Performance Engineering for the SKA telescope

Peter J Braam
SKA Science Data Processor
Cavendish Laboratory, Cambridge University
peter.braam@peterbraam.com

Mar 2018

Acknowledgement

A large group of people (~500) are working on this project
Most information is publicly available, but very technical
This presentation re-uses much from other SKA efforts

Particularly I'm using a few slides from Peter Wortmann
Background: skatelescope.org

My role: consultant & visiting acaemic for Cambridge group since 2013

Message from this talk

1. SKA telescope is a grand challenge scale project

2. Synergy between scientific computing and industry for performance

    Hardware – particularly memory, energy

    Software – agility, parallelism, energy

3. General purpose tools appear insufficient, there may be fairly deep open issues

What is the SKA?

The Square Kilometre Array (SKA)

Next Generation radio telescope – compared to best current instruments it will offer …

- ~ 100 times more sensitivity
- ~ 10^6 times faster imaging the sky
- More than 5 square km of collecting area over distances of >100km

Will address some of the key problems of astrophysics and cosmology (and physics)

- Builds on techniques developed originally in Cambridge
- It is an Aperture Synthesis radio telescope (“interferometer”)

Uses innovative technologies...

- Major ICT project
- Need performance at low unit cost

SPDO/Swinbur

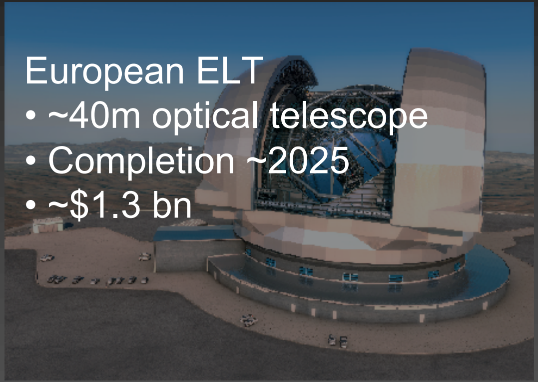
SKA International Design Consortia

Project Management and System Engineering Team based at JBO (UK)

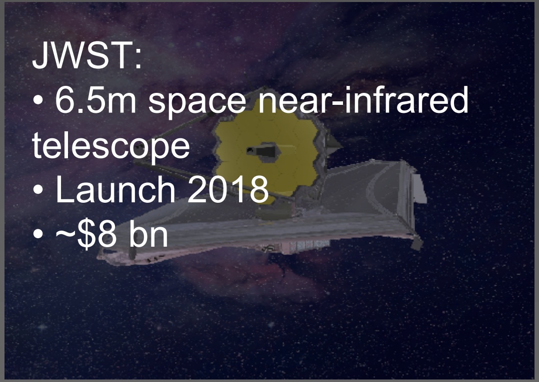
~500 scientists & engineers in institutes & industry in 11 Member countries

SKA – a partner to ALMA, EELT, JWST

Credit:A. Marinkovic/XCam/ALMA(ESO/NAOJ/NRAO)



Credit:ESO/L. Calçada (artists impression)



Credit: Northrop Grumman (artists impression)

Credit: SKA Organisation (artists impression)

In summary …

- SKA aims to be a world class “instrument” like CERN

- SKA Phase 1 – in production 2025
- SKA Phase 2 – likely 10x more antennas – 2030’s?

- This presentation focuses on SKA1

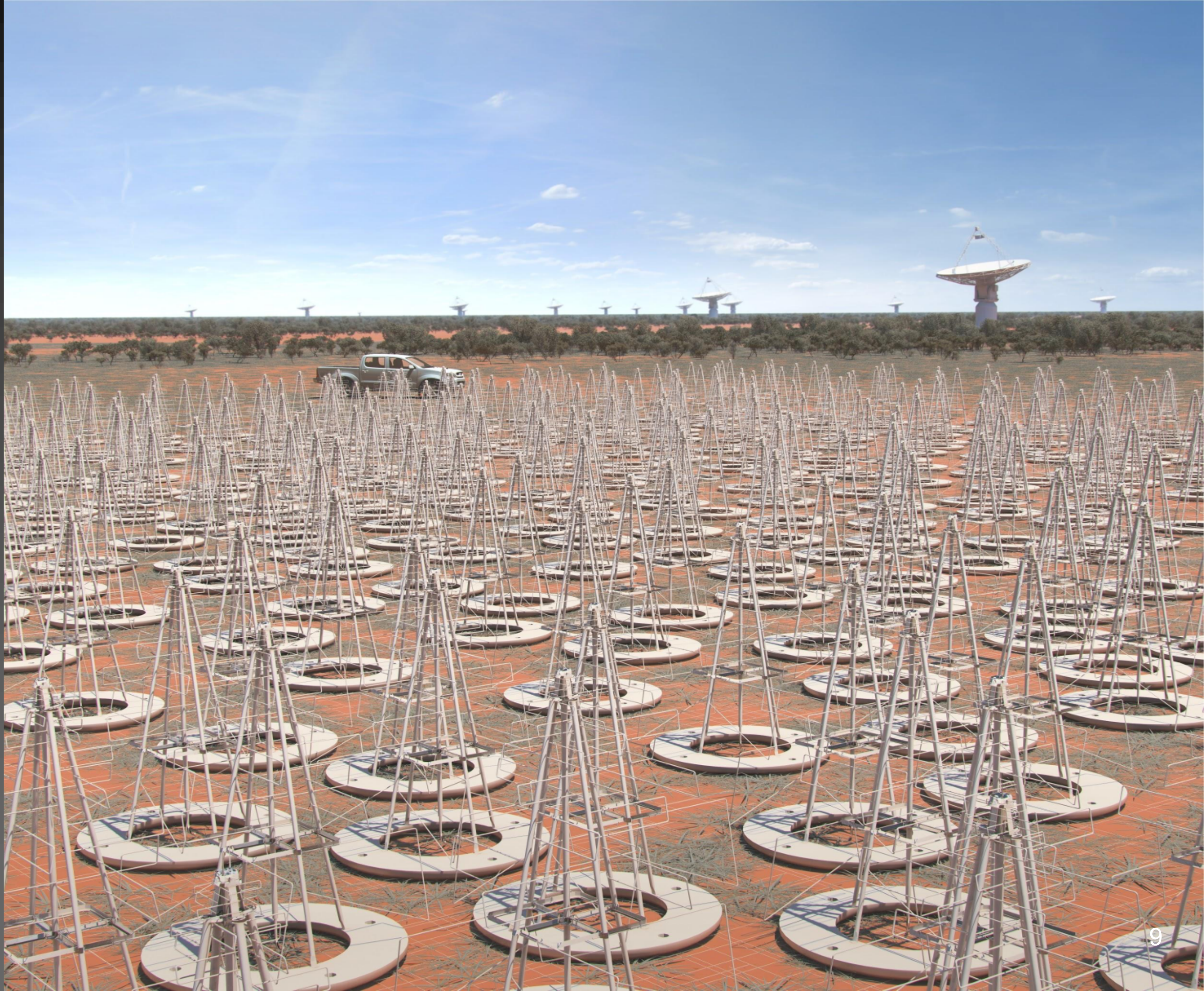
- Caveat
  - Ongoing changes
  - Some inconsistencies in the numbers

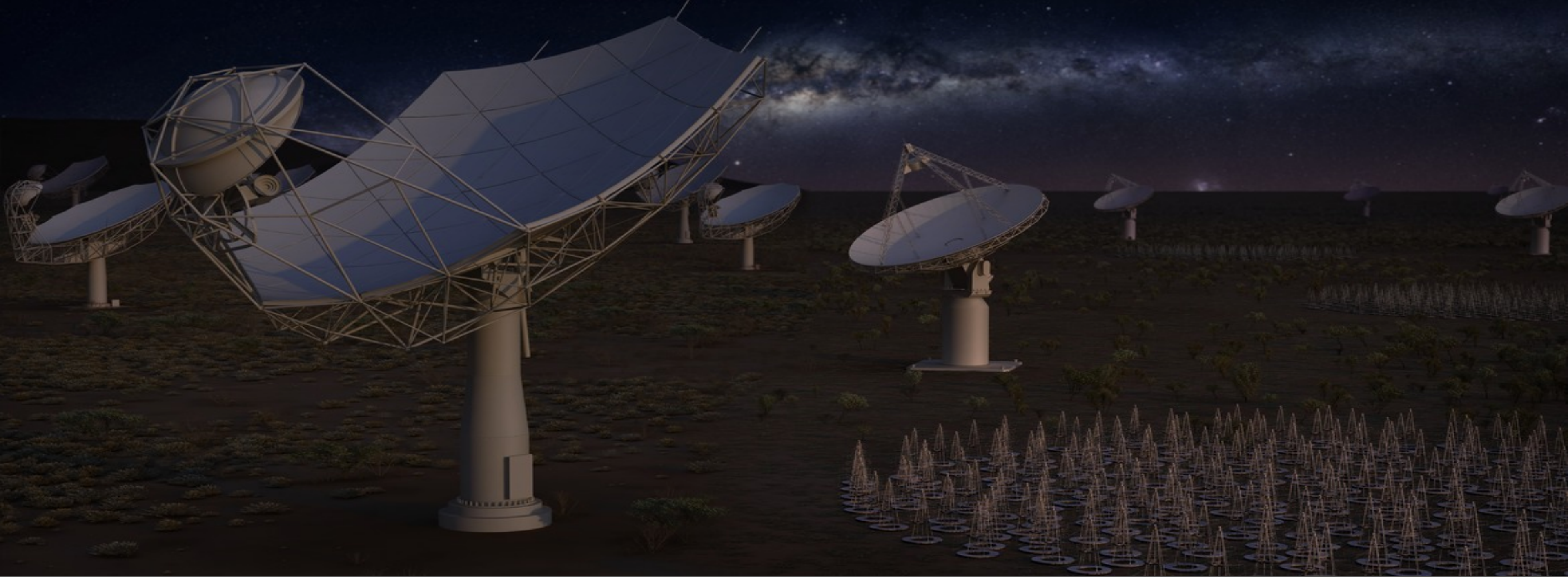
Low Frequency Aperture Array
0.05 – 0.5 GHz

Australia

~1000 stations
256 antennas each
phased array with
Beamformers

Murchison Desert
0.05 humans/km^2
Compute in Perth



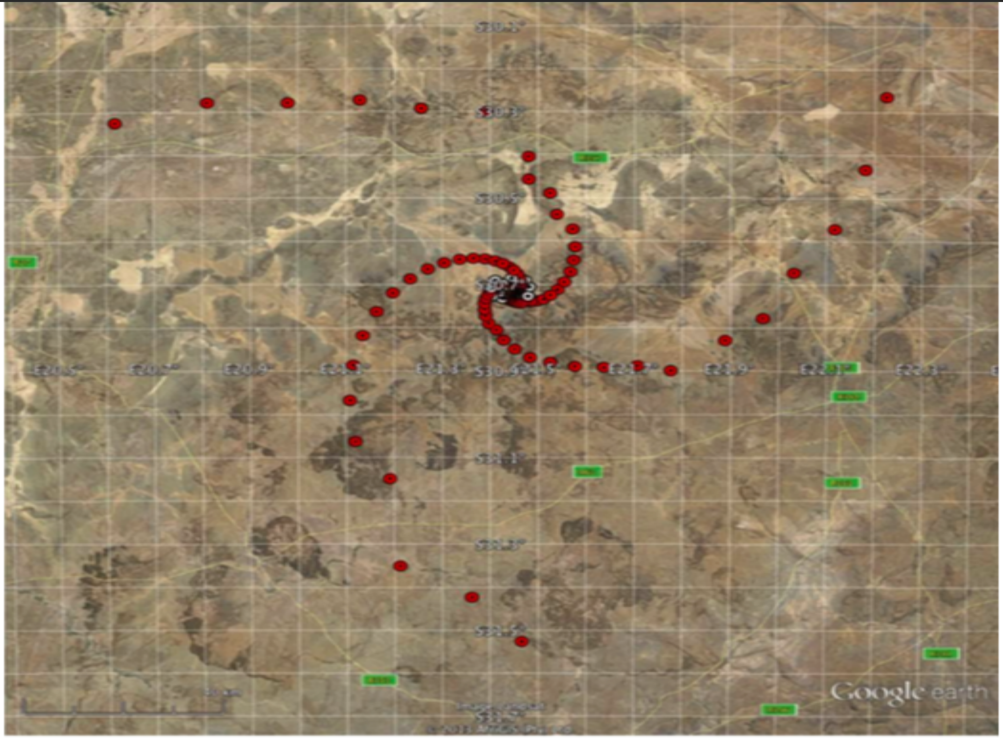


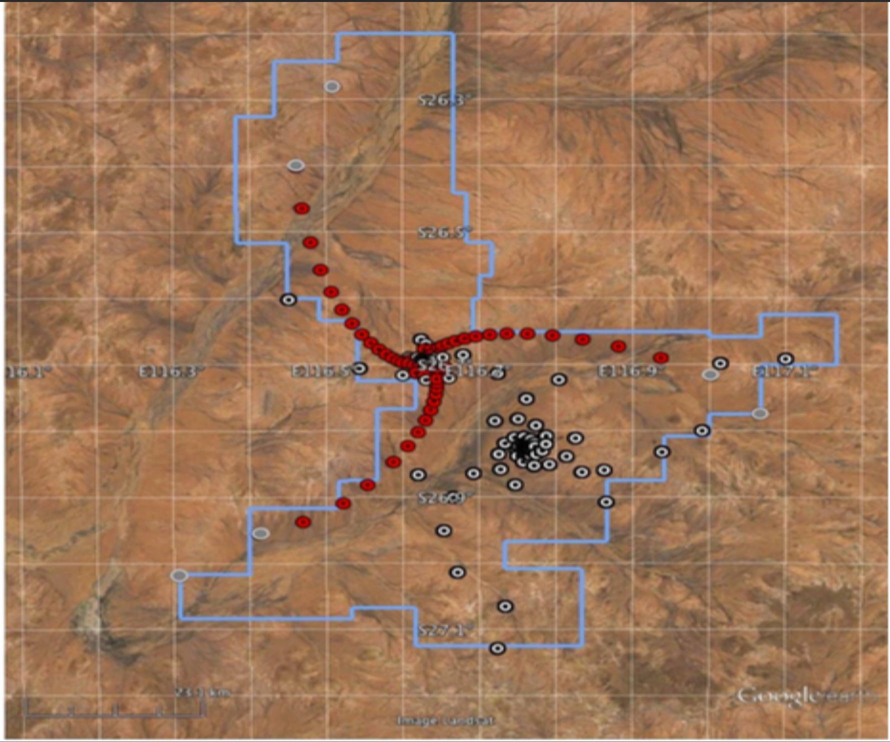
Mid Frequency Telescope

South Africa

250 dishes with single receiver
Karoo Desert, SA - 3 humans / km^2
Compute in Cape Town (400 km)

Antenna array layout





SKA1–MID, –LOW: Max Baseline = 156km, 65 km

Science

Science Headlines

Fundamental Forces & Particles

Gravity

- Radio Pulsar Tests of General Relativity
- Gravitational Waves
- Dark Energy / Dark Matter

Magnetism

- Cosmic Magnetism

Origins

Galaxy & Universe

- Cosmic dawn
- First Galaxies
- Galaxy Assembly & Evolution

Stars Planets & Life

- Protoplanetary disks
- Biomolecules
- SETI

skatelescope.org – two very large books (free!) with science research articles surrounding SKA

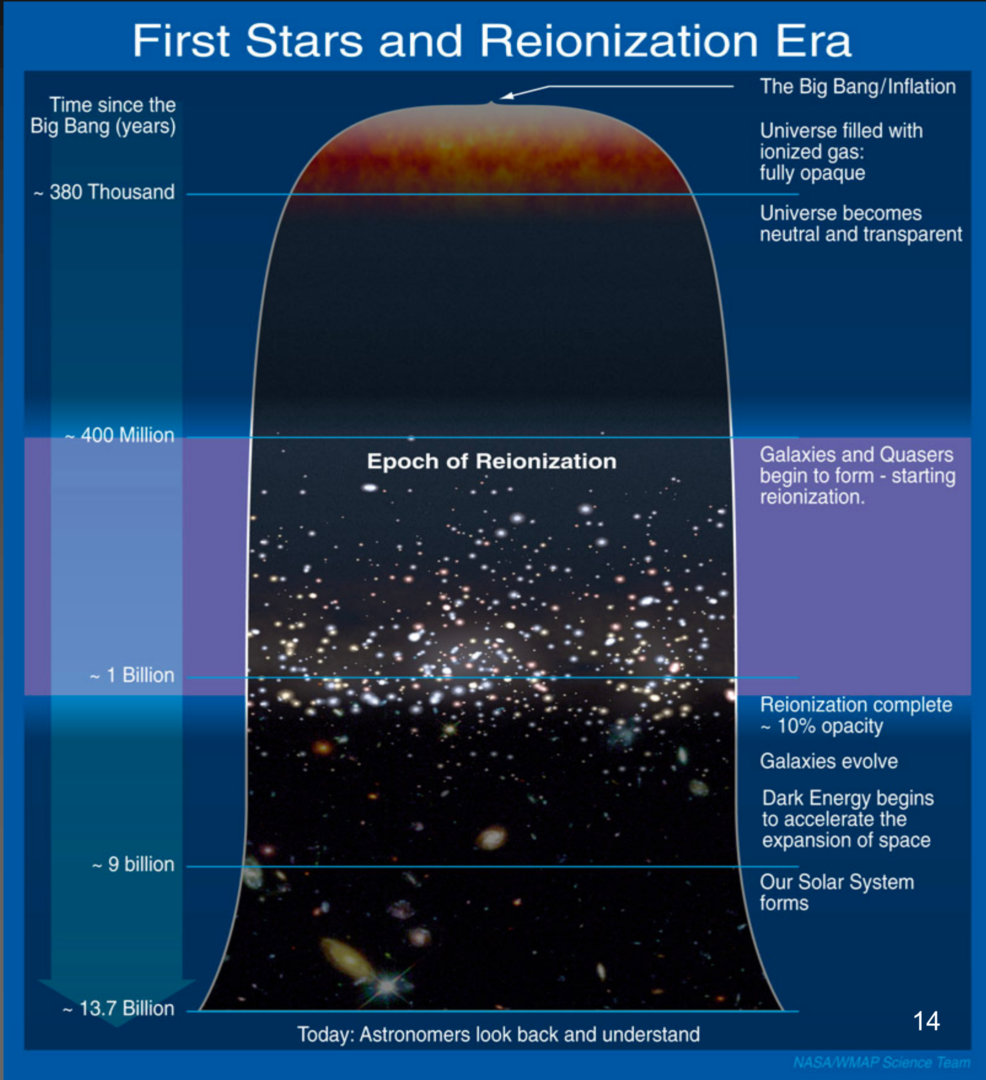
Epoch of Re-Ionisation

21 cm Hydrogen spectral line (HI)

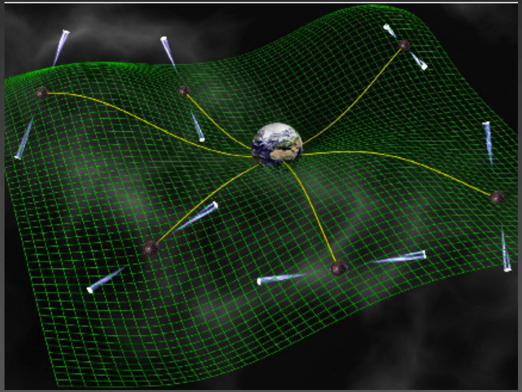
Difficult to detect

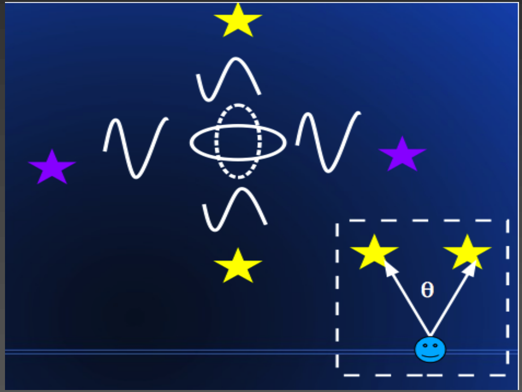
Tells us about the dark age:

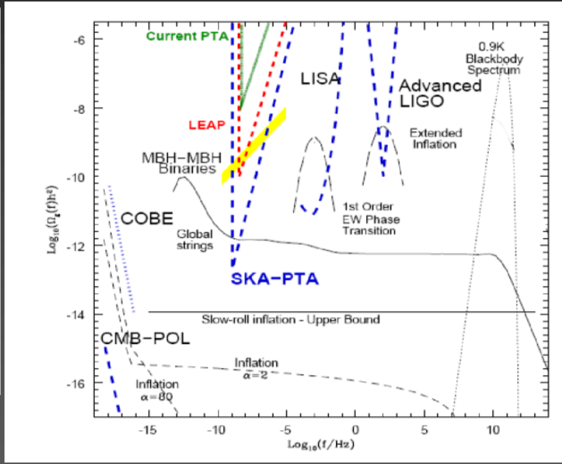
400K – 400M years (current age 13.5G year)



Pulsar Timing Array





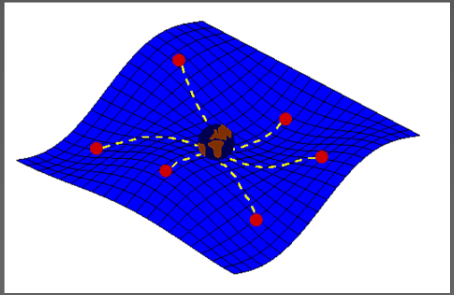


What can be found:

- gravitational waves
- Validate cosmic censorship
- Validate “no-hair” hypothesis

- **Nano-hertz frequency range**
- ms pulsars, fluctuations of 1 in 10^20

- SKA1 should see all pulsars (estimated ~30K) in our galaxy



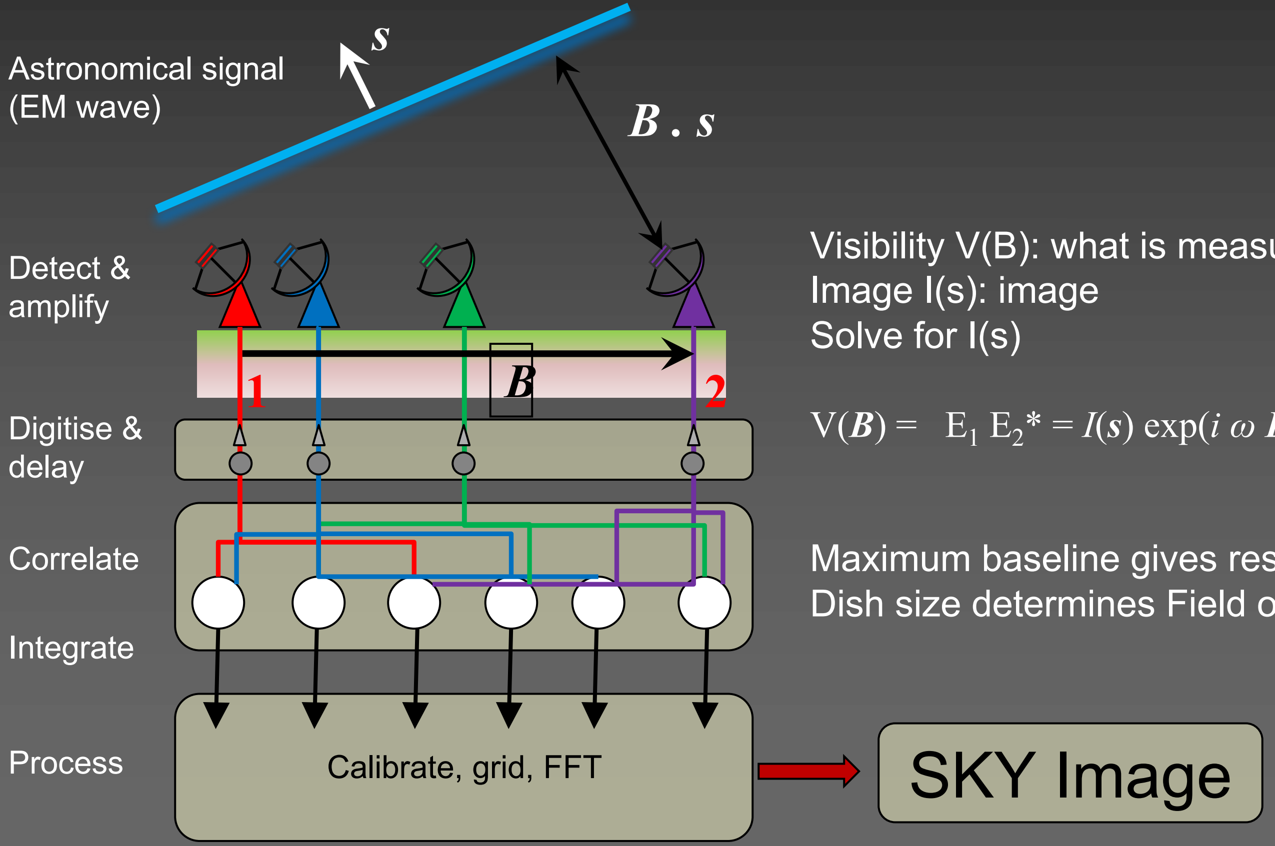
Physics & Astrophysics

Many key questions in theoretical physics relate to astrophysics

Rate of discoveries in the last 30 years is staggering

Imaging Problem

Standard interferometer

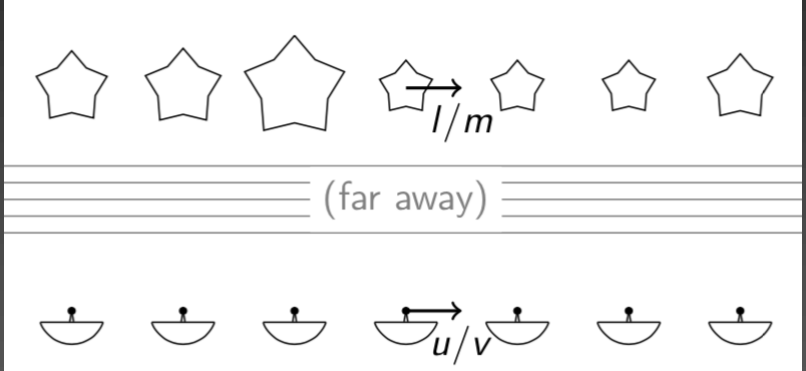


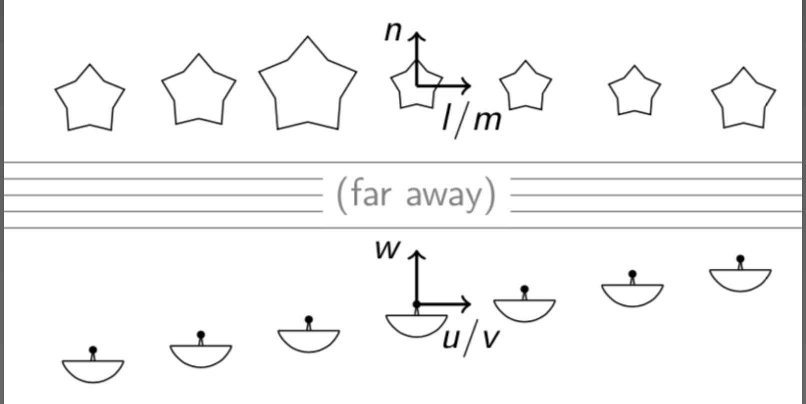
Visibility V(B): what is measured on baselines
Image I(s): image
Solve for I(s)

$V(\mathbf{B}) = E_1 E_2^* = I(\mathbf{s}) \exp(i\,\omega\,\mathbf{B.s}/c)$ – image equation

Maximum baseline gives resolution: $\theta_{max} \sim \lambda / B_{max}$
Dish size determines Field of View (FoV): $\theta_{dish} \sim \lambda / D$

Interferometry radio telescope





Simplified

Sky is flat
Earth is flat

Telescope to image is Fourier transform

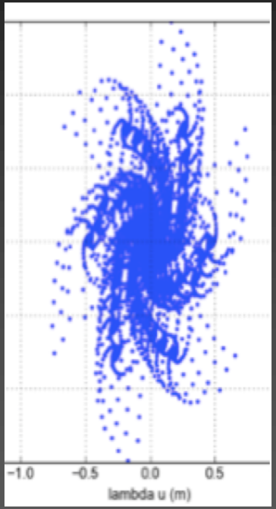
Actually

Sky is sphere, earth rotates, atmosphere distorts

Now it is a fairly difficult problem:

1. Non-linear phase
2. Direction, frequency, baseline dependent gain factor

Data in the computation



Two principal data types

input is visibility – irregular, sparse uvw - grid of baselines

Image grid - regular grid in sky image

Different kinds of locality

Splitting the stream by frequency

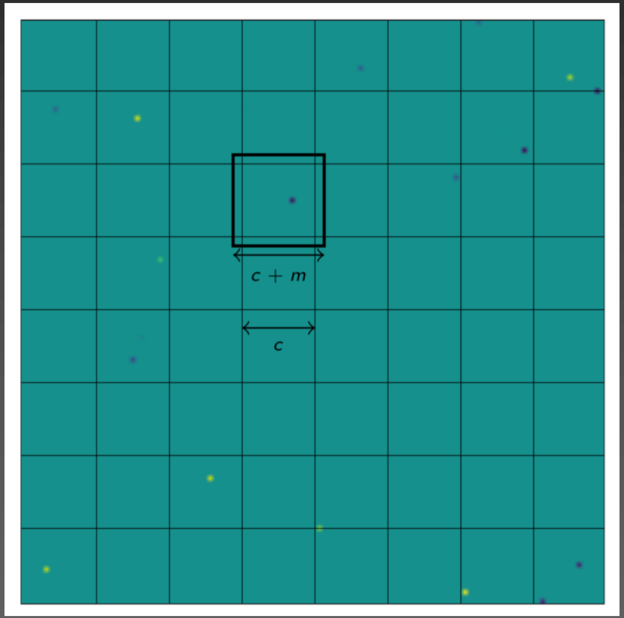
Tiling visibilities by region – but visibility “tile” data is highly irregular

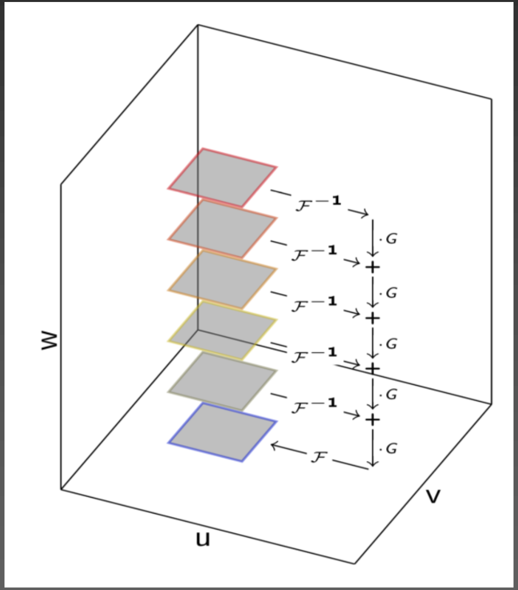
Analyze visibility structure – 0, sparse, dense: separate strategies

Remove 3rd dimension by understanding earth rotation

Data flow model with overlapping movement and computation

Reducing to 2D





Try to go back from 2D to 3D problem by relating (~100) different w values.
Domain specific optimization.

Grid size is 64K x 64K for 64K frequencies – problem is large
Full FFT is O(k log k), sparse FFT: O(#nonzero log #nonzero). This approach is close to this.

Computing in radio astronomy - 101

@Antennas: wave guides, clocks, beam-forming, digitizers

@Correlator (CSP central signal processing): == DSP for antenna data

- Delivers data *for every pair of antenna's (a "baseline")*
- Dramatically new scale for radio astronomy ~100K baselines
- Correlator averages and reduces data, delivers sample every 0.3 sec
- Data is delivered in frequency bands: ~64K bands
- 3 complex numbers delivered / band / 0.3 sec / baseline
- Do math: ~ 1 TB/sec input of so called *visibility data*

@Science Data Processor (SDP) – process correlator data

- Create images (6 hrs) & find transients (5 secs) – "science products"
- Adjust for atmospheric and instrument effects *calibration*

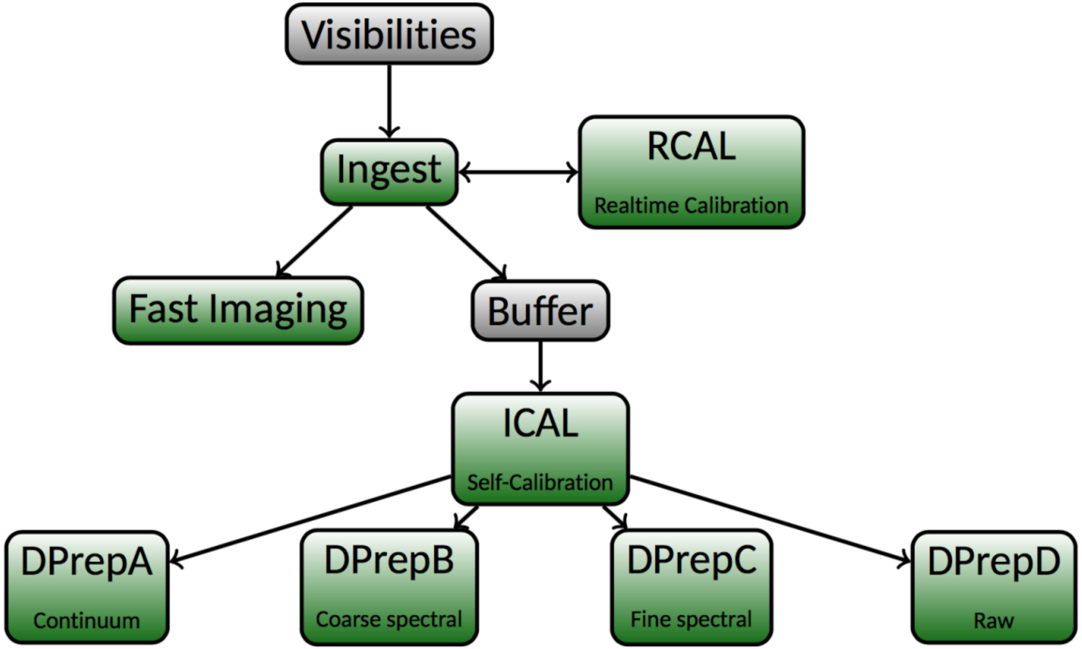
Outline of algorithm

About 5 different analysis on the data are envisaged:
e.g. spectral vs continuum imaging etc.

Imaging pipelines:

- Iterate until convergence – approximately 10 times
- Compares with an already known model of the sky
- Incorporates and recalculates calibration data

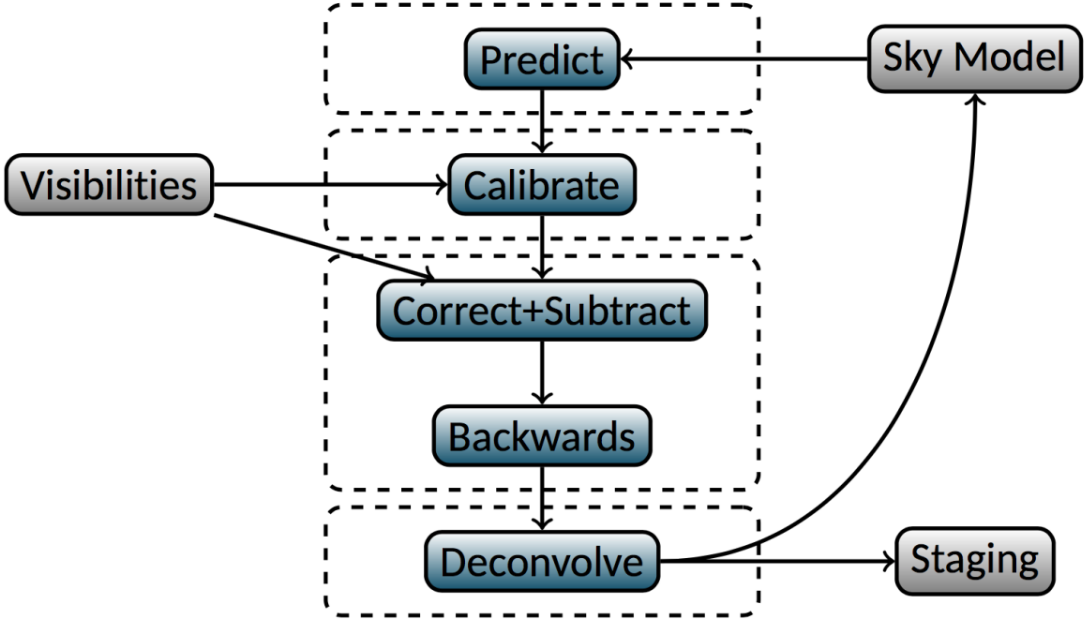
Pipelines Structure



Follows architecture, allows running multiple data preparations.

Pipeline Internals

Rough structure and distribution pattern of most pipelines:



SDP specific Pipelines

Algorithmic similarities with other image processing

Each step is

- Convolution with some kind of a “filter” – e.g. “gridding”
- Fourier transform
- All-to-all for calibration

Why new & different software?

- Data is very distinct from other image processing
- Problem is very large – much bigger than RAM
- Reconstruction dependencies: sky model & calibration

Engineering Problem

Requirements & Tradeoffs

Turn telescope data into science products soft real time

1. Transient phenomena: time scale of ~10 seconds
2. Images: 1 image ~6 hours

Agility for software development

Telescope lifetime ~50 years

SDP computing hardware refresh ~5 years

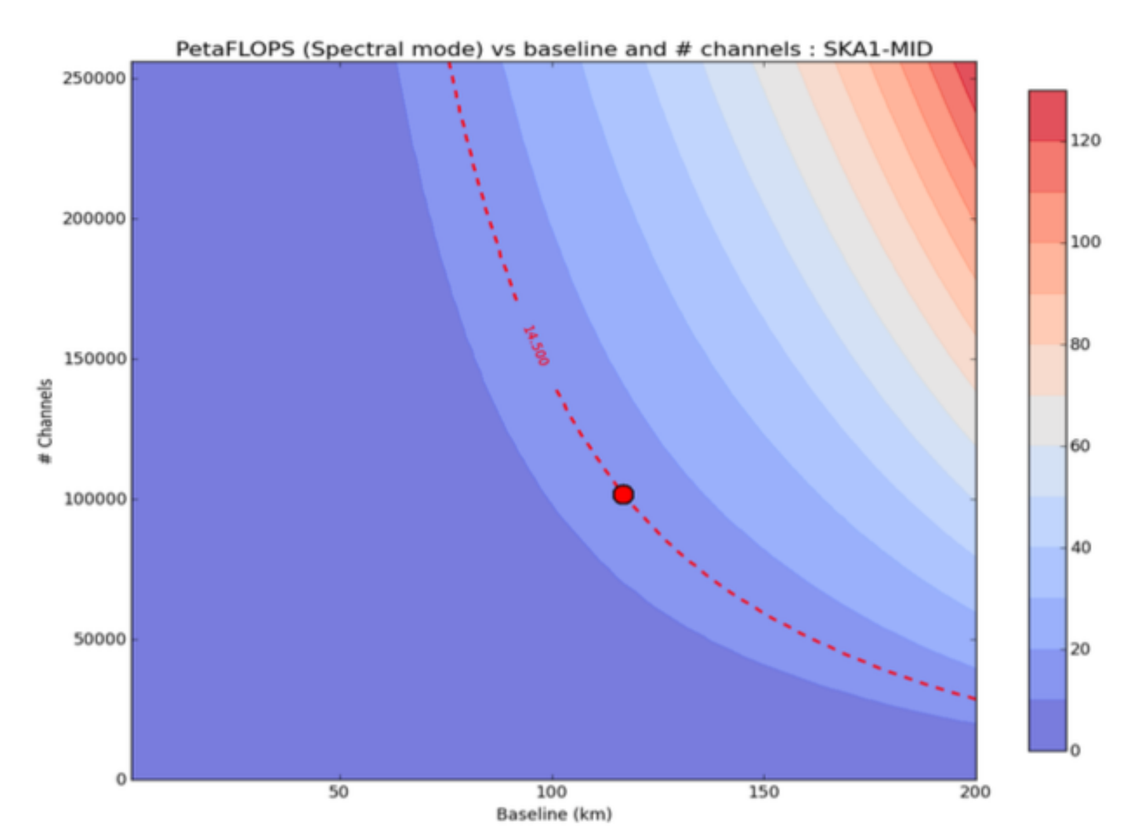
Use of large clusters is new in radio astronomy

New telescopes always need new algorithms

Initial 2025 computing system goal: make SKA #1

So – how difficult is this?

Flops vs. #channels & baseline



SDP "performance engineering" approach

Conservative - this is not computing research

Known-good algorithms, hardware

Perhaps deep math question remains: is problem really O(#antennas^2)?

Parametric model of the computation

Detailed FLOPs, memory use, data movement, energy

Key outcome: 100 PF/sec & move 200 PB/sec from HBM to CPU

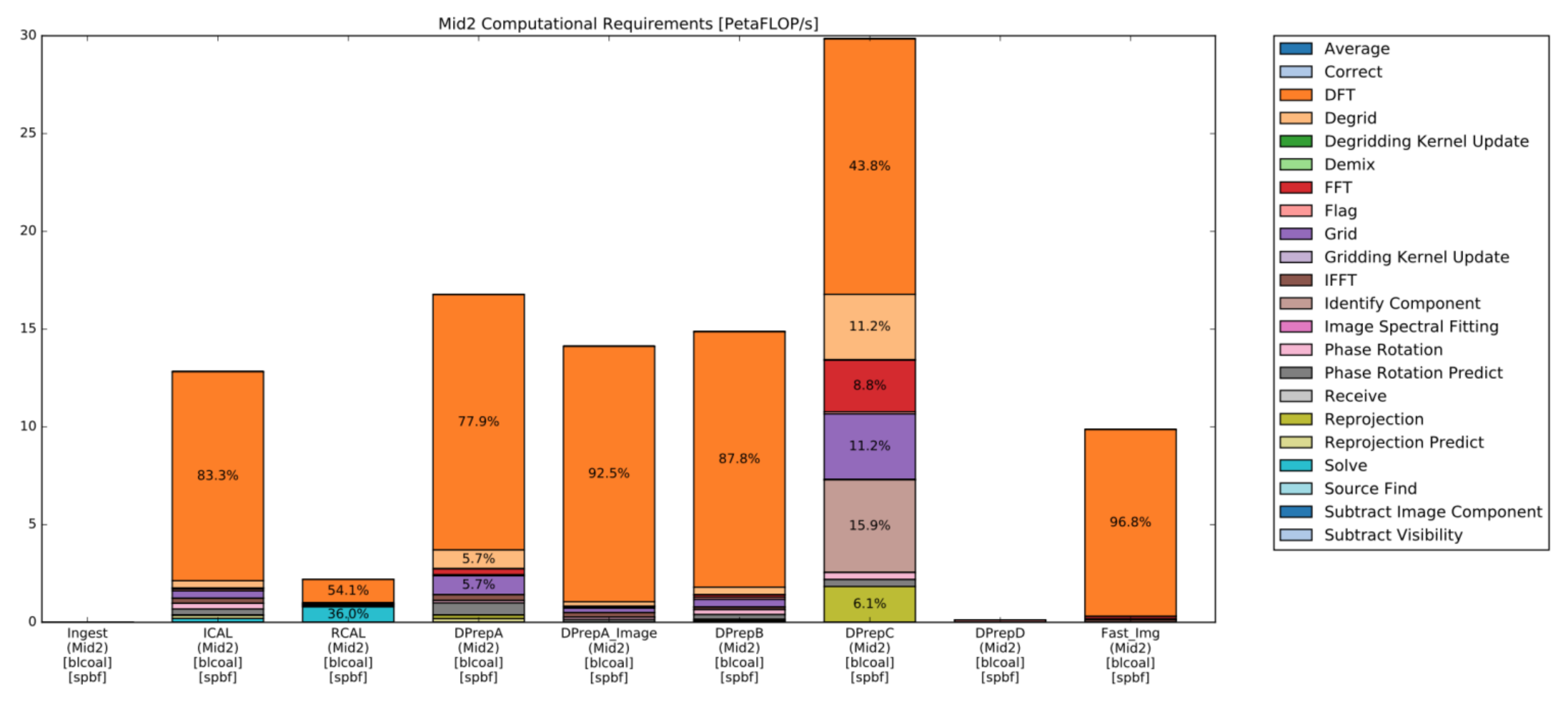
@50 PJ / byte this is ~10MW power

Software

Reference Libraries with Algorithms

Address scalability issues

Relative kernel cost



Software Framework for SKA SDP

Creating software is a very high risk part of the project

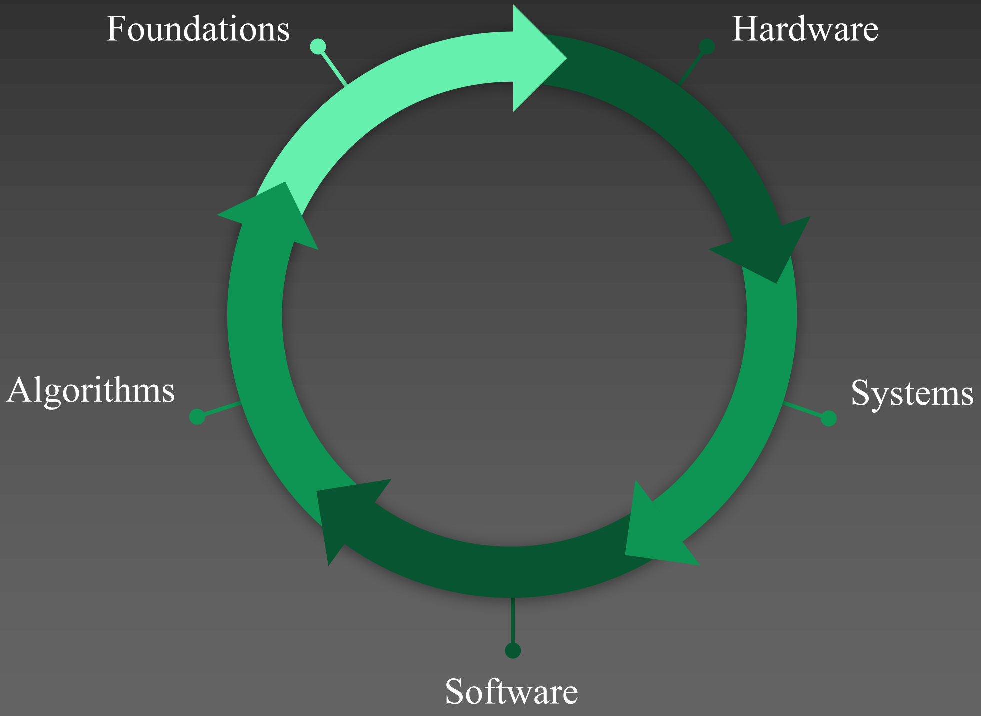
Ideal perspective:

- Execution framework from 3rd party
- Domain specific application language for pipelines
- Automatic optimization – performance & energy
  - …. this is proving less easy than we had hoped

Many approaches – excellent compilers and ….

- Adapting existing packages – MPI C++ applications
- Use a big-data framework like Spark, TensorFlow
- Use HPC frameworks like Swift/T, Legion

Co-design



Lesson:

Interactions between these domains is a very big challenge

Work Breakdown

Foundations: be conservative – no totally new approaches

Algorithms: innovation - adapt to scale

Software: prototyping, seeking mainstream solution

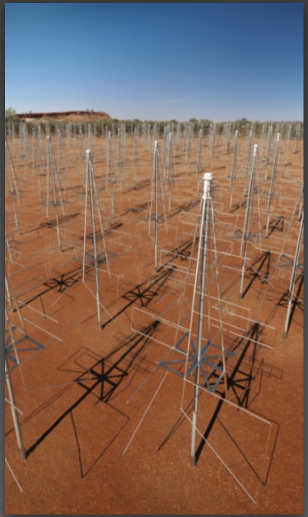
Hardware: much has been learned from working with the chip vendors

Systems: analysis by HPC experts, costing, vendor ideas etc.

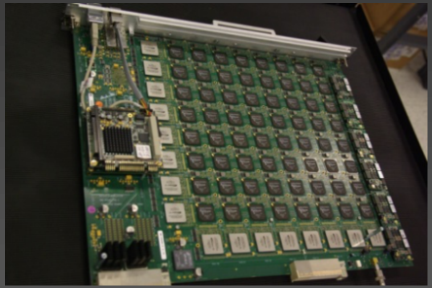
Samples of Data Processing Considerations

SKA – data schematic

Antennas



Central Signal Processing (CSP)



Transfer antennas to CSP
2024: 20,000 PBytes/day
2030: 200,000 PBytes/day

Over 10’s to 1000’s kms

Imaging (SDP) – HPC problem

2024: 100 PBytes/day
2030: 10,000 PBytes/day
Over 100’s kms

High Performance Computing Facility (HPC)

HPC Processing
2024: 300 PFlop
2030: 30 EFlop

In: 20 EB in -> out: 100 TB

SDP top-level compute challenge

Telescope Manager

CSP

Science Data Processor

SDP Local Monitor & Control

Data Processor

High Performance
~100 PF/sec
~200 PB/sec
Data Intensive
~100 PB/job
Partially real-time
~10s
Read-intensive
Constrained

Long Term Archive

EB volume

100PB annually

Infrequent access

Delivery System

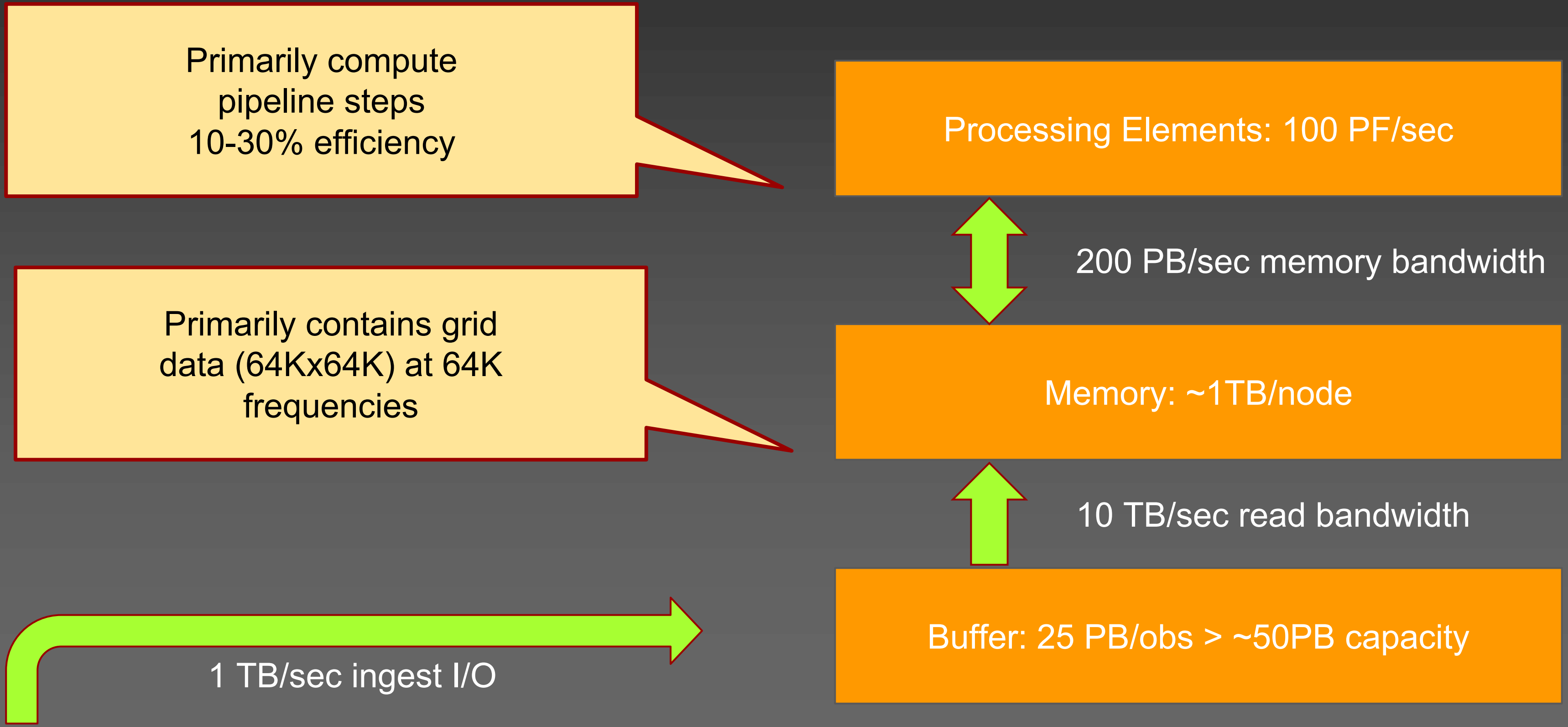
Distribution
~100PB/y

From Cape Town & Perth to rest of World

Visualization of 100k x 100k x 100k voxel cubes

Regional Centres & Astronomers

Data Movement



Supercomputer parameters

2025	LFAA (AU)	Mid (SA)
FLOPS	100 PF	360 PF
Memory bandwidth	200 Pb/sec	200 Pb/sec
Buffer Ingest	7.3 TB/s	3.3 TB/s
Budget	45 M€	3.3 TB/s
Power	3.5 MW	2 MW
Buffer storage	240 PB	30 PB
Storage / node	85 TB	5 TB
Archive storage	0.5 EB	1.1 EB

Memory Bandwidth
- Cost
- Energy
- 10x 1st EF BW

Memory … SKA's biggest challenge

High Bandwidth Memory (HBM) is becoming dominant for HPC

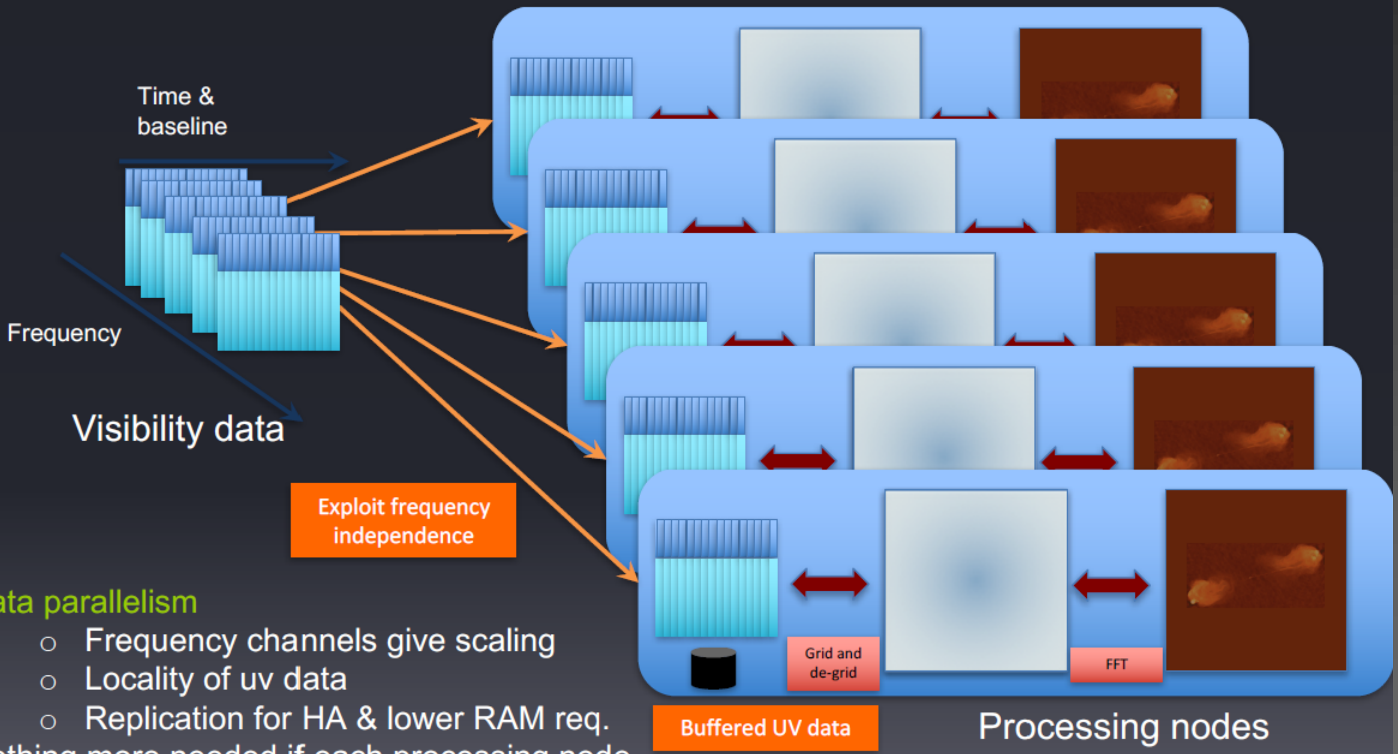
In 2013 the problem looked perhaps out of reach

HBM is 3D, on package, memory

10x bandwidth of RAM, perhaps similar cost

Delay in SKA the deliverables has been very helpful

Data Flow on System Architecture



Data parallelism

- Frequency channels give scaling
- Locality of uv data
- Replication for HA & lower RAM req.

Nothing more needed if each processing node can process a frequency channel completely

Visibilities & Baselines distribution

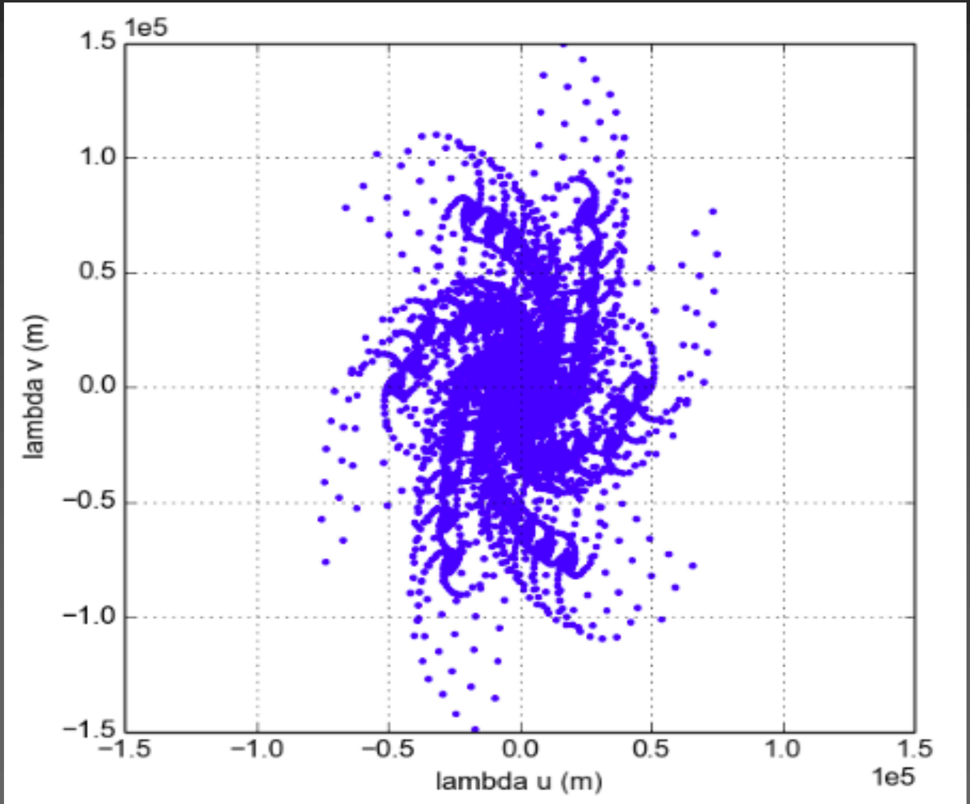
Each pair of telescopes has a baseline

Baselines rotate as time progresses

Each baseline has associated visibility data ("sample")

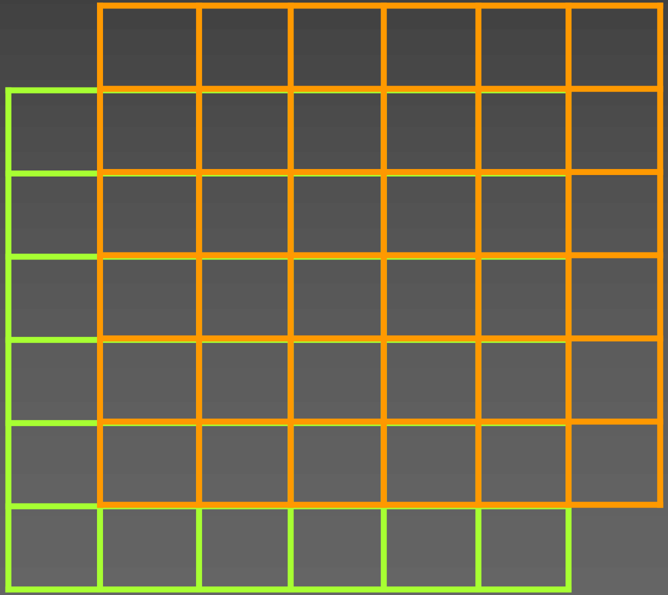
Baselines are sparse & not regular, but totally predictable

The physical data structure strongly enables and constrains concurrency & parallelism



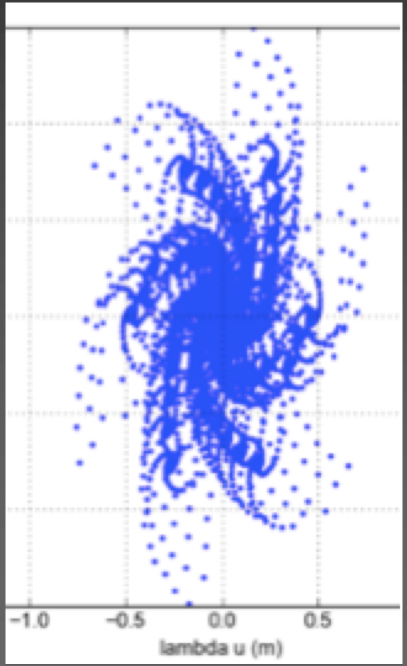
Simulated data from 250 SKA1-MID dishes

Visibility gridding & cache re-use



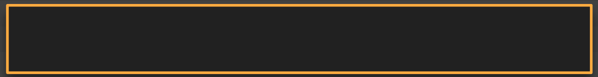
Time rotation of
UV grid.

Only fetch edges
Re-use core



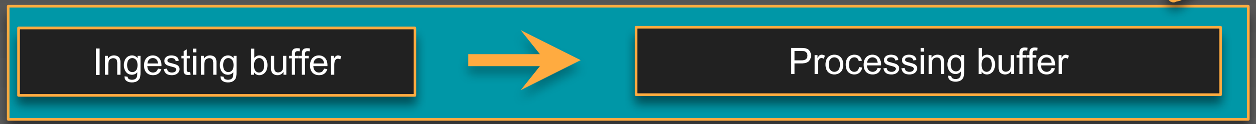
Long vs short buffer question

Processing requires up to 6 hours of ingest – buffer that.



21,600 TB – “unit of data ingest” to compute on

Overlapping ingest and compute: double buffer ?



Double Buffer: ~50PB, write 1TB/sec, read 10TB/sec

But processing time is uneven –

double buffer: minimizes storage cost,

at expense of equally quick execution of worst compute cost

Stream fusion

Some kernels exchange too much data

Solution: deviate from pipeline actors
do more operations and less data movement.

Few compilers / frameworks automatically

Doing it manually is awkward for portability

Conclusions

Conclusions

Computing is extremely central, well beyond the instrument

e.g. applying AI / ML to analyzing the science data

Astrophysics has everyone’s attention – this project must succeed

SKA will succeed based on astrophysics

but its computing lies on the frontier of big data handling

Software may is the highest risk and hardest problem of all

Thank you.

questions?

skatelescope.org

peter.braam@peterbraam.com