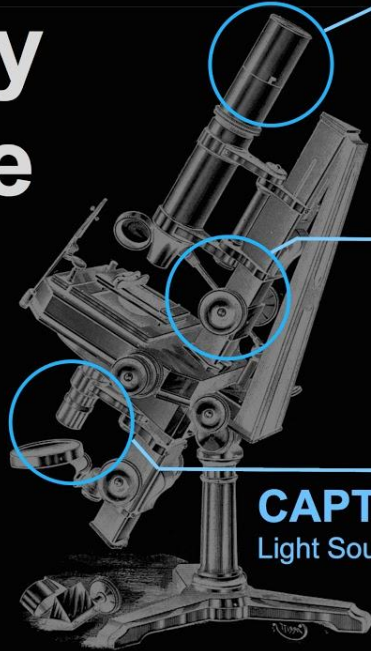


# Cryo-Electron Microscopy

Architecting the process

## 21<sup>st</sup> Century Microscope



### INSIGHT

Lens, Viewfinder



### ANALYSIS

Filters, Adjustments



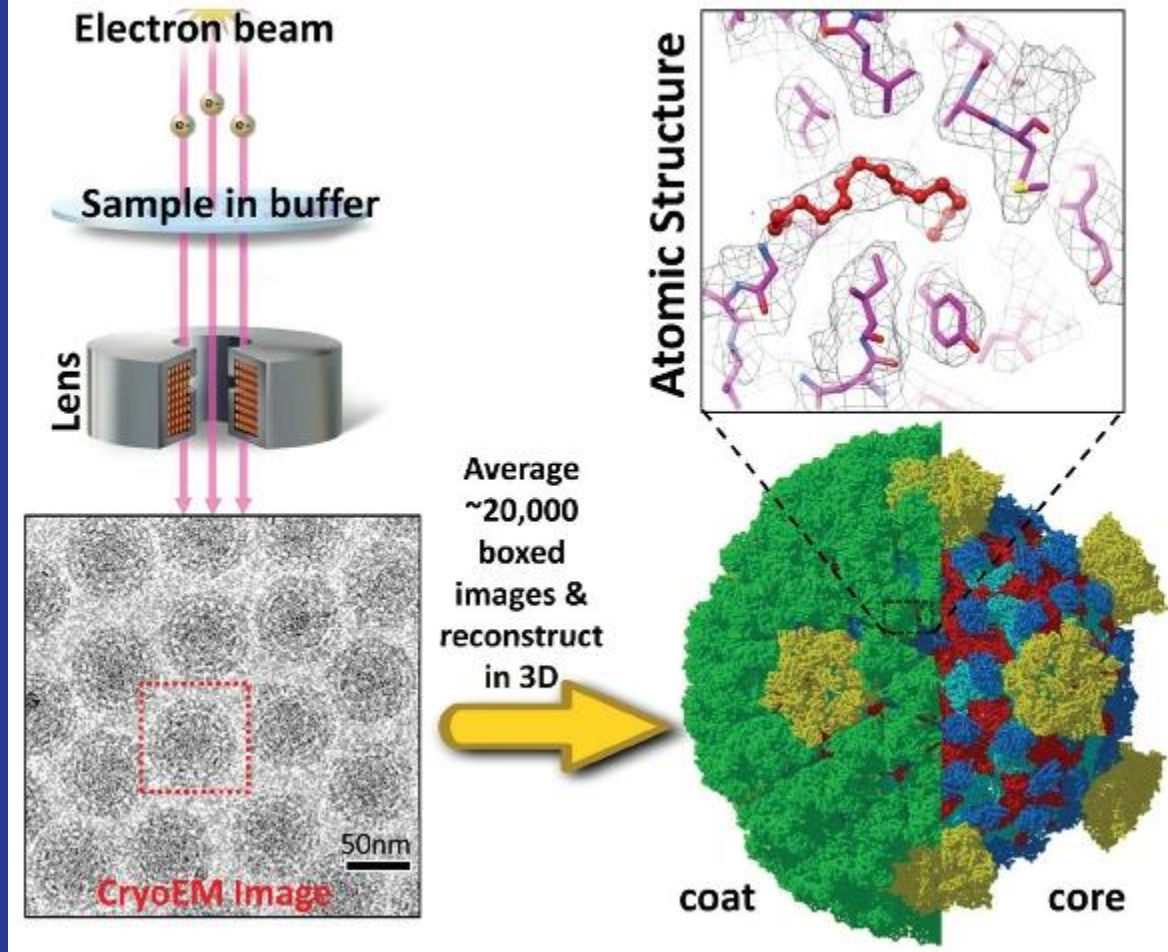
### CAPTURE

Light Source, Samples





# What is Cryo-Electron Microscopy?





What is the scope?

# Compute and Storage Requirements

- ~1TB data set/sample ~400-800 files
- Pipeline analysis with internal external tools
- Require large memory gpu > 8GB
- Require large system memory > 64GB
- Require cpu cores 200 - 400
- Parallel file reads and writes



# Options for Computation

## Workstation

Pro

Full user control

Con

Limited by single machine

## Cloud

Pro

Scales easily

Con

Cost, complexity, data movement

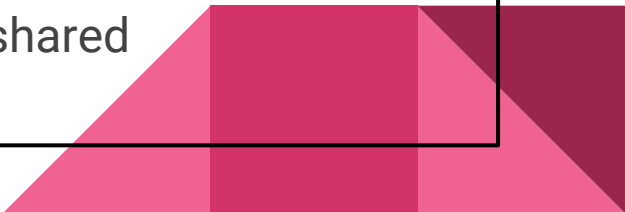
## HPC

Pro

Huge resources

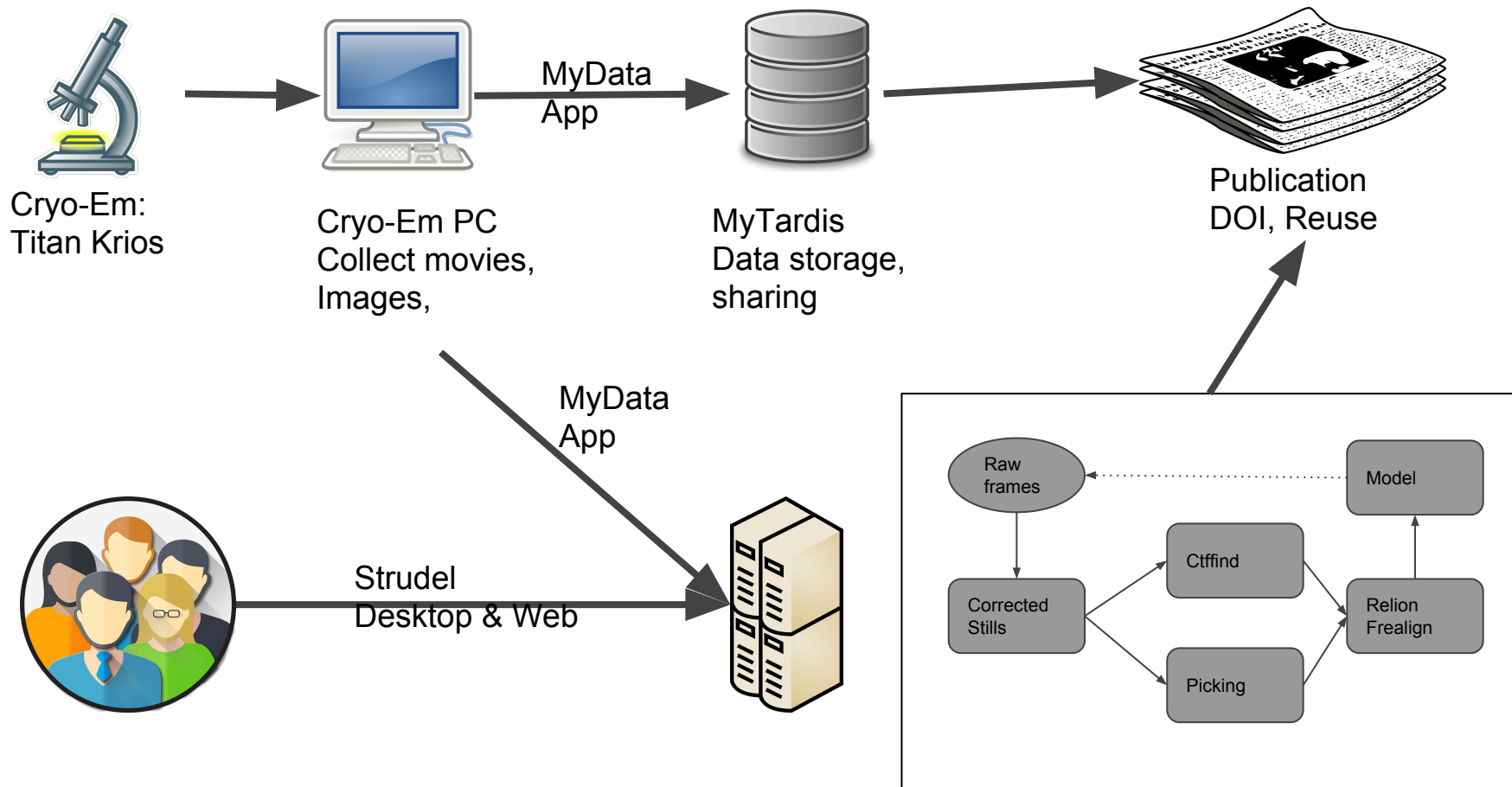
Con

Tightly controlled, shared





Sounds like a HPC problem ... or  
is it?







Where can we help?

# Motion Correction

The electron beam drifts during collection and the results need to be shifted to account for it.

<http://www.ncbi.nlm.nih.gov/pubmed/23644547>

Software: motioncorr 2.1

3.4 GB of raw movie data (15 movies)

218MB of corrected micrographs (15 images)

Processing Time:- 109s

Single Nvidia K80

How to achieve maximum processing speed?

# GPUs = # raw movies

Storage I/O > Processing I/O



# Case study - Motion Correction 10x speedup!

Using local desktop :- ~3hrs

- Limited by local storage and network access, single gpu

Using remote dekstop on MASSIVE:- ~45mins

- Limited by GPU (2 per desktop)

Using an inhouse parallel scripted version:- ~4.5mins

- Limited by file system bandwidth (1-2 GB/s)



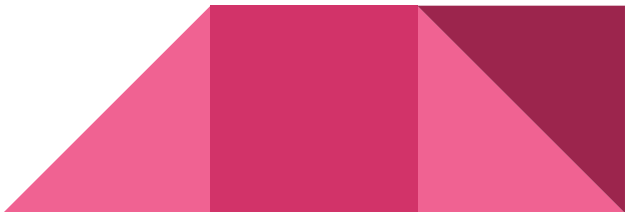


So ... what is the solution?

# Solution

- GPU cluster
- Large high performance file system
- Remote desktop with access to cluster (for GPU and storage)
- HPC and domain experts optimising pipeline for systems

# Outcome

- Data processing faster than collection!
  - Shared resource for optimal use
- 



Questions?