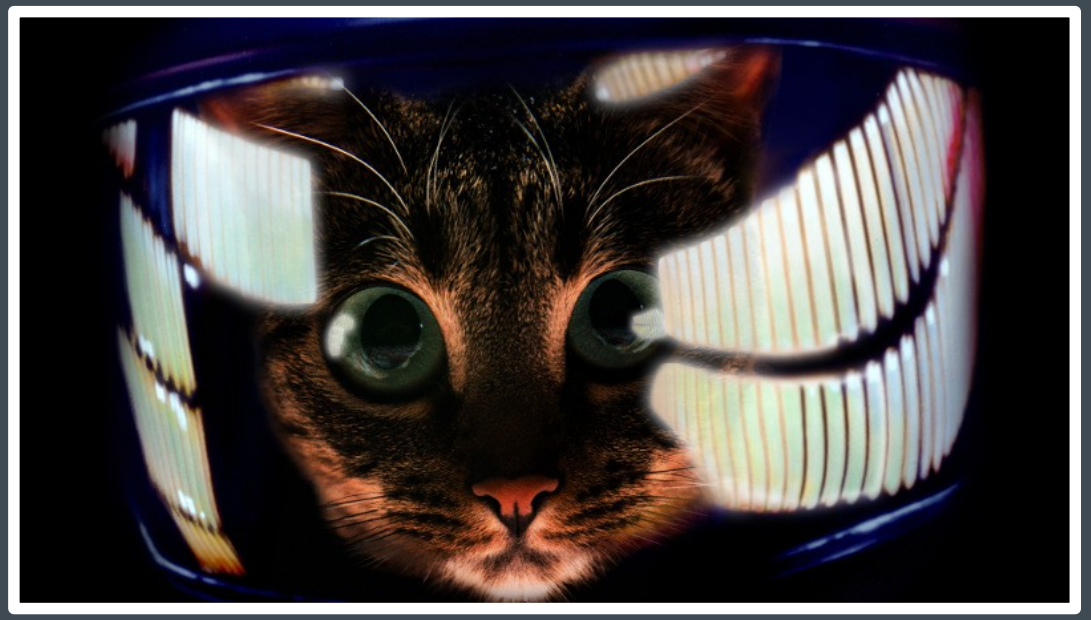
Embedded Supercomputing: Radio Astronomy at the Limit

Simon Ratcliffe

Technical Lead, Scientific Computing

JSNTB

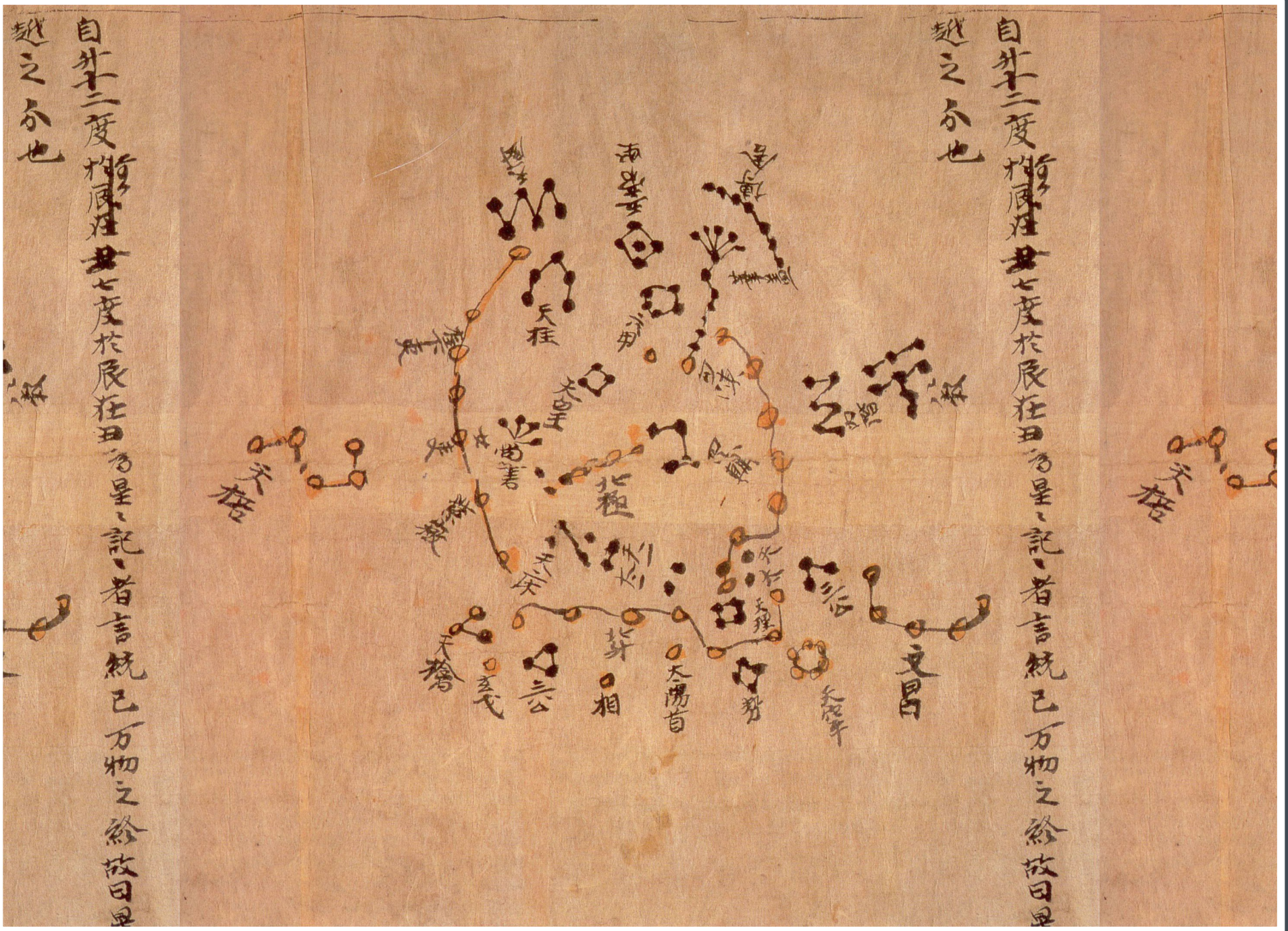
Starz ?



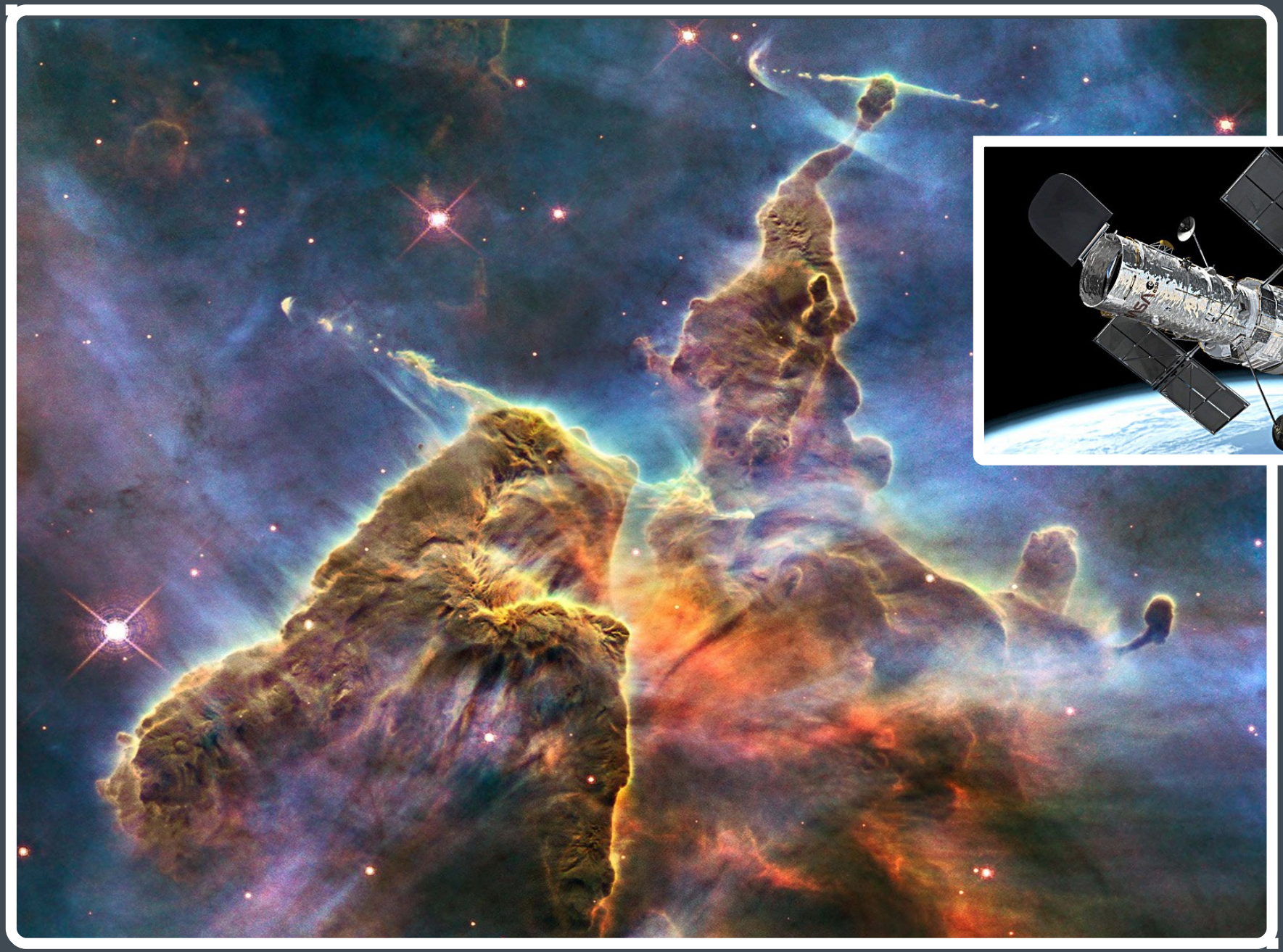
We are here

* The cake is a lie...

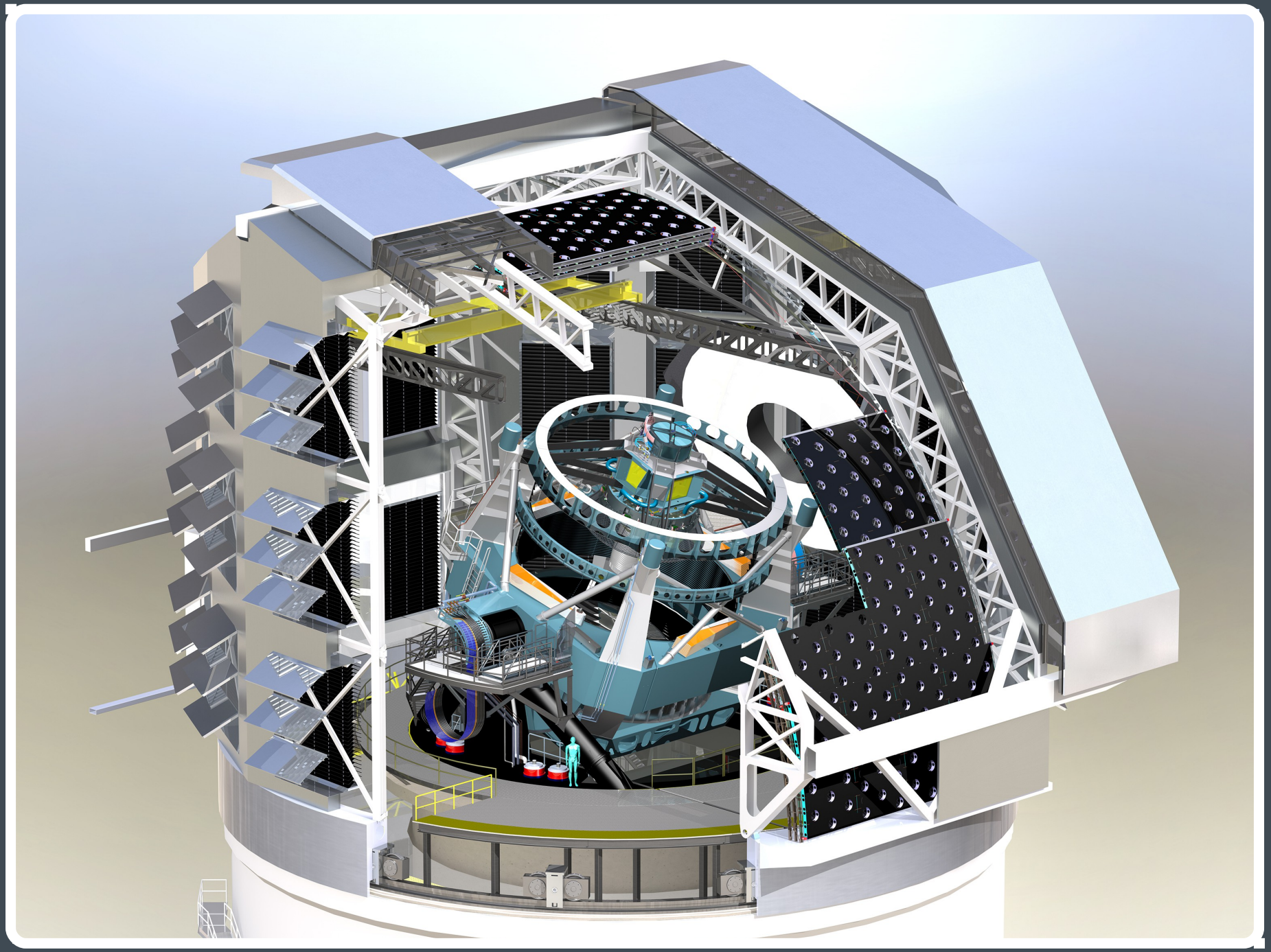
The first observational astronomer



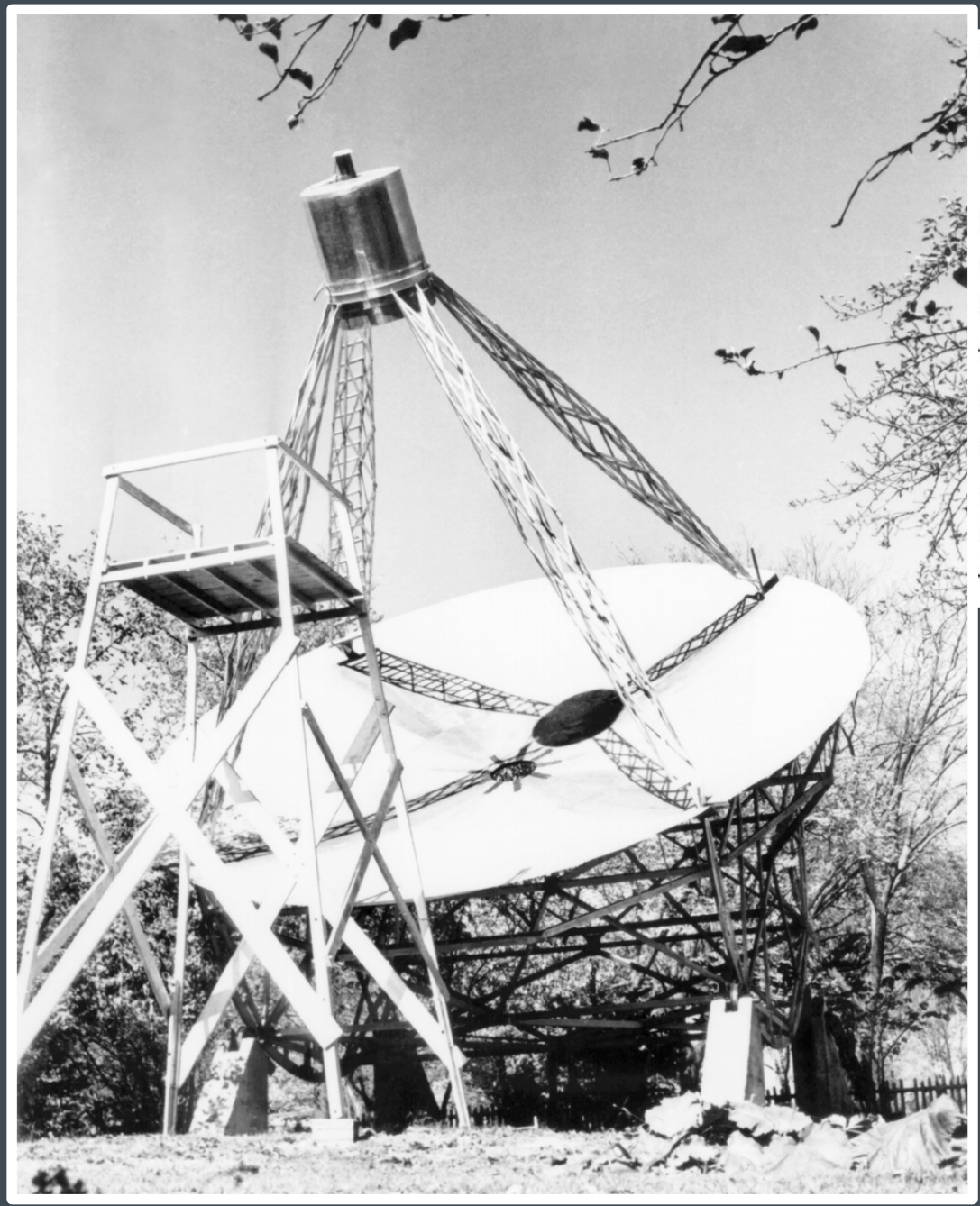
Dunhuang Map – 710 AD

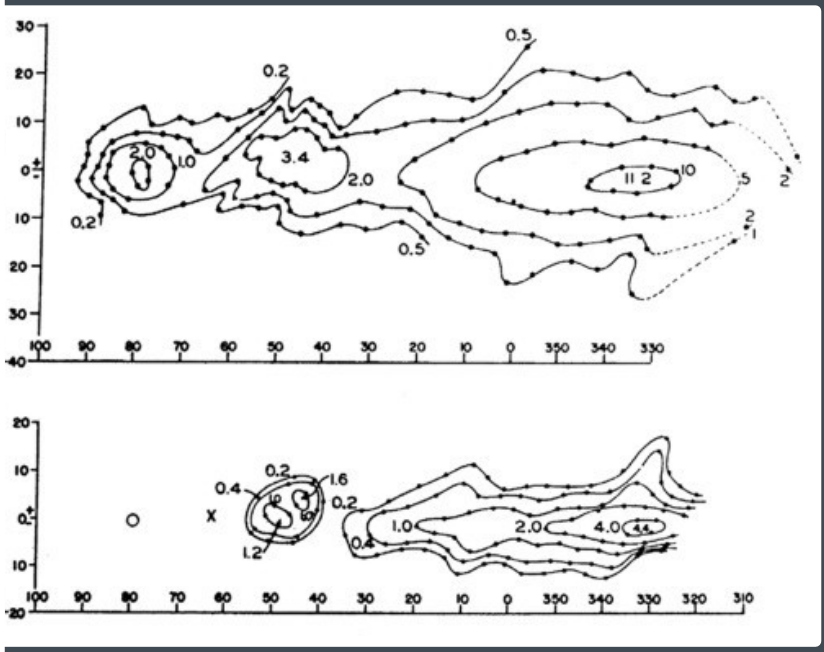


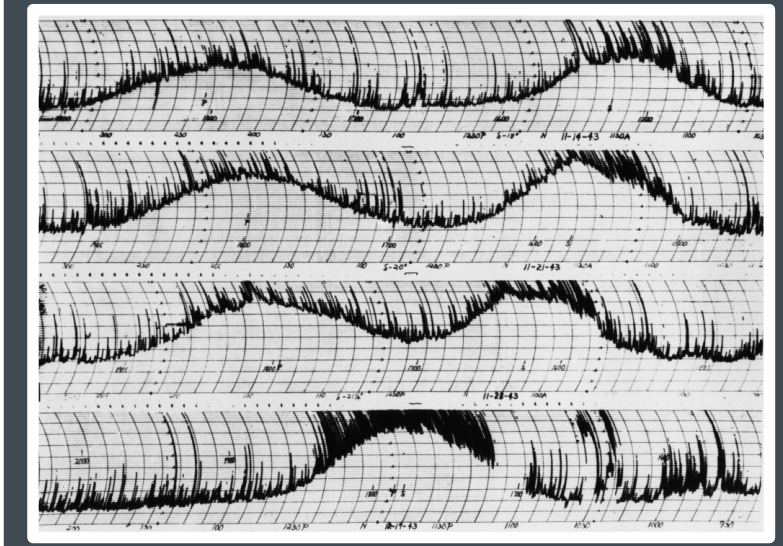
HST: Science archive around 10 TB



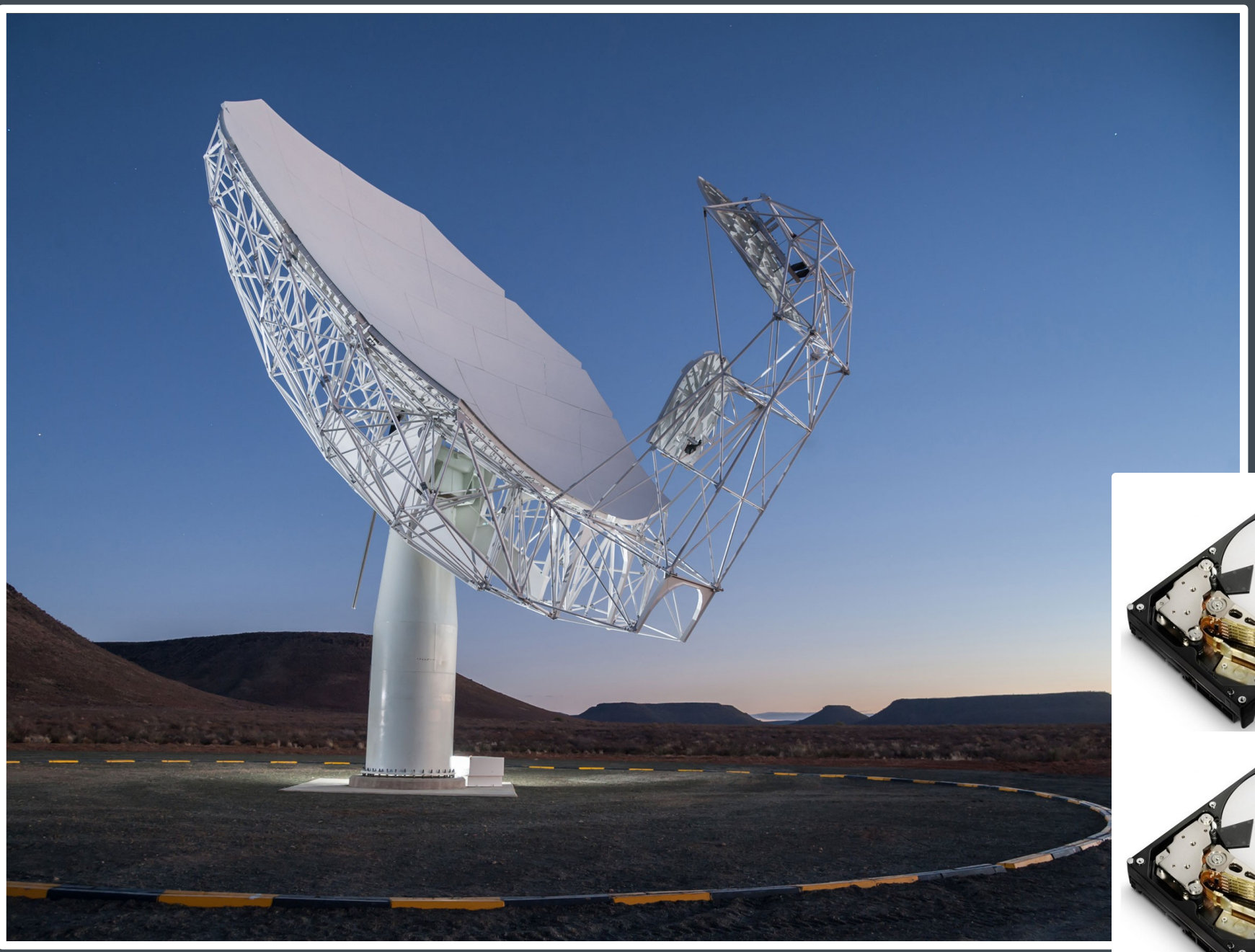
LSST: Full Sky Survey every 2 days → 37 Billion stars



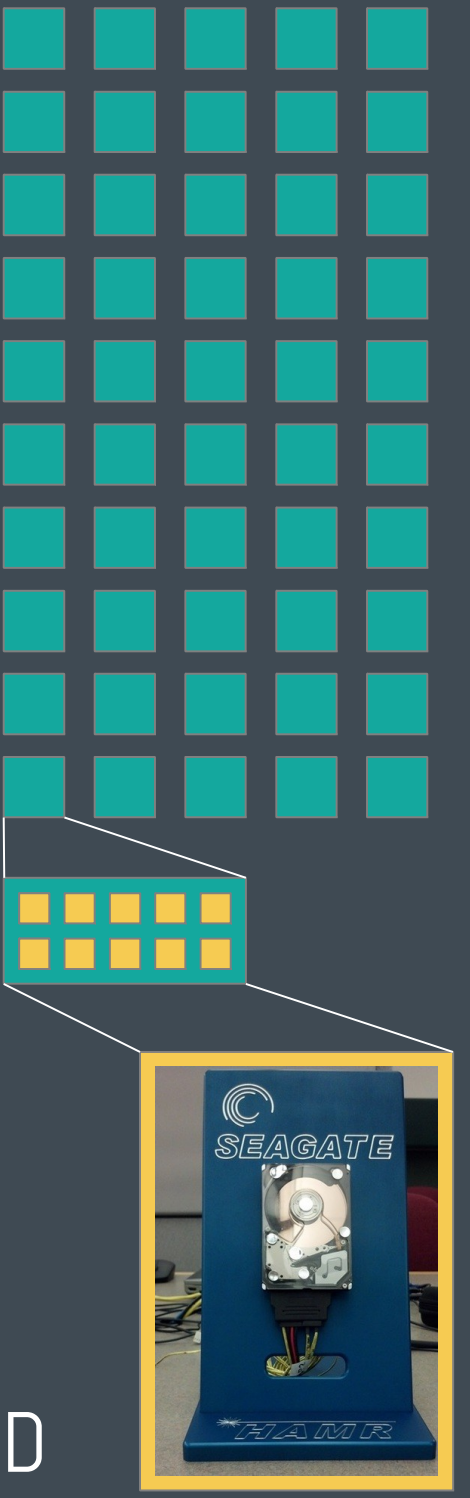




Grote Reber: Few hours - Few Feet of chart Paper (1941)



MeerKAT: 16 TB per 8 hours - 2 x 8TB HDD (2016)

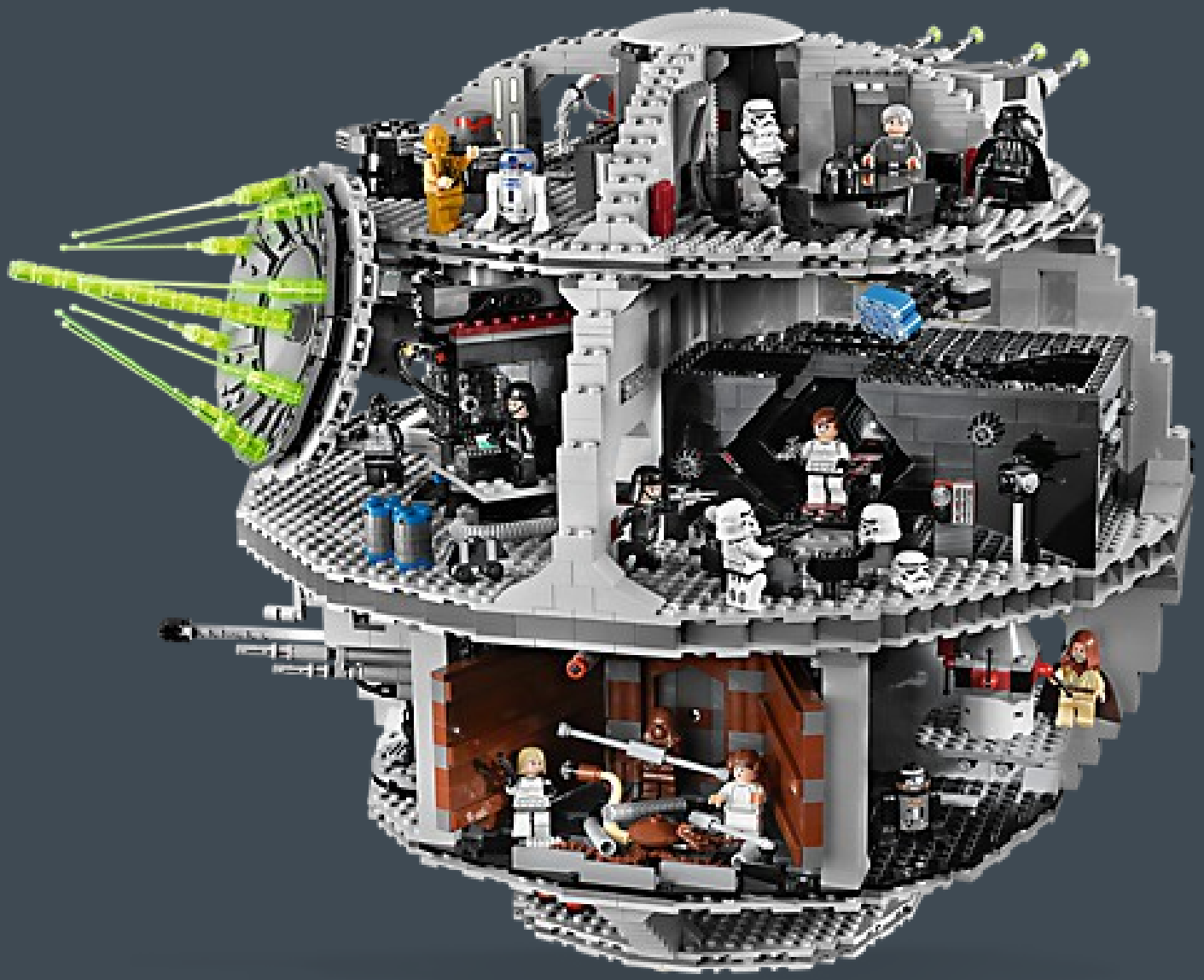


SKA1-MiD: 50 PB per 6 hours - 500 x 100TB HDD

65 exbibyte

- 65 exabyte

124,247 x



Big iron

$$\vec{V}_{ij} = M_{ij} B_{ij} G_{ij} D_{ij} E_{ij} P_{ij} T_{ij} \vec{V}_{ij}^{\,IDEAL}$$

PB / FLOPS / kW / $

Hardly the Kessel Run

29 ExaMACs
total MACs

1.1 Tibps
buffer read rate

1.7 ExaFlop
total FP operations

1.9 TB
working memory

Easy

Far, far away....

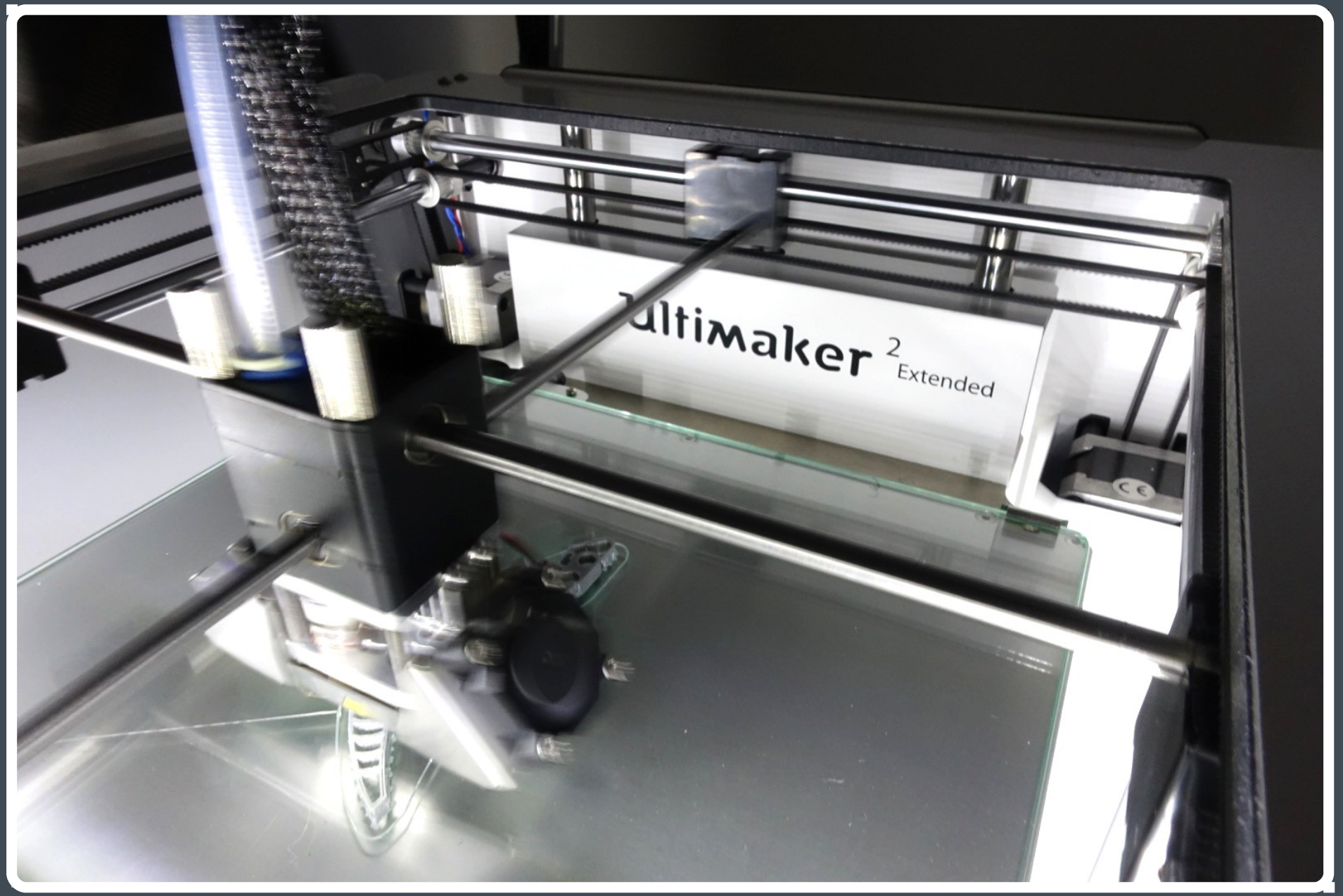
Just a flesh wound

25 kW

power target

Republic credits are no good here...

If you print it they will come

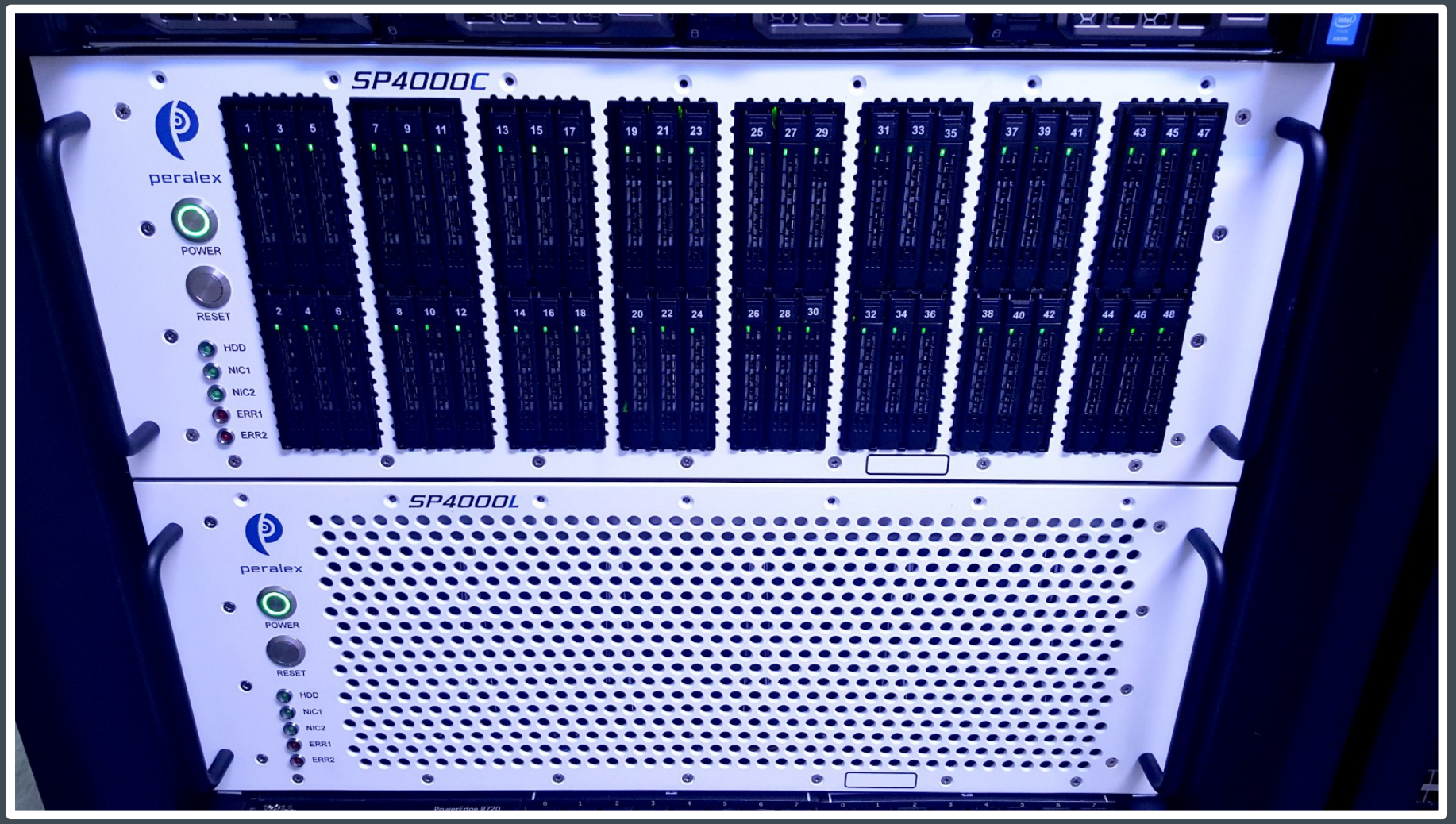


Big storage vendors love us :)

raw disk cost +

15%

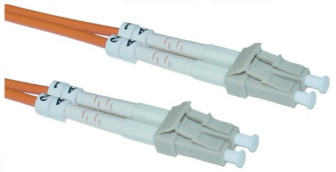
Just finished qualification....



High Speed Pod: 40 Gbps to disk / 40 TiB
Bulk Pod: 10+Gbps to disk / 360 TiB

Grand designs

Immersive cooling
No fans
Glycol ground loop

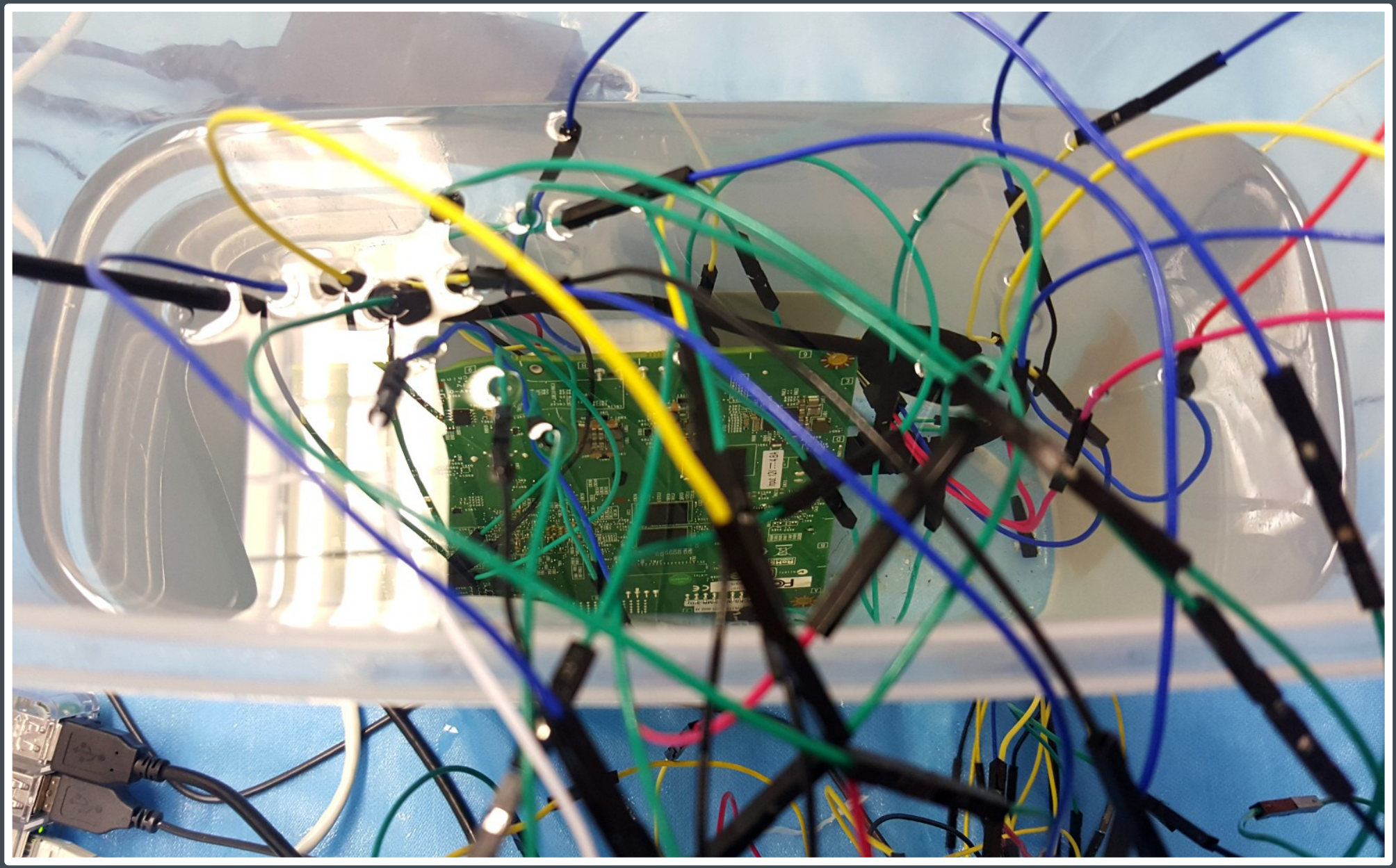


20 GBps In
2000 nodes
100 Mbpn

Tegra SoC
1W TDP
Cuda GPU
SO DIMM

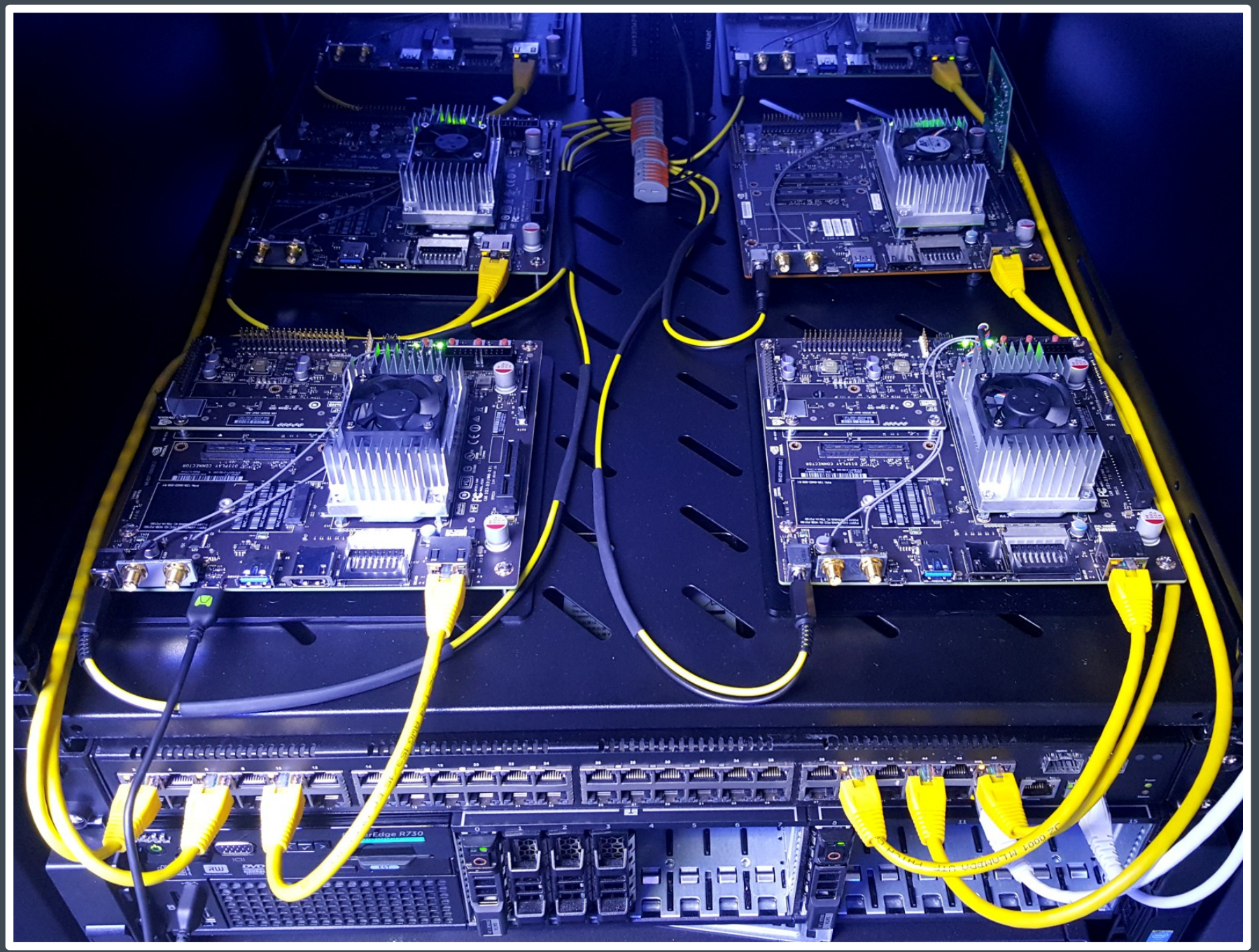
50 TFlops, 2 TB RAM, 8 TB Flash
2.5 kW Power, O kW Cooling, $350k built

Life support

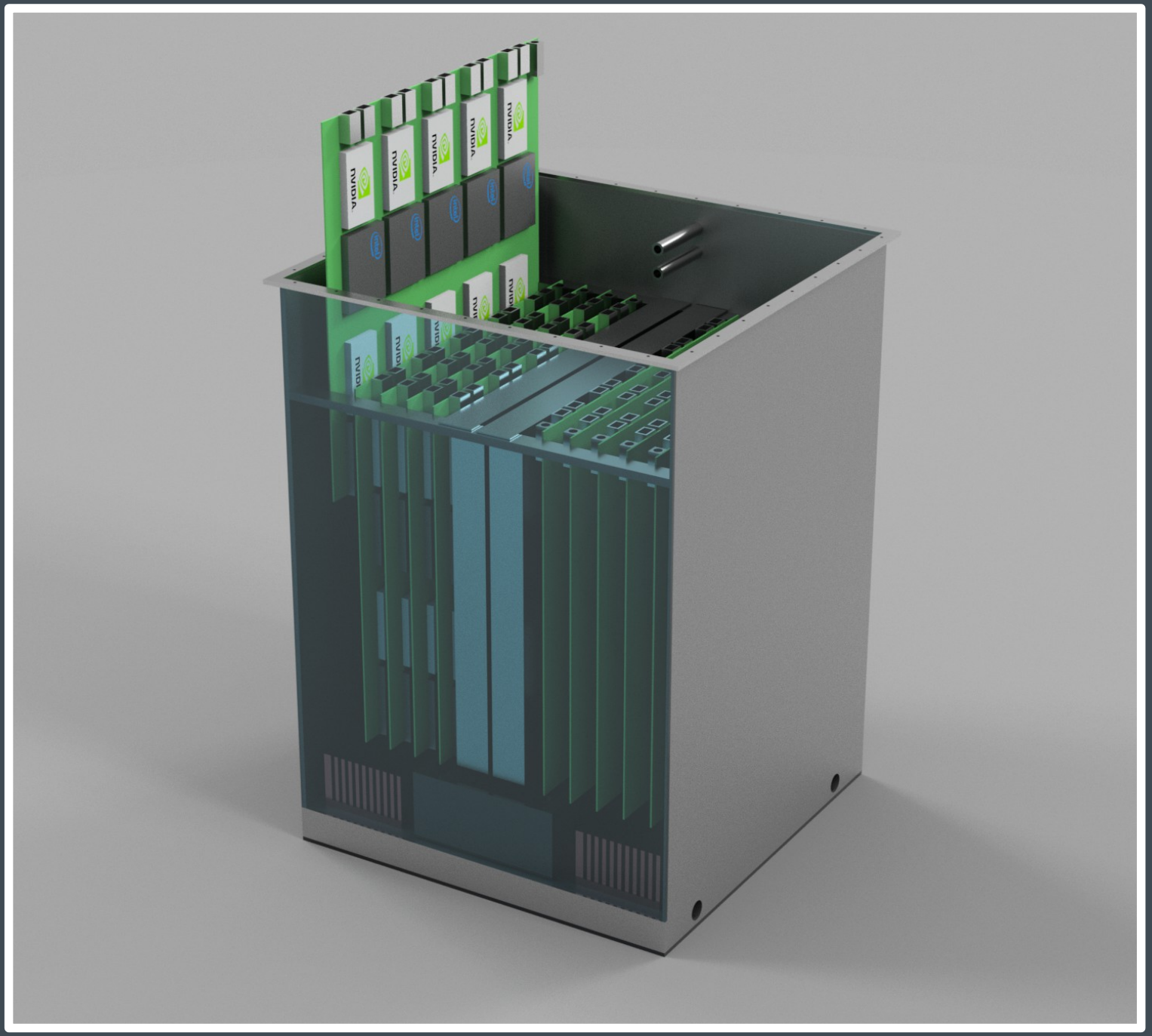


14 Months in oil

Software Stack Testbed



Coruscant Droid Spa



Same numbers as last year :)

TEGRA X1

1000 Nodes

Tegra X1
4 GB RAM
512 GB SSD

20 Switches

2 x 10 GbE SFP+
48 x 1 GbE

10 Pods

15M Ground Loop
100L Mineral Oil

Titan X

22 Servers

4 x Titan X
2 x E5-2660v3
12 x 2TB SATA
128 GB RAM

3 Switches

4 x 40 GbE QSFP
36 x 10 GbE SFP+

3 Racks

Just a rack

Not quite the same numbers as last year

TEGRA X1

$391 kilo

$313k Hardware
$40k Infrastructure

18.9 kW

18.4 kW Hardware
0.5 kW Cooling

Titan X

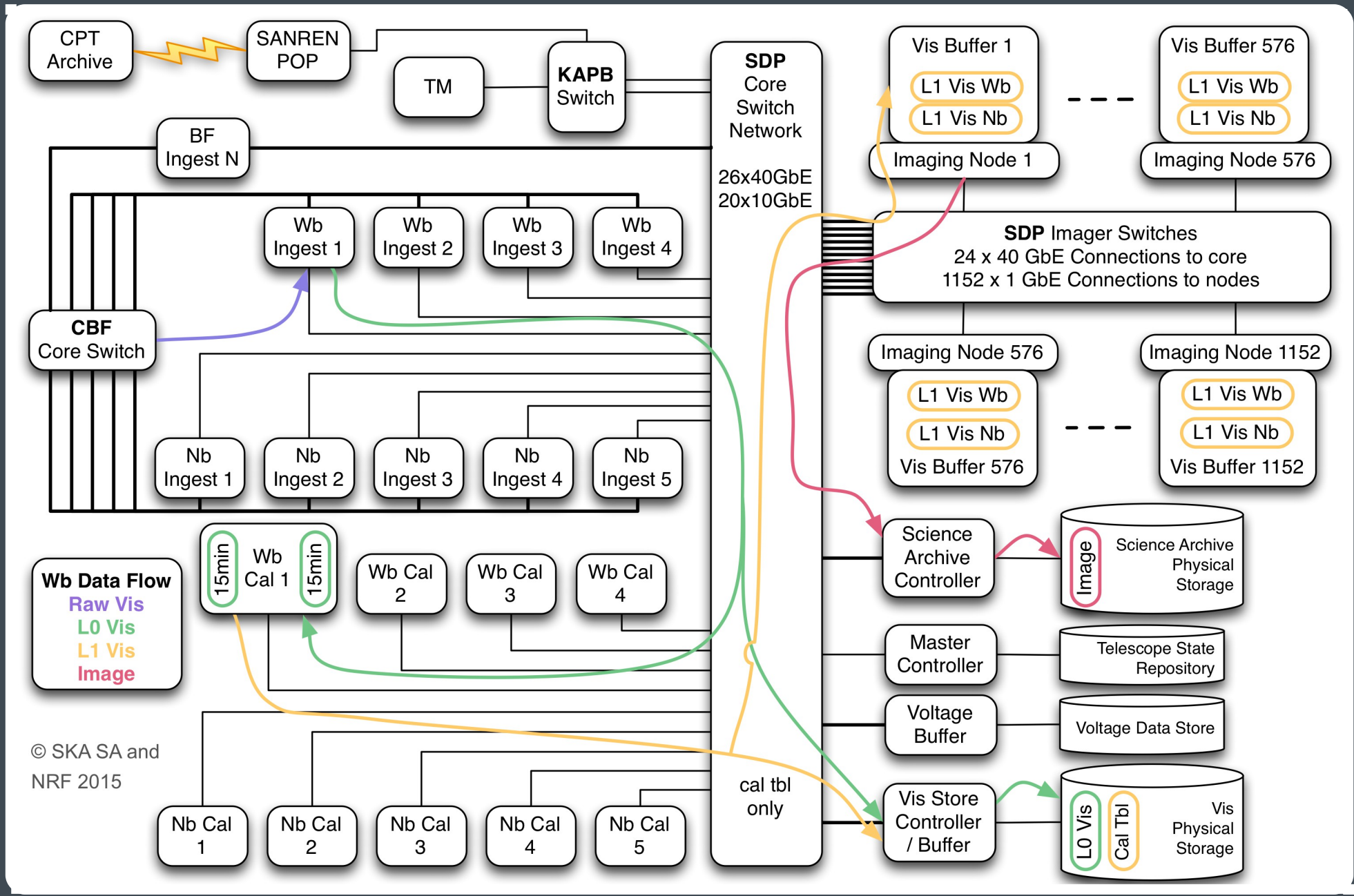
$441 kilo

$322k Hardware
$118k Infrastructure

33.7 kW

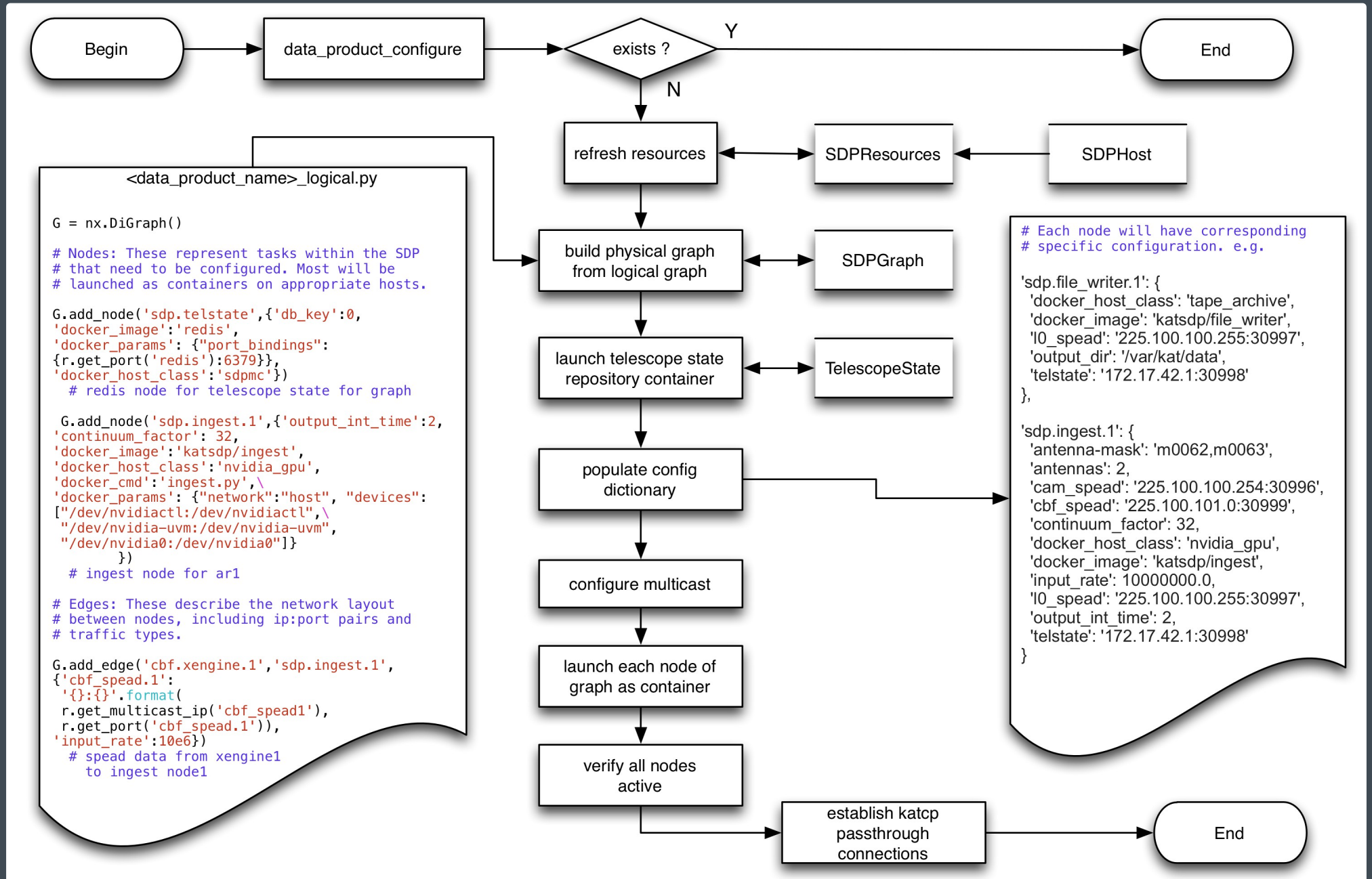
25.9 kW Hardware
7.8 kW Cooling

Top level data flow



© SKA SA and NRF 2015

Graphs all the way down until you hit the turtles



Telescope State Repository

Using redis as key→ts→value store

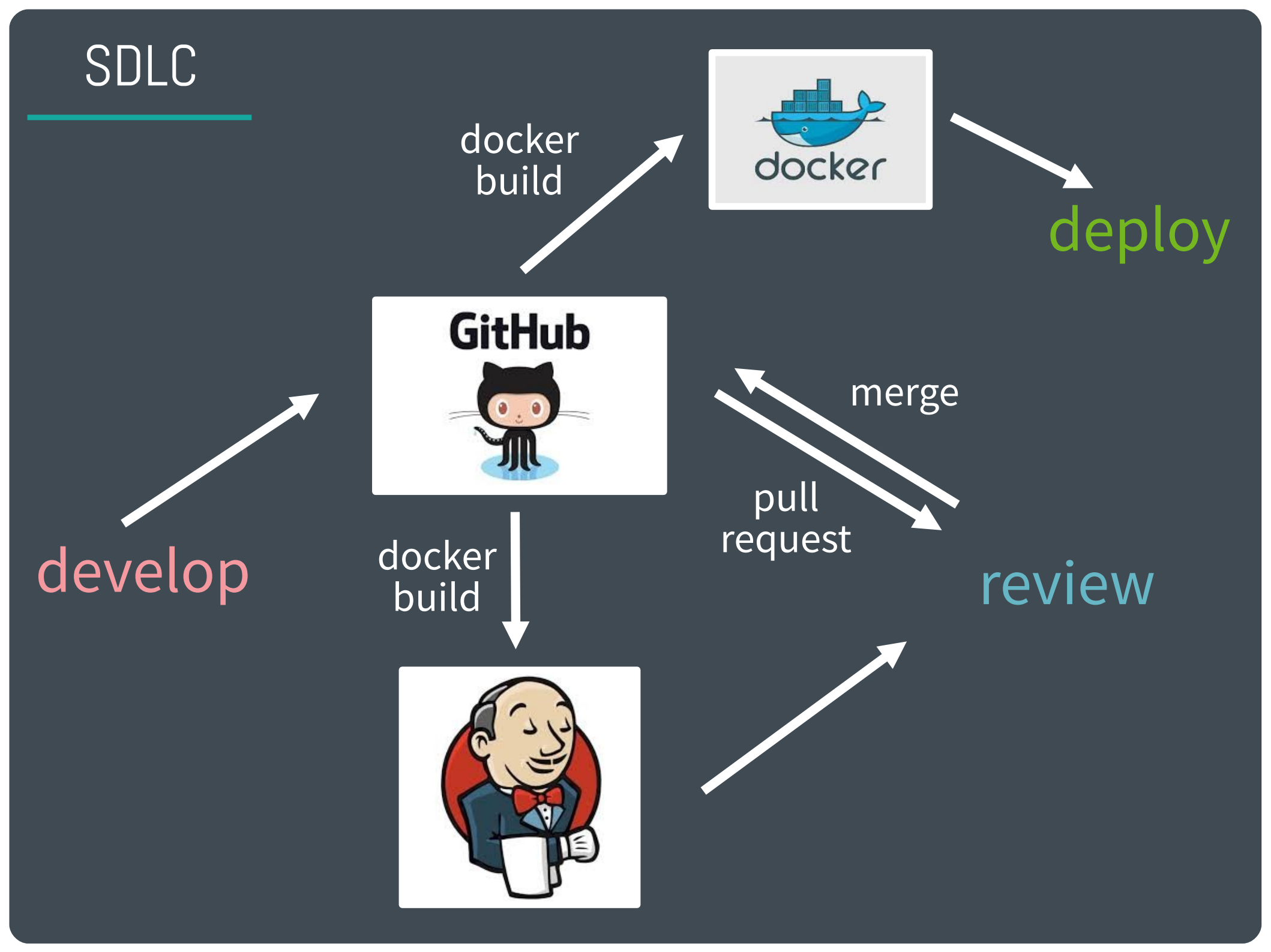
Implemented using zrange

Provides static configuration for stateless components

Dynamic data from upstream (CAM) and internal (e.g. CAL)

State repository per subarray (graph)

SDLC


docker
build
docker
deploy
GitHub
merge
pull
request
develop
docker
build
review

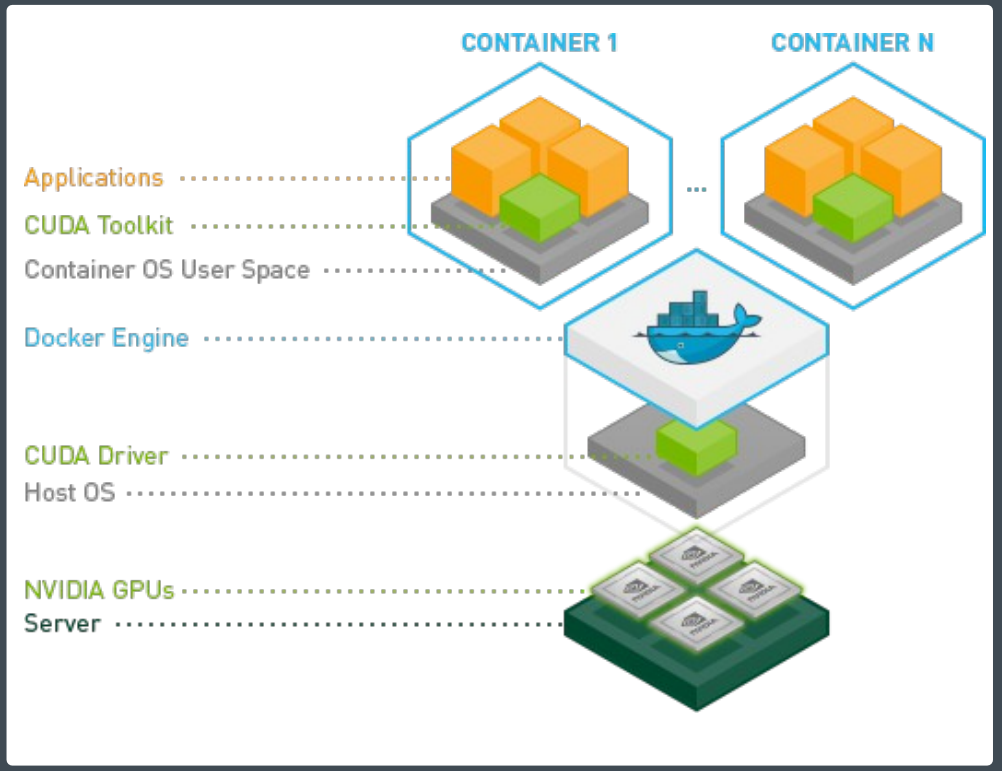
Imager Software Stack

Based on an Ubuntu 14.04 GPU container

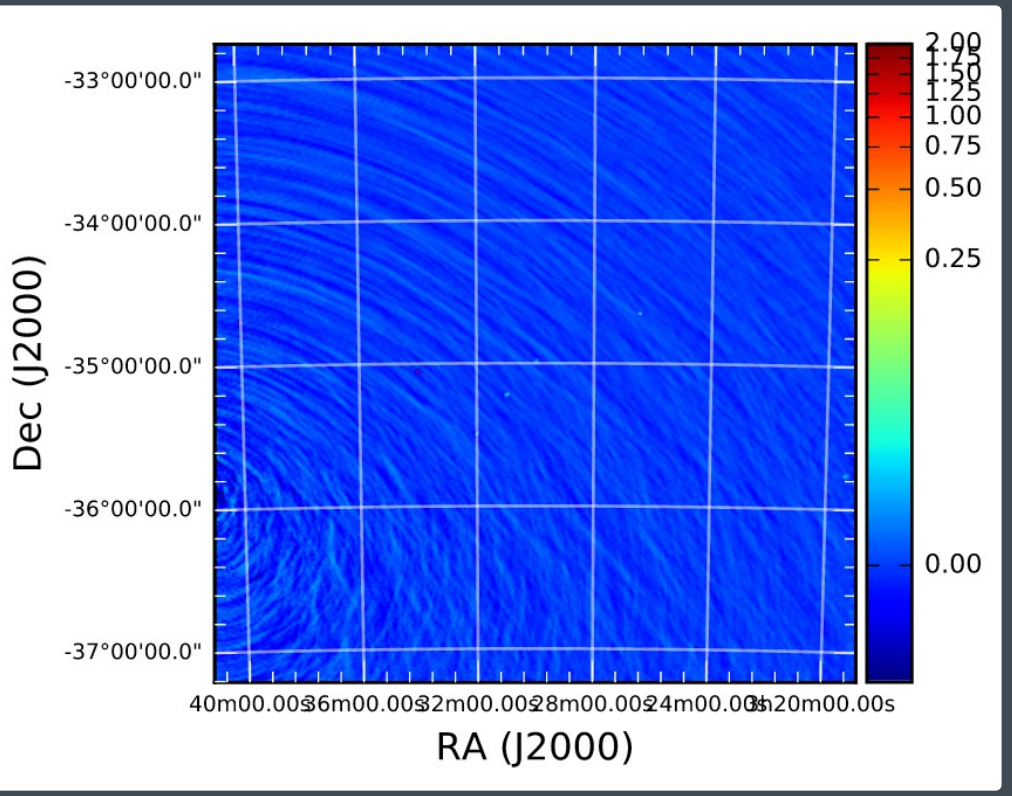
Auto-tuned for target architecture (TK1/TX1)

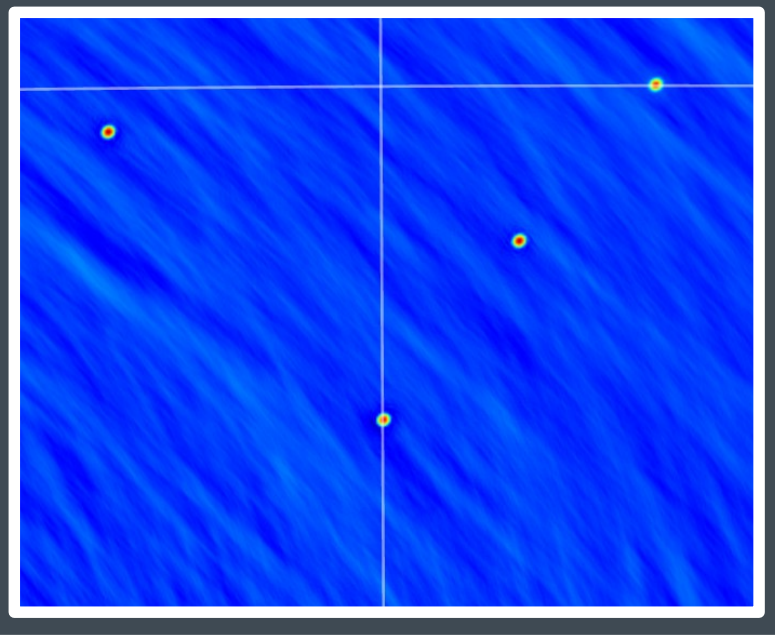
Mostly Python via PyCuda

Grid → FFT (cuFFT) → Degrid

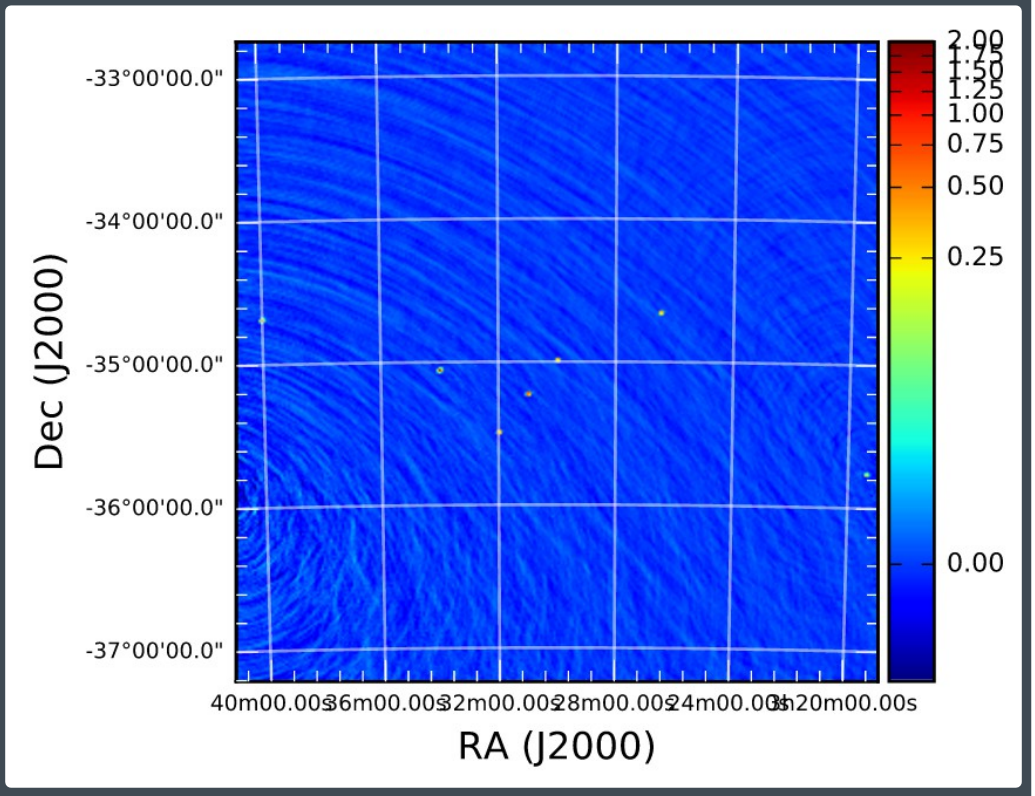


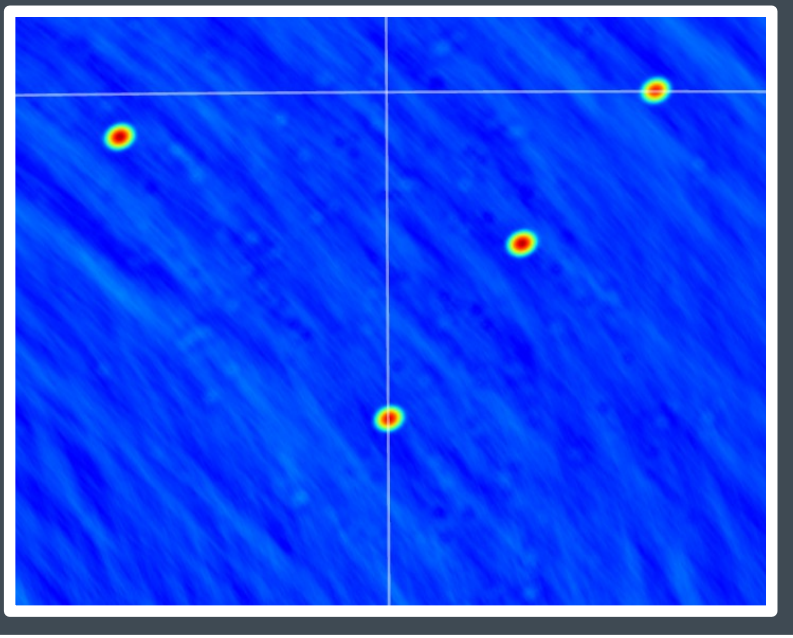
lwimager





katsdpimager





TX1 Issues

Base deployment via Ansible roles

Multiple steps to get up to date docker:

- Cross compiled custom kernel (incl aufs)
- Cross compiled Go 1.4.2 used to bootstrap 1.6
- Hack build docker *(AUTO_GOPATH=1 ./hack/make.sh dynbinary)*
- Docker build docker

```
ubuntu@tegra-ubuntu:~$ docker version
Client:
 Version:      1.10.2
 API version:  1.22
 Go version:   go1.6
 Git commit:   c3959b1
 Built:        Thu Mar 24 13:25:32 2016
 OS/Arch:      linux/arm
```

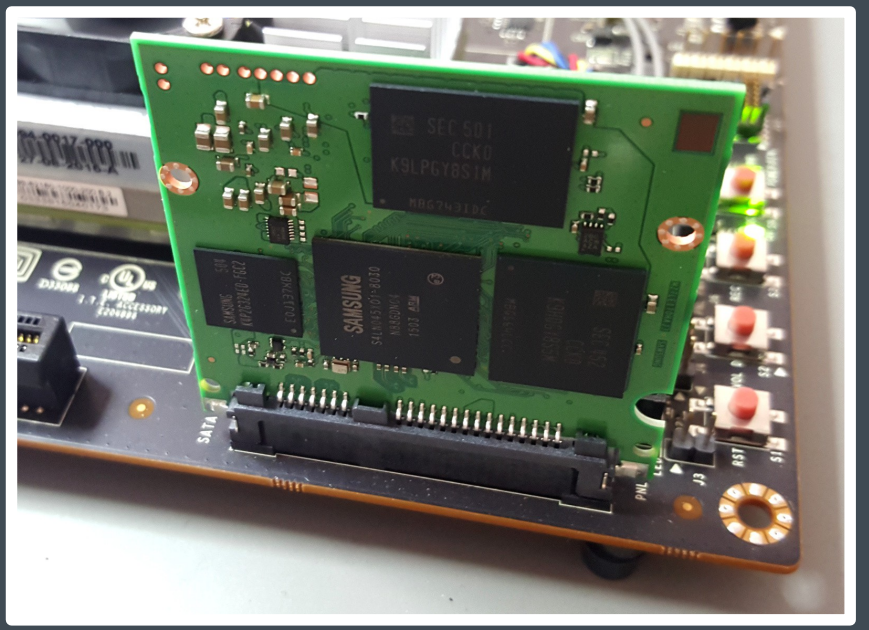
Memory

Key limiting factor → constrains image size

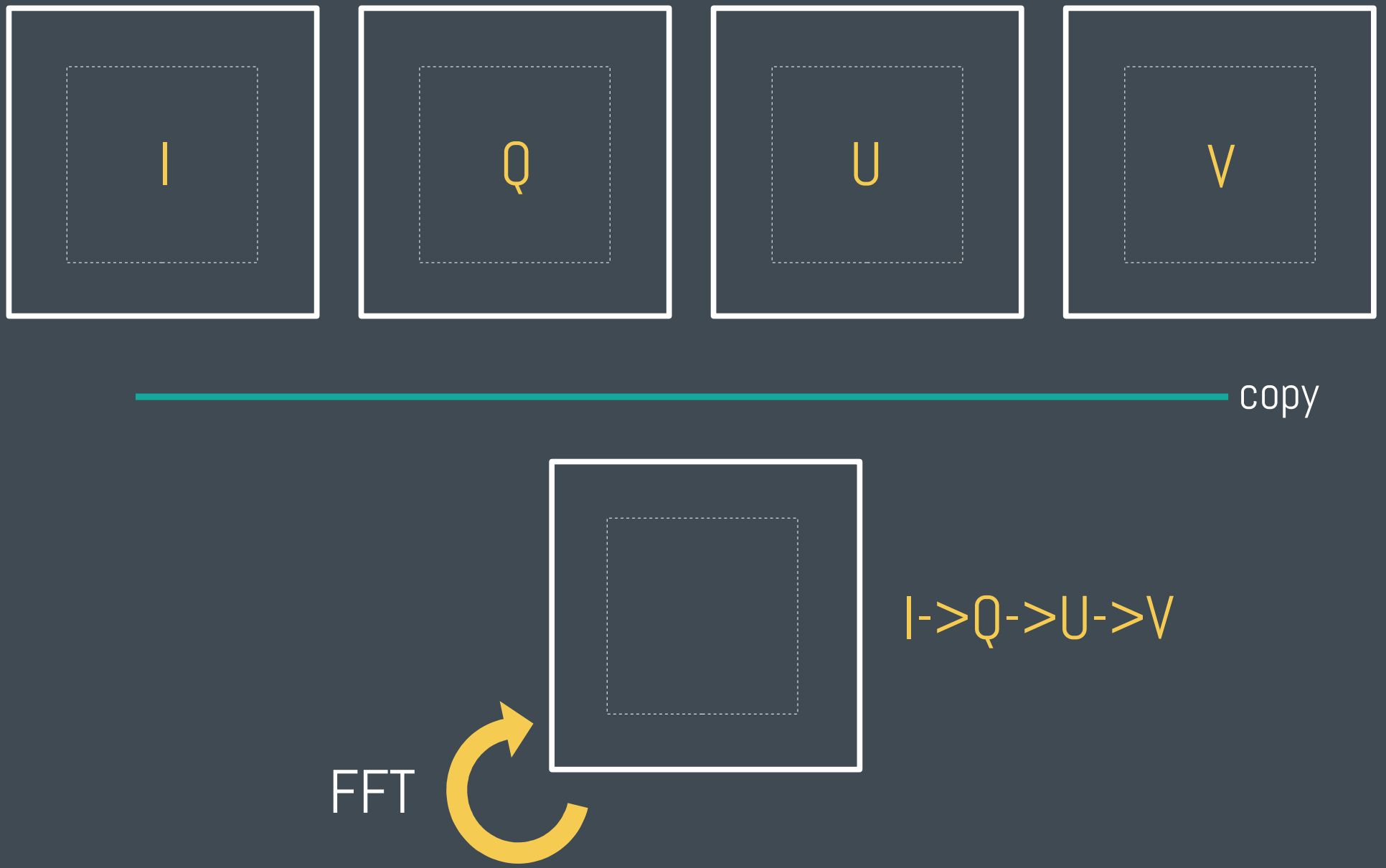
Fast(ish) local disk allows for per major cycle loads

cudaMallocManaged bug crashes driver

32-bit userland...



Memory Optimisation (Gridding)



Gridding

Romein 2012: exploit spatial coherence to reduce global memory traffic

Muscat 2014: combine visibilities that have the same subgrid coordinates

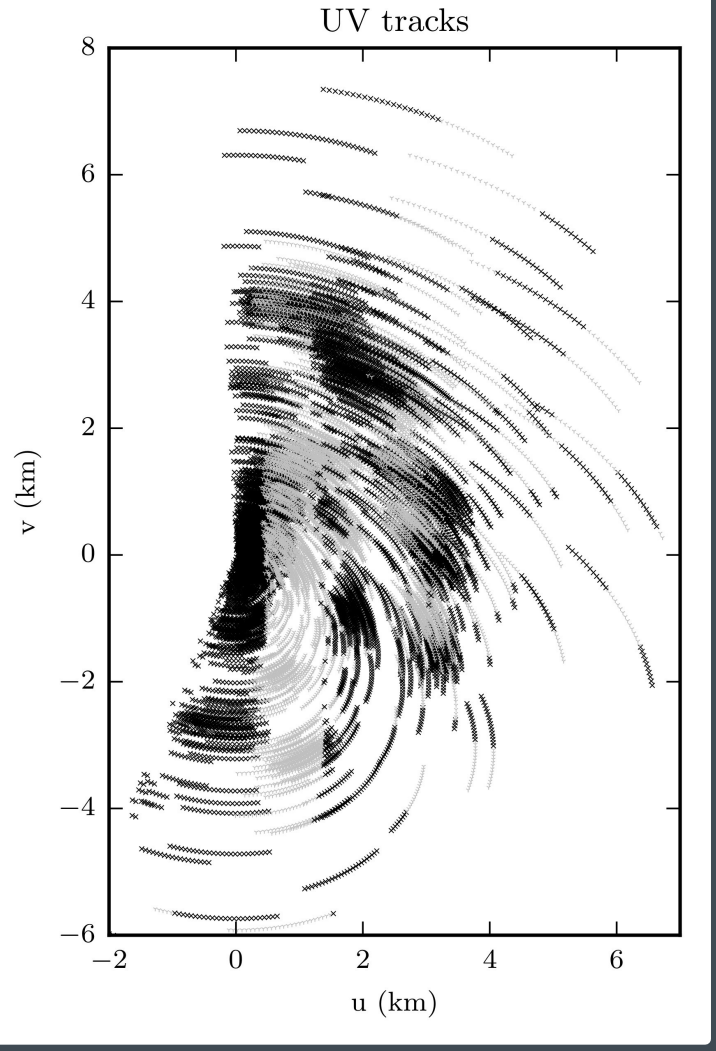
Merry 2015: approximate W kernel as separable

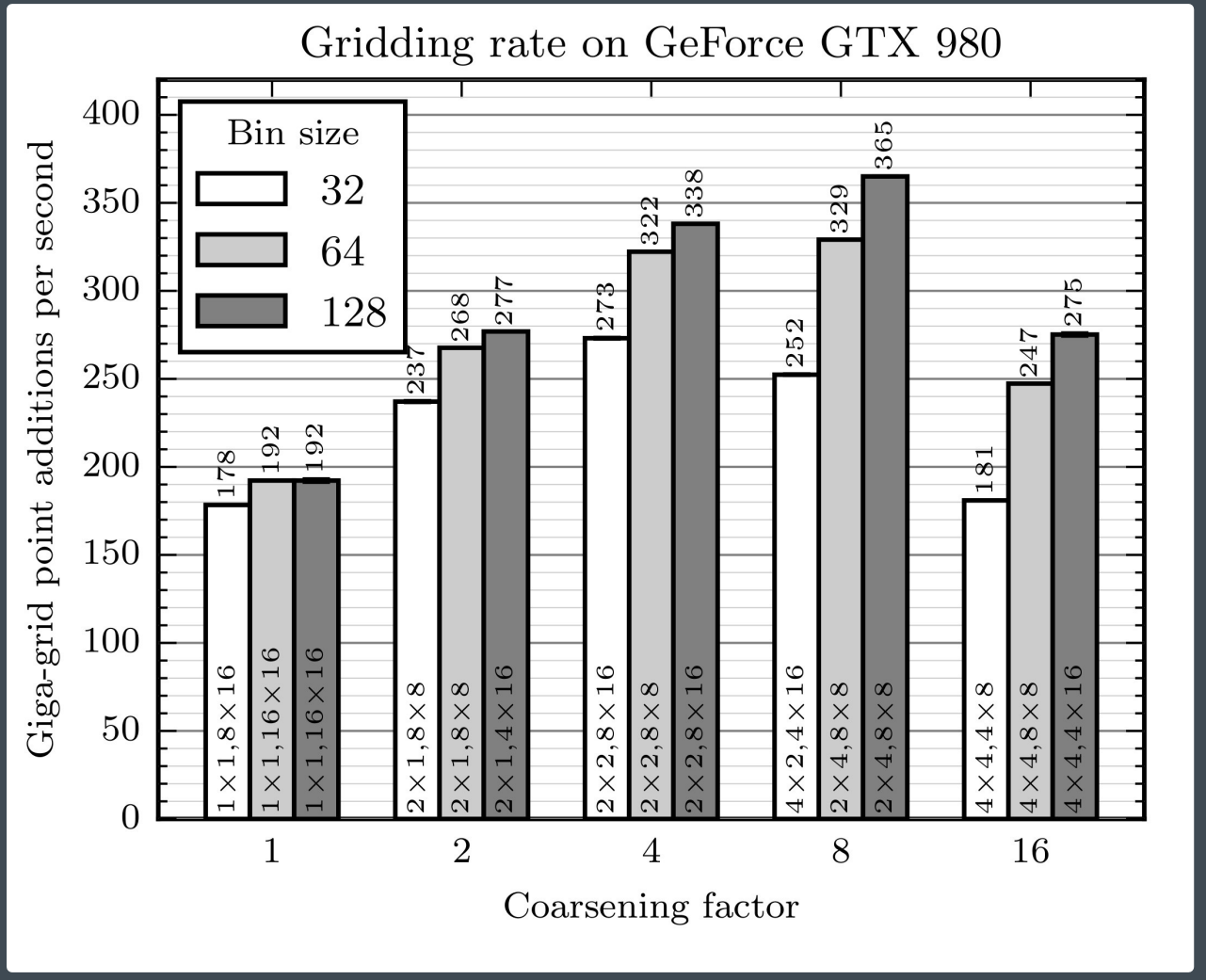
Merry 2016: use thread-coarsening to improve efficiency

>50%

theoretical peak

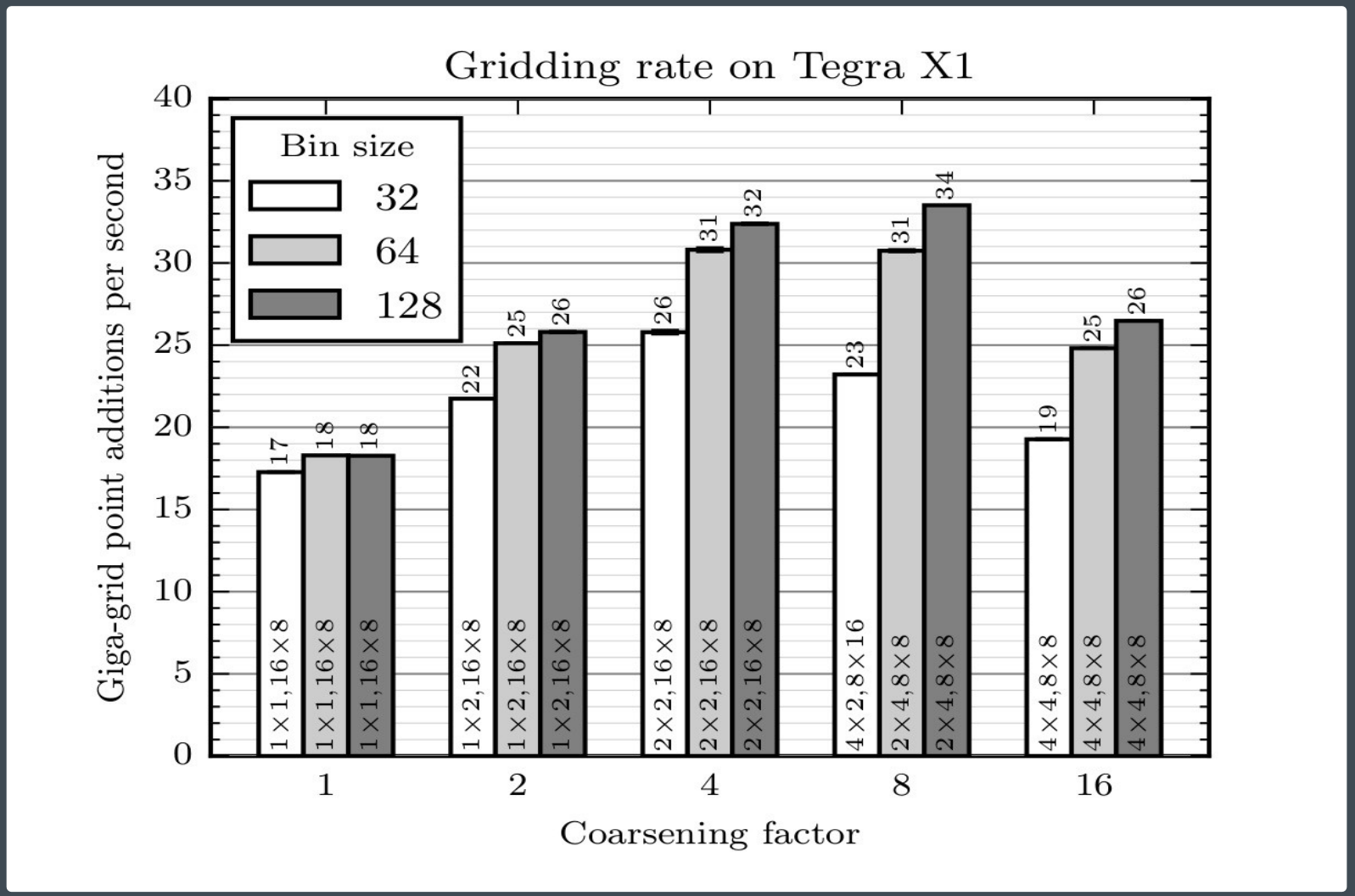
Gridding





2hr MeerKAT Simulated Data: 365 GGPA/s at 55% efficiency.

Hot off the grid(dle)



34 GGPA/s at 51% efficiency

~10% of 980ti performance

Bibliography (Gridding)

Merry, B., 2016. Approximating W projection as a separable kernel. MNRAS 456, 1761–1766.

Merry, B., 2016. Faster GPU-based convolutional gridding via thread coarsening. Under review.

Muscat, D., 2014. High-Performance Image Synthesis for Radio Interferometry. Master's thesis. University of Malta.

Romein, J.W., 2012. An efficient work-distribution strategy for gridding radio-telescope data on GPUs, in: Proc. 26th ACM Int. Conf. on Supercomputing, ACM, New York, NY, USA. pp. 321–330

& Beyond

64 antenna MeerKAT complete early 2017

194 antenna SKA Phase 1 construction starts 2019.

1500+ antenna SKA Phase 2 starts construction 2025.

www.ska.ac.za