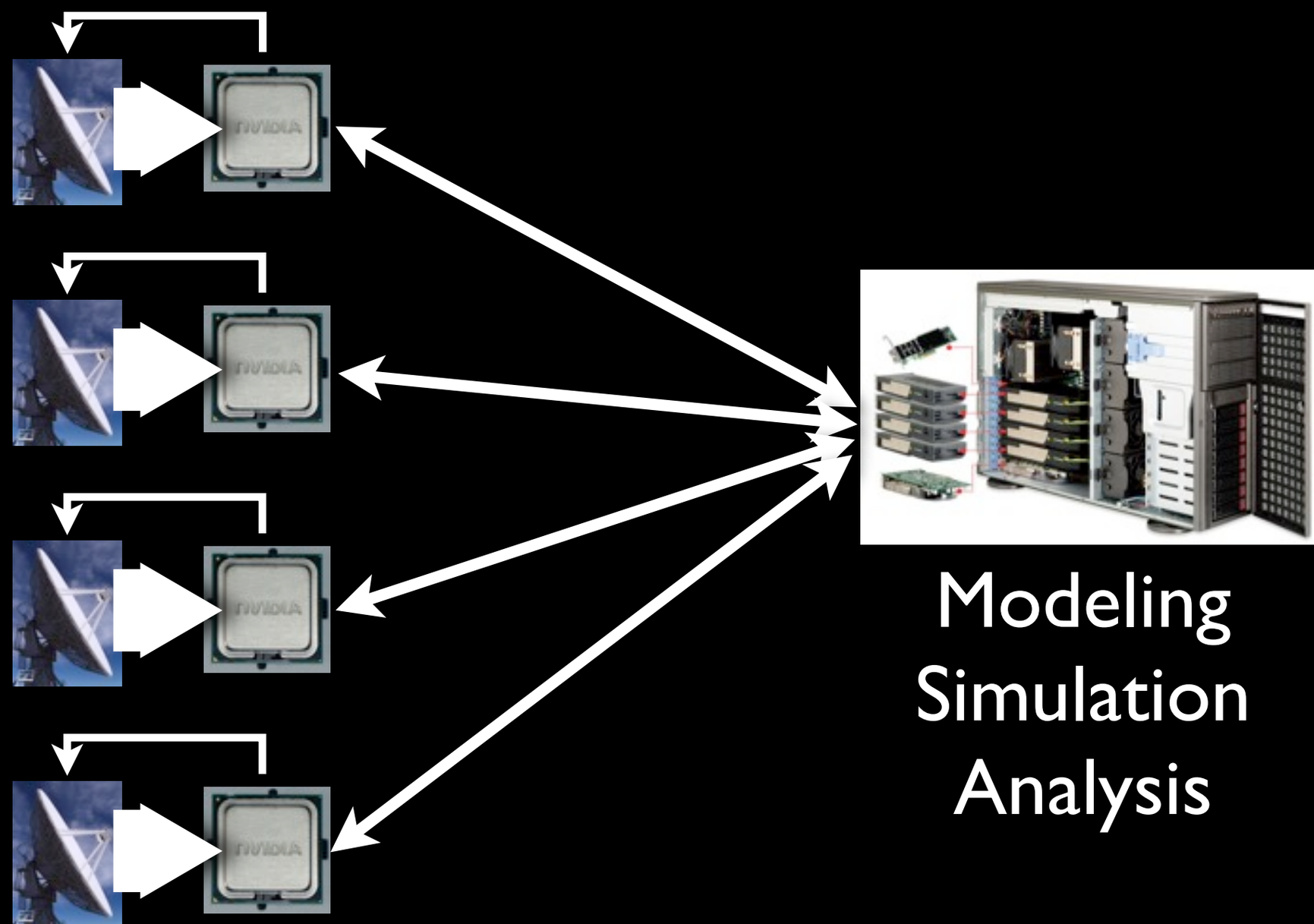
Data
Reduction

Modeling
Simulation
Analysis

Feedback

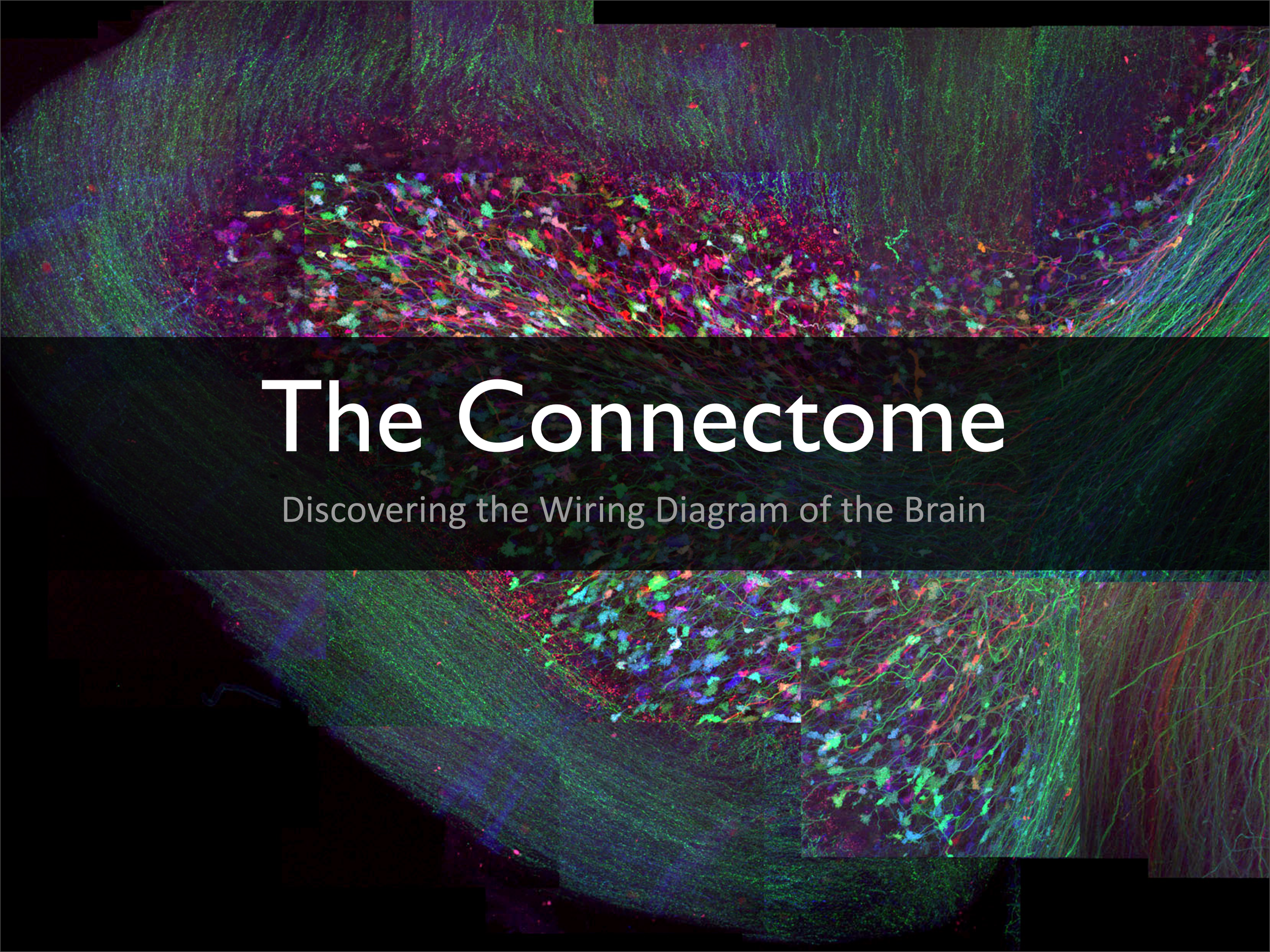

Feedback
NVIDIA
Input
Data
Reduction
Modeling
Simulation
Analysis

Scalability



Outline

- High-throughput Neuroscience
- High-throughput Astrophysics
- High-throughput Simulations

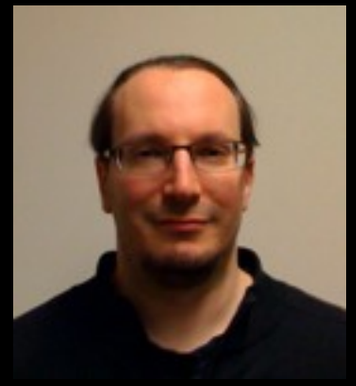


The Connectome

Discovering the Wiring Diagram of the Brain

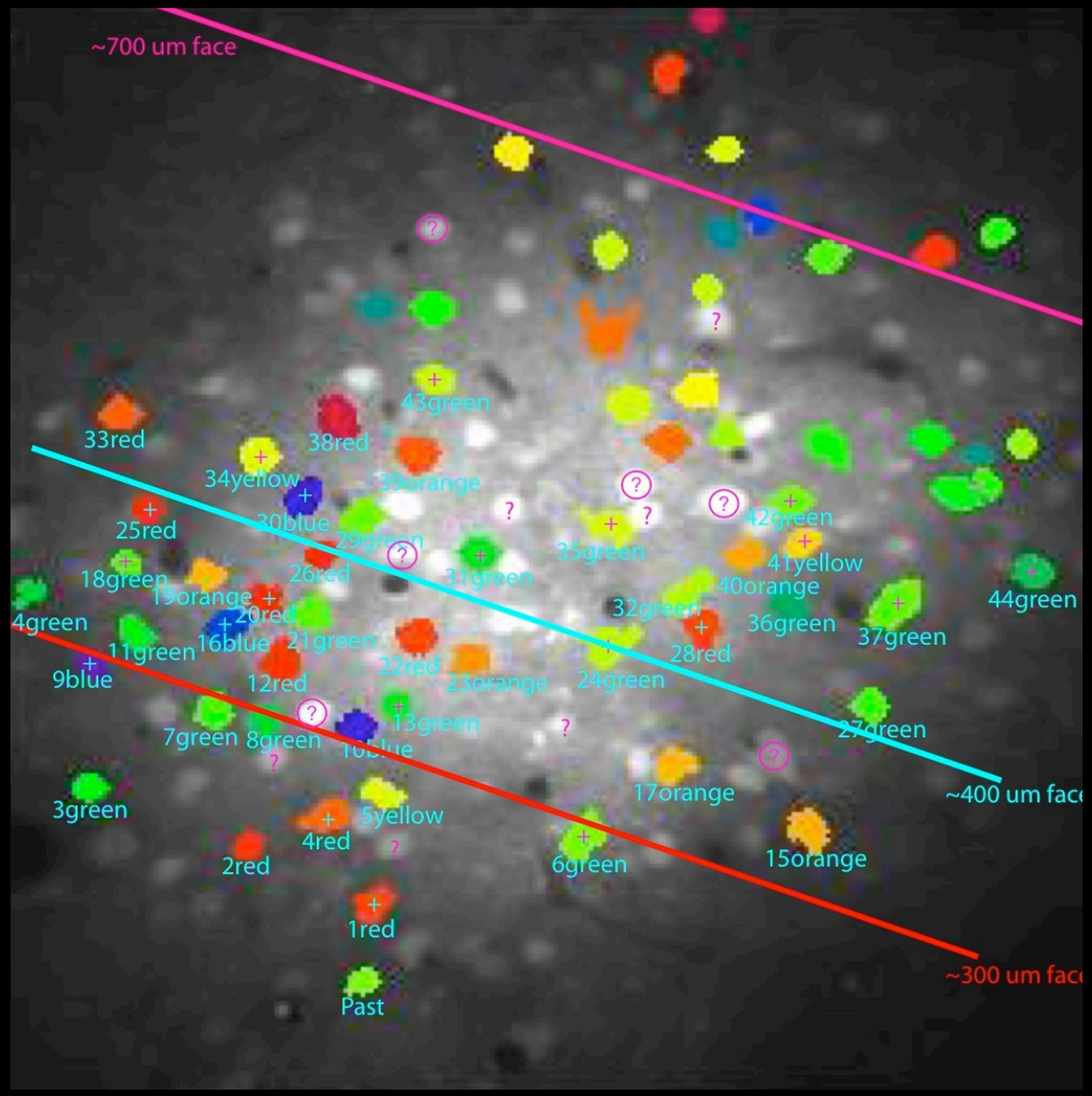
Collaborators

- Harvard Center for Brain Science
  - Jeff Lichtman & Clay Reid
- Kitware Inc.
  - Will Schroeder, Charles Law, Rusty Blue
- VRVis Vienna
  - Markus Hadwiger, Johanna Beyer
- SEAS
  - Amelio Vazquez, Won-Ki Jeong
  - Hanspeter Pfister



The Scientific Challenge

- What do large brain circuits do?

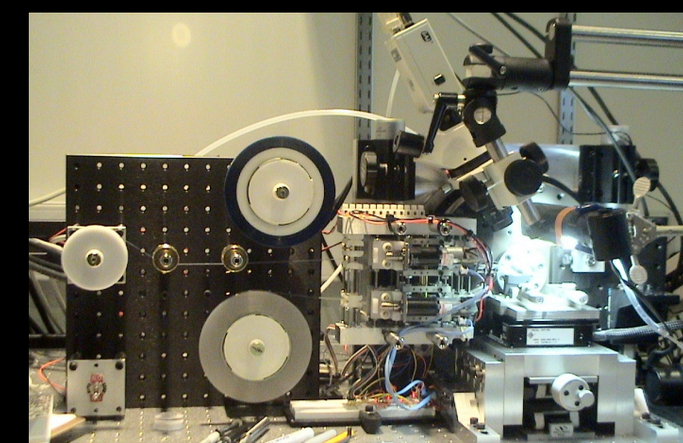


The Scientific Challenge

- What do large brain circuits do?
- What is their structure?

Ramón y Cajal, 1905

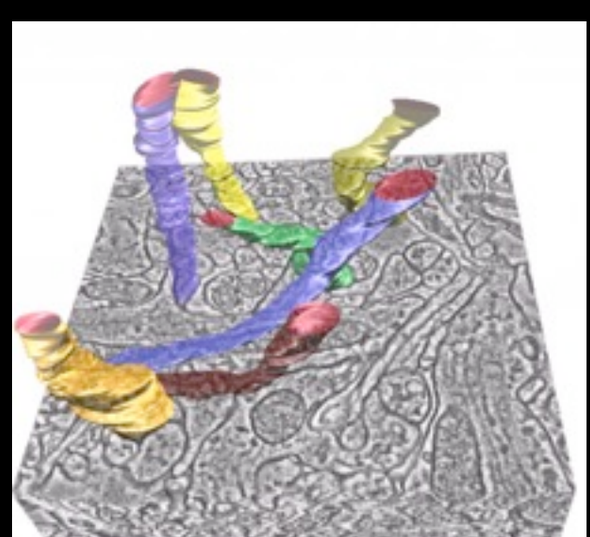
Connectome Pipeline

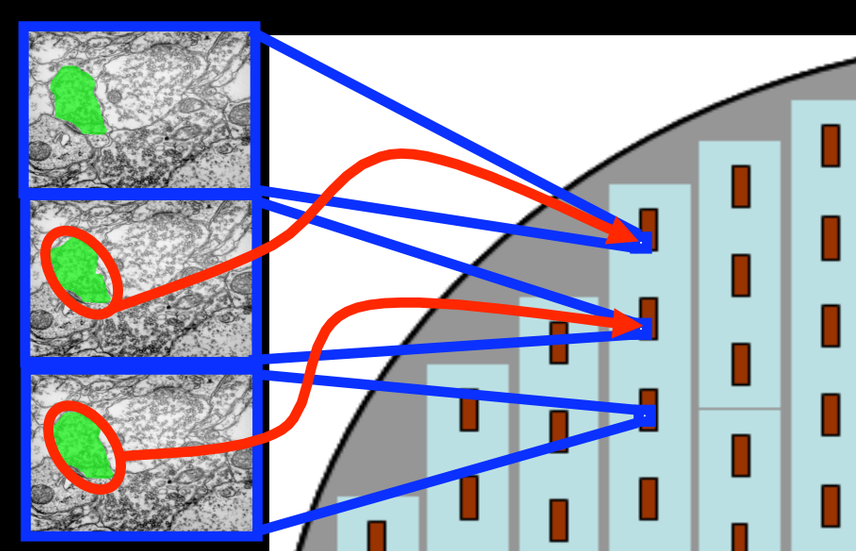


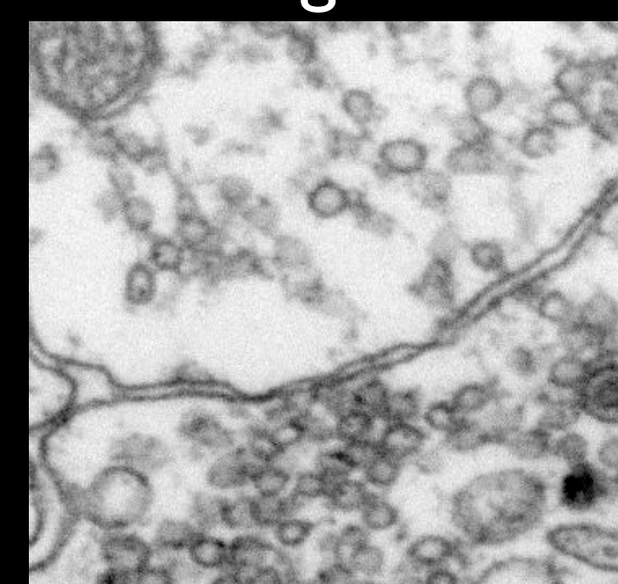
Get a brain → Get a piece → Cut it thin

↓

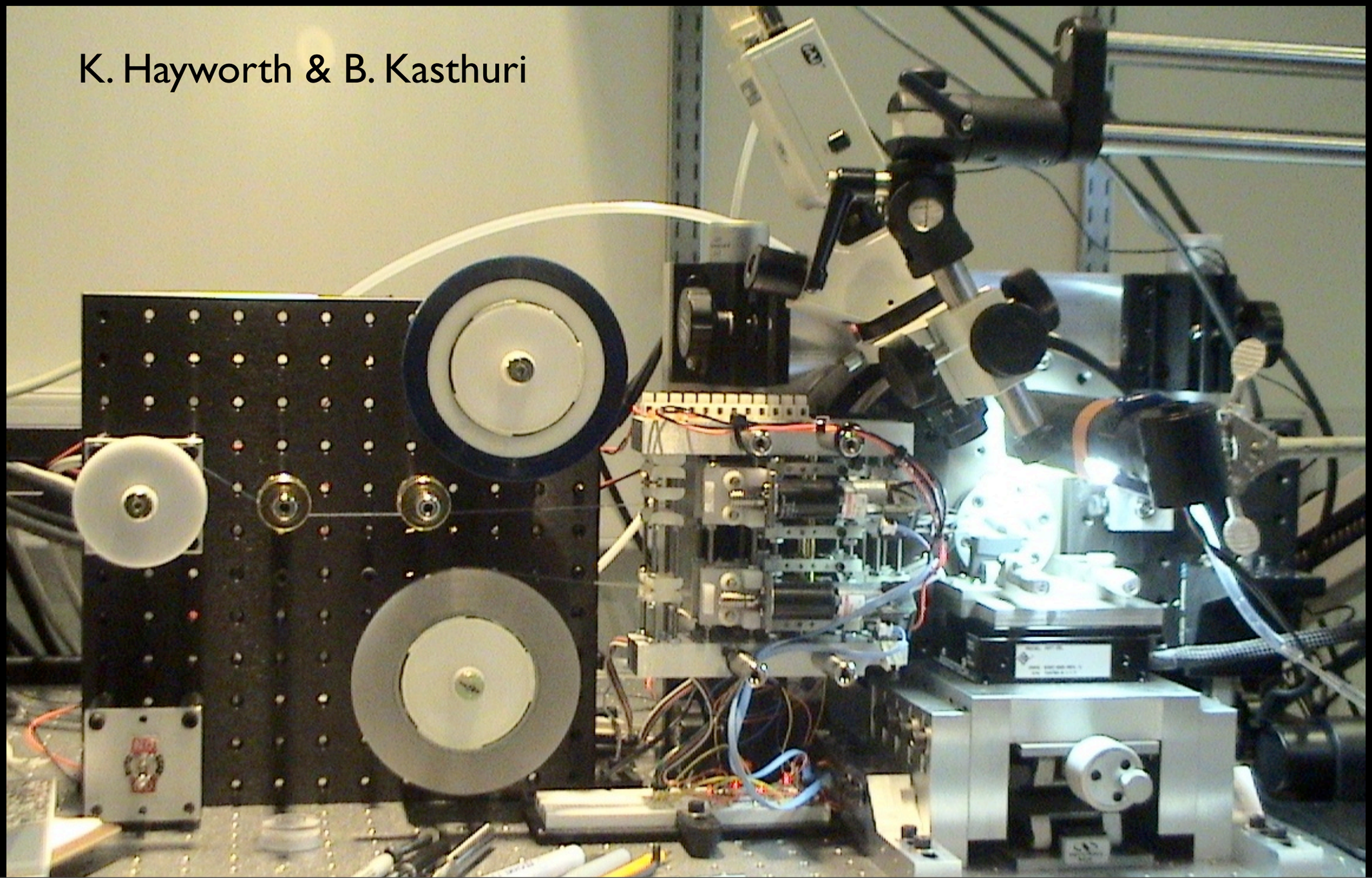
Visualize & Analyze ← Register and Segment ← Image it



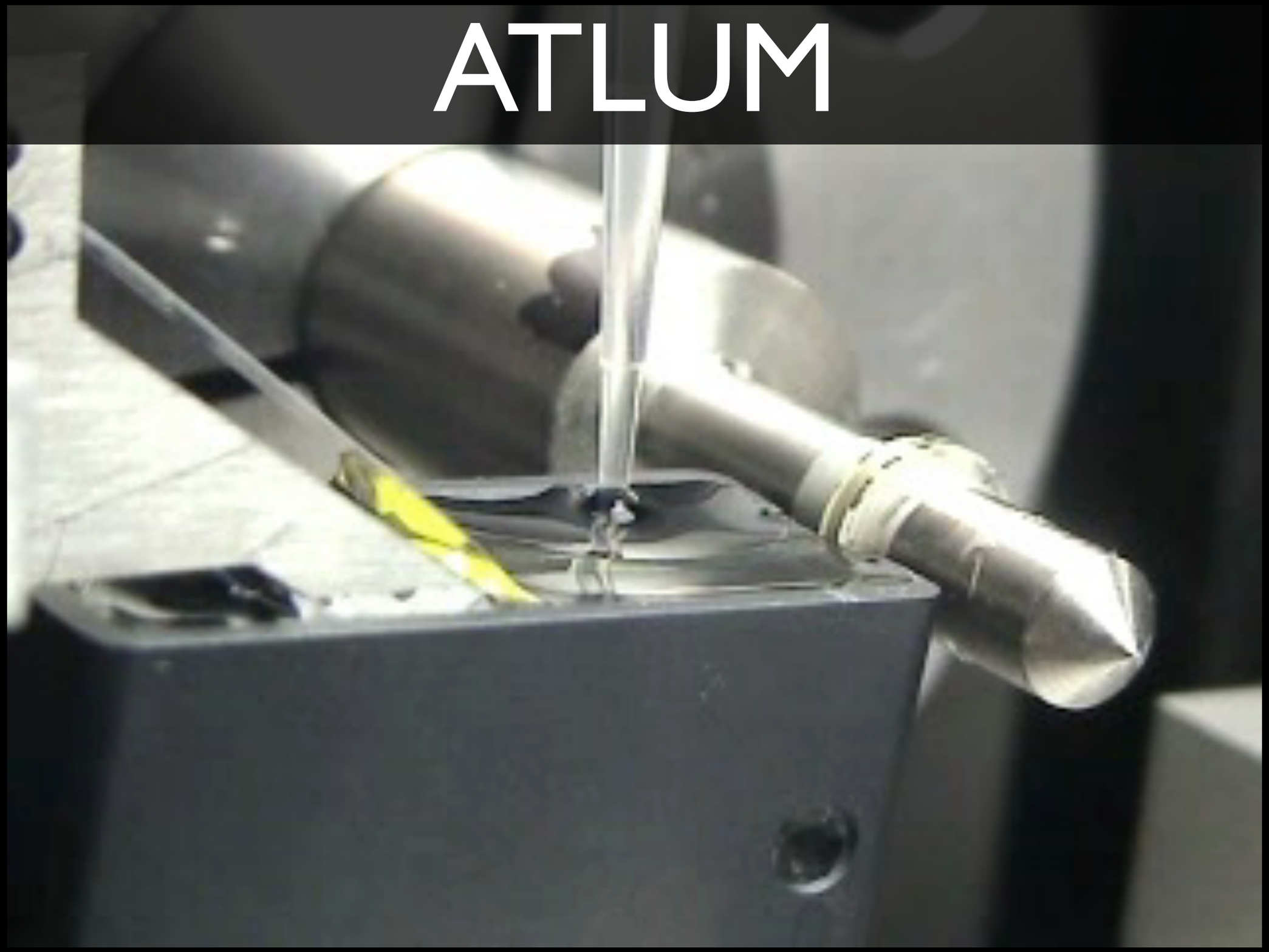




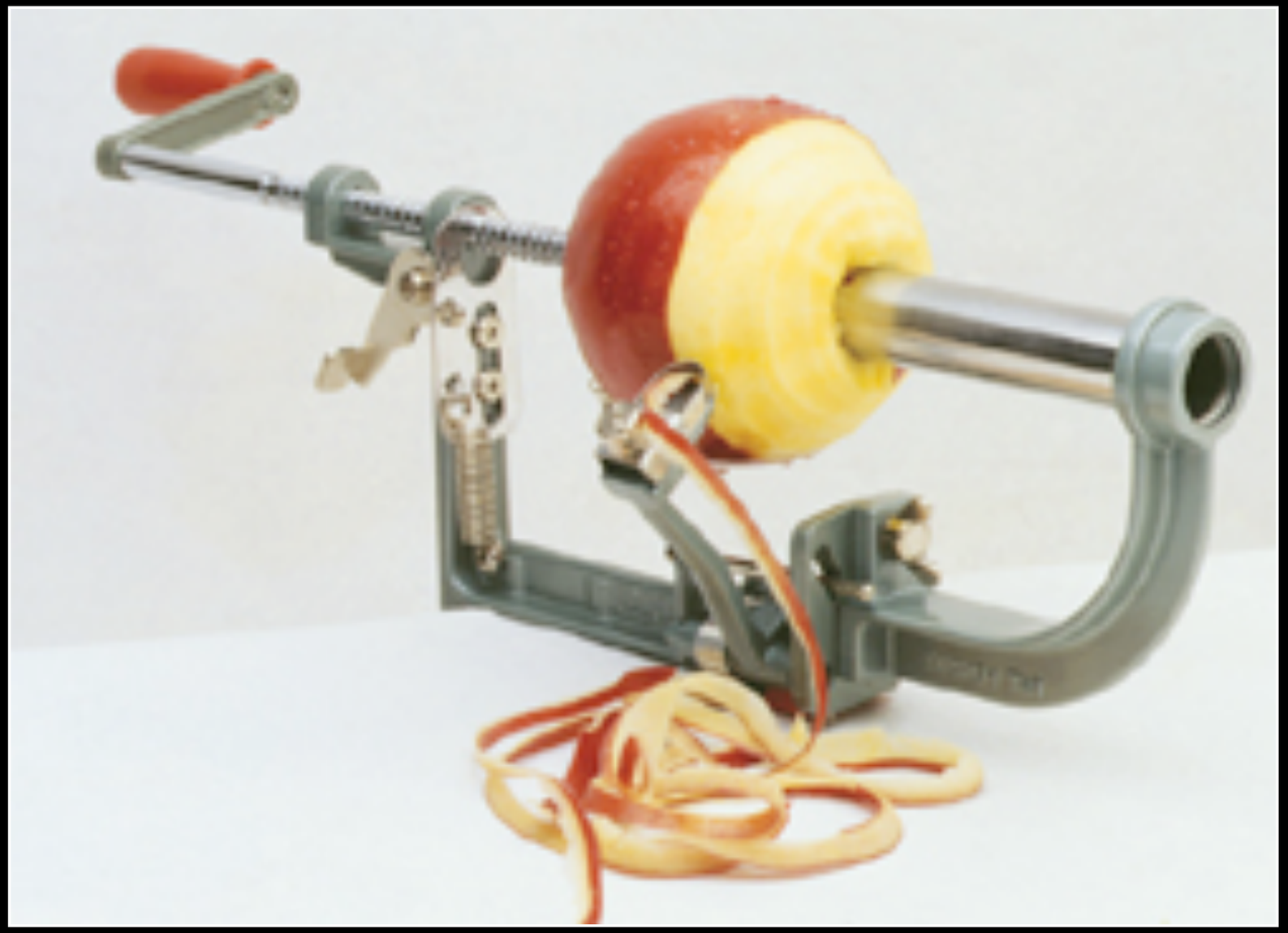
Automatic Tape-collecting Lathe Ultra Microtome



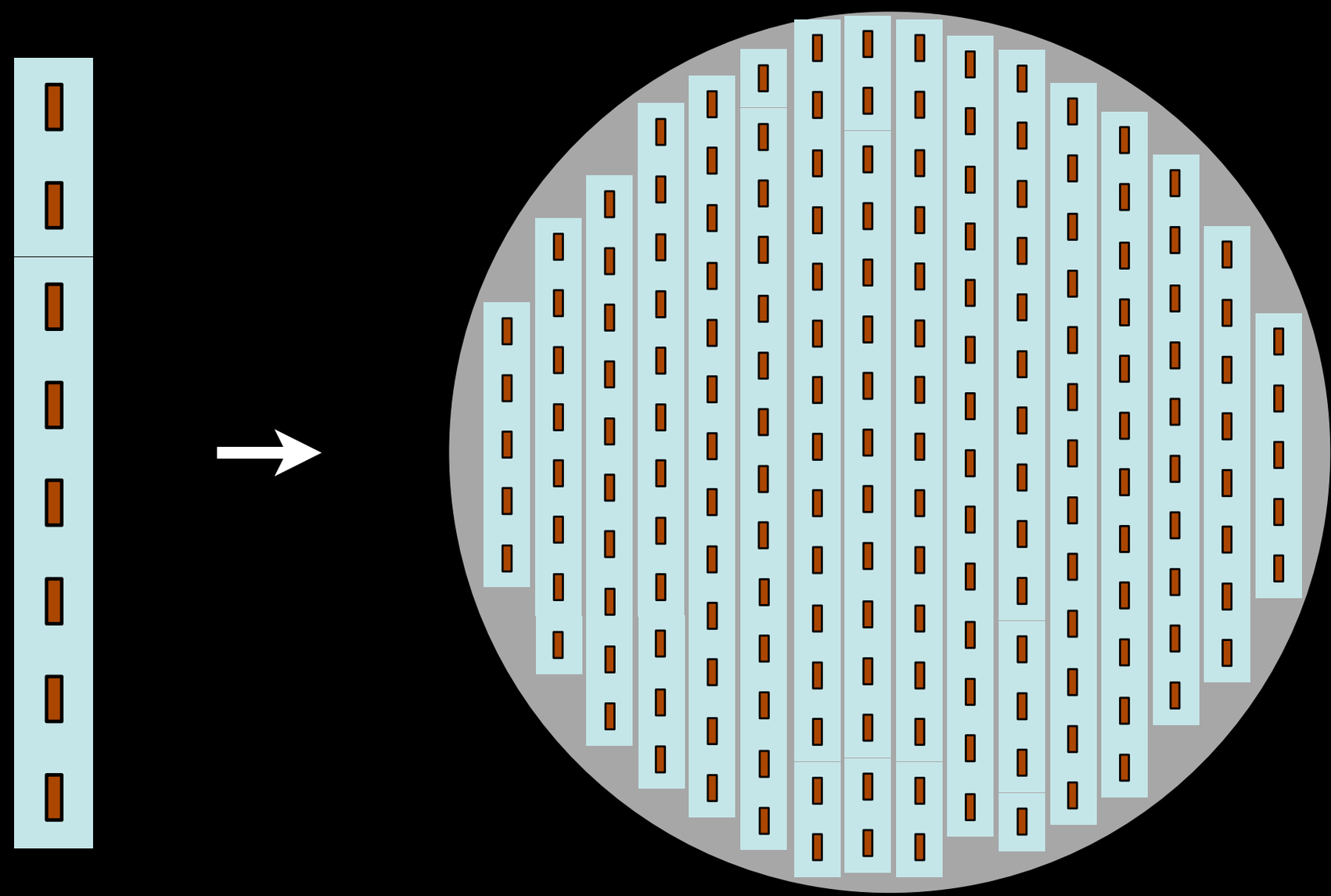
ATLUM



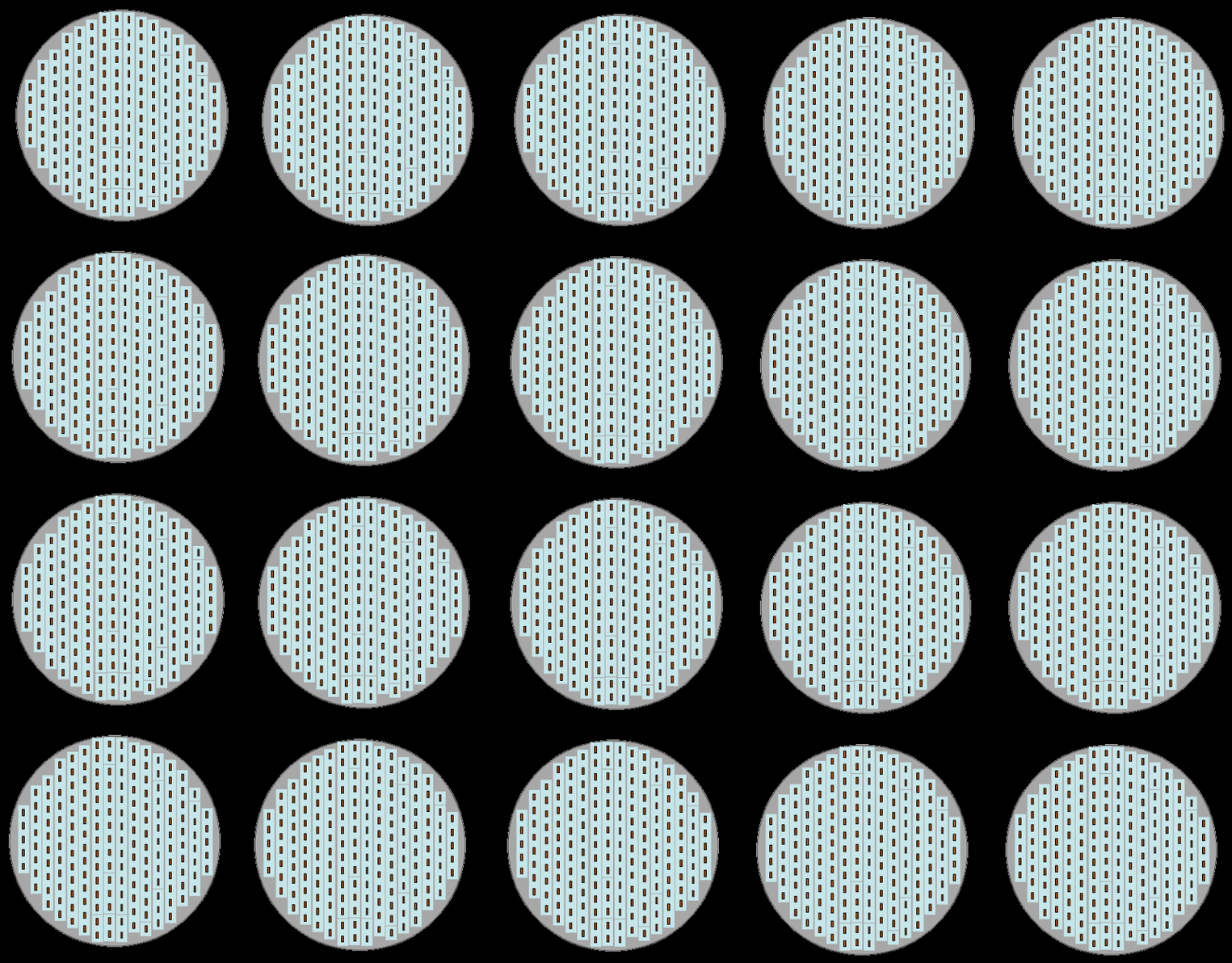
Inspiration for ATLUM



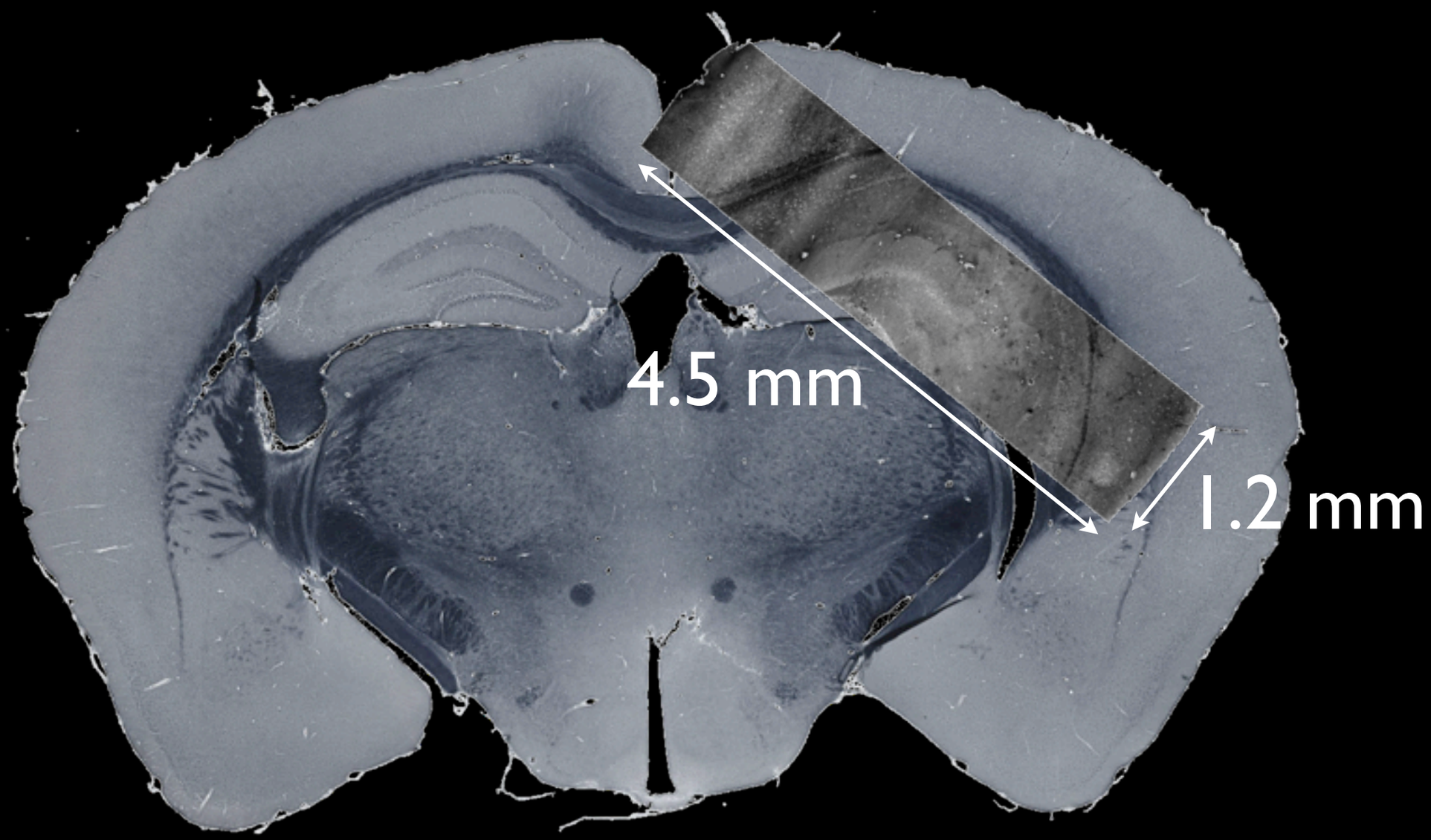
Ultrathin Section Libraries

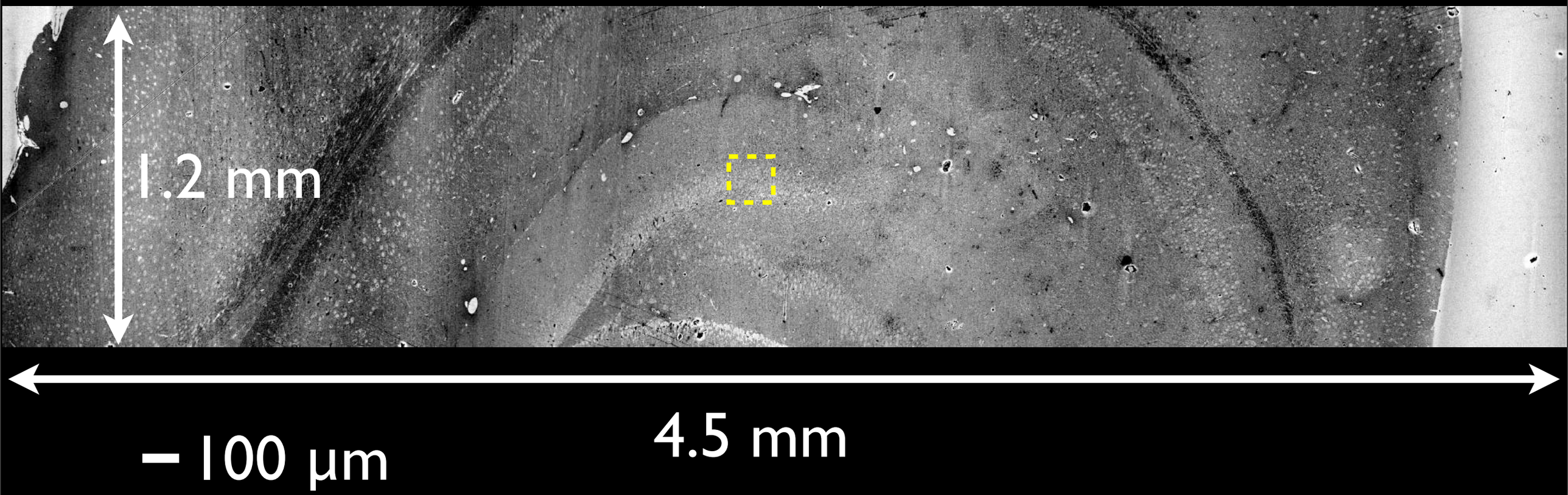


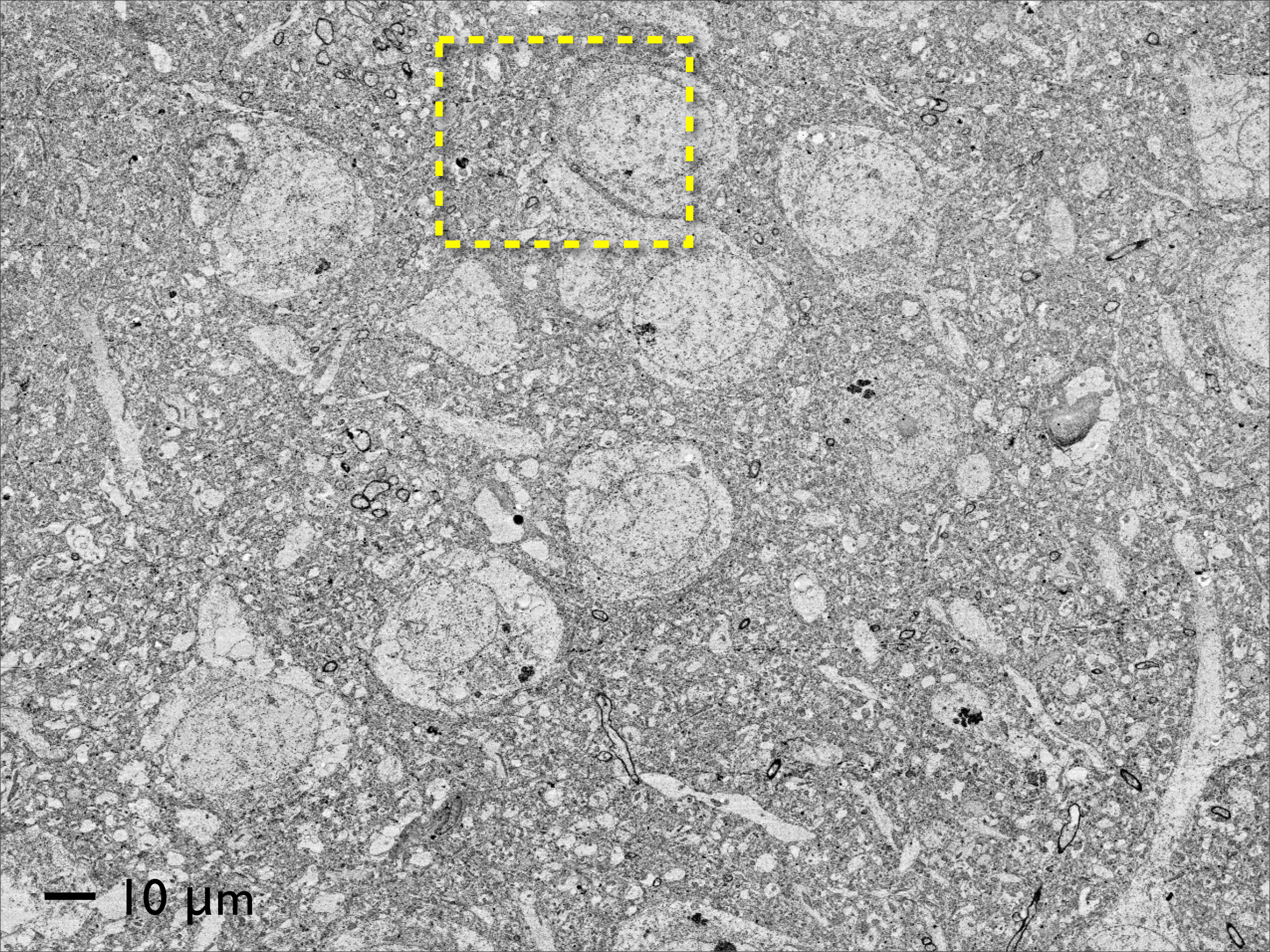
20 wavers = 1 mm^3 tissue

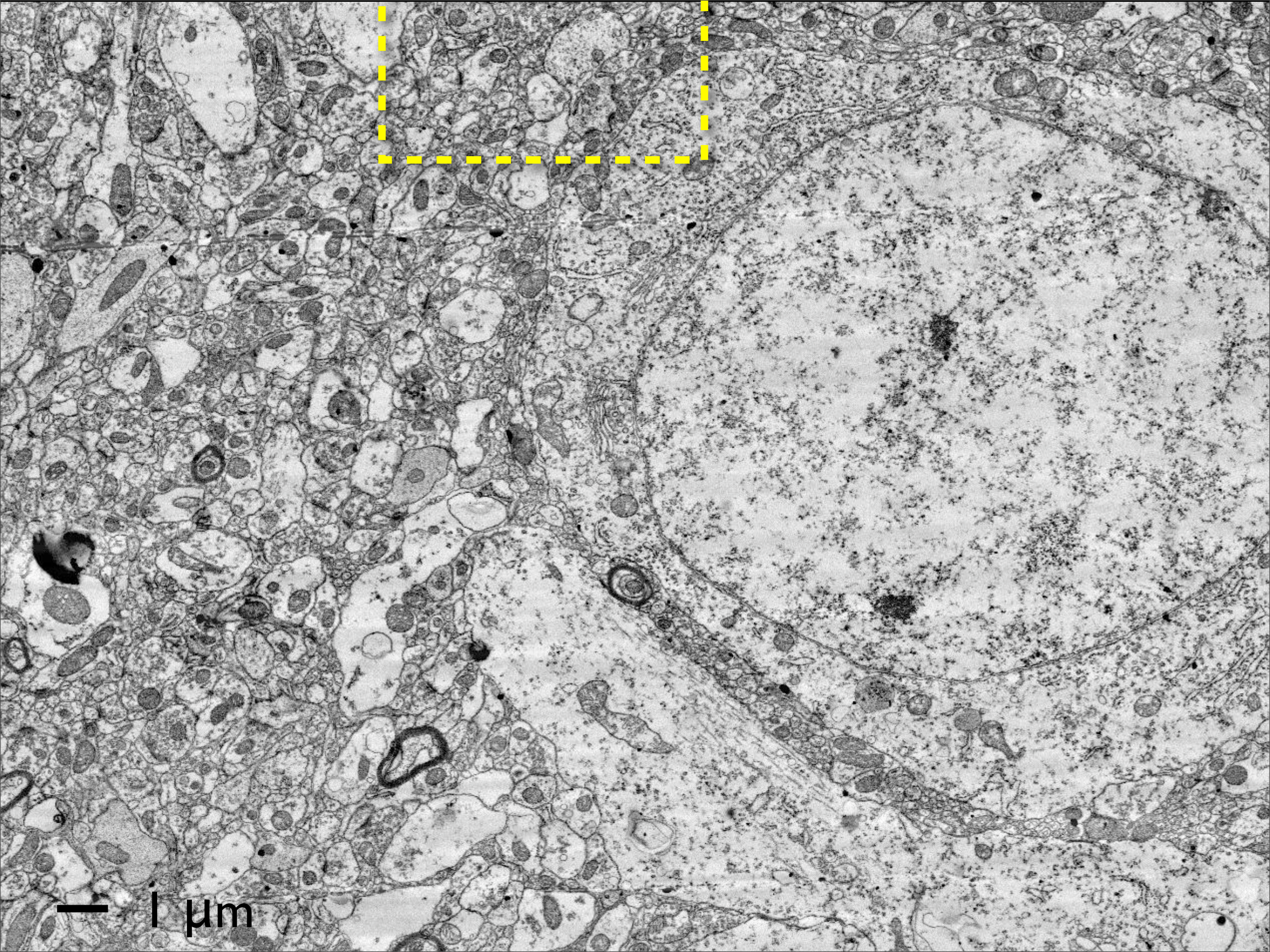


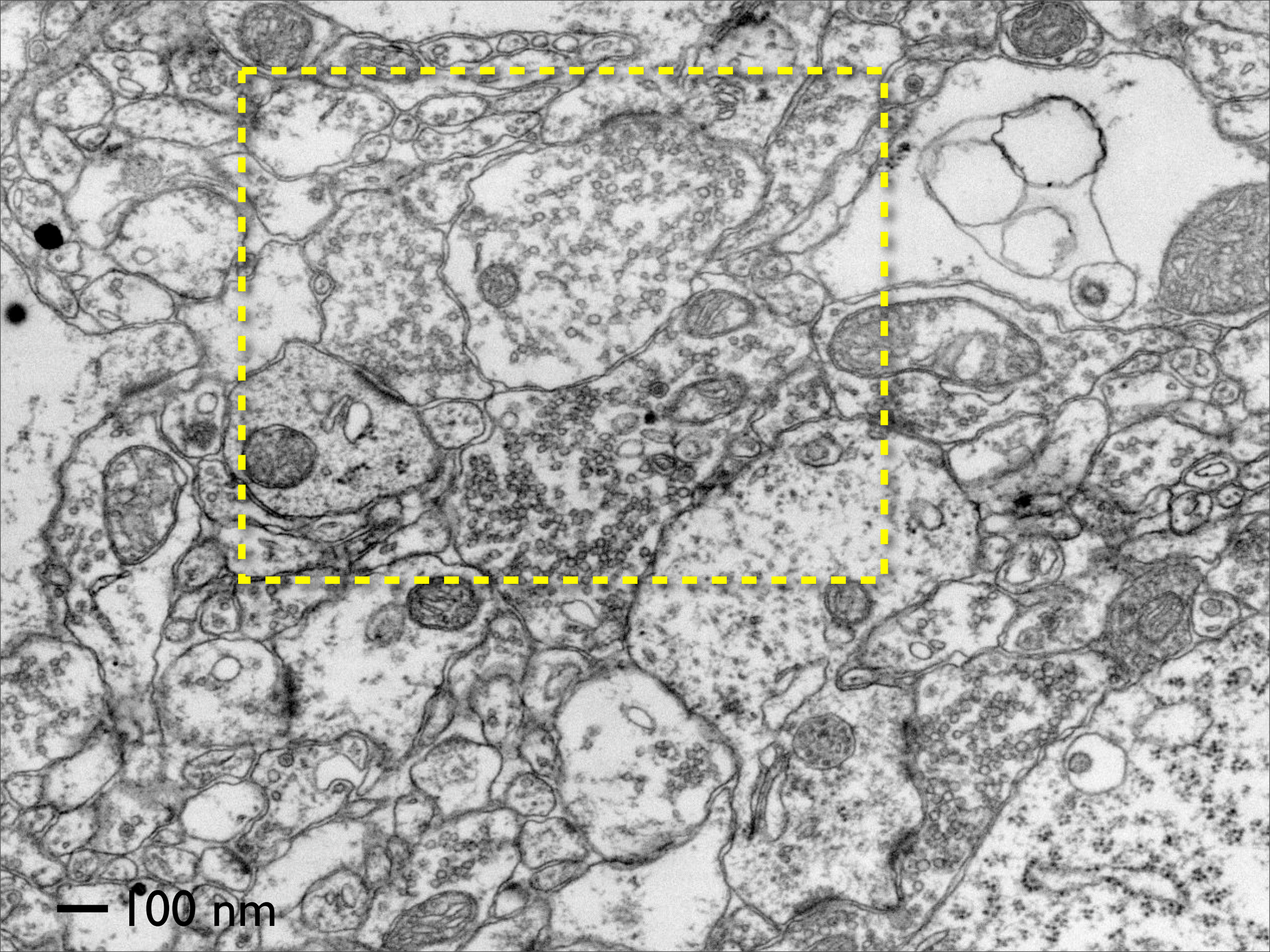
Electron Microscopy

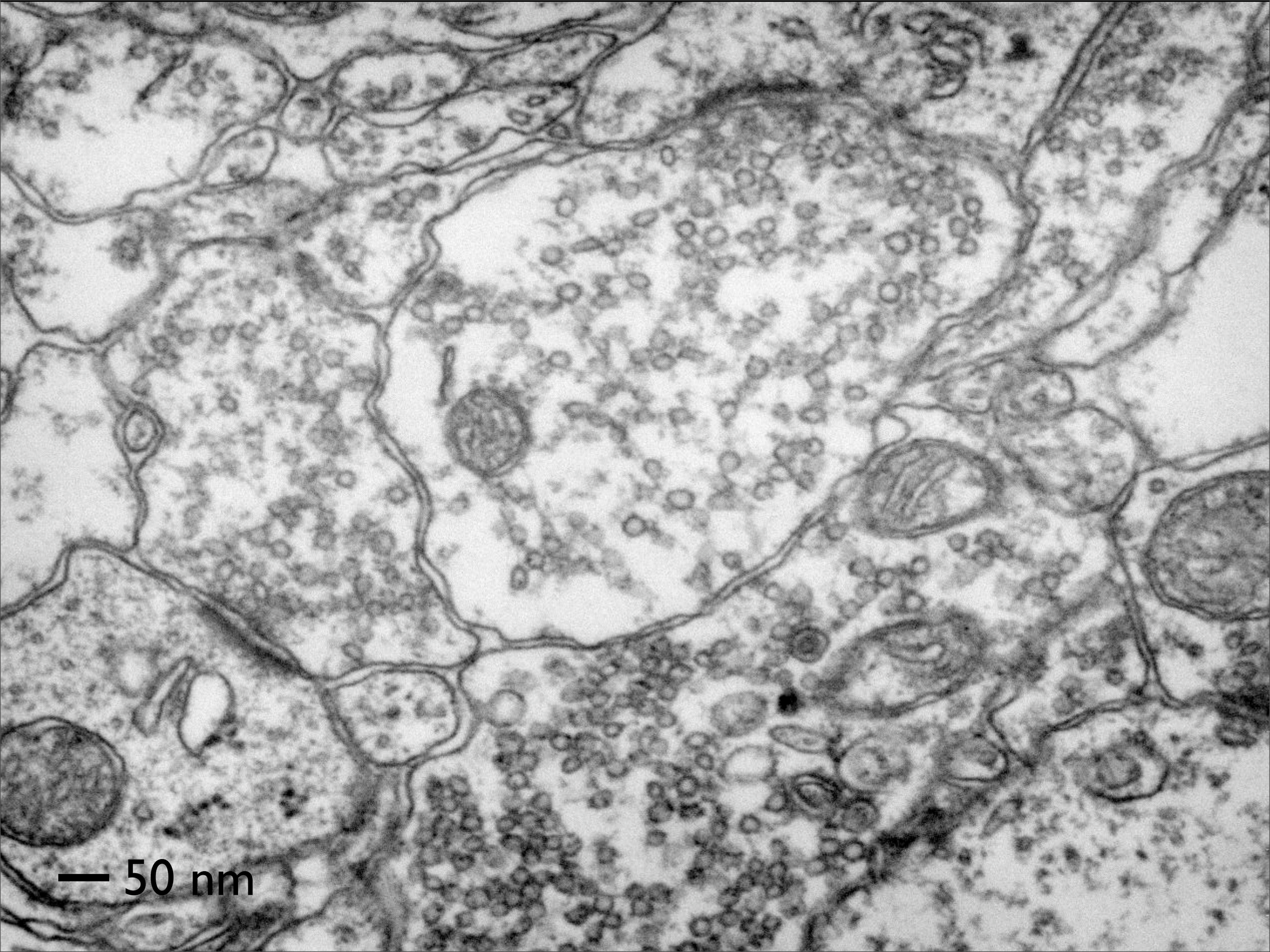



1.2 mm
4.5 mm
100 µm

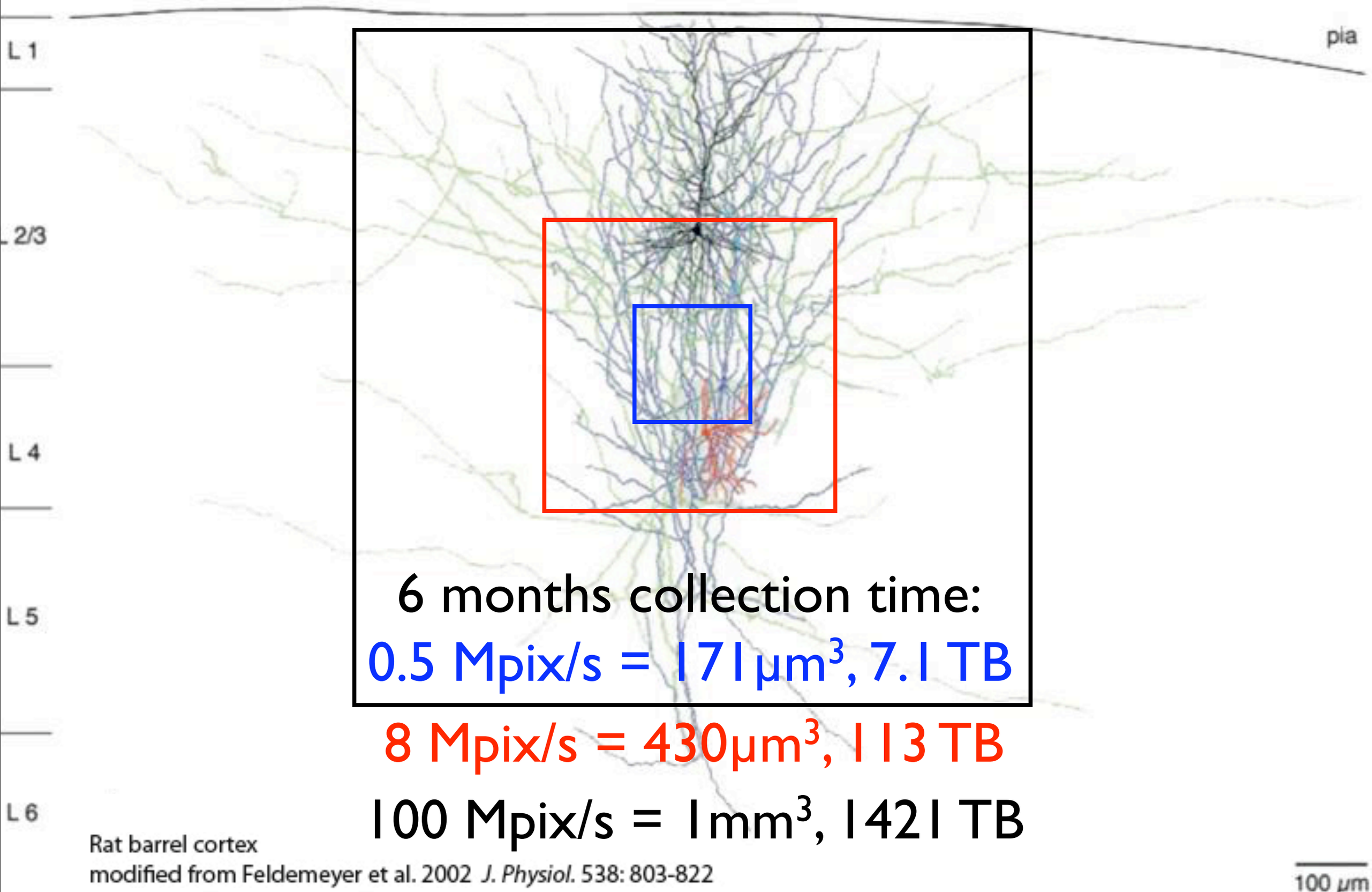

10 µm


1 µm


100 nm


50 nm

4 nm x 4 nm x 40 nm voxels



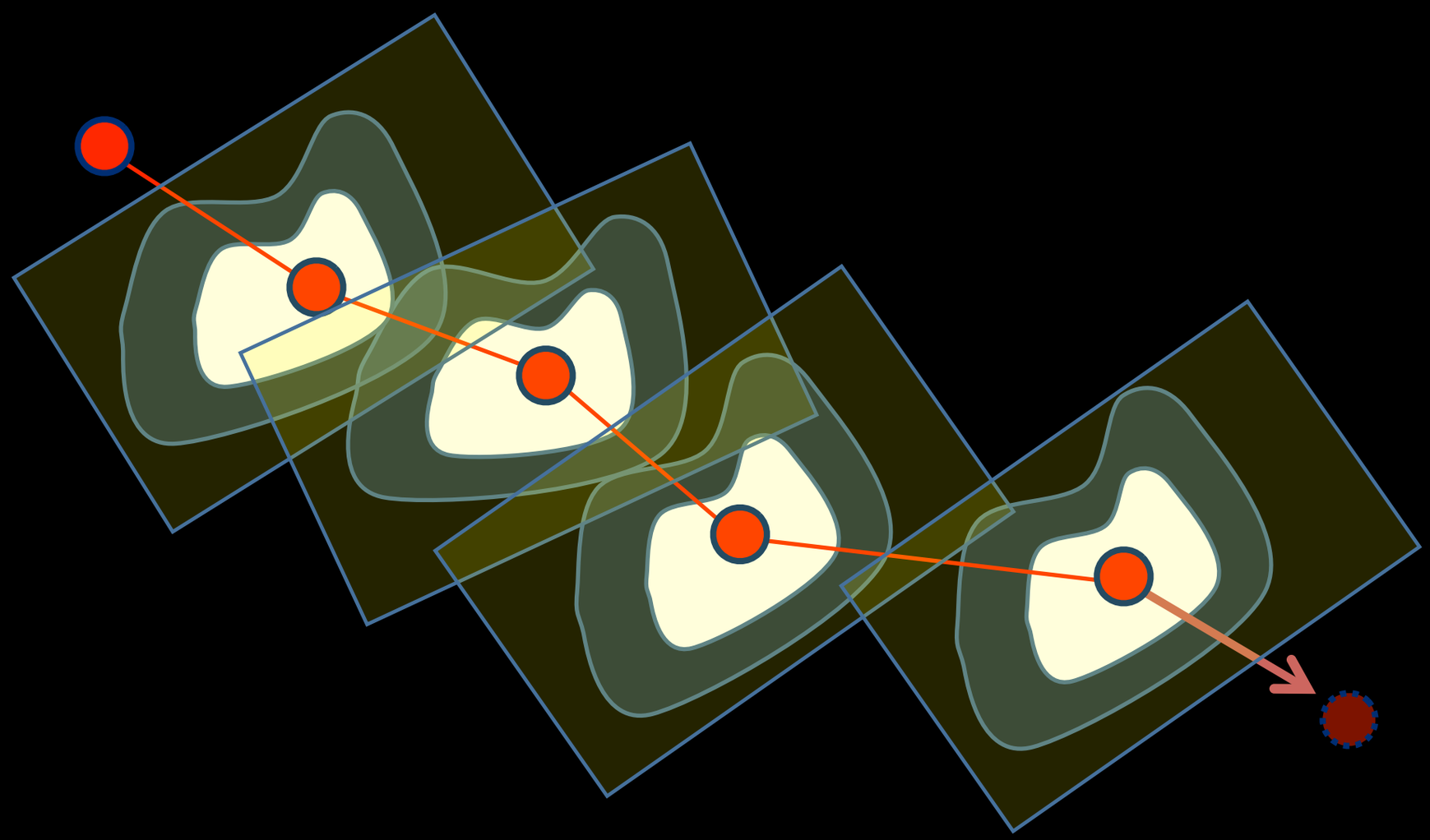
Segmentation

- Active Ribbons [Vazquez et al., CVPR 09]

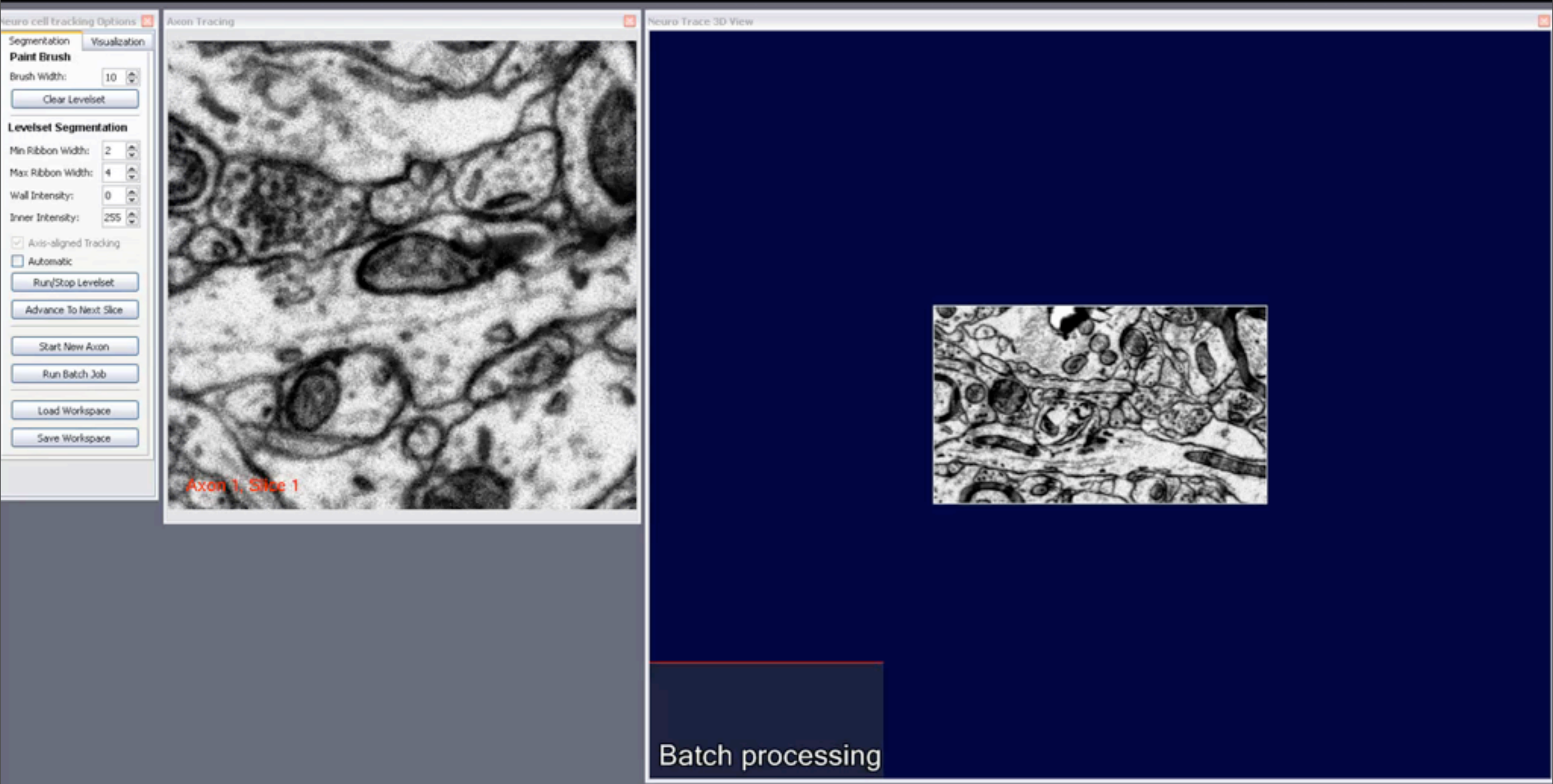
Section 1

Active Ribbons on GPUs

- 23x speedup over CPU

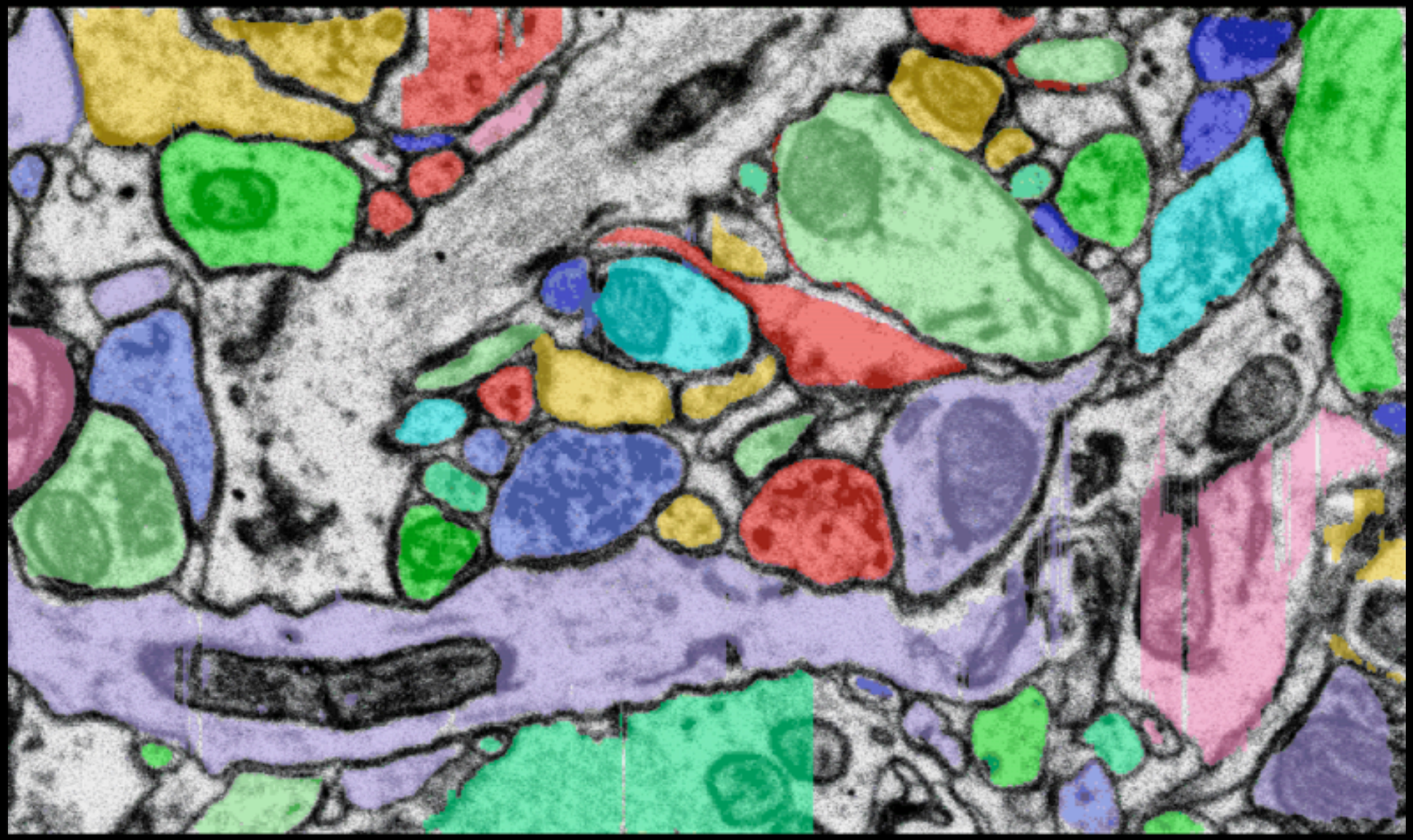


Segmentation

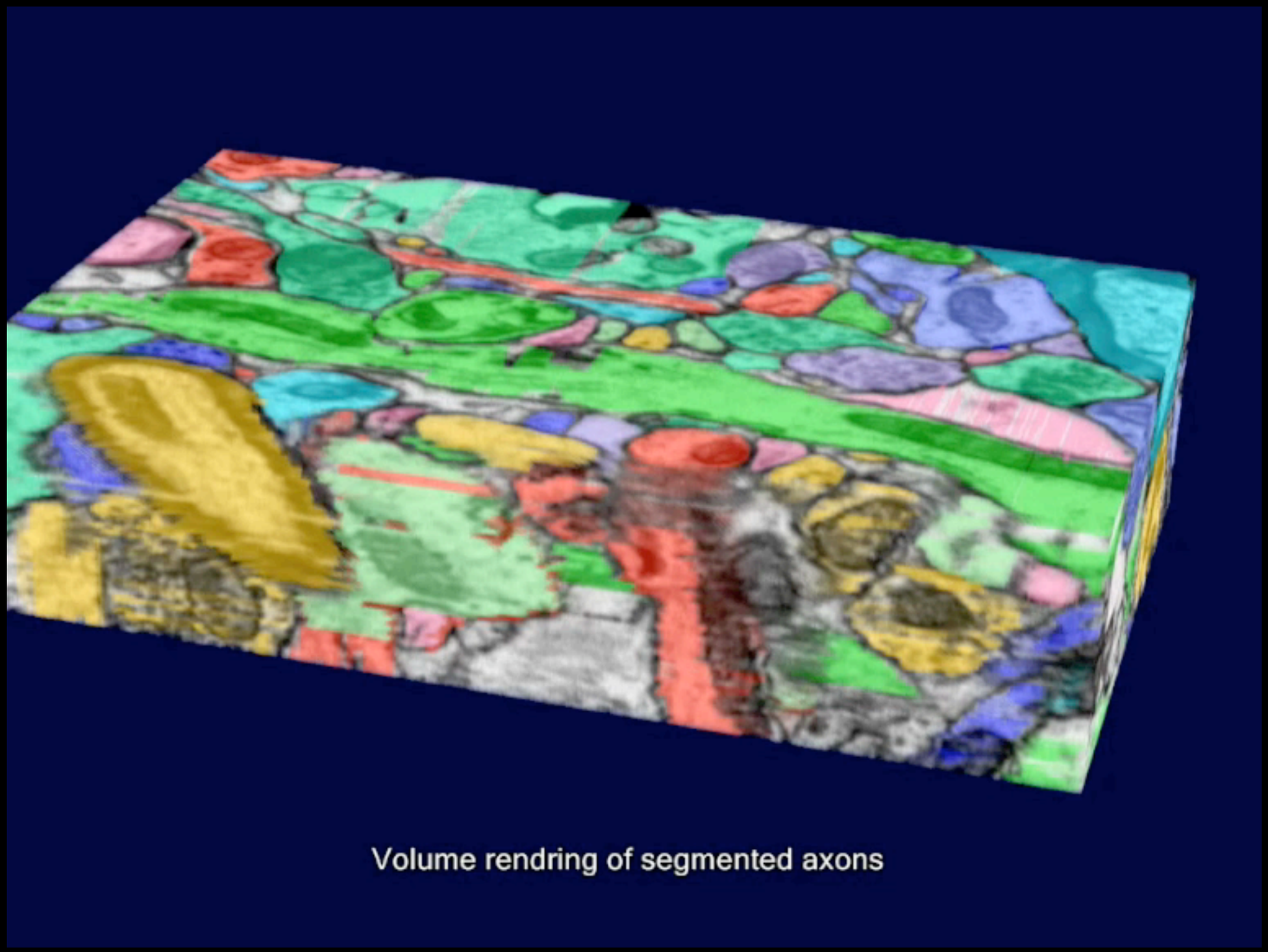


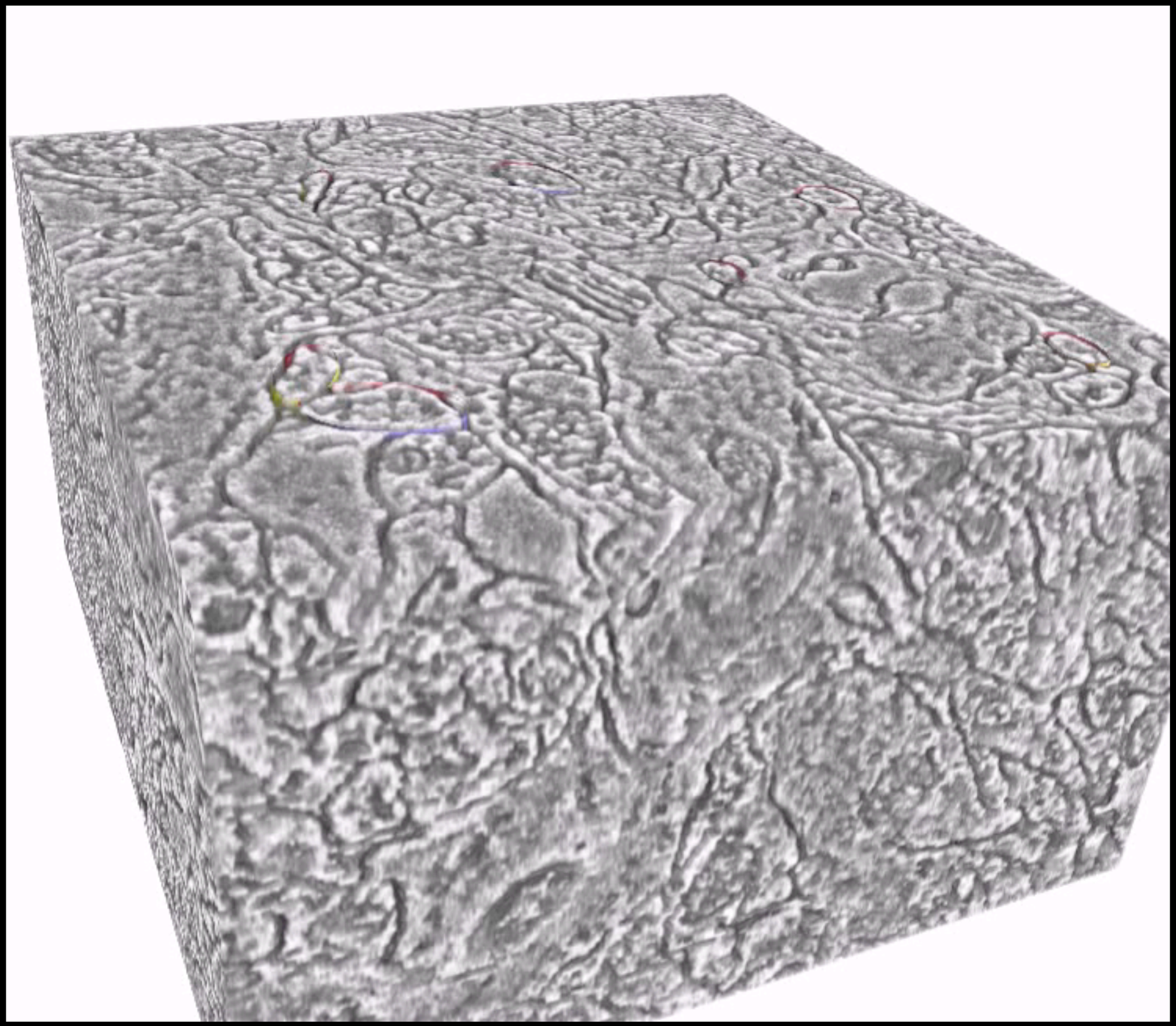
Active Ribbons [Vazquez et al. 09] NeuroTrace [Jeong et al. 09]

Segmentation

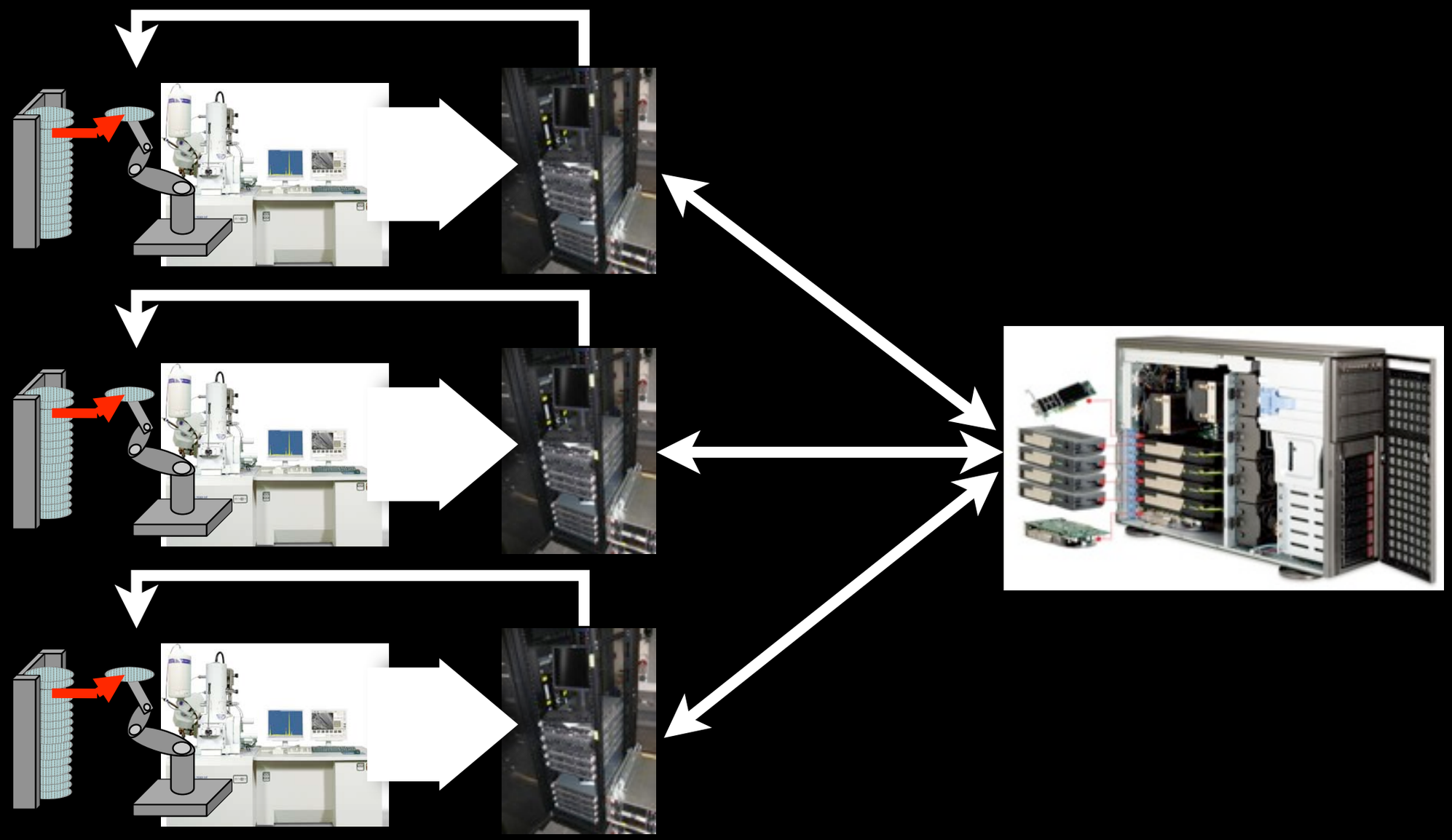


Visualization

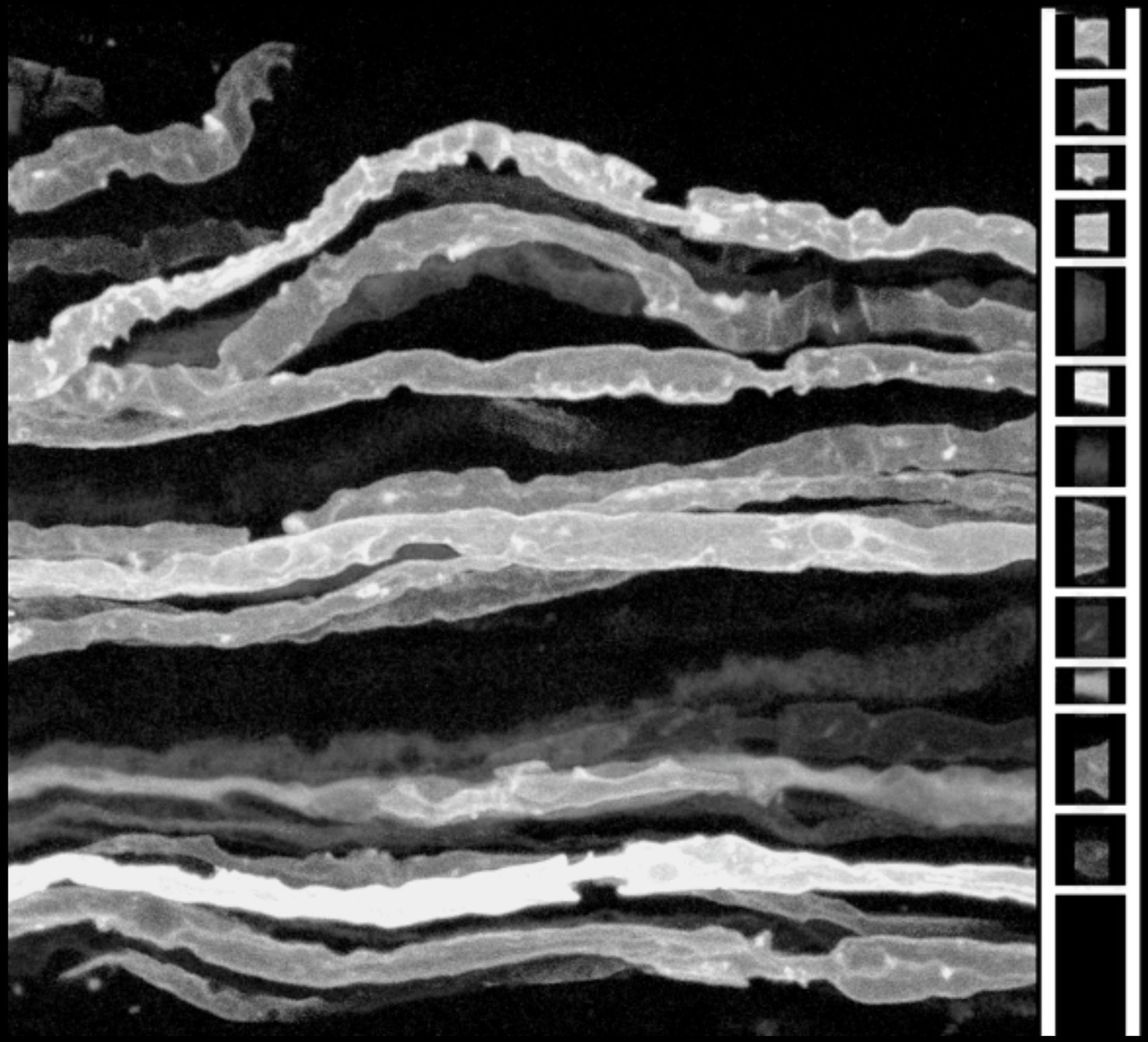




High-Throughput Connectomics

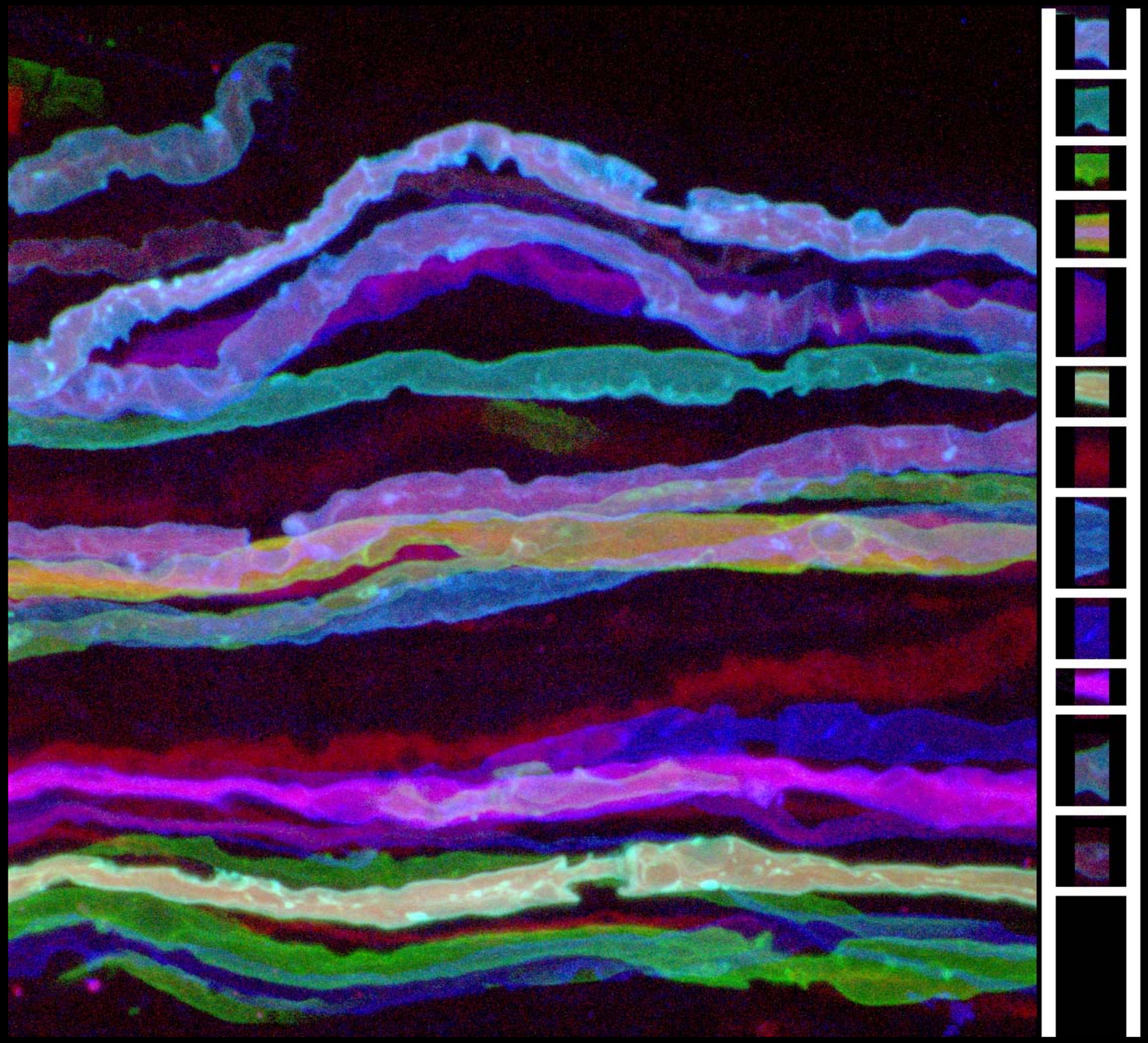


The Future

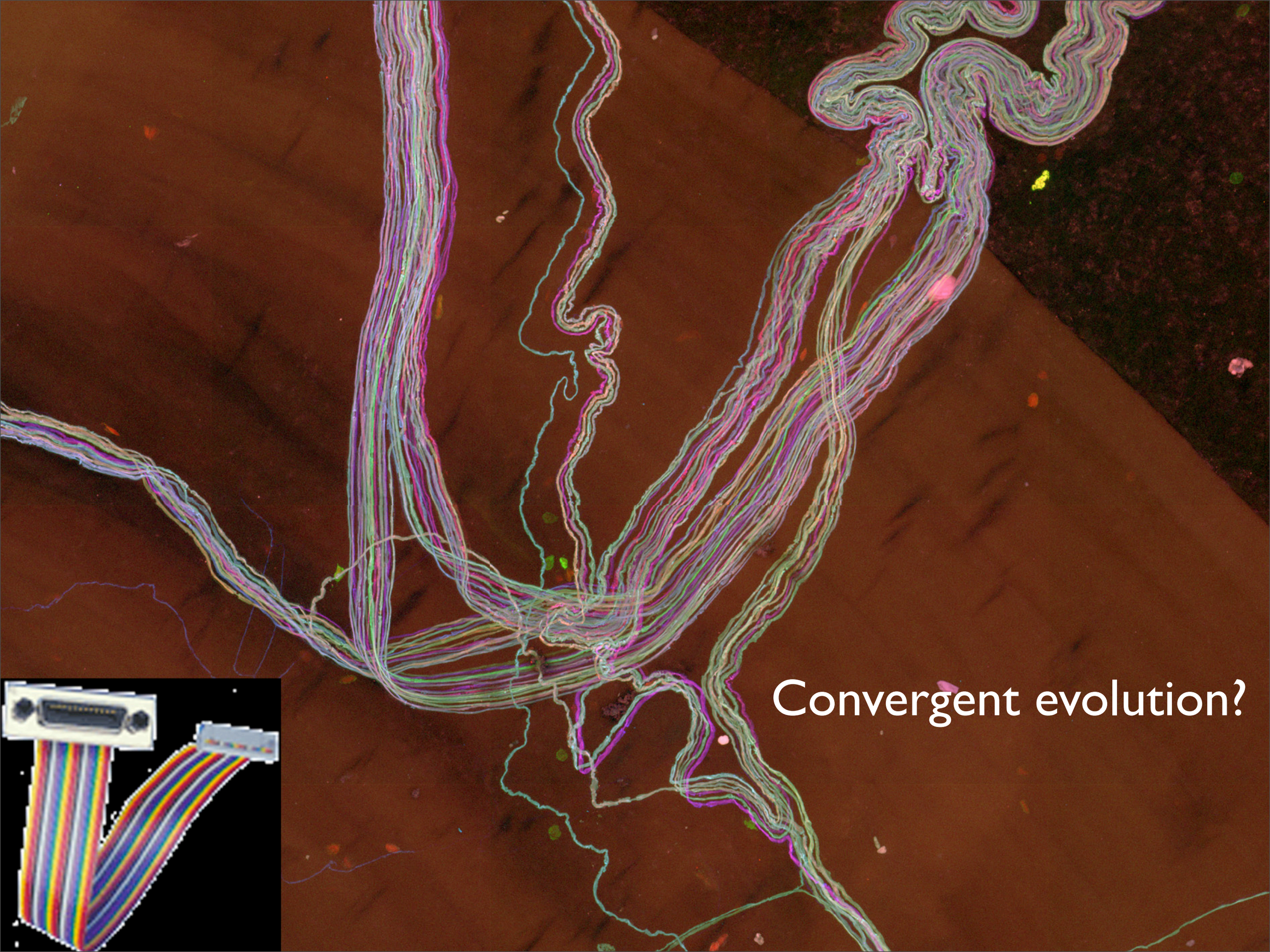


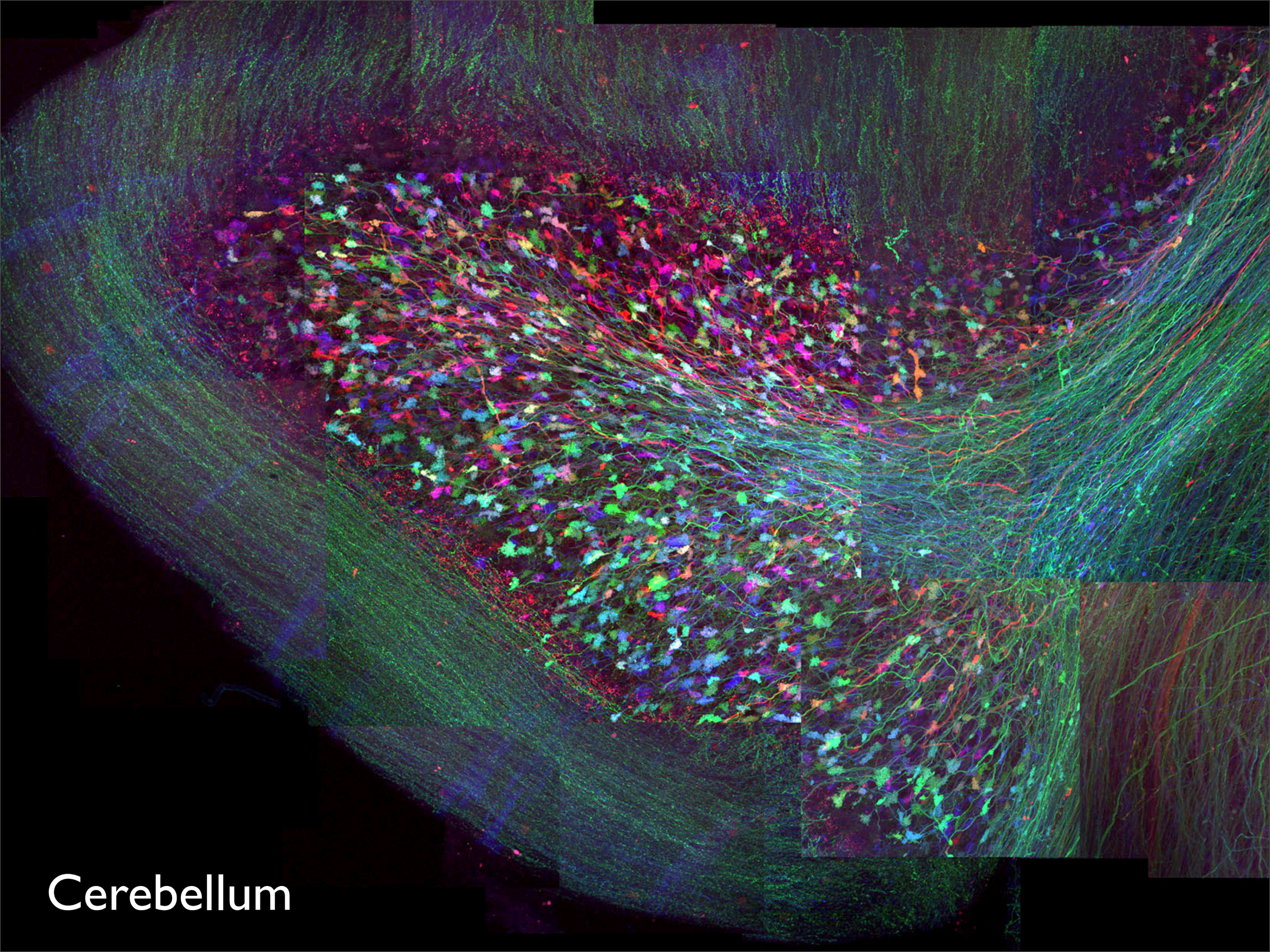
J. Livet

Brainbow



J. Livet


Convergent evolution?


Cerebellum

The MWA

The Industrialization of Astronomy

Collaborators

- Harvard Center for Astrophysics
  - Lincoln Greenhill
  - Daniel Mitchell
  - Stephen Ord
  - Randall Wayth
- SEAS
  - Kevin Dale, Richard Edgar
  - Hanspeter Pfister

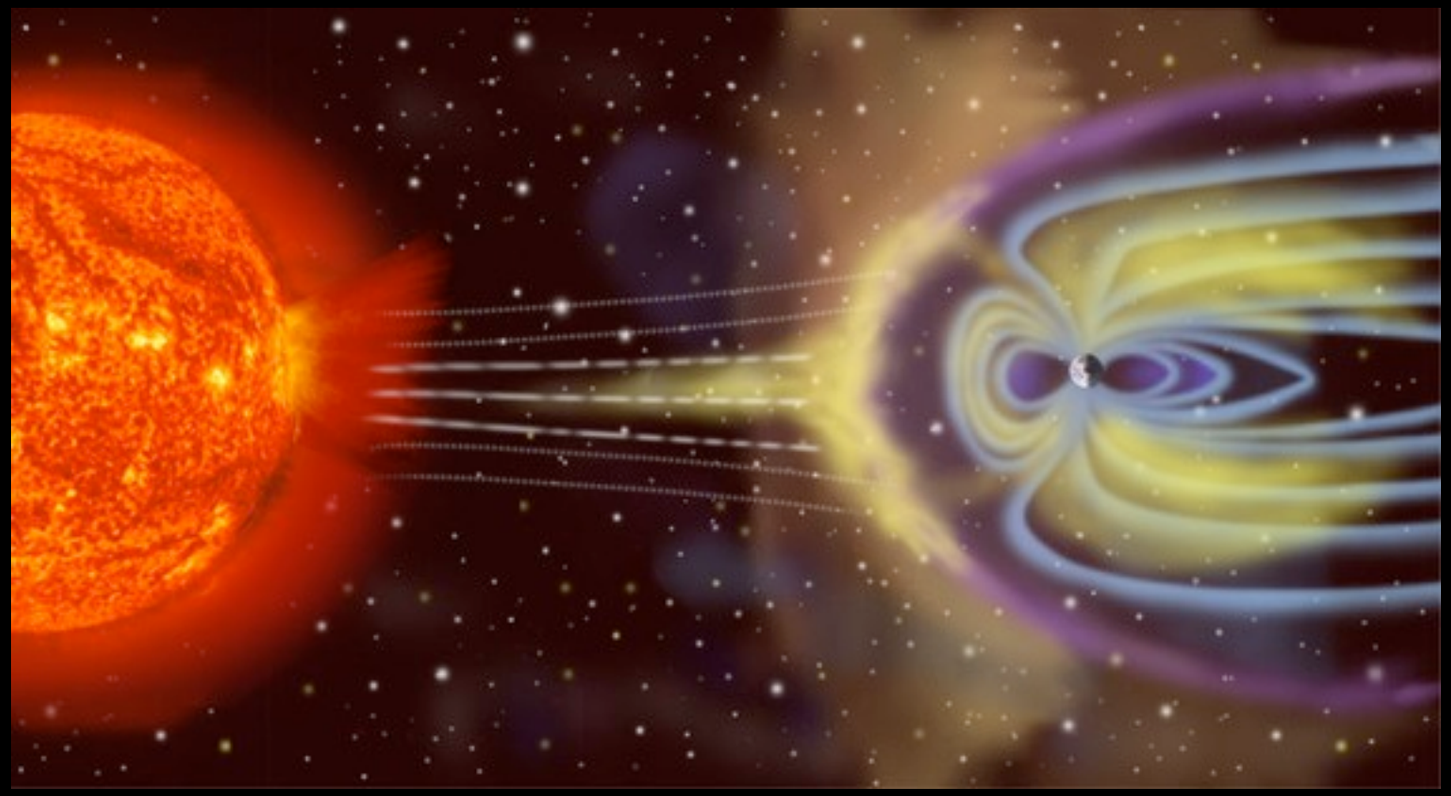
The Scientific Challenge

- What happened through the 1 billion years of the universe's dark ages?

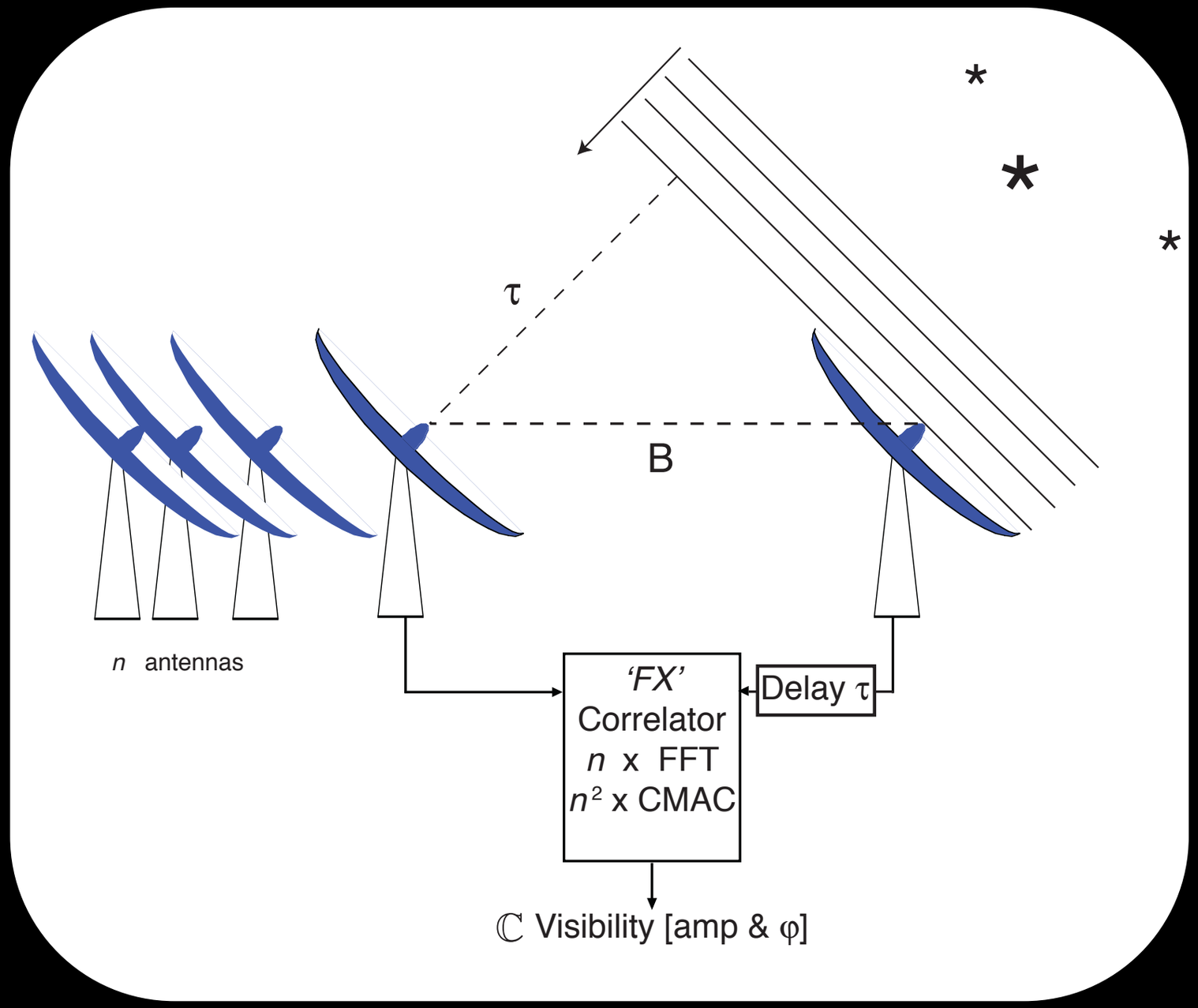
~300,000 ~1 bio

The Scientific Challenge

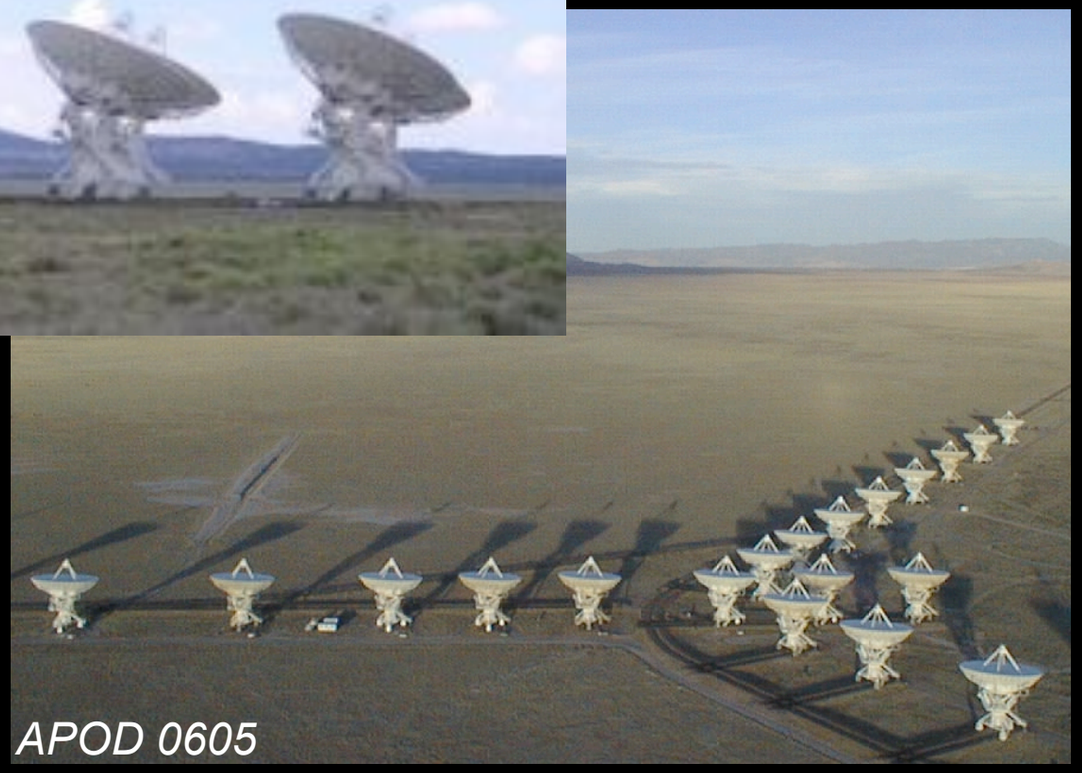
- What happened through the 1 billion years of the universe's dark ages?
- Can we predict coronal mass ejections?



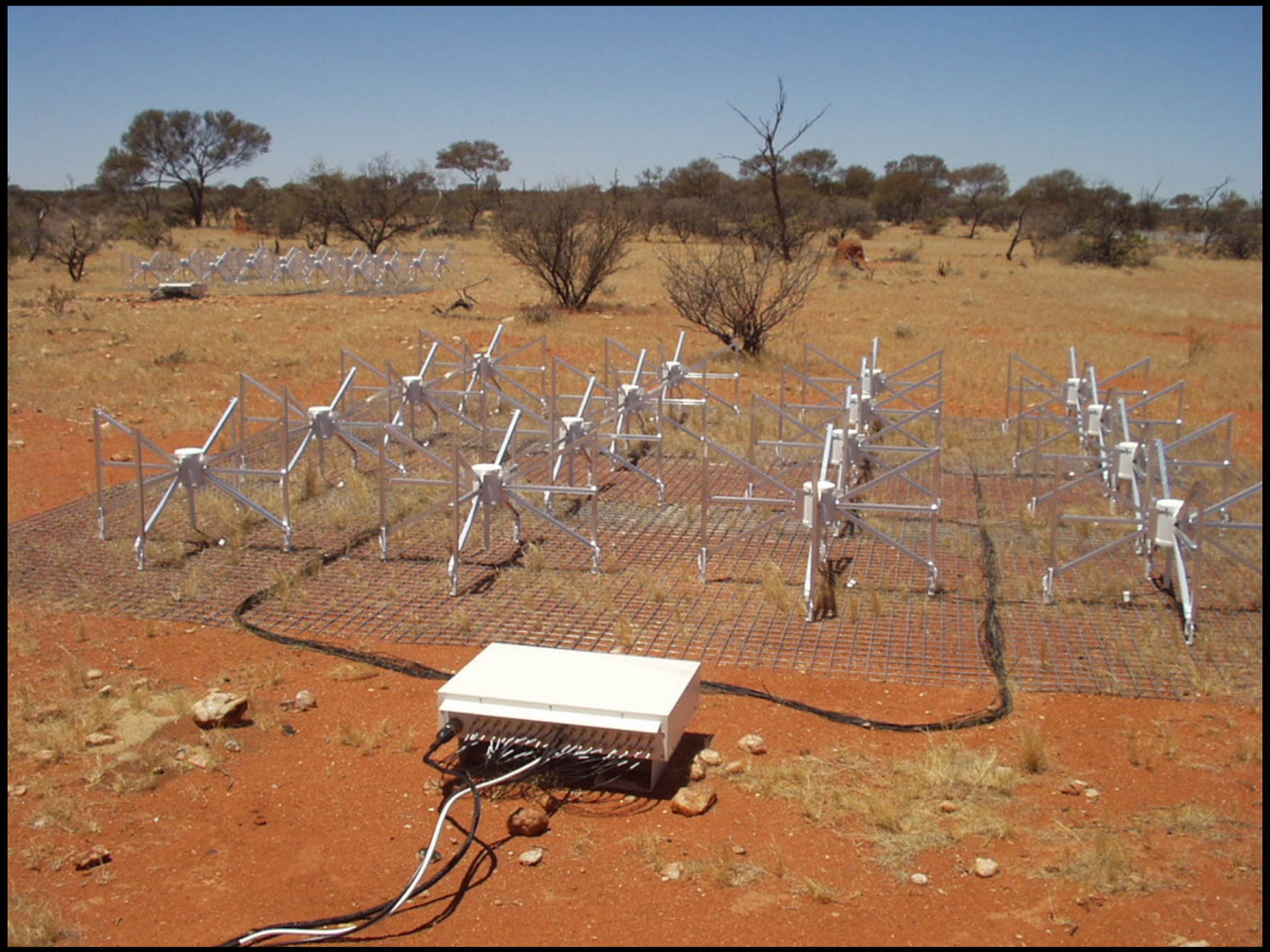
Radio Interferometry



From this...



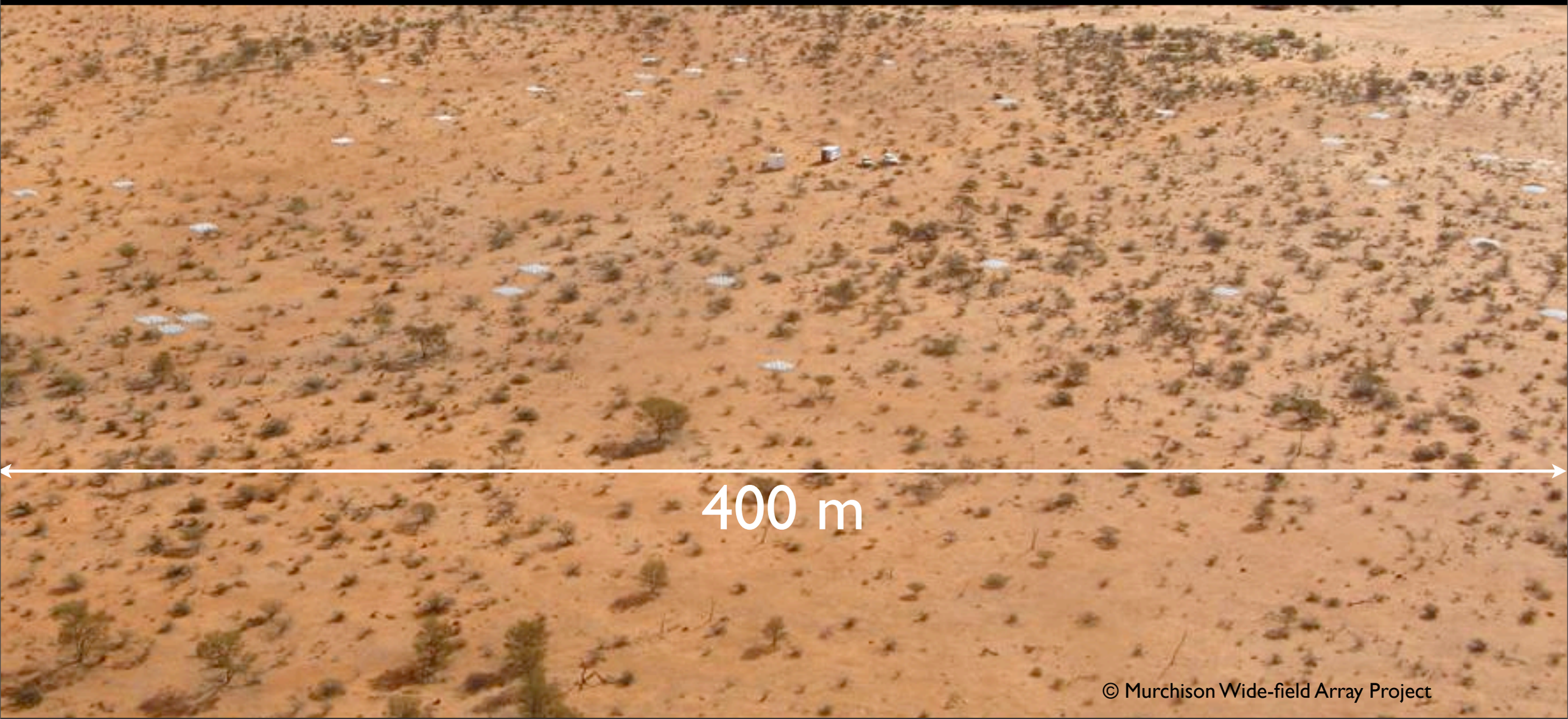
...to this



© Murchison Wide-field Array Project

MWA

- 32 antenna prototype
- Eventually 512 antennas on 1 km^2

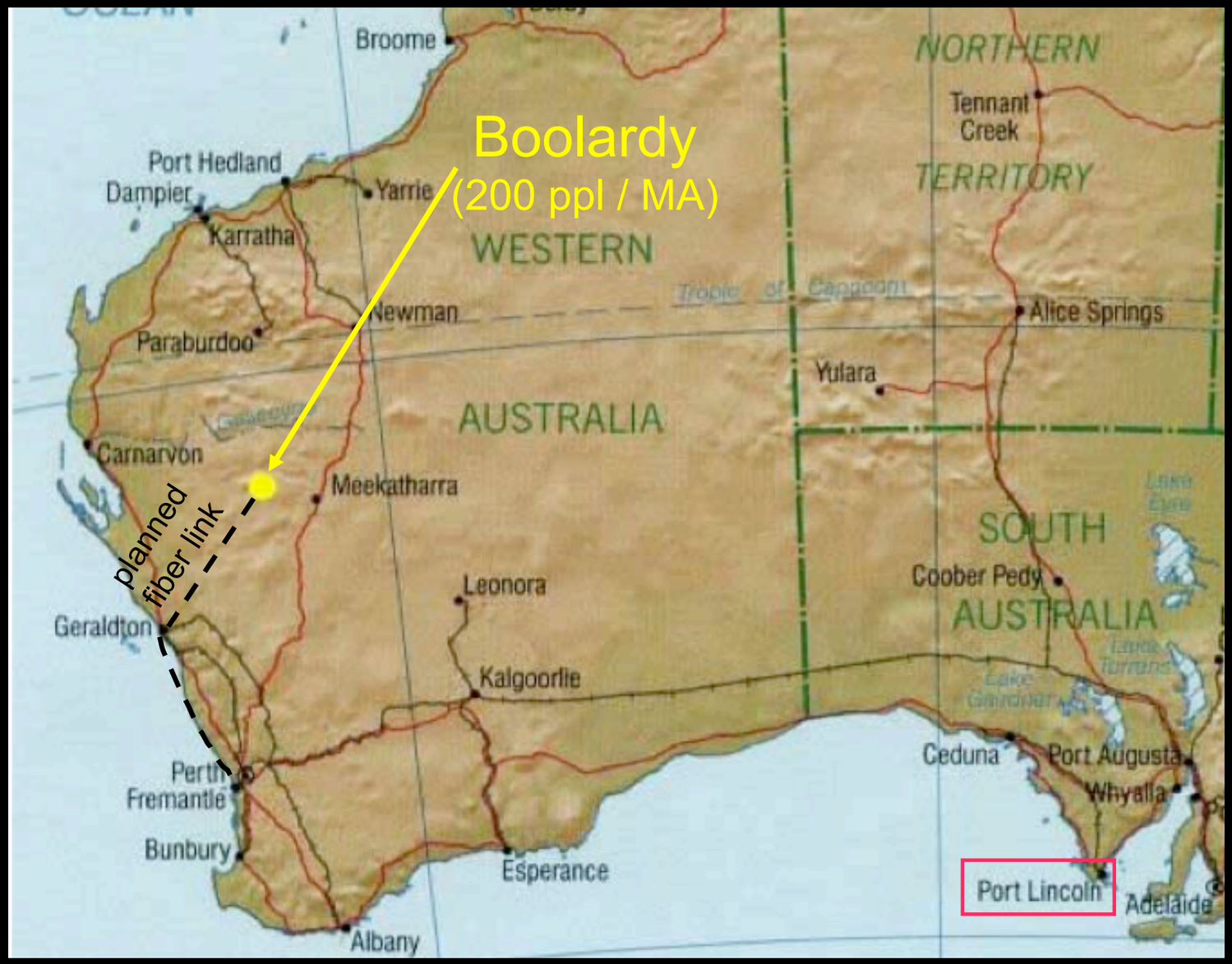


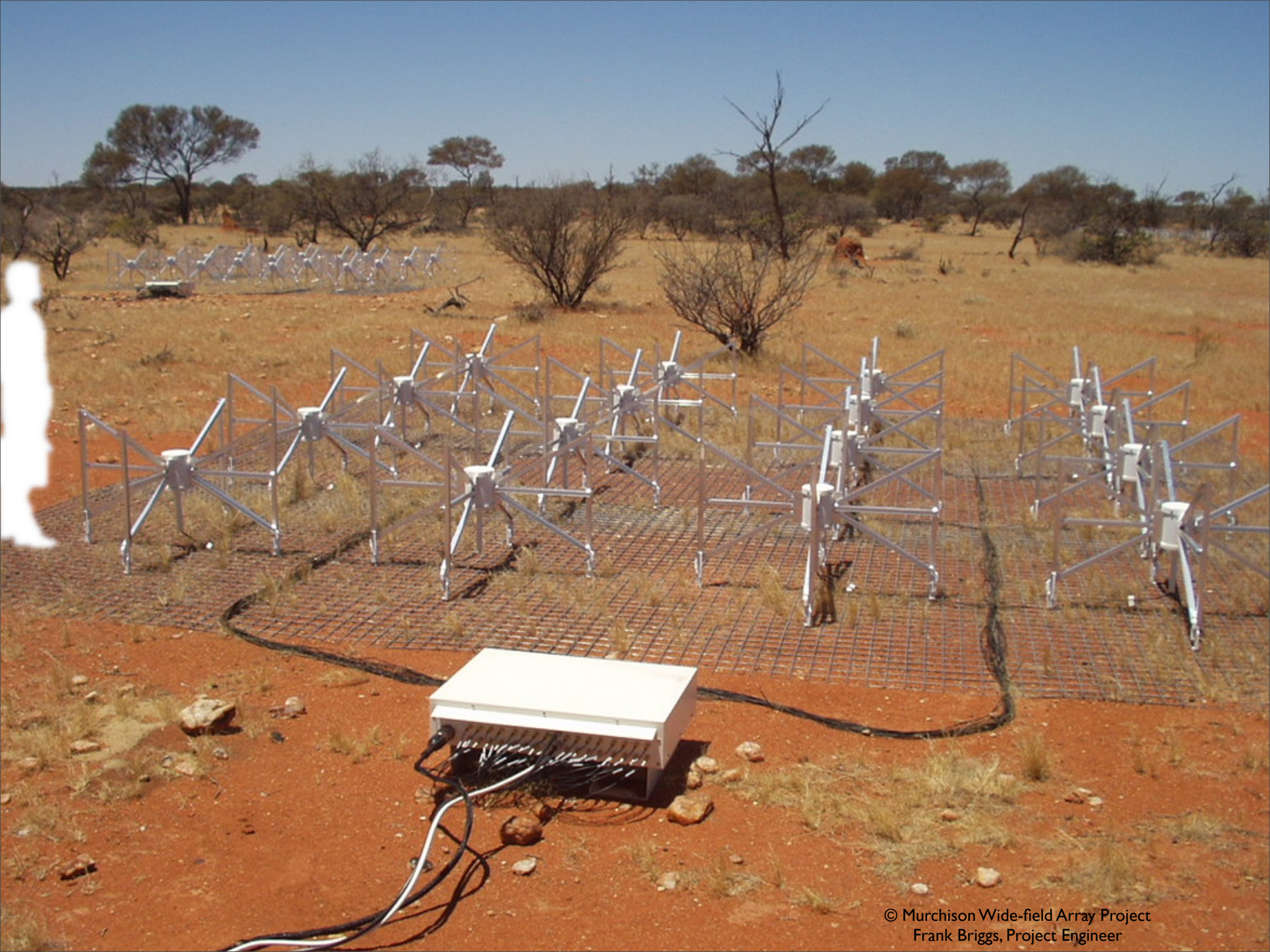
© Murchison Wide-field Array Project

Nowhere (Middle of)
300 km
THANK YOU FOR BEING RADIO QUIET

Outback Computing

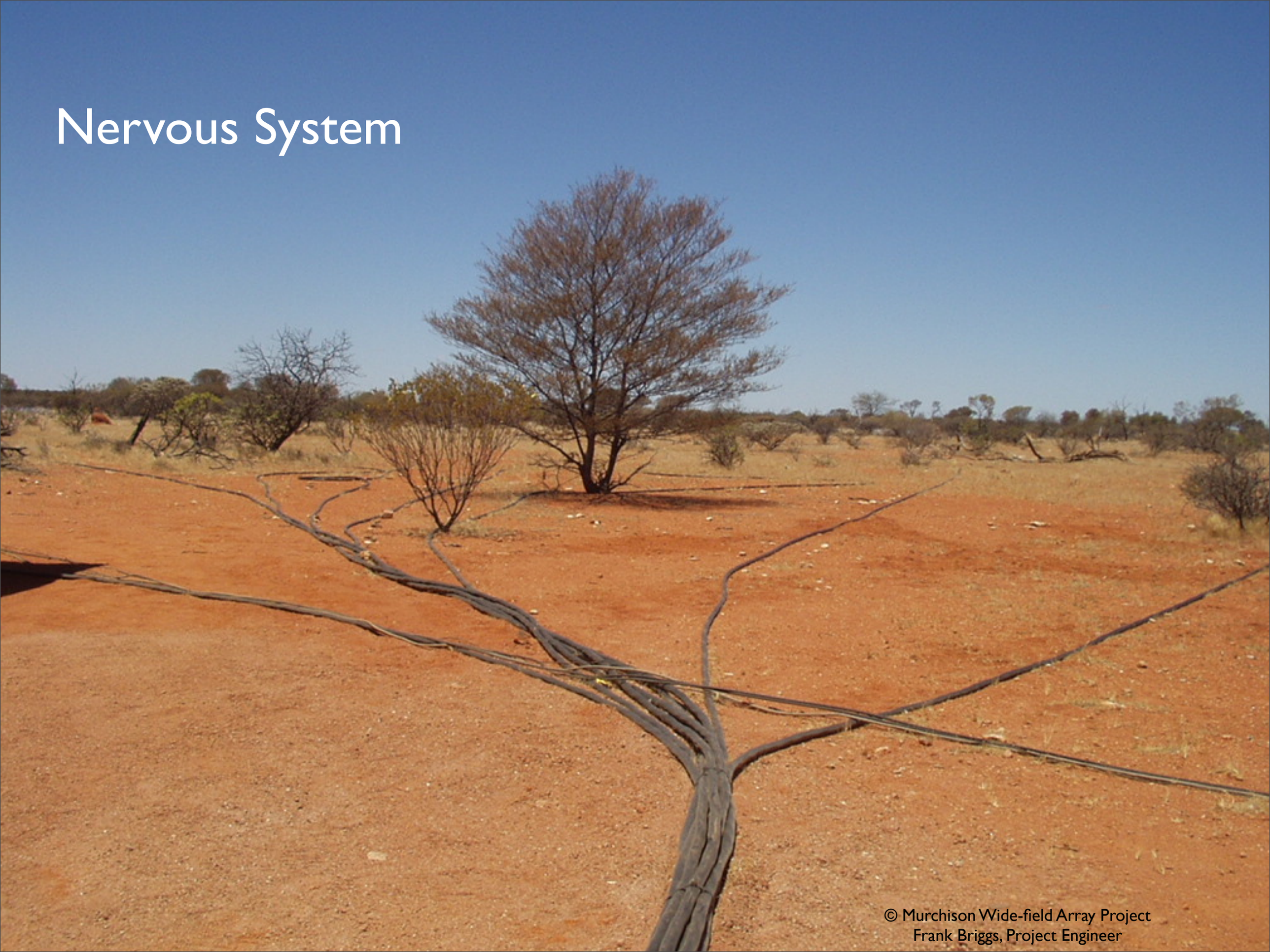
- Max. ambient on site: 50° C / 122° F





© Murchison Wide-field Array Project
Frank Briggs, Project Engineer

Nervous System



© Murchison Wide-field Array Project
Frank Briggs, Project Engineer

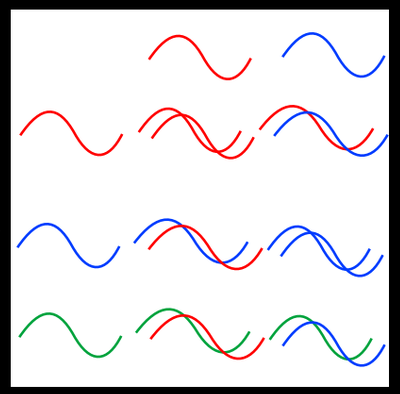
Supercomputing Center

Who is the GPU expert?

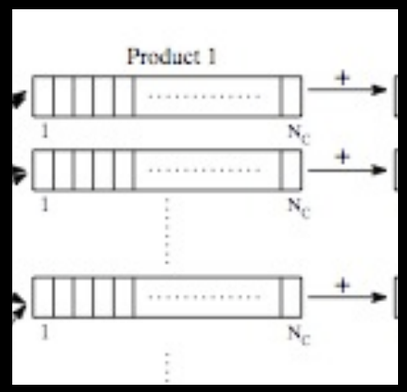
Life Support

© Murchison Wide-field Array Project
Frank Briggs, Project Engineer

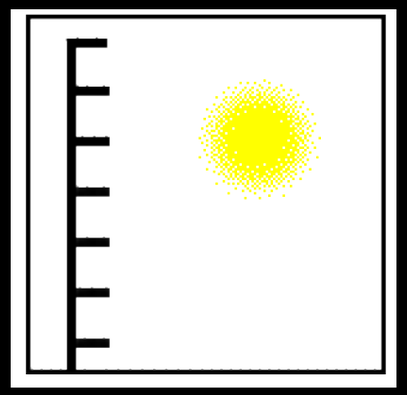
MWA Pipeline



Correlator



Integrator



Calibrator

Gridding

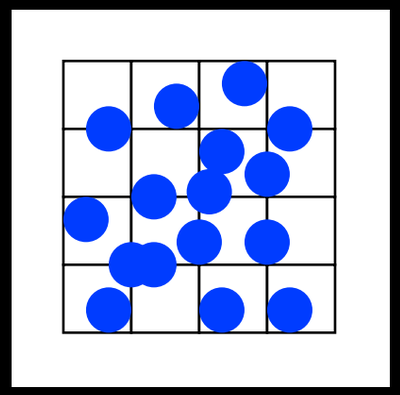
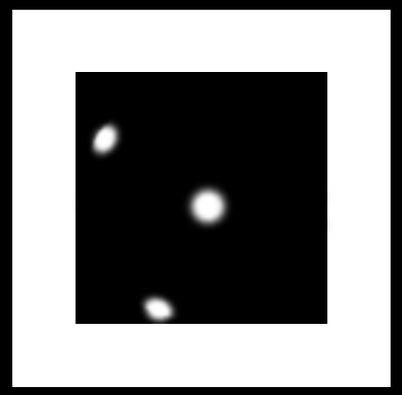
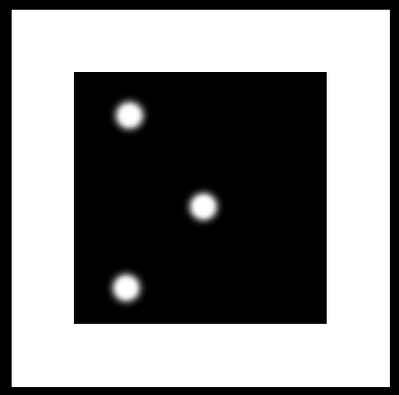


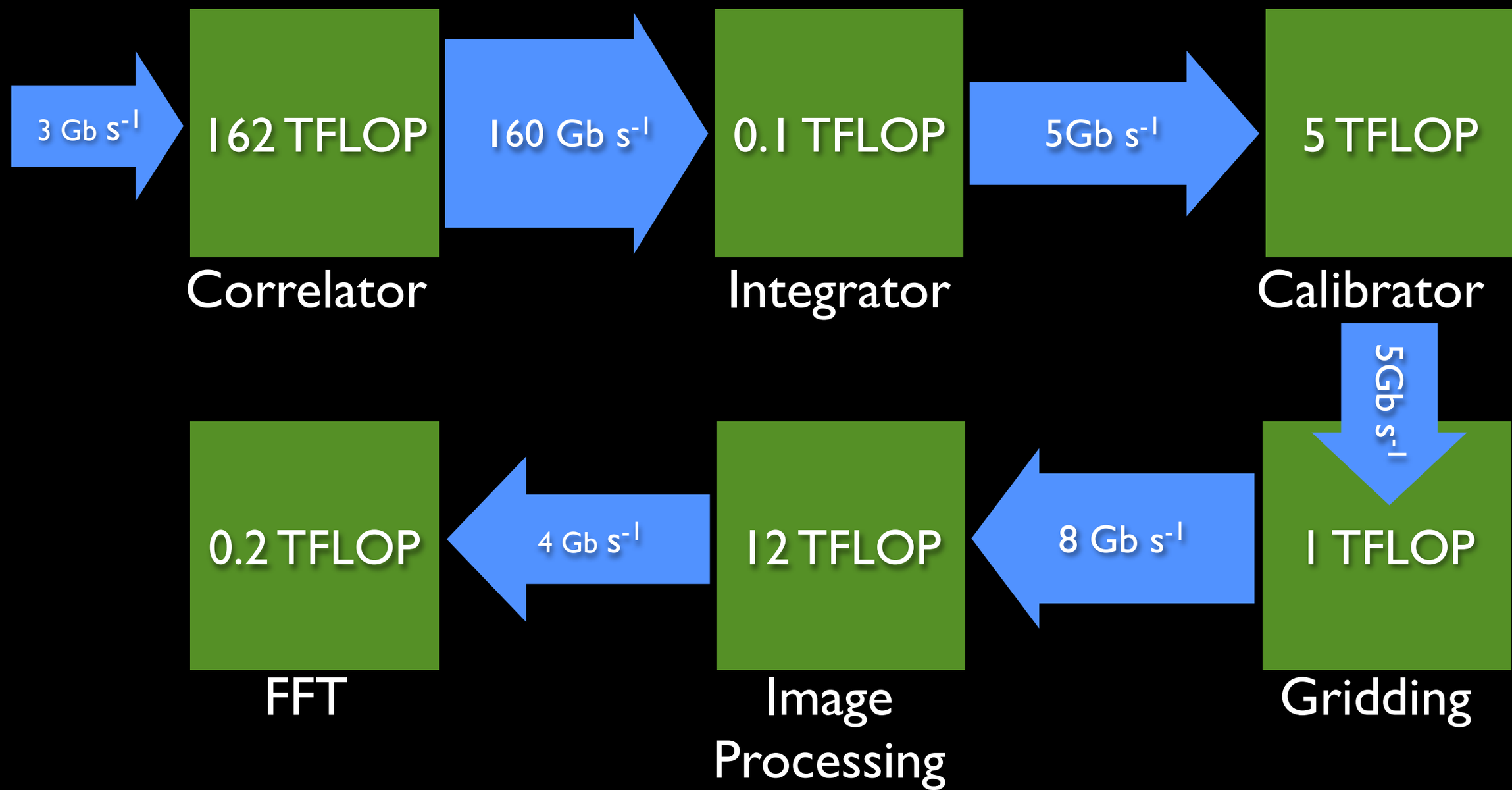
Image Processing



FFT



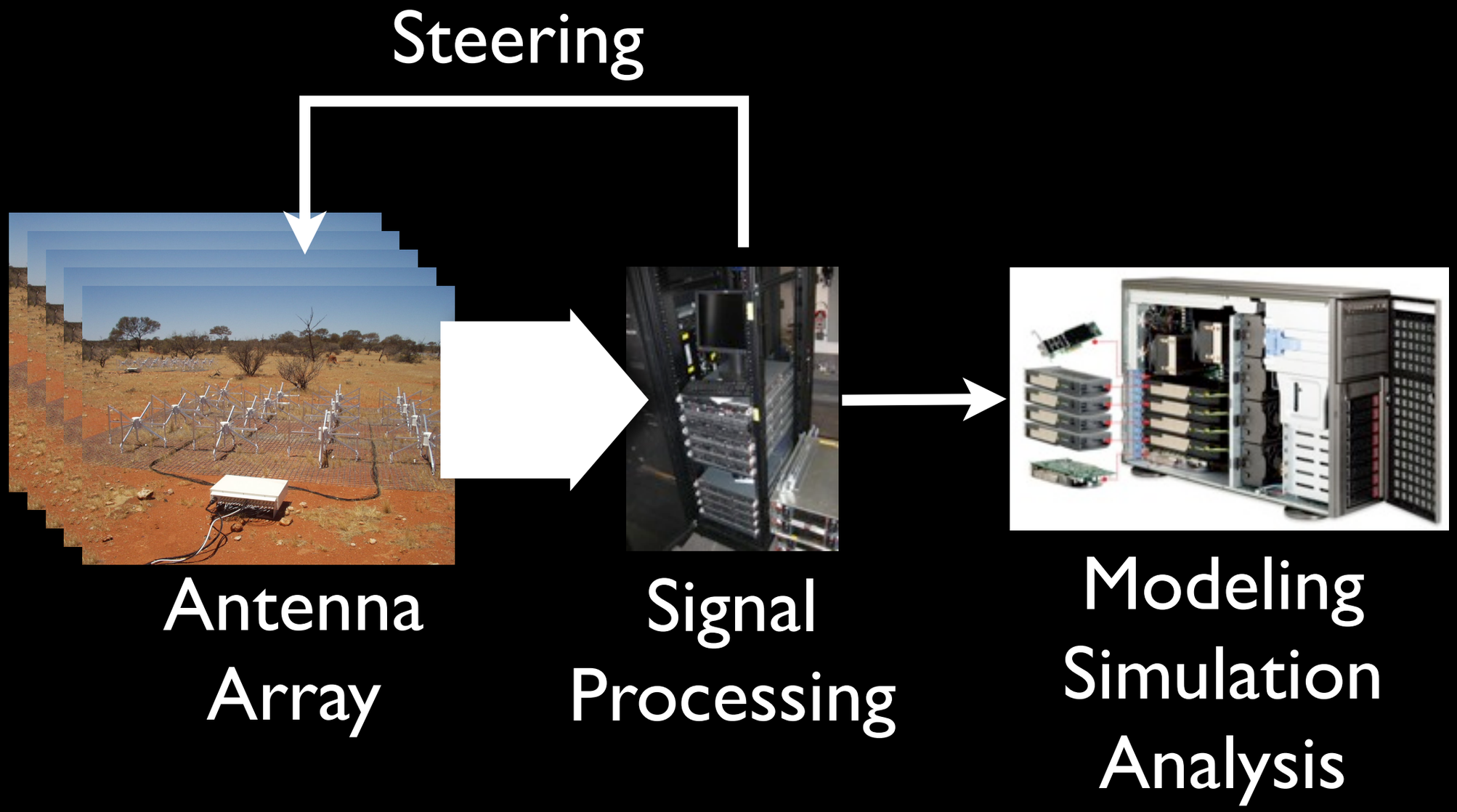
MWA Pipeline



Real Time System

	CPU	GPU
Manufacturer	Intel	NVIDIA
Model	E5472	C1060
# Cores	1	240
Cost	$2500	$4000
Power Consumption	300 W	550 W
Relative Speedup	1x	18.7x
Performance / $	1x	11.7x
Performance / W	1x	10.2x

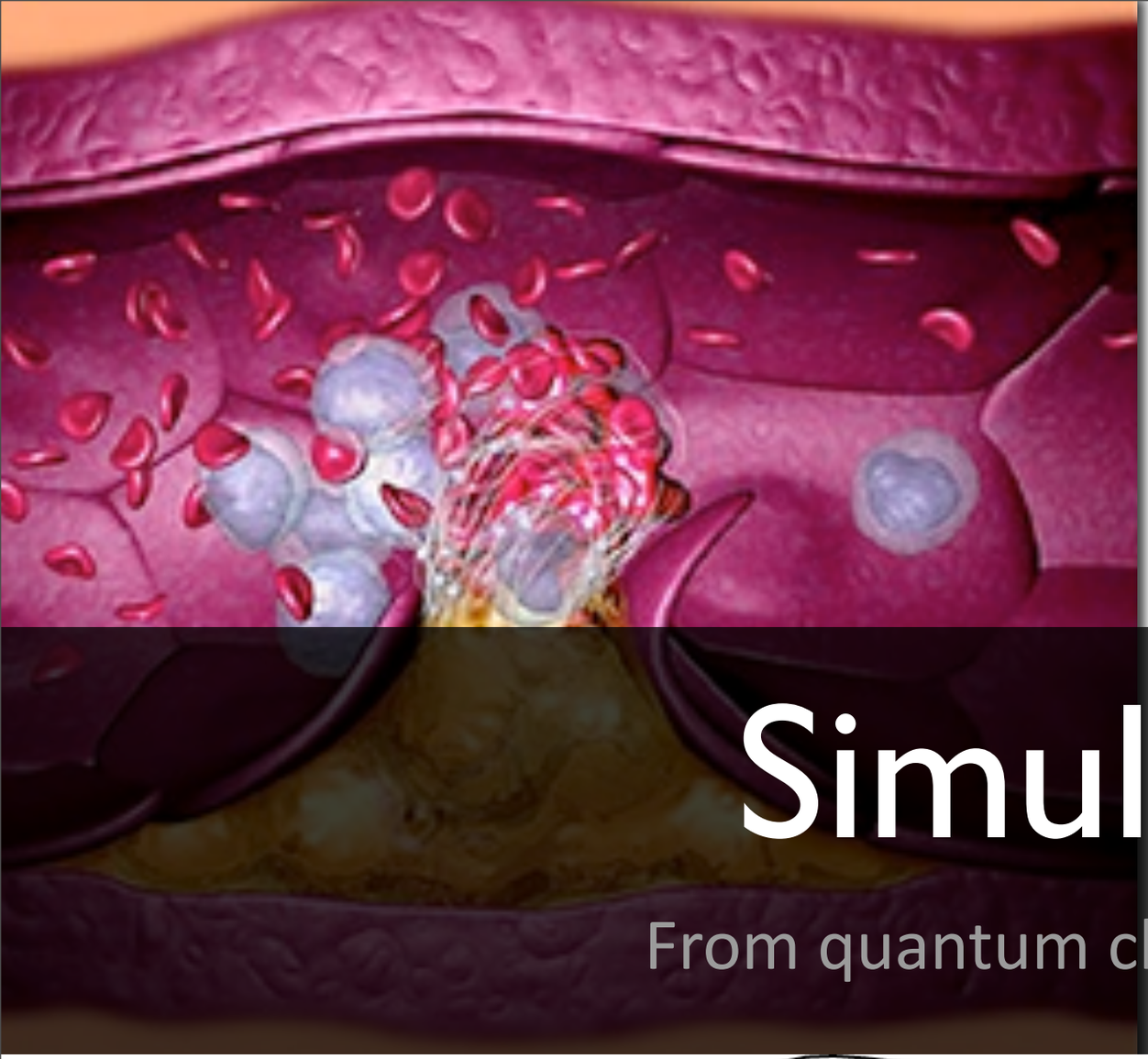
High-Throughput Radio Astronomy

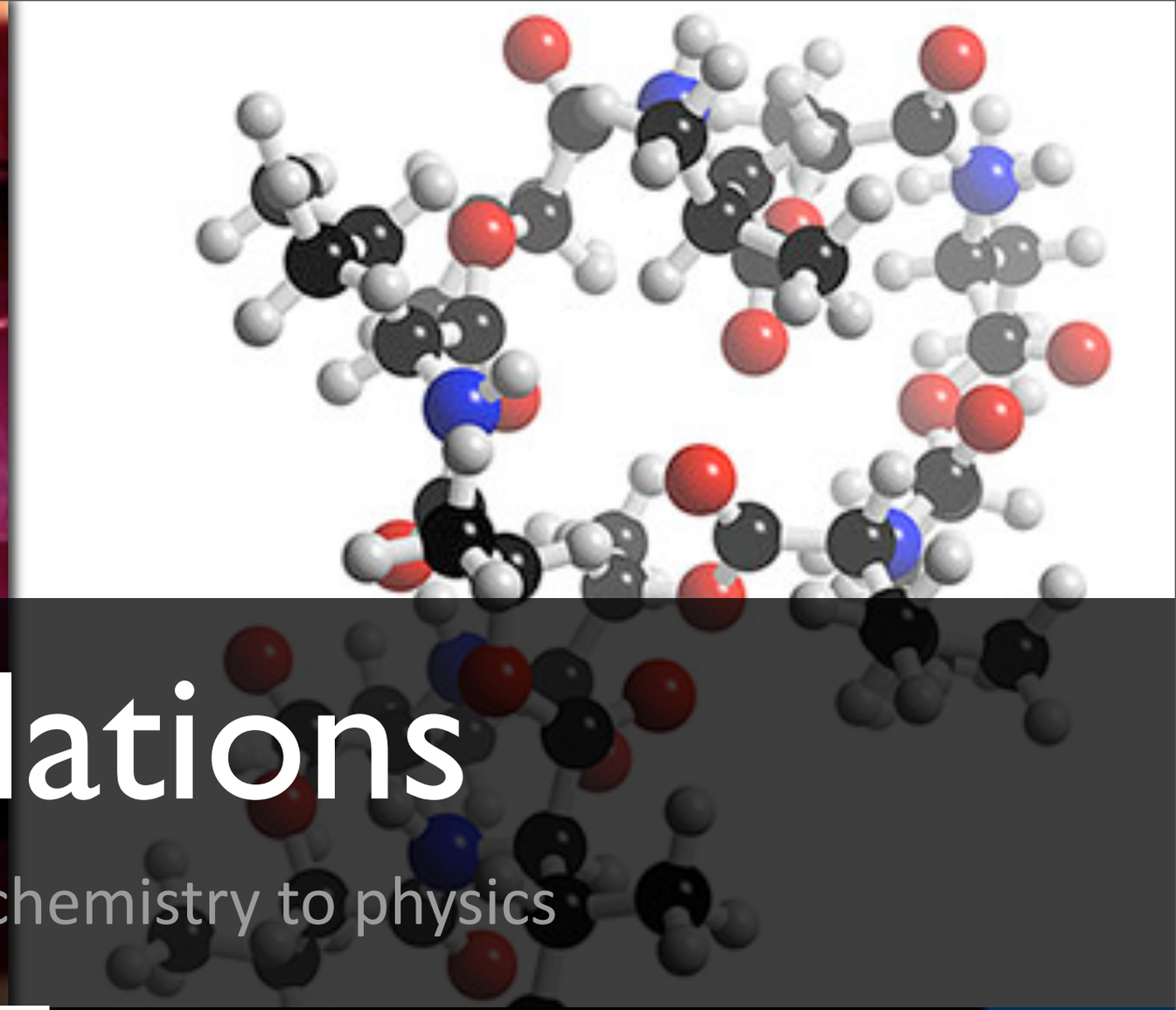


The Future?

Square Kilometer Array, ca. 2020

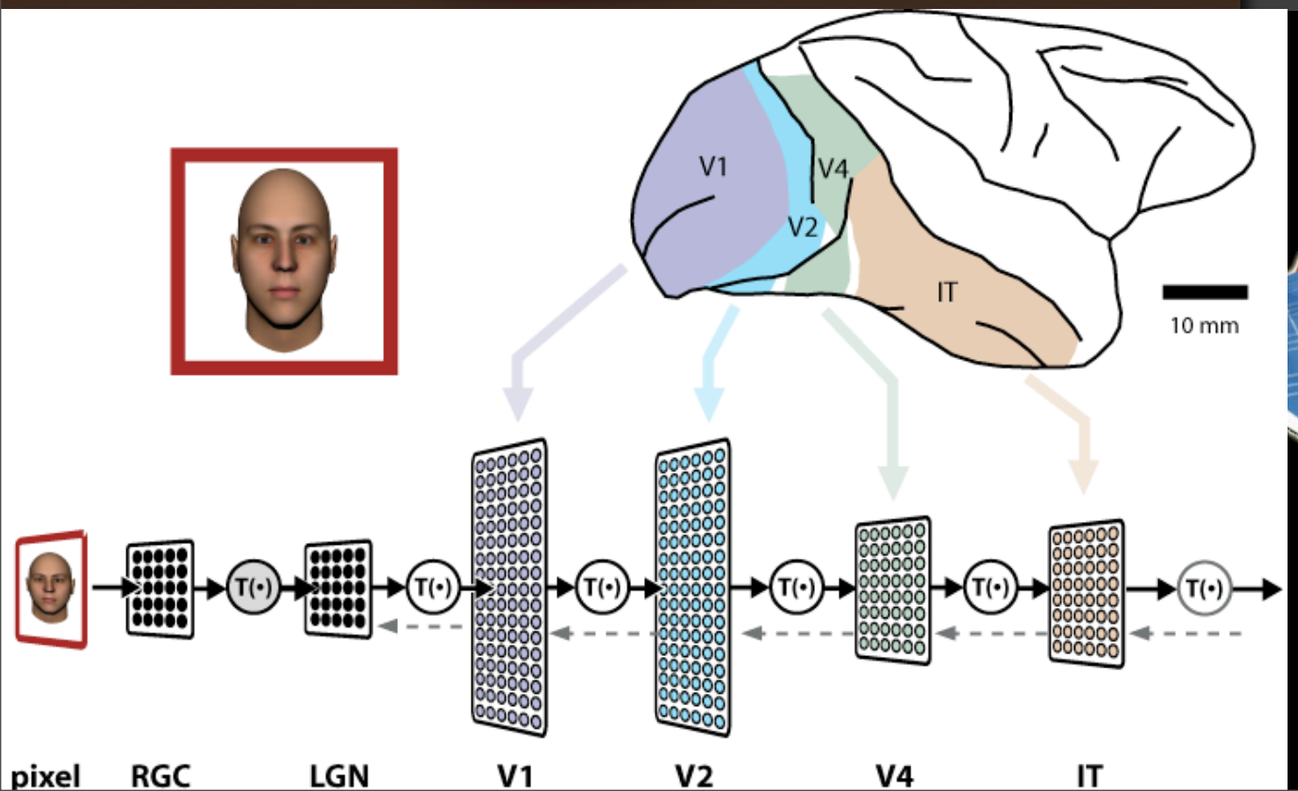
1 EFLOP / 1 Tb s^{-1}

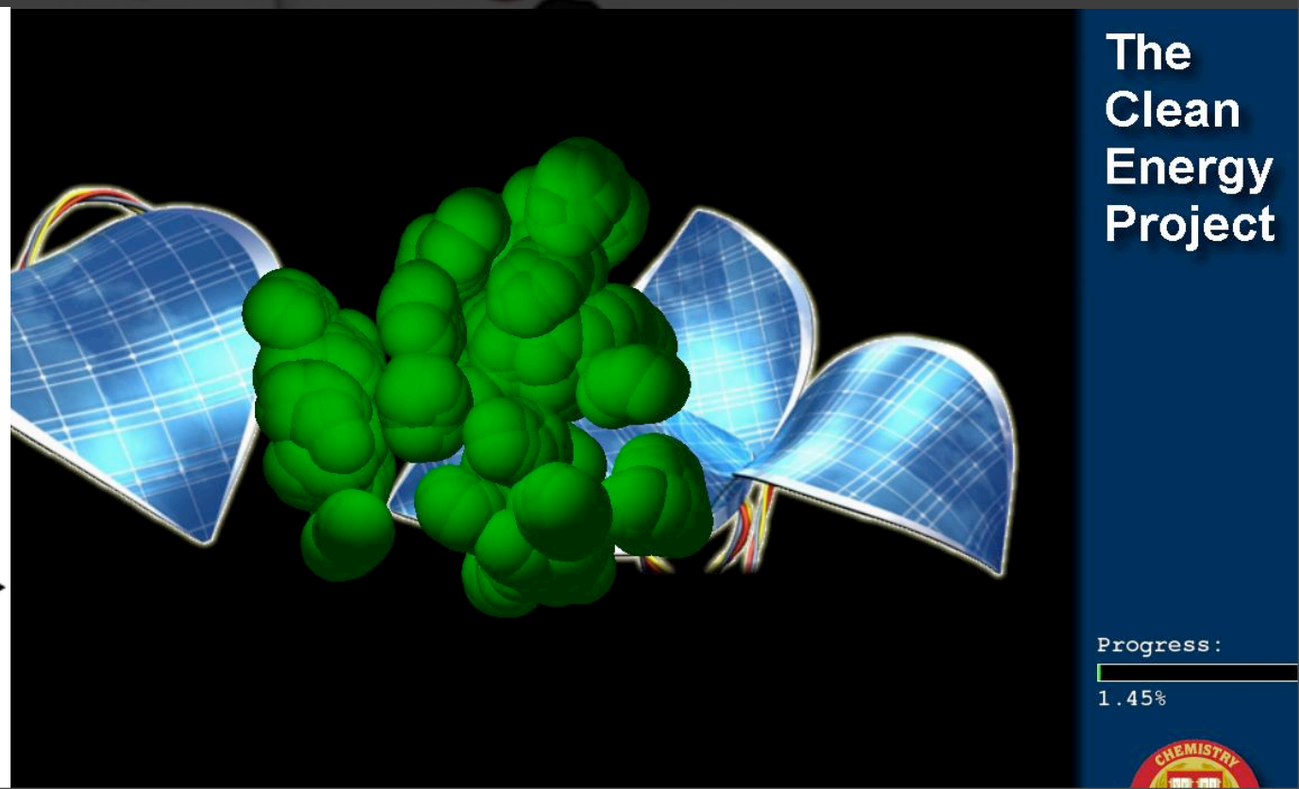




Simulations

From quantum chemistry to physics





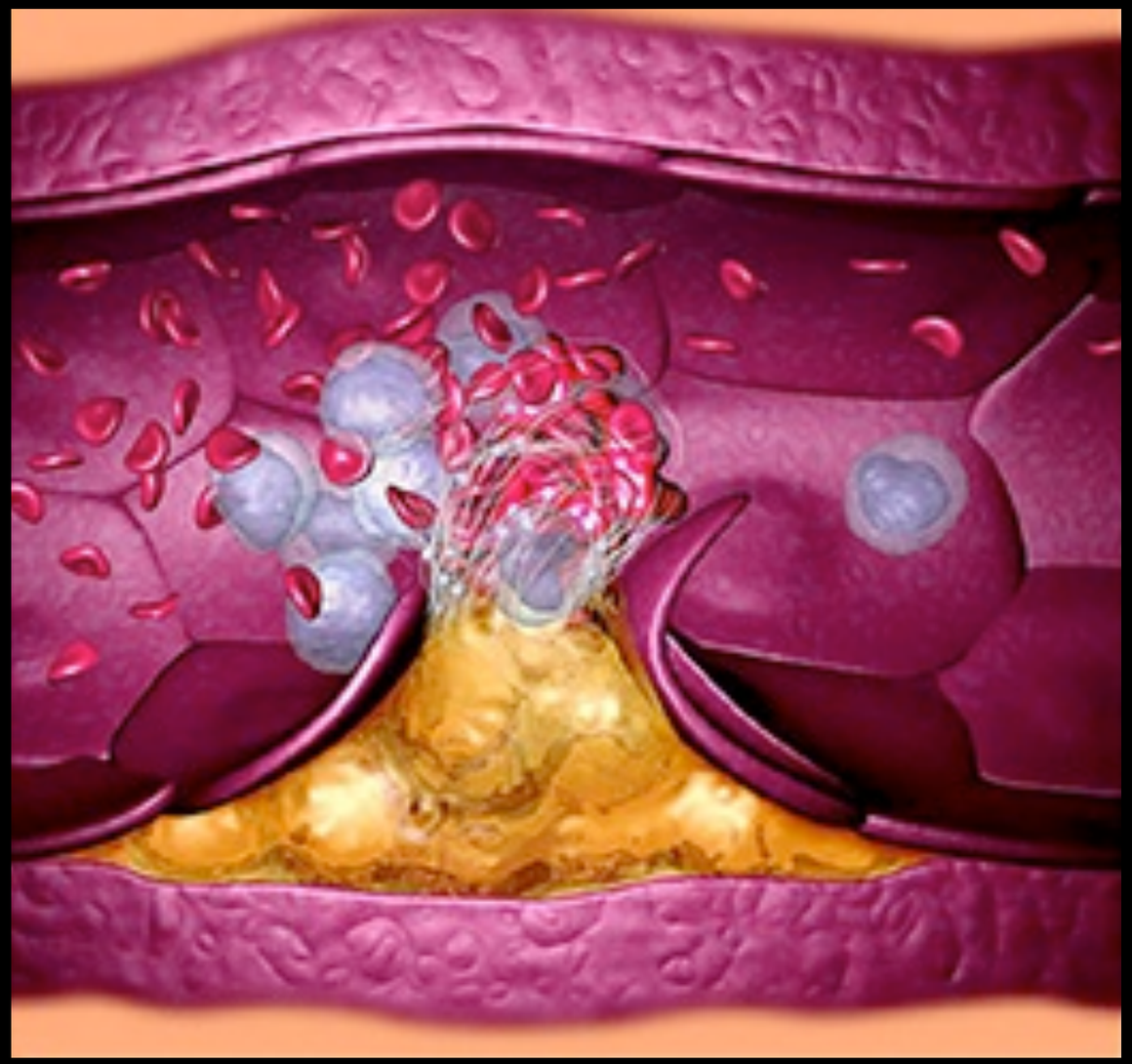
Multiscale Hemodynamics

Massimo Bernaschi - Michelle Borkin - Ahmet Coskun - Charles Feldman - Efthimios Kaxiras - Simone Melchionna - Dimitris Mitsouras - Hanspeter Pfister - Frank Rybicki - Joy Sircar - Michael Steigner - Peter Stone - Sauro Succi - Frederick Welt - Amanda Whitmore

Tim Kaxiras

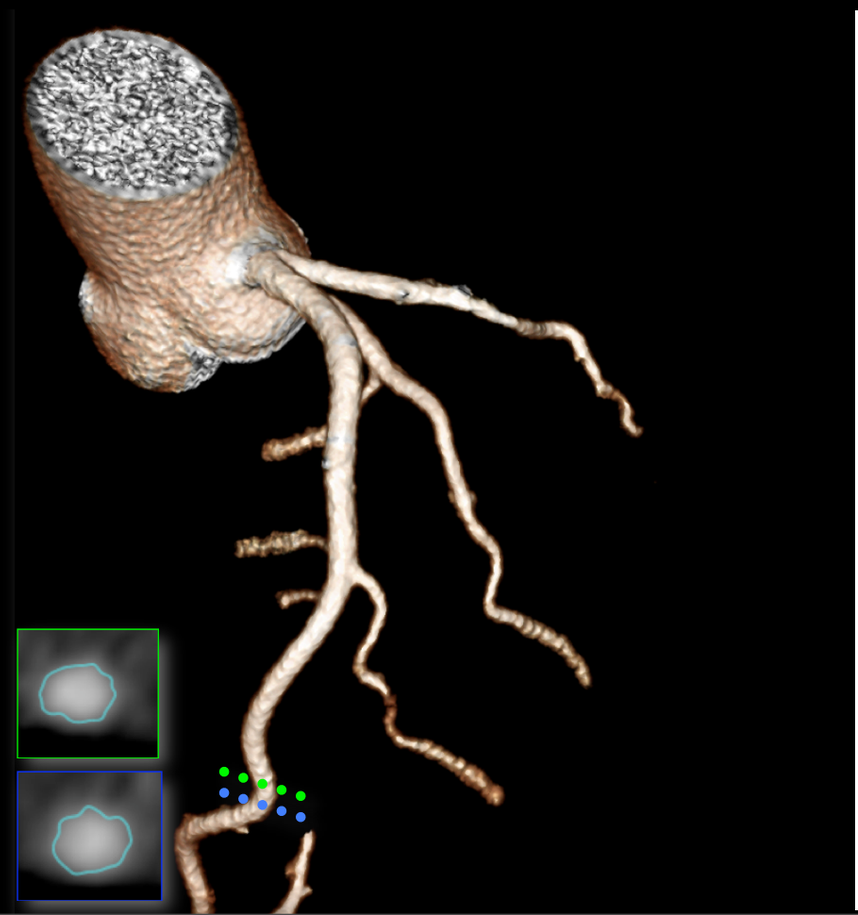
The Scientific Challenge

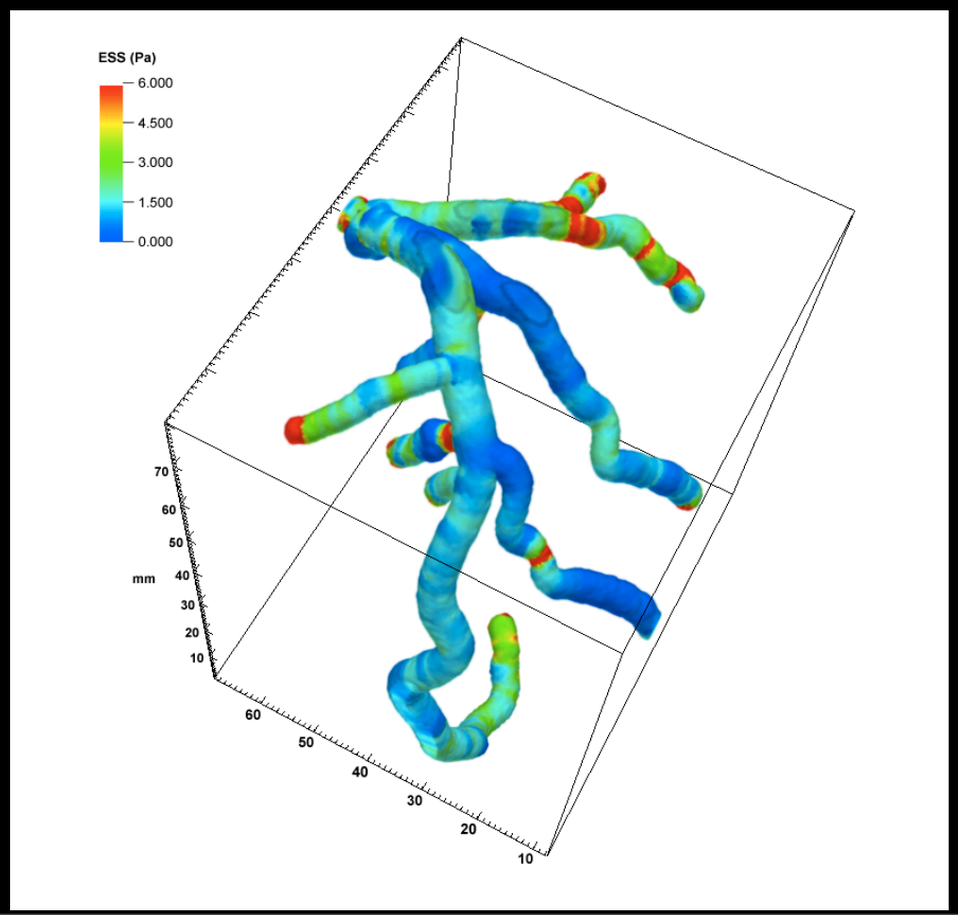
- Predict and prevent heart attacks



The Scientific Challenge

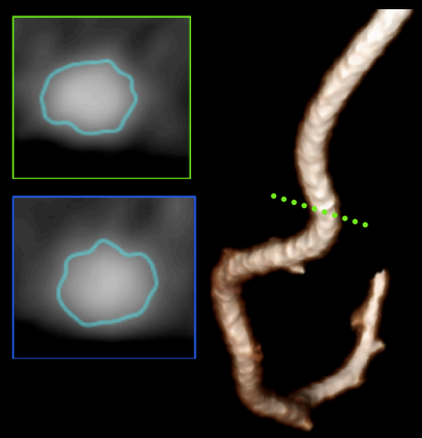
- Predict and prevent heart attacks
- Simulate blood flow in coronary arteries from cellular to plasma level



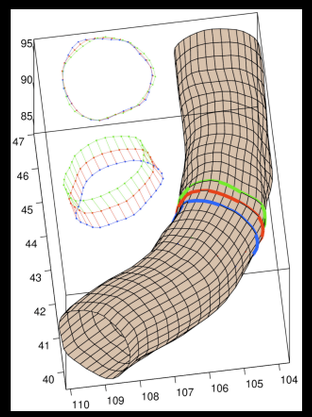


Hemodynamics Pipeline

CT Patient Data →



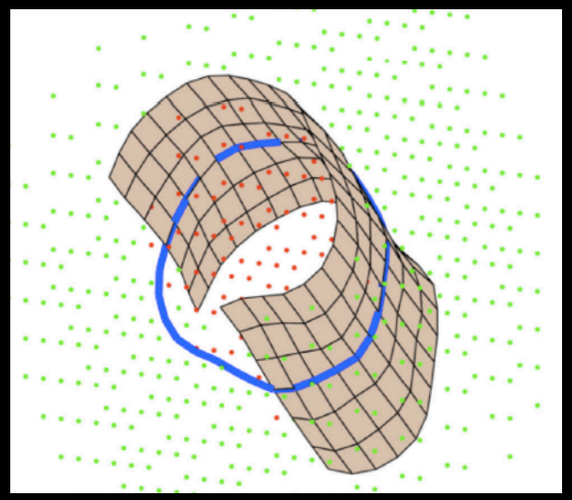
Segmentation →



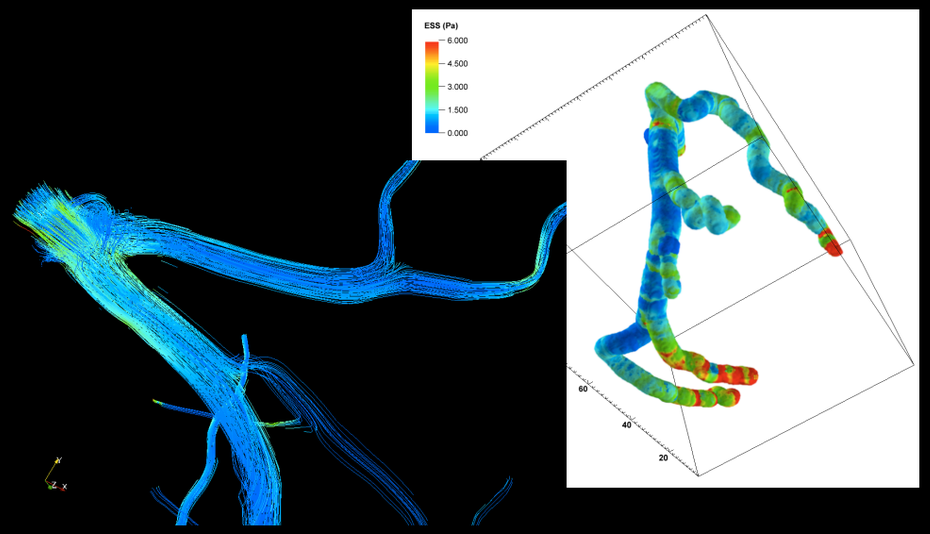
Post-Processing

↓

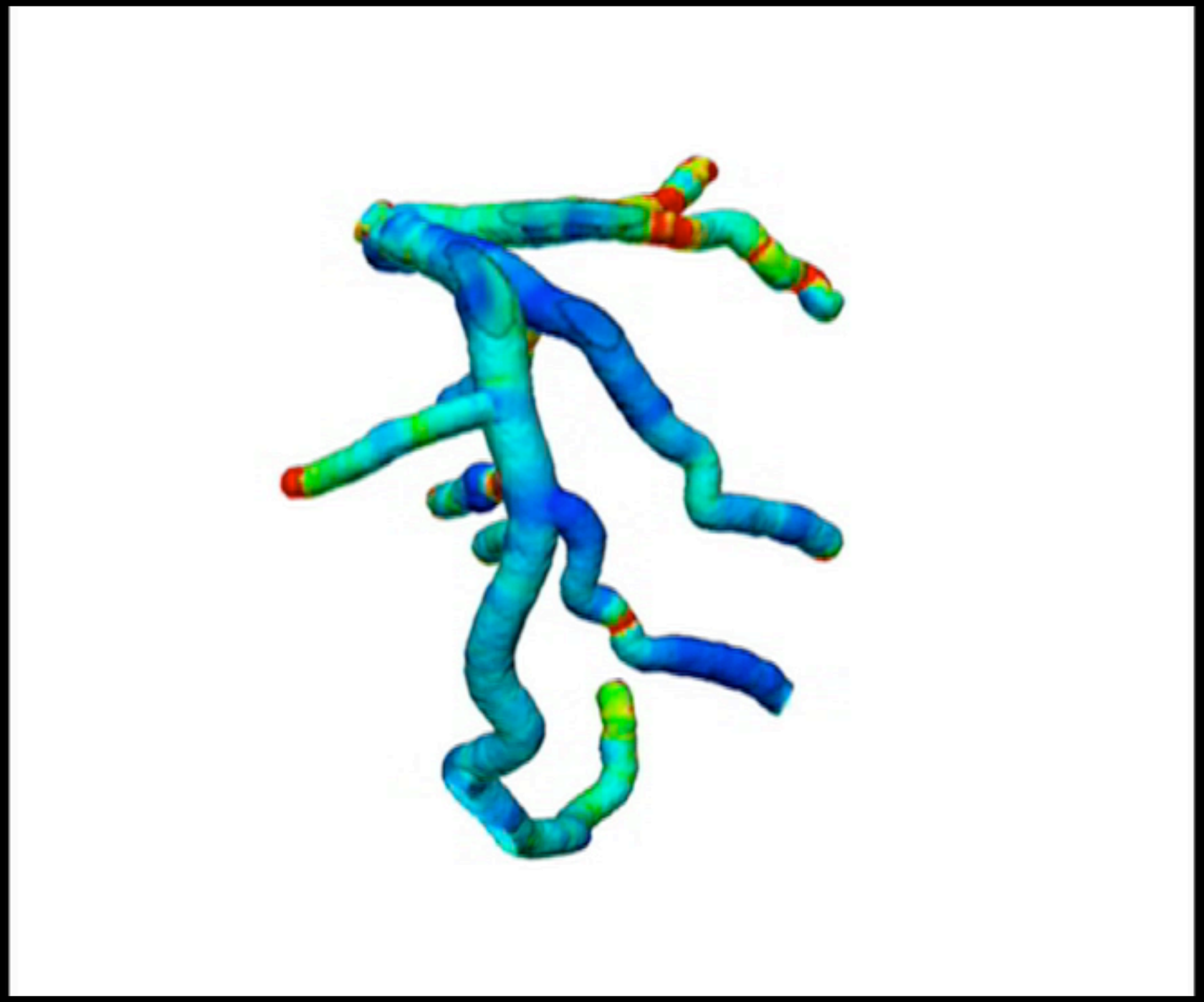
MUPHY Lattice-Boltzman Fluid Simulation



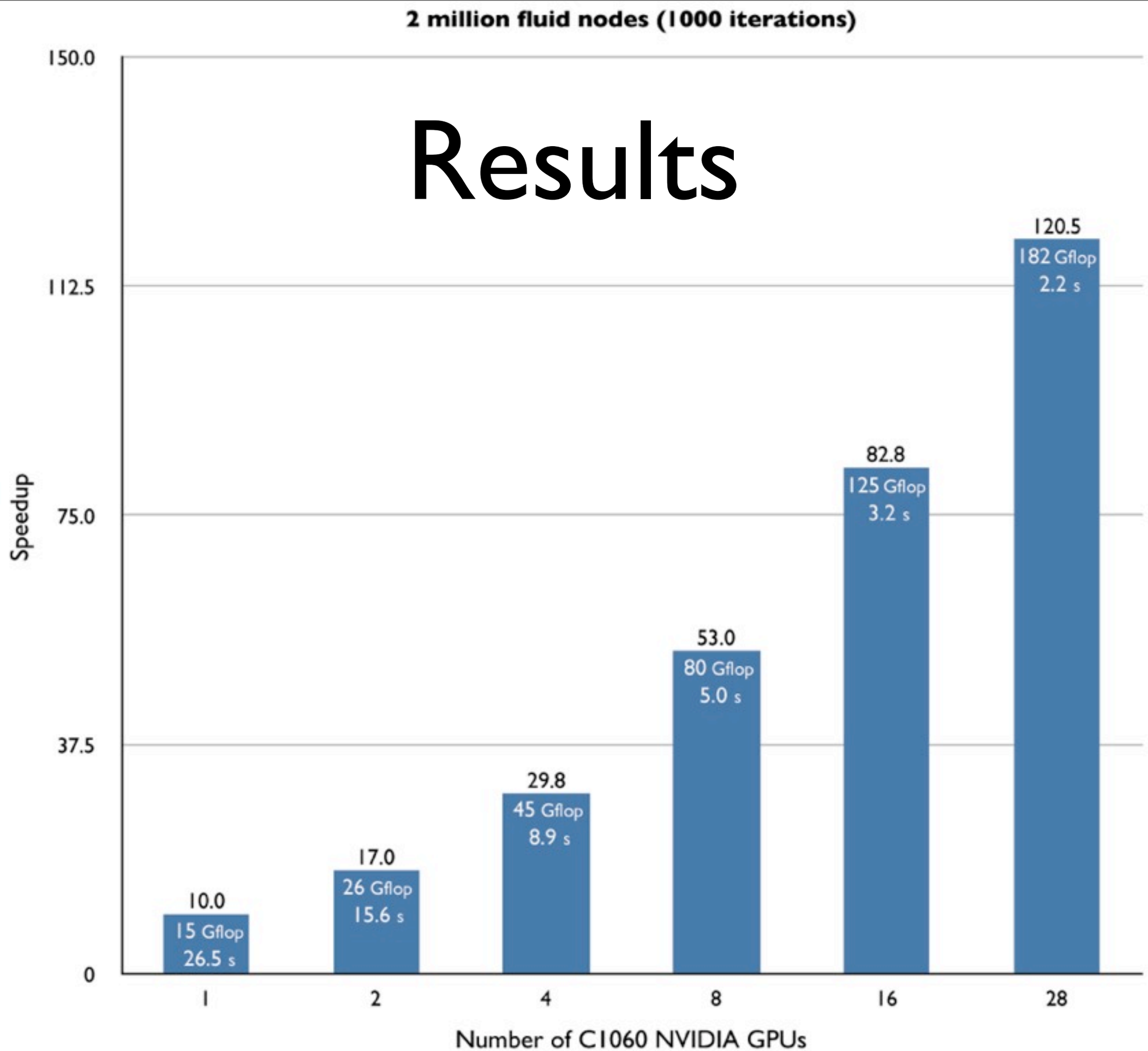
← Visualization & Analysis



MUPHY

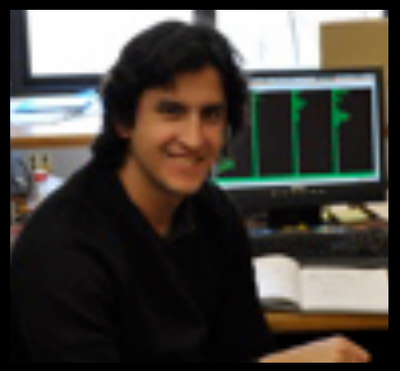


Results


2 million fluid nodes (1000 iterations)
150.0
112.5
75.0
37.5
0
Speedup
10.0
15 Gflop
26.5 s
17.0
26 Gflop
15.6 s
29.8
45 Gflop
8.9 s
53.0
80 Gflop
5.0 s
82.8
125 Gflop
3.2 s
120.5
182 Gflop
2.2 s
1
2
4
8
16
28
Number of C1060 NVIDIA GPUs

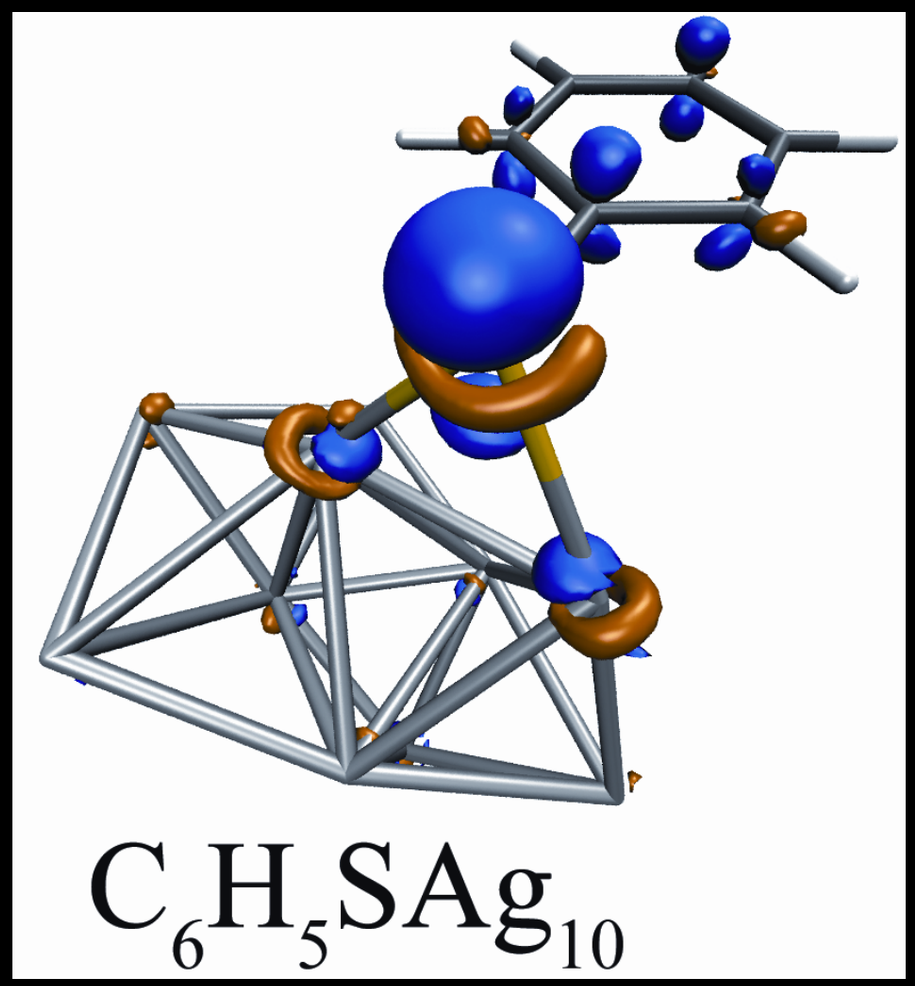
Quantum Chemistry

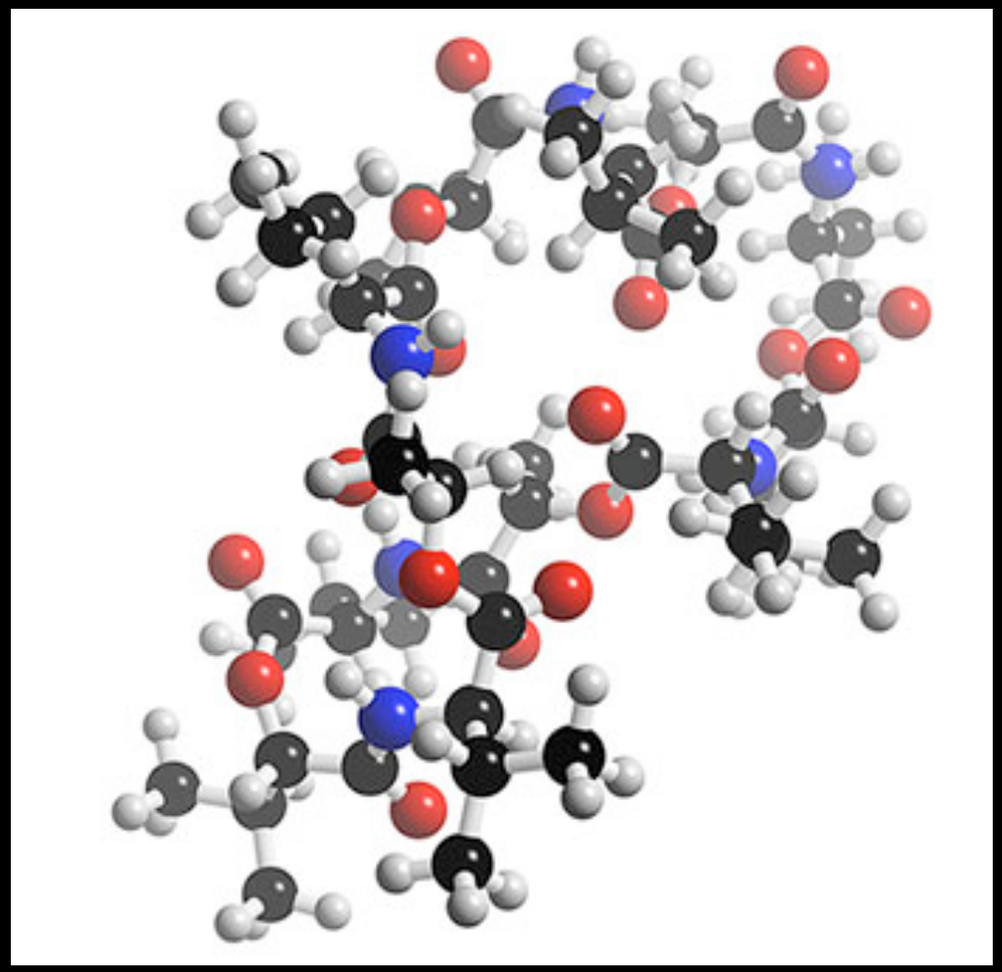
- Alan Aspuru-Guzik
- Mark Watson
- Richard Edgar
- Kenta Hongo
- Roberto Olivares
- Leslie Vogt
- Sean Kermes



The Scientific Challenge

- High-accuracy calculations to determine properties of organic molecules

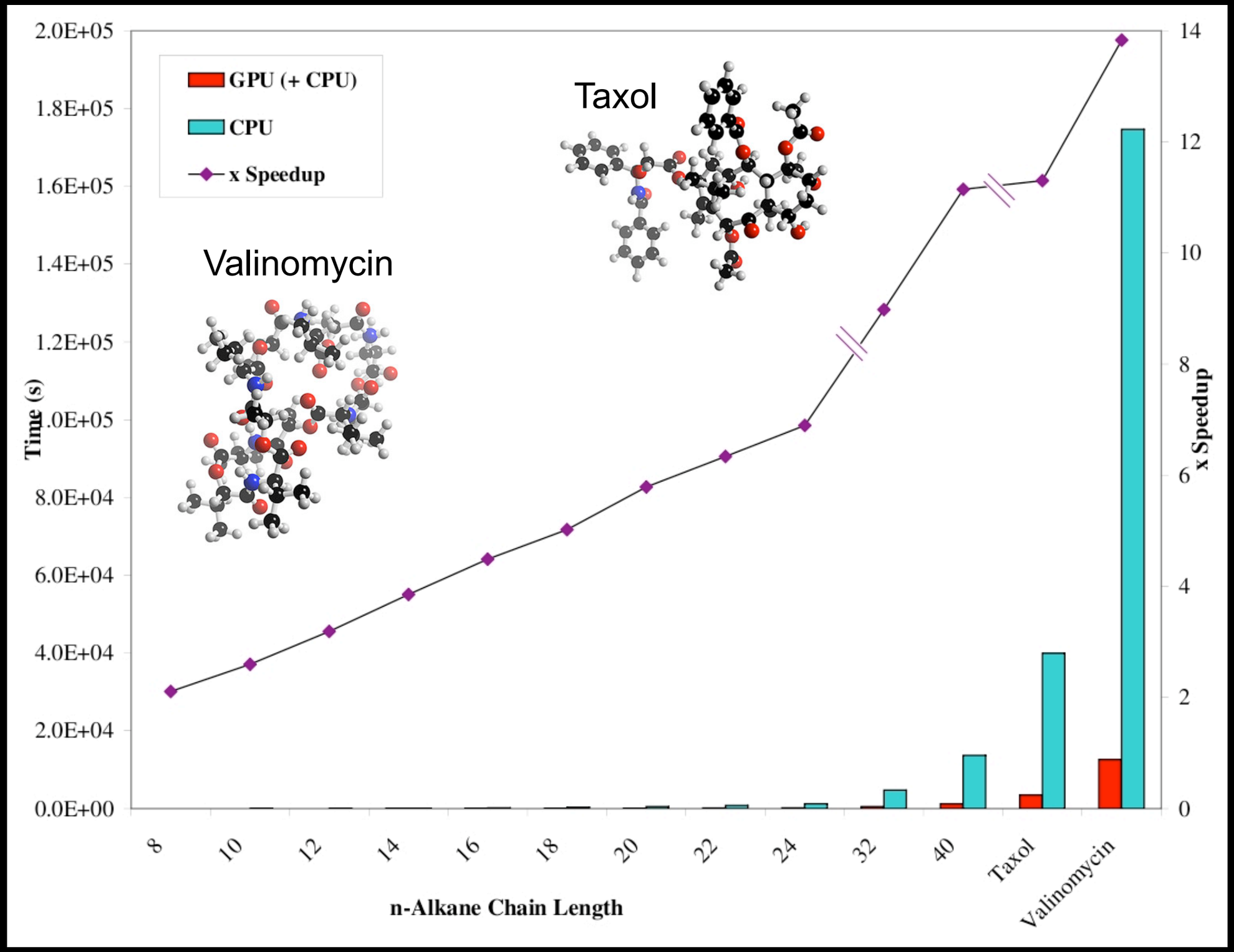




The Scientific Challenge

- High-accuracy calculations to obtain properties of organic molecules
- Improve organic photovoltaic materials

Results


GPU (+ CPU)
CPU
x Speedup
Taxol
Valinomycin
Time (s)
x Speedup
2.0E+05
1.8E+05
1.6E+05
1.4E+05
1.2E+05
1.0E+05
8.0E+04
6.0E+04
4.0E+04
2.0E+04
0.0E+00
14
12
10
8
6
4
2
0
8
10
12
14
16
18
20
22
24
32
40
Taxol
Valinomycin
n-Alkane Chain Length

Visual Neuroscience

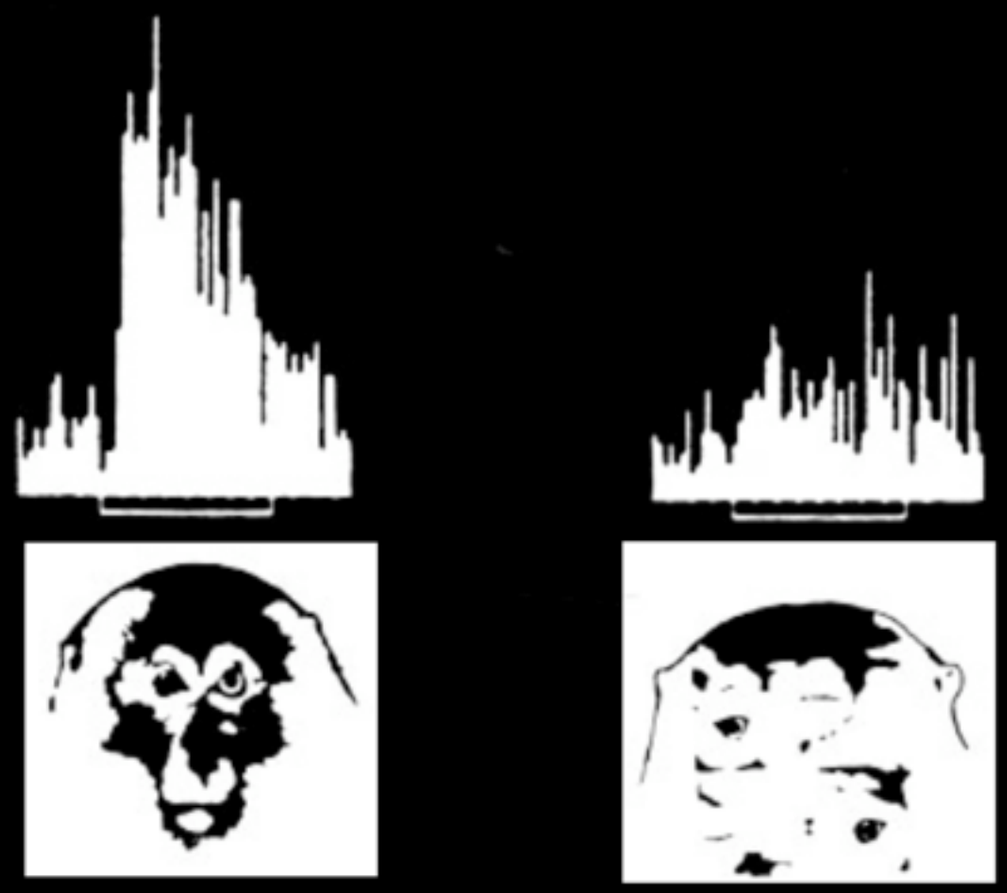
- David Cox, The Rowland Institute at Harvard

- Nicolas Pinto, MIT

- Jim DiCarlo, MIT

The Scientific Challenge

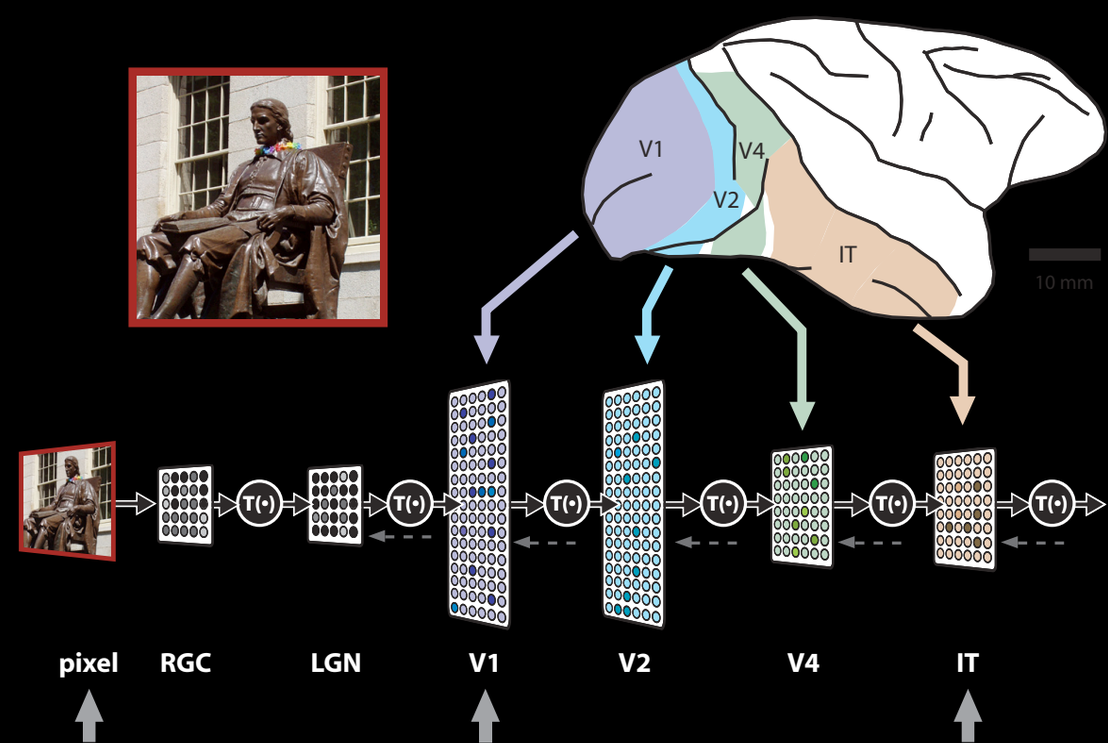
- How does the brain perform object recognition?



Desimone et al. 1984

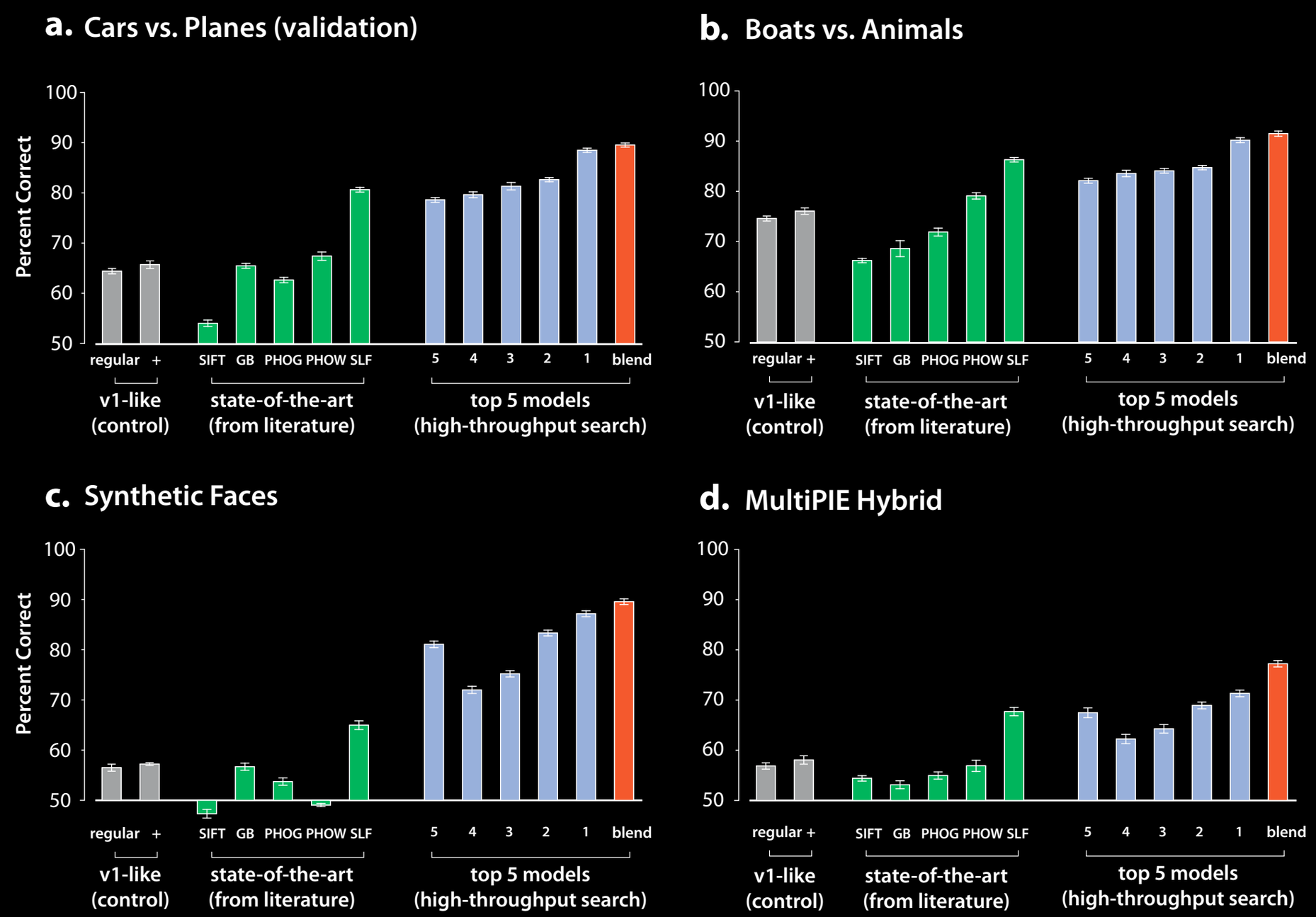
The Scientific Challenge

- How does the brain perform object recognition?
- What are good models of the human visual system, and what are their parameters?



Simulations

Test 10,000 to 100,000 different models



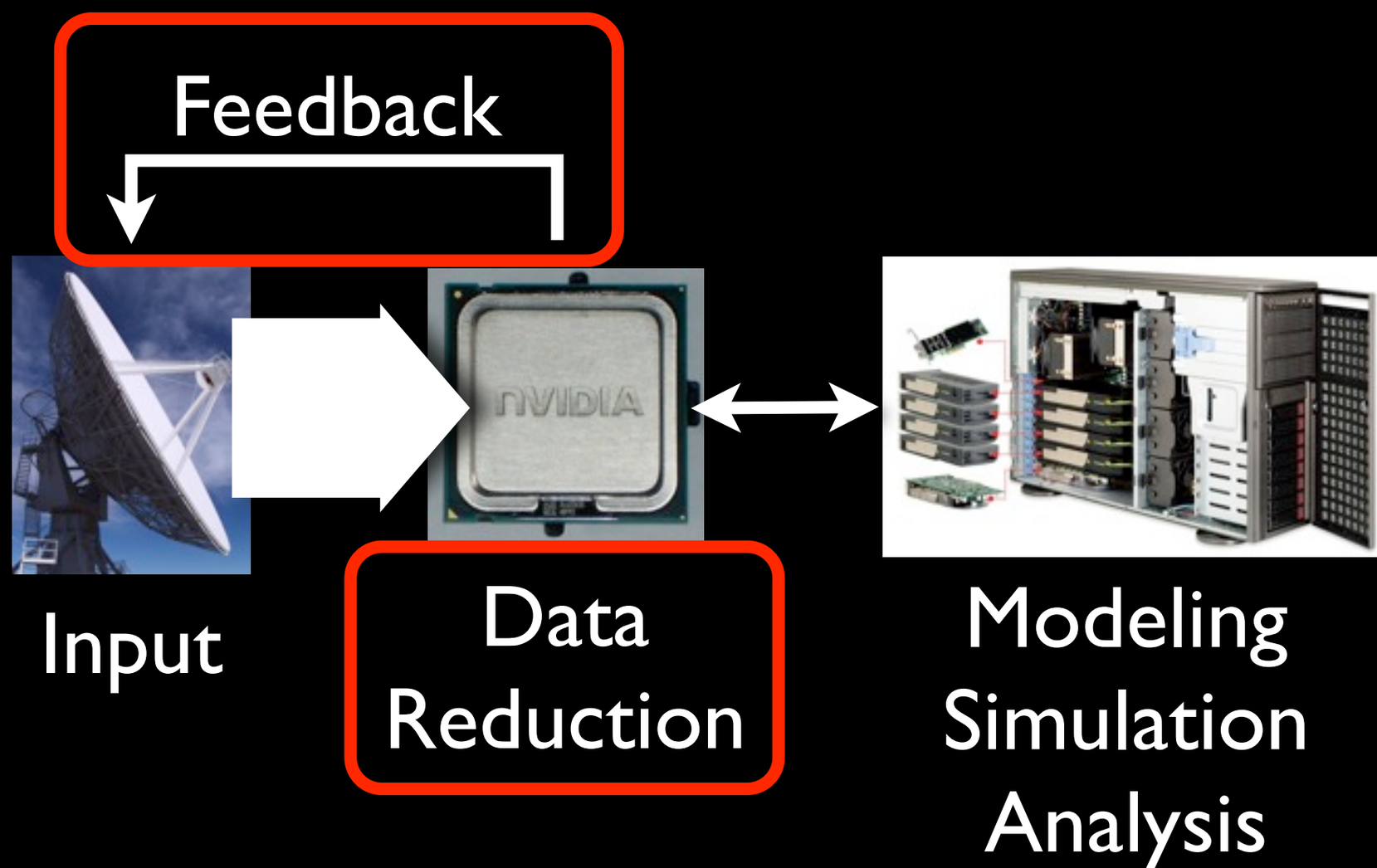
Results

3D Filterbank Convolution

	CPUs			GPUs		
Hardware						
Manufacturer	Intel	Intel	Intel	NVIDIA	NVIDIA	NVIDIA
Model	**Q9450**	**Q9450**	**Q9450**	**7900 GTX**	**8800 GTX**	**GTX 280**
# cores used	1	4	4	4x96	4x128	4x240
Implementation	MATLAB	MATLAB	SSE2	Cg	CUDA	CUDA
Year	2008	2008	2008	2006	2007	2008
Performance / Cost						
Full System Cost (approx.)	$1,500**	$2,700**	$1,000	$3,000*	$3,000*	$3,000*
Relative Speedup	**1x**	**4x**	**80x**	**544x**	**1544x**	**2712x**
Relative Perf. / $	**1x**	**2x**	**120x**	**272x**	**772x**	**1356x**

Next Steps

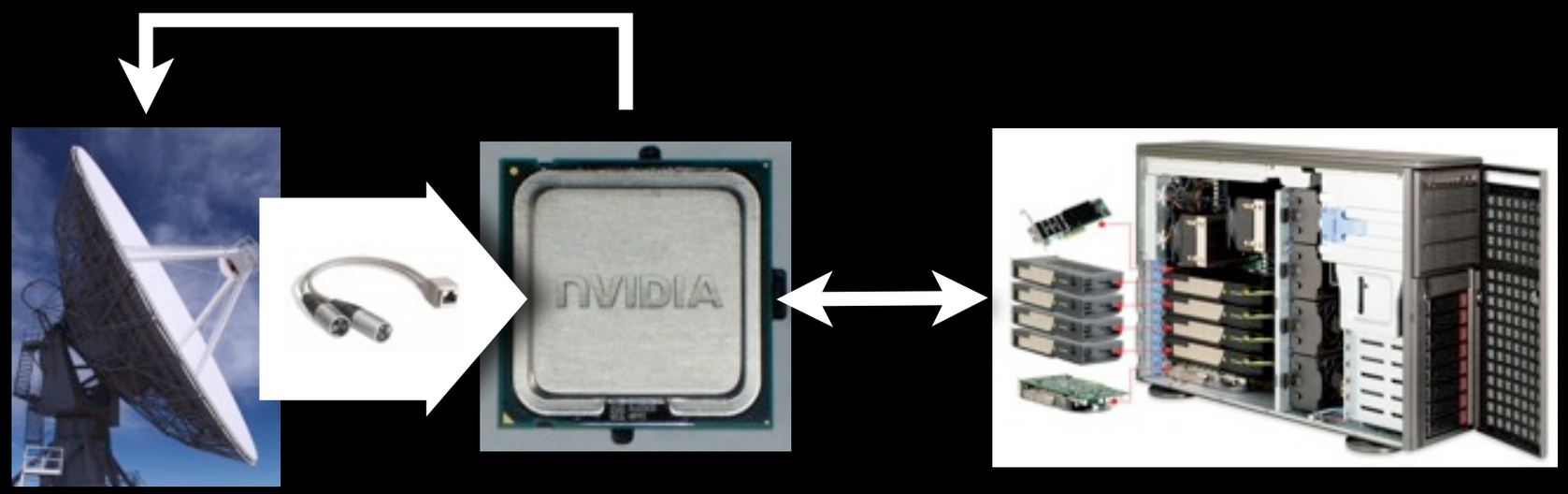
High-Throughput Science



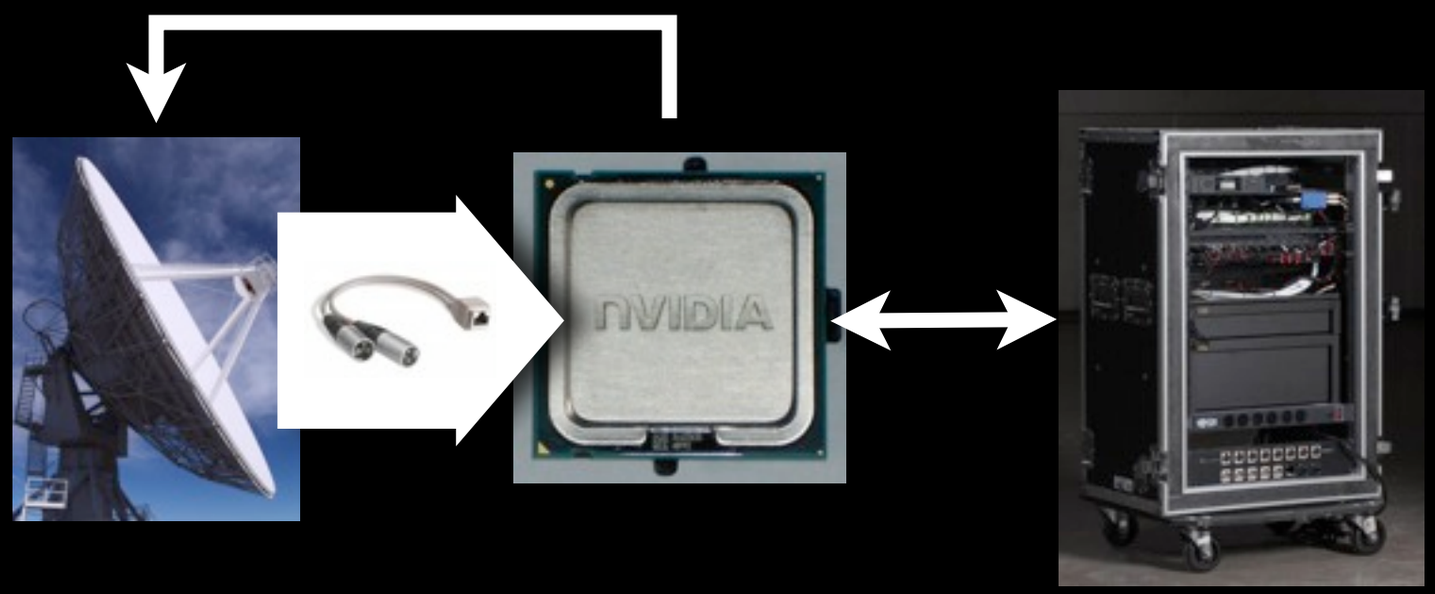
General Programmability

- Higher-level programming models
- Domain Specific Languages (DSLs)
- Easier scalability

Standardized High-Bandwidth I/O

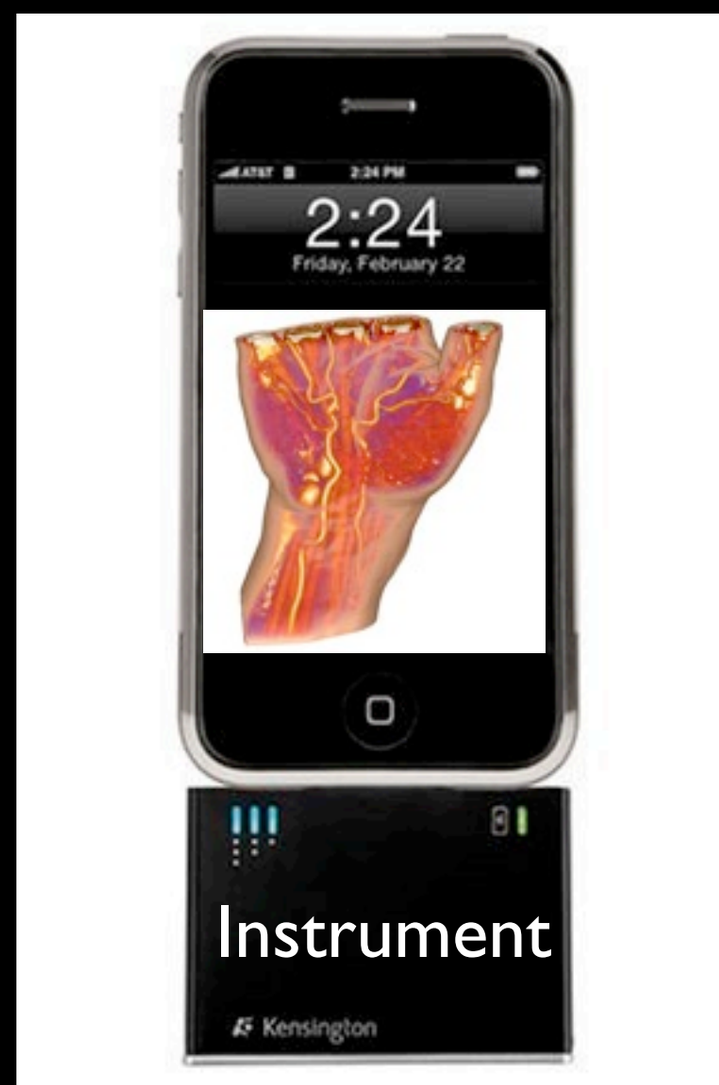


Compute Appliances



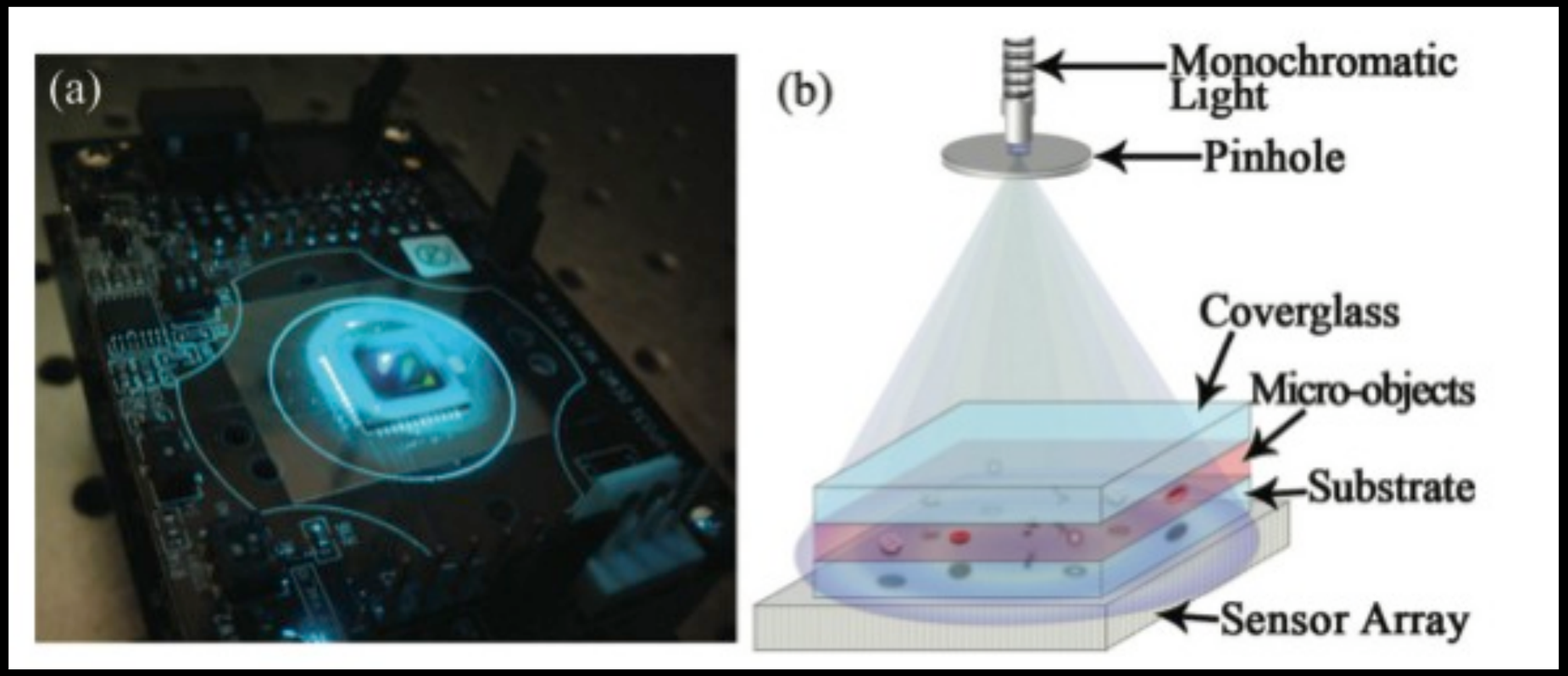
The Future?

Tricorders?



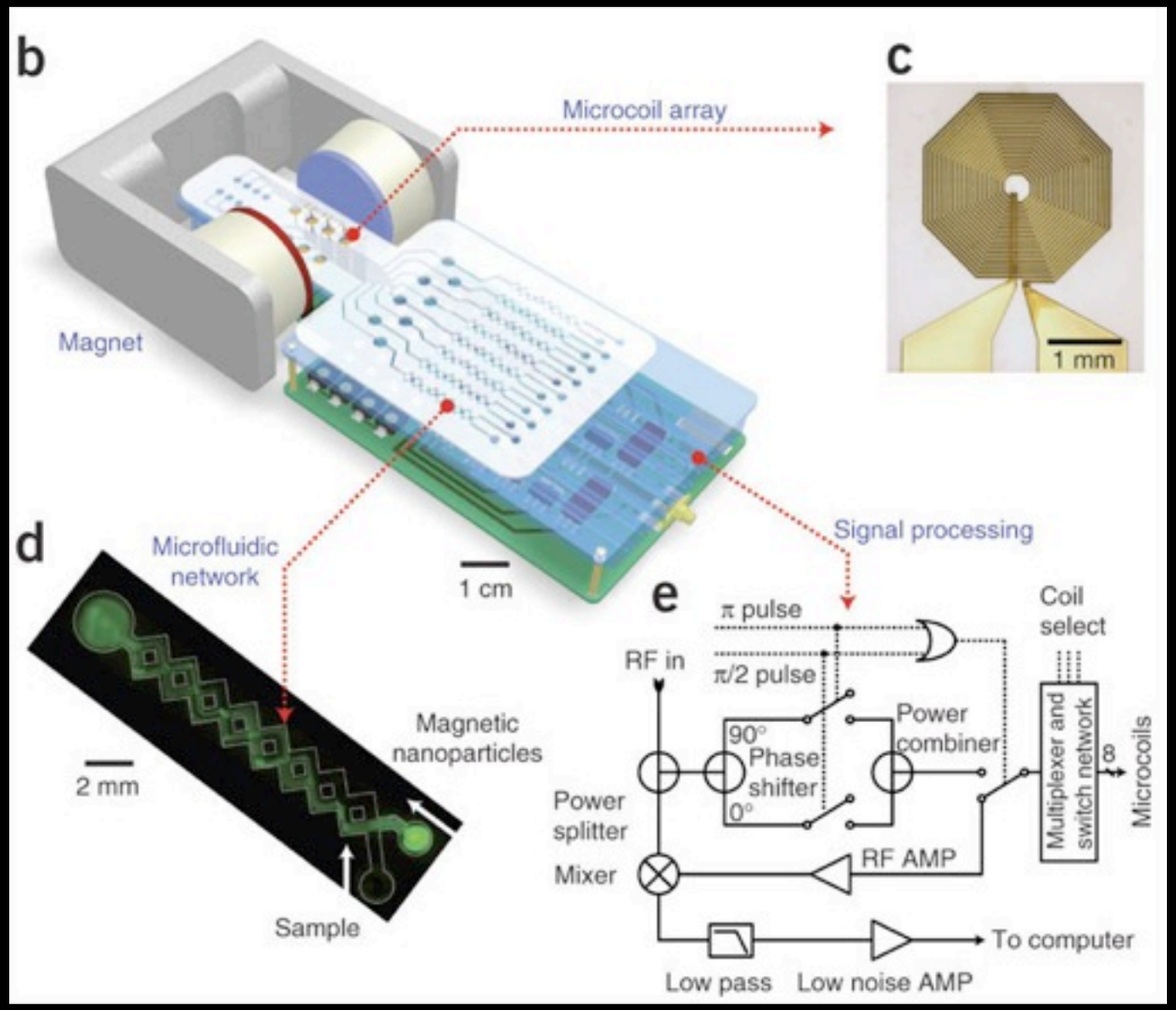
Chip Microscopy

- Sungkyu Seo, Ting-Wei Su, Derek K. Tseng, Anthony Erlinger and Aydogan Ozcanabc, UCLA



Chip-NMR Biosensor

- Hakho Lee, Eric Sun, Donhee Ham & Ralph Weissleder, SEAS & HMS



Wish List

- Better programming models
- Standardized high-bandwidth I/O
- High-throughput compute appliances

Acknowledgements

- Microsoft
- NVIDIA
- NSF CDI PHY-0835713, Austrian Research Promotion Agency FFG, Vienna Science and Technology Fund WWTF