An Architectural Journey to Scalable Supercomputing for AI

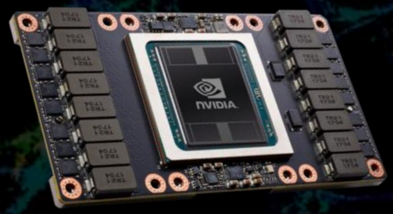
The Inside Story.

Jake Carroll, *Associate Director – Research Computing (Institutes)*

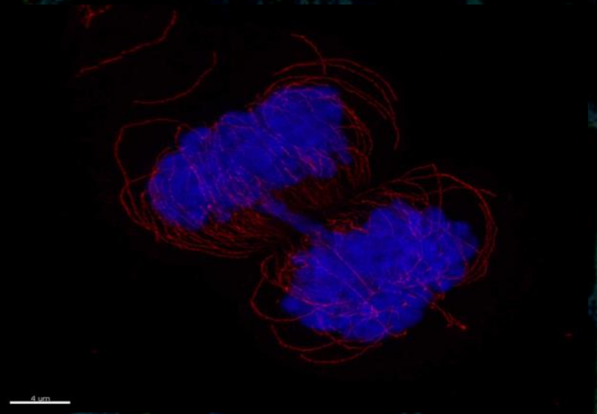
The University of Queensland

jake.carroll@uq.edu.au

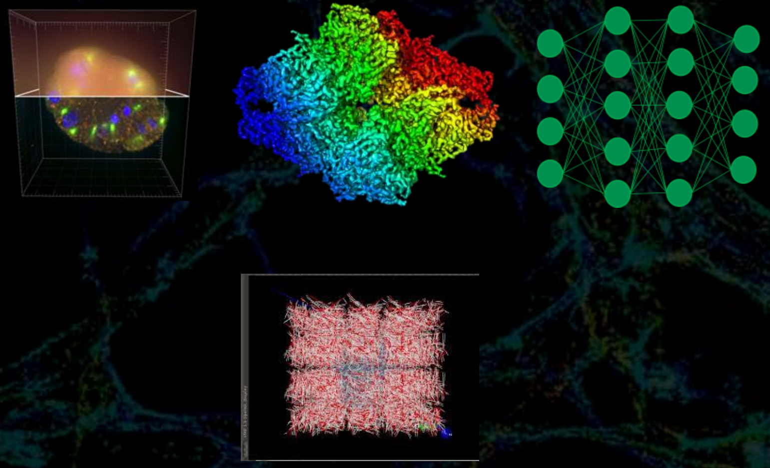
Today…



Is about technology.



Is about science.



Is about how worlds mix and collide.

Is about people.

Meet the family.

CAI

QBI

100's of … enerated.

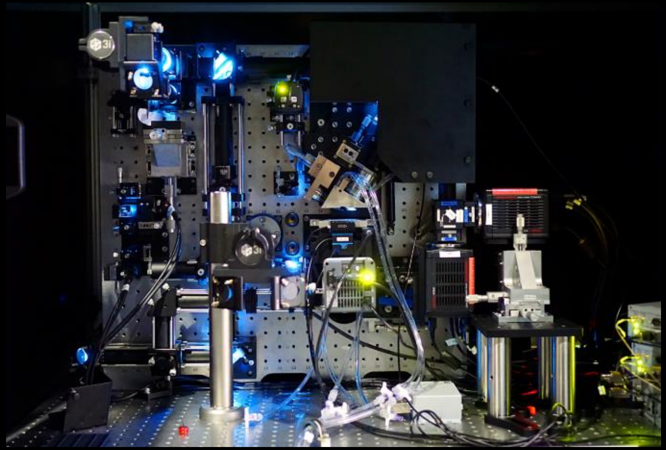
Eclectic mixture … eering, physics,

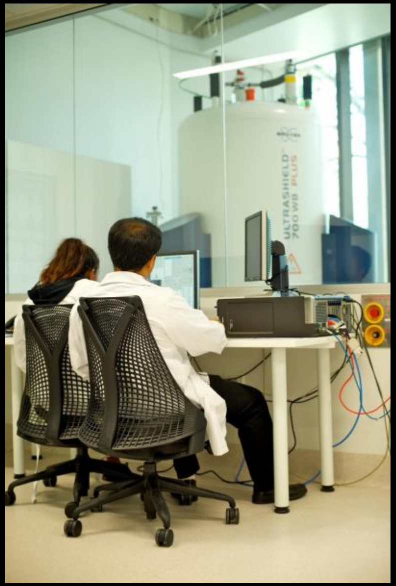
na … ....

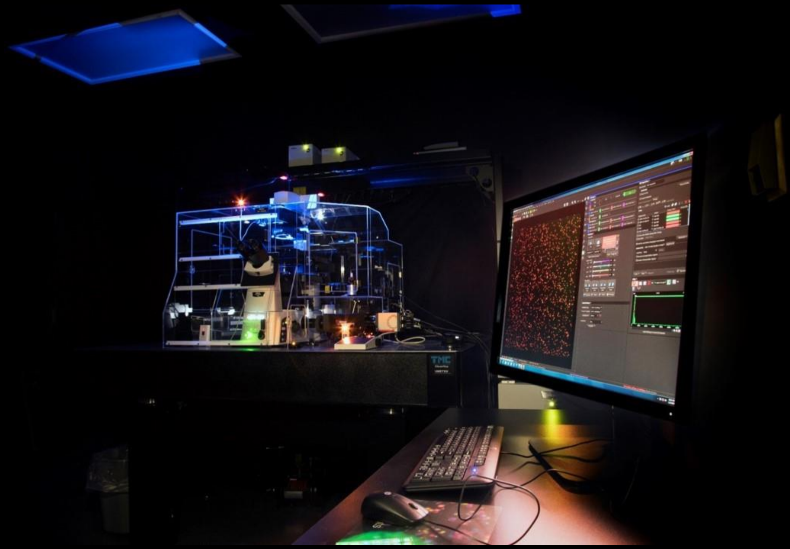
IMB

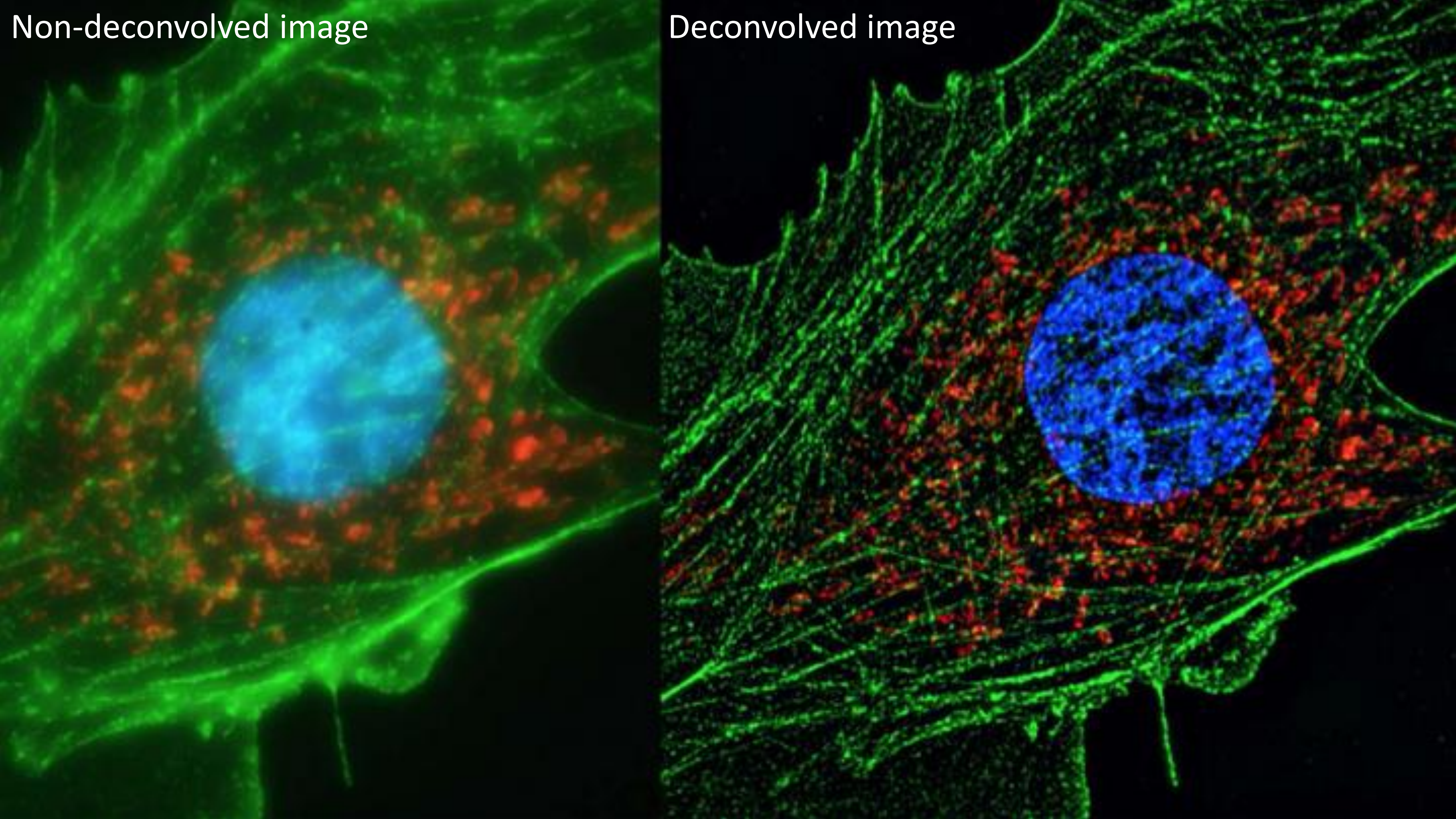
AIBN

Scientific infrastructure of immense scale



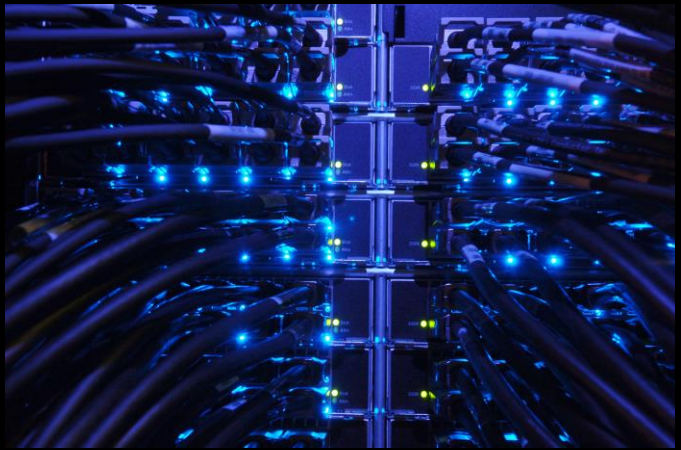





Non-deconvolved image
Deconvolved image

RCC RESEARCH COMPUTING CENTRE

UQ’s supercomputing strategy - Right supercomputer for the right task. “Best fit”



Tinaroo - 7000 cores of Intel Broadwell. Tight MPI, massively parallel, Infiniband FDR connected “Traditional” HPC.

FlashLite - 1632 cores of Intel Haswell. High memory footprint, virtual SMP (ScaleMP), high throughput. SSD /tmp in each node.

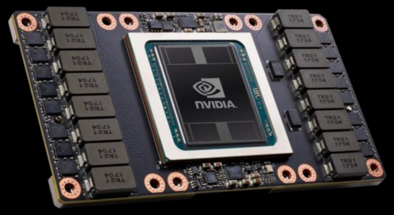
Awoonga - 1032 cores of Intel Broadwell. Loosely coupled, embarrassingly parallel, high latency tolerant workloads. *Ethernet* connected HPC.

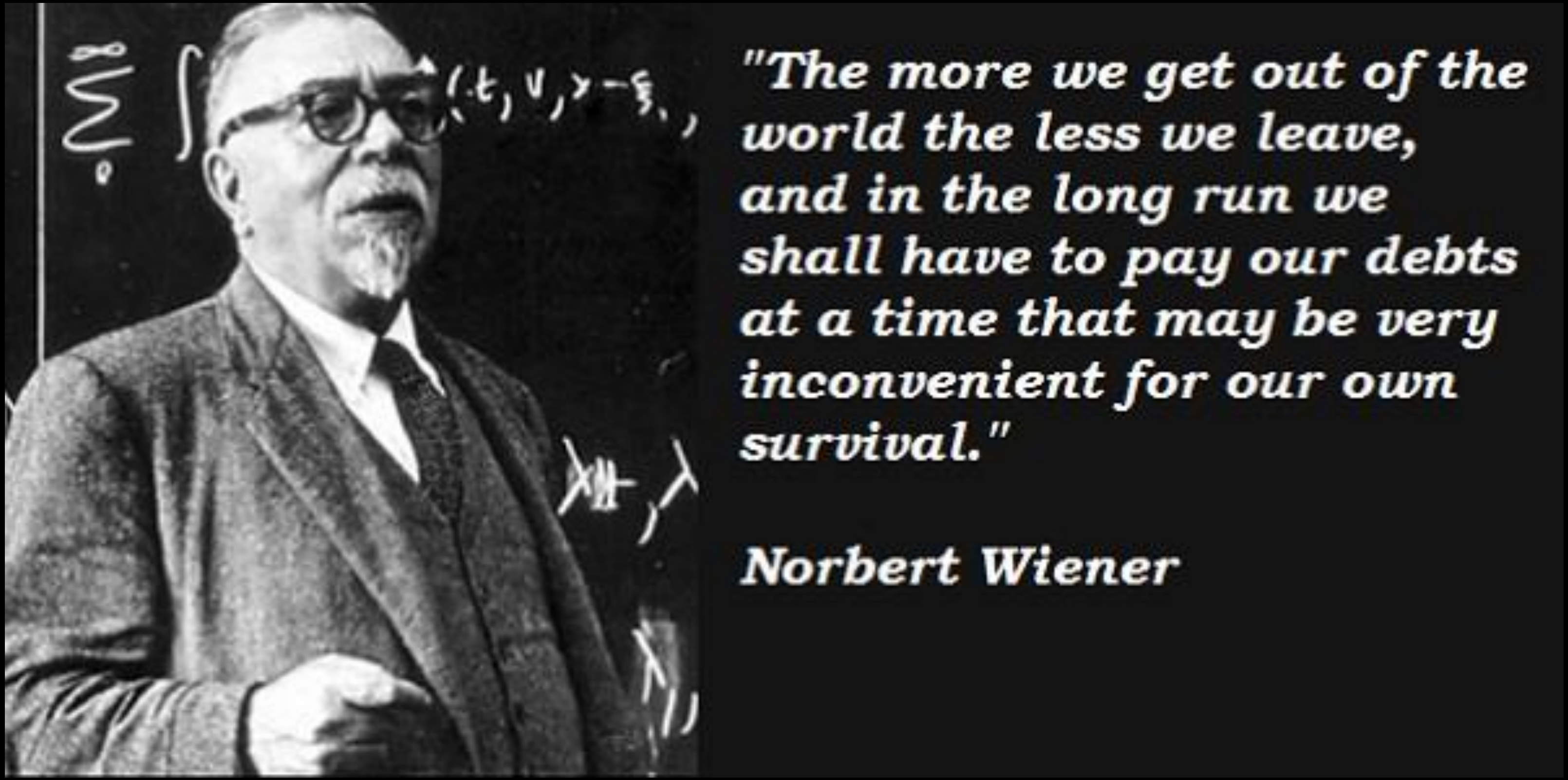
Accelerator based supercomputing strategy.

To cope with 100’s of terabytes per day of imaging, genomics and sensor data, UQ turned to GPU accelerated supercomputing to solve its significant and complex scientific problems.

Wiener – announced Nov 2017 @ SC2017

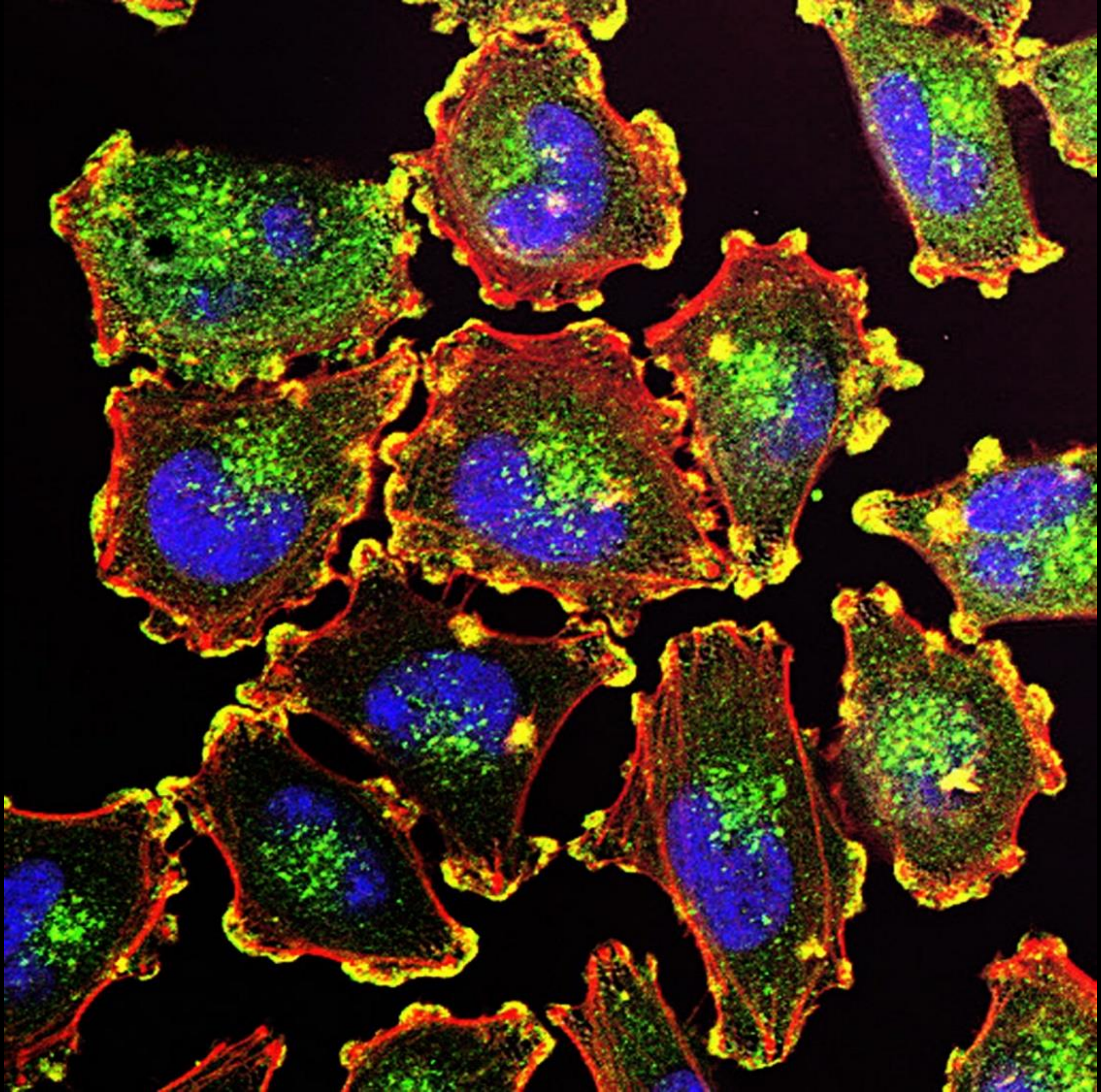
First nVidia Volta in Asia Pacific





Norbert Wiener (1894 - 1964)

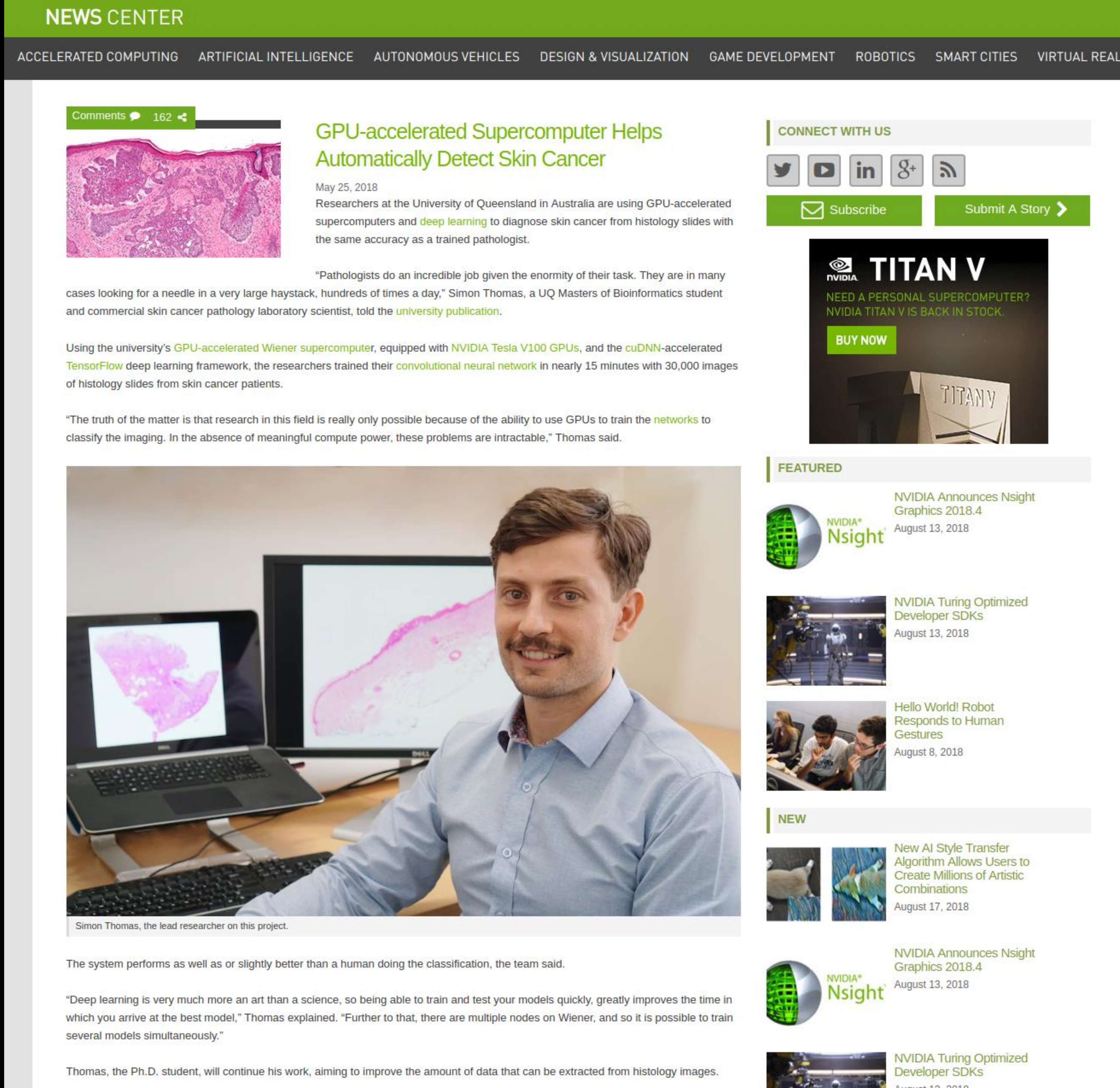
Christiaan Huygens (1629 - 1695)



Better than human vision cancer pathology and recognition accuracy.

30,000 pathology slides categorised in an CNN in 13 minutes.

TensorFlow function on nVidia Volta.



NEWS CENTER

ACCELERATED COMPUTING | ARTIFICIAL INTELLIGENCE | AUTONOMOUS VEHICLES | DESIGN & VISUALIZATION | GAME DEVELOPMENT | ROBOTICS | SMART CITIES | VIRTUAL REAL

Comments 162

GPU-accelerated Supercomputer Helps Automatically Detect Skin Cancer

May 25, 2018

Researchers at the University of Queensland in Australia are using GPU-accelerated supercomputers and deep learning to diagnose skin cancer from histology slides with the same accuracy as a trained pathologist.

"Pathologists do an incredible job given the enormity of their task. They are in many cases looking for a needle in a very large haystack, hundreds of times a day," Simon Thomas, a UQ Masters of Bioinformatics student and commercial skin cancer pathology laboratory scientist, told the university publication.

Using the university's GPU-accelerated Wiener supercomputer, equipped with NVIDIA Tesla V100 GPUs, and the cuDNN-accelerated TensorFlow deep learning framework, the researchers trained their convolutional neural network in nearly 15 minutes with 30,000 images of histology slides from skin cancer patients.

"The truth of the matter is that research in this field is really only possible because of the ability to use GPUs to train the networks to classify the imaging. In the absence of meaningful compute power, these problems are intractable," Thomas said.

Simon Thomas, the lead researcher on this project.

The system performs as well as or slightly better than a human doing the classification, the team said.

"Deep learning is very much more an art than a science, so being able to train and test your models quickly, greatly improves the time in which you arrive at the best model," Thomas explained. "Further to that, there are multiple nodes on Wiener, and so it is possible to train several models simultaneously."

Thomas, the Ph.D. student, will continue his work, aiming to improve the amount of data that can be extracted from histology images.

CONNECT WITH US

Subscribe | Submit A Story

FEATURED

NVIDIA Announces Nsight Graphics 2018.4
August 13, 2018

NVIDIA Turing Optimized Developer SDKs
August 13, 2018

Hello World! Robot Responds to Human Gestures
August 8, 2018

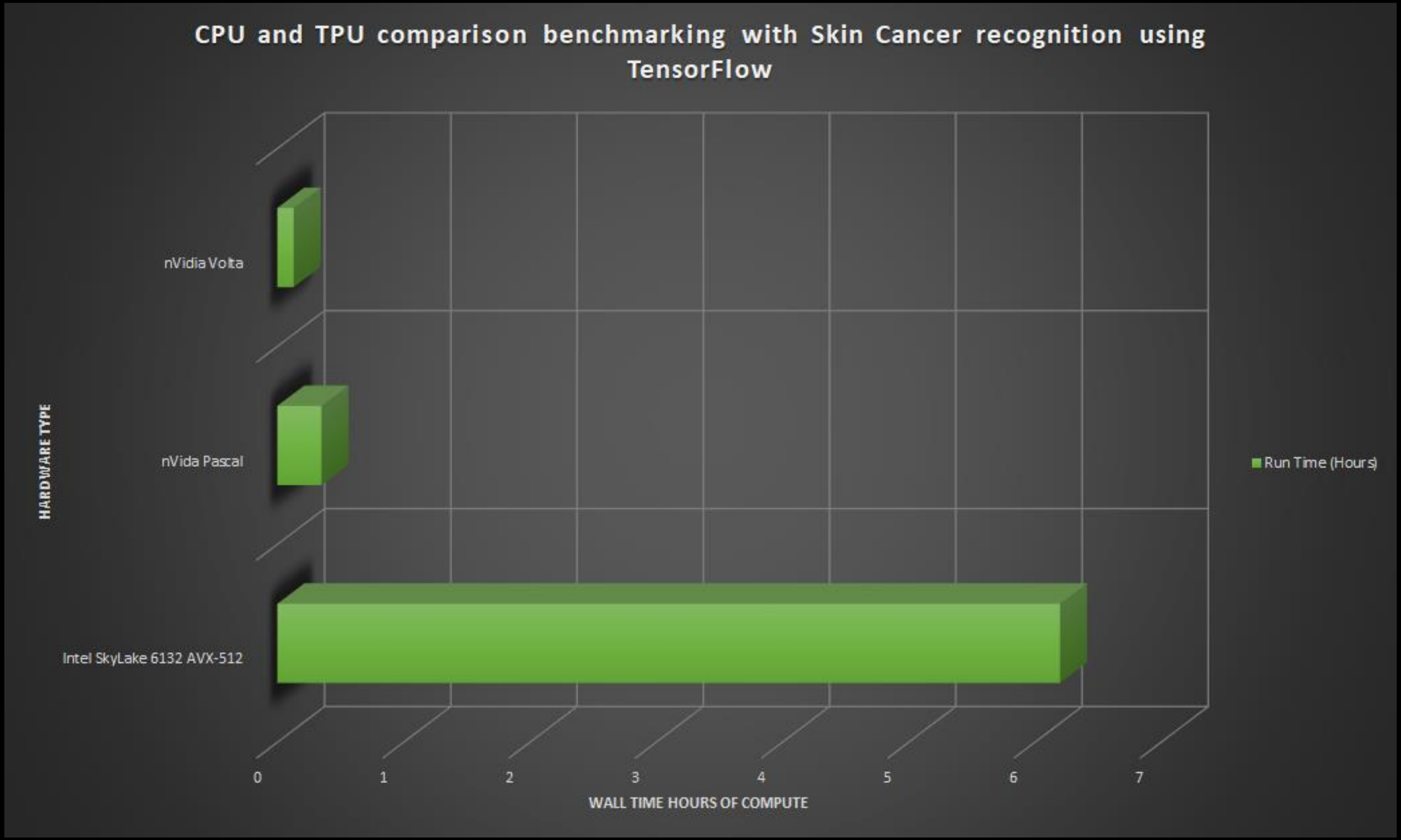
NEW

New AI Style Transfer Algorithm Allows Users to Create Millions of Artistic Combinations
August 17, 2018

NVIDIA Announces Nsight Graphics 2018.4
August 13, 2018

NVIDIA Turing Optimized Developer SDKs

TensorFlow workload for skin cancer recognition inference



Abstract

There is considerable interest in designing guidance systems for UAVs that use passive sensing (such as vision), rather than active sensing which can be bulky, expensive and stealth-compromising. Here we describe and evaluate a sensor that uses optic flow for measurement and control of height above the ground. A video camera is used in conjunction with a specially shaped reflective surface to simplify the computation of optic flow, and extend the range of aircraft speeds over which accurate data can be obtained. The imaging system also provides a useful geometrical remapping of the environment, which facilitates obstacle avoidance and computation of 3-D terrain maps. Laboratory tests of the performance of the device, whilst undergoing complex motions, are described.

1. Introduction

moving image that is captured by the camera through the mirror should exhibit a constant, low velocity everywhere, thus simplifying the optic flow measurements and increasing their accuracy (see Fig. 1).

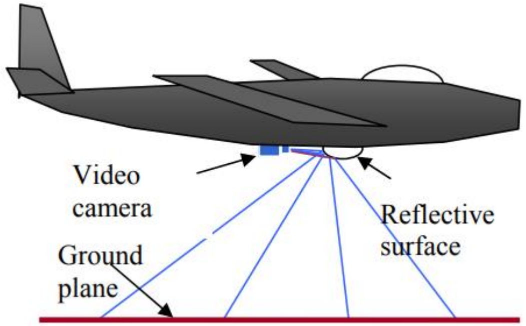
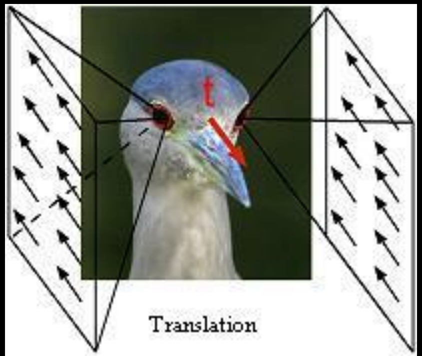


Fig. 1. Illustration of a system for visually guided terrain following and landing. The optical system is shown on an enlarged scale relative to the aircraft in order to clarify its configuration

Machine vision of optic flow for UAV workloads. Novel use of machine vision techniques, partnership with US DoD.



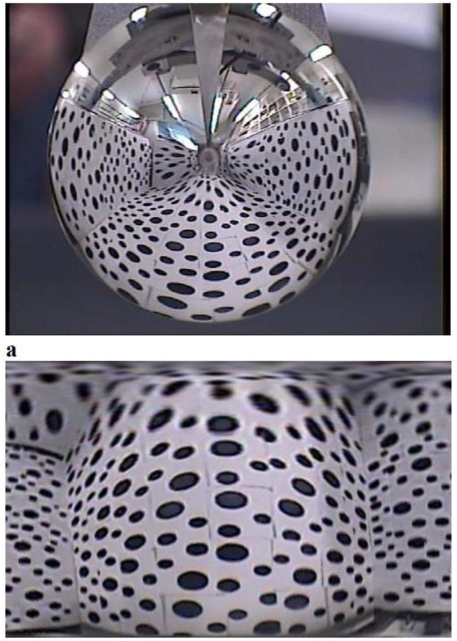


Fig 9: Raw (a) and unwarped (b) images of the arena as viewed by the system.

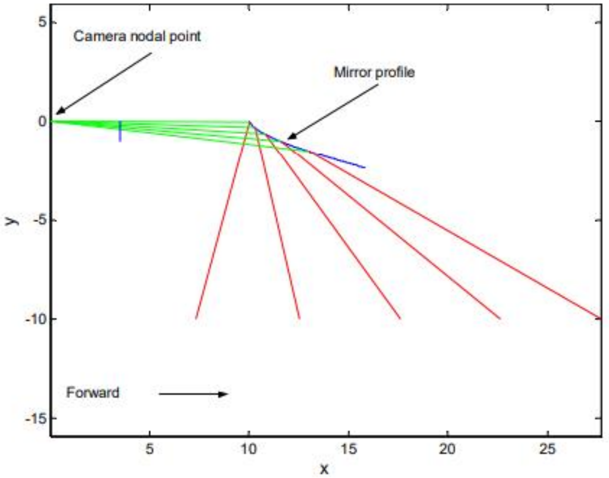


Fig. 3. Example of computed mirror profile

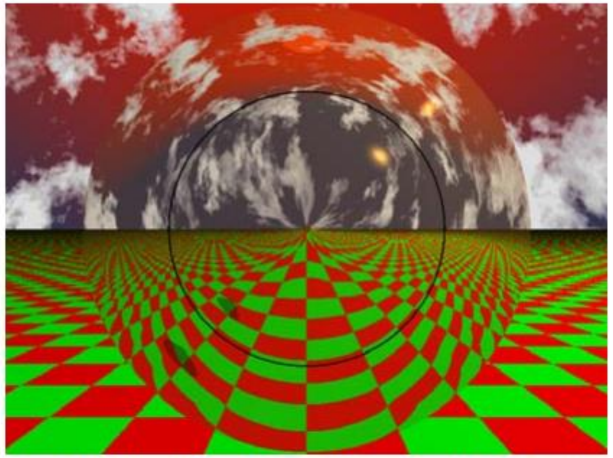


Fig. 4. Illustration of imaging properties of mirror

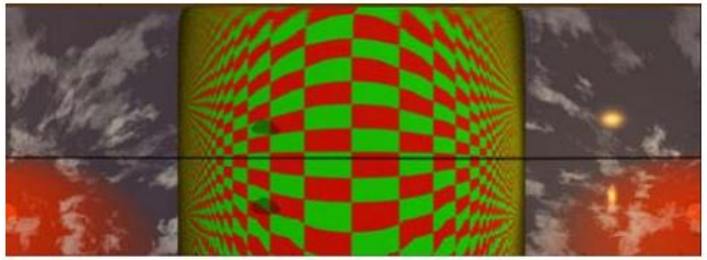
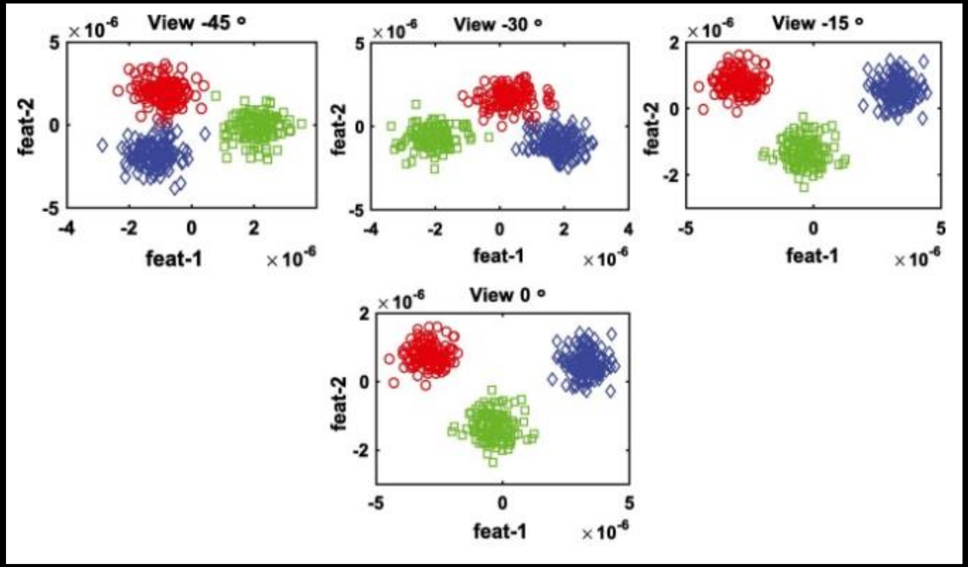
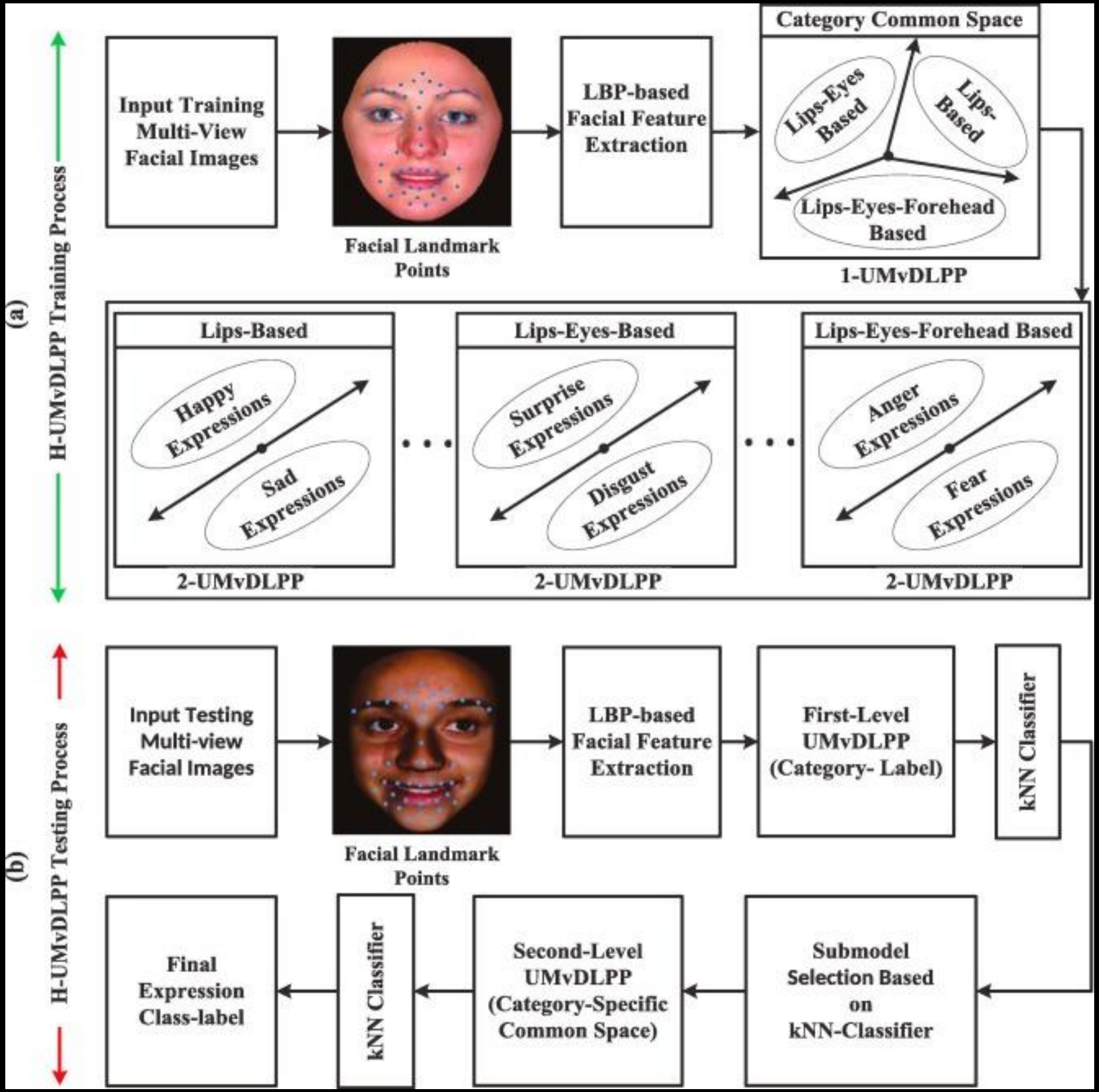
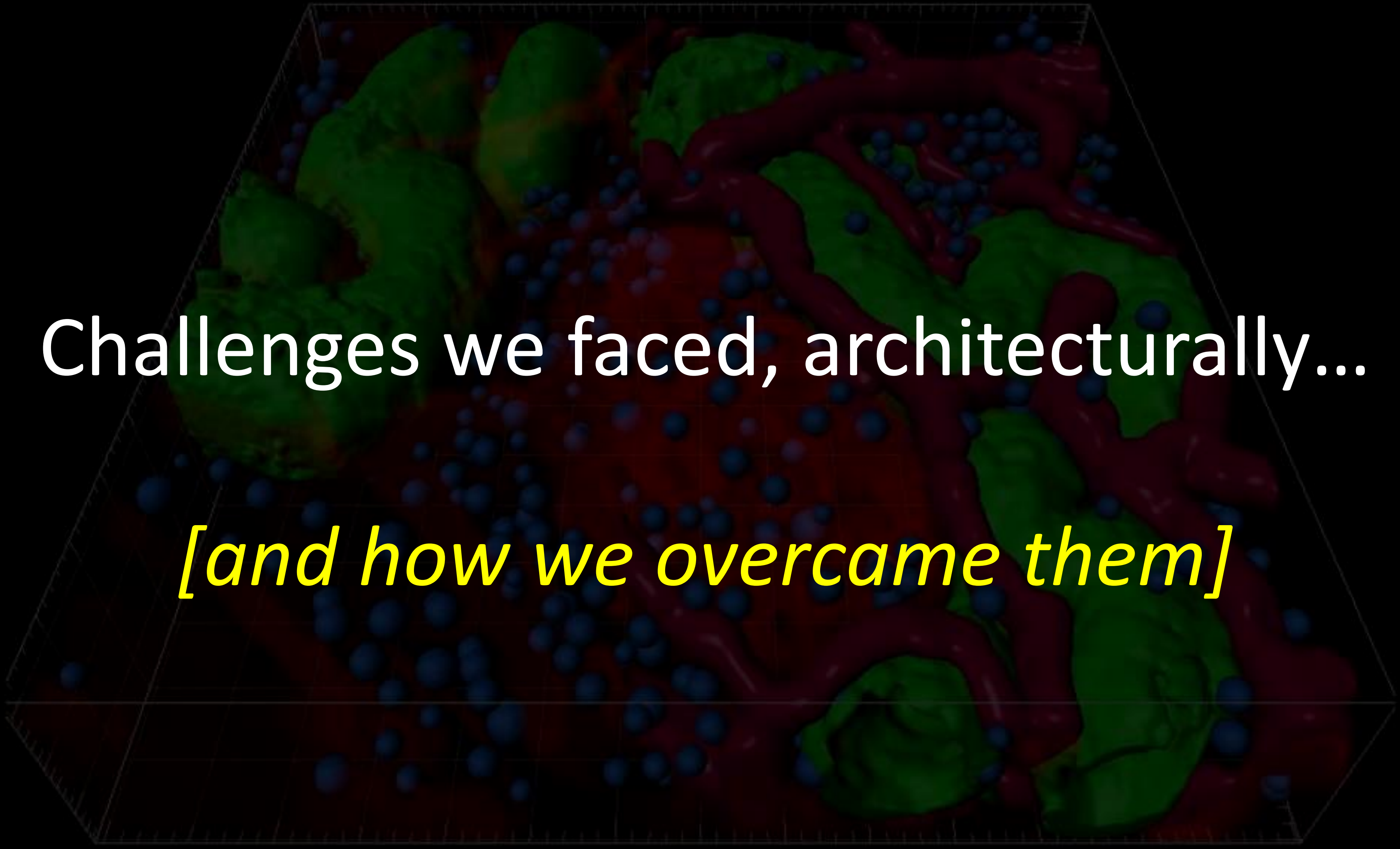


Fig. 5. Remapped version of image in Fig 4.

Artificial intelligence techniques in multi-view real time facial recognition for sentiment analysis, security, surveillance and threat analysis.







Challenges we faced, architecturally…

[and how we overcame them]

Long gone is the era of "building a capability" and expecting it to satisfy users for 5 years…3 years…or even 1 year. This mentality is *not* going to win you that **Nature** front page in a computationally intensive "arms race".

If that's the mindset – you've already lost the game. You've already come second. You've already ended up a less impactful research organisation than the crew down the road who pushed that little bit harder.

Some might say: "Won't somebody think of the administrative overhead? That would cost so much to achieve! So many hours – so much risk!"

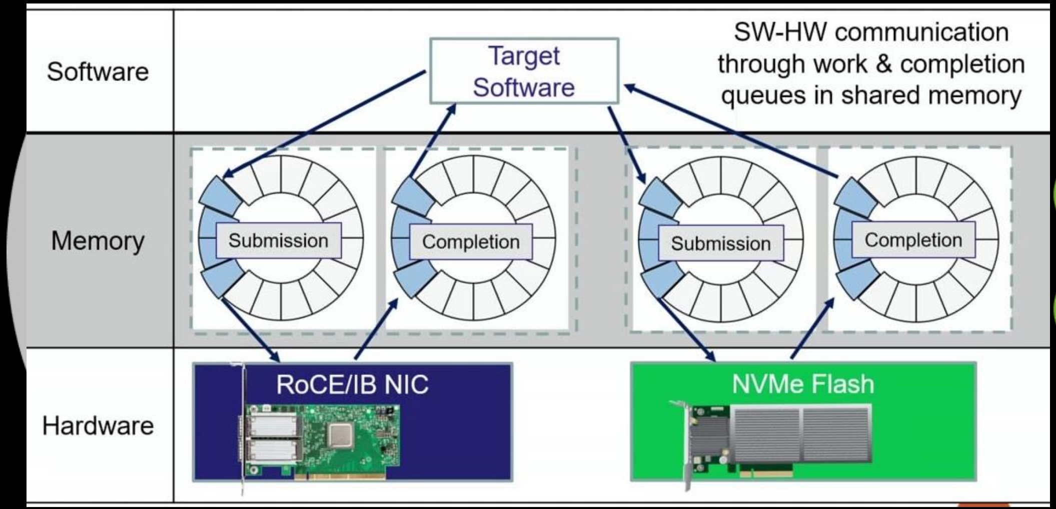
Continuous aggressive version baselines, driven upgrades, firmware and optimisation. Monthly, weekly, almost daily, in the quest for every last bit of performance. Recompile daily to squeeze out every bit of life you can.

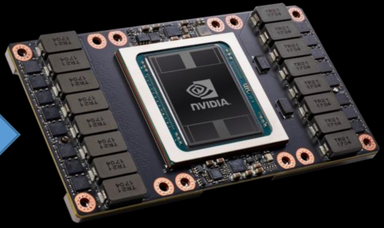
Expect more

Ye old(e) storage bottleneck effect…

BeeGFS on nVME RDMA EDR connected flash. Delivering 180GB/sec and 25m IOPS of sustained performance.

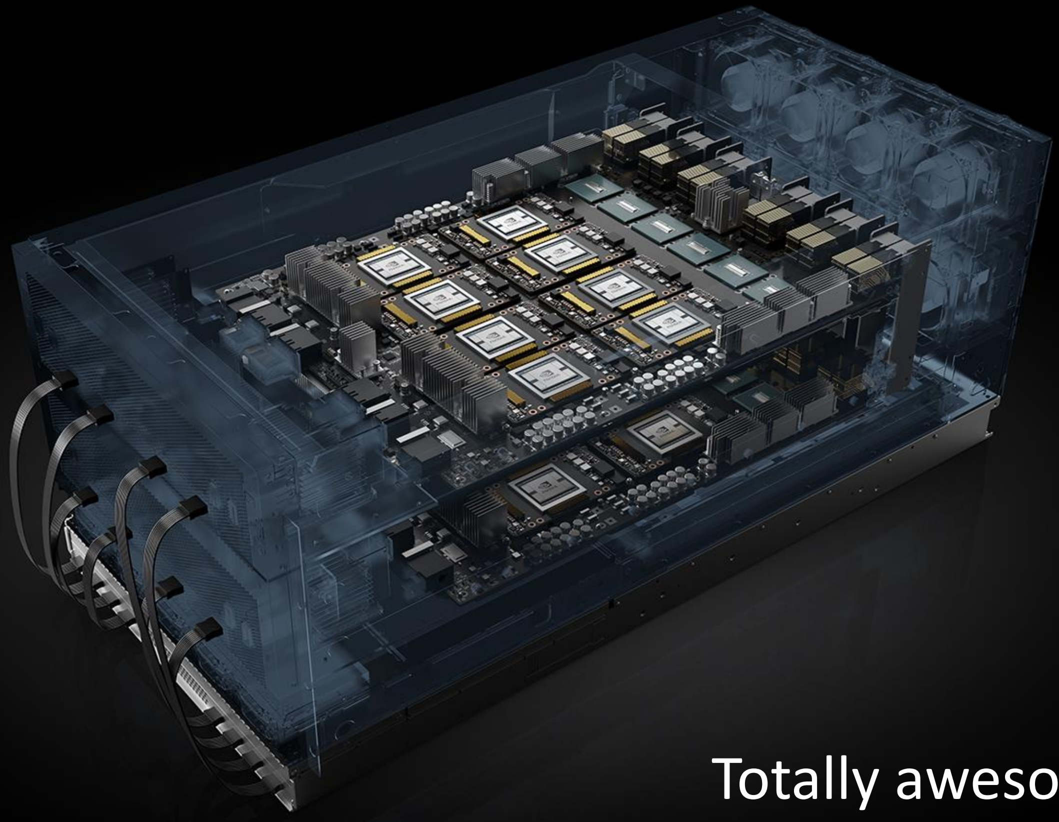
Traditional parallel filesystem, or object (GPFS, Ceph, Swift etc)





GPUs

MPI and inter-nodal comms



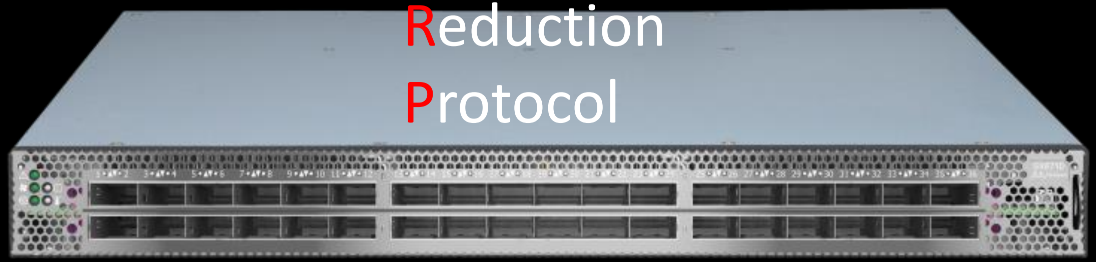
Totally awesome….but what if you need double…or triple…or quadruple this many GPU's – and they all need to communicate at scale?

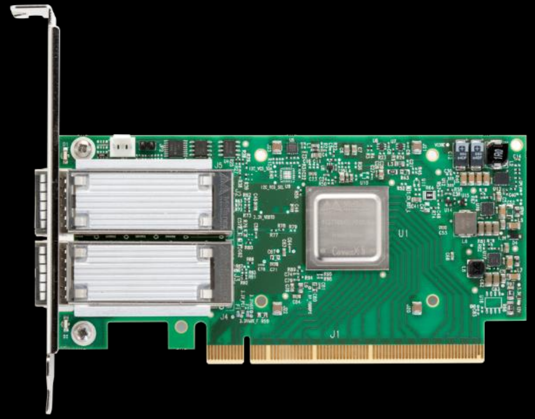
Easy, right? Just buy a few of them. Stack them together somehow…

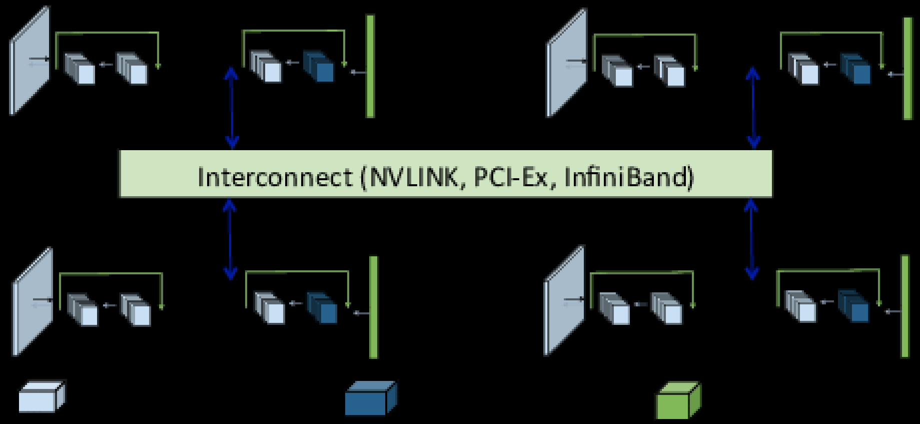
Not *quite* that simple….

In supercomputing we need to think more about high scalability scenarios, pressure on the interconnect and how to distribute workloads with least-cost/lowest latency…

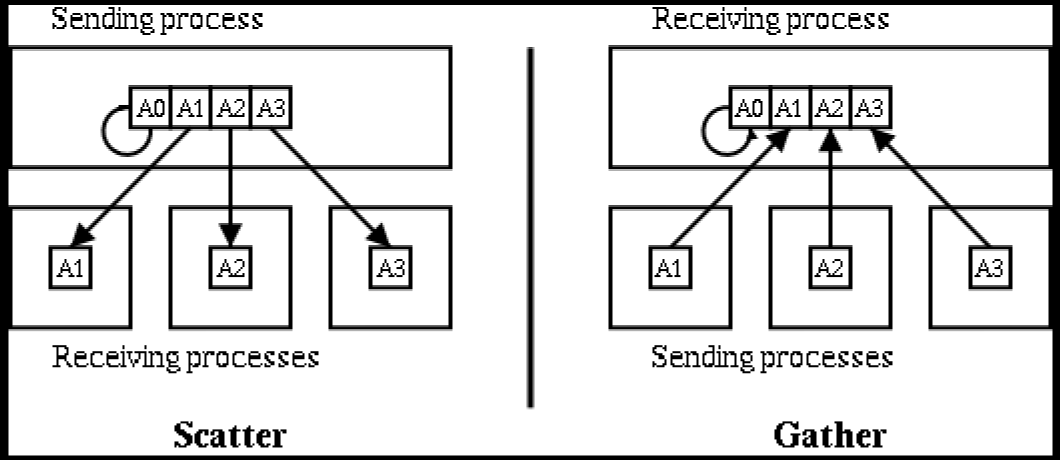
Scalable
Hierarchical
Aggregation *and*
Reduction
Protocol







Putting MPI primitives in hardware, in the network.



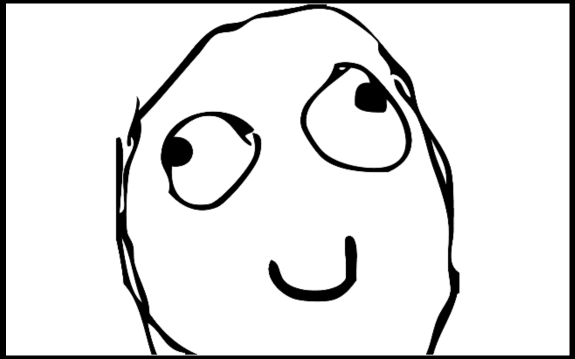
IO-gluttony. Games with 100Gb RDMA pipes…

~15.26GB/sec

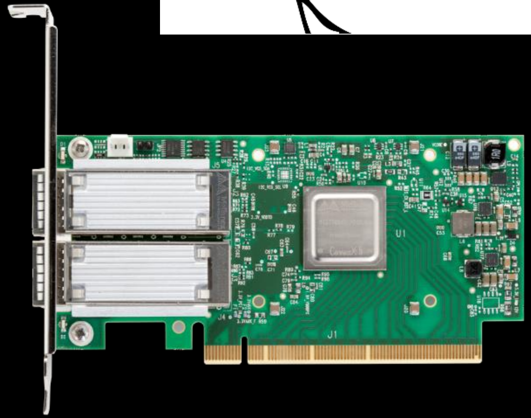
100Gb
Both RDMA storage IO and RDMA GPUDirect/MPI IO between GPU’s

A perfect IO starvation storm in a tea cup

So, let us fix this…



100Gb * 2
Dedicated RDMA for BeeGFS IO
Dedicated RDMA for MPI/GPU IO



~15.26GB/sec

But, one more big (complicated) problem…

TensorFlow™

NVIDIA CUDA

cuDNN

PYTORCH
Deep Learning with PyTorch

Caffe2

DL/ML/AI frameworks and software stacks

$\neq$

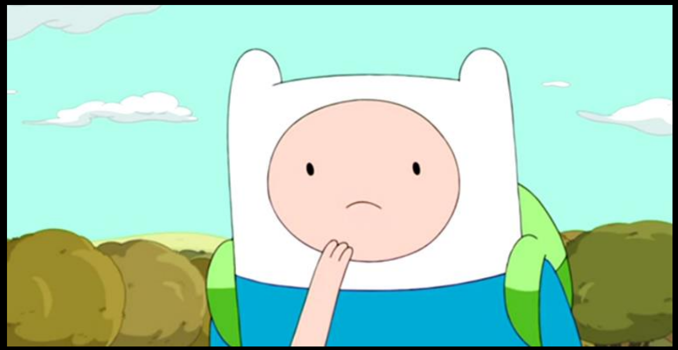
“HPC” techniques, hardware, technologies, methods, communications and principles

Different strokes, for different folks…

"This is how we've always done things! Don't mess with my world! I'll do what I want!"

"Let me at it! I'm keen to try anything if I can make an impact and get something good happening!"

"I….will take over the world with this! I must have all of it. More cores. More GPU's. All of it! Now!"



"I've no idea where to start. Help me computer-science person. I just want to do my research!"

Build the town hall. Let people come. Let them speak…

Start working with users to *co-create* value.

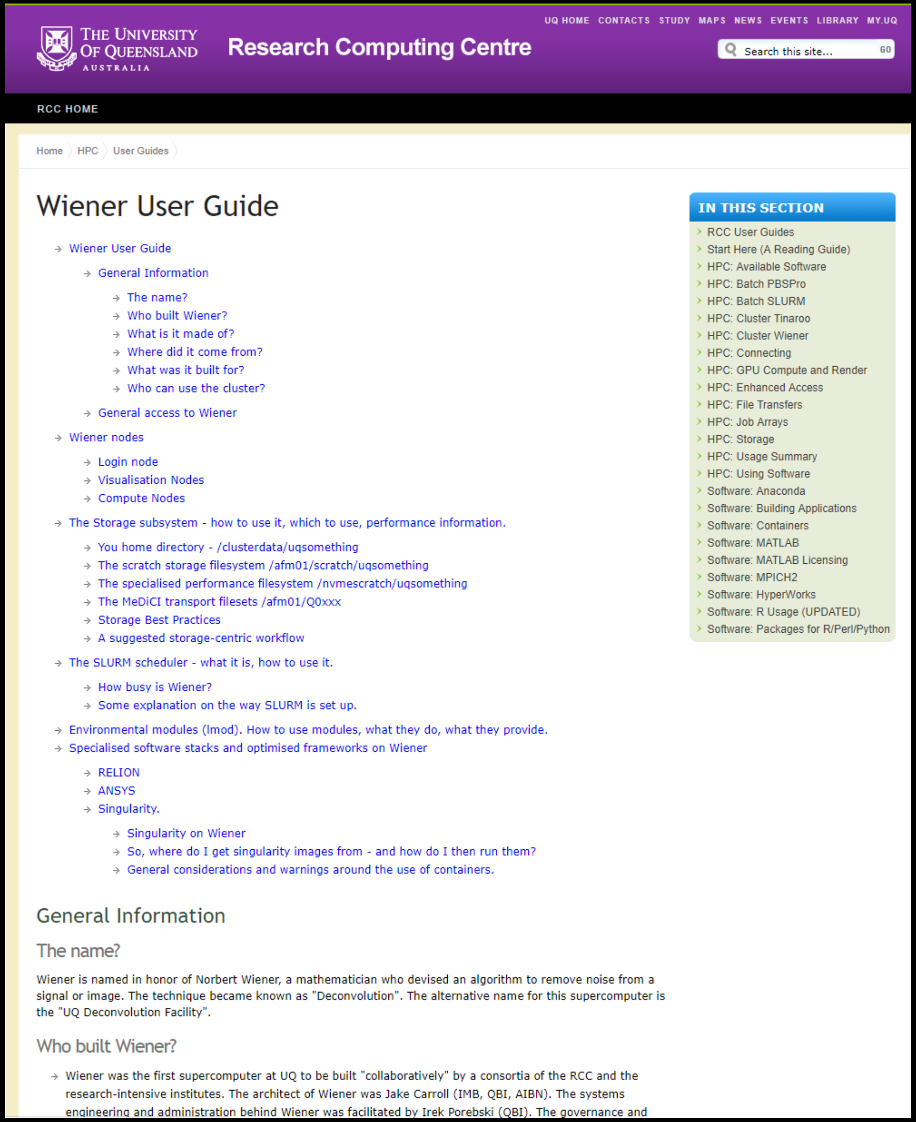

SHARP
BeeGFS
imb
sinfo
pytorch
fsl
amber
nvlink
openmpi
relion
nvidia-smi
dragonfly
digits
sbatch
low-latency
module
gromacs
singularity
sxm2
squeue
norbert
tensorflow
optimisation
fiji
javac
rcc
nvcc
huygens
load-balance
infiniband
atlas
grid
deconvolution
deep-seek
nvme
network
krios
high-throughput
imagej
aibn
docker
supercomputing
mvapich2
mpicc
gather
collective
microvolution
deep-sm
blas
nccl
scatter
tensorboard
mofed
slurm
cryosparc
mrtrix
qbi
ganglia
verbs
ofed
cudafe
pci-e
wiener

And then odd, but wonderful things started happening…

“Jake,

The new GPU supercomputer is fantastic. Thanks for giving us early access and the chance to get some early apps up to iron things out/get it working quickly. I’ve been thinking about it a bit, and I think it would be useful for the rest of my community and my school if I actually built a dedicated “how to” guide for Relion + Volta. What do you think? The performance advances we’ve seen between Volta and Pascal, combined with all the MPI tweaks we worked on together make it worth the effort I think to explain to users how different it is – and how they can take advantage of it. I can get a rough draft for the Research Computing Centre wiki to you next week, if you’re keen?”

And before we knew it, we had a very full user guide…



UQ HOME CONTACTS STUDY MAPS NEWS EVENTS LIBRARY MY.UQ

The University Of Queensland AUSTRALIA

Research Computing Centre

Search this site... GO

RCC HOME

Home › HPC › User Guides

Wiener User Guide

- Wiener User Guide
  - General Information
    - The name?
    - Who built Wiener?
    - What is it made of?
    - Where did it come from?
    - What was it built for?
    - Who can use the cluster?
  - General access to Wiener
- Wiener nodes
  - Login node
  - Visualisation Nodes
  - Compute Nodes
- The Storage subsystem - how to use it, which to use, performance information.
  - You home directory - /clusterdata/uqsomething
  - The scratch storage filesystem /afm01/scratch/uqsomething
  - The specialised performance filesystem /nvmescratch/uqsomething
  - The MeDiCI transport filesets /afm01/Q0xxx
  - Storage Best Practices
  - A suggested storage-centric workflow
- The SLURM scheduler - what it is, how to use it.
  - How busy is Wiener?
  - Some explanation on the way SLURM is set up.
- Environmental modules (lmod). How to use modules, what they do, what they provide.
- Specialised software stacks and optimised frameworks on Wiener
  - RELION
  - ANSYS
  - Singularity.
    - Singularity on Wiener
    - So, where do I get singularity images from - and how do I then run them?
    - General considerations and warnings around the use of containers.

General Information

The name?

Wiener is named in honor of Norbert Wiener, a mathematician who devised an algorithm to remove noise from a signal or image. The technique became known as "Deconvolution". The alternative name for this supercomputer is the "UQ Deconvolution Facility".

Who built Wiener?

- Wiener was the first supercomputer at UQ to be built "collaboratively" by a consortia of the RCC and the research-intensive institutes. The architect of Wiener was Jake Carroll (IMB, QBI, AIBN). The systems engineering and administration behind Wiener was facilitated by Irek Porebski (QBI). The governance and

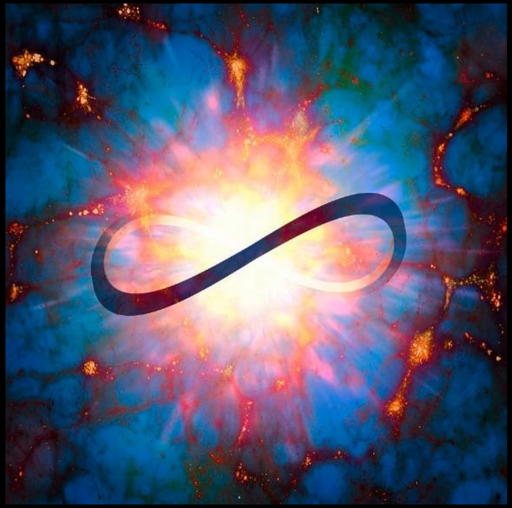
IN THIS SECTION

- RCC User Guides
- Start Here (A Reading Guide)
- HPC: Available Software
- HPC: Batch PBSPro
- HPC: Batch SLURM
- HPC: Cluster Tinaroo
- HPC: Cluster Wiener
- HPC: Connecting
- HPC: GPU Compute and Render
- HPC: Enhanced Access
- HPC: File Transfers
- HPC: Job Arrays
- HPC: Storage
- HPC: Usage Summary
- HPC: Using Software
- Software: Anaconda
- Software: Building Applications
- Software: Containers
- Software: MATLAB
- Software: MATLAB Licensing
- Software: MPICH2
- Software: HyperWorks
- Software: R Usage (UPDATED)
- Software: Packages for R/Perl/Python

The take away.

Balanced architecture is hard, actually.

We made some poor assumptions to begin with. ML/DL/AI is still quite immature when mashed with HPC.

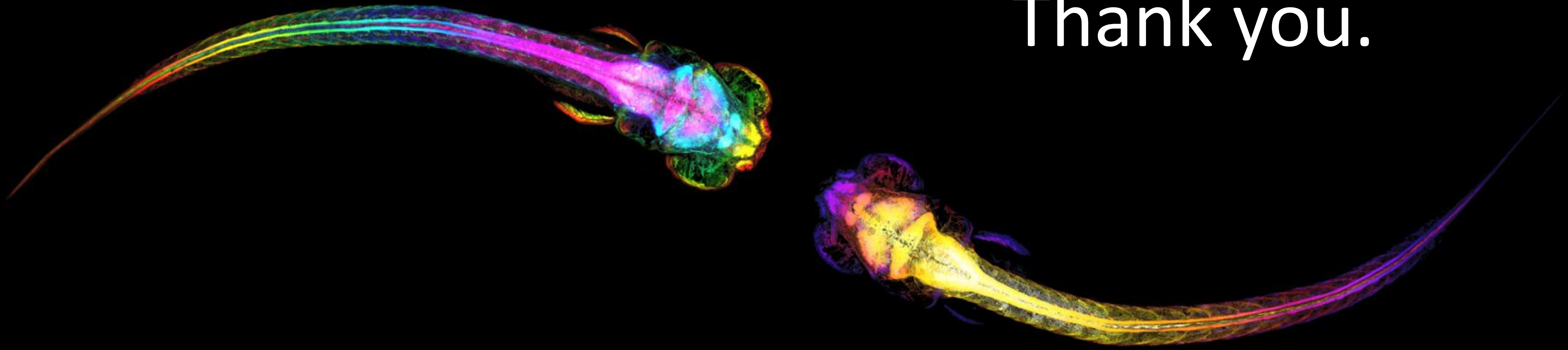


Never stop optimisation. Call it continuous integration in supercomputing…or a deep obsession with going as fast as you can, for the outcome.

The most important people in the building.

Architecture is just a stepping stone for us to innovation.

Thank you.



THE UNIVERSITY OF QUEENSLAND AUSTRALIA | QBI
Queensland Brain Institute
ART IN NEUROSCIENCE
WWW.QBI.UQ.EDU.AU