BioDAC:
Challenges of Bio-image Processing at Exascale

Matthew Archer[1,2,3] and Anita Karsa[1,2]

[1]Centre for Advanced Image Analysis (CAIC), University of Cambridge

[2]Dept. of Physiology, Development and Neuroscience, University of Cambridge

[3]University Information Services (UIS), University of Cambridge

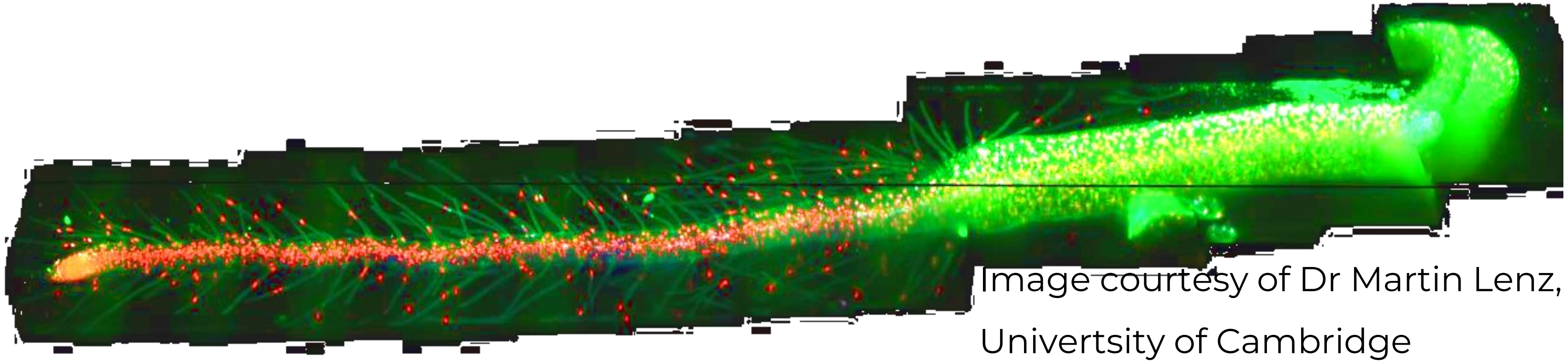
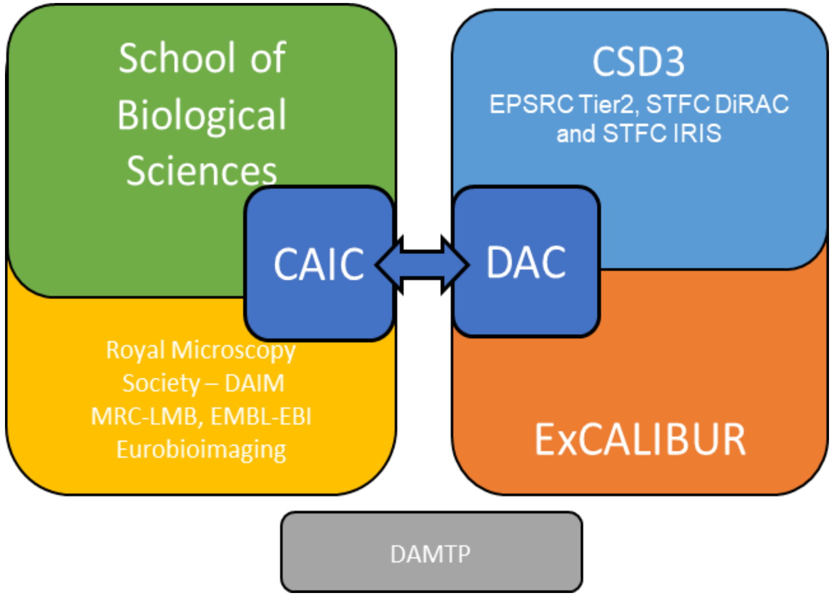


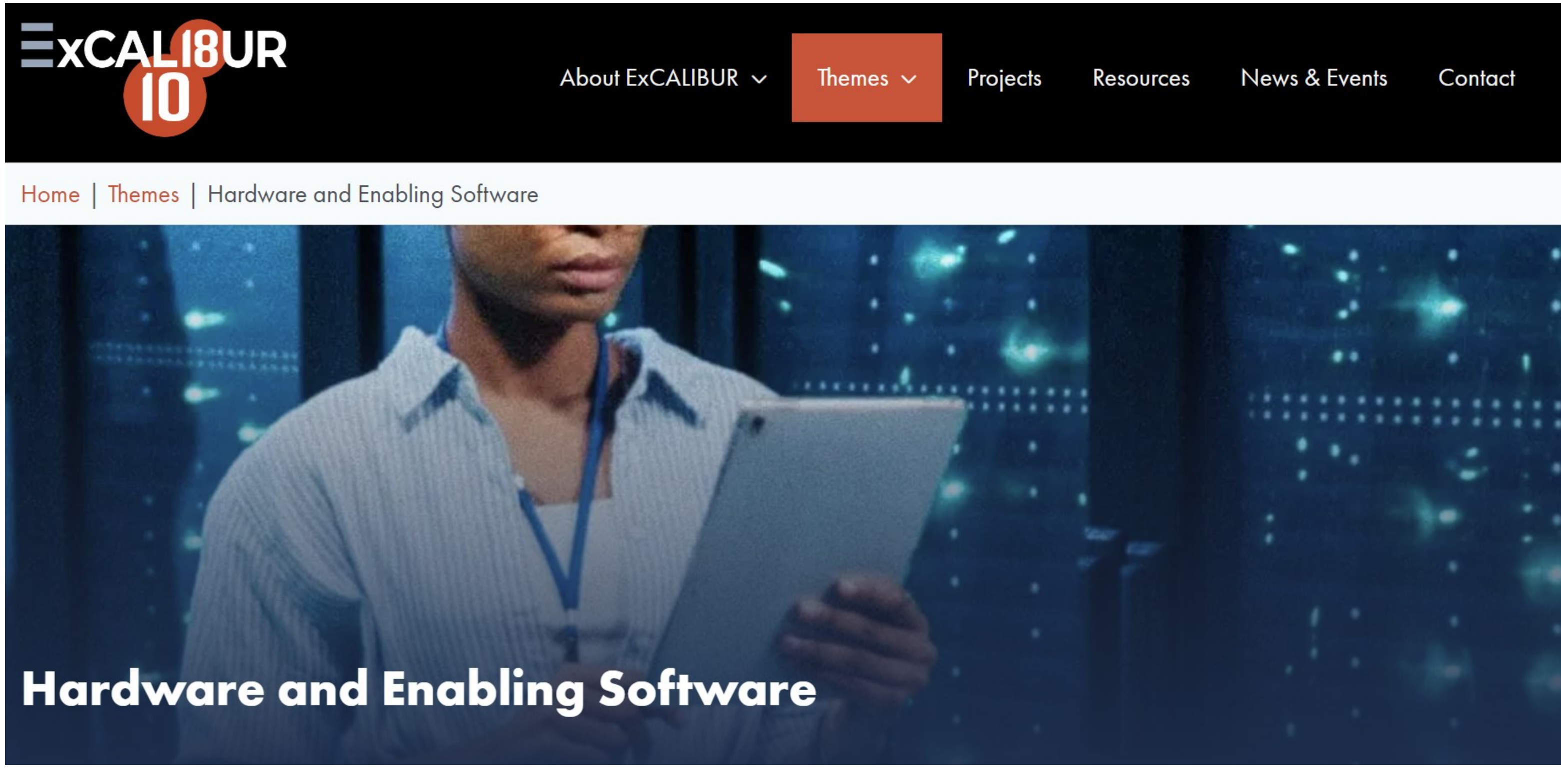
Image courtesy of Dr Martin Lenz,

Univertsity of Cambridge

Who are we? What do we do?



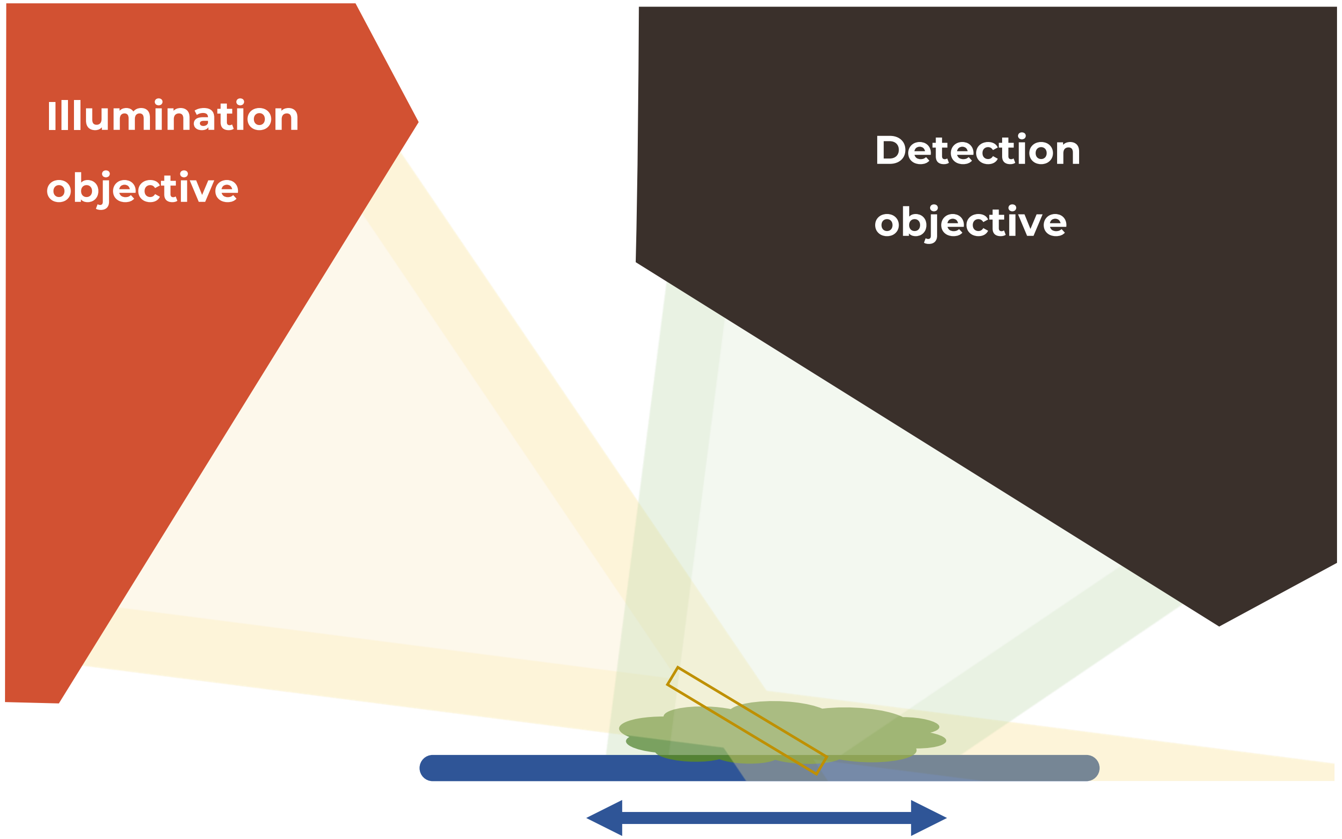
Cambridge Advanced Imaging Centre | **Data Accelerator Cambridge**

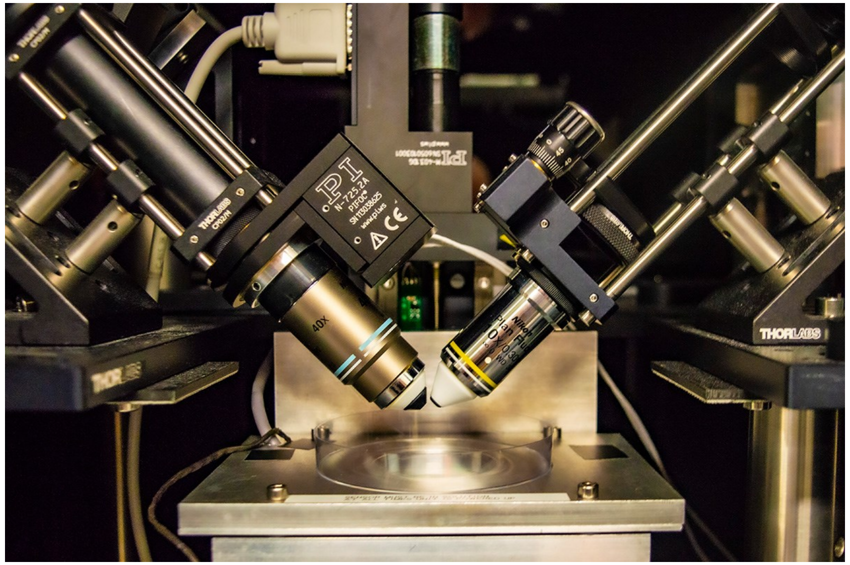


UK Research and Innovation

- Microscopy imaging (commercial and custom-built microscopes)
- Biological samples
- Image processing and analysis

Light sheet microscopy





Reduced phototoxicity

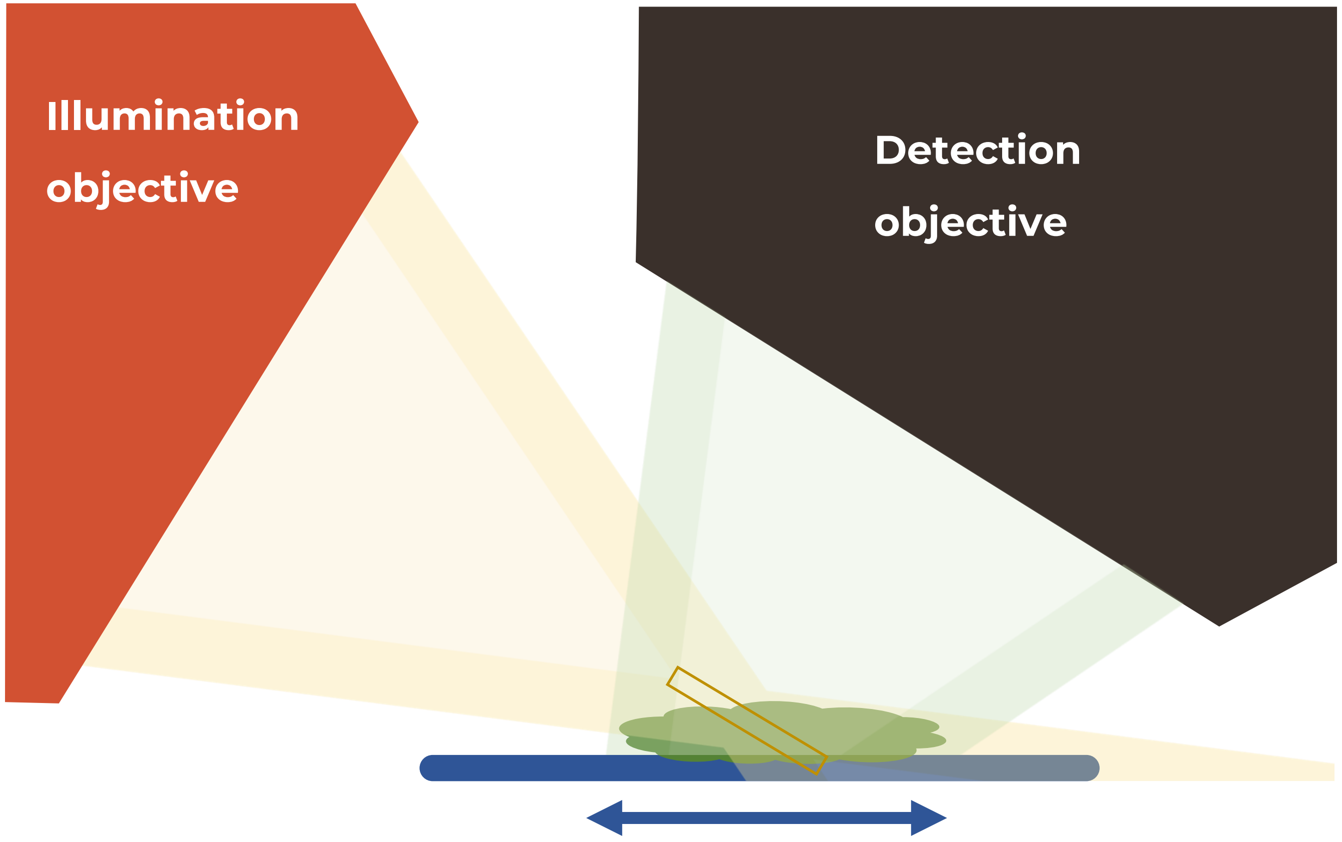
↓

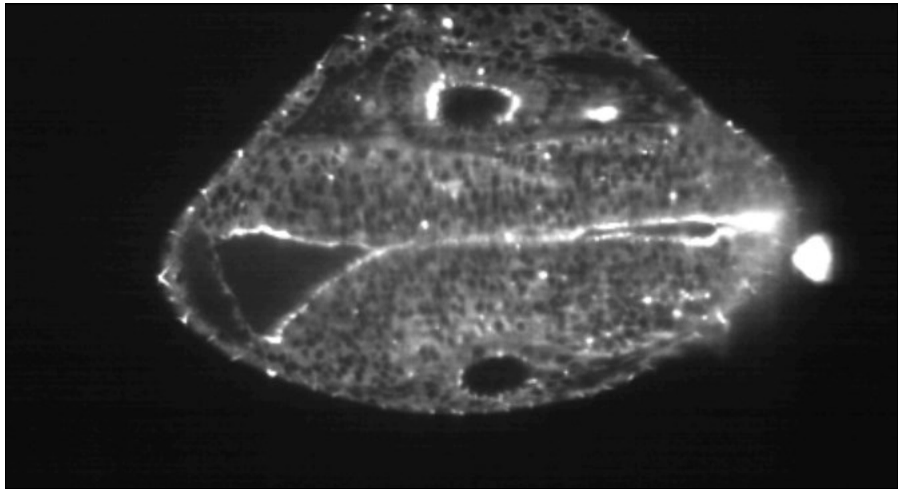
Longer sample life

↓

High-resolution 3D + time data

Light sheet microscopy

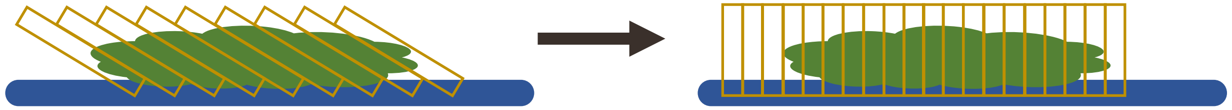




3D + time data of zebrafish (400 GB)

Image courtesy of Dr Clare Buckley
and Amelia Race

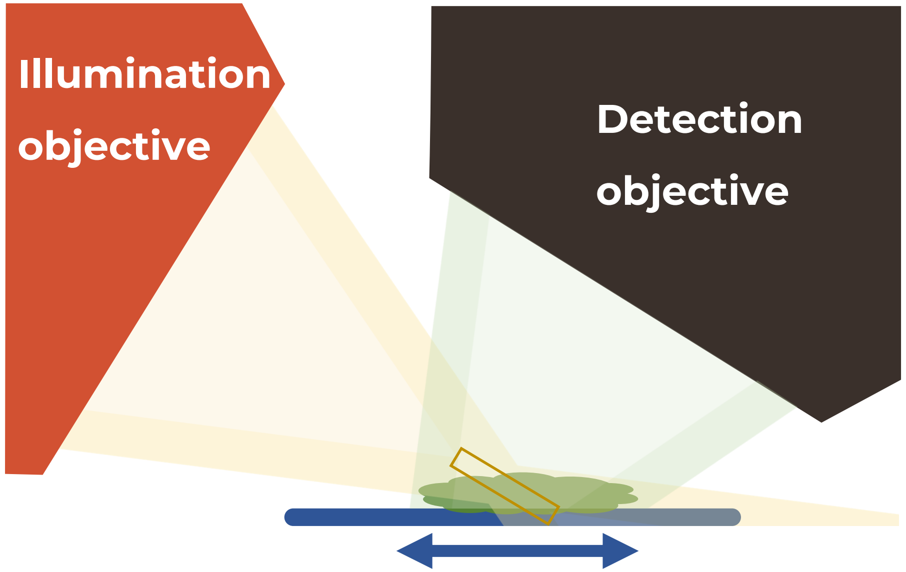
Deskewing:



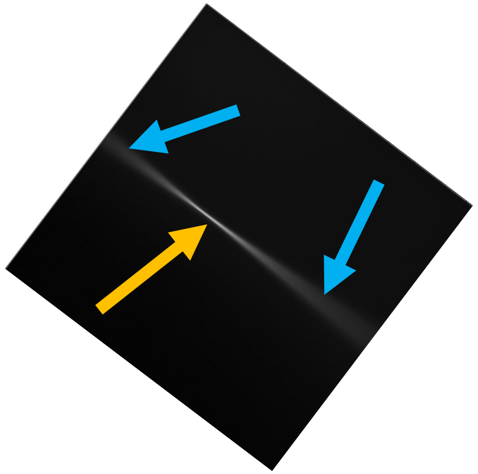
<u>Image processing challenges</u>:

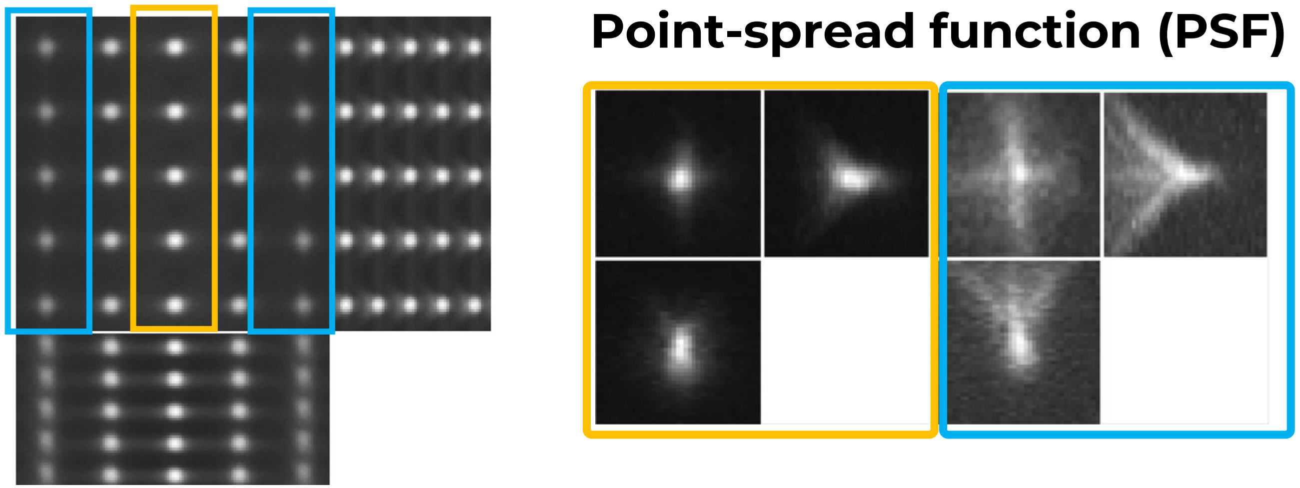
- Deskewing
- Point-spread function (PSF) correction

PSF correction in light sheet microscopy



Geometry of the light beam



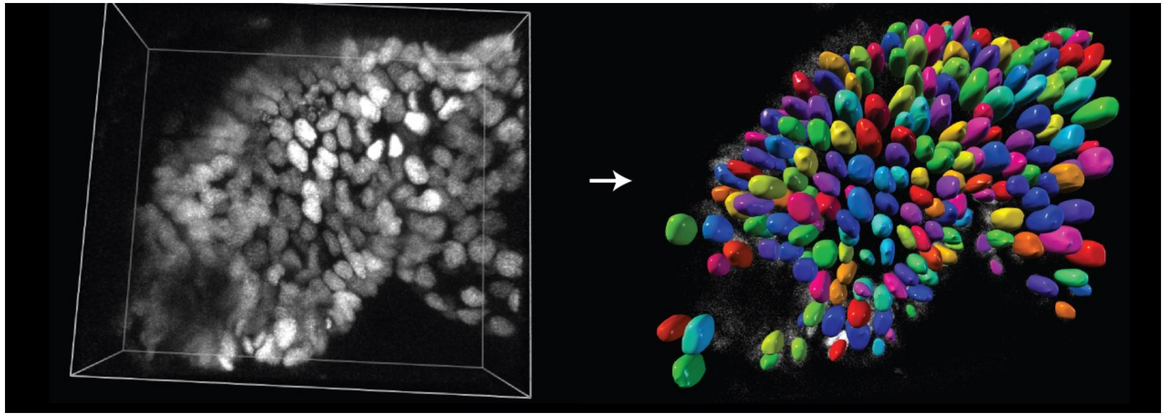


- Spatially-varying PSF
- Non-Gaussian PSF
- + Poisson (emission) and Gaussian (detection) noise

Toader et al. JMIV (2022) 64:968–992

Bespoke Image Analysis Solutions

- Acquisition parameters might be suboptimal for analysis
- Voxel size might be incompatible with existing image analysis methods (especially deep learning-based)
- Interdisciplinary communication is HARD



Weigert et al. In Proceedings of the IEEE/CVF WACV. 2020.

Research questions

↓

Experimental planning

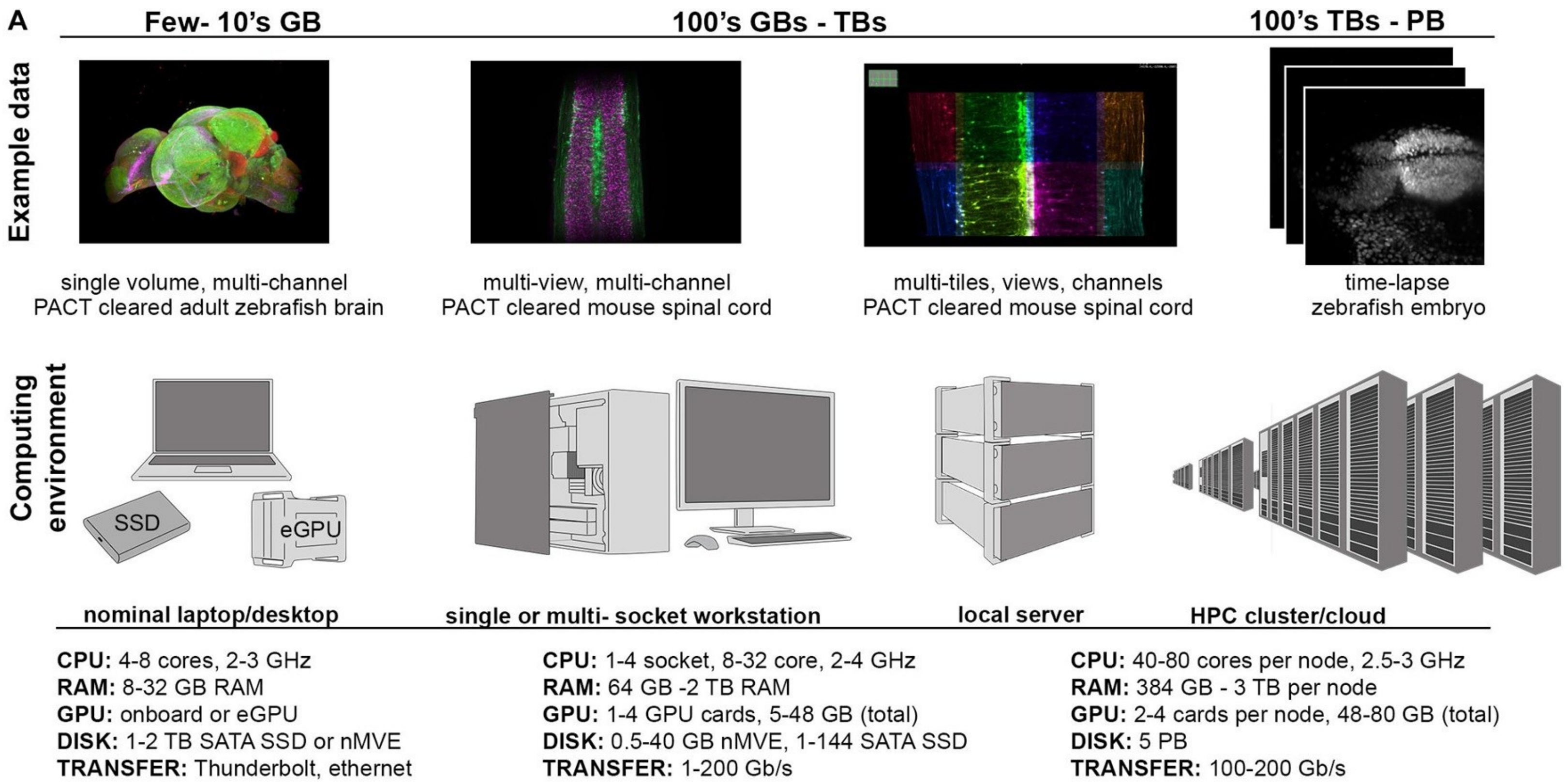
↓

Image acquisition

↓

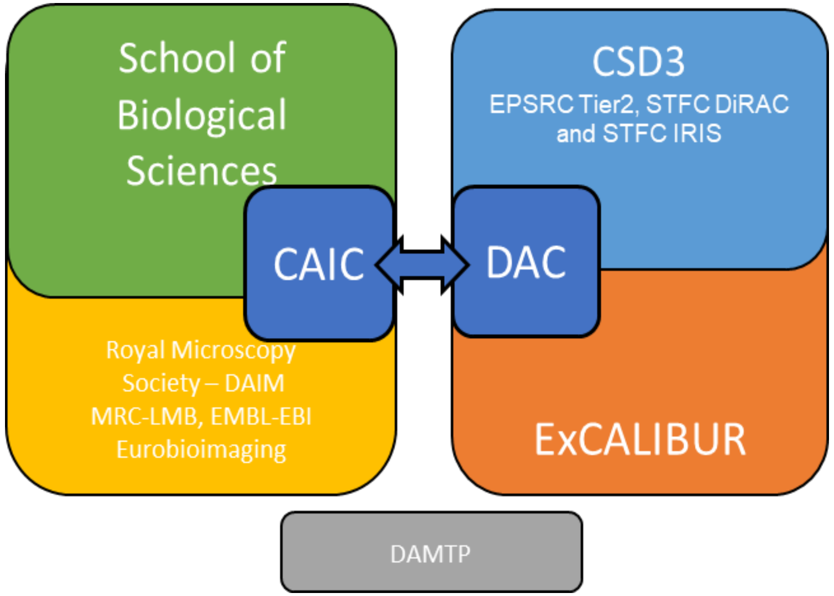
Image processing and analysis

Computational requirements



- Large data sizes
- Scale hardware accordingly

High-resolution 3D + time data



- Optimisation locally on down-sampled.
- Data transfer is a bottleneck – Globus.

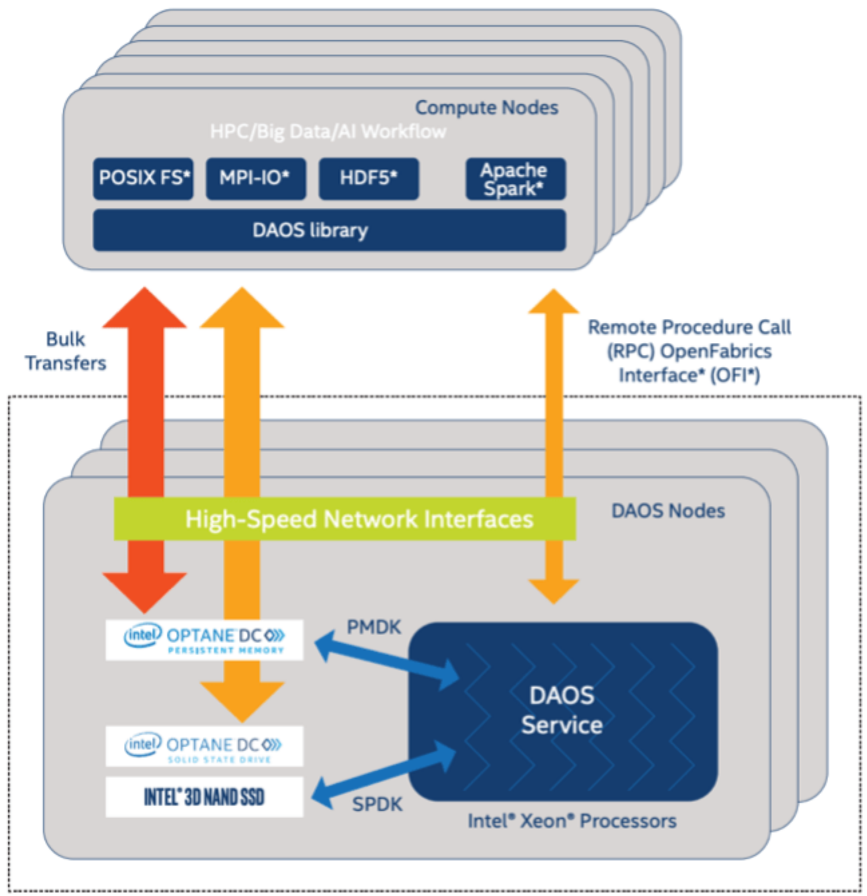
Gibbs et al. Front Cell Dev Biol, 9, 2021, 10.3389/fcell.2021.739079

Computational requirements and the Exascale Test Bed

- Reaching memory limits in processing pipeline.
  - Use high-memory nodes, or:
  - Code redesign for distributed memory system

- To improve I/O performance we can use the Exascale Test Bed:
  - 1 PB NVMe storage
  - Access to CSD3's compute
  - 500GB/s bandwidth

https://excalibur.ac.uk/projects/exascale-data-testbed



Conclusions

- Light sheet microscopy enables the acquisition of high-resolution 3D+time images
- Complex image processing algorithms and customised image analysis pipelines are often necessary to answer the relevant scientific question
- Multiple levels (PC to cluster) of advanced hardware and software are required to store, transfer, and fully process these images
- Improved communication with the biologist teams might help alleviate some of the challenges

Thank you for your attention!