SKA Computing Challenges

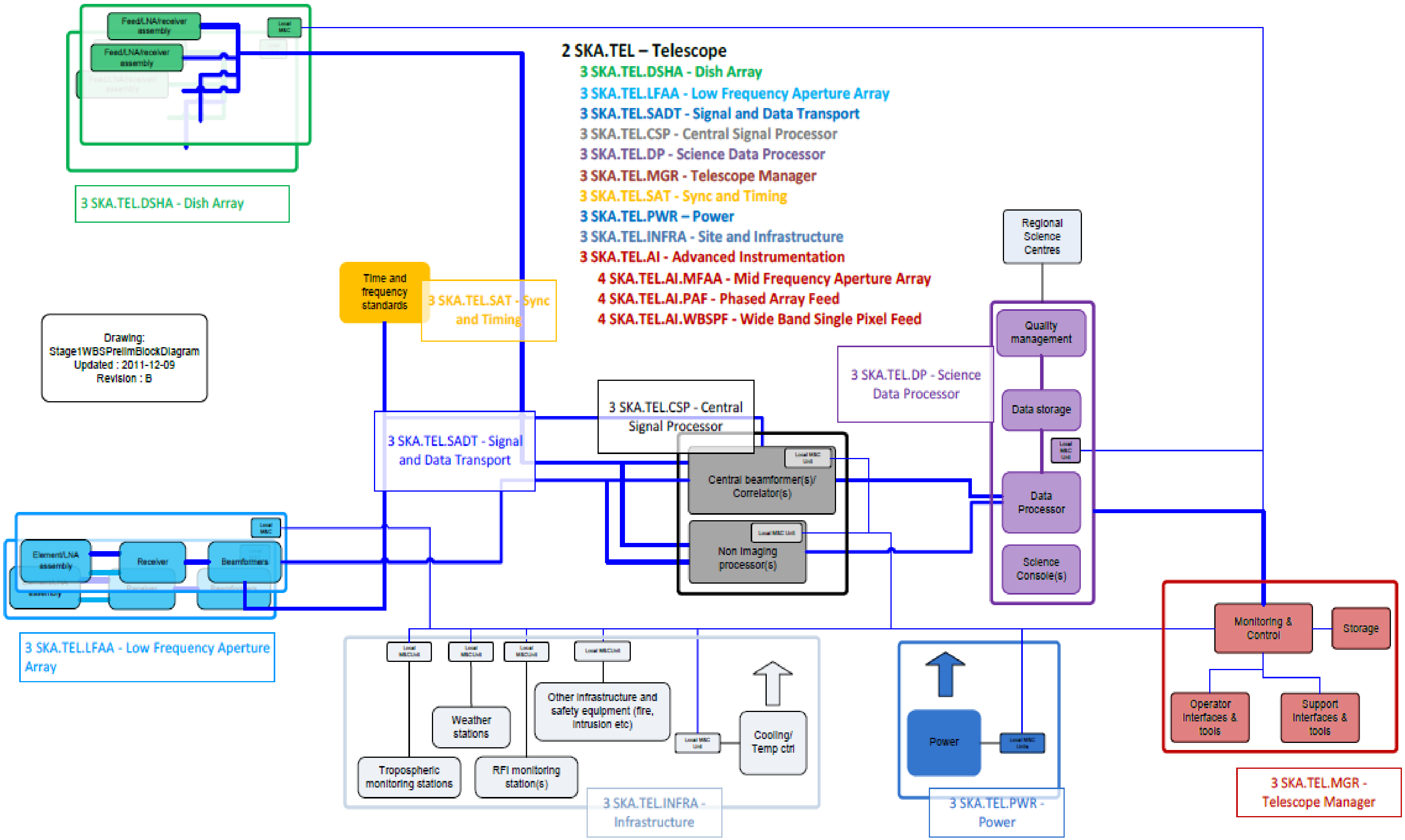
Andrew Ensor
Computing for SKA Colloquium
12 February 2015

Images credit SKA organisation/TDP/DRAO/Swinburne Astronomy

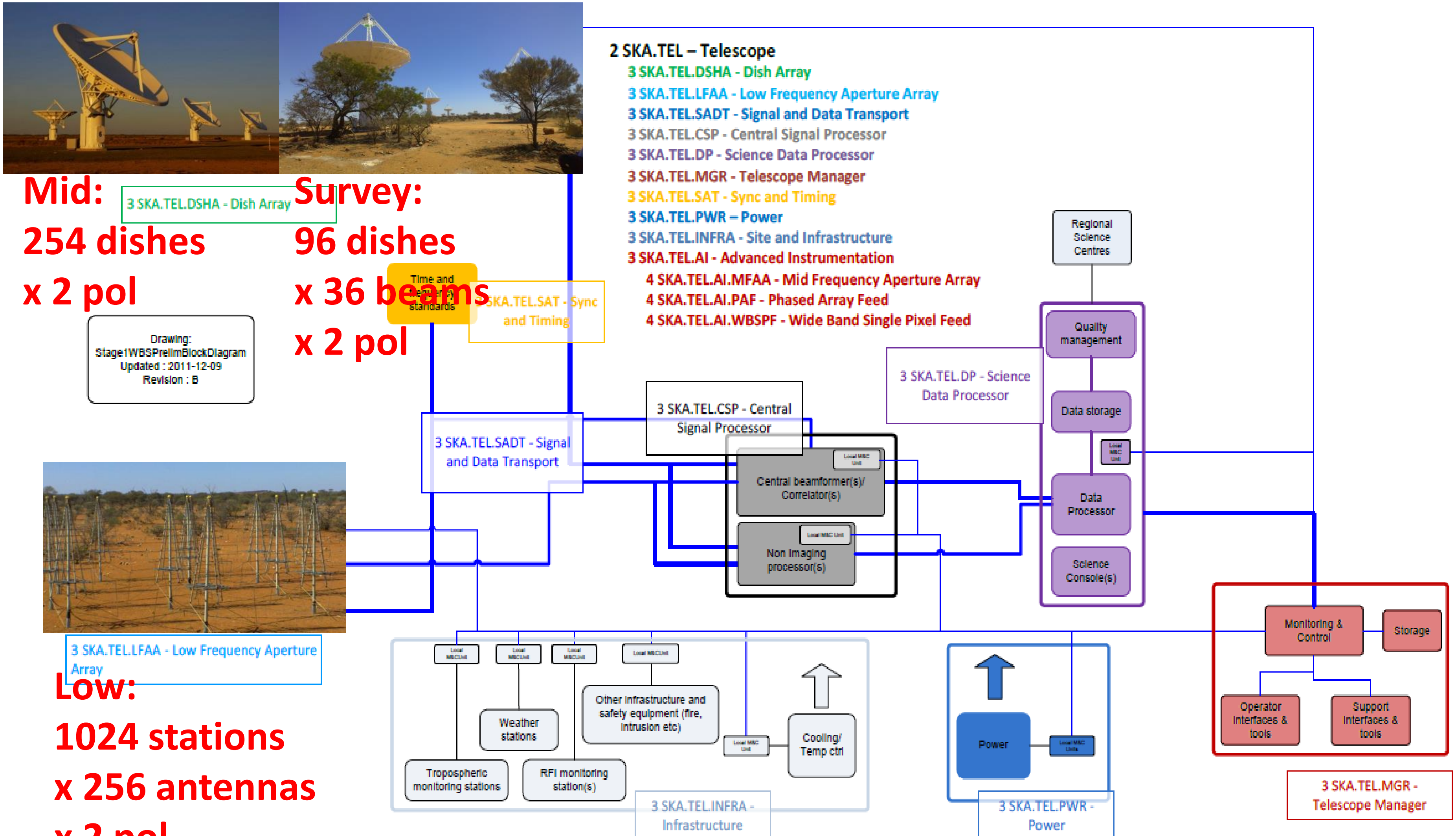
The Square Kilometre Array

- Will be World's largest and most powerful radio telescope
- Actually four types of telescope arrays:
  - Low - sparse aperture array with 1024 stations, each with 256 dipole antennas, later also dense aperture array
  - Mid – 254 single pixel dishes
  - Survey – 96 phased-array feed receivers each with 36 beams
- Cost €650M for SKA1
- Early estimates €2- €6billion for SKA2
- Preconstruction design for SKA1 commenced Nov 2013
- Two precursor arrays already being built:
  - MeerKAT in South Africa (Mid)
  - ASKAP in Australia (Survey)
- Actually an enormous engineering design and construction project

SKA1 Work Packages



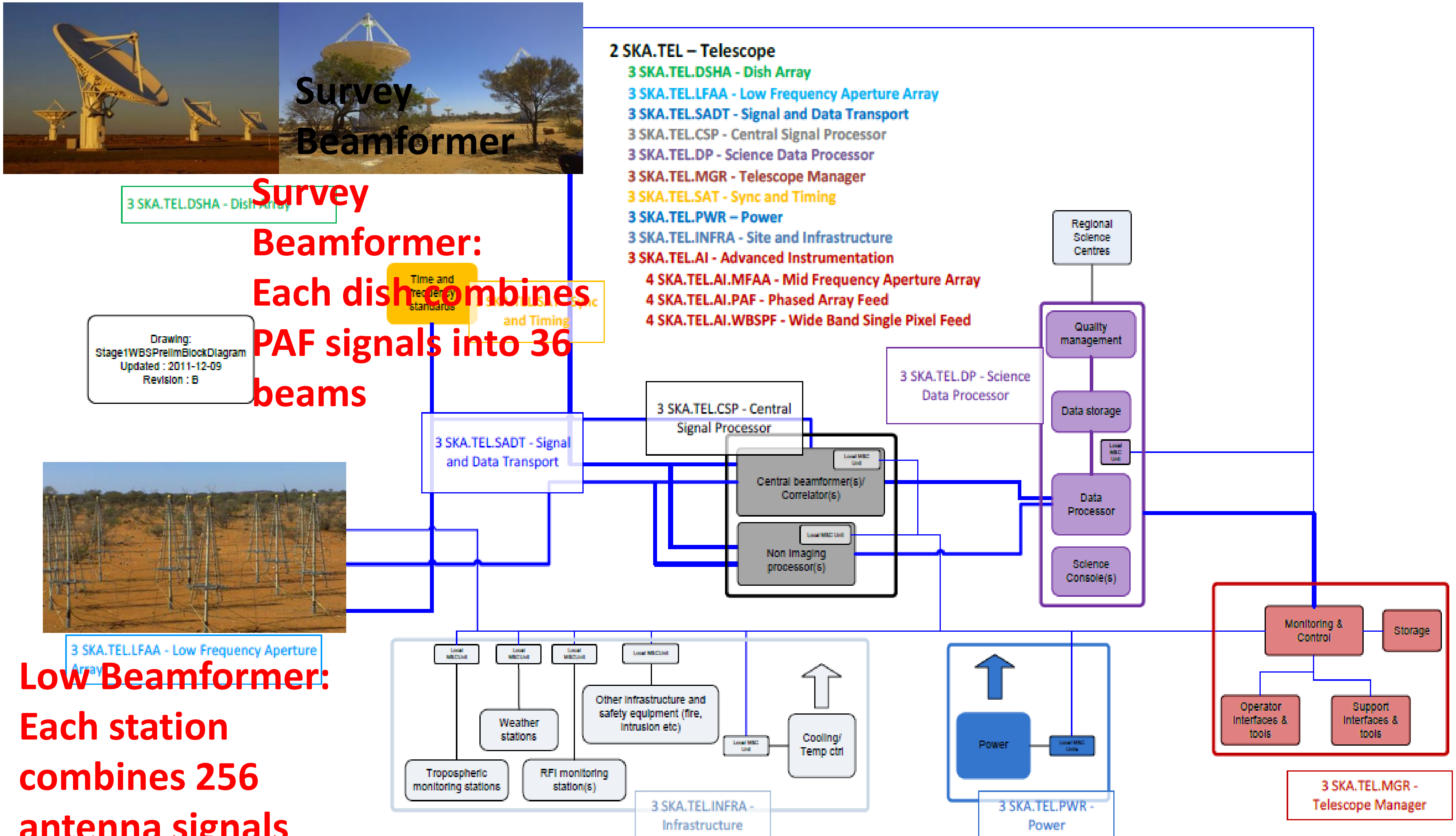
SKA1 Key Computing-Data Generation



SKA "Big Data"

The animation summarises the amount of data that would be generated by the current baseline design for SKA1……assuming the science use cases that would maximise data rates for this design. Expect that, by construction time, the figures are likely to be reduced considerably……although they will still be immense.

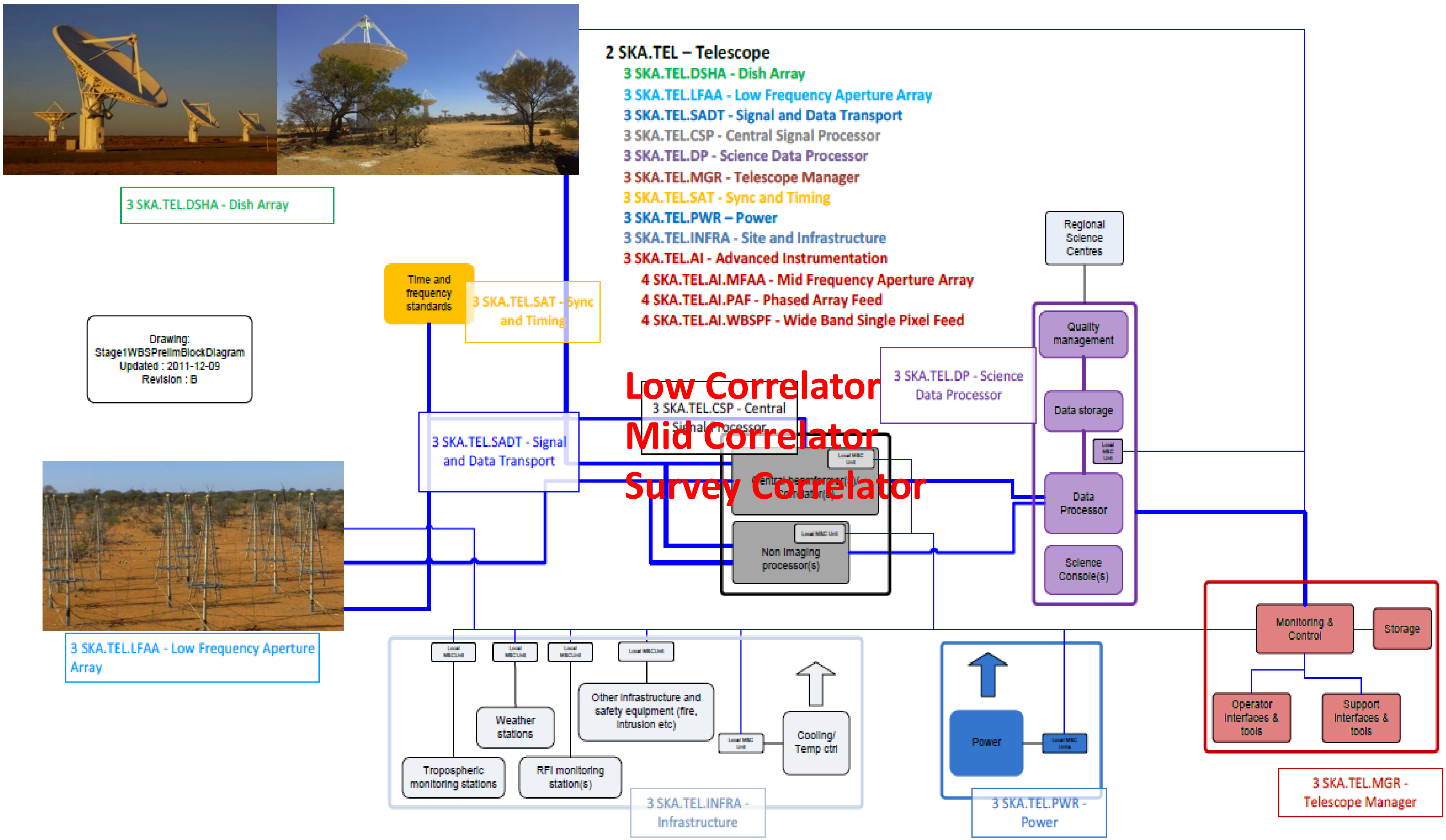
SKA1 Key Computing-Beamforming



Survey Beamformer: Each dish combines PAF signals into 36 beams

Low Beamformer: Each station combines 256 antenna signals

SKA1 Key Computing-Correlators



Correlator Overview

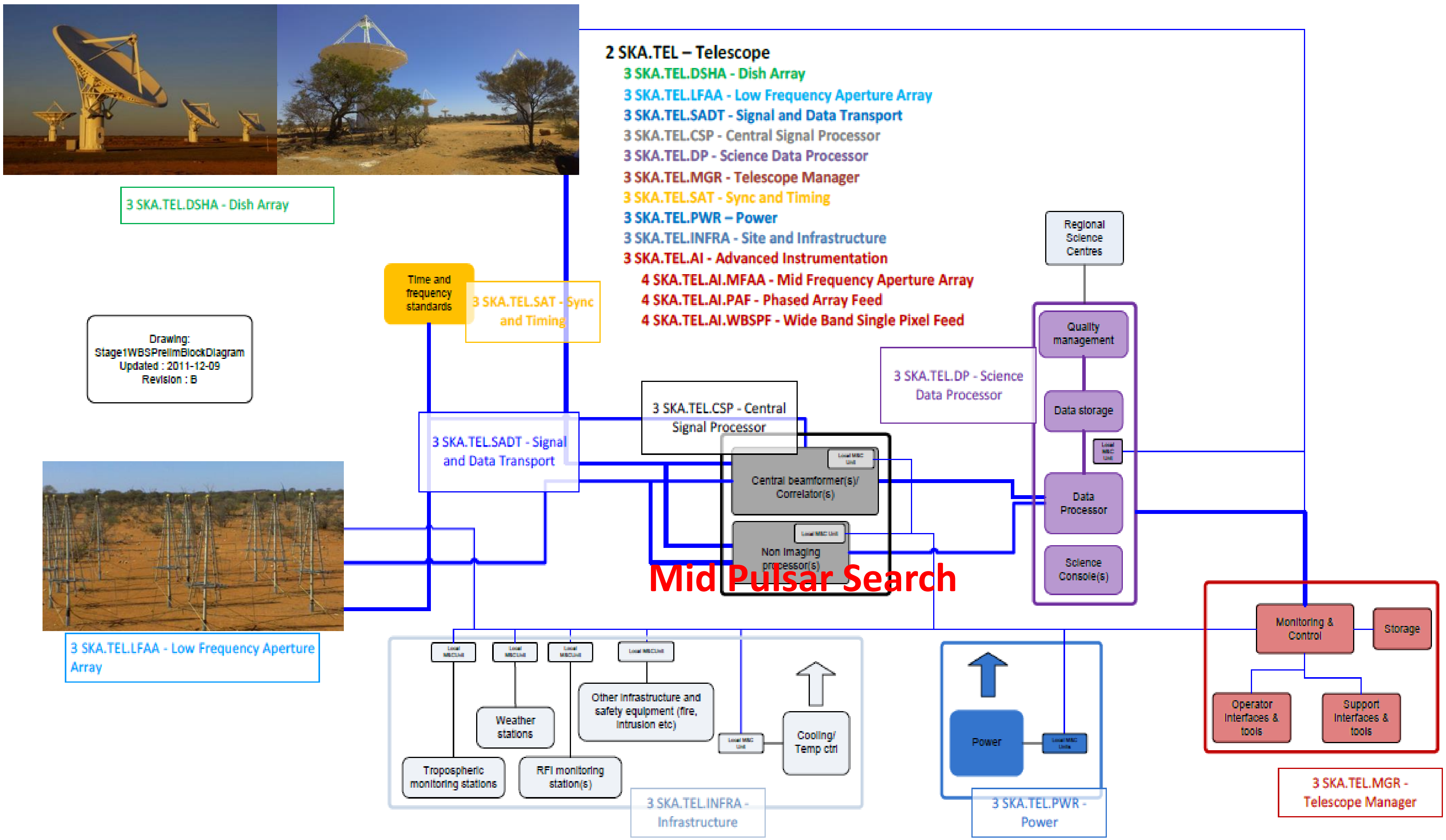
- Correlator combines signals together from each receiver so array acts like a single telescope
- FX correlator:
  - Each signal Fourier transformed
  - Each pair of transformed signals cross-correlated
  - Correlated products accumulated over brief time
  - Outputs are called visibilities
- Pairing requires interconnect between channelizer and correlator
  - To avoid enormous data duplication requires "centralised" correlator

Telescope	Data Input	Channelizer Compute	Correlator Compute	Max Data Output
Low	2338 GB/s	46.7 TMAC/s	2090 TMAC/s	7320 GB/s
Mid	580 GB/s	11.6 TMAC/s	2003 TMAC/s	3250 GB/s
Survey	7900 GB/s	158.0 TMAC/s	1313 TMAC/s	4670 GB/s

Correlator Challenges

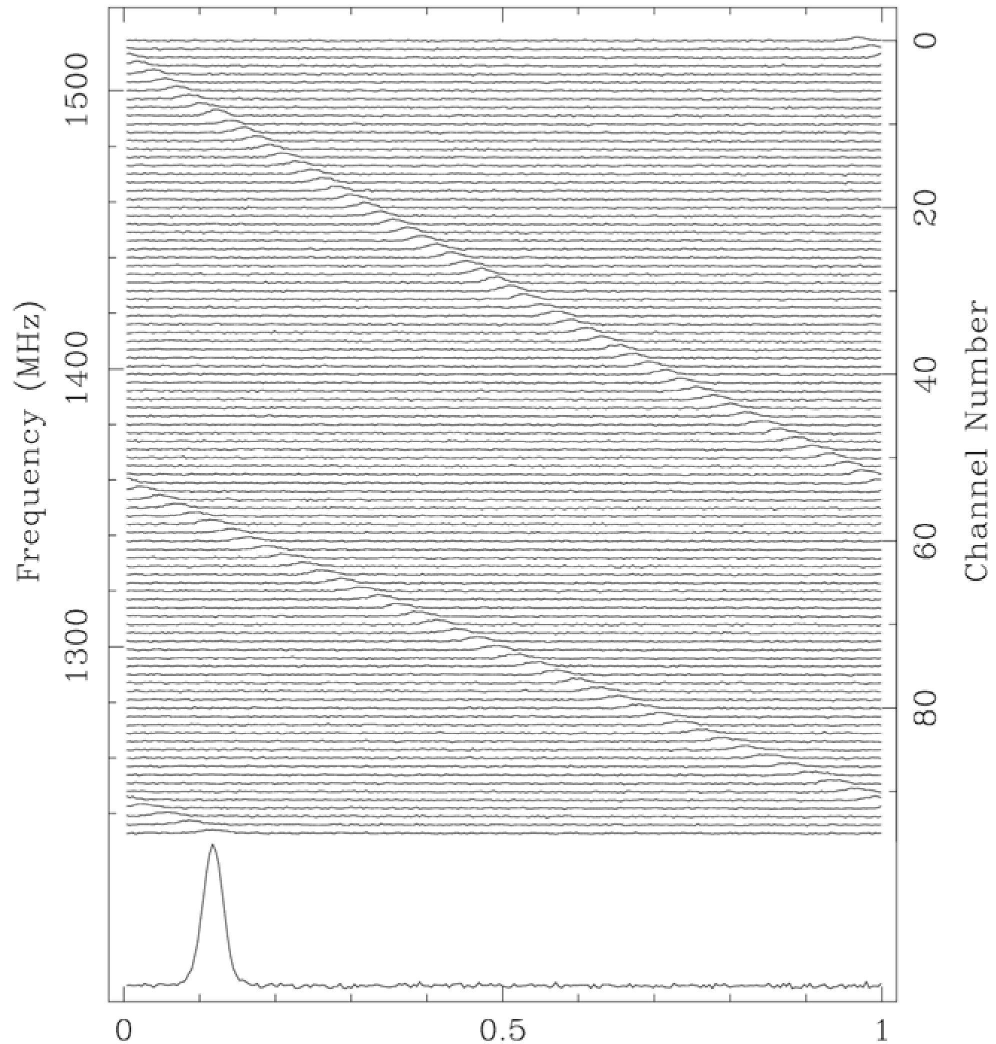
- Exascale Big Data project:
  - Large Hadron Collider expected to produce about 15 PetaByte/year
  - Correlators will produce about 40 PetaByte/hour
- Unprecedented I/O challenge:
  - Each telescope array requires up to 8 TeraByte/s channelization and correlation and outputs up to 4TeraByte/s (Cray Titan I/O about 1.4TeraByte/s total)
  - Data throughput a major concern
- Green computing requirements:
  - Two telescope arrays are deep in Australian outback, no power grid so energy consumption a major concern

SKA1 Key Computing-Pulsar Search



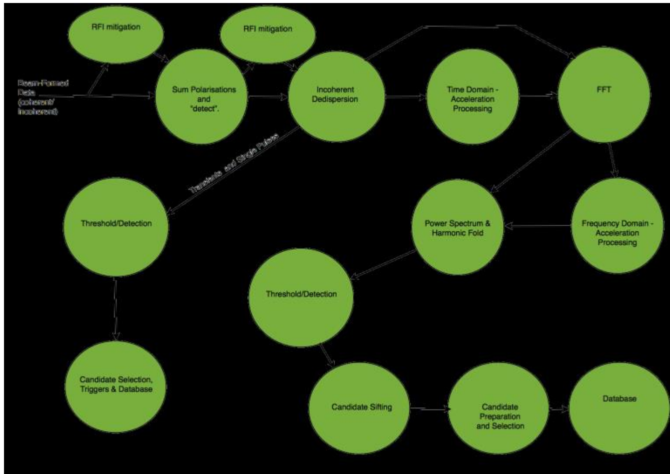
Pulsar Search Overview

- Search for pulsar signals:
  - Unknown periodicity
  - Signal deeply buried in noise
- De-dispersion:
  - Radio signals interact with free electrons, causes frequency dependent delay
  - Time signal searched over different dispersion measures to find candidate pulsars
- Acceleration search:
  - Particularly want to find pulsars in relativistic orbits around neutron stars or black holes
  - Period changes so guess acceleration

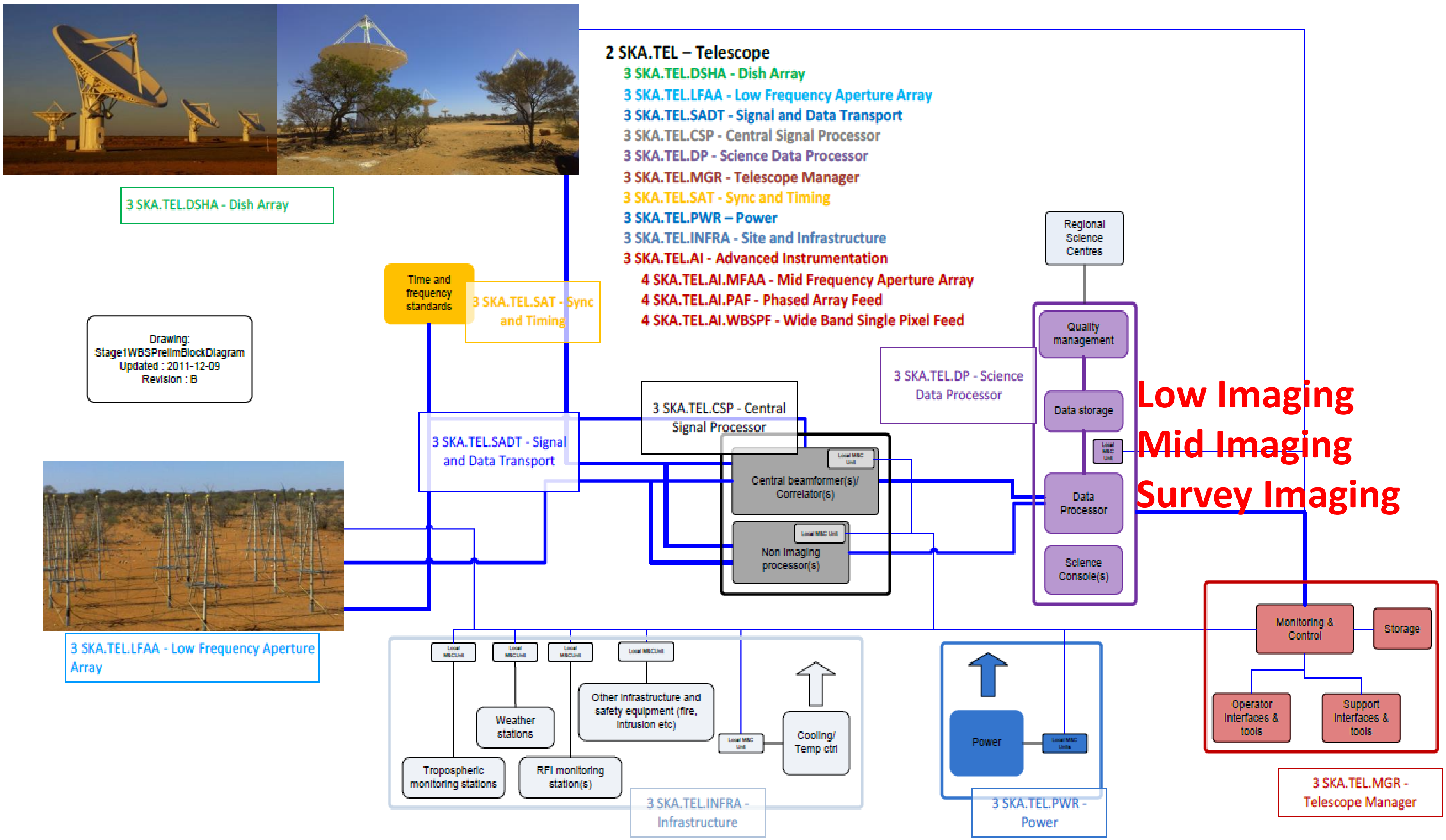


Pulsar Search Challenges

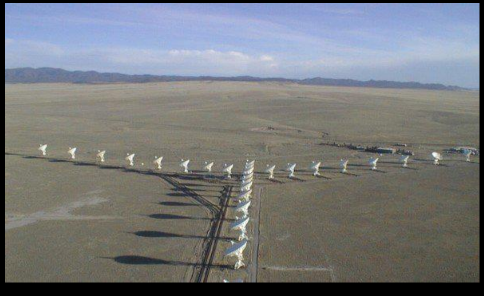
- Search for weak signals deeply buried in noise (ultimate needle-in-the-haystack)
- Simultaneously search 2222 beams
- Each beam has 4000-15000 channels, sampled every 50µs (700Gsample/s total) for 7min
- Three-dimensional search space amongst noise:
  - Unknown pulse period
  - Unknown dispersion
  - Unknown source acceleration



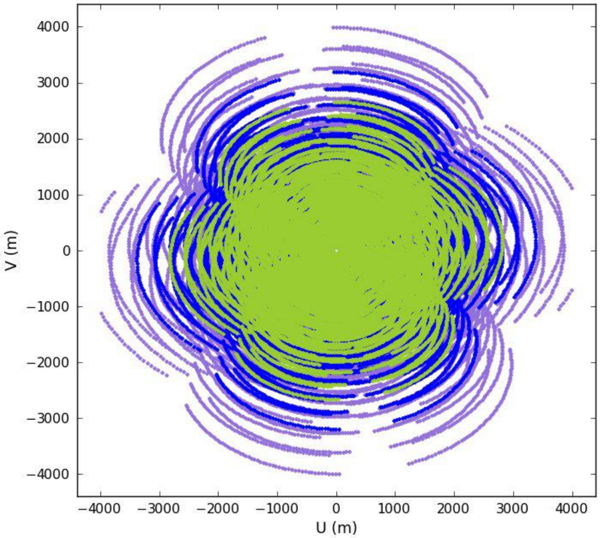
SKA1 Key Computing-Imaging



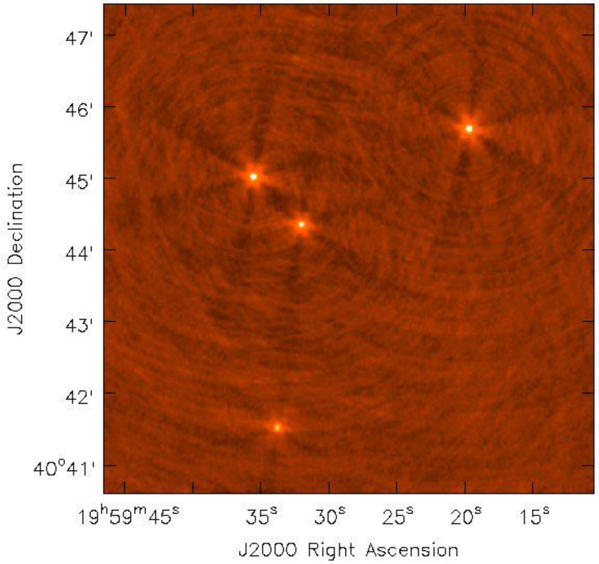
Imaging Overview



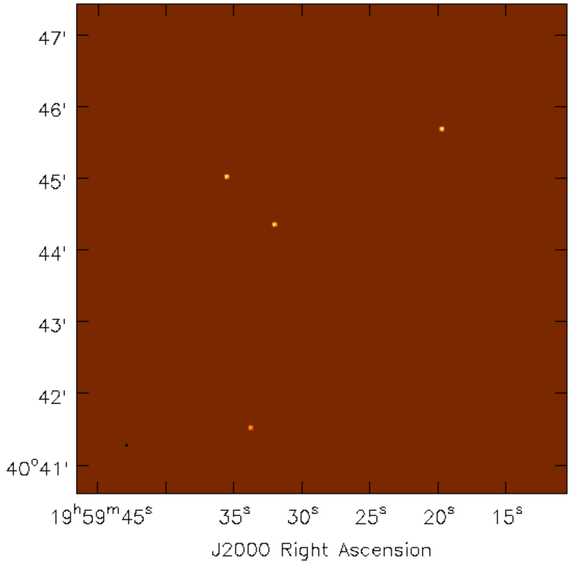
Visibilities from VLA correlator with 27 antennas over four hour observation



Dirty image after gridding visibilities and inverse FFT



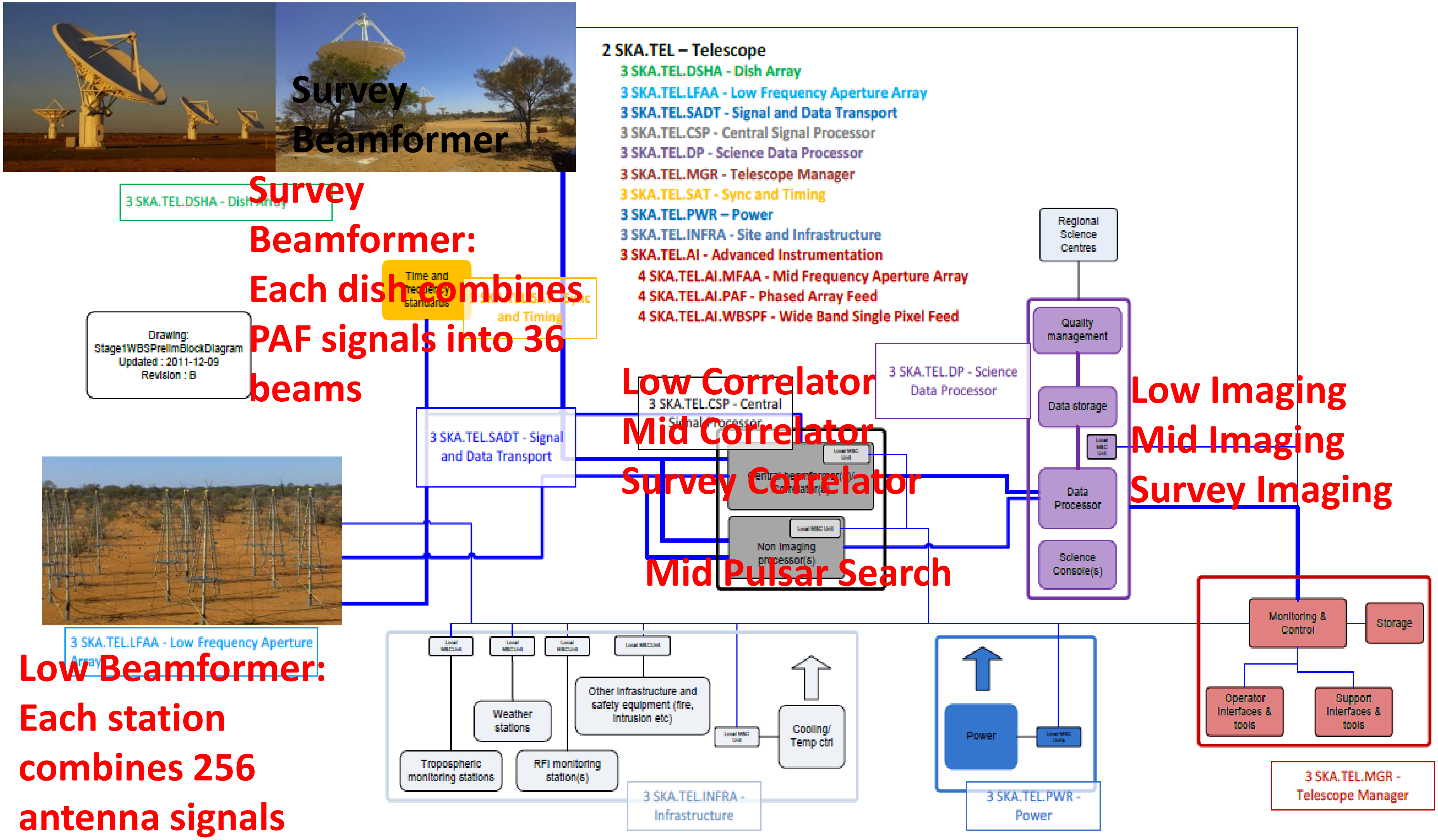
Sought sky image after deconvolution



Imaging Challenges

- Gridding and (inverse) FFT are enormous compute challenges
- Using up to 60kx60k grid and including each of up to 2^{18} channels in the imaging pipeline requires supercomputing performance:
  - SDP estimated to require up to peak 100 PetaFLOP for Low, 350 PetaFLOP for Mid, 125 PetaFLOP for Survey (Cray Titan is 17.59 PetaFLOP), within reach by 2018 but everything must be done in real time 24-7
  - Power requirements determine ultimate costs, SDP computing need to fit within 5-10 MW
- SKA-2 will increase computing demands by approximately 100 fold
- Independent "compute island" design

SKA1 Key Computing



Questions?

Thank you

NZ SKA Work Packages

Work Packages	Lead	NZ Commitment
Correlators: - Survey - Mid & PowerMX - Low	 NZ (Ensor) NRC (Carlson) Oxford (Zarb-Adami)	 5.14 FTE/year 0.81FTE/year 0.45FTE/year
Imaging Pipeline	Southampton (Scaife)	1.1FTE/year
Pulsar Search	Manchester (Stappers)	0.70FTE/year
CSP & SDP Modeling	NZ (Wilson)	1.47FTE/year
CSP Prototyping	NZ (Ensor)	0.38FTE/year
SDP Computing - Hardware & Storage - Middleware - Common Software	Astron (Broekema) NZ (Bancroft) NZ (Christie)	 0.64FTE/year 0.40FTE/year 0.61FTE/year
Software Development Environment	NZ (Erdody)	0.38FTE/year
Science Analysis Pipeline Requirements	NZ (Johnston-Hollitt)	2.50FTE/year