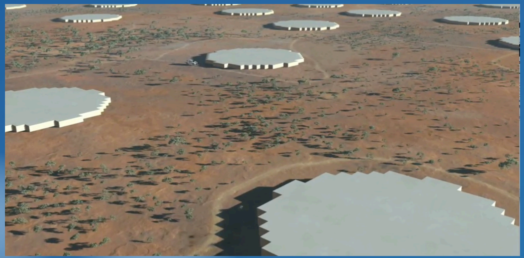


Computational challenges for SKA and SKA-Low

Tim Cornwell, Square Kilometre Array

SKA data processing challenges

- Wide field imaging
  - ASKAP, MWA, LOFAR
- Imaging with aperture arrays
  - LOFAR, MWA
- Imaging with phased array feeds
  - ASKAP
- Imaging with wide bandwidth
  - ASKAP, VLA
- Calibration and correction of direction dependent effects
  - LOFAR, MWA, MeerKAT, ASKAP, VLA
- All organizations represented at CALIM meetings
  - Challenges being addressed
- High performance computing lags

Types of processing

- SKA-Mid
  - Standard imaging
  - Pulsar timing and searching
- SKA-EOR
  - Standard imaging
  - Experimental EOR detection
- SKA-Survey
  - Standard (ASKAP-style) imaging

Costing

- The SKA Board wishes to set a cost cap
- SKA Office has to advise the Board by July 2013
- Constructing cost estimates for Baseline Design
- Baseline Design is not overly prescriptive
- Best guesses needed

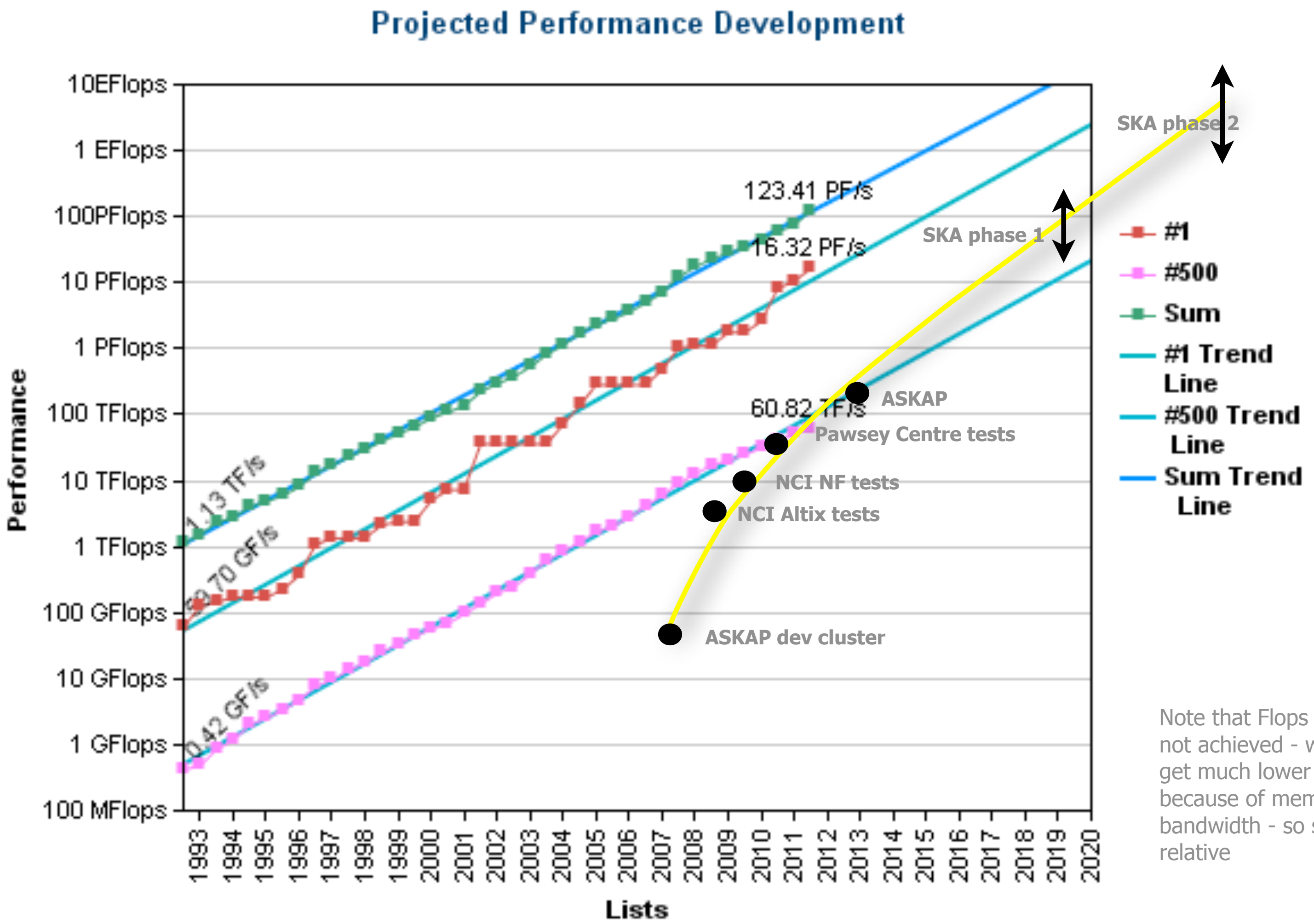
SDP software platforms

- Assume one software platform for all science cases
  - Specializations and optimizations as necessary e.g. AA, PAF
- No existing software package can be used
- Many existing algorithms will fail to scale
- Third-party apps (e.g. casacore, wcslib) must be rewritten for multi-threading
  - Roughly €10M in existing casacore
- Software platform must be supported on multiple hardware platforms
  - Mitigate risk by facilitating use of non-SKA hardware

SDP hardware platforms

- Architecture of computing platform unknown
  - Many nodes of many cores (GPU or MIC) most likely
- Could have different architectures across telescopes or applications
  - Plan explicitly for non-SKAO platforms elsewhere
- Procurement delivery is staged to maximise capabilities and match to growing needs
  - As used at Pawsey Centre
- Shift procurement into operations to procure only when techniques are mature

SKA data processing rates



Note that Flops numbers are not achieved - we actually get much lower efficiency because of memory bandwidth - so scaling is relative

Calibration and imaging cost model

- Five pipelines
  - Ingest, Calibration, Continuum, Spectral line, Transients
- Steps in processing
  - Gridding and degridding visibility data
  - Multi-Scale Multi-Frequency CLEAN

$$T_{clean} = \mu_{clean} N_{scales} N_{Taylor} N_{iterations} N_{psf}^{2}$$

- Resources
  - Processing
  - Memory
  - Fast storage

Typical costs

- Based on GPU scaling
- Runs all pipelines on longest baselines

Telescope	Diameter	Baseline	Processor	Memory	Storage	Total
	(m)	(km)	(M€)	(M€)	(M€)	(M€)
SKA1_AA_Low	35	100	14.3	0.9	1.4	16.6
SKA1_Mid	15	100	19.2	4.8	1.3	25.3
SKA1_Survey	15	100	22.6	6.5	2.7	31.8

Gridding/degridding model

$$T_{ws} = \mu_{grid} N_{vis} \left(\rho^2 \left(\frac{w_{rms}}{w_{max}} \right)^2 R_F^2 + R_A^2 \right) + \mu_{FFT} \left(2\rho R_F \left(\frac{T_{obs}}{T_A} \right) + N_{int} \right) N_{pixels}^2 \log_2 \left(N_{pixels}^2 \right)$$

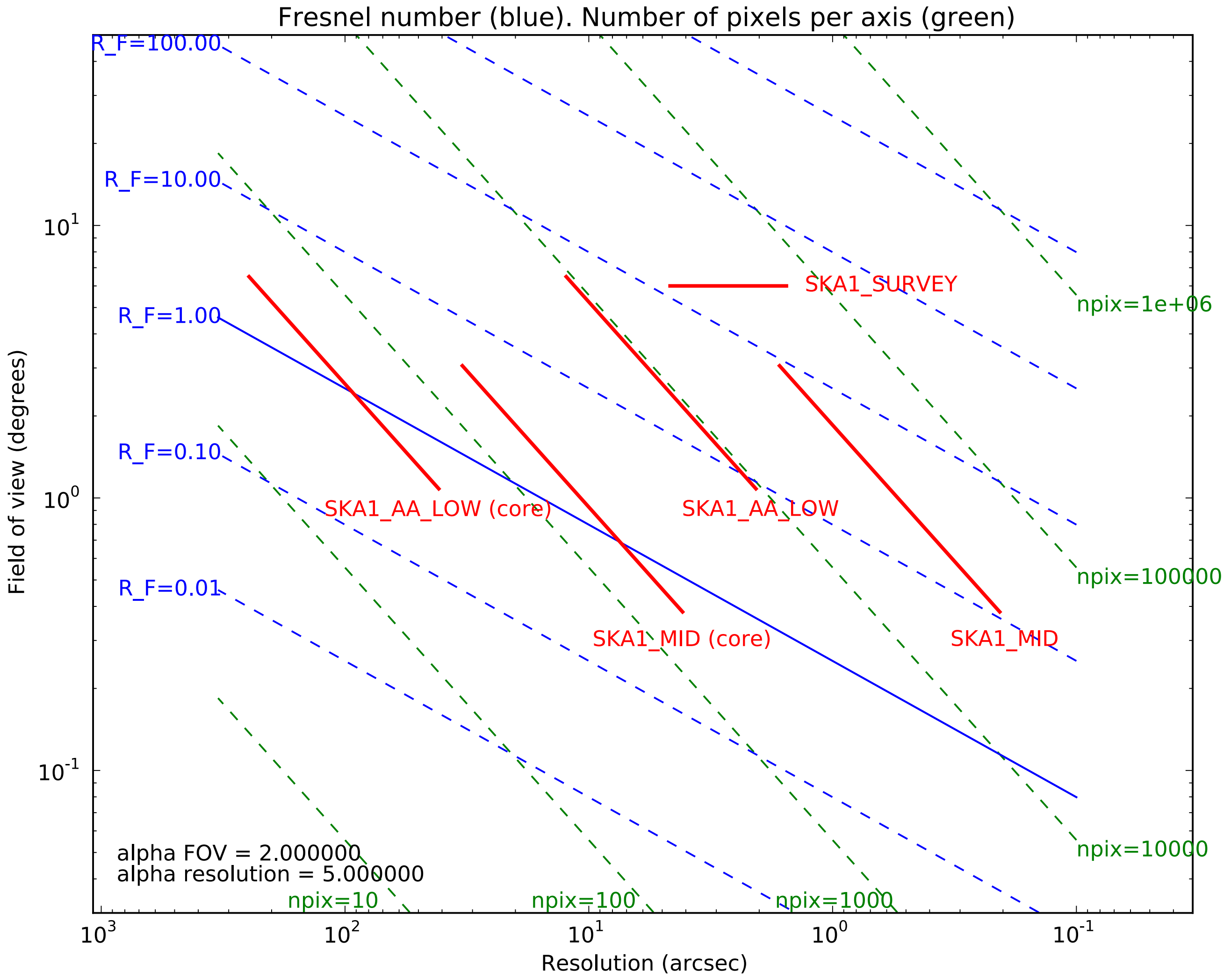
$$+ \mu_{reproj} \frac{N_{pixels}^2 h_{obs}}{\rho}$$

- AW snapshots algorithm
  - Identify best-fit plane in uvw space
  - Use AW Projection to move points onto this plane
  - Update plane at rate chosen to minimize total work
  - Some projection mandated by Earth curvature
  - Correct for coordinate distortion of each image plane
- Superior to snapshot imaging and AW Projection
  - Less CPU, memory
- Diagonal or general Mueller matrices
- Update or change model in the future as appropriate

Data handling

- Assume that single pass processing is insufficient
- Hence data must be buffered for ~ days to allow multiple calibration passes
  - Might be continuum only but could be all data
- Require large multi-day visibility data buffer for all telescopes
- Assume average throughput must = 100%

Cost of field of view



dio telescope

CPU-based scaling numbers

- Performance measured by four numbers
- Can be benchmarked by small programs
  - tConvolve $\mu_{wp}, \mu_{FFT}, \mu_{reproj}, \mu_{clean}$
- Numbers shown are from Pawsey Centre
- ASKAP Real Time Computer
  - 200TF system costing €3.2M
- Expect SDP to update cost model regularly during pre-construction and construction
  - Many assumptions that can fail at scale

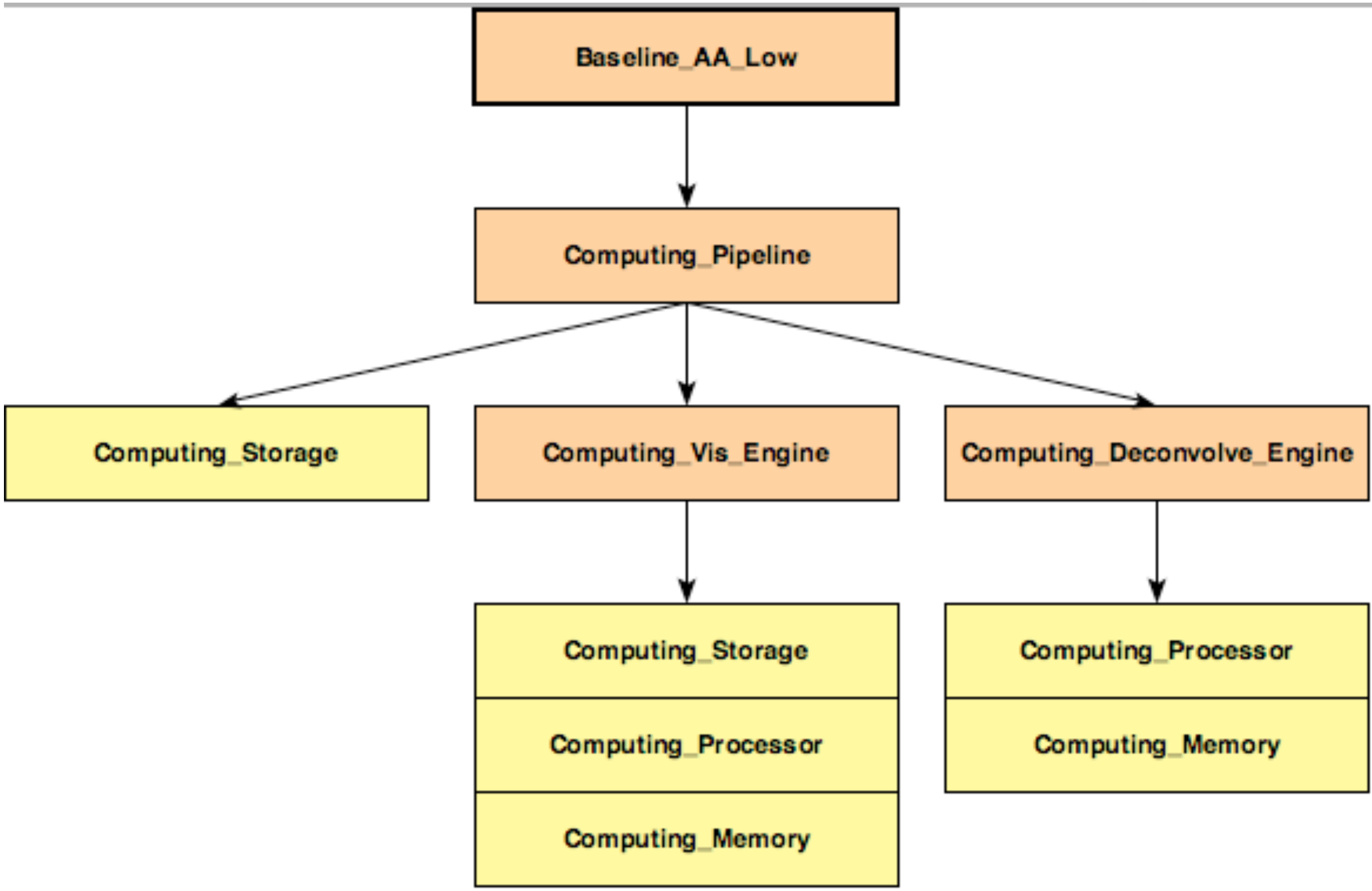
μ_{grid}	8.0e-13 s
μ_{FFT}	1.1e-11 s
μ_{reproj}	9e-10 s
μ_{clean}	1.0e-12s
Cost (2012)	€1.6M

GPU-based scaling numbers

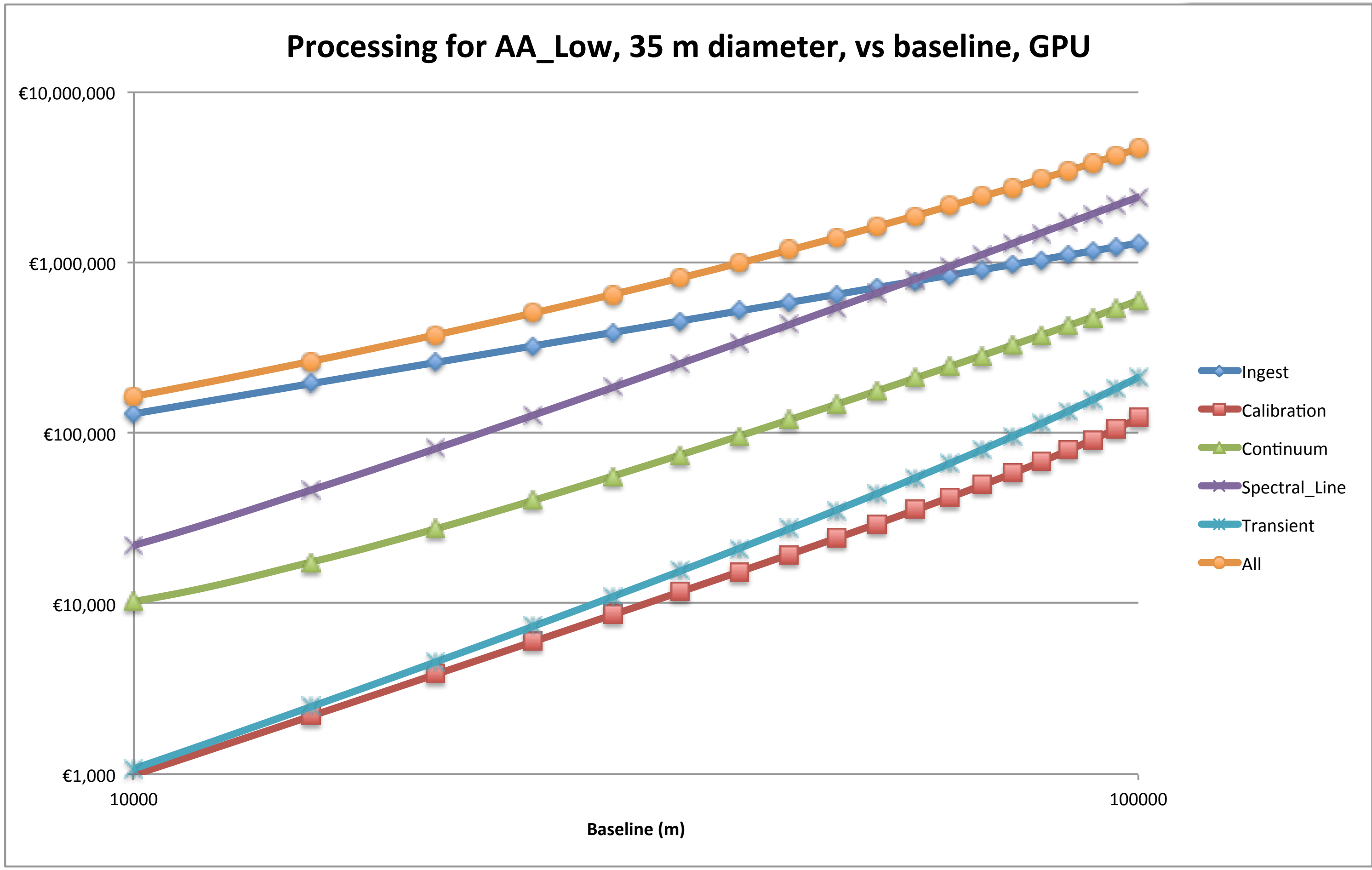
- Somewhat speculative
- Grid, clean measured
- FFT, reproject scaled
- Substantially better than CPU

μ_{grid}	8.0e-14 s
μ_{FFT}	1.1e-12 s
μ_{reproj}	9e-11 s
μ_{clean}	2.0e-13s
Cost (2012)	€1.6M

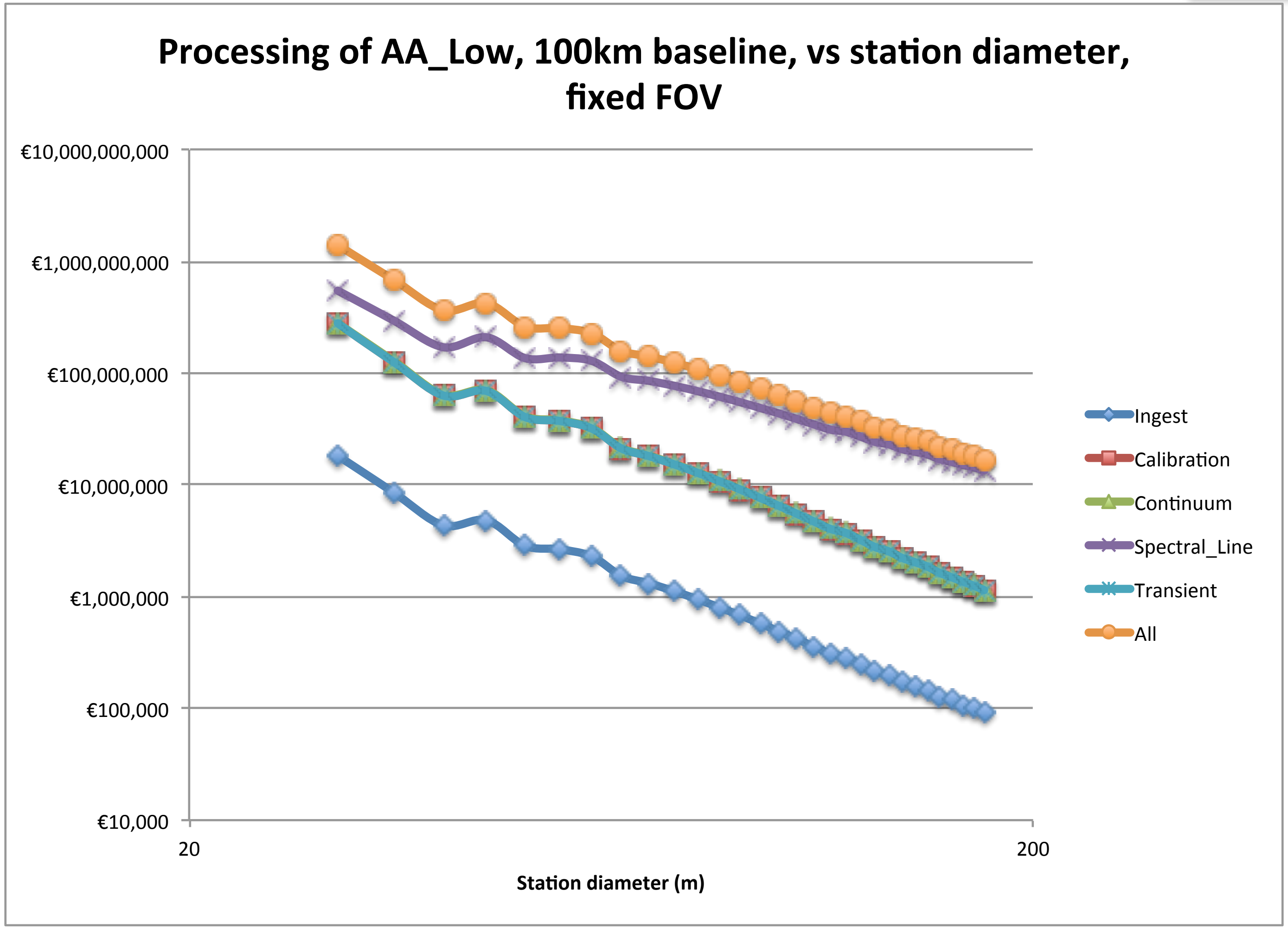
Structure of costing model



Cost of resolution



Cost of diameter, fixed field of view



Things we need to know for costing

- Observing scenarios
- Calibration of processing numbers
- Firmer numbers for
  - Maximum baseline
  - Station diameter
  - Number of mosaiced beams
  - Number of (science) spectral channels
- Thoughts on beams, polarisation, etc.
- Other science cases