Department of Applied Physics
Curtin Institute of Radio Astronomy

System Design for the Square Kilometre Array: New Views of the Universe

Timothy Maarten Colegate

This thesis is presented for the Degree of
Doctor of Philosophy
of
Curtin University

November 2012

Declaration

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgment has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

.....

Timothy M. Colegate

26 November 2012

Acknowledgements

I have many people to thank for guidance and advice through my PhD journey. My supervisors, Prof. Peter Hall and Prof. Steven Tingay, have supported me to find my niche in the SKA project. In particular, I thank Peter for his pragmatism that helped cut through the detail. I also thank Prof. Mervyn Lynch for initially getting me interested in the developing radio astronomy opportunities in Western Australia. In terms of financial support, I have been fortunate to receive an Australian Postgraduate Award and a Curtin Research Scholarship.

Although there are many other people at the Curtin University node of the International Centre for Radio Astronomy Research (ICRAR) who have contributed in one way or another, I would particularly like to thank Wayne Arcus, Ramesh Bhat, Nathan Clarke, Dave Emrich, Leith Godfrey, Aziz Jiwani, J-P Macquart, Steve Ord, Adrian Sutinjo, Steven Tremblay and Mark Waterson. Thanks also to ASTRON colleagues Jan Geralt Bij de Vaate for helping me get up to speed with aperture arrays during his time at Curtin University, and Andre Gunst for his input.

Thanks to Ron Ekers for being an early and active user of the SKA cost and performance modelling software. Thanks to Rosie Bolton, Dominic Ford and Paul Alexander for the collaboration on SKACost, and kindly hosting my visits to the University of Cambridge. Thanks to Richard Schilizzi, Billy Adams, Lisa Bell, Jo Bowler, Kobus Cloete, Phil Crosby, Peter Dewdney, Colin Greenwood, Duncan Hall, Roshene McCool, Neil Roddis and Wallace Turner for making me feel welcome at the then SPDO (SKA Program Development Office) in Manchester; I have appreciated the many discussions surrounding SKA system design and costing. Thanks to Guido Aben and John Nicholls at AARNet for helping me understand optic fibre networks.

And finally I thank those to whom I owe many years of my life. My mates, who wondered if I would ever get a 'real job'. Marnelke and Brian, who always encourage my inquisitive mind, and along with Beth and Katharine, were always interested in what I was doing, but left me alone when I really needed to get the thesis writing done! And Kate, who did not quite know what she was getting into, and put her dreams on hold to help me get through.

Abstract

The Square Kilometre Array (SKA) radio telescope is being designed as a premier scientific instrument of the 21st century, using novel technologies to maximise its scientific capability. The SKA has an aggressive project timeline, dynamic and evolving scientific requirements, and a large design exploration space with many interdependent sub-systems. These complexities increase the difficulty in developing cost-effective design solutions that maximise the scientific capability of the telescope within construction and operations funding constraints.

To gain insight into specific design challenges in this thesis, I have developed parametric models of the telescope system that relate cost to key performance metrics. I examine, as case studies, three aspects of the SKA design that have had little investigation compared to the rest of the telescope to date, but show considerable potential for discovering new astronomical phenomena.

First, I present fast transient survey strategies for exploring high time resolution parameter space, and consider the system design implications of these strategies. To maximise the scientific return from limited processing capacity, I develop a new metric, 'event rate per beam', to measure the cost-effectiveness of the various search strategies. The most appropriate search strategy depends on the observed sky direction and the source population; for SKA Phase 1, low-frequency aperture arrays tend to be more effective for extragalactic searches, and dishes more effective for directions of increased scatter broadening, such as near the Galactic plane.

Second, I compare the cost of two design solutions for low-frequency sparse aperture array observations (70–450 MHz) that achieve similar performance: a single-band implementation with a wideband antenna design; and a dual-band implementation, with each array observing approximately half the fractional bandwidth. Perhaps somewhat surprisingly, despite the dual-band array having twice the number of antenna elements, neither a representative single or dual-band implementation is cheaper *a priori*, although the uncertainties are currently high. In terms of the broader telescope system design, I show that the central processing, antenna deployment and site preparation costs are potentially significant cost drivers that have so far had insufficient attention.

Third, the recent site decision gives rise to the question of how to cost-effectively provide data connectivity to widely separated antennas, to enable high angular resolution observations with the SKA dish array in Africa. To facilitate the design of such a data network, I parametrise the performance and cost of an exemplar network using three simple metrics: maximum baseline length; number of remote stations (grouped antennas) on long baselines; and the product of bandwidth and number of station beams. While all three metrics are cost drivers, limiting the beam–bandwidth product reduces cost without significantly impacting scientific performance.

The complexities of the SKA design environment prevent straightforward analyses of cost-effective design solutions. However, the case studies in this thesis demonstrate the importance

of parametric performance and cost modelling of the telescope system in determining cost-effective design solutions that are capable of revealing large regions of unexplored parameter space in the radio Universe. The analytical approach to requirements analysis and performance–cost modelling, combined with pragmatic choices to narrow the exploration space, yields new insights into cost-effective SKA designs. Continuation of this approach will be essential to successfully integrate the forthcoming results from various verifications systems into the SKA design over the next few years.

Contents

List of Figures

List of Tables

Abbreviations, acronyms and symbols

The page number indicates where the symbol or abbreviation is first introduced.

Principal abbreviations and acronyms

AA aperture array (p 30)
AGN active galactic nuclei (p 129)
AIP Advanced Instrumentation Program (p 9)
ALFA Arecibo L-band Feed Array (p 62)
ASKAP Australian SKA Pathfinder (p 11)
ATA Allen Telescope Array (p 28)
CER cost estimating relationship (p 50)
CoDR Concept Design Review (p 36)
CPF central processing facility (p 99)
CRAFT Commensal Real-time ASKAP Fast Transients (p 62)
DM dispersion measure (p 65)
$\mathbf{DRM_1}$ SKA_1 Design Reference Mission (SSWG, 2012) (p 43)
$\mathbf{DRM_2}$ SKA_2 Design Reference Mission (SSWG, 2009) (p 43)
DWDM dense wavelength division multiplexing (p 131)
e-VLBI electronic Very Long Baseline Interferometry (p 16)
EoR epoch of re-ionisation (p 4)
FoM figure of merit (p 22)
FoV field of view (p 24)
GMRT Giant Meterwave Radio Telescope (p 3)
HLSD high-level system description (p 36)
ICRAR International Centre for Radio Astronomy Research (p v)
JVLA Karl G. Jansky Very Large Array (p 3)
KSP Key Science Project (p 4)
LCC life-cycle cost (p 48)
LNA low noise amplifier (p 116)
LOFAR Low Frequency Array (p 11)
LWA Long Wavelength Array (p 161)
MWA Murchison Widefield Array (p 11)
PAF phased array feed (p 29)
PAPER Precision Array for Probing the Epoch of Re-ionization (p 3)
PBS product breakdown structure (p 50)
PDF probability distribution function (p 58)
PoI probability of intercept (p 5)
RF radio frequency (p 27)

RFI radio frequency interference (p 32)
SEMP *System Engineering Management Plan* (Stevenson, 2011c) (p 40)
SKA Square Kilometre Array (p 1)
SKA$_1$ SKA Phase 1 (p 2)
SKA$_2$ SKA Phase 2 (p 2)
SKADS SKA Design Studies (p 15)
SKA$_1$-low low-frequency sparse aperture arrays for SKA Phase 1 (p 1)
SPDO SKA Program Development Office, Manchester, UK. Now Office of the SKA Organisation (p v)
SPF single-pixel feed (p 28)
SSFoM survey speed figure of merit (p 24)
VLBA Very Long Baseline Array (p 16)
VLBI Very Long Baseline Interferometry (p 16)
WBS work breakdown structure (p 50)

Symbols

A_{e} Effective area (p 23)
$A_{\mathrm{e-0}}$ Effective area of a receptor (p 28)
$A_{\mathrm{e-arr}}$ Array effective area (p 32)
$A_{\mathrm{e-dish}}$ Dish effective area (p 28)
$A_{\mathrm{e-st}}$ Station effective area (p 30)
A/T Sensitivity metric (p 23)
b_{max} Maximum baseline length (p 34)
b_{min} Minimum baseline length (p 35)
C_{block} Block cost (p 52)
C_{fix} Fixed cost (p 52)
C_{total} Total cost (p 52)
C_{var} Variable cost (p 52)
d_{avg} Average inter-element spacing (p 105)
D_{arr} Array diameter (p 33)
D_{dish} Dish diameter (p 28)
D_{st} Station diameter (p 30)
k Boltzmann's constant ($1.38 \times 10^{-23}\,\mathrm{J\,K^{-1}}$) (p 23)
$\mathcal{K}_0$ Receptor taper or aperture illumination factor (p 72)
$\mathcal{K}_{\mathrm{arr}}$ Array beam taper (p 33)
$\mathcal{K}_{\mathrm{dish}}$ Dish aperture illumination factor (p 29)
$\mathcal{K}_{\mathrm{st}}$ Station beam taper (p 31)
$\mathcal{L}$ Pulse luminosity (p 67)
N_0 Number of receptors (p 32)
$N_{\mathrm{b-0}}$ Number of receptor beams (p 31)
$N_{\mathrm{b-arr}}$ Number of (phased or tied) array beams (p 33)
$N_{\mathrm{b-dish}}$ Number of dual polarisation beams per dish (p 29)
$N_{\mathrm{b-st}}$ Number of dual polarisation beams per station (p 31)

$N_{\mathrm{b\,-\,tile}}$	Number of dual polarisation tile beams (p 124)
$N_{\mathrm{bit\,-\,dig}}$	Number of bits out of digitiser (p 31)
$N_{\mathrm{b\,-\,RSt}}$	Number of dual polarisation beams per remote station (p 136)
$\overline{N_{\mathrm{b\,-\,st}}}$	Average $N_{\mathrm{b\,-\,st}}$ over the band (p 105)
N_{ch}	Number of channels (p 77)
N_{det}	Number of detections (p 68)
$N_{\mathrm{dish/RSt}}$	Number of dishes per remote station (p 138)
$N_{\mathrm{e/st}}$	Number of elements per station (p 30)
$N_{\mathrm{e/tile}}$	Number of elements per tile (p 105)
N_{pix}	Number of pixels (p 71)
N_{pol}	Number of polarisations (p 20)
N_{RSt}	Number of remote stations in the array (p 134)
N_{sa}	Number of sub-arrays (p 71)
N_{st}	Number of stations in the array (p 105)
r_{max}	Maximum astronomical distance (depth) (p 25)
R_{RSt}	Total data rate from a remote station (p 136)
$\mathcal{R}$	Event rate (p 67)
$\mathcal{R}_{\mathrm{beam}^{-1}}$	Event rate per beam (p 68)
S	Flux density (p 20)
S_{min}	Minimum detectable flux density (p 23)
T_{rec}	Receiver temperature (p 27)
T_{sys}	System temperature (p 23)
T_{tot}	Total survey time (p 68)
W	Observed pulse width (p 67)
W_{i}	Intrinsic pulse width (p 67)
Δt	Temporal resolution (integration time) (p 20)
$\Delta\nu_{\mathrm{ch}}$	Spectral resolution (channel width) (p 20)
$\Delta\nu$	Processed bandwidth (p 23)
η_{ap}	Aperture efficiency (p 28)
η_{r}	Antenna element radiation efficiency (p 30)
θ_{HP}	Half-power (3 dB) beamwidth (p 28)
$\nu_{\mathrm{transition}}$	Dense–sparse transition frequency (p 75)
ξ	Spectral index (p 74)
ϕ	Angular resolution (p 20)
ρ_i	Intrinsic event rate volume density (p 6)
$\rho_{i,\Omega}$	Intrinsic event rate per unit solid angle (p 85)
τ	Post-detection integration time (p 23)
τ_{d}	Pulse broadening time (p 74)
Ω_0	Receptor beam FoV (p 28)
Ω_{arr}	Array beam FoV (p 33)
Ω_{proc}	Station processed FoV (p 24)
Ω_{req}	Required processed FoV (p 105)
Ω_{st}	Single station beam FoV (p 31)

Chapter 1

Introduction

The Square Kilometre Array (SKA) is being designed as general-purpose radio telescope, with the potential to be an iconic scientific instrument of the 21st century. Once constructed, the telescope will undertake transformational science programs, including many observations not possible with earlier instruments. This scientific capability will be realised through a combination of increased collecting area and new technologies. Furthermore, technological improvements will ensure a continuous increase in capability, with the full potential of the telescope not being realised for decades.

A cost-effective design for the SKA will maximise scientific capability within construction and operations funding constraints. In this thesis, I employ a systems view of the SKA to identify cost-effective solutions to three aspects of the telescope design that, to date, have had little investigation compared with the rest of the telescope: high time resolution observations for fast transients; dual-band alternatives to the wideband antenna element implementation for the low-frequency sparse aperture arrays (SKA-low); and data networks for the long baselines of the dish array (high angular resolution astronomy). I frame these design challenges as case studies that exemplify how parametric performance and cost models enable the development of cost-effective design solutions for the SKA. The models in the case studies use parametric equations to provide estimates of performance and cost for various design solutions, by varying key input parameters. The parameters generally describe the physical attributes of the system; for example, the number of dishes in the array and dish diameter.

This thesis is also a first attempt to explore the overlapping scientific potential of these seemingly disparate design challenges. With a suitably designed telescope signal processing architecture, searches for fast transients can leverage both low-frequency aperture arrays and long baselines to maximise new parameter space observed by the SKA. These new views of the Universe are not accessible to contemporary radio telescopes.

For the remainder of this chapter, I outline the SKA science goals and design challenges. I then describe further the thesis motivation and outline in the context of the SKA.

1.1 The Square Kilometre Array

The SKA will be a powerful and flexible mega-science instrument, with capabilities significantly greater than existing radio telescopes. It will greatly improve our understanding of the Universe with key science goals encompassing the most important pre-existing contemporary problems in radio astronomy. Answers to these problems will impact areas such as astroparticle and fundamental physics, cosmology and astrobiology (Carilli & Rawlings, 2004).

The SKA (Schilizzi et al., 2011):

- is a global project, with a target construction cost of €1 500 M (2007 value)

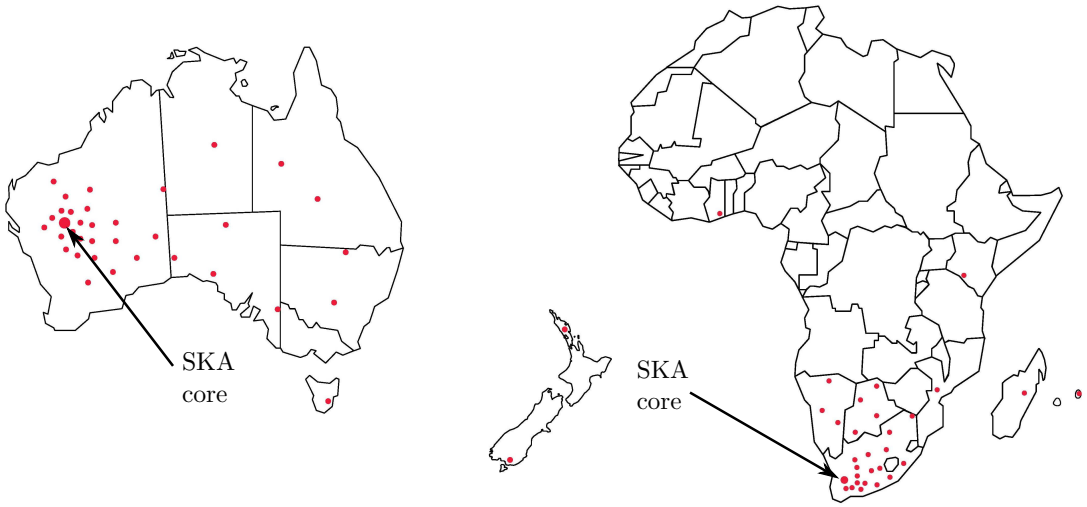


Figure 1.1: Proposed geographical layout of the antennas in Australia and Africa (continents not to scale), prior to site selection. The dots indicate antenna locations, the arrow indicates the core of the telescope. Adapted from: SPDO.

- is currently in the design phase, with construction due to commence in the next few years
- has a phased implementation approach to construction
  - Phase 1 (SKA_1), with a €350 M budget (2007 value), will realise a subset of the Phase 2 receptor (antenna) technologies with typically ∼10–15 % of the Phase 2 collecting area
  - Phase 2 (SKA_2) will contain the full collecting area at frequencies of 70 MHz–10 GHz
  - Phase 3 is planned for higher frequencies (up to 25 GHz or more)
- has a broad range of science requirements and a complex set of specifications
- will use a mix of custom-developed and commercially available technology.

Two candidate sites were shortlisted as scientifically suitable to host the core of the telescope: the state of Western Australia and the Karoo region of South Africa, with antennas to be spread across the continent of the chosen site, as shown in Figure 1.1. A recent site decision[1] splits the telescope between these sites, such that Australia hosts the low-frequency sparse aperture array (AA) component (70–450 MHz) and Southern Africa hosts the main dish component (0.45–10 GHz). Furthermore, as part of SKA_1, Australia will host an additional 'survey' array of up to 96 dishes equipped with novel phased array feed (PAF) technology, while the mid-frequency dense aperture array technology for SKA_2 will be hosted in South Africa.

The general-purpose capabilities of the SKA will be significantly greater than any other current or forthcoming radio telescope. Two key measures of radio telescope performance are sensitivity and 'processed' field of view (FoV). Figure 1.2 plots these metrics for various telescopes at SKA frequencies, where sensitivity is plotted as minimum detectable flux density $S_{\rm min}$. To show frequency and source-dependent effects, Figure 1.2 follows Fender & Bell (2011) by correcting for frequencies below 1.4 GHz using $S_{\rm min} \propto \nu^{\xi}$, where $\xi = -0.7$ for optically thin synchotron sources and $\xi = -2.0$ for coherent sources. This correction allows a more nuanced comparison

[1] http://www.skatelescope.org/news/dual-site-agreed-square-kilometre-array-telescope

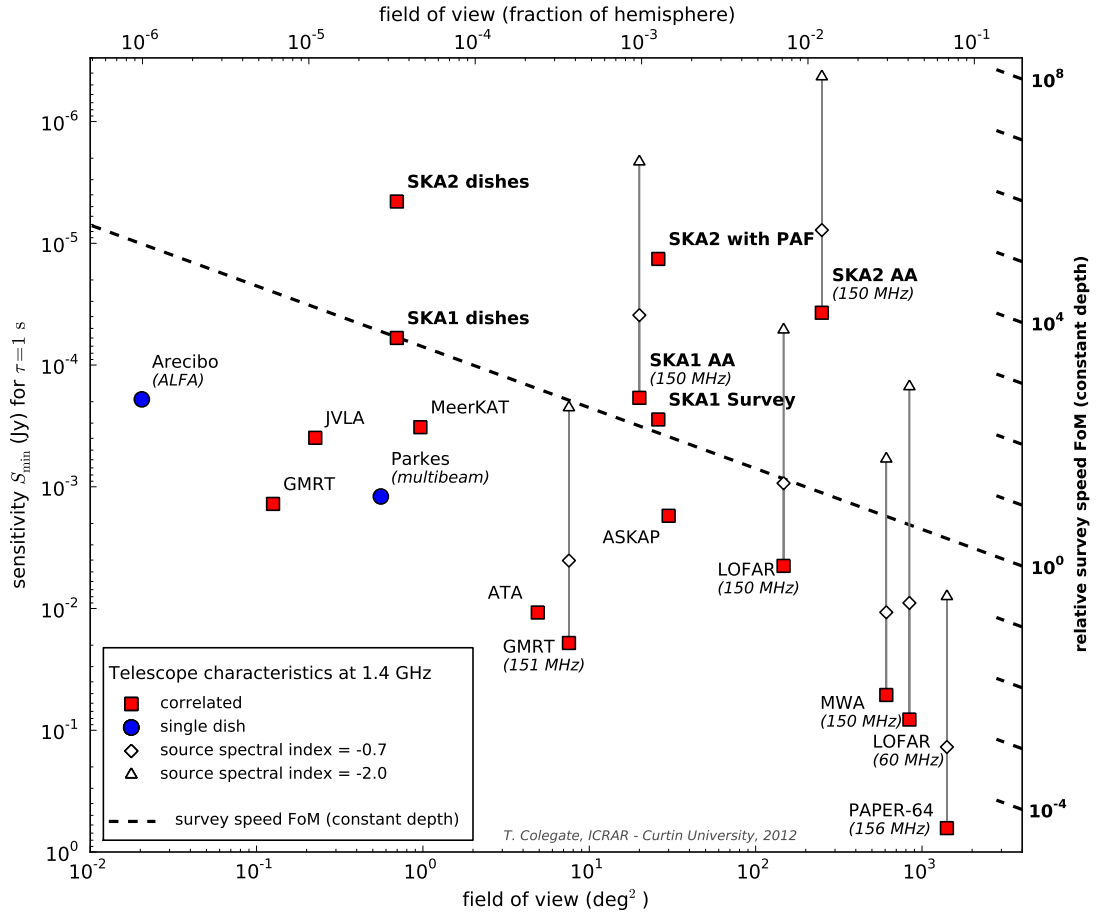


Figure 1.2: Field of view and sensitivity parameter space for present and forthcoming radio telescope arrays (red squares) and single-dish telescopes (blue circles). Higher sensitivity (towards the top of the figure) corresponds to a lower minimum detectable flux density ($S_{\rm min}$), calculated for $\tau = 1$ s integration time at a 1σ significance level. The dashed line shows constant survey speed (to fixed depth), arbitrarily centred at the SKA$_1$ dish array; the right-hand axis shows the survey speed figure of merit relative to this line. For observing frequencies less than 1.4 GHz, the open markers indicate the equivalent sensitivity at 1.4 GHz due to the source spectral index. See text for further details.

of the effectiveness of the telescopes at different frequencies; the plot shows that the SKA will excel in both sensitivity and FoV at all frequencies.

The dashed diagonal line in Figure 1.2 shows a survey speed figure of merit (SSFoM), being proportional to FoV and the square of sensitivity (Bunton, 2003a; Jones, 2004). This SSFoM is a measure of the effectiveness of a radio telescope in surveying the sky, and is commonly employed to assess surveys for steady sources; it describes the speed in which an area of sky is surveyed to a given sensitivity, or 'depth' (Cordes, 2009b). Maximising the survey speed (towards the top-right of Figure 1.2) is crucial for the SKA to detect large numbers of objects, or alternatively, rare and weak objects. Although Figure 1.2 does not cover every facet of telescope performance, it is a cogent representation of a telescope's capability in observing astronomical sources.

1.1.1 SKA science case

The science case for the SKA describes the motivation for the telescope design. It places the planned scientific capabilities in the context of other current and future astronomical observatories of all wavelengths, as well as other scientific instruments such as the Large Hadron Collider. The SKA was originally conceived to observe neutral hydrogen, requiring approximately $1\,\mathrm{km}^2$ of collecting area (DRAO, 1991; Noordam et al., 1991; Swarup, 1991; Wilkinson, 1991). The 'Square Kilometre Array Interferometer' concept evolved into a more general-purpose, next-generation radio telescope (e.g. Braun, 1996), which led to the first science case for the SKA (Taylor & Braun, 1999).

Science with the SKA (Carilli & Rawlings, 2004) was a significant revision and enlargement of the science case. This revision included five Key Science Projects (KSPs) that represent unanswered questions in fundamental physics or astrophysics which interests the government funding agencies and the broader community, and for which the SKA has a key role in answering (Gaensler, 2004). They are (Carilli & Rawlings, 2004; Hall et al., 2008):

- KSP I. The cradle of life: extra-solar searches for Earth-like planet formation and signal transmissions from other civilisations.
- KSP II. Strong-field tests of gravity using pulsars and black holes: laboratories for tests of fundamental science, such as general relativity.
- KSP III. The origin and evolution of cosmic magnetism: characterisation of magnetic fields in the galaxies and clusters of galaxies.
- KSP IV. Galaxy evolution and cosmology [and dark energy]: how stars are formed from gas within galaxies.
- KSP V. Probing the dark ages [and the epoch of re-ionisation (EoR)]: making steps to understand the transition of the intergalactic medium from a neutral to ionised state.

The KSPs drive the SKA design, through top-level scientific requirements such as frequency range, sensitivity, survey speed, angular resolution and number of independent fields of view.

However, the most important legacy of the SKA could actually be 'exploration of the unknown': serendipitous discoveries which fundamentally change our knowledge of astronomy (e.g. Gaensler, 2004; Wilkinson et al., 2004). Whereas the KSPs aim to answer *known* questions in fundamental physics or astrophysics, the answers to the *unknown* questions may be most interesting.

In a detailed analysis of astronomical discoveries, Harwit (1981) found that the use of novel technology soon uncovered new phenomena. Historical events show the validity of this proposition. An oft-used example is the serendipitous discovery of pulsating neutron stars (pulsars) by Hewish et al. (1968), for which the enabling technology was the higher time resolution telescope receiving system, implemented to study scintillation from the interplanetary medium. Another example is the high-velocity water-vapour maser emission from the nucleus of the active galaxy NGC4258 discovered by Nakai et al. (1993); this discovery was made possible by a wide-bandwidth spectrometer with many channels, enabling the measurement of a large range of velocities in a single observation (Moran, 1993). Mapping of this emission with follow-up

Very Long Baseline Interferometry (VLBI) observations in turn led to "compelling evidence" of a massive black hole at the galaxy's centre (Miyoshi et al., 1995).

Similarly, the large increase in performance capability of the SKA, compared with existing radio telescopes, greatly improves the potential for unexpected discoveries (Braun, 1996). To maximise this capability to discover scientific phenomena, the exploration of the unknown is a recommended design philosophy for the SKA (SSWG, 2009). Wilkinson et al. (2004) outline a design and operation philosophy to facilitate serendipitous discoveries: construct a telescope capable of observing new regions of parameter space, and flexibly operate that telescope to maximise access to this parameter space. This operational flexibility and evolving capability through new technology will enable innovative observers to make unexpected and, most likely, astonishing discoveries.

What is the potential for exploration of the unknown, and how should capability be measured? Chapter 2 uses Harwit's (1981) concept of parameter space as a broad measure of the scientific capability of a telescope to observe existing and new phenomena. This capability includes large-scale surveys, which are a central feature of the KSPs (Gaensler, 2004). Discoveries with these surveys are likely to be made using the statistical properties of a large population, or through the recognition of a rare, previously unknown class of object (Harwit, 1998). In terms of parameter space, the capability of a survey can be measured by rate of information gathered in the survey. Compared with existing telescopes, the increased sensitivity and FoV of the SKA significantly improves the information rate, as measured by the survey speed figure of merit in Figure 1.2. This surveying capability will be supported by high spectral and angular resolution over a large range of frequencies. The combination of these capabilities will make the SKA a powerful instrument for both analysis of existing phenomena and discovery of new phenomena.

1.1.2 The high time resolution Universe

The high time resolution Universe presents a notable opportunity for the SKA to search for new phenomena. For this reason, I have chosen, as one topic in this thesis, to analyse the capability of the SKA in this observing domain. Even though pulsar observations have been an active research area since their discovery nearly half a century ago by Hewish et al. (1968), the dynamic sky has not been well sampled at radio frequencies, compared to the X-ray and gamma-ray frequency bands (Cordes et al., 2004). This is especially true of impulsive singly-occurring or intermittent signals, emitted from high energy density events ('fast transients').

Digital signal processing and computing advances have enabled more powerful searches for pulsars and fast transients. Through such searches, a small number of energetic, single-pulse events of potentially extragalactic origin have been recently discovered by Lorimer et al. (2007), Keane et al. (2011, 2012) and M. Bailes et al. (2012, pers. comm.), using the Parkes 1.4 GHz multibeam system. These events, along with continuing technological advances, is creating a growing interest in exploration of the transient universe.

Unlike continuous or periodic sources, fast transient searches cannot use long integration times to improve sensitivity, nor can further details about the event be gained with follow-up observations (although there may be evidence of the event at other wavelengths). As a result, the

strategies to maximise the potential of fast transient searches are distinct from continuous or periodic sources. The main goal of at least first-generation fast transient searches is to maximise the number of verifiable events detected in a survey, and to accumulate sufficient data to enable follow-up investigation of detections (Macquart et al., 2010b).

The number of events detected is closely related to the 'probability of intercept' (PoI), which describes the effectiveness of a fast transient survey (Hall et al., 2012). It is a function of instrumental capability, source population characteristics and the nature of the intervening medium. Macquart (2011) formalises this relationship with an event rate metric, being the number of events detected per unit time. This allows proper comparison of fast transient surveys of different source populations, telescopes and detection systems.

Large single-dish telescopes have typically been used for pulsar and fast transient searches (e.g. Cordes et al., 2006; Keith et al., 2010; Manchester et al., 2001). For single-dish telescopes, the improved sensitivity of a larger diameter telescope is offset by its reduced FoV (although multi-beam or phased array feeds recover some of this FoV). Therefore, increases in the probability of intercepting a given population has generally been limited to longer observations and more sensitive radiometers and/or signal processors.

For radio telescope arrays, the options are more varied. For most science observations, images are produced by correlating the antenna signals and averaging (integrating) the output over a few seconds to reduce subsequent processing costs. Adding more antennas to the array increases sensitivity without compromising FoV. But correlation and imaging processing costs are currently too expensive to achieve 'fast imaging' at the time resolution of milliseconds (or better) required for the detection of fast transients. The alternatives are a highly sensitive signal combination mode with small FoV (such as the coherent combination of antenna signals) or less sensitive modes that cover more of the sky (incoherent combination, subarraying and 'fly's eye'). There are trade-offs between each of these; the best choice being also influenced by the spatial density of antennas in the array. Furthermore, additional processing capability enables multiple beams to be formed and searched, re-using the array collecting area to some extent.

Figure 1.3 presents my synopsis of contemporary and forthcoming fast transient searches at $\sim$1.4 GHz, including SKA_1 and SKA_2 (the derivation of which is discussed more in Section 4.6.1). The potential of SKA_1 to find both rare and weak events surpasses existing telescopes, through the use of both incoherent and coherent signal combination modes. SKA_2 will further increase this potential. The fast imaging mode (imaging at millisecond time resolution) is also shown to emphasise future SKA capability. To illustrate that the PoI is easily relatable to known or putative astrophysical events, the blue and green lines in Figure 1.3 show nominal event rate limits for a hypothetical population of giant-pulse emitting pulsars, extrapolated to extragalactic distances; surveys towards the top-right of Figure 1.3 have a large PoI. The characteristics of this population is based on giant pulses intermittently emitted by the Crab pulsar, which are extremely bright at radio wavelengths. Besides being an exciting discovery in itself, detecting pulses at extragalactic distances could reveal details about the physics of the intervening medium (Macquart et al., 2010b).

The bold blue lines in Figure 1.3 show, for a targeted observation, the FoV–time product and

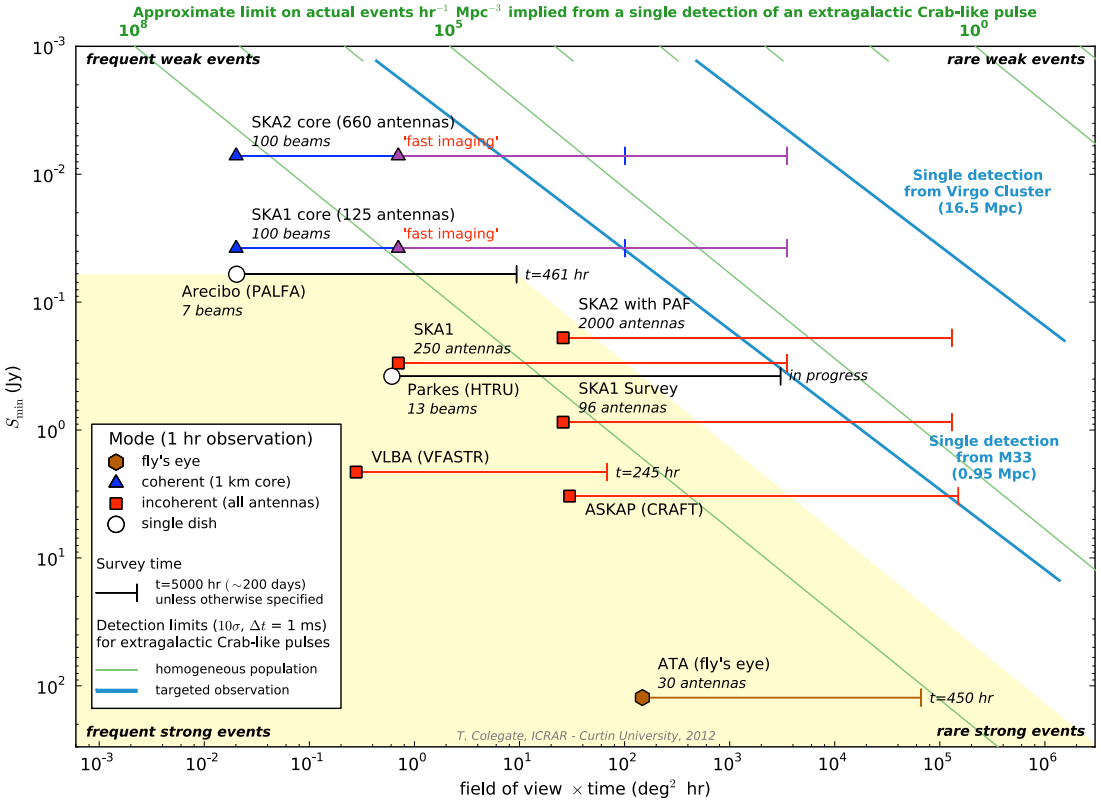


Figure 1.3: High time resolution parameter space for selected telescopes and surveys at ∼1.4 GHz. Sensitivity is plotted on the vertical axis (more sensitive observations towards the top of the figure), while the horizontal axis shows the product of FoV and observing time. For the array telescopes, the fly's eye and coherent and incoherent signal combination modes are shown as relevant. Averaging time Δt is 1 ms, $S_{\min}$ is 10σ significance. The green and blue diagonal lines are related to the probability of intercept; they show nominal event rate limits for extragalactic Crab-like pulsars. The yellow shaded region shows the highest sensitivity and event rate limit from a completed survey to date. See text and Section 4.6.1 for further details.

sensitivity required to detect a single pulse from this example population at a galaxy or cluster of galaxies at the labelled distance. The intrinsic event rate (actually occurring events, whether observable or not) for such a population is extrapolated from an estimated population of Crab-like pulsars. As an example of the application of PoI to an extragalactic population, the green lines show limits on the intrinsic event rate volume density ρ_i (events $\mathrm{hr}^{-1}\,\mathrm{Mpc}^{-3}$) due to the detection of a single event from a homogeneously distributed population in Euclidean space.

An aspect not captured by Figure 1.3 is the verification capability of the SKA or other arrays. Single-dish telescopes have limited ability to localise candidate fast transient events, which makes multiwavelength follow-up difficult (Kaplan et al., 2009). Radio telescope arrays generally have better localisation capability, along with more options to discriminate against radio frequency interference. Chapter 4 further discusses signal combination modes and strategies to maximise the effectiveness of searches of high time resolution parameter space.

1.1.3 Technical challenges for the conceptual design

The Key Science Projects, along with the exploration of the unknown, drive the specifications for the SKA design (Schilizzi et al., 2011). Constructing the SKA to accomplish the science goals while maintaining system flexibility presents a number of technical challenges. They include (Dewdney et al., 2009):

- developing cost-effective methods to build a large amount of collecting area
- reducing system noise to achieve the required sensitivity
- transmitting large data rates from the antennas to a central location
- processing wide bandwidth signals from thousands of antennas
- calibrating the telescope and forming images over large fields of view
- constructing the telescope in remote areas with limited power generation or distribution capacity.

Other factors include the risk of using new technologies, the operations versus capital cost of the instrument and site-specific costs that are affected by geographical diversity.

The SKA high-level conceptual diagram (Figure 1.4) gives an overview of the telescope's design. Figure 1.4 presents many similarities to other radio telescope arrays: radio signals are received and digitised relatively close to the antennas; transported to a central location; combined through either cross-correlation or beamforming; and then post-processed to produce radio images (from cross-correlation) or other measurements, such as pulsar and fast transient detections. In view of the dual-site decision, the central blocks (central processing facility and 'SKA HQ') in Figure 1.4 would be implemented at each site, rather than as single entities shown here (SOWG, 2012).

Like many science projects, the SKA is cost-constrained, so ultimately there will be compromises and trade-offs to achieve a cost-effective design. One such compromise is implicit in Figure 1.4, where the receptor concepts belong to one of two categories, depending on the maturity of the technologies and design for full-scale production. Two receptor types form the relatively low-risk 'baseline design': low-frequency sparse aperture arrays and dishes with single-pixel feeds, illustrated in Figure 1.5. The receptor concepts labelled 'AIP' in Figure 1.4 are higher performance technologies under development in the Advanced Instrumentation Program. In simple terms, the AIP technologies operate at the mid-band frequencies ($\sim$0.45–10 GHz) to increase the amount of FoV or bandwidth that can be processed at one time, thereby increasing survey speed. If sufficiently mature, they will be considered for inclusion in the SKA design as part of the phased approach to the telescope's design and construction (Schilizzi et al., 2011). In particular, FoV is increased using the PAF technology planned for the SKA_1 dish survey array and potentially the SKA_2 dish array; the mid-frequency dense aperture arrays are also a potential technology for SKA_2. The wider bandwidth for SKA_2 would be achieved with wideband single-pixel feeds on the dish array.

The inherent scalability of radio telescope arrays allows for the phased design and construction of the SKA; Table 1.1 shows the current timeline for the design and construction of SKA Phases 1 and 2. Because the antennas form a collection of interferometers whose signals are combined

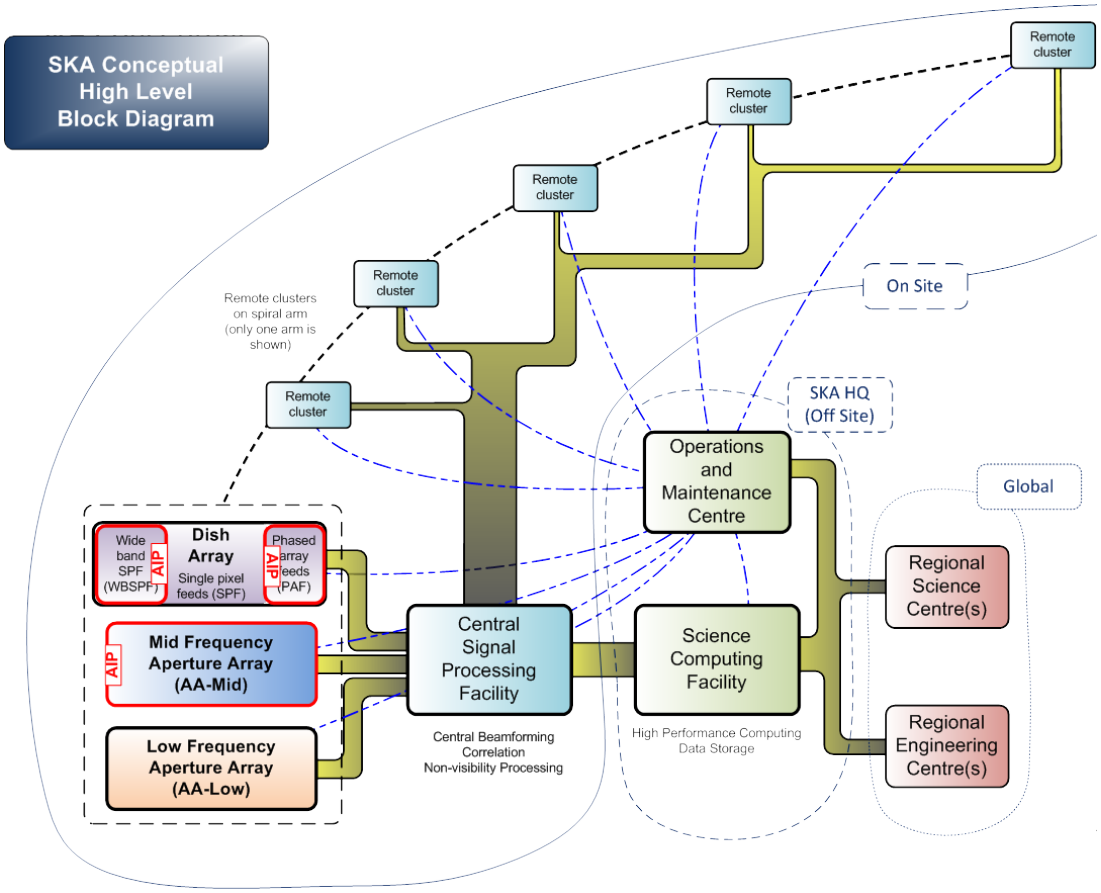


Figure 1.4: SKA high-level conceptual block diagram, prior to dual-site decsion, showing receptors linked to the centrally located facilities. The receptor technologies in the Advanced Instrumentation Program are labelled AIP. Source: Schilizzi et al. (2011).

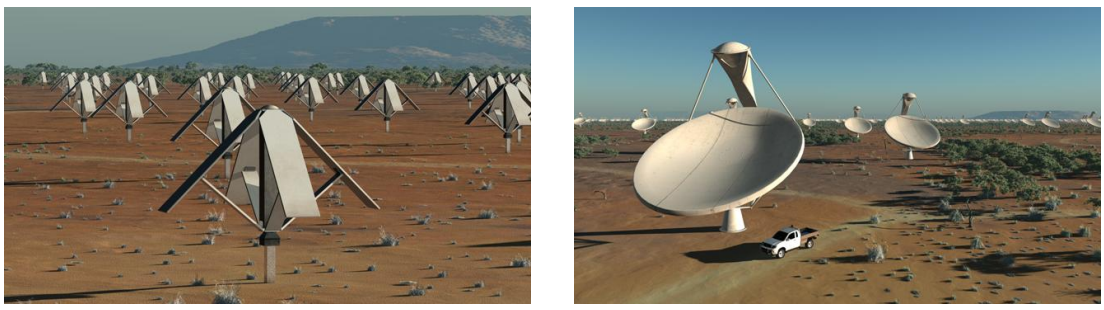


Figure 1.5: Artist impressions of the low-frequency aperture array antennas (left) and dishes with single-pixel feeds (right). Credit: SPDO/Swinburne Astronomy Productions and additionally TDP/DRAO (right).

Table 1.1: Current SKA timeline for SKA Phases 1 and 2. Sources: http://www.skatelescope.org/the-project/project and Cloete (2011).

SKA_1 activities		**SKA_2 activities**	
2008–12	System design and cost	2011–15	System definition and science and engineering trade-offs
2012	Site decision	2016	Advanced Instrumentation Program decision point
2013–15	Detailed design (pre-construction)	2016–20	Preliminary and detailed design
2016–20	Construction	2020–24	Construction
2020	Science operations	2024	Science operations

to observe the sky, only a relatively small fraction of the final number of antennas are required to create a functioning telescope (albeit with lower performance). Indeed, SKA_1 will be a powerful telescope in its own right, rivalling or exceeding the capability of other instruments (see Figures 1.2 and 1.3). The advantages of phased construction is two-fold: it recognises interest from the SKA community in conducting early scientific observations, and allows scientific and technological developments to indicate a path to the full SKA design (Schilizzi et al., 2007).

1.1.4 Contributions from pathfinding projects

In addition to a phased development approach, the scientific and technical contributions from existing radio astronomy projects are integral to the SKA design process. These are projects officially designated[2] as a pathfinder, precursor or design study, where: a pathfinder is a technology, science or operations activity related to the SKA; a precursor is a 'pathfinder' telescope on an SKA site; and a design study is a study or prototype construction of one or more SKA sub-systems (Schilizzi et al., 2011).

The precursors, pathfinders and design studies generate a wealth of knowledge and data that is directly relevant for the SKA (e.g. Hall et al., 2008; Schilizzi et al., 2011), including:

- new observing techniques and scientific results
- verification data for key technologies and sub-systems
- insight into infrastructure, capital and operations costs
- better understanding of non-technical (e.g. project management) factors that affect cost and schedule.

Of particular relevance to this thesis are the three SKA precursors and the LOFAR pathfinder telescope. Table 1.2 briefly describes these telescopes and identifies the related SKA_1 receptor technologies.

Although the collecting area (effective area) of each of these telescopes is, at most, a few percent of SKA_2, they will still probe new regions of parameter space compared to existing radio telescopes (see Figure 1.2). Furthermore, many of the scientific goals are aligned with

[2] See http://www.skatelescope.org/the-organisation/precursors-pathfinders-design-studies for those projects approved by the SKA Science and Engineering Committee (now disbanded).

Table 1.2: Summary of selected SKA precursor and pathfinder telescopes, showing the related SKA receptor technologies.

Telescope and location	Brief design description	SKA receptor technology
ASKAP (Australian SKA site)	36 dishes of 12 m diameter, equipped with PAFs to observe frequencies between 0.7 and 1.8 GHz with a wide (30 deg^2) FoV (DeBoer et al., 2009). The first 6 of the PAFs will be installed by the end of 2012[a].	Dish array with PAFs
LOFAR (The Netherlands and Europe)	48 dual-band aperture array 'stations' of diameter ∼70–90 m (30–80 MHz) and ∼30–60 m (120–240 MHz)[b]. As a 'software telescope', the station digital signal processing configuration can be tailored to suit the astronomical application. LOFAR is commencing regular proposal observing in 2012–13[c].	Low-frequency aperture arrays
MeerKAT (South African SKA site)	64 dishes of 13.5 m diameter, initially equipped (by 2016) with a cryogenically cooled single-pixel feed (0.9–1.67 GHz). A second phase adds two more feeds to achieve 0.58–14.5 GHz frequency coverage by 2018[d].	Dish array with single-pixel feeds
MWA (Australian SKA site)	Low-frequency (80–300 MHz) telescope, composed of 128 aperture array 'tiles' as inputs to the correlator, where each tile consists of 16 dual-polarisation vertical bow-tie dipoles. Science operations are to commence in 2013 (Lonsdale et al., 2009; Tingay et al., 2012).	Low-frequency aperture arrays

[a] http://www.atnf.csiro.au/projects/askap/timeline.html.
[b] LOFAR has 'core', 'remote and 'international' stations, each with different antenna element layouts and digital signal processing configurations, see http://www.astron.nl/radio-observatory/astronomers/users/technical-information/lofar-stations.
[c] http://www.astron.nl/radio-observatory/astronomers/asking-time/asking-time-0.
[d] http://public.ska.ac.za/meerkat.

those of the SKA. For example, one of the major science goals for SKA_1 is "understanding the history and role of neutral hydrogen in the Universe from the dark ages to the present-day" (Garrett et al., 2010), encompassing *galaxy evolution and cosmology* (KSP IV) and *probing the dark ages* (KSP V). At low frequencies, LOFAR and the MWA will inform the requirements for power spectrum measurements of the EoR (SSWG, 2012). Meanwhile, deep and wide surveys of neutral hydrogen gas are key science goals of MeerKAT[3] and ASKAP (Duffy et al., 2012; Johnston et al., 2008).

Each of the telescopes in Table 1.2 will be crucial for testing and verifying the receptor technologies for SKA_1. Importantly, these telescopes are variants of the 'large-N, small-D' approach for the SKA, where the signals from a large number of small-diameter receptors are input to the correlator (Lonsdale & Cappallo, 1999). A larger number of receptors improves the spatial sampling of the sky, while the small-diameter receptor increases FoV to maximise the survey

[3] http://public.ska.ac.za/meerkat/meerkat-large-survey-projects

speed of the telescope. These projects are not only developing and testing the receptor technologies, but also the signal processing technologies required to transport and correlate the data from the receptors, and to make widefield images.

Further to the various technical information available from all the pathfinding projects, the precursor telescopes (ASKAP, MeerKAT and MWA) will accumulate site-related data. This data is pertinent to the SKA because these precursors demonstrate the logistical and infrastructure challenges and costs of constructing complex systems on remote sites (Schilizzi et al., 2011; Tingay et al., 2012). The cost of operating these precursors will also provide important data for the SKA (Kellermann et al., 2006).

While the pathfinding projects are essential to the SKA, they bring with them an additional set of challenges. Not only do these precursors inform the SKA design process, but the ASKAP and MeerKAT arrays are planned to be incorporated into the SKA_1 arrays[4]. The SKA dish design is likely to differ from both the ASKAP and MeerKAT dish designs, so incorporating those existing dishes into the SKA may increase increase the complexity of the system. Even if the ASKAP and MeerKAT antennas are not themselves re-purposed, the infrastructure developed for the precursors, such as power and fibre connections and reticulation, is valuable to the SKA (ISPO, 2007). The organisations involved in the pathfinding projects are also likely to play key roles in the SKA_1 pre-construction stage (e.g. Cloete, 2012), further influencing the SKA design in a somewhat more abstract way.

1.2 Thesis outline

The remainder of this chapter describes the motivation and contributions of this thesis. Overall, the thesis is structured as follows:

Chapter 2: Exploring the radio Universe with the SKA

Chapter 2 reviews current radio astronomy metrics and SKA sub-systems, which I present in the context of regions of observational parameter space (also known as phase space) available to the telescope.

Chapter 3: SKA performance and cost exploration

Chapter 3 reviews the systems engineering process for the SKA, specifically the processes for requirements analysis and trade-off studies. This chapter also overviews parametric performance and cost modelling, and describes SKACost as a tool for this modelling.

Chapter 4: SKA as a fast radio transients telescope

Here I present the first of three case studies of SKA design challeges, in which I consider the scientific implications of the system design on high time resolution searches for fast transients. I develop metrics to quantify the cost-effectiveness of fast transient searches, and apply these to SKA_1 (although they are equally applicable to other radio telescope arrays).

[4] http://www.skatelescope.org/news/dual-site-agreed-square-kilometre-array-telescope

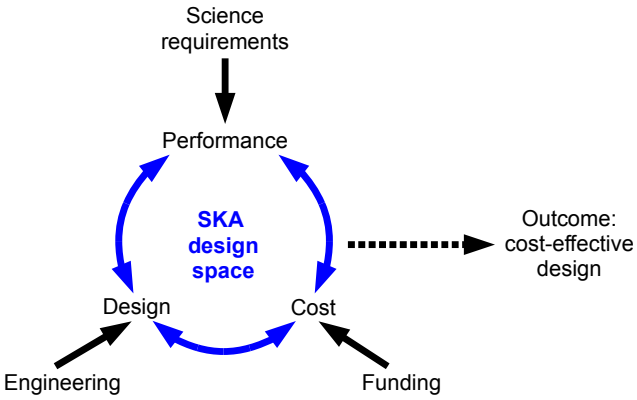


Figure 1.6: Overview of the SKA system design process, where iterative science and engineering trade-offs are made in a cost-constrained environment.

Chapter 5: Single versus dual-band SKA-low

The second case study makes a comparative analysis of single and dual-band low-frequency aperture arrays for SKA_1 (SKA_1-low). I also investigate the cost implications of different intra-station signal transport and processing architectures.

Chapter 6: The high angular resolution SKA

The third case study develops a modelling framework for the data network for the long baselines of the SKA_2 dish array, and applies this framework to an exemplar network architecture. I investigate trades between scientific performance and the design parameters that drive the cost of the network.

Chapter 7: Conclusions

Chapter 7 discusses the insights gained from the parametric performance and cost modelling used in the three case studies, and places this work in the broader context of developing cost-effective SKA designs.

1.3 Thesis motivation

In this thesis, I apply a systems engineering process to develop insight into cost-effective designs for the SKA. Systems engineering is a well-established, structured process for designing and implementing large and complex projects; the system is taken as a collection of components whose functionality is defined not only by the individual components (sub-systems), but also their interconnection. The case studies in this thesis show the utility of parametric performance and cost modelling of the telescope in determining cost-effective design solutions.

Section 3.2 outlines the stages of the systems engineering approach for the SKA project, from conception and definition, through to detailed design, construction and operation. The current system definition stage is a relatively early stage in the systems engineering context. Activities in this stage are conceptualised in Figure 1.6. An essential task of this stage is to develop of a complete set of requirements, and validate these requirements to ensure that they will

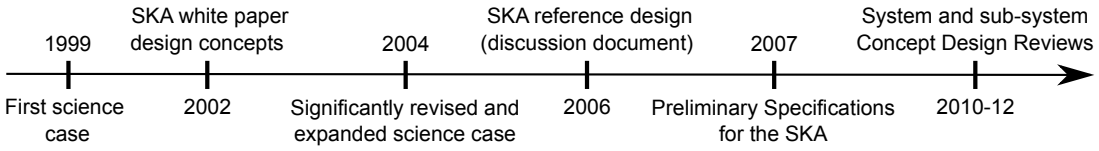


Figure 1.7: A summary timeline of previous concept definition activities for the SKA.

provide the expected functionality and performance for stakeholders (primarily astronomers). As a concurrent and complementary task, trade-off studies aim to develop a telescope design solution that meets these requirements to the maximum extent possible, and which can be carried through to the next stage. The set of requirements developed in this definition stage is pivotal to the project, because it defines the subsequent systems engineering activities that lead to the detailed telescope design, and its construction and operation (Stevenson, 2011c).

Figure 1.7 summarises previous, less formal, concept definition activities for the SKA that have narrowed down the design options to those described in this chapter. These activities are, in effect, iterations of Figure 1.6. However, there is much more work to be done to develop a telescope design solution for the upcoming detailed design (pre-construction) stage. Such work is complex, and I consider the principal causes of this complexity to be:

- an extremely large design space, due to both the inherent potential scalability of modern aperture synthesis radio telescopes, and the continually evolving capabilities of digital technologies
- changing scientific requirements, because of new and improved understanding of the science and predicted telescope performance, and emerging results from other telescopes
- sub-system concept designs that require order-of-magnitude performance improvements over existing realisations
- the multitude of interfaces between sub-systems, where changes to one sub-system due to the evolving requirements and capabilities propagate to other sub-systems.

These complexities prevent a straightforward analysis of requirements and concurrent trade-off studies. Rather, the process of developing cost-effective designs involves many iterative steps, where each iteration re-analyses results from the previous step, incorporating new scientific and technical data.

The methods for finding cost-effective design solutions depend on the depth of previous iterations. Less mature problems, such as cost-effective searches for fast transients, generally involve understanding and formalising the scientific requirements in the context of a high-level telescope design concept. As the requirements and design solutions mature, the work progresses to more detailed trade-offs and system optimisations. For some trades, I have used SKACost, a custom-developed system performance and cost modelling software tool; I am one of the SKACost developers (see Chippendale et al., 2007; Ford et al., 2009, 2010). SKACost uses parametric models of the system to enable high-level trade-off investigations of telescope performance, cost, design architectures and key design parameters (see Section 3.5). I use SKACost for the uncertainty analysis in the low-frequency aperture array study (Chapter 5) and the multi-dimensional cost and performance trades of the data network for the long baselines

(Chapter 6). Where the trade-offs and optimisations in this thesis are not sufficiently complex to warrant the use of SKACost, I have developed the models such that they are transferable to SKACost or another system modelling tool as required.

1.3.1 Case studies on developing cost-effective design solutions

I have chosen to investigate, as three scientifically interesting case studies, areas of the baseline SKA design that open up new parameter space at high time resolution, low frequencies and high angular resolution. In this context, key technologies in the baseline design are not only the receptors, but also cutting-edge signal transport and processing technologies, and high performance computing hardware and software. The specific technical challenges of these topics are outlined below.

1.3.1.1 High time resolution observations (Chapter 4)

Most of the technical challenges for fast transient searches stem from the extremely high data rates required for high time resolution observations. Each of the options to reduce the data rate inhibits the effectiveness of the observations, mainly by limiting FoV, sensitivity, time resolution or a combination of these. Unfortunately, fast imaging modes that retain both FoV and sensitivity (see Figure 1.3) are beyond the capability of current and near-future technologies. However, if searches are conducted in real-time, an innovative solution to retain the maximum FoV, sensitivity and time resolution is to temporarily store the digitised antenna voltages in a data buffer (e.g. Ekers, 2003; Stappers et al., 2011). Triggers for dumping the data from this buffer could be less sensitive observations, or alternatively, monitoring from a telescope at a wavelength outside of the radio band. The buffer also enables high angular resolution post-observation analysis (e.g. imaging) for event verification; such follow-up capability is a key feature of experiments such as V-FASTR (Wayth et al., 2011).

Despite the high time resolution Universe being relatively poorly explored, transient detection is not a key science goal for SKA_1, and "will be carried out only if it can be done with minimum additional cost or effort" (Dewdney et al., 2010a). So how should scientific return be maximised within these cost constraints? In Chapter 4, I use a new metric, 'event rate per beam formed and searched', to generalise the problem and parametrise the cost-effectiveness of a search strategy. Both the probability of intercept shown in Figure 1.3 and the event rate per beam vary with the frequency of observation, choice of receptor (antenna), signal combination mode (sensitivity and field of view) and the observed sky direction. Chapter 4 investigates how these factors impact the strategy for cost-effective fast transient searches with SKA_1.

1.3.1.2 Low-frequency aperture arrays (Chapter 5)

Single-dish telescopes and aperture synthesis dish arrays have been the dominant type of radio telescope in recent decades; these reflector telescopes are efficient instruments to spatially sample the incoming radiation in a phase coherent manner (Bregman, 2000). Recently though, there has been a resurgent interest in low frequency ($\sim$10–300 MHz) radio astronomy with aperture arrays. For example, low frequency arrays have the potential to observe the epoch of re-ionisation and significantly increase our understanding of the early history of the Universe (e.g. Furlanetto et al., 2006; Mellema et al., 2012). Detection of the faint EoR signals

requires digital advances that enable radio astronomers to calibrate and correct the instrument response, and filter out (subtract) foreground sources (see Greenhill & Bernardi, 2012, and references therein). Moreover, the ability to digitally form multiple independent fields of view makes aperture arrays a powerful tool for high time resolution observations, both for searches and monitoring of pulsars, as well as intermittent and single-pulse events (e.g. Stappers et al., 2011, see also Chapter 4).

For aperture arrays, a frequency of approximately 450 MHz has been identified as a natural breakpoint between the sky-noise dominated lower frequencies and the system-noise dominated higher frequencies (Schilizzi et al., 2007). Meanwhile, frequencies of 70 MHz or lower are required for EoR observations (Mellema et al., 2012; SSWG, 2012). Recent work in the SKA Design Studies (SKADS) program proposes a single wideband antenna element design to observe this 70–450 MHz frequency range with the SKA (Faulkner et al., 2010). However, an alternative is to use two antenna element designs, each observing approximately half the fractional bandwidth. I term these 'single-band implementation' and 'dual-band implementation' respectively. Although there have been qualitative technical analyses of the single and dual-band approaches (e.g. Bij de Vaate et al., 2011), there has been little quantitative comparison of the cost and performance of the approaches; I make this comparison in Chapter 5.

1.3.1.3 The high angular resolution SKA (Chapter 6)

Very Long Baseline Interferometry (VLBI) observations have been conducted over many decades, where the data from each antenna in the interferometer is recorded to a storage device (e.g. magnetic tape or hard disk drive) and correlated post-observation (Thompson et al., 2001). In recent years, digital advances have allowed for near real-time 'electronic VLBI' (e-VLBI) observations, where data from the VLBI telescopes are transferred to a central correlator over a fibre optic network (e.g. van Langevelde, 2010). While e-VLBI arrays are created on an ad-hoc basis, the SKA_2 dish array will have dedicated connections to the remote stations (25 groups of dishes on the long baselines, $\gtrsim 180$ km from the core; 'remote clusters' in Figure 1.4). The technical challenge is to implement these data connections, which are more numerous and of higher bandwidth than e-VLBI, in a cost-effective manner.

There has been little investigation of the cost-effectiveness of the fibre network for the long baselines, partly due to the absence of simple performance metrics for high angular resolution observations. Although existing VLBI instruments such as the Very Long Baseline Array (VLBA) allow observations with higher angular resolution, observations with the long baselines of the SKA_2 dish array will have significantly increased sensitivity and also improved (u, v) coverage (spatial sampling of the sky). In Chapter 6, I use a common set of metrics to model both the cost of a fibre network architecture for the long baselines, and the telescope performance as influenced by the chosen network design. With this modelling framework, I consider trade-offs between the cost of the fibre network and the scientific performance of various network design solutions.

1.4 Thesis contributions

In this thesis, I demonstrate that parametric performance and cost modelling can provide insight into cost-effective design solutions, even with limited performance and cost data. The key features of my approach are to:

- selectively define trade studies within the system context
- create scalable parametric performance and cost models of alternative designs
- when only limited data is available, make differential cost analyses and comparative performance analyses of alternative design solutions.

Chapter 3 presents the structure and formalisms for the parametric performance and cost trades.

The case studies in this thesis further develop the modelling framework to suit each design problem, and provide new insights for cost-effective SKA designs, described as follows:

- Chapter 4 demonstrates that both the aperture array and dish arrays of the SKA will be highly effective for fast transient searches. With appropriate planning in the design stage, a low-cost system for fast transient searches can be implemented on SKA_1 that will be sensitive to a large region of new parameter space, and have powerful event localisation and verification capabilities. Chapter 4 also finds that frequency-dependent effects have considerable impact on preferred receptor and frequency range; therefore the exploration of high time resolution parameter space is maximised by having flexible search modes that can be adjusted to suit the observing line of sight and expected source population. Juxtaposing these findings, the models in Chapter 4 also serve to predict the parameter space filtering effect of particular telescope architectures and observing strategies. These results are applicable to future planning for fast transient search systems for the SKA and other radio telescope arrays.
- Chapter 5 shows that despite a dual-band array having twice the number of antenna elements, neither a representative single or dual-band SKA_1-low sparse aperture array design solution is cheaper *a priori*; the single-band implementation is more sensitive to high digital signal processing costs, while the dual-band implementation is more sensitive to changes in costs that scale with the number of signal paths, such as those of the active antenna element hardware and deployment. In both implementations, the cost reduction due to a first stage of analogue tile beamforming is potentially significant. These results are of particular relevance to the aperture array work in the SKA_1 pre-construction phase (SKA Drafting Groups, 2012). Chapter 5 highlights that the product of processed FoV and instantaneous bandwidth (beam–bandwidth product) is a major cost-driving parameter for SKA_1-low that requires further scientific consideration. The central processing, antenna deployment and site preparation costs are also potentially significant cost drivers that have so far had insufficient attention.
- Chapter 6 develops a modelling framework for the data network for the long baselines of the SKA_2 dish array. The framework uses the number of remote stations, maximum baseline and the beam–bandwidth product of a remote station as both measures of scientific performance and design parameters that drive the cost of the network. The results for the exemplar network architecture show that significant cost reductions are achievable

by matching the scientific requirements to the available design solutions. For example, limiting the beam–bandwidth product reduces cost markedly, without significantly impacting scientific performance. Additional to these preliminary results, the modelling framework will be useful for future performance–cost trades of the long baseline network.

The case studies in this thesis show that an analytical approach to requirements analysis and performance–cost modelling gives new insights to cost-effective SKA designs. The complexities of the SKA design environment demand pragmatic choices to narrow the exploration space and relate even the top-level scientific requirements to the telescope design and cost. Thus, relatively simple models can produce insight. These models can be used to make significant trade-offs and optimisations, with results that potentially save many millions of euros. With this approach, the results and insights are more rapidly available for the next iteration of the systems analysis, and the foundations exist for further trade-offs and optimisations.

1.5 Contributions of others

The work in this thesis was carried out between 2008 and 2012, mainly at Curtin University but including visits to the UK: one month in 2008 at the Astrophysics Group, University of Cambridge and two months in 2010 at the SKA Program Development Office (SPDO), University of Manchester. I have developed the ideas for this thesis in a collaborative environment, through interaction with colleagues at Curtin University, as well those at the CSIRO Australia Telescope National Facility (ATNF), University of Cambridge and ASTRON, The Netherlands.

Appendix A provides a full list publications related to this thesis. Specific contributions by others are as follows:

- The generic fast transient search pipeline in Chapter 4 was originally developed for the Commensal Real-time ASKAP Fast Transients (CRAFT) survey by me, and other members of the collaboration (Dodson and Macquart et. al., 2009, unpublished; Macquart et al., 2010b); similar ideas were developed contemporaneously for LOFAR (e.g. Hessels et al., 2009). Chapter 4 builds on these concepts and applies them to the event rate formalism developed by Macquart (2011); however, the 'event rate per beam' metric and frequency dependent event rates is my own work. The modelling re-uses some software code written by Aaron Chippendale (the Parameter Class, see Appendix D of Chippendale et al., 2007). The majority of Chapter 4 was published as Colegate & Clarke (2011), to which Nathan Clarke contributed the material incorporated as Appendix E.1 and Appendix E.3 in this thesis. The application to a giant pulse population and the strawman transients system is my additional work developed in the Curtin University milieu and published in Hall et al. (2012).
- The high-level performance and cost analysis of the single vs. dual-band implementation for SKA_1-low (Chapter 5) is original work conducted by me, with secondary input from Peter Hall and Andre Gunst, that draws on discussions with Jan Geralt Bij de Vaate, Andrew Faulkner, Aziz Jiwani, and other Curtin University, ASTRON, SPDO and University of Cambridge colleagues. Chapter 5 has been published in the SKA Memo series (Colegate et al., 2012), with only minor changes for consistency.

Chapter 2

Exploring the radio Universe with the SKA

In this chapter, I quantify the capability of a given telescope design to meet the stated scientific goals, and to probe beyond them for future applications such as observing new astronomical phenomena. To describe the radio signals expected from astronomical phenomena, I introduce the concept of multi-dimensional parameter space, where the radio telescope array is regarded as a filter, or window function, on this parameter space. The regions of parameter space to be sampled by the SKA can be determined by top-level measures of performance (metrics and figures of merit). Because they characterise the high-level capability of the telescope, measures of performance are also essential for parametric modelling of the system and comparison of telescope designs. This chapter appraises the scientific and technical concepts and trends most relevant to the SKA performance and cost modelling framework (Chapter 3) and the trade studies in the later chapters.

Section 2.1 introduces the parameter space concept. Section 2.2 characterises the telescope by how it filters the signal parameter space, using common radio telescope measures of performance. Section 2.3 relates these measures of performance to the fundamentals of radio telescopes and the individual sub-systems in the SKA signal path. Lastly, Section 2.4 gives a brief overview of SKA system details.

2.1 Parameter space

The scientific effectiveness of the SKA to discover and analyse astronomical phenomena depends on the capability of the telescope. The SKA is requisite because these phenomena are either not observable with existing radio telescopes, or their observation would require an impractical amount of telescope time. In his book *Cosmic Discovery*, Harwit (1981) identifies novel observational approaches as being a principal driver for discovery of astronomical phenomena. These approaches are usually a product of new technologies or improvements in the capability of existing technologies by several orders of magnitude. To measure the potential for discovery through improved capability, Harwit (1981) presents the concept of parameter space, which is an elegant description of the relationship between the astronomical phenomena and the telescope observing them. Different phenomena are distinguishable by the regions of a multi-dimensional parameter space that they occupy. Therefore, the phenomena that the SKA will discover and analyse depends on the regions of this parameter space that are observed.

2.1.1 Properties of electromagnetic radiation

The information in the electromagnetic spectrum can be described using only a few parameters, forming a multi-dimensional parameter space or phase space (Harwit, 1981). Photons emitted

Table 2.1: Dimensions of electromagnetic radiation and their relationship to observational parameter space. Adapted from Harwit (1981).

Properties of photons	**Parameter space axes**
3 spatial coordinates	None[a]
Frequency coordinate	Spectral frequency ν (or wavelength λ)
Time coordinate	$\begin{cases} \text{Temporal resolution } \Delta t \\ \text{Spectral resolution } \Delta\nu_{\text{ch}}\ ^{\text{b}} \end{cases}$
Direction of motion (2 angular coordinates)	Spatial (angular) resolution ϕ
Spin parameter	Plane or circular polarisation (number of polarisations N_{pol})
Count of photons with indistinguishable properties	Spectral flux density S

[a] Assumes absolute time and location of individual events and sources is unimportant—see text for further discussion.
[b] Of a channelised signal.

by phenomena in the Universe can be described by unique points in this phase space, which I refer to as observational parameter space. Table 2.1 lists eight properties of photons, and their mapping to axes of parameter space. At time t_0, a photon of a specific wavelength is at a specific position. Its movement is in a specific direction and it has a specific polarisation. But five parameters are sufficient to describe the radiation; Harwit (1981) argues that the density of sources and events in the Universe remains approximately constant over human timescales (e.g. < 1 million years), therefore different types of phenomena can be distinguished without detailed knowledge of the absolute time and location of individual events and sources. These parameters form the orthogonal axes of a five-dimensional parameter space.

A sixth parameter in Table 2.1 is the count of photons with indistinguishable properties. It is listed separately, because it is not the property of a single photon. Rather, it describes the accumulation of photons (information) with a set of properties described by the other five parameters in Table 2.1. A common metric in radio astronomy to measure these photons is spectral flux density S, which is the flux density per unit bandwidth. It has units of Janskys (Jy), where $1\,\text{Jy} = 10^{-26}\,\text{W}\,\text{m}^{-2}\,\text{Hz}^{-1}$. Following convention, in this thesis I simply refer to S as flux density. Flux density has a dual role in parameter space: it can be used to distinguish different phenomena or increase the rate of information observed by the telescope. Section 2.2.1 discusses this dual role.

As an aside, other properties of photons potentially play a role for discovery of new astronomical phenomena. For example, there may be photon orbital angular momentum induced in astronomical signals, such as when maser emission passes through inhomogeneities in the interstellar medium (Harwit, 2003b). Interestingly, low-frequency ($<$1 GHz) antenna array designs capable of measuring photon orbital angular momentum have been modelled (Thidé et al., 2007). However, these do not feature in present SKA designs or planning documents, and I do not consider such designs in this thesis.

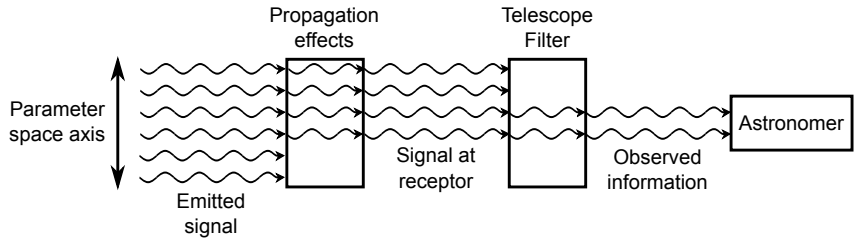


Figure 2.1: Telescope as a filter on an axis of parameter space. Adapted from Harwit (1981).

2.1.2 Telescopes as parameter space filters

A telescope will only observe an astrophysical signal if the telescope's capabilities are sufficiently well-matched to the properties of the incoming photons. The telescope can be thought of as a matched filter on the signal (Harwit, 1981), a mapping (Cordes & SWG, 2006) or a window function (Djorgovski et al., 2012). Furthermore, there may be propagation effects that transform the signal emitted from the source, forming an additional filter on the signal (Cordes & SWG, 2006). An example of this is the temporal broadening (scatter broadening) of an impulsive signal due to multipath propagation in the interstellar medium discussed in Chapter 4. Figure 2.1 shows, for a single dimension of parameter space, how these filters conceptually apply to a signal emitted from the source.

As represented in Figure 2.1, the astrophysical signal can occupy a range of values along each axis of observational parameter space. The signal can thus be described by a hypervolume, a multi-dimensional volume of parameter space. Similarly, the telescope is only sensitive to some fraction of each axis of parameter space. The capabilities of a telescope can thus be described as another hypervolume in the same observational parameter space; the telescope cannot detect incoming radiation with characteristics outside this hypervolume. For example, hypervolume of SKA-low along the frequency axis is restricted to its frequency range of 70–450 MHz. Therefore, the intersection of the signal and telescope hypervolumes describes the scientific effectiveness of the telescope.

2.1.3 Discoveries of new phenomena

A central thesis of Harwit (1981) is that new astronomical phenomena are discovered soon after access becomes available to new parameter space. Existing radio telescopes have sampled much of the available temporal, spectral and angular resolution axes of parameter space (Wilkinson et al., 2004). But these axes have not necessarily been sampled simultaneously; when appropriately combined with the greatly increased sensitivity and a large FoV for surveying, the SKA will be able to probe as-yet unexplored sub-volumes of parameter space, along with new volumes of parameter space (Cordes & SWG, 2006).

The challenge for astronomy is to design a telescope to sample the unexplored sub-volumes of parameter space in a cost-effective manner (Harwit, 1998). Herein lies the system design challenge for the SKA: which sub-volumes of parameter space should the telescope be designed to observe? The requirements from the science case provide some guidance (see Section 3.3.1), but are built on specific known or postulated phenomena. There are also bounding surfaces on

the hypervolumes of parameter space available to the telescope; these are due to known physical limits on the source emission mechanism and intervening media filtering out certain regions of parameter space (see Harwit, 1981, 2003a).

For some years now, accessing unexplored parameter space to discover new phenomena has been advocated for the SKA[1]. To identify specific sub-volumes of unexplored parameter space, Wilkinson (2007) suggests the morphological analysis approach of Zwicky (1948, 1957); the phenomenological classification of Harwit (1981) could also be useful. However, such an analysis would require a detailed review of past astronomical observations and discoveries. Hence the design requirements for the exploration of the unknown have not progressed much beyond the guiding principles in Wilkinson et al. (2004), being to:

- maximise flexibility to allow high sensitivity to be combined with high time, spatial and spectral resolution observations, or at least enable trade-offs between these
- re-use expensive components
- plan a modular system that allows for upgrades to exploit new digital technologies
- time-buffer and archive as much raw data as possible.

The 'flexibility' advocated by Wilkinson et al. (2004) describes the telescope's capability to maximise simultaneous access to the parameter space available to the SKA. However, any design change to promote this flexibility via improved capability has an associated cost. Some costs are relatively small, such as a fast transients search processor (Section 4.7.1). Other costs are significant, such as the processing required to image the full field of view of dishes on long baselines (Chapter 6). Furthermore, capability can evolve. In particular, rapid technology advances in digital signal processing and computing reduces the future cost of technology-dependent capability. However, I do not rigourously treat the cost of capability evolution in this thesis, because such treatment would require detailed analysis of the technologies, budget and schedule.

Although the SKA broadly aims to excel in all of axes of parameter space, current signal processing and computing limitations means it will not be able to excel in all of these axes simultaneously. Proposals to increase capability to improve access to new parameter space require assessment of their cost-effectiveness. There are well-defined procedures for cost estimation (see Section 3.4.1), whereas quantifying effectiveness is more difficult. The concept of parameter space provides a useful top-level view of telescope performance, hence effectiveness. Although it does not detail all the requirements to design a telescope, the parameter space approach complements, rather than replaces, the analysis of the scientific requirements; the approach enables high-level cost and performance exploration at an earlier stage in the iterative science–engineering trade-off process.

2.2 Performance metrics and figures of merit

The regions of parameter space that a telescope samples can be measured by top-level performance metrics and figures of merit (FoMs). These metrics and figures of merit describe the

[1] e.g. Braun (1996); Ekers (2009); Kellermann et al. (2009); Wilkinson (2007); Wilkinson et al. (2004).

scientific performance of the telescope in broad terms, so they are useful for modelling telescope effectiveness. For example, the top-level measures of performance used in the development of the *Preliminary specifications for the Square Kilometre Array* (Schilizzi et al., 2007) are:

- sensitivity (A/T) as a function of frequency
- accessible FoV as a function of frequency and baseline length
- processed FoV as a function of baseline length
- survey speed as a function of frequency
- angular resolution as a function of frequency
- polarisation purity
- total bandwidth and spectral resolution
- imaging dynamic range.

Angular and spectral resolution and polarisation purity each map to an axis of observational parameter space (Table 2.1). The other measures of performance describe the capability to measure flux density or make multiple observations in a survey, as described below.

2.2.1 Flux density and sensitivity

As mentioned in Section 2.1.1, flux density has a dual role of distinguishing different phenomena, or increasing the rate of information observed by the telescope. Distinguishing phenomena requires measurement of the flux density of the different signals. A signal will only be detected if the telescope is sensitive enough, such that the flux density S of the signal is greater than the minimum detectable flux density $S_{\min}$ of the telescope. As derived in Appendix B, a radio telescope detects a signal if it is statistically significant above Gaussian noise: a level n-sigma (σ) above the root mean square (rms) variation in measured flux density S is chosen. The minimum detectable flux density is given by

$$\begin{aligned} S_{\min} &= \sigma S_{\mathrm{rms}} \\ &= \frac{\sigma 2kT_{\mathrm{sys}}}{A_{\mathrm{e}}\sqrt{N_{\mathrm{pol}}\Delta\nu\tau}} \quad \mathrm{Jy}, \end{aligned} \tag{2.1}$$

where $k = 1.38 \times 10^{-23}\,\mathrm{J\,K^{-1}}$ is Boltzmann's constant, T_{sys} is system temperature, A_{e} is telescope effective area (see Section 2.3.1), N_{pol} is the number of oppositely polarised signals detected, $\Delta\nu$ is bandwidth and τ is post-detection integration or averaging time.

Equation 2.1 only holds true if the telescope filter matches the signal at the receptor for the parameter space axes of temporal resolution (τ), spectral resolution ($\Delta\nu$) and polarisation (N_{pol}); this matching was shown conceptually in Figure 2.1. For example, if an impulsive signal is observed with many hours integration time, the telescope filter does not match the incoming signal and the observation is insensitive to the impulsive signal. Conversely, a long integration time does match the signal from a steady source. Similar arguments apply to the spectral resolution and polarisation.

When the telescope filter does match the signal at the receptor, the sensitivity FoM, A/T,

describes the sensitivity of a telescope to a point source. Derived from Equation 2.1,

$$A/T = \frac{A_\mathrm{e}}{T_\mathrm{sys}} \quad \mathrm{m^2K^{-1}}. \tag{2.2}$$

The sensitivity FoM is a measure of the telescope collecting area and telescope noise contributions. Because it excludes noise reduction due to integration (averaging) time, bandwidth and summing polarisations, it is simply proportional to the inverse of system equivalent flux density (see Appendix B).

The calculation of A/T from S_min requires knowledge of integration time, processed bandwidth and the number of polarisations. These factors can be constrained by the expected emission, as described above. They may also be constrained by the telescope and its operation. For example, calculation of A/T for surveys with the SKA use a standard 2 year observing time per survey (SSWG, 2009, 2012). This limits the integration time for each pointing in the survey, hence to achieve a given S_min, sufficient A/T is required.

There are other figures of merit, such as surface brightness sensitivity (Johnston & Gray, 2006) to describe the sensitivity to extended sources (those larger than the FoV); these are less relevant to the case studies in this thesis and are not considered here.

2.2.2 Survey parameter space

Astronomical surveys are generally designed to search a large area or volume of sky for a particular class of phenomena. As mentioned in Chapter 1, discoveries from large surveys are likely to be made using the statistical properties of a large population, or through the recognition of a rare, previously unknown class of object. The key to detecting a large population or a rare astrophysical signal is to increase the rate of information gathered by the telescope (Harwit, 1981, 1998). The properties of the information gathered is also important; the phenomena to which the survey will be sensitive is defined by the region (hypervolume) of parameter space that the telescope sees.

2.2.2.1 Accessible and processed FoV

The accessible field of view describes the maximum solid angle (area) of sky instantaneously available at the receptor. The processed FoV Ω_proc describes the area of sky instantaneously available to be imaged or otherwise observed, which may be less than the accessible FoV to reduce processing costs (Schilizzi et al., 2007). In aperture synthesis arrays, the processed FoV available to the astronomer is determined by the most restrictive part of the elemental signal path, which may be the receptor or some aspect of the signal processing. The processed FoV available at each part of the signal path is discussed in Section 2.3.

2.2.2.2 Survey speed and other figures of merit

The survey capability of the SKA is often described using a survey speed figure of merit (SSFoM), which describes the speed that the telescope surveys an *area* of sky to a given sensitivity:

$$\mathrm{SSFoM} = \Omega_\mathrm{proc}\Delta\nu \left(\frac{A_\mathrm{e}}{T_\mathrm{sys}}\right)^2, \tag{2.3}$$

although the bandwidth $\Delta\nu$ is not always applicable (e.g. in a survey for a narrow spectral line). This figure of merit encapsulates the capability of the telescope to make multiple observations of a particular class of signal, described by a particular hypervolume in observational parameter space. As shown in Appendix C.1, SSFoM is independent of integration time τ, thus is only suitable for steady sources or those that vary on timescales longer than the longest integration time in the survey.

If SSFoM is held constant, then the FoV, sensitivity and bandwidth of a telescope design can be traded independently of the sensitivity limit of the survey (such as Figure D.1 in Appendix D). However, SSFoM is somewhat limited in its application to astronomical phenomena, because (a) it assumes a steady source for the duration of the observation; and (b) the sky is observed to a fixed depth.

An alternative approach to quantifying survey effectiveness considers the *volume* of sky surveyed; this approach is important for surveying the high time resolution Universe (Chapter 4). The survey volume can be used to infer the number of sources the survey will detect, or expressed per unit time as a detection rate $\mathcal{R}_{\rm det}$. Appendix C.2 derives two different figures of merit for surveying a volume of sky. If the integration time τ remains constant, $\mathcal{R}_{\rm det}$ can be expressed as

$$\mathcal{R}_{\rm det} \propto \Omega_{\rm proc}\Delta\nu^{3/4}\left(\frac{A_{\rm e}}{T_{\rm sys}}\right)^{3/2}, \tag{2.4}$$

which is of the same form as Equation 2.3, but with different exponents. If τ can be traded for A/T, as is done to derive SSFoM (Appendix C.1), and maximum survey depth $r_{\rm max}$ and source luminosity $\mathcal{L}$ remain constant,

$$\mathcal{R}_{\rm det} \propto \Omega_{\rm proc}\Delta\nu\left(\frac{A_{\rm e}}{T_{\rm sys}}\right)^{2}. \tag{2.5}$$

Equation 2.5 still describes a volume surveyed, but it is now consistent with SSFoM (Equation C.5) because the sensitivity (hence depth) to which the volume is surveyed remains constant. When determining event rates for fast transient surveys, τ cannot be traded for sensitivity, so only Equation 2.4 applies. Equation 2.4 is the basis for the trade-offs and optimisations in Chapter 4, and is discussed further in that chapter.

These two detection rate figures of merit are not clearly differentiated in the literature (e.g. Cordes, 2009a,b). However, each of these figures of merit have different astrophysical consequences, or conversely, different implications for the design and operation of survey telescopes such as the SKA. In terms of comparing alternative telescope design solutions:

- $\Omega_{\rm proc} \propto (A/T)^{-2}$ for fixed SSFoM (or fixed $\mathcal{R}_{\rm det}$ to a constant survey depth, when τ is traded for A/T)
- $\Omega_{\rm proc} \propto (A/T)^{-3/2}$ for fixed $\mathcal{R}_{\rm det}$ and τ.

The first of these relationships was plotted on Figure 1.2 to show survey speed to constant depth, the second was plotted as the green lines on Figure 1.3 to show detection limits for extragalactic pulses.

These figures of merit provides a top-level estimate of the telescope's capability to survey

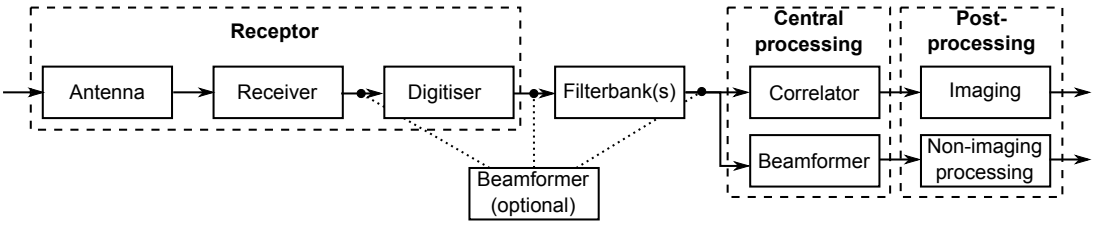


Figure 2.2: SKA elemental signal path through system. Adapted from Hall (2004a).

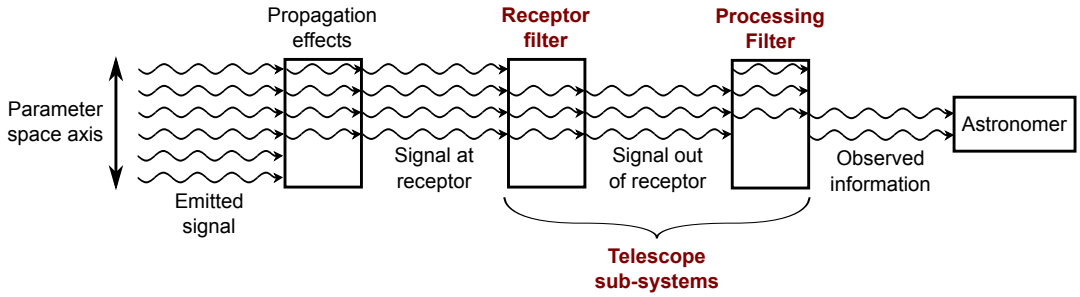


Figure 2.3: Telescope sub-systems (highlighted in red) as filters on an axis of parameter space.

the sky. The hypervolume of parameter space for a given survey (spectral resolution, angular resolution, and so on, see Table 2.1) completes the description of parameter space to be observed. Djorgovski et al. (2012) describes survey parameter space using concepts similar to Harwit's work, but in significantly more detail. While such additional details may assist to optimise the surveys to maximise the exploration of new parameter space, the metrics and figures of merit described in this section are sufficient for high-level performance and cost trades; they are easily relatable to the system design, as shown in the next section.

2.3 SKA system design and elemental signal path

This section has a dual purpose: it outlines the principal SKA sub-systems in the signal path and describes how these sub-systems act as parameter space filters on the incoming signal (Section 2.1.2). Figure 2.2 shows the principal sub-systems encountered by a signal arriving at the telescope. There is significant flexibility in how the individual sub-systems can be configured to make astronomical measurements, and the most suitable configuration often depends on the scientific application. But in general terms, the signal arrives at the antenna, and is received, processed and combined with signals from other antennas in real-time, so that the incoming signal is appropriately observed by the telescope. Post-processing with high-performance computing (HPC) is then applied to the observation. Throughout the signal path, information is transported in either digital or analogue form; the problem essentially becomes one of data transport and processing, which is a more tractable problem to non-astronomers (Hall, 2004b).

An additional viewpoint that I propose here extends the concept of the telescope as a parameter space filter (Section 2.1.2). The measurements of astronomical sources are made using an appropriate combination of antennas, receivers and digital signal processing units. Therefore, each sub-system in the signal path is actually a parameter space filter, because it can modify

the information measured by the telescope. Figure 2.3 shows this concept as an extension of Figure 2.1.

In this section I describe the role of these sub-systems as parameter space filters, using the metrics and figures of merit described earlier in this chapter. The sub-systems that I review in this section are those that I consider to be the principal filters on parameter space. These sub-systems are of particular relevance to later chapters. Other telescope sub-systems pertinent to particular cost and performance trades, but not directly on the signal path, are discussed in the relevant chapter.

2.3.1 Radio receptors

Following the trend of recent SKA documents, I use the term radio 'receptor' to describe the combination of one or more antennas and receivers, and accompanying signal processing; for example, a dish with single-pixel feed, or an aperture array 'station'. The signals from these receptors are then centrally combined with further signal processing to make the scientifically useful measurements of the sky. The technologies used in the receptors themselves define much of their capability; in this section I focus on the effective area and processed FoV of the receptor. As discussed in Chapter 1, low and mid-frequency aperture arrays (AAs) and dishes with one or more feed technologies are the receptors of interest for the SKA.

Before further detailing these receptors, I introduce the antenna, receiver, digitiser and beamformer as receptor sub-systems. I describe these sub-systems in general terms because their exact configuration depends on the intra-receptor architecture, especially for AAs (see Section 5.7.3).

2.3.1.1 Antenna, receiver and digitiser

The antenna and receiver of a receptor are matched to achieve optimum performance. The antenna is simply the part of the system that receives electromagnetic radiation (IEEE, 1993). The receiver generally describes the amplification, filtering and frequency conversion (where relevant) of the signal at the antenna; the digitiser is usually located at or near the receiver. The receiver contributes thermal noise to the observation (receiver noise T_{rec}); reducing this noise increases sensitivity, as per Equation 2.1. The antenna–receiver combination also restricts the range of frequencies to which the receptor is receptive. The digitiser may further restrict the instantaneously available processed bandwidth, however the SKA requirements specify a fractional instantaneous bandwidth of $\sim$1 (SSWG, 2012).

2.3.1.2 Receptor beamformer

Although not a new technology, beamforming is gaining a resurgence in radio astronomy by allowing more flexible observing capabilities and re-using collecting area with multiple beams (Hall et al., 2008). For aperture arrays, beamforming is also useful for reducing the intra-receptor and post-receptor signal processing requirements (Faulkner et al., 2010). A beamformer is the analogue (or radio frequency, RF) or digital signal processing unit that applies the correct phase (or time delay) and amplitude weight to form the beam. The beams are electronically

steered (scanned) by varying the phase or time delays. Multiple independently pointed beams are formed by using different phase or time delays for each beam (Mailloux, 1995).

2.3.1.3 Parabolic reflectors

The parabolic reflector is a structure that focuses radiation onto an antenna feed; dish is the generic terminology for a (usually steerable) reflector with a circular or elliptical paraboloid shape. An aperture efficiency factor η_{ap} describes how effectively the aperture collects radiation; it is usually given as a measured value. The aperture effective area $A_{\mathrm{e\,-\,0}}$ is a product of its physical size and η_{ap}. For a parabolic dish,

$$A_{\mathrm{e\,-\,dish}} = \frac{\pi}{4}\eta_{\mathrm{ap}} D^2_{\mathrm{dish}}, \tag{2.6}$$

where D_{dish} is the diameter of the dish. A higher aperture efficiency maximises the utility of the physical collecting area.

The antenna feed is also important. In light of the SKA field of view expansion technologies described below, the simple waveguide antenna feed used on most contemporary dish telescopes is termed a single-pixel feed (SPF). If a single feed has a frequency ratio of order 4:1 or higher, it is termed a wideband single-pixel feed (Dewdney et al., 2010b). When accompanied by appropriate receiving and signal processing systems, the wideband single-pixel feed enables a larger instantaneous bandwidth. For example, the 42 dish Allen Telescope Array (ATA) uses a log-periodic wideband single-pixel feed, designed to cover a 0.5–10 GHz frequency range (Welch et al., 2009).

Another generic term is 'beam', used to describe the power pattern of the antenna, receptor or telescope array. The half-power beamwidth θ_{HP} describes the angular width of the beam at the level of half the maximum power:

$$\theta_{\mathrm{HP}} = \frac{\mathcal{K}\lambda}{D}, \tag{2.7}$$

where $\mathcal{K}$ is the aperture illumination factor (also known as current grading or taper), λ is the wavelength of observation and D is the diameter of the antenna, receptor or telescope array. The aperture illumination factor is related to η_{ap}; it describes the beam size and field of view of the antenna. For a circular aperture with uniform illumination, $\mathcal{K} = 1.02$ (Rohlfs & Wilson, 2004).

The receptor field of view Ω_0 describes the angular area of sky visible to the half-power point of the receptor beam. For dishes in particular, this is termed the primary beam. The FoV of a circular aperture is the solid angle of the half-power beamwidth, derived from sky geometry:

$$\Omega_0 = \frac{\pi}{4}\theta^2_{\mathrm{HP}}. \tag{2.8}$$

For a parabolic dish,

$$\Omega_{\mathrm{dish}} = \frac{\pi}{4}\left(\frac{\mathcal{K}_{\mathrm{dish}}\lambda}{D_{\mathrm{dish}}}\right)^2. \tag{2.9}$$

Substitution of Equation 2.6 gives the well-known trade between effective area and FoV for a dish: $\Omega_{\mathrm{dish}} \propto A^{-1}_{\mathrm{e\,-\,dish}}$.

Figure 2.4: Schematic of a single dish with: (a) single-pixel feed; and (b) phased array feed.

As a survey telescope, achieving wide FoVs is a key design goal for the SKA. An alternative to trading effective area for FoV is to achieve both by using an FoV expansion technology to increase the dish FoV for given effective area. The dish FoV can be expanded by using a focal plane array to simultaneously form multiple beams (Veidt, 2006). A multibeam cluster feed is one such focal plane array, where multiple feeds are placed at or near the focus of the dish (e.g. Staveley-Smith et al., 1996; Bird & Cortes-Medellin, 2003). Each feed samples (illuminates) a slightly different area of the focal plane of the reflector, thereby forming multiple beams on the sky.

Alternatively, a phased array feed (PAF) can be used to sample the focal plane with multiple beams to increase the FoV, as shown in Figure 2.4. A PAF is an array of small antennas whose complex signals are appropriately weighted (in amplitude and phase) and combined, such that each set of weights produces a signal that samples a slightly different part of the focus. Although 'phased array feed' is the adopted SKA terminology, there is a range of alternative nomenclature for the technology, such as 'dense focal plane array' and 'array feed' (Hayman, 2011). Rather than the one-to-one mapping between the feeds and beams in the multibeam system, the PAFs can be used to fully (Nyquist) sample the focal plane out to the size of the PAF. This can be done in a single pointing, whereas the horn feeds used in a multibeam system requires interleaved pointings to achieve this sampling (Veidt, 2006).

For the SKA mid-frequencies from 0.5–1.5 GHz, PAFs may be a cost-effective method to increase the FoV and hence survey speed of dishes (Schilizzi et al., 2007). Alongside increased FoV, the benefit of PAFs is primarily due to the flexibility to dynamically customise the aperture illumination characteristics by changing the beamformer weights. This allows for the dish sensitivity and sidelobes to be customised for the scientific application (Schilizzi et al., 2007). Compared to single-pixel feeds, the drawbacks of PAFs are mainly due to technology-related challenges, as the PAFs and their accompanying beamformers are more complex and costly to design and implement (e.g. DeBoer et al., 2009).

As for dishes with single-pixel feeds, the effective area of each PAF beam is determined by Equation 2.6, although $\eta_{\rm ap}$ varies, depending on how the aperture is illuminated by the PAF. The processed FoV $\Omega_{\rm proc}$ depends on the number of PAF beams formed and how they sample

the focal plane. If each beam observes an independent part of the sky,

$$\Omega_{\mathrm{proc}} \approx N_{\mathrm{b\,-\,dish}}\Omega_{\mathrm{dish}}, \tag{2.10}$$

where $N_{\mathrm{b\,-\,dish}}$ is the number of beams formed at the dish ($N_{\mathrm{b\,-\,dish}} = 1$ for a dish with single-pixel feed). If the beams overlap, as they may do for PAF beams, a FoV efficiency factor can be included in Equation 2.10 to account for the overlap (Bunton & Hay, 2010).

For a given dish size, the phased array feed allows a greater processed FoV than for the single-pixel feed. However, PAF technology is still under development. PAFs are being implemented on the Australian SKA Pathfinder (ASKAP) to achieve a FoV 30 times larger than that of a single-pixel feed (DeBoer et al., 2009). ASKAP itself will be a part of the larger SKA_1-survey array[2]. Another in-progress PAF project for radio telescope arrays is APERture Tile In Focus[3] (APERTIF), an SKA pathfinder project to upgrade the Westerbork Synthesis Radio Telescope with PAFs.

2.3.1.4 Aperture arrays

At lower frequencies, phased arrays of electrically small antennas (aperture arrays, AAs) become an attractive solution to providing the sensitivity and FoV for the SKA. Instead of a focusing reflector, a planar array of simple antennas can be used to form the aperture. In modern radio astronomy, an aperture array is what IEEE (1993) describes as an active array antenna system, where some or all of the individual antennas are connected to their own receiver. The voltages measured from the planar array of antennas are adjusted in phase (or time) and weighted in amplitude to form a beam towards a specific direction on the sky.

The principle advantage of AAs is the flexibility to create multiple independent FoVs. The simple antennas can 'see' most of the observable sky, which means that the phased array beam can be 'pointed' within this large FoV. The term 'station' is generally used to describe the aperture array receptor; a schematic is shown in Figure 2.5. With sufficient signal processing, many station beams can be formed from the input antenna signals (van Ardenne et al., 2009). To further reduce signal processing costs, an intermediate stage of 'tile' beamforming can be introduced, as also shown in Figure 2.5.

The station effective area $A_{\mathrm{e\,-\,st}}$ depends on whether the array is 'dense' or 'sparse' at the frequency in question. There is no single definition for when an array is dense or sparse. The broad definition used in this thesis is that an array is dense when the aperture is fully sampled, and inter-element mutual coupling is significant. Effective area is then approximately equal to the physical (geometric) area of the station for a zenith pointing; this is the maximum effective area. When the array is sparse, the array effective area is approximately the sum of the effective area of the individual antenna elements.

The effective area of a dense circular aperture array station of uniformly distributed antenna elements is similar to a dish:

$$A_{\mathrm{e\,-\,st}} = \frac{\pi}{4}\eta_{\mathrm{r}} D_{\mathrm{st}}^2, \tag{2.11}$$

[2] http://www.skatelescope.org/news/dual-site-agreed-square-kilometre-array-telescope
[3] http://www.astron.nl/general/apertif/apertif

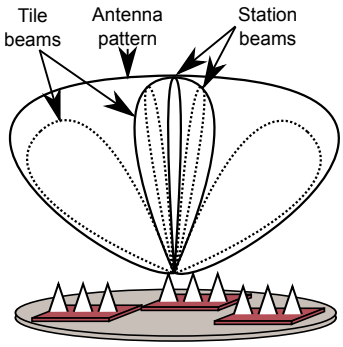


Figure 2.5: Aperture array schematic, showing the antenna elements (triangles), tiles (squares) and the station (disc), and their corresponding beam pattern.

where η_{r} is the antenna element radiation efficiency and D_{st} is the station diameter (Balanis, 2005). When the array is sparse, the effective area of each isolated element contributes to the station effective area, such that

$$A_{\mathrm{e-st}} = N_{\mathrm{e/st}} \frac{\lambda^2}{4\pi} \eta_{\mathrm{r}} \mathcal{D}, \tag{2.12}$$

where $N_{\mathrm{e/st}}$ is the number of elements per station and $\mathcal{D}$ is the directivity of an isolated antenna element (Mailloux, 1995). Wideband aperture arrays have a dense to sparse transition region, which occurs over an inter-element spacing of 0.5-1.5 λ for dipole-type antennas (Braun & van Cappellen, 2006). At these transition frequencies, effective area calculations generally require simulation of the antenna patterns and array factor, and analysis of mutual coupling effects (e.g. Craeye & González-Ovejero, 2011). However, Equations 2.11 and 2.12 are suitable first-order approximations for the high-level measures of performance used in this thesis.

For the same station, the FoV of a single station beam can be approximated by

$$\Omega_{\mathrm{st}} \approx \frac{\pi}{4} \left(\frac{\mathcal{K}_{\mathrm{st}} \lambda}{D_{\mathrm{st}}} \right)^2, \tag{2.13}$$

where $\mathcal{K}_{\mathrm{st}}$ is a frequency-dependent station beam taper determined by the station configuration. For a sparse station, the FoV is not directly related to the station effective area. But when the station is dense, $\Omega_{\mathrm{st}} \propto {A_{\mathrm{e-st}}}^{-1}$. The spacing of the antenna elements in the station is an important factor here, and is discussed in Chapter 5.

As for the PAFs, if multiple independently pointed beams are formed, the processed FoV is

$$\Omega_{\mathrm{proc}} \approx N_{\mathrm{b-st}} \Omega_{\mathrm{st}}, \tag{2.14}$$

where $N_{\mathrm{b-st}}$ is the number of station beams.

2.3.2 Signal transport

The SKA is a distributed telescope array, with many receptors. The cost and design limitations of transporting the digitised astronomical signals from the receptors is a potential filter on parameter space. The data rate R from the receptor scales as

$$R \propto N_{\mathrm{b-0}} N_{\mathrm{pol}} N_{\mathrm{bit-dig}} \Delta\nu \ , \tag{2.15}$$

where $N_{\mathrm{b\text{-}0}}$ is the number of receptor beams formed and $N_{\mathrm{bit\text{-}dig}}$ is the number of bits out of the digitiser. If the data rate is a limiting factor, then the product of these scaling factors are also limited. The primary trade-off here is processed FoV ($\Omega_{\mathrm{proc}} \propto N_{\mathrm{b\text{-}0}}$) and bandwidth $\Delta\nu$. The LOFAR telescope implements such a trade-off, where station beams and bandwidth are tradeable (de Vos et al., 2009).

The available data rate depends on cost and the choice of digital signal transport technology, and is discussed in the context of aperture arrays and long baselines in Chapters 5 and 6 respectively.

2.3.3 Digital signal processing

Like other modern radio telescopes, much of the SKA's capability will be realised in the digital domain. Once digitised, the data can be flexibly processed to best suit the scientific application. The principal use of the digitised data is to form images from aperture synthesis, using cross-correlated antenna pairs. Put succintly, aperture synthesis is the "spatial, spectral, and temporal sampling of the incoming radio-radiation field so as to match the expected structure of the field in those three domains" (Dewdney et al., 2009). However, there is also an explicit role for non-imaging processing in the SKA for pulsar observations, and other high time resolution observations such as those proposed in Chapter 4. Section 2.3.3.4 outlines the non-imaging processing.

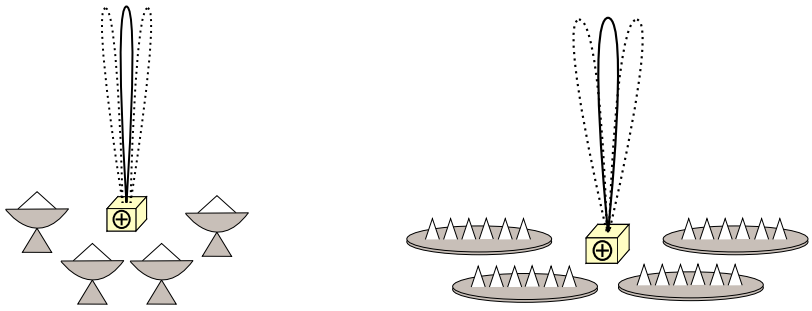
2.3.3.1 Filterbank

The filterbank transforms the digitised signal from the time to frequency domain. For the SKA, a time to frequency domain transformation and cross-correlation 'FX' correlator is the most cost-effective architecture, as opposed to other correlator topologies such as 'XF'. In an FX correlator architecture (e.g. Bunton, 2000), the signal from each element or tile input is filtered into frequency channels ('F') and for each channel, the input signals are cross-correlated ('X'); the architecture is more cost-effective with an increasing number of correlatable inputs (receptors in this thesis). An FX architecture also allows signal processing actions, such as beamforming and radio frequency interference (RFI) excision, to be efficiently performed (Hall et al., 2008).

Bunton (2003b) shows that the filtering can be efficiently implemented with a polyphase filterbank (PFB). Furthermore, cascaded or multi-stage filterbanks allow flexibility in the implementation architecture, where each filterbank stage is appropriately located to suit the architecture. The final filterbank in the cascade determines the spectral resolution $\Delta\nu_{\mathrm{ch}}$.

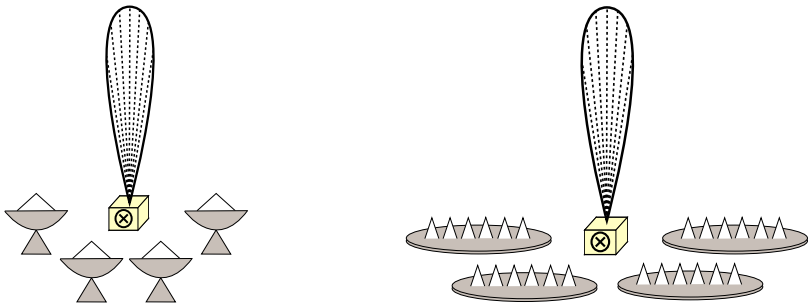
2.3.3.2 Phased array of receptors

Prior to discussing the correlator, it is conceptually simpler to first consider a phased or tied array of receptors, where voltages measured at each receptor are aligned in phase (or time) towards a specific direction on the sky, and perhaps amplitude weighted, in a manner similar to aperture arrays consisting of elemental antennas. The resulting array beam has higher sensitivity than an individual receptor, but a smaller FoV. Multiple beams can be formed within the FoV of the receptor beam, making array beams particularly useful for targeted compact



(a) Dishes. **(b)** Aperture arrays.

Figure 2.6: Phased arrays of receptors. Dotted lines show multiple beams.



(a) Dishes. **(b)** Aperture arrays.

Figure 2.7: Correlated arrays. Dotted lines show angular resolution.

source observations, such as pulsar timing (Kramer et al., 2004). Figure 2.6 shows a schematic of phased arrays of dishes and AAs, while Figure 2.7 shows the correlation counterpart.

Phased arrays of receptors follow a similar theoretical background as sparse aperture arrays. The effective area of each receptor contributes to the array effective area:

$$A_{\mathrm{e\,-\,arr}} = N_0 A_{\mathrm{e\,-\,0}}, \tag{2.16}$$

where N_0 is the number of receptors of effective area $A_{\mathrm{e\,-\,0}}$. The FoV of a single array beam is independent of $A_{\mathrm{e\,-\,arr}}$, and is given by

$$\Omega_{\mathrm{arr}} = \frac{\pi}{4}\left(\frac{\mathcal{K}_{\mathrm{arr}}\lambda}{D_{\mathrm{arr}}}\right)^2, \tag{2.17}$$

where $\mathcal{K}_{\mathrm{arr}}$ is the array beam taper, which depends on aperture distribution defined by the geographical layout of the receptors, and D_{arr} is the diameter of the array. The $\pi/4$ factor assumes a circularly symmetric distribution of receptors.

If $N_{\mathrm{b\,-\,arr}}$ independently pointed array beams are formed, the processed FoV is

$$\Omega_{\mathrm{proc}} = N_{\mathrm{b\,-\,arr}}\Omega_{\mathrm{arr}}. \tag{2.18}$$

2.3.3.3 Correlation and Imaging

The correlator is a processing device fundamental to aperture synthesis; it cross-correlates (multiply and average) signals from pairs of antennas of varying separation, to measure the brightness distribution of sources in the sky. The radio telescope array is a spatial frequency filter: the lowest spatial frequency is defined by the interferometer with the shortest baseline (the closest-spaced antennas), the highest by the longest baseline. The cross-correlation of an antenna pair results in a complex visibility, which is the interferometer's response (via Fourier transform) to the source brightness distribution. The correlator outputs measurements of these complex visibilities (see e.g. Thompson et al., 2001).

These measurements, or sampled visibilities, are used to form the resulting map or image of the sky. The range of spatial frequencies measured can be mapped onto a Cartesian coordinate system, the (u, v) plane. The baseline vector of each antenna pair occupies a point on the (u, v) plane, and the range of these spatial frequencies is called the (u, v) coverage (Thompson et al., 2001). The (u, v) coverage describes the aperture distribution; its Fourier transform gives the synthesised beam (or point spread function). The maximum angular resolution of the synthesised beam depends on this aperture distribution.

A sufficient number of these spatial frequencies allows for synthesis of the aperture, and imaging of the source brightness distribution. The basics steps to forming an image are:

(a) sample visibilities with the correlator

(b) Fourier transform the visibilities to create a 'dirty image'
- For an FFT to be used, 'gridding' is required, where each visibility is placed onto a discretised 2-D (u, v) plane.
- The dirty image is the convolution of the actual image (source brightness distribution) by the synthesised beam of the array.

(c) perform deconvolution on the dirty image to remove the effects of the synthesised beam and recover the source brightness distribution
- The deconvolution step may interpolate the missing spacings of (u, v) coverage to improve the recovered image.
- Deconvolution steps are repeated until sufficient image quality is achieved.

Standard texts (e.g. Taylor et al., 1999; Thompson et al., 2001) describe this process in detail.

The sensitivity of a correlation array of N_0 identical receptors increases proportional to the number of correlated baselines. For $N_0(N_0 - 1)/2$ baselines, the point source sensitivity of an image is (Crane & Napier, 1989):

$$S_{\min} = \frac{\sigma 2kT_{\rm sys}}{A_{\rm e\,-\,0}\sqrt{N_0(N_0-1)N_{\rm pol}\Delta\nu\tau}}. \tag{2.19}$$

For large N_0,

$$S_{\min} \approx \frac{\sigma 2kT_{\rm sys}}{N_0 A_{\rm e\,-\,0}\sqrt{N_{\rm pol}\Delta\nu\tau}}. \tag{2.20}$$

Given $A_{\rm e} = N_0 A_{\rm e\,-\,0}$, the sensitivity to a point source is similar for a single dish, phased array and correlated array of equivalent total effective area $A_{\rm e}$.

A key feature of interferometers is that unlike a single dish telescope, angular resolution is independent of receptor FoV. The correlator is an efficient mapping device, because with a single observation interval, the Fourier transform of the visibilities results in an image that is sensitive to the range of spatial frequencies defined by the (u, v) coverage (Thompson et al., 2001). Thus the processed FoV of the receptor can be imaged, with angular resolution of $\phi \approx \lambda / b_{\max}$ radians, where $b_{\max}$ is the maximum baseline length. To achieve a similar mapping with a phased array requires many array beams to be formed. The computational cost of each approach depends on the actual application (see Section 2.3.3.4).

Spatial frequencies less than that observed by the shortest baseline can be observed with a single dish (total power), or by a technique called mosaicing to recover the information from baselines shorter than the closest-spaced interferometer (Holdaway, 1999). These low spatial frequencies are sensitive to source structure larger than $\sim \lambda / b_{\min}$ radians, where $b_{\min}$ is the minimum baseline length (Thompson et al., 2001).

However, there are technical limitations on imaging the full receptor FoV. The imaging steps described earlier consider the baselines to either be coplanar or approximated as coplanar; this allows the use of a 2-D FFT to transform the sampled visibilities into dirty image. When

$$\frac{D_0^2}{\lambda b_{\max}} \lessapprox 1, \tag{2.21}$$

non-coplanar baselines affect the image (Cornwell et al., 2008). Wide-field imaging describes the problem of imaging a large FoV when this coplanarity approximation no longer holds. For the SKA, the relatively small apertures, long wavelengths (low frequencies) and long baselines make wide-field imaging an important consideration for particular observations. Solutions such as processing sub-images before recombining them as a larger image ('faceting') are computationally expensive (Cornwell & Perley, 1992); these are being superseded by more efficient algorithms (e.g. Cornwell et al., 2008, 2012). Even so, wide-field imaging still presents a significant challenge for the SKA (Cornwell, 2012).

There are other imaging challenges, such as various calibration routines, bright source removal, RFI mitigation and correcting for the rotation of beam patterns on the sky (Dewdney et al., 2009). Therefore, most astronomers will likely use astronomical quality data products (corrected images) from the SKA, rather than the visibilities or (u, v) data that is the main data output from current telescopes (Alexander et al., 2009). These proposed data products reduce the volume of data, but potentially place further implicit parameter space filters on the signal.

2.3.3.4 Non-imaging processing

In the SKA context, the non-imaging processing describes the signal processing required to search for pulsars and time them (a key science goal for SKA_1), along with additional high time resolution science such as fast transients. The standard correlation and imaging procedures usually average the visibilities over a period of order seconds. This is not suitable for high time resolution experiments, hence other signal processing procedures are required. To retain the full sensitivity of the array, the visibilities can be averaged over much shorter timescales. Alternatively, many array beams can be used to observe the sky (Section 2.3.3.2). Cordes (2009a) compares the computational cost pixelising the full FoV through these two approaches

(correlation and beamforming), and finds that the cost depends primarily on the filling factor (spatial density of the receptor layout), where higher filling factors favour beamforming over correlation. Chapter 4 discusses how a higher filling factor improves the effectiveness of the array beams for high time resolution observations.

The non-imaging processor in Turner et al. (2011) proposes the beamforming approach to search for pulsars, using a processed FoV of approximately 1.25 $\deg^2$ formed from tens of thousands of array beams. Once the FoV is pixelised, each beam (or pixel) is searched for pulsars. Turner et al. (2011) details the steps in the pulsar search, where the key procedures in the search are to dedisperse the signal to correct for dispersive delays due to the cosmic medium, and conduct acceleration searches to detect binary pulsars. However, the exact implementation details for pulsar observations with the SKA are a work-in-progress.

For fast transient searches, there are a variety of current and future non-imaging processing options, including the correlation and beamforming described above. In Chapter 4, I analyse these options and compare their cost effectiveness.

2.4 SKA system details

For completeness, I give an overview here of the current representative system design for the SKA. The most recent detailed documentation of the SKA system are the high-level system descriptions (HLSDs). At the system Concept Design Review (CoDR) and subsequent system delta-CoDR, a HLSD was developed for SKA_2 (Dewdney et al., 2010b) and SKA_1 (Dewdney et al., 2011a) respectively. These documents describe the SKA concept design through 'representative implementations' of the sub-systems, based on the known requirements at the time. They provide a basis for advancing the requirements analysis, but leave sufficient scope for optimisations and trade-offs as part of this requirements analysis. Table 2.2 provides an overview of the system details SKA_1 and SKA_2. In Table 2.2, the details for SKA_2 are based on a 'maximal system' of candidate receptor technologies; in due course, there will be technology down-selections for SKA_2.

The main design implications from the recent SKA site decision[4] are:

- Some duplication of infrastructure will be required, although existing infrastructure at both sites will be used.
- The central signal processing cannot be shared.
- Dishes with PAFs will be implemented as a third receptor type for SKA_1.

Only the last of these significantly changes the HLSD. Up to 60 SKA_1 dishes will be equipped with PAFs and added to the ASKAP's 36 dishes, forming an SKA_1-survey array (SOWG, 2012). Although the ASKAP dishes have a smaller 12 m diameter, I have assumed for simplicity in Figure 1.2 that all 96 dishes are of 15 m, with PAF specifications as per the SKA_2 HLSD.

[4] http://www.skatelescope.org/news/dual-site-agreed-square-kilometre-array-telescope

2.5 Chapter summary

The chapter reviewed the top-level measures of performance as as descriptors of the information-gathering capability of the SKA. Using the concept of parameter space, I showed that this information-gathering capability depends on how a given telescope design filters the incoming signal. These measures of performance therefore indicate the scientific capability of a given design.

Section 2.3 showed how the sub-systems in the elemental signal path act as filters on the incoming signal, and hence limit scientific capability. The quantity and characteristics of SKA receptors significantly impact telescope performance, as they are the principal determinant of the observing frequencies, sensitivity and field of view; these are the key factors for survey experiments. However, post-receptor signal transport and processing also act as filters of field of view and instantaneous bandwidth, as well as spectral and temporal resolution. These are key limitations of, for example, fast transient searches. Assessment of telescope performance requires consideration of all sub-systems in the elemental signal path, reinforcing strongly the need for the systems engineering approach outlined in Chapter 1.

Table 2.2: Overview of system details for SKA Phase 1 (P1) receptors and SKA Phase 2 (P2) candidate receptor technologies, as per Dewdney et al. (2010b, 2011a) and SOWG (2012).

	Low-frequency aperture array		SKA_1-survey	Mid-frequency aperture array	Dish array		
	P1	P2	P1: PAF[a]	P2	P1: SPF[b]	P2: WBSPF[c]	P2: PAF[d]
Frequency range (GHz)	0.07–0.45		0.5–3	0.4–1.4	0.45–2	0.3–12	0.5–3
Number of receptors	50	250	96	250	250	3300	2000
Receptor diameter (m)	180	180	15	56	15	15	15
Processed FoV (deg^2)	20	250	161 / 26 [e]	250	Dish beam	Dish beam	161 / 26 [e]
Distribution of receptors in array configuration regions							
Core (radius<0.5 km)	50 %	66 %	N/A	66 %	50 %	20 %	25 %
Inner (0.5<radius<2.5 km)	20 %		N/A		20 %	30 %	30 %
Mid (2.5<radius<180 km) [f]	30 %	33 %	N/A	33 %	30 %	26 %	45 %
Outer (180<radius<3000 km)	0 %	0 %	N/A	0 %	0 %	24 %	0 %

[a] Dishes with phased array feeds, details extrapolated from SKA_2 HLSD PAF specifications.
[b] Dishes with single-pixel feeds.
[c] Dishes with wideband single-pixel feeds.
[d] Dishes with phased array feeds.
[e] 0.5–1.2 GHz and 1.2–3 GHz respectively.
[f] Maximum radius of 100 km for SKA_1.

Chapter 3

SKA performance and cost exploration

For the SKA and science mega-projects in general, the interrelationship between telescope design, performance and cost is complex. Systems engineering provides a well-defined process to meet the challenge of achieving a cost-effective design, where, in the SKA context, the scientific and other requirements are met within a fixed cost ceiling. These requirements are primarily defined in the current system definition stage, where conflicting requirements, and cost, schedule and performance risks are identified and resolved through systems analyses and trade studies.

This chapter relates the measures of telescope performance in Chapter 2 with the systems engineering process of requirements analysis. Sections 3.1 and 3.2 review the rationale for systems engineering in the SKA context. Section 3.3 outlines the system definition phase, which includes requirements definition, system modelling and trade-off analysis. Section 3.4 details the performance and cost modelling approach used in later chapters. Developing these models is an on-going effort for the SKA; Section 3.5 presents SKACost as a tool assist in the design process, and shows examples of previous uses. Finally, Section 3.6 discusses performance and cost uncertainties.

3.1 The rationale for systems engineering

Mega-projects are at high risk of cost overruns, but systems engineering is one tool to mitigate this. A brief review of cost overruns in other large mega-project systems rationalises the importance of the SKA systems engineering approach. A general definition of a mega-project is a project of significant public interest, with a cost greater than €1 billion or US$1 billion. Cost overruns of more than 50% are common in infrastructure mega-projects, for which the root cause is attributed to systematic under-estimation of cost (Flyvbjerg et al., 2002, 2003). This occurs when planners do not accurately account for the risk of unforeseen events in complex projects, which can increase the cost and lengthen the schedule. Contributing factors include imperfect or inadequate information to make a decision, changes to requirements, and management issues (Flyvbjerg & COWI, 2004).

Sufficient and proper planning is critical in avoiding cost overruns. The James Webb Space Telescope (JWST), currently under-construction, is a specific example of such significant cost overruns. A review panel report on the JWST (Casani et al., 2010) judged the total life-cycle cost of the project to be US$6.2–6.8 billion, and a NASA revision puts the cost at US$8.8 billion (Howard, 2012). In contrast, the cost estimate measured at project confirmation (after the preliminary design stage) was nearly US$5 billion. The cost estimates in the earlier stages of the project are not considered in the JWST report, but earlier projected costs increased (in 2006 dollars) from US$1 billion in 1996 to US$3.5 billion in 2006 (Reichhardt, 2006). The JWST panel report ascribed the telescope project's cost increase and launch delay to budgeting

and management problems; the baseline cost at the time of project confirmation was "...flawed because it was not based upon a current bottoms-up [cost] estimate and did not include the known threats [unforeseen work that arose during earlier project stages]" (Casani et al., 2010).

To determine the principal success drivers in high-technology mega-projects, Crosby (2012a) recently conducted a meta-study of such projects. He analysed publications and case studies encompassing 2820 mega-projects, derived 18 success factors and ranked these factors according to their contribution to project success. Project management systems and a clear project definition ranked as the key success drivers. Clarifying the project definition is critical in the SKA's conceptual planning stage; this includes defining the requirements and scope of the project, and developing a sound business case Crosby (2012b). According to GAO (2011), a business case that matches the project requirements and resources results in programs that "...are more likely to be successfully completed within cost and schedule estimates and deliver anticipated system performance."

The relatively smaller-scale SKA pathfinder and precursor projects also present systems engineering lessons for the SKA. For example, LOFAR, with a combination of a large number of antennas and a predominantly digital signal chain, is a successful demonstration of industrial collaboration and mass-production techniques (Schilizzi et al., 2011). Furthermore, LOFAR has used a phased prototyping and deployment approach (de Vos et al., 2009). These considerations were generally not important in the realisation of previous radio telescopes, but are likely to be applicable for the SKA.

The funding agencies expect that the science mega-projects are cost-effective; that is, they achieve the requirements for the lowest cost. Or to rephrase for a cost-constrained project such as the SKA, a cost-effective design maximises scientific capability within construction and operations funding constraints. From the examples given here, an early focus on both understanding the requirements and producing a design that meets these requirements within the cost-ceiling is important to a project's success.

3.2 Systems engineering in the SKA context

Systems engineering is being employed to ensure that SKA meets the expectations of stakeholders, such as the scientists and funding agencies. The systems engineering approach is used to design, realise and operate a system, where the system is something which achieves results that cannot be achieved by its constituent components individually (NASA, 2007). Systems engineering is particularly beneficial for large and complex projects, because it provides a process to design and implement the project through assessing and balancing the system requirements and constraints. For the SKA, this means optimising the design and making trade-offs, so that the system requirements, such as scientific, environmental and operational requirements, are fulfilled within engineering and funding constraints.

Systems engineering is process-driven. The project timeline is divided into stages, from the initial idea to the project's closure, with activities specified for each stage. An International Standard (ISO/IEC, 2008) exists for systems engineering; a complementary standard (IEEE,

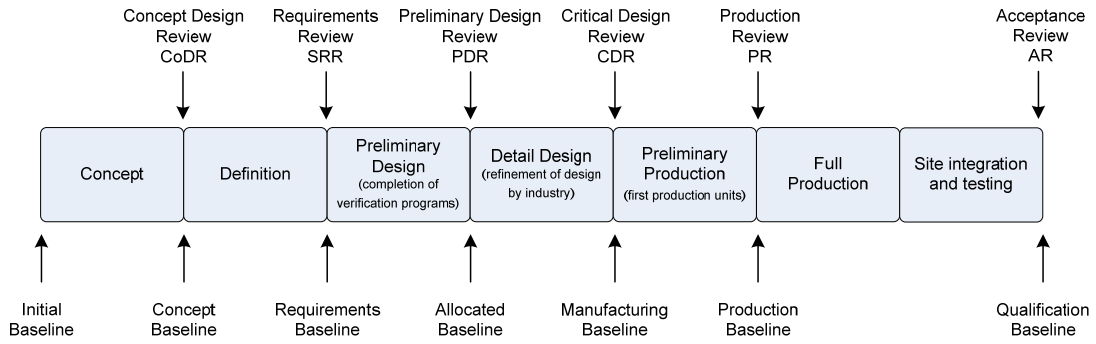


Figure 3.1: SKA systems engineering stages, review points and baselines. Source: Stevenson (2011c)

2005) provides a more detailed description of the systems engineering processes. Furthermore, individual organisations produce their own handbooks (e.g. INCOSE, 2010; NASA, 2007).

The SKA *System Engineering Management Plan* (SEMP, Stevenson, 2011c) describes the systems engineering stages for the SKA, shown in Figure 3.1 and summarised as follows:

Concept: Develop an understanding of the problem and a draft high-level concept, through background investigation into technology options and preliminary investigations into requirements and risk.

Definition: Ensure a complete set of requirements exists, through requirements analysis and validation, and select candidate technical solution(s).

Preliminary design: Refine candidate technical solution(s) to a single low risk baseline solution. Draft final designs and retire major risks.

Detailed design: Verify designs for requirements compliance; refine design for production.

Production (preliminary and full): Manufacture, assemble, integrate, test and verify sub-systems.

Site integration and testing: Install sub-systems; integrate, test and verify the system.

Each stage concludes with a design review, although in reality there will be some overlap between the end of one stage and the beginning of the next.

Within the systems engineering process, these stages can be split into two groups: the first four encompass *system design*, the last two encompass *product realisation* (NASA, 2007). Each of the system design stages follow a set of basic systems engineering processes, shown in Figure 3.2. These processes are described in detail in IEEE (2005); they aim to ensure that the system is designed without gaps. System optimisation and trade studies are a key feature of Figure 3.2, and have an important role in determining a cost-effective design solution, as discussed in the next section.

3.3 System definition stage

The SEMP describes the aims of the current system definition stage as being:

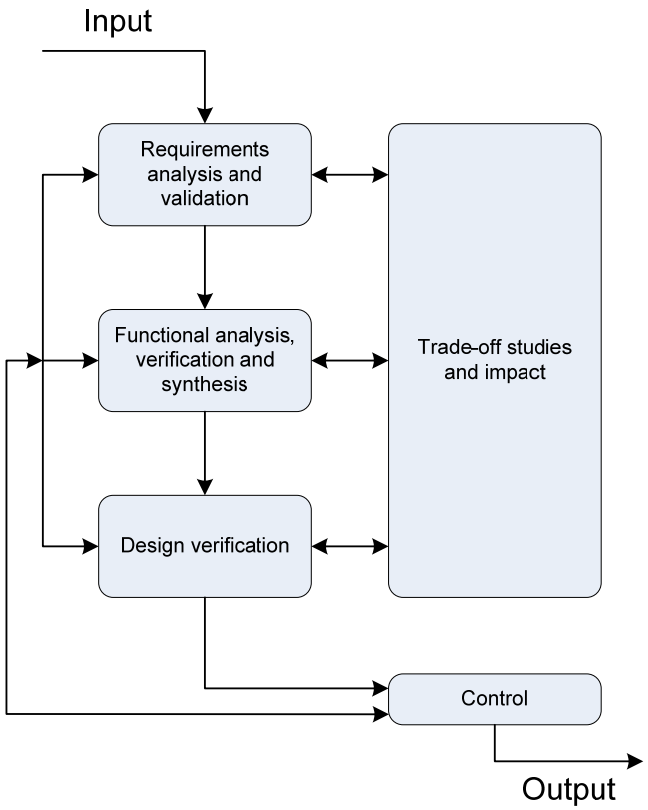


Figure 3.2: Simplified outline of systems engineering processes for the system design stages. Source: Stevenson (2011c).

- requirements analysis and validation to compile a complete set of requirements
- further investigation of the technology options that were confirmed during the Concept Design Review (CoDR) as continuing options
- trade-off studies between possible design solutions, with the goal being to identify and select a preferred solution (or more than one preferred solution for those sub-systems where it is not possible to arrive at a single solution in the definition stage)
- production of a draft architectural design document, which is a logical, rather than physical, description of what the system must do to satisfy requirements.

The subsequent systems engineering activities for the telescope's design, construction and operation activities are based on the complete set of requirements developed during the system definition stage. The qualifier 'complete' is important, because the life-cycle cost can be influenced much more easily in the early stages of the project. If there are design changes to be made (due to unidentified or changed requirements, or mistakes), the cost of implementing the changes late in a project is two orders of magnitude higher (INCOSE, 2010; Stevenson, 2011c).

3.3.1 Requirements analysis

The system is described by a set of requirements, including functional and performance requirements that respectively define *what* is done by the telescope, and *how well* it is done. The

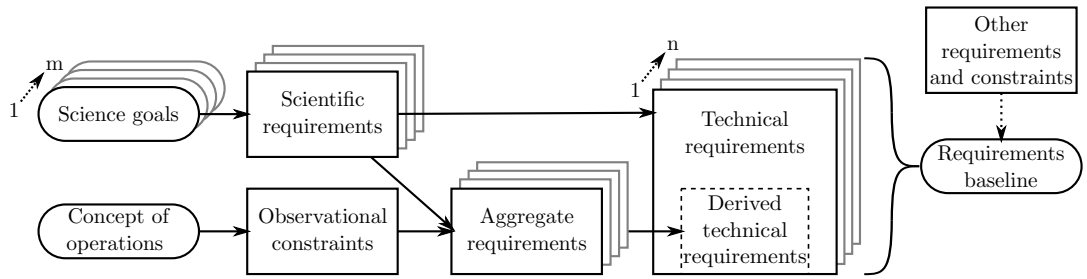


Figure 3.3: Relationship between requirements, and the path to deriving technical requirements.

functional and performance requirements are primarily derived from the scientific goals of the SKA, but within the context of funding constraints. A summary of the tasks for requirements analysis are (IEEE, 2005):

- analyse the inputs, such as stakeholder expectations and constraints
- identify cost, schedule and performance risks
- define functional and performance requirements
- define other (non-functional) requirements and constraints
- identify conflicting requirements, and conduct system trade-off analyses to resolve such conflicts.

The flow of tasks in the requirements analysis process is not linear; there are actually many iterations and recursive steps, especially during the system trade-off analyses (Section 3.3.2).

The goal of these tasks is to produce a requirements baseline at the end of the system definition phase (see Figure 3.1), where the requirements and constraints of the system are captured in a Systems Requirements Specification (Cloete, 2011). Once the requirements baseline is produced, it is then verified to ensure that it meets the expectations of the stakeholders (e.g. funding agencies, astronomers), and used for the subsequent functional analysis process in Figure 3.2.

3.3.1.1 The Design Reference Mission

Inputs to the requirements analysis come from a variety of sources, such as astronomers and engineers, technical standards and project documents (Stevenson, 2011a). A key input to the requirements analysis is the Design Reference Mission (DRM), a document that represents the SKA science goals. The set of science observations described in the DRM provide an 'envelope of requirements' for the telescope. There is currently a DRM for SKA_1 (SSWG, 2012) and SKA_2 (SSWG, 2009), hereafter DRM_1 and DRM_2 respectively.

The first draft of SKA_1 Systems Requirements Specification (Stevenson, 2011b) sets out different types of requirements, based on the DRM_1. I use these basic concepts, but apply them such that the tradable and non-tradable requirements are more clearly distinguished. My definition of the requirements and their relation to each other is shown in Figure 3.3, and described as follows:

Scientific requirement: Developed from the science cases, and presented in scientific terms.

Table 3.1: Summary of principal requirements analysed in the case studies in this thesis (Chapters 4–6). See text for explanation of the categories.

Fast transients	**Single and dual-band $\mathbf{SKA_1}$-low**	**Long baselines of the $\mathbf{SKA_2}$ dish array**
	Scientific requirements	
• search for singly occurring or intermittent impulsive signals • exploration of the unknown	• as per DRM_1	• as per DRM_2 • Godfrey et al. (2011)
	Observational constraints	
• observing time	• as per DRM_1	• as per DRM_2
	Aggregate requirements	
• event rate (probability of intercept)	• as per DRM_1	• none
	Technical requirements	
• frequency ν • sensitivity A/T • processed FoV Ω_{proc} • processed bandwidth $\Delta\nu$ • temporal resolution Δt	• sensitivity A/T • processed FoV Ω_{proc} • others as per DRM_1	• (u, v) coverage • processed bandwidth $\Delta\nu$ • angular resolution ϕ • processed FoV Ω_{proc}

Technical requirement: Functional and performance requirements; each technical requirement requirement is mutually independent (not tradeable).

Aggregate requirement: Representation of multiple observations. Derived from a combination of one or more scientific requirements and observational constraints such as time allocation and scheduling.

Derived technical requirement: Derived from an aggregate requirement, a derived requirement only takes precedence as a design driver if it is more demanding than the technical requirements derived directly from the scientific requirements.

Table 3.1 summarises the requirements analysed in this thesis. If an aggregate requirement gives rise to more than one derived requirement, then trade-offs can be made between the derived requirements that form the aggregate requirement. For example, I make such trade-offs in Chapter 4, for the aggregate and technical requirements shown in Table 3.1.

Regardless of how they are derived, the end result is a set of technical requirements for each science case, as shown by the 1..n boxes in Figure 3.3. These technical requirements are input into the Systems Requirements Specification.

3.3.1.2 Requirements represented in parameter space

The process for deriving functional and performance requirements for the SKA is not clearly defined in the SEMP. The presupposition of the DRMs is that if the system meets the envelope

of requirements, then the system can potentially be used for many other science observations, including the exploration of the unknown (SSWG, 2012). However, the DRMs themselves are not a complete analysis of the technical requirements. For example, an analysis of DRM_1 version 2.0 by Alexander et al. (2012) shows that there are areas of insufficient detail to translate that DRM into a complete set of technical requirements.

I present here a more general approach to understanding scientific requirements in the context of parameter space, and relate this approach to the DRMs. To this author's knowledge, Harwit's parameter space analysis (Section 2.1) has never been linked to a systems requirements analysis of a telescope. I do so here, because not only does it describe the astronomical observations defined by the science goals, it is a useful tool to characterise the requirements of the 'exploration of the unknown'.

The top-level requirements provide a general description of the regions of parameter space to which the telescope is sensitive. The top-level *technical requirements* can be described by the parameter space axes in Table 2.1, while the *aggregate requirements* can be described by survey metrics (Section 2.2.2). With this approach, each set of technical requirements describes a multi-dimensional filter (hypervolume) on the observational parameter space (Section 2.1.2). Therefore, the science to be done with the SKA is defined by the boundary (or envelope) of all of these hypervolumes. In addition, such an approach gives a reasonable first-order description of the parameter space explored by an SKA designed to meet the envelope of requirements.

The same approach can be used in reverse to identify the technical requirements for the exploration of the unknown. The exploration of the unknown is not adequately defined by the science goals in the DRM, because they are based on known or expected phenomena. But specific regions of parameter space can identified as candidate areas for exploration; the high time resolution Universe being one such example (see Section 2.1.3). With the candidate regions of parameter space defined, system design trade-offs and optimisations can be made with the goal of maximising the exploration of these regions in a cost-effective manner.

3.3.2 SKA systems analysis

Although the technical requirements for the SKA are not yet complete, there is sufficient information about the high-level technical requirements to explore aspects of the SKA design through systems analysis, shown as 'trade-off studies and impact' in Figure 3.2. To give some context to the trade-offs, Figure 3.4 shows a generalisation of the systems analysis process that happens during each of the system design stages. For each stage, the process begins at the centre of the spiral. The top-level requirements of the system (or sub-system) are identified, along with suitable design concepts. Cost, performance, risk and schedule conflicts are resolved through trade studies. The impact of alternative solutions are analysed, with the aim of selecting the solution or solutions which best meet the objectives as defined by the stakeholders (IEEE, 2005); these objectives are generally a combination of cost, performance, risk and schedule. The results then input back into the systems engineering process in Figure 3.2.

The SKA project has already undertaken the activities in Figure 3.4 a number of times. The designs selected at different stages are documented as follows:

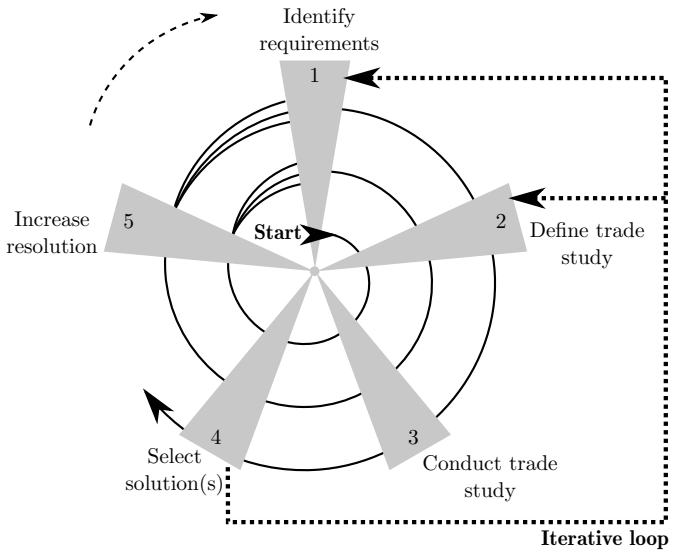


Figure 3.4: Generalised systems analysis activities. Each of the grey triangles show an activity. The dotted line indicates the iterative loop that can occur between activities. The design space is recursively more deeply explored with successive turns of the circle. Adapted from NASA (2007).

- Pre-concept stage:
  - *Reference Design for the SKA* (ISPO, 2006)
  - *Preliminary Specifications for the Square Kilometre Array* (Schilizzi et al., 2007).
- Concept stage, at the Concept Design Review (CoDR) and delta CoDR (dCoDR):
  - *High-level SKA System Description* (Dewdney et al., 2010b)
  - *SKA1: High Level System Description* (Dewdney et al., 2011a).

In the concept stage, the design was explored in more detail through subsequent CoDRs of the major sub-systems.

The case studies in this thesis are each an iteration of this systems analysis process. Table 3.2 summarises the systems analysis activities for each case study; the activities themselves are detailed below.

3.3.2.1 Define trade study

The purpose of the trade study is to achieve a convergence of requirements and potential design solutions. IEEE (2005) defines a trade-off analysis as being 'judgemental', 'informal' and 'formal'. The judgemental trade-off analysis is subject to the expert opinion of the designer. This expert opinion is used when one option is obviously superior, or when there is insufficient time to make a detailed analysis. The informal and formal analyses occur when the consequences of the trade-off are more important; the formal is simply documented more thoroughly than the informal. The case-studies in this thesis are informal analyses.

Activities in IEEE (2005) for defining trade-off analysis include:

Select methodology: The chosen methodology depends on the importance of the trade-study, and the resources available. In this thesis, I use parametric modelling to explore

Table 3.2: Summary of systems analysis activities undertaken for the case studies in this thesis.

Fast transients	**Single and dual-band SKA_1-low**	**Long baselines of the SKA_2 dish array**
	Define trade study	
	Methodology	
Model fast transient observing strategies and determine the experimental and system design implications for SKA_1 and radio telescope arrays in general.	Compare single vs. dual-band SKA_1-low implementations.	Explore cost-effective data networks for SKA long baselines.
	Success criteria	
Cost: low cost for SKA_1. Performance: maximise the number of verifiable events detected in a survey.	Cost: lowest cost. Performance: similar processed FoV and sensitivity between implementations.	Present alternative design solutions of varying cost and performance.
	Identify alternatives	
Evaluate incoherent and coherent signal combination modes.	Single and dual-band SKA_1-low implementations; RF tile and all-digital beamforming.	Alternative designs vary the beam–bandwidth product, number of remote stations and maximum baseline.
	Trade study environment (modelling framework)	
Simple parametric models of SKA_1 cost effectiveness: event rate per beam and probability of intercept.	Detailed parametric models of SKA_1-low station costs and broader system costs.	Simple parametric models of the long baseline data network
	Conduct trade study	
	Analyse life-cycle costs (LCCs)	
Beam 'formed and searched' as a proxy for cost in the absence of a detailed system architecture.	Analysis of differential costs. LCCs not analysed (see Section 5.3.2).	Capital costs representative of LCCs (see Section 6.2).
	Analyse cost-effectiveness	
Trade-offs as a function of receptor type, frequency, sky direction and expected population characteristics.	Comparative study of the design solutions. Consider examples of reduced FoV and of more stations of smaller diameter.	Trades between performance and cost.
	Quantify risk factors	
No analysis of risk factors.	Uncertainties calculated from two cost data sources developed with different methodologies. Simple sensitivity analysis for some broader system costs.	Sensitivity analysis of technology capability and network costs.

performance and cost trends, and to conduct trade-offs, as detailed in Section 3.4. The parametric analysis involves defining the system by a set of parameters, and then modifying one or more of these parameters to model the effect on system performance and cost.

Select success criteria: The success criteria defines how the design solution are compared. The criteria may assess factors such as cost, performance, schedule and risk.

Identify alternatives: Identify the design solutions to be studied.

Establish trade study environment (modelling framework): Establish the models used for the trade study, as well as metrics of the success criteria.

3.3.2.2 Conduct trade study

The trade studies analyse life-cycle costs, cost-effectiveness, safety and environmental impacts and quantifies risk factors (IEEE, 2005):

Analyse life-cycle costs: The life-cycle cost (LCC) is the total cost of ownership over the lifetime of the system (NASA, 2007). The *Draft SKA costing strategy* (McCool et al., 2010) identifies the sources of cost for the SKA. It breaks costs into three types:

- sub-system costs, such as hardware, software and operational costs
- infrastructure and siting costs
- project overheads, such as contingency, taxes and system integration costs.

A life-cycle cost estimate requires significant effort to develop, and has not yet been done for the SKA. However, rough estimates that only cover some of the life-cycle costs is an acceptable approach for the investigation of high-level alternatives (GAO, 2009). In this thesis, I make differential cost estimates using available data, where I exclude those costs that do not significantly differ between designs or architectures. I primarily consider the sub-system hardware costs, but also consider costs such as infrastructure and siting costs as appropriate.

Analyse cost-effectiveness: The system cost-effectiveness is a joint metric of costs and measures of effectiveness, where the latter usually depend on system performance. Particular design solutions may reduce costs without changing effectiveness, or increase effectiveness for a given cost. These optimisations of the system design increase cost-effectiveness. Alternatively, both cost and effectiveness can be increased or reduced (NASA, 2007). These latter trade-offs result in changes to the capability of the telescope to meet the requirements. For the case studies, I use parametric modelling to explore the cost-effectiveness of a variety of design solutions by varying key input parameters. The input parameters may be technical requirements, or design parameters that affect performance.

Quantify risk factors: The trade-off analyses assess the impact of risk on the system cost-effectiveness (IEEE, 2005). The main risk factors considered in this thesis are those uncertainties associated with performance and cost estimation (see Section 3.6). I evaluate some of these risks with sensitivity analyses, where the effect of varied input values are assessed (INCOSE, 2010). As the work breakdown structure is developed, a more complete analysis of cost, schedule, technical and programmatic risks can be made (GAO, 2009).

Safety and environmental impacts: I do not consider these impacts in this thesis, as the safety and environmental requirements are yet to be fully developed (Dewdney et al., 2011a).

3.3.2.3 Select solutions

In Figure 3.4, Step 4 is a decision point for the iteration or recursion of the systems analysis. If the stakeholders' objectives cannot be met then either the requirements or designs are reconsidered, via the iterative loop. If one or more appropriate design solution are found, the resolution of the design is increased, such that the next level down in the system is considered, and the steps repeated. The case studies in this thesis represent a single iteration in Figure 3.4 and do not represent a final convergence. However, they do provide insight and guidance for further iterations, as discussed in Chapter 7.

3.4 Performance and cost modelling

This section discusses the background to the modelling; it is most relevant to the SKA_1-low (Chapter 5) and long baseline (Chapter 6) case studies, because these chapters undertake more detailed cost modelling than the fast transients case study (Chapter 4).

The high-level modelling ensures that the system is represented with sufficient accuracy while maintaining scalability. The models use the best information available at any given time; this allows for the refinement of costs as technology develops and uncertainty reduces. Development of the models requires a thorough understanding of the system performance and design. The cost models are developed with bottom-up estimates where the detailed designs are available, and a top-down, parametrised approach for those costs which are better estimated from existing trends. Chapter 5 is an example of model development for SKA-low.

3.4.1 Cost estimation methodologies

A robust cost estimation methodology is required to determine the cost-effectiveness of alternative designs in the trade-off analyses. The appropriate methodology and detail of the cost estimate depends on the project stage and the trade-offs being considered. The *2008 NASA cost estimating handbook* (NASA, 2008) lists three cost estimation methodologies: parametric, analogous and engineering:

Parametric: As the primary method in the earlier stages of the project, parametric cost estimation facilitates high-level trade-offs when there is insufficient data for a detailed approach (NASA, 2007). Parametric estimates use a set of mathematical models that describe how the cost of the system scales with various input parameters, such as physical or performance characteristics. Parametric models allow for 'what if' questions to be asked and explored (NASA, 2008).

Reference class (analogous): Uses cost data from similar projects. The cost data is adjusted, depending on the relative complexity of the projects, technological improvements, inflation and other factors. The term 'reference class forecasting' is used by Flyvbjerg (2008) to describe how the cost of a proposed transport infrastructure mega-project is compared to a statistically meaningful number of similar projects that have already been implemented.

Table 3.3: Recommended use of cost estimation methodologies at various stages in a project. SKA is currently in the design definition stage. Adapted from NASA (2008).

	Early concept definition	Design definition	Preliminary and detailed design	Production and deployment	Operations, support and disposal
Parametric	●	●	◑	◑	○
Reference class (analogous)	●	◑	◑	◑	○
Bottom-up (engineering)	◑	◑	●	●	●
● Primary	◑ Applicable	○ Not applicable			

Such an approach is difficult to undertake for the SKA telescope as a whole, because there is no directly comparable project. However, certain sub-systems may be comparable between projects, thus providing the reference class cost data discussed in Section 3.4.2.2.

Bottom-up (engineering): Being the most detailed method, is more useful later in the project as actual cost data is accumulated (NASA, 2007). The bottom-up methodology builds up a cost estimate by adding up all the individual cost elements in the system, where each element represents a product or service that can be procured (IEEE, 2005; NASA, 2008). Engineering cost estimation can give insight into the major cost contributors, but not the cost drivers in the system. Such estimation also requires a good understanding of all the costs involved ('cost coverage'), as outlined in the *Draft SKA costing strategy* (McCool et al., 2010); incorrect estimates can result if significant costs are not included (NASA, 2008).

The project stages where these methodologies should be used are shown in Table 3.3. I use parametric estimation in this thesis; it is recommended as the primary cost estimation method in the design definition stage.

3.4.2 Complex parametric models

Parametric cost estimation requires appropriate research and documentation to develop the models, termed cost estimating relationships (CERs). Very simple CERs use only a few input parameters, whereas a 'complex model' uses many more input parameters, and requires some expert judgement to develop the models. The complex models use interrelated CERs, combined with other rules and assumptions (ISPA, 2008).

The sub-systems that are described by these cost estimating relationships can be represented in a hierarchical structure. For example, the product breakdown structure (PBS) describes hardware, software and information (data) items in the system (NASA, 2007); the requirements are traceable through this hierarchical structure. The SKA system hierarchy (Figure 3.5) represents the major telescope sub-systems, and is a form of PBS.

The PBS in turn sits within a work breakdown structure (WBS), which is a project management tool that describes the work to be done (GAO, 2009). A WBS covers all parts of the systems, including responsibilities and deliverables for work to be performed (Stevenson,

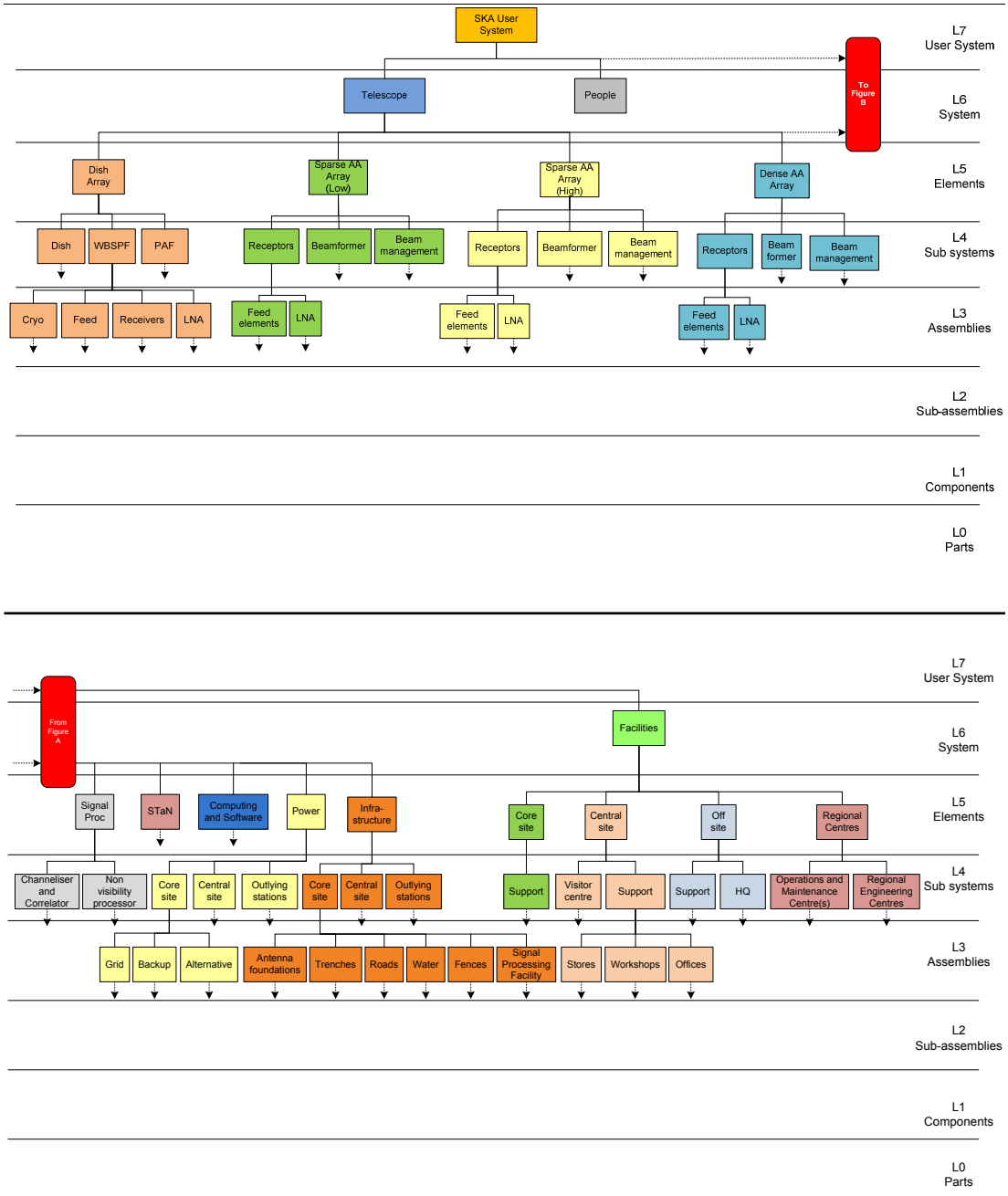


Figure 3.5: Example SKA system hierarchy. Source: Stevenson (2011c).

2011c); a draft WBS has been developed for the SKA (see SKA Drafting Groups, 2012). The number of levels (the depth) of the PBS and WBS will depend on the amount of detail in the design concept, which in turn depends on how far the project has progressed through the design process. The blocks that I model in Chapters 5 and 6 generally sit at the L4 sub-system level in Figure 3.5.

3.4.2.1 Cost models

The systems view encapsulated by the PBS is a useful basis to develop the parametric models. Each sub-system is related to their parent and children sub-systems, therefore changes in one sub-system can affect others. The system can be modelled by describing the sub-systems using scalable models, and applying a common set of input variables, or parameters, to these models. A system design is therefore the combination of these inter-linked sub-system models and their input parameters. The advantage of this approach is that rather than producing discrete design solutions, the scalability of parametric models allow many solutions, within the bounds of the models.

Sub-systems from the system hierarchy (Figure 3.5) can be represented by a small number of blocks, where each block describes the quantity and cost scaling relationships of one or more sub-systems. Relatively simple algebraic equations are used to describe the cost of these blocks as a function of one of more variables (parameters). For example, the parametric equation of a particular block may be

$$C_{\text{block}} = C_{\text{fix}} + C_{\text{var}}\, x, \tag{3.1}$$

where C_{block} is the cost of a single 'instance' (occurrence) of the block, C_{fix} and C_{var} are the cost coefficients and x is a scaling parameter.

The cost coefficients may represent products that can be procured, or a cost-scaling value. A procurable component, such as a cable or processing chip, allows for a bottom-up, engineering cost estimate. The cost-scaling values are used for parametric cost estimates, where an estimate is obtained by specifying a unit cost, and multiplying by the quantity of units required. The costs themselves are derived from one or more cost estimates, and reference class and bottom-up cost estimates can be suitable sources of data, as discussed in Section 3.4.2.2.

The total cost C_{total} is simply the summation of the product of the quantity (N) and cost (C) of each block:

$$C_{\text{total}} = N_{\text{blockA}} C_{\text{blockA}} + N_{\text{blockB}} C_{\text{blockB}} + \dots \tag{3.2}$$

If the blocks represent the system hierarchy at more than one level, then the cost a particular block in Equation 3.2 may be the summation of blocks lower in the hierarchy:

$$C_{\text{blockA}} = N_{\text{blockA.A}} C_{\text{blockA.A}} + N_{\text{blockA.B}} C_{\text{blockA.B}} + \dots \tag{3.3}$$

If costs are calculated by starting at the leaves at the bottom of the hierarchy and rolling up the costs at each level, this becomes a bottom-up cost estimate, where the quantities in Equation 3.3 count the number of components.

This approach allows flexibility for a combination of both parametric (Equation 3.1) and

bottom-up (Equation 3.3) cost estimates to be made. Initial models can use only a small number of parametric models, as in this thesis. As the system design progresses, the block can be broken up into sub-blocks, and those blocks modelled.

The total cost in Equation 3.2 is calculated with a set of common input parameters, used for all blocks. The parameters generally describe the physical attributes of the system; for example, the number of dishes in the array and dish diameter. The blocks, combined with the input parameters, are then a model of a particular telescope design. Although these algebraic equations do not capture all the nuances of a design, they do provide a scalable, first-order cost estimate.

The trade-off studies are made by varying one or more of the input parameters. Because these parameters are used by the models in the design blocks, in effect, every variation models a new telescope design with different cost and performance properties.

3.4.2.2 Cost data

The cost data for the parametric models for each sub-system come from a variety of sources, such as new and existing instruments and knowledge, research and development programs, industry trends and industry quotes (ISPA, 2008). The cost data is developed using methodologies similar to Section 3.4.1, but is applied to particular, large sub-systems, rather than the system as a whole:

- The parametric method uses a number of cost data points to develop the parametric equation (cost estimating relationship); it derives the cost coefficients and scaling parameters from a database of historical costs using statistical techniques such as regression analysis.
- In contrast, the reference class method uses expert opinion to develop the parametric equation, where the cost coefficients are extrapolated from the cost of a similar sub-systems.
- The bottom-up method similarly uses expert opinion to develop the parametric equation, but the cost coefficients are built up from individual cost elements.

For the SKA_1-low (Chapter 5) and long baseline (Chapter 6) case studies, I use cost data developed with the reference class and bottom-up methods; these are described in more detail in the relevant chapter. Regardless of the method for developing the parametric equations, normalisation of the cost data is required for accurate estimates (see Section 3.6.1). Also, as is done in the SKA_1-low analysis, a direct comparison of different cost data sources can be made by running the model with the different cost coefficients, but using the same parametrisation. Such sensitivity analyses highlight areas of the design where costs differ, thus are most uncertain and require further investigation.

3.4.2.3 Performance analysis

In addition to the cost analyses, performance analyses contribute to trade-offs and optimisations of system cost-effectiveness. However, performance analysis receives less attention in the systems engineering literature than the cost estimation; this may be because the performance estimation depends on the system under consideration.

Performance analyses, which include simulations and modelling, are used to verify that the system design is likely to meet its requirements (NASA, 2007). But prior to this verification,

performance analyses are also used in trade studies to evaluate design alternatives (Stevenson, 2011c). For the trade studies in this thesis, I have employed performance models based on the contemporary understanding of the scientific performance of the various telescope technologies (from the receptor through to the high-performance computing). In these models I use the top-level performance metrics and figures of merit described in Section 2.2. They are generally sufficient for a high-level analysis of the system, because of the simple relationships that exist for interferometric arrays to describe the performance of the telescope as a function of receptor type and number. As discussed in Section 3.3.1, these metrics of system performance are also compatible as descriptors of the technical requirements. Thus the analysis of changed system performance is, in effect, an analysis of changed technical requirements.

Detailed performance analyses generally require modelling and simulations of the telescope and its response the sky; these activities are more complex and time-consuming and beyond the scope of this work. However, as the design progresses, simulations and accompanying prototypes will be increasingly used for design verification (Stevenson, 2011c). Some examples of the simulation tools presently used to model radio telescope arrays are OSKAR (Dulwich et al., 2009) and Xarray[1], that model the response of beamformed aperture arrays to the sky, and MeqTrees (Noordam & Smirnov, 2010), that models instrumental effects and provides calibration solutions.

3.5 SKACost

This section introduces SKACost as a costing and design tool to model the SKA system and assist with trade-off and optimisation analyses. SKACost implements the complex parametric models described in Section 3.4.2, such that top-level inputs can be specified and performance and cost metrics reported. I have been a key developer of the SKACost and the models therein. SKACost has been custom-developed by a small collaborative team (Ford et al., 2009, 2010) to enable SKA decision-makers to more easily understand the performance–cost trade-offs for the telescope.

3.5.1 Background

SKACost is the synergy of the complementary approaches to performance and cost analysis presented in SKA Memos 92 (Chippendale et al., 2007) and 111 Bolton et al. (2009c). SKA Memo 92 describes SKAcost, a high-level cost and performance modelling tool that mainly focuses on the engineering performance and monetary cost for given system designs and implementations. This tool has the flexibility for investigating technology and design trade-offs. The tool's design is based on building up concept descriptions of the SKA from an elementary signal path, using common sub-systems (Hall, 2004a; Horiuchi et al., 2004). Memo 92 demonstrates, for a putative SKA design composed only of dishes with either single-pixel feeds or phased array feeds, the utility of the tool in comparing design concepts, and making performance-cost trade-offs. This SKAcost tool was used for the performance–cost trade-offs of the dish-based receptors in *Preliminary Specifications for the Square Kilometre Array* (Schilizzi et al., 2007).

[1] http://sites.google.com/site/xarraytool/

SKA Memo 111, a revised version of SKA Memo 93 (Alexander et al., 2007), documents a SKADS benchmark scenario that matches the 'single-pixel feed dishes and dense aperture arrays' mid-frequency SKA concept in *Preliminary Specifications for the Square Kilometre Array*. The Memo 111 approach is based on predominantly bottom-up cost estimates, developed in the SKADS project and captured within a spreadsheet, and some costs from Memo 92.

SKACost incorporates the major aspects of these approaches to performance and cost estimation and embodies the following philosophies:

- flexible, extensible tool capable of trade-off exploration and optimisation
- expandable architecture and scalable designs
- presentation of a common (sub)-system view
- signal path and geographical array zone analysis
- uncertainty and contingency analysis
- engagement with the wider community, and input from domain experts.

The tool is extensible, so other cost-driving metrics can be incorporated. Although monetary costing is the primary goal, the tracking of power consumption and data transmission rates are two such examples of where the tool has been extended. A full description of SKACost is given in Ford et al. (2010). SKACost was used for cost modelling and trade-off analysis in the development of the SKADS-SKA implementation (Faulkner et al., 2010), consisting of low and mid-frequency aperture arrays, and dishes with wideband single-pixel feeds.

3.5.2 SKACost: the costing engine and the telescope designs

SKACost is an interactive calculation engine written in Python to allow quick development and prototyping. The tool is structured with a sharp division between the 'costing engine' and the 'telescope designs': the costing engine performs the mathematical calculations, while the telescope designs hold the parametric equations as a series of hierarchically interconnected sub-system models. The costing engine acts upon a given telescope design, using common input parameters (such as dish diameter) to calculate cost and performance characteristics of the design. The inputs, models and outputs are readily available to the user, whereas the underlying programming complexity of the costing engine is not visible.

Figure 3.6 shows a screenshot of the graphical user interface, displaying the actual model I developed for the SKA long baselines (Chapter 6). The hierarchical blocks are displayed on the left-hand side. These, in combination with some top-level global inputs and an index file, contain all the information to model the system. The components and design blocks are stored as a simple database of XML (extensible markup language) files with snippets of Python code in the design blocks for the performance calculations. The buttons and text on the right-hand side provide further information on the sub-system, and lets the user drill down the system hierarchy.

SKACost also implements a number of strategies to allow for a more complex cost estimation, which is difficult to implement in a spreadsheet. Example features are:

- input parameter variation over a range of values to model changes to cost and performance

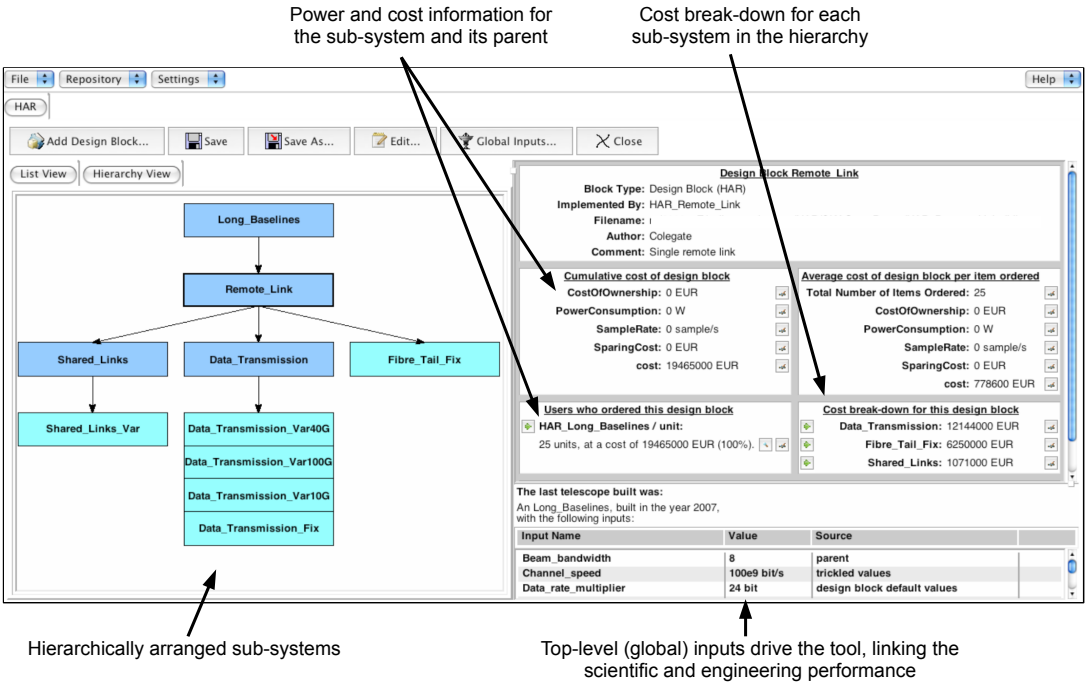


Figure 3.6: A screenshot of SKACost, displaying the model developed for the SKA long baseline data network (Chapter 6).

- Monte Carlo uncertainty analysis
- tracking of non-recurring engineering (NRE) and manufacturing costs
- financial accounting, such as
  - net present value calculations
  - currency movement
  - inflation and other exponential cost improvements, such as Moore's law (Mollick, 2006) or similar
- ranges or sets of permitted values for each parametric equation. If the values fall outside this range, an error message is sent to the user.

For the first-order analyses in this thesis, I only use the Monte Carlo uncertainty analysis (Chapter 5) and input parameter variation (Chapter 6). However, all of these features will be become more relevant as the SKA project progresses.

3.5.3 Previous system analyses

Systems analyses of varying complexity have been used to refine the SKA design to the present high-level system descriptions for SKA_1 (Dewdney et al., 2011a) and SKA_2 (Dewdney et al., 2010b). For example, Weinreb & D'Addario (2001) consider a dish array, with the dishes grouped into phased array stations. A small number of station beams are formed, as input to the correlator. For a fixed A/T, cost as a function of dish diameter or number of antennas per station are explored. Bregman (2004) and Bunton & Hay (2004) optimise systems with multibeam antennas, within A/T and SSFoM targets.

Variants of SKACost have also been used for design investigation activities, using the refined antenna (receptor) concepts presented in the reference design (ISPO, 2006): low and mid-frequency AAs and dishes with wideband single-pixel feeds and phased array feeds. Appendix D collates some examples of trade-offs made with SKACost, that show trends and cost drivers for the telescope, and demonstrate the utility of the performance and cost modelling approach enabled by this custom-developed software. Not all of these types of analysis are made for the case studies in this thesis. However, they serve as examples of future use for the performance and cost models, as the SKA design progresses and further information becomes available.

3.6 Uncertainties

The modelling of uncertainties is an often overlooked aspect of the performance and cost analysis. However, the uncertainties and limitations of the performance and cost modelling manifest themselves as project risks. Early in the project, the uncertainties are large, and as the project progress and the models refined, the uncertainties will be reduced (GAO, 2009). There are multiple sources of uncertainty in the parametric performance and cost models. Some of these uncertainties can only be quantified on a macro basis (i.e. at a system level), while other uncertainties arise from cumulative effects throughout the system.

3.6.1 Sources of uncertainty

Although there are many sources of uncertainty, those of relevance to parametric performance and cost modelling are as follows:

Cost data uncertainties: For the parametric, reference class and bottom-up cost estimation methodologies, there are many sources of errors in the cost data. These are primarily due to data normalisation, cost adjustments and extrapolation from historical data (GAO, 2009). In particular, cost data extrapolated from analogous systems or sub-systems can be require a subjective, 'expert' opinion. As the projects become less similar, the uncertainty of the extrapolation increases (NASA, 2008).

Performance uncertainties: The uncertainties in estimating the performance or functional requirements of a project are not well-defined in cost-estimation and systems engineering handbooks. In the same way as cost estimators for transport infrastructure show a systematic bias in under-estimating costs (Section 3.1), it is not unreasonable to expect that engineers show a bias in under-estimating telescope performance by selecting the ideal performance characteristics. Tools, such as error budgets used in optical telescopes, can track these uncertainties (Duren, 2006). A similar approach may be applicable for the SKA.

Uncertainties in the parametric models: The cost estimating relationships in the parametric models can be a source of uncertainty, where their incorrect use can result in 'serious' estimation errors (NASA, 2008). An example of this is where the models are used outside the range of inputs for which it was designed.

3.6.2 Sensitivity analysis

The parametric model allows for sensitivity analysis of the inputs, to determine which blocks in the system significantly affect the total cost. By identifying these cost drivers, extra attention can be paid to them during the design phase, thereby reducing risk (NASA, 2008). The sensitivity analysis is made by varying one of more input parameters and comparing the results. Sensitivity analyses can also be used to assess the cost impact of changed requirements or performance, again by varying the input parameters.

3.6.3 Uncertainties in SKACost

Because the parametric models are necessarily simple to encompass all the design options, the cost data in the model contains uncertainties that can be modelled explicitly. Each input parameter to the model may also have an associated uncertainty. These uncertainties can be described by probability distribution functions (PDFs). Monte Carlo simulation provides a statistical approach for analysis of these uncertainties, using PDFs for each 'uncertain' input in the parametric model. An input, for example, may be a variable parameter (such as dish diameter) or a unit cost. The simulation then randomly picks values from these input PDFs and calculates a point cost estimate. Over thousands of iterations, this results in a new PDF which is the cost estimate with a statistical distribution (NASA, 2008). Monte Carlo uncertainty analysis has been implemented in SKACost (Chippendale et al., 2007; Ford et al., 2010); Figure 3.7 shows the input PDFs available in SKACost. A simple uncertainty analysis using SKACost is undertaken for SKA_1-low (Section 5.6.3.3).

3.7 Chapter summary

This chapter invoked systems engineering processes to describe the interrelationship between SKA design, performance and cost. Although these processes are only one of many factors in a successful science mega-project, the iterative science and engineering trade-offs are necessary to ensure cost-effective designs are developed for the SKA.

To explore these science and engineering trades, an appropriate modelling framework is required. Section 3.4 presented the general structure and formalisms for parametric performance and cost modelling, and Section 3.5 introduced SKACost as an example tool to implement the modelling. The case studies in the following chapters use top-level measures of performance (Chapter 2), and the modelling framework presented in this chapter, to develop cost-effective design solutions for high time resolution, low-frequency and high angular resolution observations with the SKA.

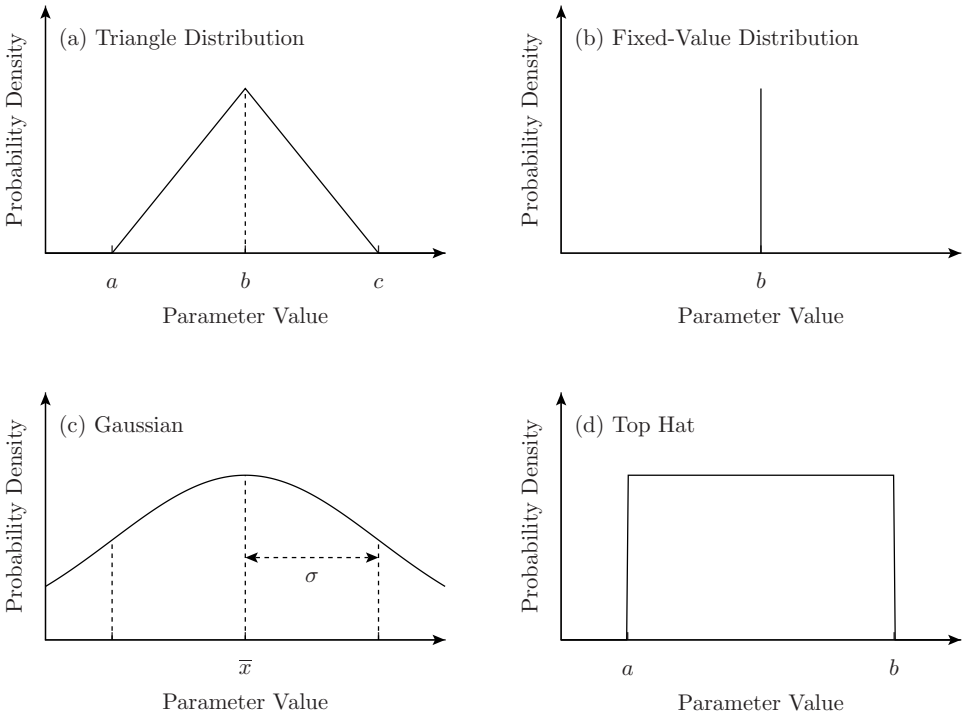


Figure 3.7: Input probability distribution functions (PDFs) available in SKACost. The PDF of a value may be described by (a) minimum a and maximum c possible values, with a peak-likelihood value of b; (b) single-value at b; (c) value with mean $\overline{x}$ and and standard deviation σ; or (d) value ranging between minimum a and maximum b. Source: Ford et al. (2010).

Chapter 4

SKA as a fast radio transients telescope

As the first case study, this chapter presents fast transient survey strategies for exploring high time resolution parameter space, and considers the system design implications of these strategies. The search strategies aim to maximise the probability of intercepting (observing) verifiable fast transient events in a survey. I also determine the comparative cost-effectiveness of these strategies using event rate per beam as a new metric.

4.1 Introduction

To date, the study of pulsars has been the principal motivation for probing high time resolution parameter space at radio frequencies. Pulsar surveys proposed for the SKA (e.g. Cordes et al., 2004b; Smits et al., 2009) involve computationally expensive, systematic searches for radio pulsar populations in the Galaxy, where the observer 'drives' the telescope. These surveys are designed to use pulsar periodicity to improve sensitivity and allow acceleration searches for binary pulsars. In addition to periodicity searches, most present-day pulsar surveys concurrently search for single pulses (e.g. Cordes et al., 2006; Keith et al., 2010). However, pulsar surveys are not necessarily designed to maximise the probability of intercepting fast transient events, hence the motivation for the investigations in this chapter.

I define fast transients as impulsive, singly occurring or intermittent signals, emitted from high energy density events; a search for such events assumes an observed pulse width less than the normal correlator averaging time of a few seconds. In this context, pulsar signals can be classed as periodic fast transients. Other known sources of fast transients at radio frequencies include giant-pulse emitting pulsars (Knight et al., 2006, and references therein), magnetars (Camilo et al., 2006) and rotating radio transients (McLaughlin et al., 2006). Meanwhile, higher energy events, such as annihilating black holes, have been speculated as candidates for extragalactic searches (e.g. Cordes & McLaughlin, 2003; Cordes, 2009a; Macquart et al., 2010b).

Recently, events of potentially extragalactic origin have been discovered by Lorimer et al. (2007), Keane et al. (2011, 2012) and M. Bailes et al. (2012, pers. comm.). These energetic, single-pulse events have been found in surveys undertaken with the Parkes 1.4 GHz multibeam system, and provide tempting glimpses of what may exist in the high time resolution Universe.

However, Parkes, as a single-dish telescope, has only limited ability to verify and localise these detections. Therefore, radio telescope arrays such as the SKA are likely to have an increasingly important role in exploring the high time resolution Universe. Table 4.1 compares Phase 1 of the SKA (SKA_1) to other searches with existing and future telescopes. The metric of event rate per beam $\mathcal{R}_{\mathrm{beam}^{-1}}$ in Table 4.1 is developed in this chapter; it is a measure of the cost-effectiveness of a survey strategy, where a high value of $\mathcal{R}_{\mathrm{beam}^{-1}}$ is desirable. The event rate per beam parametrises the choice of receptor (antenna), the performance, cost and efficiency of the signal combination mode and transients search system, and the observed sky direction.

Table 4.1: Recent radio searches of the high time resolution universe and a comparison of event rate per beam ($\mathcal{R}_{\text{beam}^{-1}}$).

Experiment[a]	Telescope and Status	ν_{centre} (MHz)	$\Delta\nu$ (MHz)	Max. baseline (km)[b]	$\mathcal{R}_{\text{beam}^{-1}}$ (normalised)[c]	Max. beams available
Archival searches[d]	Parkes	N/A	-	-	-	-
Fly's eye fast radio transient search[e]	ATA (completed)	1420	210	N/A	10^{-3} (fly)	42
Effelsberg Northern Sky Pulsar Survey[f]	Effelsberg (operational)	N/A	-	-	-	-
High Time Resolution Universe Pulsar Survey[g]	Parkes (operational)	1352	340	N/A	10^{-2}	13
Pulsar ALFA (PALFA) Survey[h]	Arecibo (operational)	1440	100	N/A	10^{-2}	7
V-FASTR[i]	VLBA (operational)	1400	64	6000	10^{-2} (inc.)	1
LOFAR Transients Key Science Project[j]	LOFAR (in progress)	120	32	< 100	10^{-1} (inc.) 10^{-4} (coh.)	$1^{\dagger}$ thousands†
Commensal Real-Time ASKAP Fast-Transients (CRAFT) Survey[k]	ASKAP (planned)	1400	300	6	10^{-2} (inc.) 10^{-6} (coh.)	36 N/A
SKA_1-low		260	380	200	1 (inc.) 10^{-1} (coh.)	hundreds† thousands†
SKA_1 low band dishes		725	550	200	1 (inc.) 10^{-2} (coh.)	1 thousands†

[a] Only experiments within SKA_1 frequencies (70 MHz – 3 GHz) are listed. Pulsar surveys insensitive to single pulses are excluded. N/A is not applicable or information not available.
[b] For event localisation using triggered buffers, see Section 4.2.
[c] Order of magnitude estimation as per Equation 4.19, normalised to the incoherent combination of SKA_1 low band dishes. For radio telescope arrays, the calculation is for fly's eye (fly), incoherent combination (inc.) or coherent combination (coh.), see Section 4.2. A flat spectrum and no scatter broadening is assumed.
[d] Bagchi et al. (2012); Burke-Spolaor & Bailes (2010); Keane et al. (2010, 2011); Lorimer et al. (2007); McLaughlin et al. (2006).
[e] Siemion et al. (2012). Note only 30 of 42 antennas were ultimately used.
[f] Barr (2011).
[g] Keith et al. (2010).
[h] Cordes et al. (2006); Deneva et al. (2009).
[i] Wayth et al. (2011, 2012).
[j] Dutch LOFAR as in Hessels et al. (2009); van Leeuwen & Stappers (2010). More scenarios are discussed in Stappers et al. (2011).
[k] Macquart et al. (2010a); Macquart et al. (2010b).
[†] Limited by available beamformer processing and data transport.

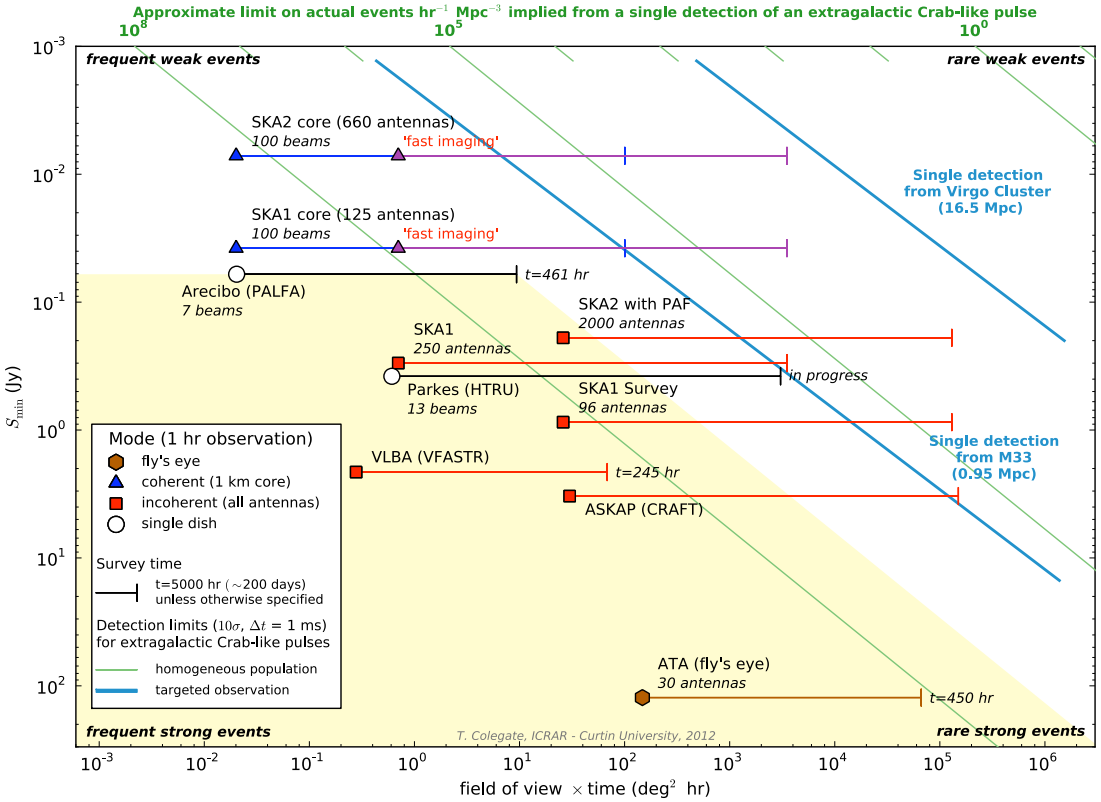


Figure 4.1: [Duplication of Figure 1.3.] High time resolution parameter space for selected telescopes and surveys at ∼1.4 GHz. Sensitivity is plotted on the vertical axis (more sensitive observations towards the top of the figure), while the horizontal axis shows the product of FoV and observing time. For the array telescopes, the fly's eye and coherent and incoherent signal combination modes are shown as relevant. Averaging time Δt is 1 ms, $S_{\rm min}$ is 10σ significance. The green and blue diagonal lines are related to the probability of intercept; they show nominal event rate limits for extragalactic Crab-like pulsars. The yellow shaded region shows the highest sensitivity and event rate limit from a completed survey to date. See text and Section 4.6.1 for further details. See Section 4.6.1 for further details.

The probability of intercept (PoI), introduced in Chapter 1, must also be high enough to be of scientific benefit and open new volumes of parameter space. Figure 1.3, duplicated here as Figure 4.1, shows the parameter space available to selected telescopes and surveys at 1.4 GHz. PoI increases (towards the top-right) for more sensitive surveys, when more beams and a larger beam field of view (FoV) is searched, and for longer total survey times.

Both event rate per beam and PoI are a function of frequency-dependent factors such as minimum detectable flux density, FoV, pulse luminosity and scatter broadening. This chapter determines the cost-effectiveness of fast transient survey strategies, taking into account these frequency-dependent effects. The analysis is applied to the SKA_1 system description to determine optimal search strategies, but the modelling framework is equally applicable to SKA_2 and other radio telescope arrays.

Section 4.2 sets out a high-level 'use case' for searching for fast transients with low frequency aperture arrays and low band dishes. Section 4.3 presents the event rate per beam metric for

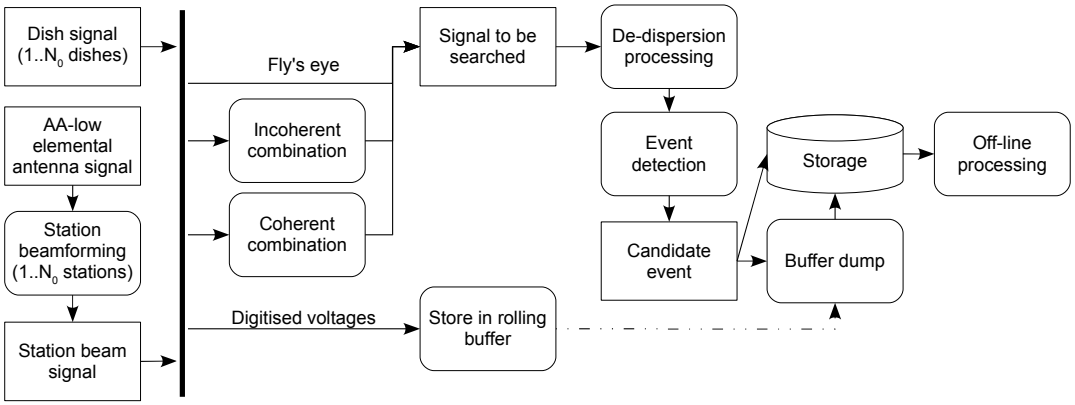


Figure 4.2: High-level flow diagram for a generic fast transient pipeline, for SKA_1 receptors. Rounded boxes are signal processing actions, rectangles describe the information flow. Some actions may be conducted in a different order to maximise the potential for detections. The solid vertical line is the data spigot point in the signal chain (see text for details).

comparing the cost-effectiveness of survey strategies. Section 4.4 makes a detailed analysis of the impact of receptor choice, signal combination modes, sky direction, observing frequency and bandwidth on the event rate per beam; Section 4.5 discusses these results. Section 4.6 considers preferable search strategies, using an example population of Crab-like pulsars to place the extragalactic searches in an astronomical context. Section 4.6 also presents a low-cost fast transient search system for SKA_1, based on the results from this chapter. Finally, Section 4.7 makes specific recommendations for SKA_1.

4.2 Use case for fast transient searches

The specific implementation of a search for fast transients depends on the target or expected source population, performance and cost factors of components in the processing pipeline, and the signal processing architecture. To provide context for subsequent optimisation of survey strategies, in this section I outline a generic fast transient search use case for SKA_1. For specific implementations on other telescope arrays, see Macquart et al. (2010b), Stappers et al. (2011) and Wayth et al. (2011), for ASKAP, LOFAR and the VLBA respectively.

Because the data volumes are too large to store cost-effectively[1], in this use case a real-time search for transient events is conducted on a data stream which is a continuous observation of the sky. However, a rolling buffer recording a small period of data as it is observed allows candidate events containing potential fast transient detections to be saved and further processed off-line. Figure 4.2 shows a generic fast transient pipeline and the signal processing steps (actions) in this pipeline are as follows:

Signal reception: Radio signals are collected by the two SKA_1 receptor types: low frequency (70–450 MHz) aperture arrays (AA-low) and a dish array equipped with low (0.45–1 GHz) and

[1]For example, $\sim$ 20 GB per station beam per second would need to be written to storage for SKA_1-low (380 MHz bandwidth $\times$ 2 Nyquist $\times$ 4 bits $\times$ 50 stations).

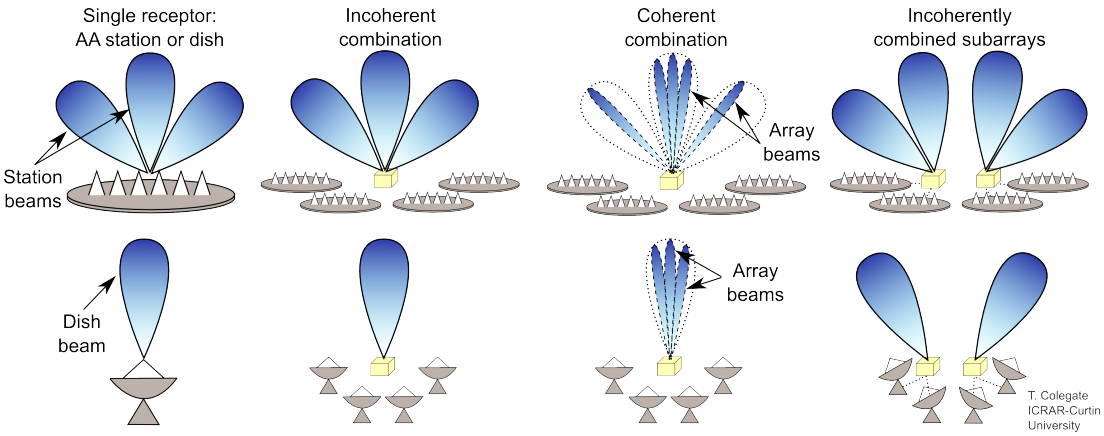


Figure 4.3: Signal combination modes, resultant beam patterns, and beam terminology for dishes and aperture array (AA) stations (beam sizes not to scale).

high (1–2 GHz) band single-pixel feeds. Only the low band feed on the dishes (hereafter referred to as low band dishes) is discussed in this chapter, because it generally achieves a higher event rate than the high band feed, unless the pulse scatter broadening is significant compared the integration time (see Section 4.4.3). SKA_1-survey (dishes with phased array feeds) is not considered in detail because, on a *per beam* basis, it is a subset of the single-pixel feed dish array.

Station beamforming: The complex signals of many SKA_1-low elemental antennas are combined vectorially to form one or more station beams, as shown in Figure 4.3. The output station beam can then be processed in the same way as a dish signal. Though not considered here, signals from a group of dishes (instead of elemental antennas) can similarly be coherently combined into dish station beams. Note that both dish and AA stations are simply groups (subarrays) of coherently combined antennas.

Signal combination: The signals from the dish or station beams can be combined coherently, incoherently or not at all. These signals are then searched for fast transients. Figure 4.3 shows the signal combination modes and resultant beam patterns; they are further discussed in Appendix E.1. Incoherent (phase insensitive) combination sums the detected signals (powers) from receptors pointing in the same direction. Coherent combination of receptor signals forms a phased or tied array beam—voltages measured at each receptor are aligned in phase towards a specific direction on the sky, in a manner similar to station beamforming. Smaller groups of receptors—subarrays—can be incoherently combined and each subarray pointed in a different direction. The extreme of this is the so-called fly's eye mode, where every receptor is pointed in a different direction.

Event localisation and the spatial discrimination of astronomical signals from radio frequency interference (RFI) is possible for coherent combination and, using buffered voltages, for incoherent combination and subarrays of three or more incoherently combined receptors. Multiple beams (incoherently or coherently combined) can also be used to discriminate RFI, where a candidate event in most or all beams indicates the presence of RFI.

Dedispersion processing: The signals pass through a cosmic medium of unknown dispersion measure (DM). This means that the detection is trialled for many DMs, each of which has a computational cost. The DM range to be trialled depends on the location on the sky. Clarke et al. (2011) discusses dedispersion for SKA_1 in detail.

Event detection: An event detection algorithm is applied to the signal from each trial DM, where optimal detection is achieved with an appropriate matched filter (Cordes & McLaughlin, 2003).

Store in rolling buffer: The digitised voltages from the dishes or stations are stored in a circular memory (rolling) buffer. In the case of a candidate event, the data from the buffer can be saved to another location (dumped) and processed off-line. The amount of memory required in the buffer depends on the sampling rate, sample size and the expected maximum (dispersed) pulse duration. The maximum pulse duration is a function of the range of frequencies to be captured and varies linearly with the maximum DM to be trialled (see Appendix E.7). For a maximum DM of $4000\,\mathrm{pc\,cm^{-3}}$ and a bandwidth of a few hundred MHz, a buffer of order a minute is required for dish frequencies (e.g. 0.45–1 GHz). For lower frequencies (e.g. 0.1–0.45 GHz), this increases to tens of minutes.

Buffer dump and off-line processing: On receipt of a trigger, the buffer will dump the original voltage data to storage for off-line processing, which could include RFI filtering, analysis of the candidate detection and correlation of the dish or station beams for source localisation and imaging.

Commensal and targeted surveys: A commensal survey greatly increases observation time by conducting the survey in parallel with normal telescope operations. It is passive; it uses dish or station beam signals from the primary user observation, placing little extra demand on the telescope. Such a survey is suitable for extragalactic searches, given the information about the population of such fast transients is not known *a priori*; hence one direction on the sky is as good as another. To observe specific areas of the sky, such as the Galactic plane and nearby galaxies, a targeted transients survey (which is the primary user observation) may be required.

Data spigot: A data spigot to the dish and station beam signals is useful for transients surveys, especially those which are commensal. If the signal chain is considered to be the signal path from the antennas of a radio telescope array to the correlator, a spigot defines a point in the signal chain where users can tap off data via a well defined interface. The spigot for fast transient searches may output either coherent (phase-preserved) data at high rates or, alternatively, incoherent data where the dish or station beam voltages are squared and integrated to a time resolution of order milliseconds to reduce the data rate and subsequent dedispersion processing load. The latter approach is being taken by the CRAFT survey to access beams from the ASKAP beamformer (Macquart et al., 2010a). The solid vertical line in Figure 4.2 shows the point in the flow diagram where the spigot for fast transients would be placed to enable the signal combination modes in this chapter.

The pipeline after the spigot point is not part of the normal imaging mode of the telescope. Based on the results in this chapter, Section 4.7.1 describes a representative low-cost fast

transient processing system for SKA_1 suitable for commensal observations, where signals from a data spigot are incoherently combined and searched, and voltages buffered for coherent follow-up of candidate events.

The post-spigot pipeline may be implemented internally or with user-provided processing. An example of processing being implemented internally is the 'non-imaging processing' for pulsar observations with SKA_1 (Dewdney et al., 2010a); the approach being taken by CRAFT is an example of user-provided processing. Note that a spigot to the digitised SKA-low elemental antenna signals could conceivably be implemented, but the data rates make this option prohibitively expensive for SKA_1.

4.3 Survey strategies

Suitable measures of performance are required to describe the probability of intercepting fast transients, and hence the number of events detected in a survey. The survey speed figure of merit (SSFoM), which measures the speed at which an *area* of sky is surveyed to a certain sensitivity, is typically used to determine the effectiveness of surveys of steady sources (Cordes, 2009a). However, SSFoM is an inadequate metric for fast transients, because the trade-off between integration time and sensitivity (A/T) cannot be made for short-duration events (see Section 2.2.2). For example, SSFoM does not differentiate between the dissimilar rates of detection for incoherent combination and fly's eye modes (Macquart, 2011).

This section develops measures of performance based on the number of events detectable in a *volume* of sky, and the processing cost of sampling the sky or searching the data. It draws on the event rate calculations in Macquart (2011); the derivation is shown in Appendix E.2.

4.3.1 Event rate

The simplest case is an extragalactic survey: a search, away from the Galactic plane, for a homogeneously distributed population of isotropically emitting fast transients of fixed intrinsic luminosity. For such a population, the total event rate observable by the telescope is

$$\mathcal{R} = \frac{1}{3}\rho_i \Omega_{\rm proc} \left(\frac{W_{\rm i}}{W}\right)^{3/4} \left(\frac{\mathcal{L}}{4\pi S_{\rm min}}\right)^{3/2} \quad \text{events}\,\text{s}^{-1}, \tag{4.1}$$

where ρ_i ($\text{events}\,\text{s}^{-1}\,\text{pc}^{-3}$) is the intrinsic event rate per unit volume, $\Omega_{\rm proc}$ is the processed FoV (instantaneously available to search), $W_{\rm i}$ is the intrinsic pulse width, W is the observed pulse width, $\mathcal{L}$ ($\text{Jy}\,\text{pc}^2$) is the intrinsic luminosity of the population and $S_{\rm min}$ is the minimum detectable flux density of the telescope for an integration time of $\tau = W_{\rm i}$. The $(W_{\rm i}/W)^{3/4}$ term approximates the loss in signal-to-noise (S/N) due to pulse broadening (see Appendix E.2).

Rather than having a fixed intrinsic luminosity, the population is likely to follow some luminosity distribution, thus affecting the event rate. However, Equation 4.1 is sufficient to compare telescopes and their signal combination modes in the present analysis; Macquart (2011) shows that the event rate remains proportional to $\Omega_{\rm proc}$ and $S_{\rm min}^{-3/2}$ for an extragalactic population with a luminosity distribution that follows a power-law or lognormal distribution, where the maximum luminosity in the distribution is much greater than the minimum luminosity.

For fast transient searches within the Galaxy, scatter broadening due to multipath propagation in the interstellar medium makes the event rate per beam dependent on frequency and direction. Although comprehensive direction-dependent modelling is beyond the scope of this thesis, the loss in sensitivity due to scattering is incorporated in the W term in Equation 4.1 and modelled for some representative sky directions in Section 4.4.3; see Section 4.6.3 for further discussion of Galactic objects.

4.3.2 Cost-effective surveys

Given each search strategy has a different processing cost, a simple FoM to measure the cost effectiveness of a fast transient search strategy is *event rate per unit cost* ($\mathcal{R}_{\mathrm{cost}^{-1}}$). Put simply, one search strategy may have a higher total rate of detection than another, but the processing cost of the strategy also needs to be considered. Although signal and search processing costs are architecture specific, they are mainly performed on a 'per beam' basis. In the absence of sufficiently accurate design for costing, I generalise the problem and parametrise the effectiveness of a search strategy with a new metric, *event rate per beam formed and searched* ($\mathcal{R}_{\mathrm{beam}^{-1}}$), as a proxy for cost. It assumes $\mathcal{R}_{\mathrm{cost}^{-1}} \propto \mathcal{R}_{\mathrm{beam}^{-1}}$, which is valid when cost increases linearly with the number of beams (independent FoVs) formed and searched. This is true for first-order beamforming and data transport costs for the SKA (Chippendale et al., 2007; Faulkner et al., 2010). The search costs also increase linearly because each beam signal is searched individually; I consider the efficiency gain from using a single processing unit to process multiple beams to be a second-order effect.

The total event rate $\mathcal{R}$ (Equation 4.1) is calculated for searches with processed FoV Ω_{proc}, being the product of the number of beams N_{beam} and the FoV of each beam. The event rate per beam is thus

$$\mathcal{R}_{\mathrm{beam}^{-1}} = \frac{\mathcal{R}}{N_{\mathrm{beam}}}, \tag{4.2}$$

where N_{beam} may be the number of receptor ($N_{\mathrm{b\text{-}0}}$) or array ($N_{\mathrm{b\text{-}arr}}$) beams (see Appendix E.1). For the independently pointed, incoherently combined subarrays in Figure 4.3, N_{beam} is the product of the number of subarrays and receptor beams formed in each subarray ($N_{\mathrm{beam}} = N_{\mathrm{sa}} N_{\mathrm{b\text{-}0}}$) .

4.3.3 FoV–time product

Although cost-effective survey strategies are important, if the goal is maximise the total number of events detected, then survey time becomes an additional tool in the survey strategy toolbox. The total number of events detected in a survey of time T_{tot} is

$$N_{\mathrm{det}} = \mathcal{R} T_{\mathrm{tot}}. \tag{4.3}$$

Deneva et al. (2009) use an approach similar to Equation 4.3 to compare the number of detections expected for surveys by different telescopes of the same putative population.

Alternatively, Equation 4.3 describes the probability of intercept for a survey. For example, Figure 4.1 shows the sensitivity, FoV and survey time required to detect a single event. Re-

arranging Equation 4.3 gives the FoV–time product required to observe a single event:

$$\Omega_{\mathrm{proc}} T_{\mathrm{tot}} = \frac{3}{\rho_i} \left(\frac{W}{W_{\mathrm{i}}} \right)^{3/4} \left(\frac{4\pi S_{\mathrm{min}}}{\mathcal{L}} \right)^{3/2} . \tag{4.4}$$

Equation 4.4 assumes that the field of view Ω_{proc} observed over time T_{tot} remains pointed within the region of interest (be it the whole sky or a smaller, targeted region).

4.4 Modelling event rates

Using the SKA_1 system description (Table 4.2), this section models event rate $\mathcal{R}$ and event rate per beam $\mathcal{R}_{\mathrm{beam}^{-1}}$ for the low-frequency aperture arrays (SKA_1-low) and the dish array with single-pixel feeds. First, Section 4.4.1 lists the simplifying assumptions used for the trade-offs. Section 4.4.2 shows how signal combination mode and filling factor (spatial density of the receptor layout) impact $\mathcal{R}_{\mathrm{beam}^{-1}}$ for a radio telescope array. I then make a first-order analysis of how SKA_1 will perform as a function of frequency and direction (Section 4.4.3), and consider the effectiveness of searching large bandwidths (Section 4.4.4).

4.4.1 Assumptions

The trade-offs in this section make the following simplifying assumptions:

- The population of fast transients is homogeneously spatially distributed and of fixed intrinsic luminosity.
- A matched filter is used to detect the dedispersed, but scatter broadened pulse (as per Cordes & McLaughlin, 2003).
- The effects of scintillation on source intermittency and optimum search bandwidth are ignored.
- The dedispersion processing system does not contribute to pulse broadening (see Appendix E.3 for a description of these instrumental contributions).
- The intrinsic pulse width is 1 ms. Shorter duration pulses would be more sensitive to S/N loss due to pulse broadening, longer duration pulses would be less sensitive.
- Events are broad-band such that the intrinsic spectral bandwidth of the pulse is greater than the processed bandwidth. Thus all channels across the band contain contributing signal.
- The beam has constant (maximum) sensitivity between the half-power beamwidth points, and zero sensitivity outside of that.
- Beamformer calibration costs are not considered.
- A time to frequency domain transformation (channelisation) and cross-correlation 'FX' correlator is used (see Section 2.3.3.3).
- The processing cost of forming and searching a beam is independent of frequency, bandwidth and signal combination modes. In practice, lower frequencies (where the maximum dispersed pulse duration is longer) and larger bandwidths will increase processing costs; the magnitude of the increase is specific to the processing architecture and the effect of sky direction on the DM range to be trialled.

Table 4.2: SKA_1 system details, as per Dewdney et al. (2011a).

Low frequency aperture arrays (SKA_1-low)	
Aperture	
Frequency range[a]	70–450 MHz
Station diameter	180 m
Number of stations	50
Number of antennas	11 200 per station
Station beam taper	1.3
Dense–sparse transition	115 MHz (2.6 m)
Array configuration regions[b]	
Core (radius<0.5 km)	$\sim$ 50 % (25 stations)
Inner (1<radius<2.5 km)	$\sim$ 20 % (10 stations)
Mid (2.5<radius<100 km)	$\sim$ 30 % (15 stations)
Core filling factor	0.81
Performance	
Receiver temperature	150 K
Bandwidth per beam	380 MHz
Single pixel feed dishes	
Aperture	
SKA_2 dish frequency capability	0.3–10 GHz
Parabolic dish diameter	15 m
Number of dishes	250
Total physical aperture	44 179 m^2
Dish illumination factor	1.15
Array configuration regions[b]	
Core (radius<0.5 km)	$\sim$ 50 % (125 ant.)
Inner (0.5<radius<2.5 km)	$\sim$ 20 % (50 ant.)
Mid (2.5<radius<100 km)	$\sim$ 30 % (75 ant.)
Core filling factor	0.03
Antenna RF system[c]	
Feed/LNA low band	0.45–1.0 GHz
Bandwidth ($\Delta\nu_{\rm low}$)	0.55 GHz
Feed/LNA high band	1.0–2.0 GHz
Bandwidth ($\Delta\nu_{\rm high}$)	1.0 GHz
Performance	
Antenna/feed efficiency[d]	70 %
Average $T_{\rm sys}$ in low band[e]	$\sim$40 K
Average $T_{\rm sys}$ in high band	$\sim$30 K

[a] Single dual polarisation antenna over frequency range
[b] Fractional number in each region.
[c] One dual polarization feed available at a time.
[d] Average over frequency.
[e] Higher at the low frequency end of this band.

4.4.2 Signal combination mode comparisons

This sub-section compares the per beam event rate for incoherent and coherent combination and fly's eye (Figure 4.3 and Appendix E.1), and applies these results to SKA_1 aperture arrays and low band dishes. A single dish or AA station receptor is designated with the subscript 0. The combination of an array of N_0 receptors may refer to the number of receptors in the total array, or some subset of the total array, such as the SKA_1 core region in Table 4.2

The effect of signal combination mode on sensitivity and FoV are as follows:

Incoherent combination An array of N_0 incoherently combined receptors increases sensitivity by a factor of $\sqrt{N_0}$ over a single receptor while retaining its FoV, Ω_0. Forming $N_{\mathrm{b\text{-}0}}$ station beams increases processed FoV as $\Omega_{\mathrm{proc}} = N_{\mathrm{b\text{-}0}}\Omega_0$.

Incoherently combined subarrays To further increase the FoV, N_{sa} subarrays of incoherently combined receptors can be formed. Each subarray is pointed in a different direction, increasing the FoV by a factor of N_{sa}, but only increasing the sensitivity of the array by a factor of $\sqrt{N_{0/\mathrm{sa}}}$ over a single receptor, where $N_{0/\mathrm{sa}}$ is the number of receptors per subarray. Fly's eye pertains to the case $N_{0/\mathrm{sa}} = 1$.

Coherent combination An array of N_0 coherently combined receptors is more sensitive than the incoherent combination and subarraying modes; sensitivity increases proportional to N_0. However the FoV of the array beam, Ω_{arr}, is much smaller; it is proportional to D_{arr}^{-2}, where D_{arr} is the diameter of the array of receptors being combined. Forming $N_{\mathrm{b\text{-}arr}}$ array beams increases processed FoV as $\Omega_{\mathrm{proc}} = N_{\mathrm{b\text{-}arr}}\Omega_{\mathrm{arr}}$.

Applying these relationships to Equation 4.1, and expanding S_{min}, gives the total event rate for each signal combination mode:

$$\begin{aligned}\mathcal{R} =& \frac{1}{3}\rho_i \left(\frac{W_{\mathrm{i}} N_{\mathrm{pol}} \Delta\nu\tau}{W}\right)^{3/4} \left(\frac{\mathcal{L} A_{\mathrm{e\text{-}0}}}{4\pi\sigma 2 k_{\mathrm{B}} T_{\mathrm{sys}}}\right)^{3/2} \mathcal{M},\\ \mathcal{M} =& \begin{cases} N_{\mathrm{b\text{-}0}}\Omega_0 N_0^{3/4} & \text{Incoherent combination} \\ N_{\mathrm{b\text{-}arr}}\Omega_{\mathrm{arr}} N_0^{3/2} & \text{Coherent combination} \\ N_{\mathrm{sa}}^{1/4} N_{\mathrm{b\text{-}0}}\Omega_0 N_0^{3/4} & \text{Subarraying,} \end{cases}\end{aligned} \tag{4.5}$$

where N_{pol} is the number of polarisations summed, $\Delta\nu$ is the processed bandwidth, τ is the post-detection integration time (which also defines the time resolution of the observation), $A_{\mathrm{e\text{-}0}}$ is the effective area of a receptor (dish or station), σ is the S/N ratio required for event detection and T_{sys} is the system temperature.

4.4.2.1 Filling factor efficiency

The coherent combination mode is more effective if the dishes or stations are closely spaced, thus achieving a higher filling factor. In this case, the same number of receptors are being combined, but the array beam FoV is larger. D'Addario (2010) considers the number of coherently combined array beams required to achieve an event rate FoM equivalent to one incoherently combined beam for ASKAP. I modify this analysis and apply it to Equation 4.5.

Following Cordes (2009a), I define the number of pixels (N_{pix}) as the maximum number of independently pointed, coherently combined array beams that can be formed within the FoV

of a single receptor beam. It is frequency independent, and given by

$$\begin{aligned} N_{\rm pix} &= \frac{\Omega_0}{\Omega_{\rm arr}} \\ &= \left(\frac{\mathcal{K}_0}{\mathcal{K}_{\rm arr}} \frac{D_{\rm arr}}{D_0} \right)^2, \end{aligned} \tag{4.6}$$

where $\mathcal{K}_0$ and $\mathcal{K}_{\rm arr}$ are the receptor and array beam tapers respectively.

A measure of the effectiveness of the coherent combination mode is the number of array beams which need to be formed and searched to achieve a coherent combination event rate ($\mathcal{R}_{\rm coh}$) equal to that of incoherent combination ($\mathcal{R}_{\rm inc}$). From Equation 4.5, $\mathcal{R}_{\rm coh} = \mathcal{R}_{\rm inc}$ results in

$$\begin{aligned} N_{\rm b\,-\,arr} &= \frac{N_{\rm b\,-\,0} \Omega_0 N_0^{3/4}}{\Omega_{\rm arr} (\eta_0 N_0)^{3/2}} \\ &= \frac{N_{\rm b\,-\,0} N_{\rm pix}}{\eta_0^{3/2} N_0^{3/4}}, \end{aligned} \tag{4.7}$$

where η_0 is the fraction of receptors in the array that are coherently combined, out of a total N_0. For example, $\eta_0 = 0.5$ if only the receptors in the $\rm SKA_1$ core are coherently combined while those in the total array are incoherently combined.

Achieving the highest possible event rate is desirable, but this must be tempered by the cost of searching multiple beams. The relative event rate per beam depends on the array filling factor and is simply the inverse of Equation 4.7 when $N_{\rm b\,-\,0} = 1$:

$$\mathcal{R}_{\rm coh\,beam^{-1}} = \frac{\eta_0^{3/2} N_0^{3/4}}{N_{\rm pix}} \mathcal{R}_{\rm inc\,beam^{-1}}. \tag{4.8}$$

A higher filling factor is achieved with a smaller $D_{\rm arr}$, in which case $N_{\rm pix}$ is lower and the effectiveness of the coherent combination improves relative to incoherent combination.

For incoherently combined subarrays (when all N_0 receptors are formed into subarrays),

$$\mathcal{R}_{\rm sa\,beam^{-1}} = N_{\rm sa}^{-3/4} \mathcal{R}_{\rm inc\,beam^{-1}}, \tag{4.9}$$

where the beams of $N_{\rm sa}$ independently pointed subarrays are searched.

4.4.2.2 Coherent combination for a fully filled array

Assuming a best case scenario of an array entirely filled with stations of equal diameter and $\mathcal{K}_0 = \mathcal{K}_{\rm arr}$,

$$N_{\rm pix} = N_0. \tag{4.10}$$

Substituting into Equation 4.7 and for $\eta_0 = 1$, the theoretical minimum number of array beams required so that $\mathcal{R}_{\rm coh} = \mathcal{R}_{\rm inc}$ is

$$N_{\rm b\,-\,arr} = N_{\rm b\,-\,0} N_0^{1/4} \tag{4.11}$$

and the relative event rate per beam (Equation 4.8) is

$$\mathcal{R}_{\rm coh\,beam^{-1}} = N_0^{-1/4} \mathcal{R}_{\rm inc\,beam^{-1}}. \tag{4.12}$$

Table 4.3: Relative event rates for extragalactic observations with selected signal combination modes for SKA_1 receptors.

Signal combination mode	Input parameters[a]		Calculated values		
	D_{arr} (km)	N_0	N_{pix}	Minimum $N_{b\text{-}arr}$ required for $\mathcal{R}_{coh} \geq \mathcal{R}_{inc}$ [b]	Relative event rate per beam ($\mathcal{R}_{beam^{-1}}$)[c]
Aperture array receptors					
Incoherent: total array	-	50	-	-	1
Coherent: core	1	25	31	$5N_{b\text{-}0}$	2.15×10^{-1}
Coherent: inner + core	5	35	772	$70N_{b\text{-}0}$	1.43×10^{-2}
Coherent: total array	200	50	1.23×10^6	$6.57 \times 10^4 N_{b\text{-}0}$	1.52×10^{-5}
Fly's eye: total array	-	$N_{sa} = 50$	-	-	5.32×10^{-2}
Low band dish receptors					
Incoherent: total array	-	250	-	-	1
Coherent: core	1	125	3.49×10^3	157	6.37×10^{-3}
Coherent: inner + core	5	175	8.72×10^4	2.37×10^3	4.22×10^{-4}
Coherent: total array	200	250	1.40×10^8	2.22×10^6	4.50×10^{-7}
Fly's eye: total array	-	$N_{sa} = 250$	-	-	1.59×10^{-2}

[a] From Table 4.2.
[b] Minimum number of array beams $N_{b\text{-}arr}$ required for a coherent combination event rate $\mathcal{R}_{coh} \geq \mathcal{R}_{inc}$, the incoherent combination of the total array. For the dishes, $N_{b\text{-}0} = 1$.
[c] Relative to $\mathcal{R}_{inc\,beam^{-1}}$, the event rate per beam for incoherent combination.

This estimation is optimistic towards coherent combination, given that it is not physically possible to entirely fill a circular array with circular stations. Regardless of this, for $N_0 > 1$, *the incoherent combination will always achieve a higher event rate per beam searched than coherent combination, and this difference increases with* N_0.

4.4.2.3 Signal combination modes for SKA_1

A flexible processing system allows various signal combination modes; Table 4.3 compares the event rate for incoherent and coherent combination and fly's eye, for SKA_1 receptors. Table 4.3 calculates the number of pixels (array beams) required to fill the receptor beam FoV, and the number of array beams required to ensure a coherent combination event rate higher than the incoherent combination of the total array ($\mathcal{R}_{coh} \geq \mathcal{R}_{inc}$). For all modes, the event rate per beam formed and searched is calculated relative to incoherent combination.

Table 4.3 shows three coherent combination modes: receptors in the core region, inner and core region and the total array. Of these, using the receptors in the core achieves the highest event rate per beam, due to the higher density of collecting area. For this reason, the core will be the only coherent combination mode further analysed in this chapter.

The coherent combination of the SKA_1-low core requires approximately five array beams to equal the event rate of a single beam of the incoherently combined total array. If the dense packing is not achievable (Dewdney et al., 2010a), the number of array beams required will be higher. Indeed, using optimisations from Graham et al. (1998), the optimal packing of 25

congruent circles of diameter 180 m in a circle results in a minimum core diameter of 1036 m. This is larger than the 1000 m diameter in Dewdney et al. (2010a) and excludes any spacing that may be required for infrastructure.

Due to the lower filling factor, 157 beams formed from the coherent combination of the dishes in the core are required to equal the event rate of the incoherently combined array. On a *per beam* basis, Table 4.3 shows that the fly's eye mode is not particularly effective. In terms of total event rate, fly's eye mode is higher than incoherent combination by a factor of $50^{1/4} = 2.6$ for $\mathrm{SKA_1}$-low, and a factor of $250^{1/4} = 4$ for the low band dishes. Although not shown, values for subarrays lie between fly's eye and incoherent combination modes, and depend on the number of receptors per subarray.

4.4.3 Frequency dependence

The event rate $\mathcal{R}$ has a frequency dependence on minimum detectable flux density, FoV, pulse luminosity and scatter broadening. The wide fractional bandwidths of the SKA further complicate the effect on event rate. Looking at each of these dependencies in turn:

- The processed FoV depends on the number of beams formed and whether they are formed incoherently or coherently, but either way is proportional to ν^{-2}:

$$\Omega_{\mathrm{proc}} = \frac{\pi}{4} N_{\mathrm{beam}} \left(\frac{\mathrm{c}\mathcal{K}}{\nu D} \right)^2, \tag{4.13}$$

  where c is the speed of light. For incoherent combination, N_{beam} is the number of station beams formed, $\mathcal{K}$ is the feed illumination factor or station beam taper and D is the diameter of the dish or station. For coherent combination, N_{beam} is the number of array beams formed, $\mathcal{K}$ is the array beam taper and D is the diameter of the array (see Appendix E.1).

- Pulses are broadened due to scattering. The broadening time τ_{d} depends on the path through the Galaxy to the observer, and scales as $\tau_{\mathrm{d}} \propto \nu^{-4.4}$ (Cordes & Lazio, 2002). The observed pulse duration is given by

$$W \approx \sqrt{W_{\mathrm{i}}^2 + \tau_{\mathrm{d}}^2}. \tag{4.14}$$

  See Appendix E.3 for further details.

- Because the search is for an unknown population, the variation of luminosity with frequency is not known. I consider three source populations whose luminosity: (a) does not vary with frequency; (b) varies with $\nu^{-1.6}$, a value typical of the pulsar population (Lorimer et al., 1995); or (c) varies with $\nu^{-3.0}$, consistent with giant pulses from the Crab pulsar (Sallmen et al., 1999). The luminosity at frequency ν is thus

$$\mathcal{L} = \mathcal{L}_0 \left(\frac{\nu}{\nu_0} \right)^{\xi}, \tag{4.15}$$

  where $\mathcal{L}_0$ is luminosity at reference frequency ν_0 and ξ is the spectral index (0, -1.6 or -3).

- S_{min} is a function of T_{sys} and $A_{\mathrm{e-0}}$:

$$S_{\mathrm{min}} \propto \frac{T_{\mathrm{sys}}}{A_{\mathrm{e-0}}}. \tag{4.16}$$

For SKA$_1$-low, the station effective area is approximately

$$A_{\mathrm{e\,-\,0}} = \begin{cases} \frac{\pi}{4} D_{\mathrm{st}}^2 & \nu < \nu_{\mathrm{transition}} \\ N_{\mathrm{e/st}} \times \frac{c^2}{3\nu^2} & \nu > \nu_{\mathrm{transition}}, \end{cases} \tag{4.17}$$

where $\nu_{\mathrm{transition}} = 115\,\mathrm{MHz}$ is the aperture array dense–sparse transition frequency. The SKA$_1$-low system temperature is the sum of the receiver noise and an approximation to the sky temperature:

$$T_{\mathrm{sys}} = T_{\mathrm{rec}} + 60 \left(\frac{c}{\nu} \right)^{2.55}. \tag{4.18}$$

A graphical breakdown of the frequency dependencies of SKA$_1$-low is shown in Appendix E.4.

Figure 4.4 shows the frequency dependence of event rate, for three representative sky directions and spectral indices of $\xi = 0$, -1.6 and -3.0. The normalised event rate per beam ($\mathcal{R}_{\mathrm{beam}^{-1}}$, Equation 4.2) is plotted at 1 MHz intervals for centre frequency ν and processed bandwidth $\Delta\nu = 1\,\mathrm{MHz}$, spanning the SKA$_1$ system description (Table 4.2) frequency range of the aperture arrays and low band dishes ($70\,\mathrm{MHz} \leq \nu \leq 1000\,\mathrm{MHz}$). The actual event rate per beam, for the 1 MHz processed bandwidth, is $6.94 \times 10^{-8} \rho_i \mathcal{L}_0^{3/2} x\,\mathrm{events\,s^{-1}}$, where x is the normalised $\mathcal{R}_{\mathrm{beam}^{-1}}$ (plotted), ρ_i has units of $\mathrm{events\,s^{-1}\,pc^{-3}}$ and $\mathcal{L}_0$ has units of $\mathrm{Jy\,pc^2}$.

In Figure 4.4, the simplest case is a search for extragalactic fast transients (solid line). In this case there is no sensitivity loss due to scatter broadening; $W \gg \tau_{\mathrm{d}}$ is assumed. Cordes & McLaughlin (2003) find that for a given sky direction, the scatter broadening of an extragalactic source will be approximately six times the broadening from the Galaxy alone. This assumes equal scatter broadening in the host galaxy, if there is one, and no contribution from the intergalactic medium. For directions away from the Galactic plane where τ_{d} is low and for an intrinsic pulse width of 1 ms, the exclusion of scatter broadening is a reasonable first-order assumption.

A first-order analysis of a search for Galactic transients is possible by invoking the simplifying assumptions listed in Section 4.4.1. I calculate relative event rates taking into account estimates of scatter broadening at a distance of 30 kpc and frequency of $\nu_0 = 1\,\mathrm{GHz}$ from the NE2001 model of Cordes & Lazio (2002), where the broadening scales as $\tau_{\mathrm{d}} \propto \nu^{-4.4}$. A distance of 30 kpc determines the maximum broadening due to interstellar scattering for that direction; broadening is less at shorter distances. Two representative sky directions for Galactic transients are:

- above the Galactic plane: $\tau_{\mathrm{d}} = 0.03\,\mathrm{ms}$ ($DM = 215\,\mathrm{pc\,cm^{-3}}$) at l = 300, b = 10
- on the Galactic plane: $\tau_{\mathrm{d}} = 1.78\,\mathrm{ms}$ ($DM = 628\,\mathrm{pc\,cm^{-3}}$) at l = 300, b = 0.

For these directions, the normalised $\mathcal{R}_{\mathrm{beam}^{-1}}$ in Figure 4.4 shows how increased scatter broadening reduces the event rate at lower frequencies.

Because the maximum (30 kpc) broadening is applied in this first-order analysis, the effect of scatter broadening on event rate is independent of the signal combination mode chosen; the frequency-dependent $(W_{\mathrm{i}}/W)^{3/4}$ term in Equation 4.5 is common to all modes, as is the frequency-dependent luminosity (Equation 4.15). Therefore, the relative event rate per beam

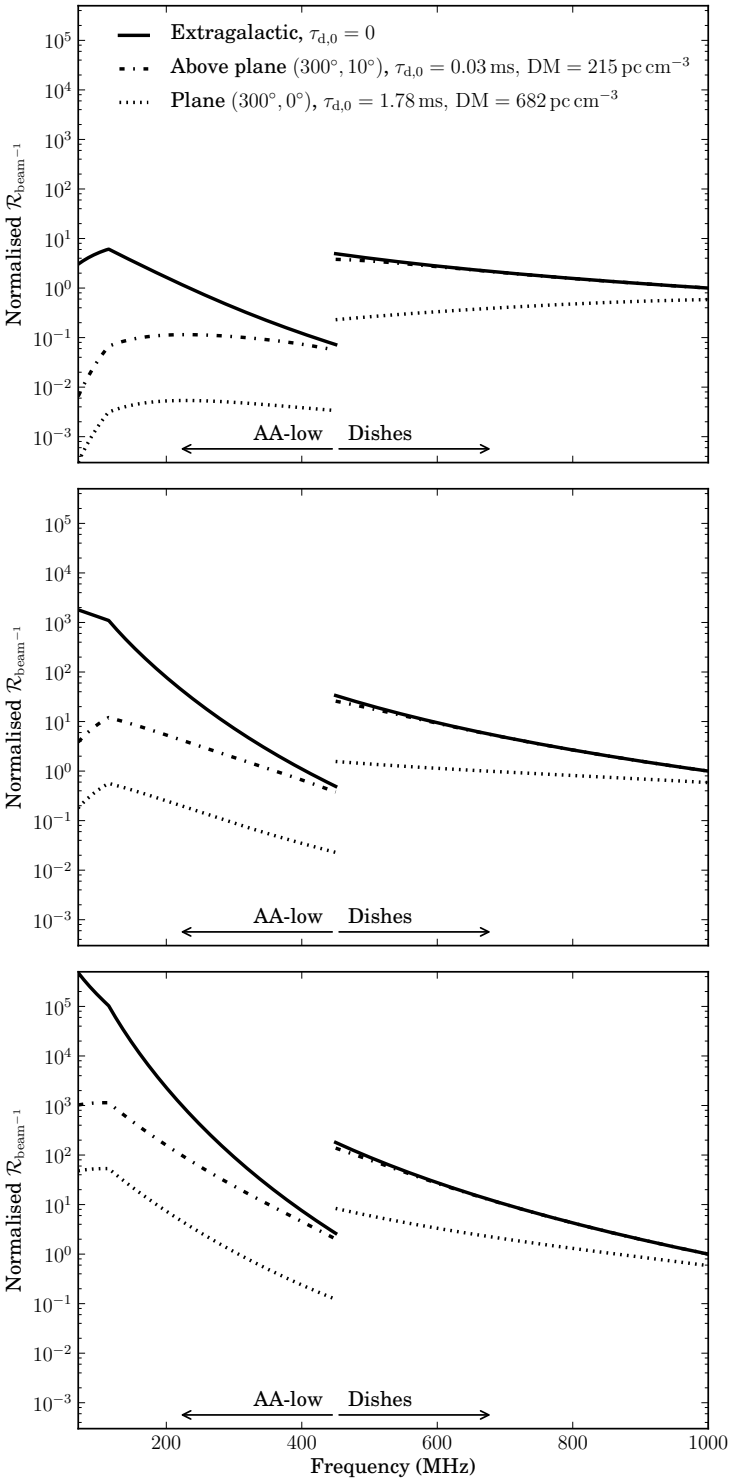


Figure 4.4: Normalised event rate per beam for the *incoherent combination* of the total array for $\Delta\nu = 1\,\mathrm{MHz}$, a post-detection integration time equal to the intrinsic pulse width ($\tau = W_{\mathrm{i}} = 1\,\mathrm{ms}$) and a source spectral index of 0 (top), -1.6 (middle) and -3.0 (bottom), at three representative sky directions. Scatter broadening and DM is calculated at a distance of 30 kpc and frequency of $\nu_0 = 1\,\mathrm{GHz}$ from NE2001 (Cordes & Lazio, 2002) and scaled using $\tau_{\mathrm{d}} \propto \nu^{-4.4}$. Data is normalised to $\mathcal{R}_{\mathrm{beam}^{-1}} = 1$ at ν_0 for the extragalactic case.

between signal combination modes (the rightmost column of Table 4.3) still applies as a multiplicative factor to the data in Figure 4.4, regardless of sky direction and frequency. Thus for Figure 4.4, the rate for the coherent combination of the receptors in the array core is less by a factor of approximately five for SKA_1-low and 157 for the low band dishes. More detailed analyses of Galactic transients searches are discussed in Section 4.6.3.

From Figure 4.4, the incoherent combination of dish signals would be most efficient for searching for extragalactic sources with a low spectral index; SKA_1-low would be more efficient for extragalactic sources with high spectral indices. For a Galactic population, the preferred receptor depends on the amount of scatter broadening. For flatter spectrum sources at directions near or on the Galactic plane, low band dishes show a higher $\mathcal{R}_{\mathrm{beam}^{-1}}$ than SKA_1-low. However, the lower frequencies become increasingly cost-effective for steeper spectrum sources. For steep spectrum sources, even those on the Galactic plane, the higher event rate is at the lower end of the frequency band of each receptor. For shallow spectrum sources near or on the Galactic plane, there is no strong maximum within a receptor frequency band. For dishes, the maximum may be at the low (450 MHz) or high (1 GHz) end of the frequency band, for small or large τ_{d} respectively. For SKA_1-low, the event rate quickly reduces for frequencies below $\nu_{\mathrm{transition}} = 115\,\mathrm{MHz}$, except for directions of very low scattering.

4.4.4 Large processed bandwidths

The strong dependence of event rate on frequency is also important for searches with large fractional bandwidths. In this chapter, processed bandwidth ($\Delta\nu$) refers to the bandwidth of the astronomical signal at the fast transient detection system. For observations of a continuum source, an increase in processed bandwidth is usually assumed to produce a commensurate $\sqrt{\Delta\nu}$ increase in signal-to-noise (Equation 2.1), which would produce a $\Delta\nu^{3/4}$ increase in event rate (Equation 4.5). However, unless the event rate is approximately constant across frequency, this increase does not hold for large processed bandwidths.

To calculate the event rate over a large processed bandwidth, I sum the frequency-dependent channel contributions shown in Figure 4.4. Where pulse broadening (due to propagation and instrumental effects) is not significant, the event rate can be calculated numerically for a processed bandwidth of $\Delta\nu = N_{\mathrm{ch}}\Delta\nu_{\mathrm{ch}}$ using

$$\begin{aligned}\mathcal{R}_{\Delta\nu} &= \left(\sum_i^{N_{\mathrm{ch}}} \mathcal{R}_i^{4/3}\right)^{\frac{3}{4}} \\ &= \frac{1}{3}\rho_i \left(\frac{\mathcal{L}_0}{\nu_0^{\xi}}\right)^{\frac{3}{2}} \left(\sum_i^{N_{\mathrm{ch}}} \left(\frac{\Omega_{\mathrm{proc},i}^{2/3}\nu_i^{\xi}}{S_{\mathrm{min},i}}\right)^2\right)^{\frac{3}{4}},\end{aligned} \tag{4.19}$$

where N_{ch} is the number of frequency channels of width $\Delta\nu_{\mathrm{ch}}$ (see E.5 for the derivation).

Table 4.4 makes these calculations for different modes and spectral indices, for the full bandwidth of AA-low and low band dishes. The relative event rate per beam between signal combination modes of a given receptor (Equations 4.8 and 4.9) still hold for large processed bandwidths. Strikingly, for the combined AA-low and low band dish event rate, the contribution from the dishes is only significant for the incoherently combined total array case when $\xi = 0$. Of course,

Table 4.4: Normalised extragalactic event rate per beam[a] for the full SKA_1 receptor bandwidth.

Receptor	Incoherently combined: total array			Coherently combined: core		
	ξ=-3.0	ξ=-1.6	ξ=0	ξ=-3.0	ξ=-1.6	ξ=0
Low band dish ($\Delta\nu = 550\,\mathrm{MHz}$)	3.83×10^3	989	262	24.4	6.30	1.67
AA-low ($\Delta\nu = 380\,\mathrm{MHz}$)	4.61×10^6	3.13×10^4	164	9.94×10^5	6.74×10^3	35.3
AA-low and low band dish	4.61×10^6	3.15×10^4	361	9.94×10^5	6.74×10^3	35.7

[a] Normalised to $\mathcal{R}_{\mathrm{beam}^{-1}} = 1$ ($\Delta\nu = 1\,\mathrm{MHz}$) for the incoherent combination of dishes at $\nu_0 = 1\,\mathrm{GHz}$. The actual event rate per beam is $6.94\times10^{-8}\rho_i\mathcal{L}_0^{3/2}x$ events s^{-1}, where x is the normalised $\mathcal{R}_{\mathrm{beam}^{-1}}$ shown in the table, ρ_i has units of events $\mathrm{s}^{-1}\,\mathrm{pc}^{-3}$ and $\mathcal{L}_0$ has units of $\mathrm{Jy\,pc}^2$.

the advantage for populations with steep spectral indices will be lessened with increased scatter broadening. But regardless, for coherent combination the dishes contribute little to the per beam event rate.

Given $\Delta\nu = N_{\mathrm{ch}}\Delta\nu_{\mathrm{ch}}$, plotting $\mathcal{R}_{\Delta\nu}$ as a function of the number of contributing channels shows the decreasing contribution of higher frequency channels to the event rate. (The decreasing contribution comes from lower frequency channels in the dense AA regime where $\nu < 115\,\mathrm{MHz}$, and spectral indices are $-1 \lesssim \xi \leq 0$.) Consider two cases for extragalactic searches, plotted in Figure 4.5: the AA-low band from 115 to 450 MHz and the low band dishes where the whole 550 MHz bandwidth is available. The ideal case of a $\Delta\nu^{3/4}$ increase over the $\Delta\nu = 1\,\mathrm{MHz}$ event rate is also shown. As expected, this plot shows that the maximum bandwidth achieves the highest event rate. However, the event rate curve flattens out well before the maximum bandwidth, especially for AA-low and also steeper spectrum sources.

Assuming a limited amount of signal processing is available, the following question arises: at what point could the processing be more effectively used elsewhere in the fast transient pipeline, and how is this quantified? One method is to arbitrarily set a threshold beyond which additional channels contribute very little to the event rate. Channels of increasing frequency are included while the following is true:

$$\frac{\mathcal{R}_{\Delta\nu+\Delta\nu_{\mathrm{ch}}}}{\mathcal{R}_{\Delta\nu}} > \text{threshold}. \tag{4.20}$$

For example, say the threshold is set to 0.5%. Then for the ideal case, $N_{\mathrm{ch}} = 150$ channels contribute to $\mathcal{R}_{\Delta\nu}$. Adding a 151^{st} channel will contribute less than 0.5 % to $\mathcal{R}_{\Delta\nu}$. For the two cases plotted, Table 4.5 shows the maximum channel for which the improvement in event rate over the rate without that channel is greater than 0.5%.

To interpret this table, compare the maximum contributing channel number when $\xi = -1.6$. For AA-low, 47 channels ($\Delta\nu = 47\,\mathrm{MHz}$) contribute above the threshold. For dishes, this is achieved with $\Delta\nu = 87\,\mathrm{MHz}$, implying that the AA-low processed bandwidth becomes less useful more quickly. This is expected, given the steeper spectral dependence of AA-low over dishes, shown in Figure 4.4.

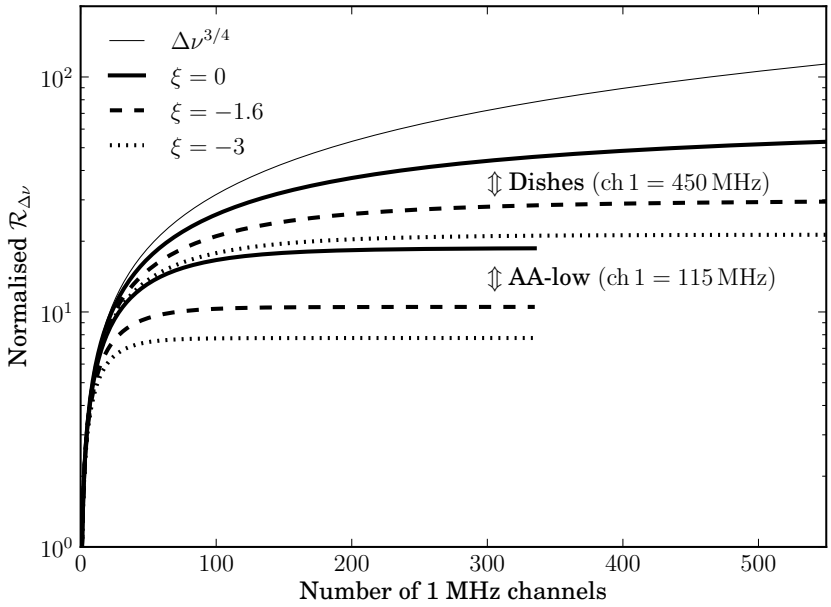


Figure 4.5: Extragalactic event rate as a function of bandwidth, calculated for 1 to 335 AA-low channels of width 1 MHz, where channel 1 is at 115 MHz, and 1 to 550 low band dish channels, where channel 1 is at 450 MHz. Thick lines show the calculated $\mathcal{R}_{\Delta\nu}$ versus processed bandwidth $\Delta\nu = N_{\rm ch}$ MHz, for spectral indices of 0, −1.6 and −3.0. The thin line is the expected $\Delta\nu^{3/4}$ increase in event rate over the rate using only one channel ($\mathcal{R}_{\Delta\nu=1\,\rm MHz}$). The rate for each curve is normalised such that $\mathcal{R}_{\Delta\nu=1\,\rm MHz} = 1$; the slope of each curve applies equally to the incoherent or coherent combination of the receptor.

Table 4.5: Maximum number of channels of bandwidth $\Delta\nu_{\rm ch} = 1$ MHz contributing more than 0.5% of the cumulative event rate.

Case [a]	Channel number		
	$\xi = -3.0$	$\xi = -1.6$	$\xi = 0$
AA-low, ch 1 = 115 MHz	37	47	69
Low band dishes, ch 1 = 450 MHz	74	87	112

[a] For the ideal case ($\Delta\nu^{3/4}$), $N_{\rm ch} = 150$.

4.5 Discussion of search strategies

So one must consider when planning a survey: is it preferable to form and search more array beams, subarrays or incoherently combined station beams? From an instrument perspective, this will be influenced by the frequency of observation, station beamforming (for AAs) and array beamforming costs, search costs and the array filling factor. The direction of observation and the characteristics of the expected population are also significant factors. This section discusses these general trends.

4.5.1 Cost-effective combination modes

For a given receptor, event rate per beam ($\mathcal{R}_{\text{beam}^{-1}}$) captures the first-order processing costs and allows a comparison of signal combination modes and array configurations. For searches of extragalactic populations and the first-order analysis of Galactic populations, these trade-offs are independent of frequency. Incoherent combination always achieves a higher $\mathcal{R}_{\text{beam}^{-1}}$ than coherent combination, as shown in Section 4.4.2.2. The difference increases with the number of receptors or with a reduced filling factor.

Although exact costs are unavailable, the preferred combination for extragalactic searches can be generalised as follows:

Incoherent combination: Achieves the highest event rate per beam, making it preferable in most cases. If the cost of searching a beam signal is high, incoherent combination presents a further advantage over other modes. However, the total event rate is limited by physical and processing constraints on the number of dish or station beams available for incoherent combination.

Coherent combination: Requires an array with a high filling factor and low beam search costs, but enables the most sensitive observations. Coherent combination can achieve the highest total event rate if many array beams are be formed and searched, although for SKA_1-low, the option of many incoherently combined station beams is also possible.

Incoherently combined subarrays: This mode is only preferable when the beam search cost is low. A fly's eye mode, being single-receptor subarrays, excludes the buffering and source localisation advantages of an array, and commensality with most observations (see Section 4.2). Sets of three-receptor subarrays counter the localisation problem and could employ receptors unused by the primary user observation. For example, splitting 24 SKA_1-low stations outside the core into 8 of these three-receptor subarrays results in a total event rate approximately equal to the incoherent combination of 50 stations.

The SKA_1-low results show a weak preference for the incoherent combination of station beams: five array beams formed from the coherent combination of the SKA_1-low stations in the core are required to equal the event rate of a single beam of the incoherently combined total array, assuming the stations are very closely packed. If multiple station beams are formed 'at no cost' for normal (imaging) array observing, or the cost of array beamforming is high, the advantage of incoherent combination is increased. In practice, other effects such as RFI mitigation, and station and array beam quality also need to be taken into account.

For the SKA_1 low band dishes, 157 array beams from the coherent combination of dishes in the core are required to equal the event rate of the incoherently combined total array. Coherent combination with dishes would only be optimal for hundreds of array beams, and even then it is likely that the processing power would be more effectively spent on SKA_1-low. However, an incoherent commensal survey with low band dishes would be a cost-effective method to cover parameter space.

Although I do not consider alternative filling factors for the core and inner regions, Chapter 5 does consider a representative dual-band SKA_1-low implementation, where for the high-band SKA_1-low (180–450 MHz), the station diameter and core diameter is halved. Because of the low filling factor of the aperture array elements at the higher frequencies, the sensitivity at frequencies above ∼230 MHz is similar to the single-band implementation considered in this chapter. The consequence of the smaller station and core diameters is that the beam FoV, hence event rate per beam, increases by a factor of four.

4.5.2 Frequency effects

As a first-order approximation for a given receptor, the *relative* event rate per beam between signal combination modes is frequency-independent. However, Figure 4.4 showed that for a given signal combination mode, the event rate per beam changes with frequency. For directions of low scattering where the intrinsic pulse width is much greater than the scatter broadening ($W_\mathrm{i} \gg \tau_\mathrm{d}$), such as the extragalactic case, the slope on the dish event rate is due to the spectral index of the source and increasing FoV at lower frequencies; for aperture arrays, the frequency dependence of T_sys (due to sky noise) and A_e are also factors.

Looking at these dependencies in more detail, the scatter broadening time modelled in this chapter scales as $\tau_\mathrm{d} \propto \nu^{-4.4}$. When the broadening dominates the intrinsic pulse width ($\tau_\mathrm{d} \gg W_\mathrm{i}$), then the observed pulse width is defined by the scatter broadening ($W \approx \tau_\mathrm{d}$). The event rate (Equation 4.1) thus decreases for lower frequencies, where $\mathcal{R} \propto \nu^{3.3}$ due to scatter broadening. But FoV and the spectral index of the expected population provide a counter-effect. The event rate due to FoV scales as $\mathcal{R} \propto \nu^{-2}$, while the effect of spectral index ξ on event rate is $\mathcal{R} \propto \nu^{1.5\xi}$. Combining these three factors, when $\xi < -0.87$, the event rate increases with lower frequency. For low-frequency aperture arrays, this is only approximate as A/T has no simple frequency scaling (see Equations 4.17 and 4.18). As per the simplifying assumptions (Section 4.4.1), a filter matched to the observed pulse width is still required to make the detection; the pulse will be wider at lower frequencies due to scatter broadening.

For SKA_1, the event rate per beam is generally higher for AA-low than the low band dishes, however the opposite is true for sources with low spectral indices and directions of larger scatter broadening. Also, lower frequencies have an increased memory cost for dedispersion, because the pulse is dispersed over a longer timescale. For an intrinsic pulse width of 1 ms, the low band dishes give a higher event rate than SKA_1-low for directions on or near the Galactic plane where the scatter broadening is large. Although the effectiveness of SKA_1-low is reduced at these directions, the increased luminosity of steep spectrum sources at lower frequencies somewhat counters this effect, as discussed above. Observations using the high band feed (1–2 GHz) on dishes are not modelled here, but in areas closer to the Galactic Centre, where scatter

broadening is increased, such observations would achieve an event rate higher than the low band feed for spectrally shallow sources and sources with smaller intrinsic pulse widths. When low spectral indices, increased scatter broadening or low-frequency turnovers are factors and only a few AA station or array beams are available, the incoherent combination of the low band dishes gives a higher total event rate.

Other astrophysical effect influence this result. For example, Bhat et al. (2004) have found departures from the -4.4 scatter broadening exponent, where a global fit results in an exponent of -3.9 ± 0.2, thus slightly reducing the broadening effect at lower frequencies. Also, most pulsars display a turnover (a break in the spectrum where pulsar brightness is maximum) around 100–200 MHz (Malofeev et al., 1994), and a small number have a turnover above 1 GHz (Kijak et al., 2011). The event rate for such sources would decrease below the turnover frequency. These effects are additional to the sensitivity and FoV factors that make the event rate in Figure 4.4 frequency-dependent, even for the extragalactic case when $\xi = 0$.

The intergalactic medium (IGM) may also be a source of significant pulse broadening, which could limit the detectability of extragalactic transients. The recent work of Koay (2012) extends scattering models of the interstellar medium (ISM) to cosmological distances, and determines observational and modelled constraints on temporal smearing. For example, the observational constraint on temporal smearing is up to $\sim$100 ms at 1 GHz and $\sim$2700 s at 100 MHz (scaling as $\tau_{\rm d} \propto \nu^{-4.4}$), although some of the modelled limits are considerably lower. Smearing this large places significant limitations on the detectability of all but the brightest extragalactic transients. But conversely, a detection verified to be extragalactic origin provides an excellent probe of the IGM.

4.5.3 Bandwidth effects

The frequency-dependent effects change how event rates are calculated for large fractional bandwidths. Section 4.4.4 shows that an optimal frequency range for searching for fast transients with SKA_1 may be smaller than the full band of the receptor. As the frequency increases, the FoV reduces and the source luminosity is expected to decrease for sources with pulsar-like emission characteristics. Processing more channels (hence bandwidth) increases the event rate, however this increase is less than $\Delta\nu^{3/4}$.

For some threshold beyond which extra bandwidth contributes little to the event rate, alternative search strategies may improve the probability of intercept. For example, the processing capacity for channels above this threshold could be more effectively used to form and search extra beams (increasing the event rate through FoV), trial more DMs or increase the detection S/N (sensitivity) through more optimal dedispersion techniques. The threshold depends on the costs for forming and searching the beams, which increase with bandwidth. Also, the number of channels (hence bandwidth) required to reach this threshold reduces for steeper spectrum sources (see Table 4.5).

If the spectral index of the target population and the scatter broadening in the direction of observation are not well-known, an alternative approach could be to search two or more smaller bandwidths of $\sim$50–100 MHz, at either end of the frequency range of the receptor. This is

most applicable to receptors with large fractional bandwidths; this observational approach has already been implemented with multiband feeds and receivers on the Green Bank Telescope to study wideband emission from pulsars (Maan et al., 2012). To search a wider range of frequencies with SKA_1, observations with half the dishes using low band feeds, and half using high band feeds is conceivable; with incoherent combination, each band would be a factor of $\sqrt{2}$ less sensitive.

Even if the search does not use the full bandwidth of the receptor, recording it in a buffer is desirable to enable improved verification and analysis of detected transients with dedicated processing—this is especially important when relatively rare candidate events are detected. The extra spectral information is also a useful analysis tool. For example, the presence or absence of multipath propagation which causes scatter broadening would be more evident with a larger bandwidth, given the frequency scaling of approximately $\nu^{-4.4}$. The larger bandwidth would help differentiate between the intrinsic pulse width and a pulse that has been scatter broadened.

4.5.4 Increasing the probability of intercept

The discussion thus far has focused on the event rate per beam in determining cost-effective search strategies. But ultimately, maximising the probability of intercepting verifiable fast transient events is desirable. The total event rate $\mathcal{R}$ of a search strategy is

$$\mathcal{R} = \mathcal{R}_{\mathrm{beam}^{-1}} N_{\mathrm{beam}}, \tag{4.21}$$

where N_{beam} is the number of beams formed and searched. Depending on the signal combination mode used, N_{beam} may be the number of station ($N_{\mathrm{b\text{-}0}}$) or array ($N_{\mathrm{b\text{-}arr}}$) beams, or $N_{\mathrm{sa}} N_{\mathrm{b\text{-}0}}$ beams for N_{sa} incoherently combined subarrays. Electromagnetic and signal processing system design and cost considerations will put maxima on each of these. For example, the number of station beams which can be formed within the elemental antenna FoV is limited by performance degradation, the diameter of the station and the beamforming processing power available. Once the limit on forming more station beams is reached, $\mathcal{R}$ will eventually become higher for coherent rather than incoherent combination. This is because N_{pix} array beams can be formed within each of the $N_{\mathrm{b\text{-}0}}$ station beams (Equation 4.6), allowing a maximum of $N_{\mathrm{pix}} N_{\mathrm{b\text{-}0}}$ array beams to be formed, and more than that if array beams are formed beyond the half-power point of the receptor beam.

Dishes with single-pixel feeds only have a single beam (i.e. $N_{\mathrm{b\text{-}0}} = 1$), so the event rate for incoherent combination of dishes cannot be increased by forming more beams. However, if phased array feeds (PAFs) are available on dishes, then $N_{\mathrm{b\text{-}0}}$ dish beams can be formed, increasing the total event rate. This is why in Figure 4.1, SKA_1-survey shows a higher probability of intercept than the dish array with single-pixel feeds.

The total survey time is the other factor to increase the number of detections, hence probability of intercept (Equation 4.3). This is where commensal surveys present a distinct advantage, because ideally, the survey time is equal to the telescope operation time. The V-FASTR experiment is an example of an on-going commensal survey, where $\sim$1300 hr of time was observed in its first year of operation (Wayth et al., 2012).

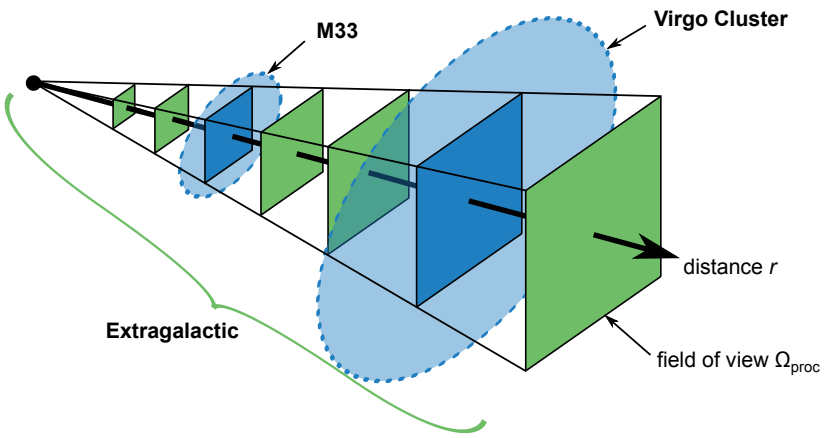


Figure 4.6: Geometry of detecting Crab-like giant pulses, for populations located at a fixed distance (blue circles), or a homogenously distributed extragalactic population (green squares).

4.6 Astrophysical implications

Although the modelling results in Section 4.4 are presented as comparative performance analyses, these results correspond to actual event rates as indicated in Table 4.4 and the text. In this section I apply these event rates to an example population of giant pulses, and discuss other factors for fast transient searches.

4.6.1 Searches for Crab-like giant pulses

Searches for a population of Crab-like pulsars illustrate how the results in this chapter relate to exploration of parameter space. The discovery of the Crab pulsar was due to the detection of its energetic 'giant' pulses (Staelin & Reifenstein, 1968). The strength of these giant pulses, along with their power-law statistics (e.g. Lundgren et al., 1995; Cordes et al., 2004a), make a hypothetical population of giant-pulse emitting pulsars an an astronomically interesting class of fast transients.

Figure 4.6 shows the geometry involved in placing a Crab-like pulsar population at extragalactic distances. An observation may target a population in a host region such as the Triangulum Galaxy M33 and the Virgo Cluster (of galaxies), shown by the blue circles, or it may be a blind search at extragalactic distances along the line of sight (green squares), as already considered in this chapter. Figure 4.1 showed nominal event rate limits for both scenarios, being the blue and green lines respectively; this section details the calculations for these limits.

4.6.1.1 Giant pulses

The giant pulses from the Crab-pulsar display a power-law energy distribution shown in Figure 4.7, with a negative exponent and no evidence for a high-energy cut-off (e.g. Bhat et al., 2008; Majid et al., 2011; Popov & Stappers, 2007); representative characteristics are given in Table 4.6. The 'pulse energy' E is the time-integrated flux of a pulse of width W; with units of Jy s (or $\mathrm{J\,m^{-2}\,Hz^{-1}}$), this is simply spectral energy fluence (Bradt, 2003). I make a simplifying

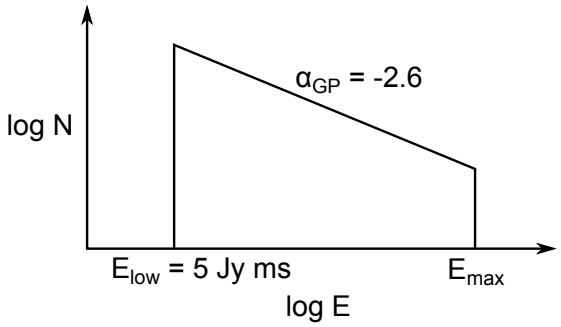


Figure 4.7: Representative energy distribution for giant pulses from the Crab pulsar.

Table 4.6: Representative characteristics of Crab giant pulses at 1.3 GHz.

Parameter	Value
Distance[a] $D_{\rm Crab}$	2.0 kpc
Low energy cut-off[b] $E_{\rm low}$	5 Jy ms
Probability distribution power-law index[b] $\alpha_{\rm GP}$	-2.6
Frequency of giant pulse occurrence[c] $f_{\rm GP}$	$4 \times 10^{-2}\,{\rm s}^{-1}$

[a] Nominal distance with uncertainty of ±0.5 kpc (Kaplan et al., 2008).
[b] From Figure 3 of Bhat et al. (2008). The slope of this cumulative giant pulse energy distribution is -1.6, or $\alpha_{\rm GP} = -2.6$ for the probability distribution.
[c] Fraction of pulses with energy equivalent to 5 Jy ms or higher is ∼0.0013, pulse period is 33 ms (Bhat et al., 2008).

assumption that $W < \tau/2$, such that the pulse's energy is measured by a single integration. For $\tau = 1$ ms, this is a reasonable assumption, given the largest pulse widths have been found to be of order 10 μs (Bhat et al., 2008) to 100 μs (Majid et al., 2011) at ∼1.4 GHz, and approximately 300 μs for the brightest pulse observed by Bhat et al. (2007) at 200 MHz.

The fraction of detectable pulses is determined by the complementary cumulative distribution function of the energy distribution. The energy distribution can be equally described by a luminosity distribution with the same power-law index $\alpha_{\rm GP}$ (see Appendix E.6); in which case, $\mathrm{P}(\mathcal{L} > \mathcal{L}_{\rm min})$ gives the fraction of pulses of luminosity greater than the minimum detectable luminosity $\mathcal{L}_{\rm min}$, at some distance r:

$$\mathrm{P}(\mathcal{L} > \mathcal{L}_{\rm min}) = \begin{cases} 1, & \mathcal{L}_{\rm min} \leq \mathcal{L}_{\rm low}, \\ (\mathcal{L}_{\rm min}/\mathcal{L}_{\rm low})^{1+\alpha_{\rm GP}}, & \mathcal{L}_{\rm min} > \mathcal{L}_{\rm low}, \end{cases} \tag{4.22}$$

where

$$\mathcal{L}_{\rm min} = 4\pi S_{\rm min} r^2, \tag{4.23}$$

and the low-luminosity cut-off $\mathcal{L}_{\rm low}$ is

$$\mathcal{L}_{\rm low} = \frac{4\pi E_{\rm low} r_{\rm Crab}^2}{\tau}, \tag{4.24}$$

such that $\mathcal{L}_{\rm low} = 80\pi$ Jy kpc^2 for $\tau = 1$ ms.

Table 4.7: Representative characteristics for targeted observations.

Parameter	M33	Virgo Cluster
Number of Crab-like pulsars per galaxy	10[a]	10[a]
Distance r (Mpc)	0.95[b]	16.5[c]
Intrinsic event rate per solid angle $\rho_{i,\Omega}$, for $E > 5\,\mathrm{Jy\,ms}$ (events $\mathrm{s^{-1}\,deg^{-2}}$) at 2 kpc	0.43[d]	3.6[e]

[a] McLaughlin & Cordes (2003).
[b] Bhat et al. (2011).
[c] Mei et al. (2007).
[d] Galaxy of size $73' \times 45'$ (Bhat et al., 2011).
[e] For 1277 galaxies in $140\,\mathrm{deg^2}$ (Binggeli et al., 1985).

4.6.1.2 Targeted observations of a host region

The details of the host regions in this example are given in Table 4.7. By assuming the host region as a plane at a fixed distance (see Figure 4.6), the intrinsic event rate density ρ_i can be approximated as an *intrinsic* event rate per unit solid angle $\rho_{i,\Omega}$ (events $\mathrm{s^{-1}\,deg^{-2}}$), calculated as the product of the giant pulse frequency f_{GP} and the number of pulsars per unit solid angle. Only those pulses with luminosity $\mathcal{L} > \mathcal{L}_{\mathrm{min}}$ will be observable, such that the detected event rate is

$$\mathcal{R}_\Omega = \Omega_{\mathrm{proc}} \rho_{i,\Omega} \mathrm{P}(\mathcal{L} > \mathcal{L}_{\mathrm{min}}). \tag{4.25}$$

Equation 4.25 assumes that the target region extends beyond the processed FoV Ω_{proc}. Note that this event rate is for a plane (hence the subscript $_\Omega$), rather than the event rate in a volume of sky used for previous calculations in this chapter.

To describe Equation 4.25 in terms of probability of intercept, the detection of $\mathcal{R}_\Omega T_{\mathrm{tot}}$ events can be expected over total survey time T_{tot}. Conversely, the FoV–time product required to observe a single event is

$$\Omega_{\mathrm{proc}} T_{\mathrm{tot}} = \frac{1}{\rho_{i,\Omega} \mathrm{P}(\mathcal{L} > \mathcal{L}_{\mathrm{min}})}. \tag{4.26}$$

Equation 4.26 is used for the blue lines in Figure 1.3; an illustrative result from that plot is that a single giant-pulse detection would be expected from ∼5000 hours of observation of M33 with the Parkes HTRU system (13 beams), the incoherently combined $\mathrm{SKA_1}$ dishes (1 beam), or 100 beams of the coherently combined $\mathrm{SKA_1}$ dishes.

4.6.1.3 Blind extragalactic searches

A comparison of the probability of intercept for blind searches requires further assumptions about the source population. I assume a homogeneously distributed population that produces events with characteristics similar to the Crab giant pulses (Table 4.6); the power-law energy distribution being a key feature. The purpose of these simplistic assumptions is to illustrate how the event rate calculations in this chapter applies to an example population, rather than implying that such a population exists.

The observed event rate for this example population is the summation of event rates for a series

of infinitesimally small volumes along the line-of-sight, as conceptualised in Figure 4.6. The green squares represent the volume $\Omega_{\rm proc} dr$ at distance r. The observable event rate for each of these volumes depends on $\mathrm{P}(\mathcal{L} > \mathcal{L}_{\rm min})$, the fraction of events observable at distance r. The total detected event rate is simply an integration along the line-of-sight.

To calculate the detected event rate $\mathcal{R}$, the fixed luminosity $\mathcal{L}$ in Equation 4.1 is replaced with a distribution of luminosities, such that $\mathcal{L}_{\rm low} \leq \mathcal{L} \leq \mathcal{L}_{\rm high}$. When $\mathcal{L}_{\rm high} \gg \mathcal{L}_{\rm low}$, as it is for this population of giant pulses, $\mathcal{R}$ is given by four cases, depending on the power-law index of the luminosity distribution (see Equation 16 of Macquart, 2011). From Table 4.6, the power-law index is $\alpha_{\rm GP} = -2.6$ and the relevant event rate calculation from Macquart (2011) is

$$\mathcal{R} = \frac{\Omega_{\rm proc}\rho_i}{3} \left(\frac{\mathcal{L}_{\rm low}}{4\pi S_{\rm min}} \right)^{3/2} \left| \frac{1+\alpha_{\rm GP}}{5/2+\alpha_{\rm GP}} \right|. \tag{4.27}$$

Conversely, FoV–time product required to observe a single event is

$$\Omega_{\rm proc} T_{\rm tot} = \frac{3}{\rho_i} \left(\frac{4\pi S_{\rm min}}{\mathcal{L}_{\rm low}} \right)^{3/2} \left| \frac{5/2+\alpha_{\rm GP}}{1+\alpha_{\rm GP}} \right|. \tag{4.28}$$

A detection or non-detection with a survey of given $\Omega_{\rm proc} T_{\rm tot}$ and $S_{\rm min}$ will therefore provide a limit on the intrinsic event rate per unit volume ρ_i. Because the value of ρ_i is unknown, the green lines in Figure 1.3 signify limits on ρ_i inferred from a single detection of an event from this example homogeneously distributed population.

4.6.1.4 Survey implications

The Crab-like giant pulses provide a useful illustration of the astrophysical application of the calculations in this chapter. If a search was specifically undertaken for such a population, there are some further factors to consider.

For example, the high energy cut-off of the power-law energy distribution is an unknown factor. The largest pulse energy observed above 1 GHz which is known to this author equates to ∼800 Jy ms (Figure 1 of Popov & Stappers, 2007), but there have been insufficient observations to to determine a true high-energy cut-off. Most published observations are of only several hours and are at non-overlapping centre frequencies, ranging from 0.1 GHz to 8.8 GHz (see Table 2 of Majid et al., 2011, Karuppusamy et al., 2010). A longer observation by Lundgren et al. (1995) of 100 hours at 812 MHz was affected by instrumental and scatter pulse broadening (Bhat et al., 2008). The short observation times, and therefore a small sample size of the most energetic observed pulses, reduces the reliability of the statistics.

To calculate the upper bound on maximum giant pulse luminosity (hence energy), Cordes (2009a) uses the spin-down luminosity, being the rate of rotational energy loss. Cordes (2009a) estimates this upper bound to be about two orders of magnitude larger than the brightest giant pulses from the Crab pulsar, for unspecified 'nominal' parameters. However, beams of smaller solid angle, and smaller bandwidths of emission radiation result in much larger luminosities. Such limits to luminosity will determine whether a particular telescope or signal combination mode is sensitive enough to detect the target population. For example, Bhat et al. (2011) calculates that a 20 kJy pulse of width 0.1 ms from M33 would be detectable by Arecibo at 1.4 GHz with 5σ confidence. A 20 kJy pulse equates to an energy of ∼2 kJy ms, which stronger

than the pulse detected by Popov & Stappers (2007). Telescopes and combinations in Figure 4.1 with less sensitivity than Arecibo, such as the incoherent combination of SKA_1 dishes, would only detect stronger pulses, if they exist. In such a scenario, using the coherent combination of dishes may be a prudent approach.

Observing at a lower frequency is also an option; the pulses are stronger, although scatter broadening acts as a counter-effect (see Section 4.5.2). Giant pulse observations at 0.6 GHz and 1.4 GHz by Sallmen et al. (1999) have found spectral indices within a range of -2.2 to -4.9, although a number of detections at 1.4 GHz, but non-detections at 0.6 GHz, indicated that some giant pulses have a flatter spectral index than -2.2. The lower frequencies show potential. For example, McLaughlin & Cordes (2003) searched M33 for giant pulses using Arecibo at 430 MHz, with $\tau = 0.1$ ms, and more recently, Rubio-Herrera et al. (2012) searched the Andromeda Galaxy M31 with the Westerbork Synthesis Radio Telescope (WSRT) at 328 MHz, with $\tau \approx 0.5$ ms. Both searches detected single-pulse events repeated at particular dispersion measures, which may have originated from the respective target galaxy, but these results were not conclusive. Meanwhile, a higher-frequency (1.4 GHz) search of M33 by Bhat et al. (2011) using the Arecibo multibeam system returned a null result. The SKA will observe the lower frequencies over a wider bandwidth with higher sensitivity, therefore improving the probability of intercept for such search.

Temporal resolution may also be a factor which influences the sensitivity of a search; for Crab-like giant pulses, a smaller time resolution increases the signal-to-noise ratio (SNR), assuming that scatter broadening is sufficiently low as to not affect the SNR. For example, the strongest pulse in Popov & Stappers (2007) has a peak flux density of $\sim$100 kJy averaged over 8 μs. If detected at with a time resolution of $\tau = 10\,\mu$s instead of $\tau = 1$ ms used in this chapter, the energy from the signal remains the same, but system noise decreases by a factor of $\sqrt{100}$, resulting in a factor of 10 higher SNR.

4.6.2 Exploring new parameter space with the SKA

The giant pulses are a specific example of how high time resolution surveys can be optimised to increase the number of events detected in a survey, for a given class of fast transients. More generally, the discoveries by Lorimer et al. (2007), Keane et al. (2011, 2012) and M. Bailes et al. (2012, pers. comm.), along with the discovery of rotating radio transients and pulses from magnetars, show that the high time resolution Universe is much more interesting at radio frequencies than was considered even a decade ago. A point of note is that all these observations were initially made with the Parkes multibeam feed (Staveley-Smith et al., 1996). Radio telescope collecting area itself did not lead to these discoveries, given that large single-dish telescopes have been available for many decades. Arguably, technical innovations such as the multibeam feed, a new filterbank with 288 MHz bandwidth and 250 μs time resolution (Manchester et al., 2001) and improved computational processing capability and techniques have led to such discoveries. Furthermore, the sky surveys in which these fast transients have been found total $\sim$3200 hr of observing time, and some of these surveys have been analysed multiple times (see Table 4.1, and Table 4 of Siemion et al., 2012).

These technical innovations and the significant total survey time have expanded the region of

high time resolution parameter space available to astronomers. However, a limiting factor for candidate fast transient events found with single-dish telescopes has been the lack of tools to verify the event; RFI discrimination and source localisation being significant factors.

Siting telescopes in RFI-quiet areas is most effective to reduce RFI contamination, and the SKA has a significant advantage in this regard. However, candidate detections still require RFI discrimination. Multibeam feeds on single dishes do provide some discrimination between events of astrophysical and terrestrial origin (e.g. Burke-Spolaor et al., 2011), while dual-receptor (Bhat et al., 2005) and multi-receptor (Bhat, 2011; Thompson et al., 2011) transient detection methods improve RFI discrimination. If the raw voltage data is available for an array telescope, this provides excellent RFI discrimination capability (e.g. James et al., 2010).

Spatial resolution allows for verification through improved localisation of a detected event. Although the search strategies analysed in this chapter have lower spatial resolution than that available with the longest baselines, transient buffers can recover the full spatial resolution of the SKA. The higher angular resolution improves the prospects for multiwavelength follow-up (e.g. Kaplan et al., 2009).

Temporal resolution influences sensitivity, but it also plays a role in understanding new phenomena. For example, pulsars have been found to display structure on the scale of microseconds (Johnston et al., 2001) and even nanoseconds (Hankins et al., 2003). Again, accessing the raw voltage data will allow for very high time resolution analysis of candidate events.

The efficiency of the search processing will affect search sensitivity, hence detection rate. Although the modelling in this chapter has assumed that the signal arriving at the telescope is detected with 100 % efficiency, the actual survey efficiency will be less than that. As outlined in Appendix E.3, the dedispersion system broadens the pulse and reduces the SNR; the degree of broadening depends on the dedispersion system. Incoherent dedispersion has a lower SNR than coherent dedispersion, but requires significantly less processing.

In summary, the availability of all information from the original observation (via the voltage buffer, for example) maximises the region of available parameter space. If a relatively rare candidate event is detected, the ability to verify and interpret such an event will be enhanced with this information. For the SKA, even though the survey may not search the whole region of parameter space, the full bandwidth of buffered data maximises the spatial and temporal resolution and polarisation states available to interpret the event and the intervening medium.

4.6.3 Galactic populations

The effects on event rate for Galactic populations are more subtle that the first-order analysis considered thus far. The population will be observable to the edge of the Galaxy if the telescope is sufficiently sensitive and the source population sufficiently luminous. In that case, the event rate for a sensitivity limited volume (as used for extragalactic populations) is not valid, because the limit is instead imposed by the boundary of the Galaxy. Furthermore, scatter broadening is integrated along the line of sight (Cordes & Lazio, 2002) and is therefore distant dependent; this creates a ‘sensitivity horizon’ for directions of large scatter broadening, where increased sensitivity no longer increases the event rate (Macquart, 2011).

Macquart (2011) captures both of these effects by introducing a direction dependent factor δ, such that

$$\mathcal{R} \propto \Omega_{\rm proc} S_{\rm min}^{-3/2+\delta} \quad \text{events}\,\text{s}^{-1}, \tag{4.29}$$

where $0 \leq \delta \leq 3/2$. A key result in Macquart (2011) is that FoV is more strongly preferred over sensitivity when δ increases, due to scatter broadening or volume boundary limits. Further modelling of the directional dependence of δ is shown in that paper.

The value of δ depends on the survey depth; this in turn depends on the sensitivity of the signal combination mode. For example, the more sensitive coherent combination mode may be more strongly impacted by scatter broadening or the volume boundary than the less sensitive incoherent combination mode. In which case, δ would be larger for coherent combination, and the relative event rates in Section 4.4.2 would no longer hold. Rather, the cost-effectiveness of coherent combination would reduce.

The value of δ may also vary as a function of frequency, due to a frequency-dependent luminosity distribution and broadening time. At lower frequencies, either the boundary of the Galaxy or the sensitivity horizon are more likely to be reached; the former for steep spectrum sources in directions of low scattering, the latter for shallower spectrum sources in directions of high scattering.

While a large δ increases the preference for FoV, it equally increases the preference for total survey time, so that the total number of detected events is maximised. Strategies to increase the probability of intercept for cases when $\delta > 0$ include: forming multiple beams (if available); incoherently combining the array or subarray signals (resulting in a much larger FoV that coherent combination); or simply spending more time observing the sky. A low cost commensal survey using the incoherently combined array increases both FoV and survey time. Therefore, the effectiveness of such a survey increases for populations near the Galactic plane.

4.7 Recommendations for SKA$_1$

This chapter has shown that relatively low-cost design solutions significantly increase the probability of intercepting verifiable fast transient events. Specific design recommendations are as follows:

- **Flexible search modes maximise the exploration of the unknown**

  The preferred receptor and signal combination mode depends on direction, especially for Galactic populations. Incoherent and coherent combination modes are both effective with SKA$_1$; their effectiveness depends on the array filling factor, signal processing costs and the anticipated spectrum of the source population.

- **The re-use of signal processing required for SKA imaging modes enables low-cost fast transient searches**

  Commensal surveys using the incoherent combination of dish signals or AA station beams are low cost options for searching for fast transients with SKA$_1$. The incremental cost of implementing such a search is small because it uses beams formed for the primary user

observation. A commensal survey using both receptors is a simple method to increase the total number of events detected, by increasing the total observation time. Access to the dish and station beam data, via a spigot with a well defined interface, enables the implementation of flexible search modes. The processing for commensal surveys could be implemented internally or with user-provided processing as it becomes affordable or available.

- **Requirements for fast transient searches with SKA_1**

  Until the processing costs are further explored, the base requirements for SKA_1 are:
  - Availability of incoherent and coherent combination modes for AA-low and low band dishes.
  - Processing for low-cost commensal survey modes; or provision for access to the dish and AA station beam data via spigots.
  - Voltage (coherent) buffering capability of the full band; of order a minute for dish frequencies and possibly minutes for lower frequencies.

  For extragalactic searches, processing the full available bandwidth is not required. Bandwidths of 50–100 MHz are sufficient on the basis of the simplified investigation undertaken in this chapter; a more detailed study of the trade-offs could be made. However, buffering the full band is desirable for analysis of candidate detections.

4.7.1 A low-cost SKA_1 transients search system

Although determining system-optimised costs for SKA_1 requires more detail about the signal processing architectures, I consider here the approximate cost of a commensal system that maximises the probability of detecting fast transients with blind searches. Figure 4.8 shows a representative implementation of such a system, where signals from the dish array are incoherently combined for real-time detection, and the digitised voltages are buffered to enable coherent localisation. This architecture is similar to the V-FASTR system on the VLBA (Wayth et al., 2011).

The representative implementation uses spigots onto an FX correlator architecture, and has an incremental cost of order €1 M. The A first-order estimate of the component costs for such a system is given in Table 4.8 and detailed in Appendix E.7. These costs are for today's technology; a Moore's law type of exponential cost reduction of the digital components can be expected. Conversely, there will be additional costs for supporting hardware for these components.

A similar system could be built for SKA_1-low; the prime differences being a factor of 5 less inputs, but a longer time delay due to dispersion (Equation E.22) increases the cost of the transients processor and buffer memory. For each additional station beam that is searched, an additional duplicate set of components would be required. An alternative with SKA_1-low is coherent combination; the event rate per beam for coherent combination of the core stations is only a factor of $\sim$5 less than the incoherent combination of the total array. Therefore, time-averaged data accessed from a central beamformer operating commensally with the correlator may be more cost-effective. However, even though the search is conducted with the core, buffering of the station beam data would still be required to maximise the capability for event

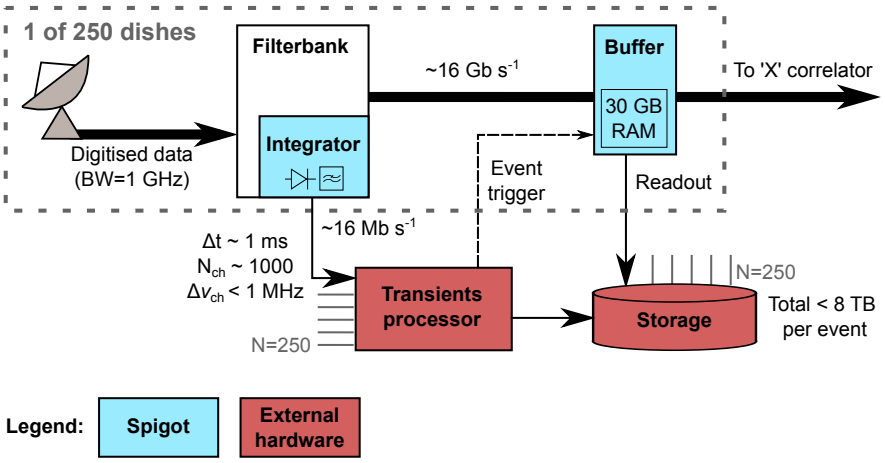


Figure 4.8: Representative commensal fast transient processing system for the SKA_1 dish array.

Table 4.8: First-order costs for the representative commensal fast transient processing system.

	Quantity	Unit cost (€)	Total cost (€)
Filterbank spigot access components	64 (serving 250 dishes)	1 k	64 k
Transients processor	250 streams	1 k	250 k
Buffer spigot	64 Uniboards, 4×30 GB RAM per Uniboard	11 k	704 k
Storage	750 TB (100 candidate events)	75 k	75 k
	Total cost of components		**1.1 M**

localisation and verification.

4.8 Chapter summary

This chapter has demonstrated that both SKA_1-low and the SKA_1 dish array will be powerful instruments to search for, and verify, fast transient events. For the SKA_1 and radio telescope arrays in general, I have shown that the probability of intercepting such events depends on instrumental effects, such as the choice of receptor, signal combination mode and frequency range. Additionally, I have analysed how the direction of observation and the expected spectrum of the source population impacts the probability of intercept. For commensal observing, where the direction of observation is determined by the primary telescope user, a fast transient pipeline must dynamically adjust the search strategy to achieve the highest event rate. For a targeted survey, the choice of receptor requires careful consideration.

To measure the cost-effectiveness of a survey strategy in detecting transient events in a volume of sky, I presented event rate per beam ($\mathcal{R}_{\mathrm{beam}^{-1}}$) as a new figure of merit. The per beam event rate enables a frequency independent first-order analysis of the optimal signal combination mode and array configuration for a given receptor type. Regardless of the telescope array used,

incoherent combination always achieves a higher event rate per beam, making it preferable to coherent combination, subarraying and fly's eye in many cases. For an extragalactic search with the core region of SKA_1 dishes, more than 157 coherently combined array beams must be formed and searched to achieve a total event rate higher than the incoherently combined SKA_1 dishes. For the more densely packed core region of SKA_1-low, five coherently combined array beams are required to achieve the event rate of the incoherently combined SKA_1-low stations.

The advantage of incoherent combination increases with a lower filling factor, the combination of an array with more receptors and when Galactic effects limit the survey depth. Incoherent combination also allows for low-cost commensal searches for fast transients, where voltage buffers enable coherent follow-up of candidate events; Section 4.7.1 presented a representative implementation of a processor to achieve this. Furthermore, both incoherent combination or incoherently combined subarrays could usefully employ receptors unused by the primary user observation (e.g. low surface brightness density imaging), therefore increasing probability of intercept through increased survey time.

As wideband or multiband feeds and receivers are implemented on a larger number of telescopes, more sophisticated survey strategies will be required to maximise the probability of intercept. Using SKA_1 as an example, this chapter shows that simply searching the full available bandwidth is not an optimal use of the processing system, especially for extragalactic searches and for steep spectrum populations. The contribution to event rate from processing additional bandwidth decreases as the signal frequency increases. Beyond a threshold, the processing could be more effectively used to form and search more beams, trial more DMs or increase the detection S/N ratio. However, buffering or storing data from the full band is desirable for verification and analysis of candidate detections, especially for rare, weak events.

Chapter 5

Single versus dual-band SKA-low

Chapter 4 presented the event rate per beam figure of merit as a high-level approximation of a telescope's cost-effectiveness in searching for fast transients, and showed that SKA_1-low has the potential to be a powerful instrument for fast transient searches. However, the capability of SKA_1-low is founded on cost-effective design solutions; in this chapter I consider the cost implications of a single and dual-band aperture array implementation for SKA_1-low, chosen for comparable performance characteristics.

5.1 Introduction

Despite the release of a high-level concept design (Garrett et al., 2010) for the first phase of the SKA (see Section 1.1.3), an outstanding question in the SKA community is whether the 70–450 MHz frequency range (SKA_1-low) should be observed with an array composed of a single wideband antenna element design (single-band implementation), or with two arrays, each observing approximately half the fractional bandwidth (dual-band implementation).

The choice of a single or dual-band implementation is a key design decision for SKA_1-low. The system descriptions present an overview of the telescope as a complex set of inter-connected parts (sub-systems). Although the design of the system and its sub-systems have been refined through Concept Design Reviews (CoDRs), most of the recent system-level studies (including the CoDRs) for the SKA have assumed a single-band implementation for SKA_1-low (e.g. Bij de Vaate et al., 2011 and references therein). Unfortunately, this means that the effects of a dual-band implementation on the system design are not currently well documented. However, the dual-band approach used in the LOFAR telescope gives many insights into a putative SKA_1-low instrument.

The large fractional bandwidth ($\sim 6.5 : 1$) of the single-band implementation presents a number of design challenges that are less burdensome for the dual-band implementation. As Section 5.6.2 discusses, designing the active antenna element and optimising the spacing between these elements is particularly challenging for the single-band implementation. An optimal inter-element spacing requires a balance of competing effects at the lowest and highest frequencies in the band. At the lowest frequencies, where sky noise dominates, sufficiently large spacing between antenna elements is required to maintain effective area, hence sensitivity (Braun & van Cappellen, 2006). However, a larger inter-element spacing makes station calibration increasingly difficult at the highest frequencies (Wijnholds et al., 2011a; Wijnholds, 2012). In a dual-band system, the fractional bandwidth of each band is lower, hence there is more design flexibility to meet these challenges.

The *SKA AA CoDR panel report* (Dewdney et al., 2011b) recommends that the impact of the dual-band option on system design should be considered. Indeed, not meeting the SKA_1-low requirements with a single-band implementation is identified as a risk in van Es et al. (2011).

While the *AA Concept Descriptions* document (Bij de Vaate et al., 2011) makes a quantitative technical analysis of the single and dual-band approaches, to date there has been no comparison of the cost-effectiveness of each approach. Such a comparison is important, because the chosen approach has consequences for antenna element design, manufacture and deployment costs, and influences the downstream signal processing costs.

This chapter uses an elementary parametric analysis to determine if a dual-band implementation, with twice as many antenna elements, is significantly more expensive than the single-band implementation described in the SKA_1 high-level system description (HLSD, Dewdney et al., 2011a). I use simple algebraic equations to model the cost and performance of each implementation, and consider the cost drivers of the single and dual-band SKA_1-low at two levels. The first level is simply the cost of the hardware required specifically for the low-frequency aperture array sub-systems ('stations'). But because these stations are inter-linked with other sub-systems to realise SKA_1-low as a telescope, the effect of design choices within the stations is considered throughout the system. Thus the second, higher level, analysis incorporates costs which differ between the two implementations, such as those of the correlator, imaging processor and non-imaging processor sub-systems, as well as site-related costs specific to SKA_1-low.

The representative single and dual-band implementations are chosen for comparable sensitivity, field of view (FoV) and survey speed performance. The single-band implementation is that which is described in the SKA_1 HLSD, but there is no similar guidance to the design of the dual-band implementation. The illustrative comparison in this chapter uses a canonical form of the dual-band implementation, composed of two single-band arrays, each of which simultaneously observes approximately equal bandwidth ratios of 2.5:1. This produces a low-band array (70–180 MHz) with the same physical layout as the HLSD, and an additional high-band array (180–450 MHz) with a 0.75 m spacing between antenna elements. Alternative implementations could, for example, have overlapping bands or a 50 MHz minimum frequency. However, consideration of the system implications of such comparisons is beyond the scope of the present work.

The parametric modelling presented here shows that the smaller inter-element spacing for the high-band array is a key driver in reducing the cost of the dual-band implementation, via reduced digital data transport and processing loads throughout the system. Although 0.75 m inter-element spacing is arbitrarily chosen as being half that used for the low-band and single-band arrays, such a spacing maintains similar sensitivity performance to the single-band array, at least for the antenna elements described in the HLSD.

The goal of the trade-off and decision making processes for the SKA is to refine the design options by linking performance, cost and risk to science returns (Stevenson, 2011c; Dewdney, 2010). Because not all life-cycle costs are currently available or sufficiently understood, this work does not give a final cost, but instead makes a differential cost analysis of the single and dual-band SKA_1-low implementations. This work is not an analysis of expected telescope performance and total cost, nor are the examined systems optimised for performance and cost. However, by drawing upon the existing documented studies of aperture arrays and the SKA, this analysis is intended to assist these trade-off and decision making processes.

5.2 Chapter structure

Section 5.3 outlines the parametric cost modelling approach, along with the models and cost data sources used. Section 5.4 details representative single and dual-band SKA_1-low implementations and compares the station sub-system costs. Other selected SKA_1-low sub-system costs, which vary between implementations, are considered in Section 5.5. Section 5.6 discusses the performance and cost trends, uncertainties and the relevance to SKA_2. Section 5.7 investigates some topical additional trade-offs: smaller station diameter, reduced beam-bandwidth product and changed intra-station architecture. Recommendations for further work are made in Section 5.8 and conclusions set out in Section 5.9. A summary of major assumptions listed in Appendix F.1.

Although the comparisons and trade-offs are progressively developed in each section, there may be aspects of the system which are of interest to particular readers. These are cross-referenced as follows:

- station hardware sub-systems
  - cost data sources: Section 5.3.2
  - derived unit costs and models: Appendix F.2
  - station design details: Section 5.4.1
  - representative implementation costs: Section 5.4.2
  - RF tile beamforming vs. all-digital beamforming: Section 5.4.3
  - uncertainty in representative implementation costs: Section 5.6.3
  - alternative intra-station architectures: Section 5.7.3.3
  - station diameter variation: Appendix F.6.1
  - reduced processed FoV through a fixed beam–bandwidth product: Appendix F.7.2
- variable system costs
  - site-related costs and models: Appendix F.4.1
  - central processing facility sub-system costs and models: Appendix F.4.2
  - representative implementation costs: Section 5.5
  - station diameter variation: Appendix F.6.2
  - reduced processed FoV through a fixed beam–bandwidth product: Appendix F.7.3
- station power demand
  - power demand models: Appendix F.4.3
  - representative single and dual-band implementations: Section 5.5.4
  - alternative intra-station architectures: Section 5.7.3.3
  - station diameter variation: Appendix F.6.2
- general cost trends
  - single vs. dual-band comparison: Section 5.6.1
  - station diameter variation: Section 5.7.1

  - reduced processed FoV through a fixed beam–bandwidth product: Section 5.7.2
- station performance
  - single vs. dual-band comparison: Section 5.6.2 and Appendix F.5
  - hierarchical beamforming Section 5.7.3.2

5.3 Parametric cost modelling

Defining a scalable model for the SKA is challenging because the project is currently in the design definition phase, where many design options and architectures are available, and a complete set of requirements is still being developed (Section 3.3.1). For these reasons, the model for this parametric analysis of SKA_1-low is necessarily simple; the design is captured in a dozen scalable blocks. Although such a model has high uncertainties, it provides useful insight at this point in the project and also indicates prime areas for further study (see Section 5.8).

5.3.1 SKA_1-low parametric models

The SKA_1 high level system description (HLSD, Dewdney et al., 2011a) is a useful starting point to create a parametric cost model for SKA_1-low, because it defines a system hierarchy which describes how the sub-systems relate to each other. For example, the low-frequency aperture arrays, signal transport and networks, signal processing, computing and software, and infrastructure are all immediate sub-systems of the telescope system. Although the HLSD will evolve as the system requirements are refined, it forms a 'representative system' as a common basis for the analysis of sub-system performance and cost.

To model the differences between the single and dual-band SKA_1-low, I decompose the low-frequency aperture array sub-system into another level of sub-systems. As mentioned in Section 5.1, the term 'SKA_1-low' encompasses hardware specifically related to the low-frequency aperture array sub-system as well as other SKA sub-systems. Although not explicitly defined in the HLSD, the low-frequency aperture array sub-system approximately describes the hardware for the SKA_1-low stations.

As shown in Section 2.3.1.4, the SKA_1-low antenna elements are grouped into 'tiles' and 'stations' to reduce the post-sensor data transport and processing loads. This grouping of antenna elements may affect the physical layout or may only change the signal processing architecture; the partitioning of the processing is termed hierarchical beamforming. The 'processed FoV' is synthesised from the formation of multiple station beams, which are cross-correlated in the telescope's signal processing system.

The HLSD also describes the geographical layout (configuration) of the stations. Approximately half the AA stations are located in a closely packed 'core' region, and the others placed with exponentially increasing density away from the core. A similar layout applies to the dishes. Because the AA and dish cores are densely packed, these are located nearby to each other but separate. At larger radii from the core, the stations and dishes are co-located, so that the data transport, timing signals and power distribution infrastructure can be shared.

Although the SKA_1-low sub-systems are captured in a small number of blocks to maintain clarity, it is important that the quantity of each of these blocks still scales correctly when the

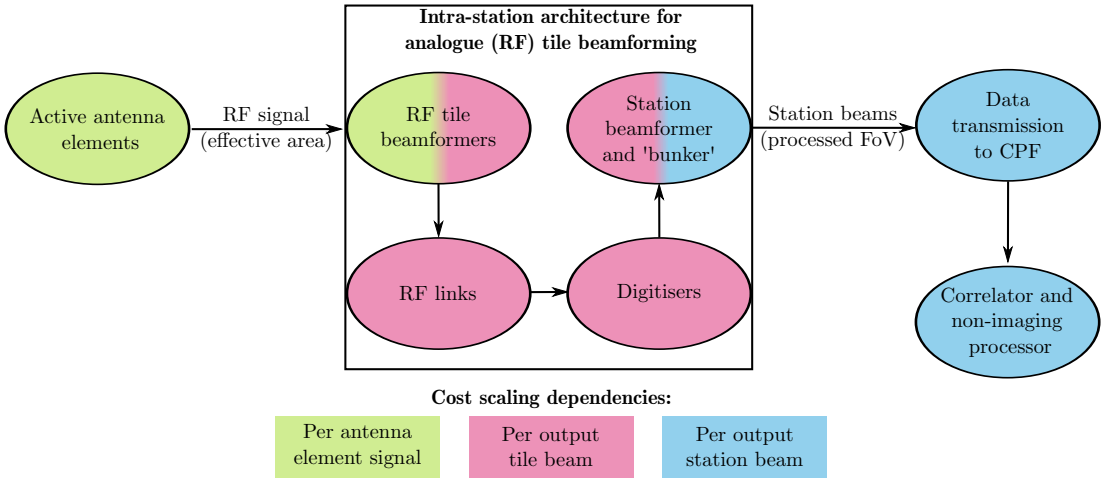


Figure 5.1: Conceptual diagram showing the general cost scaling dependencies of key blocks (ovals), for the RF tile beamforming intra-station architecture. The gradient indicates that both scaling dependencies apply.

parameters are varied. This approach is similar to some of the previous SKA costing efforts discussed in Section 3.5, except the blocks in this analysis describe the system at a higher level. Because only a small number of blocks are used, most of the results in this analysis have been calculated using a spreadsheet. However, the parametric models have been developed with a view to transferring them to SKACost (Section 3.5), which will allow further exploration of trade-offs in the SKA_1-low design space, and enable a more thorough statistical treatment of uncertainties. SKACost is used to make the preliminary uncertainty analysis in this chapter (Section 5.6.3.3).

The blocks (sub-systems) follow the elemental signal path, in an approach similar to Hall (2004a) and Horiuchi et al. (2004). The blocks used in this analysis to describe the elemental signal path are:

active antenna element: reception and amplification at the antenna element

RF tile beamformer: analogue beamforming of the elements in a tile. Digital tile beamforming is also possible, see Section 5.7.3.1

RF link: analogue signal transport from the active antenna element or the tile beamformer

digitiser: digitisation

station beamformer: coarse channelisation (filterbank) and digital beamforming of elements or tiles in the station

station 'bunker': controlled environment and infrastructure at the station processing node to house the beamformer hardware

digitiser–bunker link: digital signal transport from the digitiser to the bunker

central processing facility (CPF): central signal processing and science computing subsystems (specifically, the correlator, imaging and non-imaging processing). The CPF serves both the AA and dish arrays. Some parts of the CPF may be located off-site

data transmission to CPF (station–CPF link): digital signal transmission from the station processing node to the central processing facility (excludes network infrastructure).

These blocks assume a time to frequency domain transformation and cross-correlation 'FX' correlator architecture, being the most cost-effective architecture for the SKA (Section 2.3.3.3).

The blocks are combined to form a specific signal transport and processing architecture within a station, shown conceptually in Figure 5.1. A station will have some number of antenna elements (to realise effective area, hence sensitivity) and produce some number of station beams (to form the processed FoV). The rectangular box in Figure 5.1 encompasses the intra-station signal transport and processing architecture. The exact path of the signal from one block to the next depends on the architecture chosen; this is discussed in Section 5.7.3.

Figure 5.1 also shows the general cost scaling dependencies of each block, as relevant to the single versus dual-band comparison. The dependencies are linear parametric equations, with the key variable parameters being the number of antenna element signals, output beams formed by the tile beamformer and output beams formed by the station beamformer. For the digital signal transport links, the number of signals or beams transmitted is a proxy for the data rate transmitted. The scaling relationships and unit costs of each block are detailed in Appendix F.2. The two intra-station signal transport and processing architectures analysed in this work are discussed in the next section.

5.3.2 Cost data sources

Chapter 3 (Section 3.4.2.2) discussed the methodologies for developing cost data for use in cost estimation; the cost data sources for this analysis are shown in Table 5.1. Two cost estimates were developed in the AA CoDR (Faulkner et al., 2011) for the SKA_1-low stations; one based on the SKA Design Studies (SKADS) work (e.g. Faulkner et al., 2010), the other extrapolated from the existing Low Frequency Array (LOFAR) telescope[1]. Although these estimates describe stations which achieve similar sensitivity and FoV performance, they present two alternative intra-station signal transport and processing architectures, and also use different cost estimation methodologies and assumptions, as summarised in Table 5.2.

The principal architectural differences between the two estimates are how and where the digitisation and hierarchical beamforming is performed. The SKADS architecture uses all-digital beamforming, where both the tile and station beamforming are done digitally, while the LOFAR architecture uses analogue tile beamforming. These are the architectures considered in this work; the cost and performance implications of some other intra-station architectures on the SKA_1-low sub-system costs are discussed in Section 5.7.3.

The parametric models for the stations are based on these two cost estimates and their architectures. The all-digital beamforming architecture, as broken down into the sub-system blocks for this analysis, is shown in Figure 5.2. In this architecture, the signals are digitised close to the antenna elements, but no beamforming occurs at the tiles. Instead, two or more stages of hierarchical beamforming are assumed to occur within the station beamformer block. This differs slightly from Faulkner et al. (2011), where a first stage of beamforming is done at the

[1] www.lofar.org

Table 5.1: Sub-system cost data sources

Sub-system	Cost data source			
	SKADS[a]	LOFAR[a]	Signal Processing CoDR[b]	Professional opinion
Active antenna element	×	×		
RF tile beamformer		×		
RF links	×	×		
Digitiser	×	×		
Digitiser–bunker links	×			
Station beamformer	×	×		
Station bunker	×	×		
Data transmission to CPF	×[c]			
Correlator			×	
Non-imaging processor			×	
Correlator–computing data transport				×
Computing (imaging processor)				×
Deployment				×
Site preparation				×

[a] Faulkner et al. (2011).
[b] Turner (2011).
[c] Bolton et al. (2009a).

Table 5.2: Costing methodologies and data sources

		Cost data source	
		SKADS	LOFAR
Costing methodology	Bottom-up	×	
	Reference class		×
Intra-station architecture	Tile beamforming	digital	RF
	Tile–station signal transport	digital	RF
	Station beamforming	digital	digital

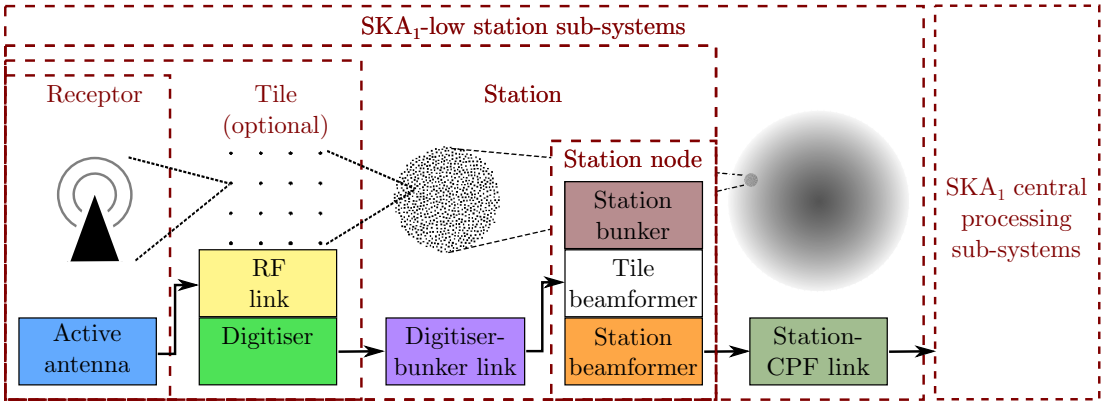


Figure 5.2: Schematic of the all-digital beamforming architecture, showing the signal path through the SKA_1-low sub-systems to the central processing facility. Two or more stages of hierarchical beamforming are assumed to occur within the station beamformer block.

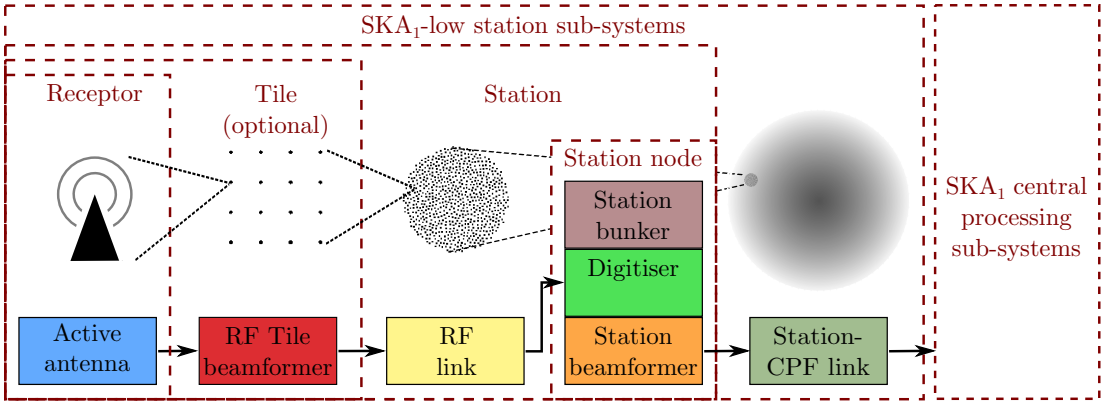


Figure 5.3: Schematic of the RF tile beamforming architecture, showing the signal path through the SKA_1-low sub-systems to the central processing facility.

tile. Figure 5.3 shows the sub-system blocks for the second architecture, which uses a first stage of analogue (RF) beamforming at the tile. In this architecture, the analogue signals are transported to the station processing node and digitisation occurs at that node.

Alongside the all-digital and RF tile beamforming architectures, I consider the different cost estimation methodologies and assumptions used for the two estimates in Faulkner et al. (2011). The cost estimate for the all-digital beamforming architecture primarily uses the bottom-up cost method described earlier; I term this the 'bottom-up SKADS' estimate. The cost estimate for the RF tile beamforming architecture uses LOFAR station costs to make an analogous cost estimate of the SKA_1-low station, hence I term it the 'reference class LOFAR' estimate. Although it is an example of a reference class cost estimate, it is implemented at a more detailed sub-system level, instead of at the system level where the reference class estimation methodology is often applied.

Understanding the epoch of the source cost estimate is also important, especially in regards to the digital hardware. Digital technology advancements can often be generalised by exponential laws; the most common being Moore's law, where the cost of an equivalent digital product

halves every 1.5–2 years. Thus the year for which the cost data applies, and the technical capability of the hardware to which it applies, are relevant factors. The bottom-up SKADS estimate makes the assumption that technology advances reduce costs and that SKA_1-low construction commences in 2016. In contrast, the reference class LOFAR estimate is based on 2007 technology and is only for an output bandwidth of approximately 50 MHz. I make the assumption that by 2016, newer technology will allow for the processing of the full 380 MHz bandwidth for the same cost.

The unit costs for the station sub-systems are derived from the cost data in Faulkner et al. (2011), as described in Appendix F.2. The derivation method is that described earlier, where the cost data is aggregated for each block, and the unit costs are derived from these block aggregates and the parameters in the HLSD (see Table 5.3). The costs, especially some of the digital costs, are necessarily simplistic in order to transcend the design details and multitude of options for each sub-system; however, the simplifications do introduce another level of uncertainty into the analysis. The costing of the dual-band array uses a similar parametrisation process, but applies multipliers (discounts) to the unit costs to account for the design differences between sub-systems of the single-band and low and high-band arrays, as discussed further in Section 5.4.1.

The bottom-up SKADS unit costs are consistently lower than the reference class LOFAR unit costs. From Table F.2 of Appendix F.2, the relative difference in cost is a factor of 2–3 for the active antenna element and fixed cost portion of the station bunker, 4.5 for the digitiser, 26 for the station beamformer and 43 for the variable cost portion of the station bunker. While useful for this first-order analysis, these cost differences show that more work is required to confirm the accuracy and precision of the unit cost derivation and the cost estimates themselves, as Sections 5.6.3 and 5.8 discuss. Note that the bottom-up methodology does not necessarily produce a lower cost estimate than the reference class methodology. Adjustments such as technology improvements and learning curves for mass production can be used to change the cost estimate (NASA, 2008). Thus the relative costs also depend on what cost adjustments are applied to each cost estimate, as well as the method of arriving at the cost data.

Many of the station sub-systems have a cost data source from both the bottom-up SKADS and reference class LOFAR estimates (Table 5.1). Not all aspects of each cost estimate are comparable because of the different architectures used. However, the active antenna element, digitiser, station beamformer and station infrastructure sub-systems are comparable and these costs comprise most of the SKA_1-low station costs. These comparable sub-systems mean that the unit costs derived from the bottom-up SKADS and reference class LOFAR estimates can be applied to either the all-digital or RF tile beamforming architectures, allowing four different cost scenarios to be modelled:

- reference class LOFAR, RF tile beamforming
- bottom-up SKADS, RF tile beamforming
- bottom-up SKADS, all-digital beamforming
- reference class LOFAR, all-digital beamforming.

Because the bottom-up SKADS, RF tile beamforming and reference class LOFAR, all-digital beamforming architectures are extrapolated, they are less optimised for technical performance

and cost than the other two scenarios; this is an extra source of uncertainty for those scenarios. Uncertainties are analysed in Section 5.6.3.

The costs discussed thus far are for the station sub-system hardware. Section 5.5 considers hardware cost for the central processing facility sub-systems. But even with these costs included, this does not represent the total telescope cost. The *Draft SKA costing strategy* (McCool et al., 2010) discusses some of the other costs to be considered for the SKA. The sub-system hardware cost is included in the present analysis, although with some caveats discussed in Appendix F.2. The costs excluded in the analysis are:

- sub-system hardware operations†
- temporary construction and integration facilities
- site operations infrastructure
- construction† (including network trenching)
- annual fibre costs
- antenna siting costs (inclusive of foundations)
- land acquisition
- power infrastructure†
- software development

†these costs can depend on the intra-station architecture, see Section 5.7.3.

Also listed in the McCool et al. (2010) are project overheads, which are outside the scope of this analysis.

However, many of the excluded costs remain approximately constant between the single and dual-band implementations. Hence Section 5.5 makes some zeroth-order estimates of those excluded costs that will vary between implementations, namely site preparation and antenna element deployment costs. This allows for a comparison to be made in the absence of all the cost information.

5.4 Single and dual-band representative implementations

This analysis is based on a single and dual-band representative system, rather than an optimised system. Most of the recent SKA_1 design work, as presented in the sub-system concept design reviews, has been developed with the SKA_1 high-level system description in mind. For this reason, the single-band SKA_1-low in the HLSD is used as the starting point for a comparison of the single and dual-band implementations. No dual-band implementation is described in detail in the *AA Concept Descriptions* document; in this work, the canonical (not optimised) dual-band design has been chosen so that its scientific performance will be similar to the single-band system.

5.4.1 SKA_1-low station design details

Using a dual-band implementation with similar performance characteristics to the single-band implementation ensures the like-for-like comparison. In particular, the low-band array (70–180 MHz) has the same physical layout as the single-band array to achieve the same sensitivity

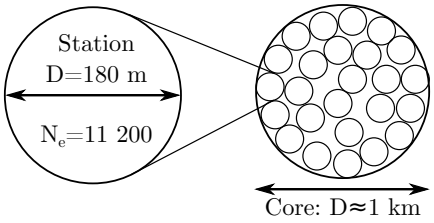


Figure 5.4: Representation of a single-band station within the densely packed core region.

Table 5.3: Single-band SKA$_1$-low system details, as per the HLSD except where noted.

Frequency range	70–450 MHz
Number of stations N_{st}	50
Average spacing between elements d_{avg}	1.5 m ($\lambda/2$ at 100 MHz)
Number of elements per tile[a] $N_{\mathrm{e/tile}}$	16
Station beam taper[b] $\mathcal{K}_{\mathrm{st}}$	1.02[c]
Number dual-polarisation beams per station (averaged over the band) $\overline{N_{\mathrm{b\,-\,st}}}$	210[d]
Dense-sparse transition[e]	115 MHz
Gain of an isolated element	6.2 dBi

[a] For the RF first-stage beamforming architecture.
[b] Where station beam FoV $\Omega_{\mathrm{st}} = \pi/4(\mathcal{K}_{\mathrm{st}}\lambda/D_{\mathrm{st}})^2$.
[c] A uniform aperture distribution is assumed (Rohlfs & Wilson, 2004), whereas the HLSD specifies $\mathcal{K}_{\mathrm{st}} = 1.3$.
[d] This differs from the 160 beams specified in the HLSD.
[e] See Appendix F.5.1.

at the lower frequencies. The high-band array (180–450 MHz) has the same number of antenna elements as the low and single-band arrays, so despite the smaller inter-element spacing, sensitivity is maintained as equivalent to the single-band at most frequencies (see Appendix F.5.1), while reducing the geometrical area occupied by the station. The required processed field of view (FoV) is a minimum of $\Omega_{\mathrm{req}} = 20\,\mathrm{deg}^2$, observed concurrently across the frequency band. Dual-polarisation ($N_{\mathrm{pol}} = 2$), or full-Stokes signals, are assumed throughout this analysis.

The representative single-band implementation is that which is described in the HLSD. The pertinent features of the system are shown in Figure 5.4 (station diameter D_{st} and number of elements per station $N_{\mathrm{e/st}}$) and Table 5.3. As shown in Appendix F.3, the required FoV Ω_{req} is synthesised from an average number of station beams over the band $\overline{N_{\mathrm{b\,-\,st}}}$; enough processing is costed to form these beams. An irregular intra-station element layout of approximately uniform element distribution is assumed.

The key differences in the dual-band implementation are the separate low and high-band stations, and the average inter-element spacing of 0.75 m in the high band. Figure 5.5 and Table 5.4 show the details of this system. The stations and the two cores are assumed to be separate, as portrayed in Figure 5.5. Rather than the two cores shown in the HLSD (the second core being composed of dishes), an SKA$_1$ with a dual-band SKA$_1$-low implementation would have three cores. The separated cores means each core can be densely packed, resulting in a higher filling factor. Such densely packed cores allow for more efficient searches of pulsars and other high

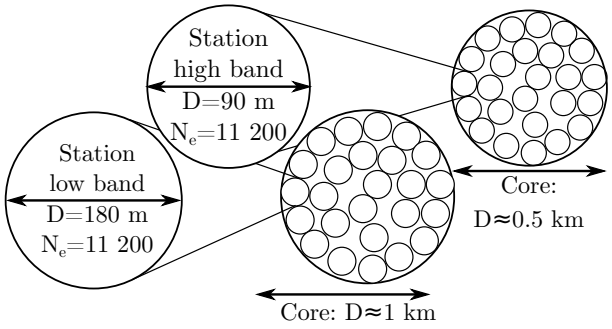


Figure 5.5: Representation of a low and high-band station within two densely packed core regions.

Table 5.4: Dual-band SKA$_1$-low system details.

	Low band	High band
Frequency range	70–180 MHz	180–450 MHz
Number of stations N_{st}	50	50
Average spacing between elements d_{avg}	1.5 m ($\lambda/2$ at 100 MHz)	0.75 m ($\lambda/2$ at 200 MHz)
Number of elements per tile $N_{\mathrm{e/tile}}$	16	16
Station beam taper $\mathcal{K}_{\mathrm{st}}$	1.02	1.02
Number dual-polarisation beams per station (averaged over the respective band)	44	70
Dense-sparse transition	115 MHz	230 MHz
Gain of an isolated element[a]	6.2 dBi	6.2 dBi
Observing mode	Simultaneous (i.e. 70–450 MHz)	

[a] Gain is assumed to be the same for each band, to ensure that the first-order station A/T estimates are comparable. Actual gain values will depend on the antenna element designs.

time resolution events, as Section 4.5.1 outlined. However, other science applications for the high-band (180–450 MHz) would require evaluation to ensure that the array configuration composed of the smaller, 0.5 km diameter high-band core remains suitable. Other configurations, such as interspersed or interleaved stations, is beyond the scope of the present analysis; they require more detailed modelling of the electromagnetic interaction between antenna elements.

The cost of the dual-band implementation is estimated by costing the low and high-band stations separately. This means determining the total cost for each band, then summing the costs. To determine the total cost for each band, the unit cost of every block is given as some fraction of the single-band cost, as detailed in Appendix F.2.9. These costs are considered to be reasonable estimates but are not based on detailed investigation.

For this analysis, infrastructure such as housing for the station processing node is not shared. However, it is assumed that stations are co-located beyond the core, hence the trenching and cables for the data transmission and power to these stations can be shared. Costing the systems separately ensures clarity for comparison purposes; an actual implementation could share some infrastructure and possibly signal processing units, while still being capable of observing the

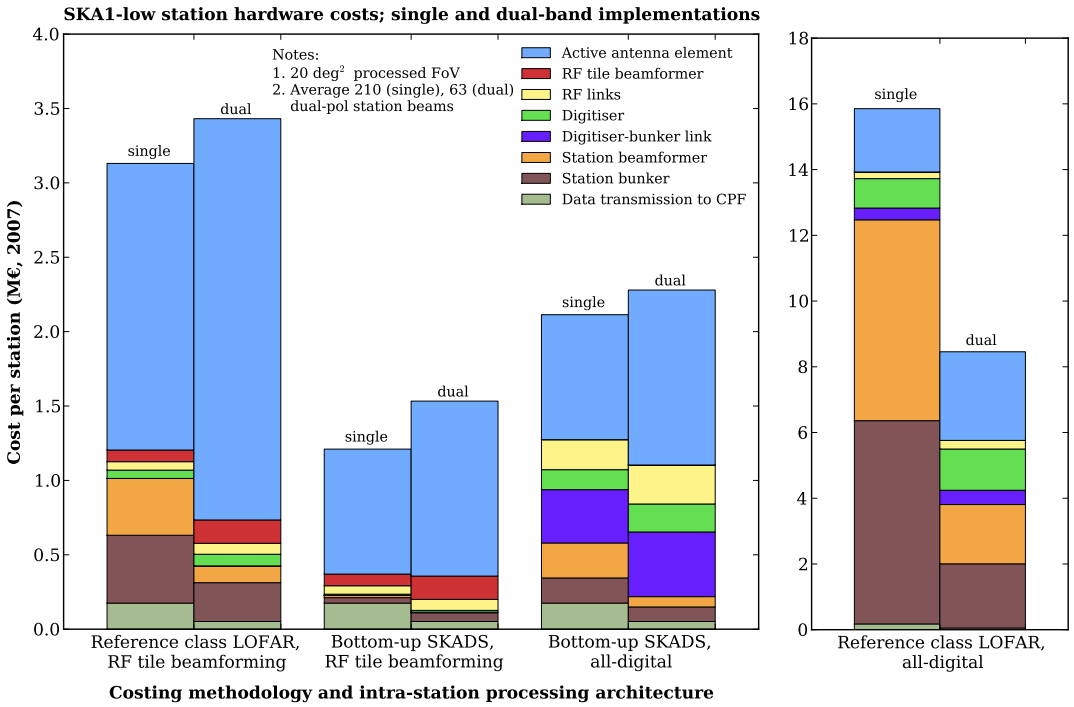


Figure 5.6: SKA_1-low station hardware cost, for permutations of single or dual-band implementations, costing methodologies and intra-station processing architectures. (Reference class LOFAR, all-digital beamforming, right, has a different y-axis scale.) The reference class LOFAR, RF tile beamforming and bottom-up SKADS, all-digital beamforming scenarios are derived from the cost estimates and architectures in Faulkner et al. (2011); the other two scenarios are the extrapolations from those derivations. Station bunker refers to the station beamformer housing and infrastructure (racks, power supply etc.). Sub-systems are colour-coded to match Figures 5.2 and 5.3.

full 70–450 MHz frequency range. Some of the implications of station co-location are discussed in Bij de Vaate et al. (2011) and Faulkner et al. (2010).

5.4.2 Comparison of the station sub-systems

Figure 5.6 plots the SKA_1-low station hardware costs for the four different cost scenarios (Section 5.3.2), each scenario being a combination of:

- reference class LOFAR or bottom-up SKADS cost estimate; and
- RF tile beamforming with digital station beamforming, or all-digital beamforming.

These cost scenarios are plotted for both the single and dual-band implementations. The costs shown here are for the SKA_1-low station sub-systems, from the active antenna element up to and including the data transmission to the central processing facility (see Figures 5.2 and 5.3).

The sub-system costs which dominate the station hardware cost in the single-band implementation differ between the RF tile beamforming and all-digital architectures. For the all-digital architecture, the active antenna element cost is less than half the total station cost. With RF tile beamforming, the majority of the cost is active antenna element costs. All the other blocks have a decreased cost, due to the factor of 16 decrease in the number of signal chains after

the RF tile beamformer (with the exception of the station–CPF link, which remains constant). This implies that understanding the active antenna element costs is more important for the RF tile beamforming architecture. The dual-band array displays the similar broad trend as the single-band array, where the active antenna element cost is more dominant in the RF tile beamforming architecture than the all-digital architecture.

Comparing the single and dual-band implementations shows the driving costs for each. Because the dual-band implementation has twice the number of stations, hence twice as many antenna elements and therefore signal chains, the active antenna element costs are significantly higher. However, the increase is less than double, because of the cost discounts applied in Table F.4 (Appendix F.2.9). These discounts reflect the less onerous requirements on the components in the dual-band implementation, due to the smaller fractional bandwidth and the 180 MHz frequency split between the low and high bands.

The opposing trend is that the smaller average number of beams across the band puts downward pressure on the dual-band cost. This is reflected in the lower cost of the station beamformer, station bunker and station–CPF data transmission sub-systems. The cost reduction is significant for the all-digital architecture, but is less pronounced for the RF tile beamforming architecture, where the reduced number of digital signal paths already decreases the digital processing and data transport costs somewhat.

A conspicuous feature of Figure 5.6 is that the station beamforming and related bunker cost in the reference class LOFAR estimate is significantly larger than the bottom-up SKADS estimate. This is due to the large difference in unit cost estimates (see Appendix F.2.6.1). The most likely reasons for this cost discrepancy are related to the type of beamforming processing technology and architecture, and the technology advancements which have been assumed. For example:

- The LOFAR beamformer is not optimised for processing the larger number of inputs, hence the unit cost derivation may over-estimate the processing cost.
- The unit cost derived from the bottom-up SKADS estimate already includes a cost discount for hierarchical beamforming (see Appendix F.2.6.2).
- The LOFAR station beamformer uses field-programmable gate array (FPGA) processors, while the SKADS design (Faulkner et al., 2010) uses more customised processing chips.
- The technology advancements may be more optimistic for the SKADS estimate than those assumed for the LOFAR estimate (p 103), and production would be for larger quantities than for LOFAR.

The station beamformer cost is an example of the potential for further investigation to determine: the accuracy and uncertainties of each of the cost data sources and their subsequent derivation into unit costs, the accuracy of the first-order station beamformer model, the requirements that contribute to the beamformer cost and design solutions or trade-offs to reduce this cost.

5.4.3 Cost reduction from analogue (RF) tile beamforming

While not the main focus of this work, the parametric analysis allows a preliminary cost comparison to be made between the two intra-station architectures: RF tile beamforming and

Table 5.5: SKA_1-low station cost for RF tile beamforming, as a percentage of all-digital beamforming cost.[a]

Cost estimate	Implementation	
	Single-band	Dual-band
Reference class LOFAR	20 %	41 %
Bottom-up SKADS	57 %	67 %

[a] Percentage shown applies to that cost estimate and implementation combination.

all-digital beamforming. The results show that for tiles composed of 16 elements, a first stage of analogue tile beamforming significantly reduces the station hardware cost. The cost reduction is irrespective of whether a single or dual-band system is implemented, or the bottom-up SKADS or reference class LOFAR cost estimate is used.

The station costs for RF tile and all-digital beamforming architecture can be compared in Figure 5.6, for the single and dual-band implementations and reference class LOFAR and bottom-up SKADS cost estimates. Table 5.5 makes a direct comparison for each implementation and cost estimate combination. The cost of the SKA_1-low sub-systems with analogue tile beamforming is 20–67 % of the all-digital beamforming; a factor of approximately 1.5–5 reduction in cost. The cost reduction, both in relative and absolute (euro) terms, is most significant when the cost of the digital sub-systems is high (e.g. the single-band implementation of the reference class LOFAR cost estimate). However, no cost reduction applies to the sub-systems in the central processing facility; because those sub-systems act on station beams, their costs are independent of the intra-station architecture.

The cost reduction is due to the FoV accessible at the station beamformer (and inherent observational flexibility) being restricted early on in the signal path; the number of digital signal chains is reduced by a factor equal to the number of elements per tile, in this case 16. The cost of the digitiser and station beamformer blocks is thus reduced, as can be seen in Figure 5.6. This is because fewer digitiser and digitiser–bunker links are required, and fewer inputs into the station beamformer reduces the processing load.

Faulkner et al. (2010) present a qualitative comparison of the all-digital and RF tile beamforming approaches. The all-digital beamforming is more flexible, in terms of generating multiple beams, RFI excision and calibration of the antenna elements, if needed. With upgraded digital signal transport and processing, the correlation of every antenna element is also be possible; whereas only the RF beamformed tiles can be correlated, not the individual elements. The main disadvantage for the all-digital beamforming approach is cost, as outlined, as well as the increased power demand of the digital components, and the distribution (or local generation) of this extra power throughout the array. Section 5.7.3 further discusses the cost and performance implications of the beamforming approaches, in the context of the intra-station architecture.

Table 5.6: Attributes of the dual-band implementation compared to the single-band.

Attribute	Percent of single-band
Number of antenna elements	200 %[a]
Physical area	125 %
Average number of station beams[b]	30 %

[a] Half these elements are physically smaller than the single-band elements.
[b] Formed over the full 70–450 MHz band to achieve 20 $\deg^2$.

5.5 System implications of variable costs

The results thus far present the costs of the SKA_1-low station hardware for each scenario and representative implementation. However, there are other cost implications on the telescope system, some of which are considered in this section. To put the costs analysed here in the context of the whole SKA budget, the simplest comparison between the single and dual-band implementations is to say that some costs remain approximately constant, such as project overheads, and some costs vary between the two implementations, such as the SKA_1-low stations costed in Section 5.4.

The variable costs considered in this analysis to most significantly impact on the total system cost are:

- station sub-systems
- antenna element deployment
- site preparation
- central processing sub-systems
- power provision and distribution.

There is no published work on the relative cost between these, hence the full effect of each implementation on the total cost is difficult to determine. However, the different attributes of the two implementations, as shown in Table 5.6, gives some insight into the cost trends. Appendix F.4 makes a more detailed analysis of these variable costs and estimates some zeroth-order costs.

5.5.1 Antenna element deployment and site preparation costs

The cost of deploying the antenna elements (i.e. building the array on-site) will be higher for the representative dual-band array, because twice the number of antenna elements need to be deployed. Because the element size is defined by being sufficiently electrically large at the lowest observing frequency, it is reasonable to expect that the size of the low-band element will be similar to the single-band element, such that comparable antenna gain is obtained. However, the high-band element will be significantly smaller, and cheaper manufacturing and deployment

options may be available. This means that the increased deployment cost for the dual-band array would be less than 200 %. Deployment is further discussed in Faulkner et al. (2011).

On a related topic, it is reasonable to expect that some fraction of the site preparation cost will increase with the physical area occupied by SKA_1-low. However, whether this cost is significant relative to the total site preparation costs is not known. Site-related data is being collected as a part of the site selection process (Schilizzi et al., 2011) and is not currently available. I have used the initial deployment and site preparations cost estimates outlined in Section 5.5.3 to illustrate their potential significance.

5.5.2 Central processing facility sub-systems

The central processing facility is a broad term to encompass the signal processing and science computing sub-systems in the HLSD. The processing is centralised, because it acts on signals from all the antennas (AAs and dishes) in the array. However, it does not necessarily imply that all these sub-systems will be on-site; the on-site processing is required to sufficiently reduce the rate of data sent to the off-site processing. The principal sub-systems are the correlator and imaging processor, and the non-imaging processor. The correlator and non-imaging processor costs are derived the Signal Processing Concept Design Review, as collated in Turner (2011). These costs focus on the processing units required (the sub-system hardware), rather than total cost of the sub-systems. The parametric cost equations of the major sub-systems within the central processing facility are derived in Appendix F.4.2; these sub-systems are costed on a 'per station beam' basis.

Table 5.7 summarises the relative costs between the representative single and dual-band implementations. The difference in cost between the two implementations is due to the larger beam size of the smaller (90 m) diameter high-band station. This reduces the number of station beams required to fill 20 $\deg^2$ FoV, resulting in lower central processing facility costs. To ensure comparable performance between the single and dual-band implementations, 20 $\deg^2$ FoV over the full 380 MHz bandwidth is correlated and imaged. Following Alexander et al. (2009), the cost of the imaging processor is assumed to be dominated by the cost of the data buffer rather than the imaging operations. The non-imaging processor acts on phased or 'tied' array beams formed from the densely packed core stations. Only array beams formed with the high-band core of the dual-band implementation are considered, because the current required frequency range for pulsar surveys with the non-imaging processor is 0.3–3 GHz (SSWG, 2012).

A key aspect of the correlator–imaging data transport and imaging processor cost is that for SKA_1-low, the correlator frequency resolution requirement is derived from the more stringent science requirements, rather than from the requirement to keep radial (u, v) smearing below an acceptable threshold (see Appendix F.4.2.1). If instead the latter dominates, then the correlator output data rate becomes independent of station diameter, for a fixed processed FoV. In that case, the imaging processor is therefore the same for the single and dual-band implementations.

Table 5.7 only applies to items such as processing components (and associated hardware and cooling), the cost of which, as a first-order approximation, scales linearly with the processing load. Correlator and non-imaging processing costs are summarised in Turner (2011), representing a range of processor technologies and architectures. Those costs are highly dependent

Table 5.7: Dual-band implementation central processing sub-system costs, compared to the single-band.

Central processing sub-system	Percent of single-band
Correlator	30 %
Correlator–imaging data transport[a]	53 %
Imaging processor[a, b]	53 %
Non-imaging processor (NIP)[c]	25 %

[a] Required correlator frequency resolution is derived from the science requirements (Appendix F.4.2.1).
[b] Cost dominated by data buffer (Appendix F.4.2.4).
[c] Only the high-band core is used in the NIP, and the processing for the AAs, not the dishes, dominates the cost (Appendix F.4.2.5).

on the processing technologies and the trade-off between efficient processing devices with larger development (non-recurring engineering) costs, and less efficient but more flexible processing devices.

5.5.3 Overall SKA_1-low costs

Although obtaining accurate total costs of the single and dual-band implementations is not yet possible, some zeroth-order estimates can be used to illustrate the system-level costs outlined in this section. Figure 5.7 plots these significant variable system costs (excluding power) for single and dual-band implementations. The plot includes the four different station cost scenarios, reflecting the different cost estimates, and intra-station signal transport and processing architectures. The variable and 'other' costs remain unchanged for each scenario; they are independent of the intra-station architecture and station cost estimates. The correlator–imaging data transport cost is not significant ($<1\,\%$) and is not plotted. To indicate the sensitivity of the comparison to changes in the deployment and site preparation costs, Figure 5.8 is a similar plot, but the deployment cost is doubled to €100 per antenna element, and the site preparation cost increased by an order of magnitude to $€100\,\mathrm{m}^{-2}$.

5.5.4 Power costs

Investigations for the provision and distribution of power for the SKA are on-going (e.g. Hall, 2011), alongside analyses of the power demand of the telescope sub-systems. Although the details of supplying and distributing power are beyond the scope of this work, simplified power demand estimates are possible with the parametric models. To make a zeroth-order estimate of the power costs for the SKA_1-low stations, it is reasonable to expect that the power demand of each sub-system is linearly proportional to one or more of the following:

- number of antenna elements
- digital processing load
- number of station beams formed, hence data rate transmitted from the stations to the CPF.

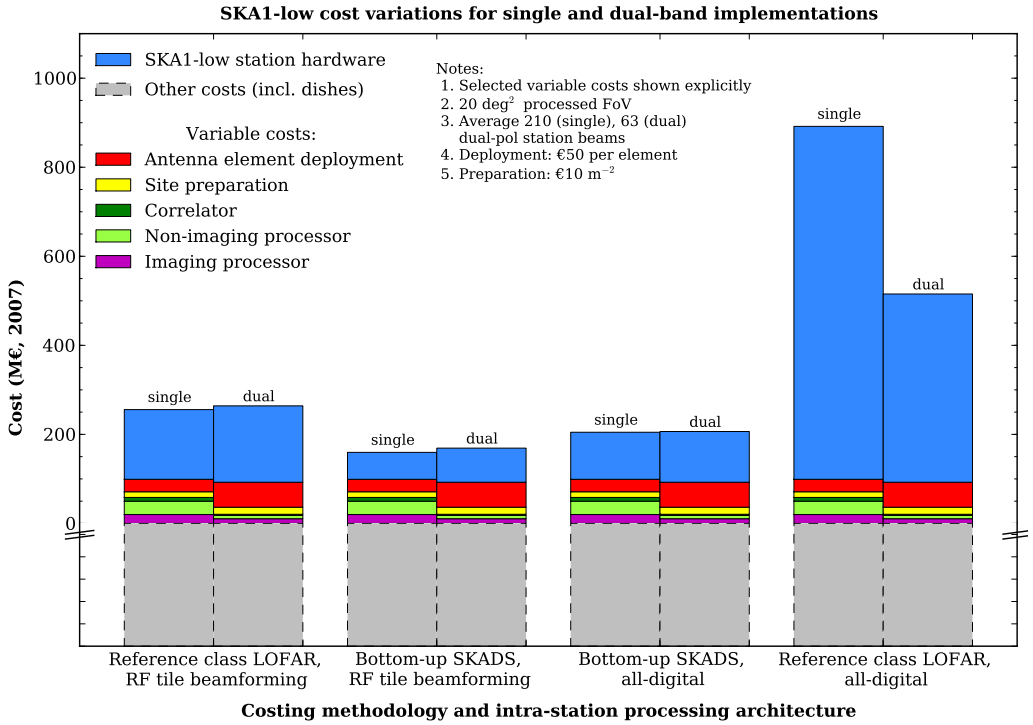


Figure 5.7: Significant variable costs (excluding power) for the representative single and dual-band implementations of SKA_1-low, for the different cost estimates and intra-station architectures. A deployment cost of €50 per element and site preparation cost of €10 m^{-2} is assumed. 'Other costs' is a placeholder for the costs which do not differ between implementations.

Figure 5.9 shows power demand, estimated from the bottom-up SKADS power budget in Faulkner et al. (2011). The 'unit power demand' is parametrised, as was done for the cost data, and listed in Appendix F.4.3. The power budget is for the all-digital station architecture, but is extrapolated to the RF tile beamforming architecture by including an estimate for the RF beamformer power demand.

The station power demand depends on both the implementation and intra-station architecture. The dual-band implementation reduces the station beamformer power demand, but the demand of the other station sub-systems is increased. For the all-digital architecture, this results in a significantly lower power demand for the dual-band implementation. But for the RF tile beamformer architecture, there is little difference between implementations. Rather, the demand reduction from RF tile beamforming is the dominant effect.

The method, hence cost, for supplying the power demand of sub-systems within the station will depend on the intra-station architecture, as discussed in Section 5.7.3.3. The intra-station power distribution cost and power supply inefficiencies depend on load and distance, making the cost specific to each intra-station architecture. For example, in the RF tile beamforming architecture, power at the antenna elements is supplied via the RF link to the station. In contrast, the all-digital architecture transmits data via optic fibre, hence requires additional power distribution cabling (which is not costed here). Some architectures do not even have

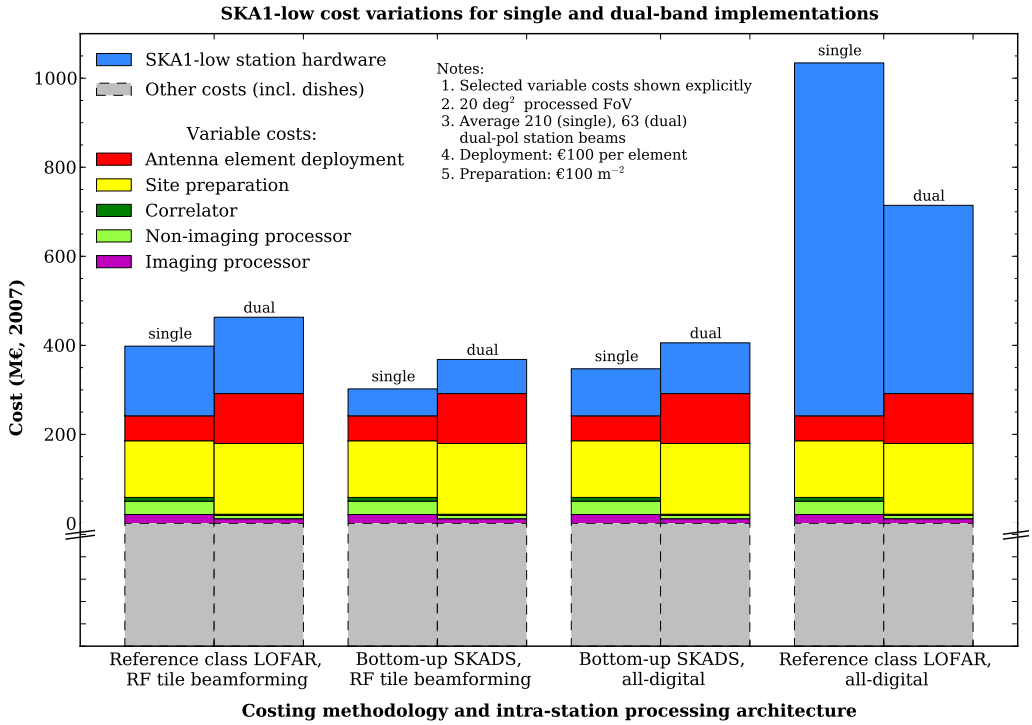


Figure 5.8: As per Figure 5.7, except a deployment cost of €100 per element and site preparation cost of €100 m^{-2} is assumed.

intra-station distribution costs, because the active antenna elements are self-powered. Bearing these caveats in mind, Figure 5.9 assumes an 85 % supply efficiency, and includes a power consumption cost of €0.12 per kWh for illustrative purposes.

For the central processing facility, the absolute power demand depends on the technologies used; in general, there is an inverse relationship between the flexibility and power efficiency of the processing (Hall, 2011). But given that both cost and power demand increase with the amount of processing required, the values in Table 5.7 can be taken as an indication of the relative power demand between the single and dual-band implementation: the central processing facility power demand is lower for the dual-band implementation.

5.6 Discussion of principal analysis

The merits of single and dual-band implementations of SKA_1-low can be considered in terms of performance, cost and design flexibility. This section summarises the cost and performance trends, and the cost estimation uncertainties. I also discuss the challenge of designing a single-band implementation which meets the sensitivity requirements at the lower frequencies and achieves a sufficiently high filling factor (a measure of array sparsity), which may assist in calibration at the higher frequencies.

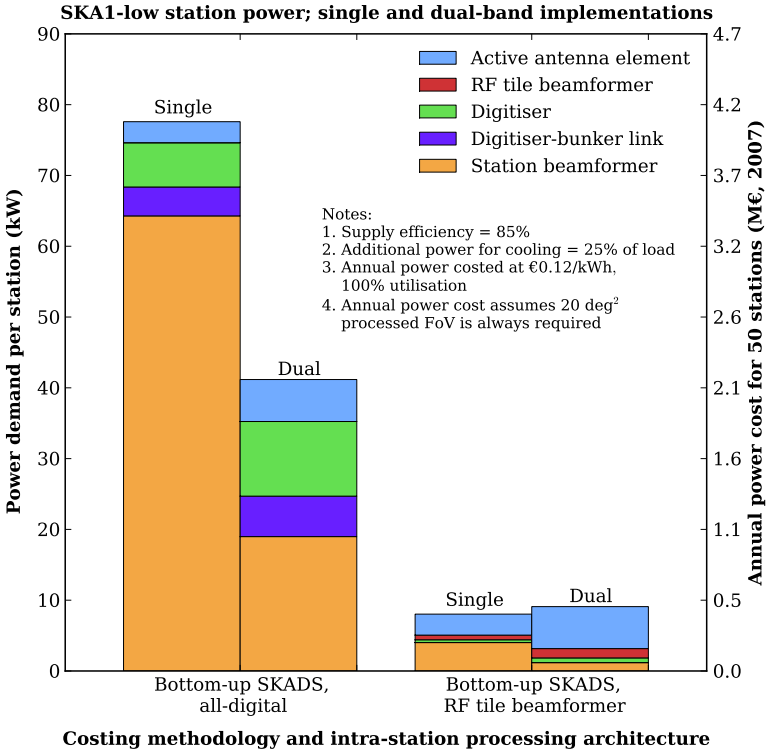


Figure 5.9: Single SKA_1-low station power demand estimate and annual power cost for single and dual-band implementations, and RF tile beamforming and all-digital intra-station processing. As per Faulkner et al. (2011): i) the power supply efficiency is assumed to be 85 % (a 1/0.85 increase in power demand); and ii) an additional 25 % power demand is applied to account for cooling the hardware. Power consumption is costed at €0.12 per kWh (Hall, 2011) and assumes 100 % (continuous) utilisation of the hardware.

5.6.1 Cost trends

The cost difference between the single and dual-band implementations is fundamentally a comparison between costs that scale with the number of signal paths prior to station beamforming and costs that scale with the number of station beams. The single-band station is very sparse at high frequencies, requiring many more beams than the dual-band station to meet a given FoV requirement. This not only increases the station beamformer and bunker cost, but also the station–central processing facility (CPF) data transmission and central processing facility sub-system costs. The dual-band implementation has twice as many signal paths prior to station beamforming, which increases cost. However, this cost is less than double that of the single-band, primarily because the smaller fractional bandwidth entails less demanding design specifications for the active antenna elements.

The dual-band array decreases the total number of station beams to be formed, thereby reducing the corresponding signal processing costs. The decrease in the number of beams comes about from the smaller high-band station diameter. However, this is predicated on the system having sufficient processing capacity to concurrently form and process the number of beams corresponding to 20 $\deg^2$ FoV across whole 70–450 MHz band. If less processing capacity is

required for the high-band array, the cost advantage when implementing the dual-band array is lessened. This is further investigated in Section 5.7.2.

A related effect is the inter-element spacing chosen for the high-band array. A smaller inter-element spacing further reduces diameter, and therefore cost. However, this ultimately begins to reduce the high-band array sensitivity, as discussed in Appendix F.5.1. The frequency split of the dual-band array is another factor. The key metric of the average number of beams per station ($\overline{N_{\mathrm{b\text{-}st}}}$) not only depends on inter-element spacing, but also the frequency split between the low and high-band arrays, as shown in Appendix F.3.2. The frequency split also affects the cost of the active antenna elements and digitiser (Appendix F.2.9), as they depend on minimum and maximum frequencies of each band, and the fractional bandwidth.

For the representative single and dual-band implementations, there are not significant differences in station hardware costs, except for the reference class LOFAR, all-digital beamforming scenario. But the variable costs which impact the system (Section 5.5) may be a discriminating factor between implementations. As discussed, the dual-band implementation lowers the cost of the CPF sub-systems. However, the increased number of active antenna elements and the extra area required for these elements increases the deployment and site preparation costs.

Power costs also bear careful consideration; capital cost may be significant and power is a major operations cost. This work shows the station power demand of the dual-band implementation equals that of the single-band when RF tile beamforming is used, and is less than the single-band with the use of all-digital beamforming. The RF tile beamforming itself significantly reduces power demand for both implementations. For the CPF, the dual-band implementation reduces the power demand.

There are also some cost trends which depend on the intra-station architecture, but which do not affect the other variable costs which impact the system. The noteworthy trend is the significantly lower station cost for RF tile beamforming compared to all-digital beamforming. In fact, the differences in station cost between these architectures is more significant than the differences between the single and dual-band implementations. Also, Bij de Vaate et al. (2011) discusses a dual-band implementation which shares a common processing system to reduce processing requirements, although exactly how this sharing occurs is not specified. While not specifically costed in this analysis, alternative intra-station architectures (including shared processing) are discussed in Section 5.7.3.

5.6.2 Performance trends

The performance characteristics of the canonical dual-band implementation used in this analysis is comparable to the single-band implementation described in the SKA_1 high level system description. This is achieved by using separate low and high-band arrays, observing simultaneously, to create a $20\,\mathrm{deg}^2$ processed FoV over the 70–450 MHz band, while maintaining a sensitivity curve across the band which is similar to the single-band implementation. However, these representative implementations are not optimised for performance.

The performance and design flexibility of the implementations is compared in Bij de Vaate et al. (2011) and potential performance issues for the single-band system are identified. In particular,

an antenna element design suitable for the wide fractional bandwidth of about 6.5 : 1 is required. The antenna must then be matched to the low noise amplifiers (LNAs) across the frequency band, which is more difficult for the wide fractional bandwidth. These steps are less challenging for the dual-band implementation, with the 2.5 : 1 fractional bandwidth in each band.

Another key performance issue is the frequency-dependent sensitivity (A/T) curve; the frequency range for which the sensitivity is optimised is a high-level design issue (Alexander & Hall, 2010). Figure 23 of the HLSD makes a simple estimation of sensitivity for SKA_1-low; this curve is replicated in Appendix F.5.1 (Figures F.1 and F.2). A key feature of the curve is that the sensitivity is lowest at the minimum (70 MHz) and maximum (450 MHz) frequencies of the band, with a peak between 100 and 200 MHz; there is also a factor of 3 difference between the highest and lowest sensitivity values.

Appendix F.5.1 shows the effect of inter-element spacing on sensitivity. In the single-band implementation, the inter-element spacing must be chosen to best satisfy requirements across a larger band, resulting in sub-optimal spacing at the lowest and highest frequencies in the band. At 70 MHz, the sensitivity is limited by the 1.5 m inter-element spacing. At 450 MHz, the 1.5 m spacing means that more beams are required to form the 20 deg^2 processed FoV, increasing the processing costs, as discussed in Section 5.6.1.

The dual-band implementation uses a smaller inter-element spacing in the high band, thus it can achieve beamformer cost savings in the high-band, while maintaining similar sensitivity to the single-band implementation. The only exception is the reduced sensitivity between 180 and 230 MHz. Another aspect of the dual-band flexibility is that a different number of antenna elements could potentially be used in the low and high band arrays, to better suit sensitivity requirements.

A single-band implementation may also present a greater calibration challenge than a dual-band array. Station calibration requires a sufficiently high filling factor so that enough calibration sources are detectable within a station beam FoV (see Appendix F.5.2). At 400 MHz, the filling factor of the single-band station is too low to detect the required number of calibration sources. The representative dual-band implementation solves this problem because the filling factor for the high-band station is greater.

5.6.3 Risk and uncertainty

As mentioned in Section 3.6, the parametric cost estimate should provide an associated uncertainty estimate, which is large in the early stages of the project. As the project progresses, this uncertainty reduces and the cost estimate eventually converges on the actual cost. This section discusses the station hardware cost uncertainties and makes an initial uncertainty estimate for the reference class LOFAR, RF tile beamforming scenario.

5.6.3.1 SKA_1-low station uncertainties

The range of cost estimates for the SKA_1-low station hardware give some indication of how sensitive the single and dual-band implementations are to different cost projections. Table 5.8 provides a summary of the single and dual-band station costs plotted in Figure 5.6. All costs

Table 5.8: Comparison of SKA_1-low station cost for the single and dual-band implementations.

Cost scenario		Implementation[a]		Dual to single-band ratio
		Single-band	Dual-band	
RF tile beamforming	Reference class LOFAR	100 %	110 %	108 %
	Bottom-up SKADS	39 %	49 %	127 %
All-digital beamforming	Bottom-up SKADS	68 %	73 %	110 %
	Reference class LOFAR	506 %	270 %	53 %

[a] Percentage of the single-band, reference class LOFAR, RF tile beamforming scenario.

are normalised to the cost of the reference class LOFAR, RF tile beamforming scenario of the single-band array. For each cost scenario, the table shows the normalised total for each implementation and the ratio of the dual-band station cost to the single-band. For three of the scenarios, the dual-band station is more expensive than the single band, while for the fourth, the opposite is true.

In Table 5.8, there is a factor of 7.5 difference between the reference class LOFAR and bottom-up SKADS cost estimates of the all-digital, single-band implementation, and a factor of 3.7 for the dual-band. For the scenarios with RF tile beamforming, the difference is a factor of 2.6 and 2.2 for the single and dual-band respectively. Some of the unit costs only have a single data source (estimate) and are used for both cost estimates; those costs do not contribute to the cost differences. Some potential reasons for the cost differences were discussed in Section 5.3.2, but this is an area for further investigation.

Importantly for the single and dual-band comparison, the range in the relative station hardware costs between the two representative implementations (as indicated by the final column of Table 5.8) is much lower than the range of absolute station hardware costs. This implies that the uncertainty of the single and dual-band station cost comparison is less than that of the individual station cost estimates.

5.6.3.2 Sensitivity analysis: cost drivers

Within the context of the SKA_1-low stations, an inspection of Figure 5.6 gives an indication of the cost drivers. The dominant blocks are the active antenna element, digitiser–bunker links and the station beamformer and bunker. For both the single and dual-band implementations, the active antenna element is the largest cost block, except in the reference class LOFAR, all-digital beamforming scenario where the station beamformer and bunker costs dominate. For the bottom-up SKADS, all-digital beamforming scenario, the digitiser–bunker links are the second most costly block. However, the large variation of the cost of many of the blocks between the scenarios indicates that the cost estimates require further refinement before a conclusive set of cost drivers can be determined. The drivers also depend on the intra-station architecture, as discussed in Section 5.7.3.

Additionally, the other variable costs listed in Section 5.5 have the potential to be cost drivers

in the comparison of single and dual-band implementations, and could exceed the most costly individual station sub-system. For example, Figures 5.7 and 5.8 in effect form a rudimentary analysis of the sensitivity of the SKA_1-low cost to changes in variable antenna element deployment and site preparation costs.

5.6.3.3 Statistical uncertainty analysis

Although the input unit cost distributions require further expert attention, I make an initial estimate of the uncertainty of the reference class LOFAR, RF tile beamforming scenario, using the Monte Carlo simulation in SKACost, described in Section 3.6.3. For this estimate, I only consider the uncertainties of the two largest costs; the active antenna elements and the station beamformer processing. For the active antenna element (which includes the ground plane), a triangular probability distribution function is applied, where the minimum unit cost is that derived from the bottom-up SKADS estimate (€75 per element), the most likely unit cost is that derived from the reference class LOFAR estimate (€172 per element) and the maximum cost is twice that (€344 per element). For the station beamformer processing and the portion of the bunker which scales with the amount of processing, a top-hat cost distribution is applied, because the most likely unit cost is unknown at present. The top-hat probability distribution function is used when only the minimum and maximum values are known, and the true cost could lie anywhere, with equal probability, between these limits. I use the unit processing cost derived from the bottom-up SKADS estimate (€0.1 per beamformer input per output beam) as the minimum value and the reference class LOFAR estimate (€2.6 per beamformer input per output beam) as the maximum value.

Figure 5.10 shows the resulting PDF of the Monte Carlo analysis of the reference class LOFAR, RF tile beamforming scenario with the above-mentioned input PDFs. The station cost is €$3.1\,(+0.8, -0.6)$ M (these percentiles are equivalent to the mean and 1 standard deviation of a Gaussian curve, 50 % and 15.9 %, 84.1 % respectively). Figure 5.11 plots this mean and error onto the station hardware cost break-down. For comparison, the single-value station cost plotted in Figure 5.6 is €3.1 M. With due consideration of the PDFs of the major costs, similar uncertainty analyses can be made of the other station scenarios, and also the other variable costs (Section 5.5).

5.6.4 Relevance to SKA Phase 2

This parametric modelling analysis is applicable to SKA Phase 2 (SKA_2), although the unit costs and scientific requirements will differ. SKA_2 will most likely use the low-frequency receptors defined for SKA_1. If the intra-station signal transport and processing architecture is similar, then the parametric equations for the station hardware costs (Appendix F.2) will still apply. The main difference will be the cost for each block. Although the cost of the digital blocks will reduce, the FoV requirements for SKA_2 will be much higher. Thus, it is possible that any digital cost reductions will be offset by the extra digital processing needed to form many more station beams to achieve the required FoV. For the other attributes shown in Table 5.6 (number of antenna elements, physical area and average number of station beams), the ratio between the two implementations still holds, regardless of required FoV. The same applies to the central processing facility costs (Table 5.7).

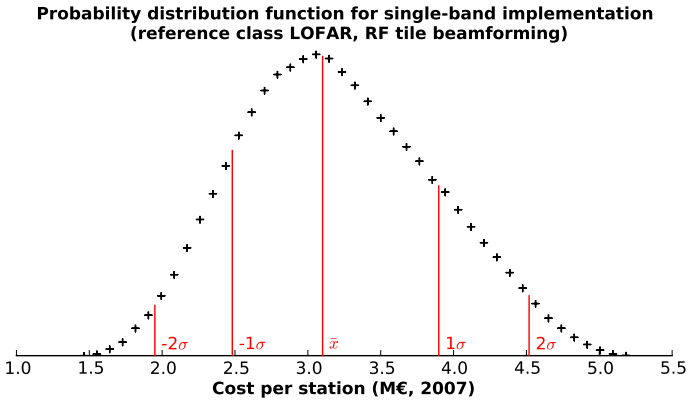


Figure 5.10: Probability distribution function for the single-band implementation of the reference class LOFAR, RF tile beamforming scenario, as calculated using SKACost. The crosses show the probability density for each sample bin and the red lines show the percentiles equivalent to the mean and 1 and 2 standard deviations of a Gaussian distribution. Refer to the text for the input probability distribution functions used.

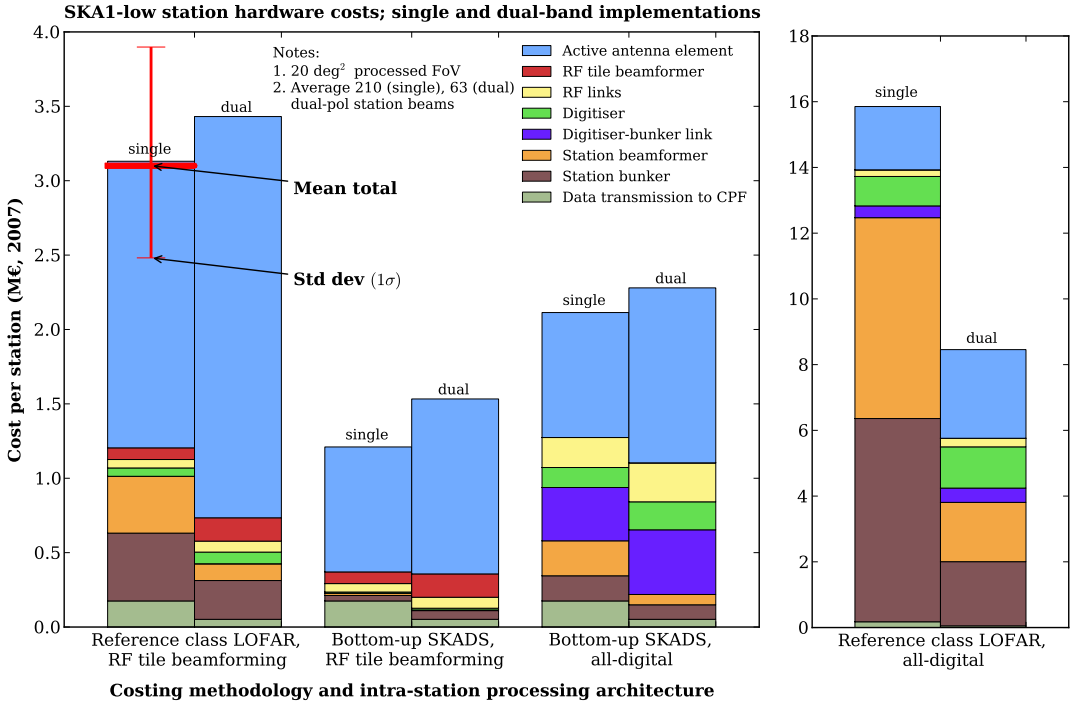


Figure 5.11: SKA_1-low station hardware cost as per Figure 5.6, but with the mean and 1σ error bars from Figure 5.10 shown for the reference class LOFAR, RF tile beamforming scenario.

The SKA_2 science requirements must also be kept in mind when comparing the single and dual-band implementations. These requirements may place a difference emphasis on the frequency-dependent sensitivity and FoV performance, compared to SKA_1; this may affect the optimal inter-element spacing. While shorter term gains might be made by optimising the design for SKA_1, there may be implications for SKA_2 cost and performance which necessitate further investigation.

5.7 Supplementary analyses

In this section I extend the analysis to include some topical factors important to the design of the SKA_1-low system.

5.7.1 Varying station diameter

The HLSD discusses a potential change to the SKA_1-low system, where diameters smaller than 180 m are used, but the number of stations N_{st} is increased to maintain the SKA_1-low array effective area $A_{\mathrm{e\,-\,arr}}$, thus sensitivity. The problem is posed as a trade-off between aperture array station diameter and the number of stations (where $N_{\mathrm{st}} \propto 1/D_{\mathrm{st}}^2$). But as Appendix F.5.3 shows, station diameter is not an independent parameter, because $D_{\mathrm{st}} \propto \sqrt{N_{\mathrm{e/st}}} d_{\mathrm{avg}}$, where $N_{\mathrm{e/st}}$ is the number of elements per station and d_{avg} is the average inter-element spacing. Assuming d_{avg} remains constant, a more exact description is that the number of elements per station is traded with the number of stations.

To illustrate the cost sensitivities of this N_{st} vs. $N_{\mathrm{e/st}}$ trade-off, Appendix F.6 considers a simple comparative example, where the diameter of every single-band, low-band or high-band station is halved (reducing $N_{\mathrm{e/st}}$ by a factor of 4), which results in N_{st} increasing by a factor of 4 to maintain $A_{\mathrm{e\,-\,arr}}$. The inter-element spacing for each band remains constant.

The hardware cost of the SKA_1-low stations shows no dominant trend between the smaller stations and their full-sized counterparts (Figure F.3, p 217). For both the single and dual-band implementations, reducing the station diameter:

- decreases the total station beamformer cost and variable bunker cost
- decreases the total cost of links from the antenna element or tile to the processing bunker
- increases the total fixed bunker cost.

The effect of these trends on the total cost of the stations depends on the implementation, intra-station architecture and cost data source. However, given the cost uncertainty discussed in Section 5.6.3, there is no dominant trend for the station hardware cost of the smaller stations relative to their full-size counterparts; only the single-band implementation of the reference class LOFAR, all-digital beamforming scenario shows a significant change in total station hardware cost. (In this case a reduction in cost.)

Of the variable costs outlined in Section 5.5, only the correlator and imaging processing costs change with the N_{st} vs. $N_{\mathrm{e/st}}$ trade-off; they increase by a factor of 4 and 8 respectively (Figure F.4, p 219). The deployment and site preparation costs are unchanged, because the total number of antenna elements in the array remains constant. Although the correlator and

the imaging processor become significant costs for the half-diameter station example, these are zeroth-order estimates. If smaller stations are being considered, more detailed investigation is required as to the accuracy of using the correlator output data rate as a proxy for the imaging processor cost scaling, and of the correlator and processor cost estimates themselves.

Cost is not the only factor to consider; there is also the effectiveness in meeting science requirements. For example, meeting (u, v) coverage requirements can depend on increasing $N_{\rm st}$ (e.g. Bolton et al., 2009b; Lal et al., 2009). But distributing the extra stations to improve (u, v) coverage will likely have additional infrastructure costs (Bolton et al., 2011). To not disadvantage the half-diameter example, I have chosen not to analyse a more widely distributed array configuration in this study. For stations within the core, Appendix F.6.3 discusses how $N_{\rm st}$ may be efficiently increased by creating a number 'logical' stations, each with fewer elements, which share a processing node and form a 'physical' station. However, any requirement to increase $N_{\rm st}$ needs to be traded with the signal processing costs; in such a trade-off, the hardware prior to the station beamformer is not a cost driver, but the imaging processor (and to some extent the correlator) has the potential to be a large cost driver.

Re-configurable station sizes through logical stations, especially in the SKA core, may allow for a cost-optimised station architecture that better meets scientific requirements. There are some real-world examples of aperture array telescopes with differing logical and physical stations. The Murchison Widefield Array (MWA) has 128 aperture array tiles, each with an RF beamformer, distributed out to a distance of $\sim$1.5 km from the array centre. A receiver node digitises and applies a coarse filterbank to the single tile beam from each RF beamformer and transmits the digital data to the correlator (Tingay et al., 2012). The receiver node serves 8 tiles, rather than each beamformer having its own node. In this context, the tiles, as inputs to the correlator, are the logical stations and the 8 tiles connected to the receiver node forms the physical station. The LOFAR stations also have different logical and physical stations. The 'core' stations at the inner region of the telescope are each composed of a set of low band antennas (LBA) and two sets high band antennas (HBA). These antennas are served by a single processing node (Gunst & Bentum, 2010). Taking the concept of re-configurable station sizes to the extreme, each logical station could simply be a single antenna element as an input to the correlator. This approach is already being implemented on LOFAR as an 'all-sky monitor' for transients (Prasad & Wijnholds, 2012).

5.7.2 Reducing the FoV requirement: defining a fixed beam–bandwidth product

Other trade-offs are emerging through analysis of the $\rm SKA_1$ Design Reference Mission ($\rm DRM_1$). $\rm DRM_1$ captures the set of observations required to achieve the $\rm SKA_1$ science goals and forms an 'envelope' of technical requirements for the telescope. One possible performance–cost trade-off is to reduce the $\rm SKA_1$-low signal processing capacity, defined by the product of the bandwidth and the average number of station beams formed. In this approach, the processing capacity only meets the beam–bandwidth product required by the most demanding science application in $\rm DRM_1$. In contrast, the $\rm SKA_1$ high-level system description (HLSD) and the results presented thus far assume sufficient signal processing capacity to concurrently observe 20 $\rm deg^2$ of processed

FoV over the entire 70–450 MHz band.

To understand the cost advantages from such a trade-off, Appendix F.7 considers the cost of a strawman design, where the representative single and dual-band implementations are modified such that the signal processing capacity is defined by the requirement to only observe $20\,\mathrm{deg}^2$ across the 70–180 MHz band, resulting in a beam-bandwidth product of 4.8 GHz. By comparison, the beam-bandwidth capacity of the canonical single and dual-band implementations presented here are 80 GHz and 24 GHz respectively, the latter being smaller because fewer beams are required in the 180–450 MHz frequency range to form the $20\,\mathrm{deg}^2$ FoV.

Figures F.6 and F.7 in Appendix F.7 show that for such a strawman, the cost of the sub-systems which scale with the number of station beams are no longer significant (those being the station beamformer and station–CPF transmission, as well as the correlator and imaging processor). The costs which dominate are those which scale with the number of antenna elements. Thus the dual-band implementation, with twice the number of elements, is more expensive in all scenarios, although the increase is less than the factor of two increase in cost which one might naively expect for twice the number of elements. However, this difference still makes the trade-off potentially significant.

5.7.3 Intra-station signal transport and processing architecture considerations

The scalable high-level view of the system, which follows the elemental signal path (Section 5.3), allows for the realisation of various signal transport and processing architectures. Section 5.3.2 describes the architectures used in this analysis, but a greater number of data transport and processing architectures are conceivable, through the combination of:

- using different technologies to perform the sub-system function
- re-arranging the order of the sub-systems in the signal path
- changing physical location the sub-system.

Table 5.9 shows the most relevant options for intra-station signal transport and processing architectures. Although comparing the cost-effectiveness of the different architectures is beyond the scope of the present analysis, this section discusses the performance and cost implications of some of these architectures, focusing on the hierarchical beamforming and power supply aspects. Some architectures are also discussed in Faulkner et al. (2010), Bij de Vaate et al. (2011) and Faulkner et al. (2011).

5.7.3.1 Digital hierarchical beamforming

Some caution needs to be applied to the comparison between the analogue tile and all-digital beamforming made in Section 5.4.3. This is because alternative intra-station architectures using hierarchical, or multi-stage, all-digital beamforming also have the potential to reduce cost. For the all-digital architecture in Figure 5.2, the digitised signal from every antenna element is transmitted to the station beamformer block and both the tile and station stages of beamforming are located in that block.

Table 5.9: Principal options for intra-station architectures.[a]

Sub-system	Physical location	Technology
Digitiser	receptor, tile or station	N/A
Tile BF (optional)	tile or station	RF or digital
Signal transport	receptor–tile and tile–station, or receptor–station	RF or digital
Station beamformer	station node, or node serving multiple stations	digital

[a] The options listed illustrate the range of signal transport and processing architectures. This does not imply that every architecture would meet all the SKA_1-low requirements, nor is every option listed.

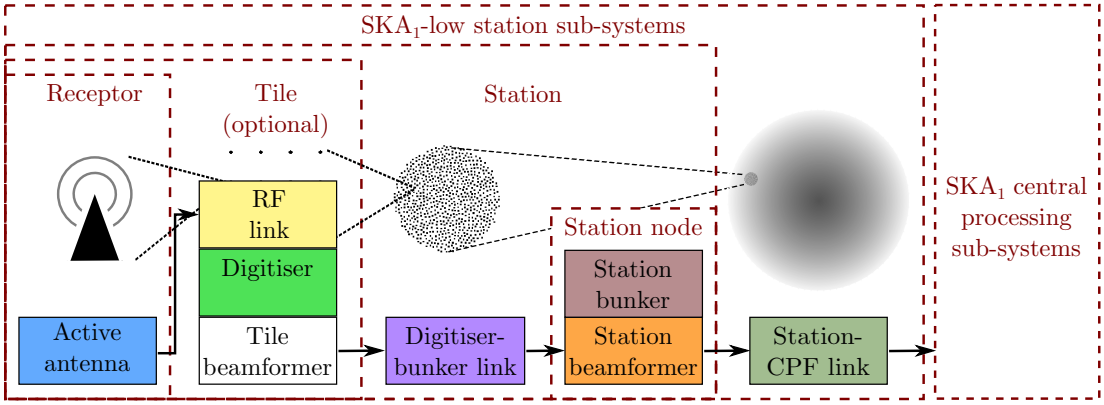


Figure 5.12: Schematic of the all-digital beamforming architecture, with a digital tile beamforming block located at the tile.

An alternative all-digital beamforming architecture is shown in Figure 5.12; this more closely represents the all-digital architecture in Faulkner et al. (2011). The architecture performs the first stage of digital beamforming at each tile to reduce the processed FoV earlier in the signal chain. Only $N_{\mathrm{b\text{-}tile}}$ digital tile beams are transmitted to the station beamformer, reducing the total data rate transmitted, hence digitiser–bunker link cost, by a factor of $N_{\mathrm{b\text{-}tile}}/N_{\mathrm{e/tile}}$. Hierarchical beamforming also reduces the beamformer processing load as described in Appendix F.2.6.2, although this applies regardless of how the hierarchical beamforming processing is physically distributed. The bottom-up SKADS cost estimate for the station beamformer processing already includes a discount from hierarchical beamforming.

However, these savings may be offset or even exceeded by the extra costs introduced by performing digital tile beamforming at distributed locations in the station signal path, rather than just at the station node. For example, the processing for distributed tile beamforming could require extra power distribution infrastructure. The total cost of controlled environment housings (including cooling and RFI shielding) for each tile beamformer would probably be more expensive than implementing both stages of beamforming within a larger controlled environment housing at each station node. Also, upgrading the digital system is easier if the processing is concentrated at a single location.

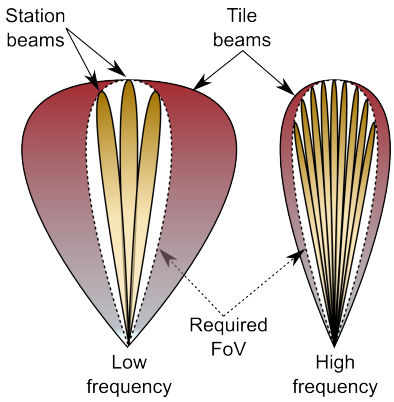


Figure 5.13: Schematic of the frequency-dependent relationship between the required FoV (dotted white) and the tile and station beams.

5.7.3.2 Hierarchical beamforming performance

Although hierarchical beamforming enables further design flexibility for the intra-station architecture, it also has the potential to decrease performance, regardless of whether the first-stage tile beamforming is analogue or digital. As mentioned in Section 5.4.3, tile beamforming reduces the accessible FoV and observational flexibility earlier in the signal path. If only a single tile beam is formed, it precludes the flexibility of pointing station beams at independent patches of the sky; the station beam pointing is restricted to within the single tile beam. Hierarchical beamforming also introduces errors in those station beams which are off-centre (i.e. not pointing in exactly the same direction as the tile beam), as discussed in Faulkner et al. (2010).

Tiles with more elements per tile further reduce the accessible FoV, because of the larger tile diameter (assuming the inter-element spacing is kept constant). If only one tile beam is formed, then the accessible FoV is defined by $\Omega_{\rm tile}$, the FoV of that tile beam. The required processed FoV, $\Omega_{\rm req}$, is limited by the accessible FoV, such that $\Omega_{\rm req} < \Omega_{\rm tile}$. As illustrated in Figure 5.13, $\Omega_{\rm tile}$ is frequency-dependent, while $\Omega_{\rm req}$ is not. Because $\Omega_{\rm tile}$ is defined to be the FoV out to the half-power point of the beam, equal sized FoVs ($\Omega_{\rm tile} = \Omega_{\rm req}$) would result in sensitivity loss away from the centre of the tile beam.

For example, consider a 4×4 element tile, with regularly (uniformly) spaced elements. From Kraus (1986), the half-power beamwidth $\theta_{\rm HP}$ at zenith for a linear array of n elements with inter-element spacing d is approximately

$$\theta_{\rm HP} = 0.88 \sin^{-1}\left(\frac{\lambda}{nd}\right); \tag{5.1}$$

away from zenith, $\theta_{\rm HP}$ is larger (Mailloux, 1995). For $d = 1.5\,\mathrm{m}$ at $450\,\mathrm{MHz}$, $\theta_{\rm HP} \approx 5.6\,\mathrm{deg}$ and $\Omega_{\rm tile} \approx 31\,\mathrm{deg}^2$, while $\Omega_{\rm req} = 20\,\mathrm{deg}^2$. Thus a single beam formed from a 16 element tile does provide sufficient FoV at the highest frequency, but there would be some sensitivity loss towards the edge of the tile beam. This loss reduces at lower frequencies, given $\Omega_{\rm tile} \propto 1/\nu^2$.

Alternatively, more tile beams could be formed. However, this would result in more signal paths to the station beamformer, and the cost reduction would be less. Because this problem similarly applies to digital tile beamforming, the full benefit of a reduction in data transmission and signal

processing cost is only realised if only one digital tile beam is formed. For example, the all-digital architecture in Faulkner et al. (2011) sends multiple tile beams from the 256 element tiles to the station beamformer, resulting in a total data rate at the station beamformer which is nearly as high as transmitting every digitised antenna element signal.

There is potential for both RF and digital tile beamforming approaches to be implemented in different phases of the telescope. If designed correctly, the analogue tile beamforming could be upgraded to all-digital beamforming at a later date, once the cost and power consumption of the digital components reduce so that such an upgrade is cost-effective.

5.7.3.3 Example alternative architectures

Alternative intra-station architectures may prove to be more cost-effective than those considered in this analysis, with implications for the cost trade-offs. Technical factors also require consideration, such as dissimilar power losses at different locations in the architecture, and the deployability of different intra-station architectures.

In the all-digital architecture (Figure 5.2), digitisation occurs relatively close to the antenna element and the digitised signal is transmitted via optic fibre. Powering the active antenna electronics, digitiser and digital optical transmission components presents an extra cost, because they would require either an extensive power distribution network which is appropriately sized to the peak power load, or the installation of self-powered (solar power and energy storage) solutions (e.g. Hall, 2009; Faulkner et al., 2011). Self-powered antenna elements would remove the need to distribute power to the electronics at every antenna element. However, to cater for the 180 m station diameter, such an architecture would require an increase to the current maximum transmission distance of 50 m for short-range, high-speed digital optical transmission (Faulkner et al., 2011).

In contrast, the RF tile beamforming architecture (Figure 5.3) can deliver power to the active antenna element electronics and RF tile beamformer from the station node via a copper-based RF link, as is done for LOFAR (Gunst, 2007). Using the RF link for both signal transport and power delivery presents a cost saving, because the dedicated power distribution network only extends as far as the station nodes, rather than to every antenna element.

An alternative all-digital beamforming architecture, if technically feasible, could put the digitisation at station node and use similar RF links to power every antenna element and transport the signal to the station node. The obvious additional cost is the extra RF links required, but a separate power distribution network is not required.

A different architecture could have only the analogue components at the antenna element, and use radio over fibre (RoF) technology to transmit the analogue signals from each antenna element (Juswardy et al., 2011). These analogue signals would be transmitted to a node containing the digitisers, channelisation and beamforming equipment, as shown in Figure 5.14. The node could serve one of more stations; for transmission distances of 200–500 m, such an architecture would not be viable for analogue transmission over copper-based cable (Perini, 2011). The fibre cables are also physically smaller than copper-based cable, which may be important if there are tens of thousands of cables entering one node. The life-cycle cost benefits of these self-powered

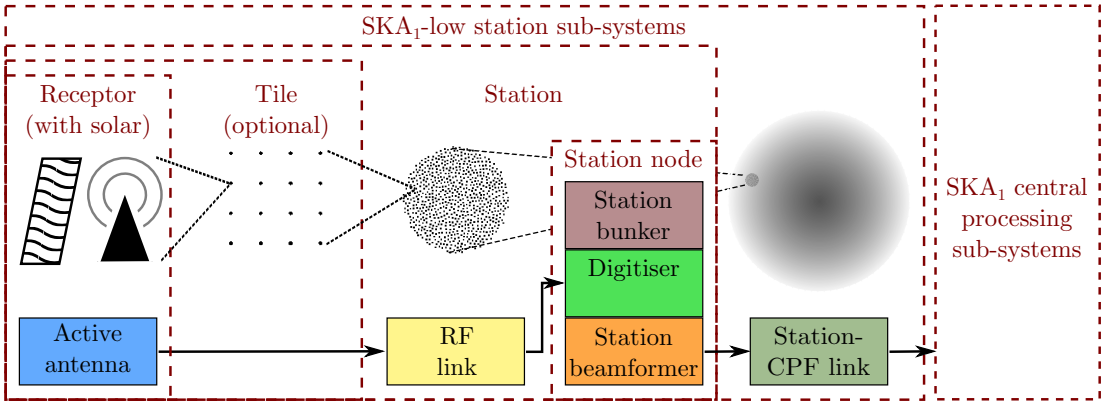


Figure 5.14: Schematic of a self-powered array, with the digital blocks located at the station node.

architectures with fibre links include simplified deployment from fewer connections (than the all-digital architecture with distributed power), potential for RFI reduction, increased resilience against lightning strikes and increased upgrade flexibility because all the received data arrives at the node (Faulkner et al., 2011).

5.8 Further work

In the absence of all the necessary cost information to make a complete cost estimate, this first-order analysis provides a simple comparison of the cost differences between representative single and dual-band implementations. Along with obtaining new cost information, there is scope for continued expert attention to refine the existing cost estimates and better understand the uncertainties. Cost refinement means improving the accuracy and precision of the cost, using reference class costing from other projects or studies, or conducting new bottom-up cost studies. Prime candidates for cost refinement are:

- active antenna elements, for both the single-band (6.5:1) and dual-band (2.5:1) elements
- antenna deployment and site preparation
- central processing sub-systems.

There will always be a level of uncertainty in parametric studies because the models use fairly simple scaling relationships for the trade-off analysis. This uncertainty can be mitigated by solidifying the SKA_1-low performance requirements, and the specific trade-offs required.

The use of the parametric model presented in this thesis need not be limited to comparing single and dual-band arrays. Section 5.7 illustrates just a few of the trade-offs which can be explored; other top-level parameters within the model which could be varied include:

- inter-element spacing
- antenna element gain
- the frequency split point of the dual-band array
- number of elements per station (to vary A/T while keeping the number of stations fixed).

The optimal number of stations and elements per station, for constant sensitivity, is an open question, but requirements other than cost must be considered. For example, the number and geographical placement of stations to achieve adequate (u, v) coverage and telescope calibration will also affect this optimisation. Such an optimisation requires a better understanding of the relative life-cycle costs between the SKA_1-low station digital sub-systems and the central processing sub-systems, in conjunction with refined performance requirements for station sensitivity and processed FoV (Appendix F.5.4). Re-configurable station sizes through logical stations, especially in the SKA core, may allow for a cost-optimised station architecture that better meets the scientific requirements; this approach requires further analysis of the design, performance and cost implications.

I mention in Section 5.1 the close link between the telescope performance, cost and risk, and the science requirements. The parametric model is sufficiently robust to consider differential cost trends between implementations, but is less suitable for determining absolute costs; calculating the monetary cost of changing a science requirement requires care.

5.9 Chapter summary

I have developed a scalable parametric model to compare the cost of implementing SKA_1-low as a single or dual-band aperture array, considering the cost impact on both the station hardware costs and the broader telescope system costs. Perhaps somewhat surprisingly, despite the dual-band array having twice the number of antenna elements, neither the representative single or dual-band implementations are cheaper *a priori*.

The cost difference between the single and dual-band implementations is, in essence, a comparison between the number and cost of the signal paths prior to station beamforming, and the downstream signal processing costs. The dual-band implementation has twice as many signal paths, but achieves a given field of view using fewer station beams, thereby reducing the downstream processing load. As a result, the dual-band implementation is more sensitive to changes in costs that scale with the number of signal paths, such as those of active antenna element hardware and deployment. Conversely, the single-band implementation is more sensitive to changes in the cost of signal processing sub-systems such as the station beamformer, correlator, and imaging and non-imaging processors. The cost difference between the single and dual-band implementations depends on the fractional cost of each of these groupings.

A particularly important SKA-low parameter requiring further scientific consideration is the product of the processed FoV and bandwidth—the beam-bandwidth product, discussed in Section 5.7.2. If the beam-bandwidth capacity of the processing can be significantly reduced, the costs which then dominate are those that scale with the number of antenna elements. This increases the cost of the dual-band implementation relative to the single-band implementation, but the difference is much less than the factor of two increase which might naively be expected from an array with twice the number of elements.

I find that implementing a first stage of RF (analogue) tile beamforming prior to the digital station beamformer enables a potentially significant reduction in station hardware costs and power demand, the effect of which can outweigh any difference between single and dual-band

implementations. Furthermore, if 90 m diameter stations are considered instead of 180 m, the correlator and imaging processor costs become cost drivers, dominating the station hardware costs.

To establish whether the single or dual-band array is the most cost-effective SKA_1-low implementation, improved cost information and further optimisation of the putative telescope designs are required. In terms of specific studies likely to assist in system design choices, the central processing, antenna deployment and site preparation costs are potentially significant cost drivers which have so far not had sufficient attention.

Chapter 6

The high angular resolution SKA

The third case study in this thesis considers cost-effective network design solutions for the SKA long baselines of the SKA_2 dish array in Africa. In this chapter, I develop a modelling framework that parametrises a realistic network architecture, and use this framework to investigate trades between scientific performance and design parameters that drive the cost of the network. Although I do not model a data network for potential long baselines for the SKA_2 mid-frequency aperture arrays, the same framework applies.

6.1 Introduction

A key technical challenge for the long baselines ($\gtrsim 180\,\mathrm{km}$) of the SKA_2 dish array is to design a data delivery network that can stream high quantities of data over large distances in a cost-effective manner (e.g. McCool, 2010). To date, the effect of the long baseline network architecture on scientific performance of the SKA has not been investigated in detail. This chapter relates the design and cost of such a network to telescope scientific performance.

Scientific goals for high angular resolution observations with the SKA dish array exploit the increased sensitivity and improved (u, v) coverage compared to existing VLBI arrays (Godfrey et al., 2011; SSWG, 2009). These science goals include:

- testing theories of gravity by measuring precise, model-independent distances to relativistic pulsar binary systems (Smits et al., 2011b)
- imaging proto-planetary disks, to help understand the process of planet formation
- resolving active galactic nuclei (AGN) and star formation in Galaxies, to track and distinguish between the contribution of star formation and AGN to the evolution of galaxies.

The long baselines will be a useful tool for verification of fast transients events, as discussed in Chapter 4. Furthermore, the long baselines will enhance lower angular resolution observations with the SKA, by improving imaging capability (Lobanov, 2012), and resolving the structure of faint extragalactic sources detected by the lower angular resolution observations (Garrett, 1999).

However, the cross-connection between the scientific arguments for high angular resolution observations and the network design solutions is low. A key development to enable this cross-connection was Godfrey et al. (2011) framing the requirements in terms of bandwidth and processed FoV; these translate more easily into data network requirements. Another requirement that influences the data network design is (u, v) coverage, but (u, v) coverage is not well-specified in DRM_2 (SSWG, 2009) nor Godfrey et al. (2011).

Rather than analysing the science and deriving a design from these requirements, this chapter aims to facilitate future network design optimisation and trades. I model a number of design

solutions, and use three simple metrics as a measure of top-level scientific performance: maximum baseline length; number of remote stations; and the beam–bandwidth product of the remote station. I develop a parametric modelling framework and use the three metrics to illustrate the representative cost and performance of a likely network architecture for the SKA long baselines. Within this framework, I identify network design solutions that reduce cost without significantly compromising access to high angular resolution parameter space.

This chapter is useful for SKA planners and scientists, and network providers. The analysis is not intended as a fully costed network design. Instead, this work aims to:

- provide a framework to link telescope performance and network cost, and contribute to a design down-selection process by identifying potential performance–cost trades
- identify the principal cost drivers for the long baseline network, with a view to stimulating further requirements specification for high angular resolution science
- inform network designers (in both industry and the SKA Organisation) how potential changes to scientific performance impacts the network design
- outline a skeleton architecture for the SKA project, against which the cost of self-build and commercially provided solutions can be evaluated.

Section 6.2 describes the scope of the analysis in this chapter. Section 6.3 presents an exemplar network architecture, and the modelling framework to parametrise this architecture in the context of the SKA scientific requirements. In contrast with the pair-wise comparisons in Chapter 5, Section 6.4 makes multi-dimensional performance and cost trades for a range of input parameters. Section 6.5 discusses the implications of these trades on design, cost and scientific performance.

6.2 Scope of analysis

To explore the performance and cost trends for long baselines, I use an exemplar network architecture. Section 6.2.1 describes this architecture and its applicability to the SKA long baselines. Section 6.2.2 briefly outlines other factors that may be relevant to the system performance and cost trade-offs of the long baselines.

6.2.1 Exemplar network architecture for SKA long baselines

Figure 6.1 shows a conceptual diagram of the exemplar network architecture modelled in this chapter, split into three cost categories described later in this section. The network is placed in the context of a 'representative implementation' of the SKA long baselines, based on the SKA_2 high-level system description (Dewdney et al., 2010b) and more recent configuration studies (Bolton et al., 2011; Millenaar & Bolton, 2011). The design specifies 600 single-pixel feed dishes, each 15 m in diameter, to provide the collecting area at long baselines. Groups of 24 dishes form a 'remote station'; these 25 remote stations are located at exponentially increasing distances from the array centre (core), from 180 km to 3000 km. To reduce the data rate, the station dishes are beamformed, in a manner similar to the antenna elements in aperture array stations. The output station beam data from each remote station is subsequently transmitted back to the central processing facility via a fibre-optic network.

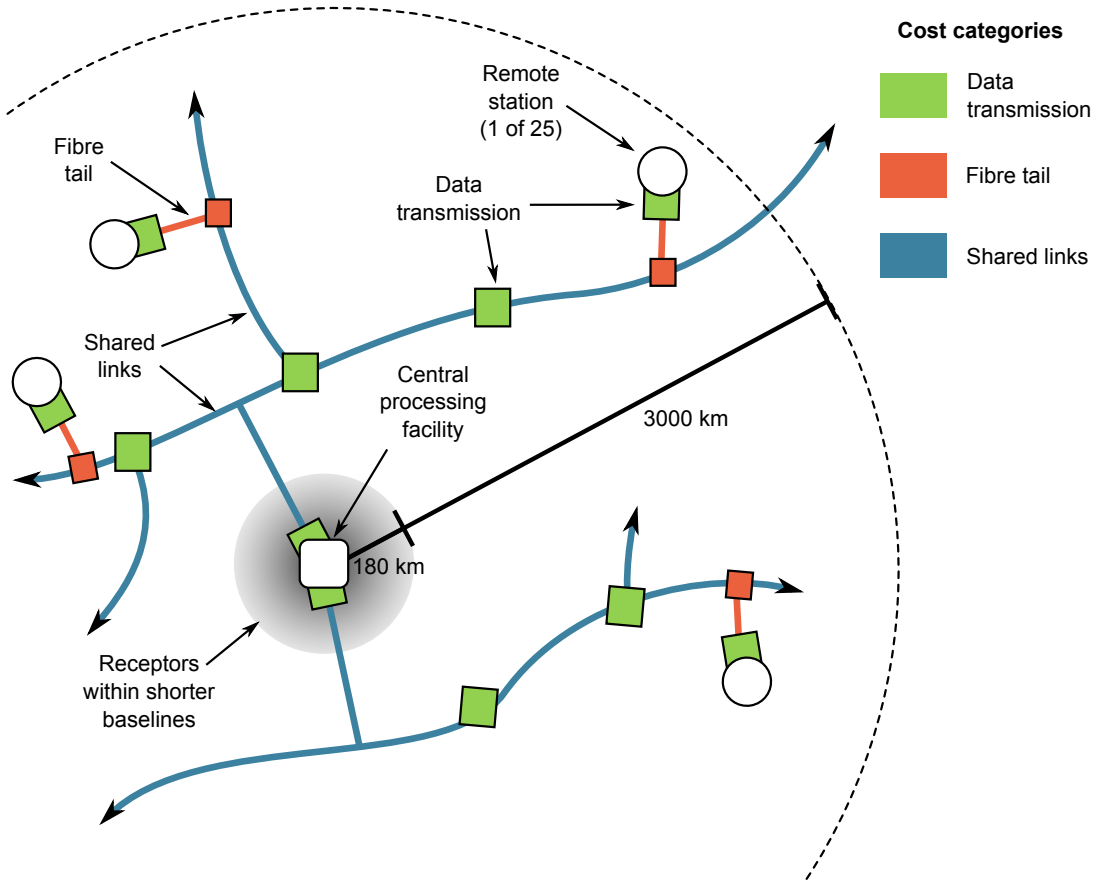


Figure 6.1: Conceptual diagram of the exemplar network architecture for the SKA long baselines (not to scale), shown in the context of a representative implementation of SKA long baselines. The network architecture is highlighted by its cost categories.

The network can be viewed as a number of point-to-point links between each remote station and the central processing facility, where data is transferred using dense wavelength division multiplexing (DWDM) technology. Used in the telecommunications industry to transfer internet and other data over intra- and intercontinental distances, DWDM technology provides the flexibility to add in links to remote stations at various locations along a fibre route, all on a single fibre pair (assuming available capacity).

The cost of transmitting such data via fibre-optic network has been the subject of previous investigations (e.g. Bolton et al., 2009c; Chippendale et al., 2007; Hall et al., 2002; McCool, 2010), and have been costed in the recent SKA site submissions (ANZSCC, 2011; SKA SA, 2011). Although these investigations and submissions propose similar network architectures based on DWDM technology, the implementation and cost of the network architectures in the above investigations and submissions vary considerably. McCool (2010) outlines some implementation options for the long baseline fibre-optic network:

- A self-build solution, where the network is constructed for exclusive SKA use.
- A dark fibre solution, where existing fibre infrastructure is leased, and the SKA supplies the networking equipment.

- A commercial bandwidth solution, where the data transport from each station is provided by a commercial provider with an existing network.

The previous investigations are primarily based on the self-build or dark fibre solutions, with remote stations placed along spiral arms out to 3000 km. In contrast, both the Australian and South African site submissions presented network architectures that maximise the use of existing telecommunication carrier networks to provide data connectivity; these architectures are a hybrid between the dark fibre and commercial bandwidth solutions described above.

Regardless of the implementation, the performance–cost trades require the network cost estimates to be scalable over a range of input parameters. The complication is that the cost of the self-build and dark fibre solutions are specific to a particular design solution (i.e. fixed input parameters). This is also true of the commercial bandwidth solution, because a simple price per unit bandwidth (€/Gbps) is unlikely to be an accurate representation of the cost of the long baseline network. For example, fibre infrastructure costs (including the fibre cable, trenching and deployment) generally vary as a function of distance and the number of individual links. Meanwhile, the DWDM technology and the network design impose discrete steps in data transmission capacity and distance, which introduce breakpoints in the cost curve.

To account for these different implementation approaches, and enable high-level performance and cost modelling, I generalise the network costs by defining three cost categories, highlighted in Figure 6.1: data transmission; 'fibre tails'; and shared links. The data transmission represents the hardware required to transmit the astronomical data and monitor and control data from the remote station to the central processing facility; this includes the intermediate signal regeneration boxes in Figure 6.1. A fibre tail is a relatively short ($<$80 km), SKA-exclusive connection, required to carry the data from the remote station and link it to the rest of the network. The data transmission and fibre tail categories are relatively simple to cost, and the costs are similar between the implementations mentioned above.

The shared links encompass most of the fibre infrastructure that carries the data from the remote station to the central processing facility; they can potentially carry data from multiple remote stations and non-SKA users. The fixed costs of building the fibre infrastructure are high, particularly the cost of trenching. For the €10 k per km trenching cost estimated in McCool (2010), this equates to €30 million for a straight-line trench out to the most distant remote station at 3000 km. More than one such trench would be required, to ensure adequate snapshot (u, v) coverage. Other costs include the capital cost of the fibre and the 'huts' every 80–100 km to house the amplification equipment; power provision for these huts are a potential operational cost. There are also costs due to regulatory overheads and maintenance of the fibre and huts (McCool, 2010). Once built and carrying data, the incremental or marginal cost for carrying additional data is relatively small. Most of the incremental cost is due to the amplification equipment when additional fibres are used.

The cost to the users for these shared links thus depends on the number of users, and the amount of fibre required. A self-build network represents the maximal cost, as all these costs are borne exclusively by the SKA. If a carrier network makes the pre-existing fibre infrastructure available to the SKA for no cost, only the incremental costs of additional fibre and amplification

components apply for the shared links. The Australian site submission resembles this minimal-cost scenario, where the national research and education network AARNet was to provide the pre-existing fibre infrastructure (ANZSCC, 2011).

The default cost of the shared links modelled in this chapter are based on this minimal cost scenario. However, a cost multiplier is included as a design input, to account for a different price structure imposed by the carrier network, which may be significantly higher than the incremental costs. Reasons for the higher costs include:

- carrier network structure and overheads (e.g. government, non-profit, commercial)
- apportioning all fibre infrastructure costs to the users
- carrier network sections purpose-built for SKA use.

The cost multiplier allows for a sensitivity analysis of increased costs; the modelling in this chapter tests the sensitivity of the network cost to an order of magnitude increase in shared link costs.

Ultimately, the cost of these shared links will depend on the self-build options and carrier networks available to the SKA at the time of SKA_2 construction. But this approach, which includes the cost multiplier, allows for the first-order parametric modelling of performance and cost trends.

6.2.2 Excluded costs

Prior to detailing the parametric modelling, I outline some factors that are not modelled, but may be relevant to the system performance and cost trade-offs of the long baselines.

Fibre tail length: The cost of connecting a remote station will be higher if there is not a carrier network (shared link) available within ∼80 km, because the remote station would require additional infrastructure and networking components. Unless there are other clients (public or private) to share the expense, this cost would be borne by the SKA. Such scenarios require case-by-case evaluation.

Remote station cost: To a first-order approximation, the hardware cost for the 600 dishes remains unchanged, irrespective of their distribution amongst the remote stations. However, the deployment, infrastructure and maintenance costs may be higher for an increased number of remote stations. These costs, and the trade-offs involved, are a function of the SKA infrastructure model, to be developed in the context of SKA funding and policy agreements.

Central processing facility costs: Imaging the dish FoV or even the station beam FoV with SKA long baselines can place a high demand on processing costs (Alexander et al., 2009), and the post-processing sub-system has previously been raised as a significant cost-driver for SKA long baselines (Faulkner et al., 2010). However, Godfrey et al. (2011) conclude that wide FoV imaging is not a requirement for SKA long baselines; therefore I assume in the present analysis that high angular resolution observations are not a cost-driver for the central processing facility.

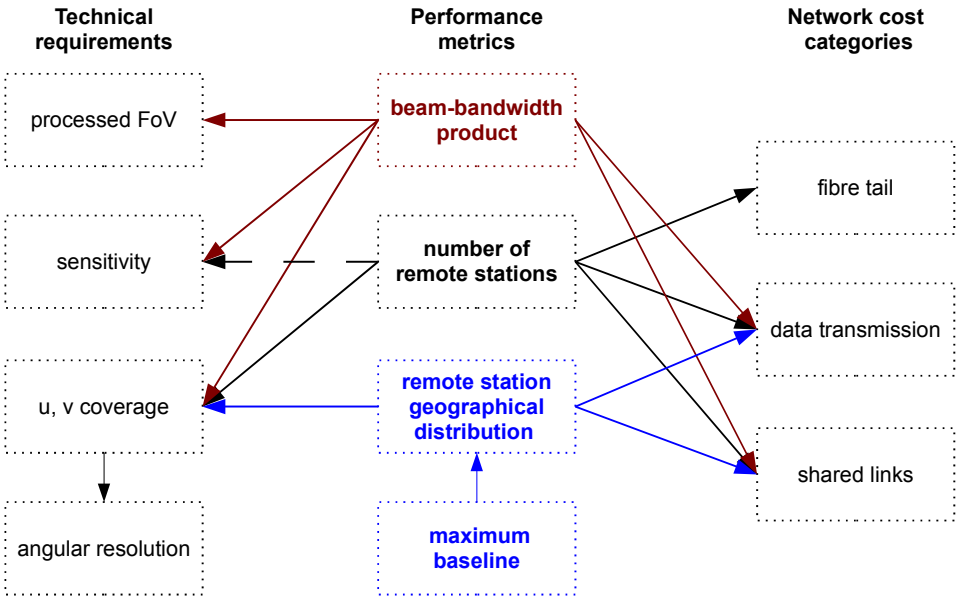


Figure 6.2: Relationship between technical requirements, performance metrics and costs for the network design.

6.3 Parametric performance and cost models

The network architecture cost is driven primarily by the number of links to remote stations, and the length and data rate of each transmission link. The parametric models express these costs as a function of three interconnected top-level performance metrics: maximum baseline length $b_{\max}$; number of remote stations N_{RSt}; and the beam–bandwidth product of the remote station. Figure 6.2 shows these metrics and their relationship to the cost categories described in the previous section. These metrics also describe the impact of the network architecture on the scientific performance of high angular resolution observations with the SKA, as shown by the technical requirements column in Figure 6.2.

In this section, I summarise the technical requirements for high angular resolution observations and place them in the context of these three performance metrics (Section 6.3.1) and describe the modelling framework that links the performance to the network design and cost (Section 6.3.2).

6.3.1 Performance metrics that describe the requirements

The technical requirements specify the telescope performance. The following is a list of generally applicable requirements for high angular resolution observations, summarised from SSWG (2009) and Godfrey et al. (2011):

Frequency range: Most experiments require mid frequencies ($0.5 - 10\,\mathrm{GHz}$), or, alternatively, the experiments can be recast to this frequency range.

Maximum baseline length: Maximum baseline length is a frequency-independent proxy for angular resolution. Lengths range from $>1000\,\mathrm{km}$ to $>5000\,\mathrm{km}$.

Table 6.1: Summary of variable input parameters modelled.

Input parameter	HLSD value[a]	Range of values	Comment
Beam–bandwidth product	9 GHz	1.5–16 GHz	Determines required data rate per station R_{RSt}.
Maximum baseline length $b_{\max}$	3000 km	500–3000 km	Distance from core.
Number of remote stations N_{RSt}	25	10–25	Placed at exponentially increasing distances between 180 km and $b_{\max}$.

[a] SKA_2 high-level system description (Dewdney et al., 2010b).

Processed field of view: Wide FoV imaging is not a requirement for SKA long baselines. Many experiments only require a single target to be observed, using a relatively narrow processed field of view compared with the dish primary beam. However, multiple station beams formed within the dish primary beam improves calibration via phase-referencing (Godfrey et al., 2011).

Bandwidth: Although DRM_2 requires an instantaneous fractional bandwidth of ∼1, a bandwidth of order 500 MHz is sufficient for many of the experiments outlined in Godfrey et al. (2011). Some science goals, such as *resolving AGN and star formation in galaxies*, require larger instantaneous bandwidths, especially at higher frequencies, to achieve a factor of 6–7 coverage in frequency range (SSWG, 2009).

Sensitivity: Because the total number of dishes in remote stations does not change, the sensitivity of the long baselines is largely independent of the network design. Sensitivity will be affected by the bandwidth of the observation, but many of the experiments in Godfrey et al. (2011) do not require a large bandwidth to achieve maximum sensitivity.

(u, v) coverage: Some experiments, such as searches for binary supermassive black holes, require "good snapshot (u, v) coverage" (Godfrey et al., 2011). However, this requirement is not quantified.

These requirements describe the target performance to be achieved with SKA_2. From the telescope design perspective, this target performance is achieved by varying the design parameters. Table 6.1 lists the performance metrics described earlier as variable input parameters to the parametric model. Although these parameters do not encompass all the technical requirements for SKA long baselines, they represent the performance that can be influenced most directly by the network architecture. A design solution for the exemplar network architecture is formed using a selected value for each input parameter in Table 6.1, along with other relevant design parameters, summarised in Table 6.2.

The following sub-sections detail the scientific implications of varying the parameters in Table 6.1.

6.3.1.1 Beam-bandwidth product

The beam–bandwidth product was introduced in Chapter 5 as a measure of signal processing capacity. In the long baseline analysis, it is a measure of the total data rate out of a station.

Table 6.2: Summary of SKA_2 design parameters relevant to the modelling

Input parameter	Values	Comment and relevant section
Data rate R_{RSt}	$24N_{\mathrm{b\,-\,RSt}}\Delta\nu$ [a]	For $N_{\mathrm{b\,-\,RSt}}$ remote station beams and $\Delta\nu$ bandwidth (Section 6.3.1.1).
Station diameter	$\sim$350 m [b]	
Total number of dishes in remote stations	600 [c]	20 % of SKA_2.
Number of dishes per station $N_{\mathrm{dish/RSt}}$	$600/N_{\mathrm{RSt}}$	Section 6.3.1.3.

[a] McCool (2011b).
[b] Millenaar et al. (2011).
[c] Dewdney et al. (2010b).

As with SKA_1-low, I assume that coarse channelisation occurs prior to station beamforming. Therefore, the total output data rate per remote station, in bits per second (bps) is

$$R_{\mathrm{RSt}} = 24N_{\mathrm{b\,-\,RSt}}\Delta\nu, \tag{6.1}$$

where $N_{\mathrm{b\,-\,RSt}}$ is the number of remote station beams formed and $\Delta\nu$ is the processed bandwidth. The factor of 24 accounts for two polarisations and multipliers for the digital signal (see Appendix G.1).

The beam–bandwidth product from each remote station, $N_{\mathrm{b\,-\,RSt}}\Delta\nu$, describes the performance of long baselines for a variety of observing modes. Options that vary $N_{\mathrm{b\,-\,RSt}}$ include:

- improving phase calibration with multiple station beams , where each additional station beam observes a calibrator source; three such beams are suggested for the SKA (Godfrey et al., 2011)
- using multiple station beams to observe more than one target within the dish FoV
- dividing the station into independently pointed subarrays, in which case $N_{\mathrm{b\,-\,RSt}}$ is the product of the number of subarrays and the number of beams formed within each subarray
- pointing each dish in an independent direction, in which case $N_{\mathrm{b\,-\,RSt}}$ equals the number of remote stations.

The definition of processed bandwidth $\Delta\nu$ depends on the astronomical application. The simplest definition is that $\Delta\nu = \nu_{\max} - \nu_{\min}$, where $\nu_{\max}$ and $\nu_{\min}$ are the highest and lowest frequencies of observation; a large $\Delta\nu$ maximises sensitivity for continuum observations. For example, $\Delta\nu = 3$ GHz for a 1–4 GHz feed. This processed bandwidth is limited by the feed and receiver, therefore $\Delta\nu$ is smaller at lower frequencies. For channelised data, $\Delta\nu = N_{\mathrm{ch}}\Delta\nu_{\mathrm{ch}}$, where N_{ch} is the number of channels and $\Delta\nu_{\mathrm{ch}}$ is the channel bandwidth. Appropriately designed station channelisation and beamforming hardware could allow the selection of specific frequency channels, reducing the data rate transmitted to the central processing facility. Observing modes for which this may be useful include multi-frequency synthesis (Conway et al., 1990) and spectral observations.

To illustrate that not all observations require the full beam–bandwidth product, Table 6.3 shows example observational scenarios for different beam–bandwidth products. Actual beam–

Table 6.3: Example scenarios for different beam–bandwidth products.

Beam–bandwidth product	R_{RSt} (Gbps)	$\Delta\nu$ (GHz)	$N_{\mathrm{b\,-\,RSt}}$	Scenario
Maximal (36 GHz)	864	9	4	Available at the highest frequency of observation. Assumes sufficient calibrators are available for the three extra beams used for in-beam calibration.
High (9 GHz)	216	9	1	'Nodding' calibration only requires a single beam.
		3	3	Maximum bandwidth of the 1–4 GHz feed, 2 beams for calibration.
		1	18	Multiple targets.
Intermediate (4 GHz)	96	4	1	25 % fractional bandwidth for 12 GHz maximum frequency ($\Delta\nu/\nu_{\mathrm{max}}$).
		1	4	Multibeam calibration, reduced bandwidth.
Low (1.5 GHz)	36	1.5	1	'Nodding' calibration, reduced bandwidth.
		0.375	4	Multibeam calibration, significantly reduced bandwidth.

bandwidth requirements could be developed through future analysis of the science goals. Note that Table 6.3 is purely illustrative, as there are some inconsistencies in the maximum bandwidth specified for SKA_2. In the SKA_2 high-level system description (Dewdney et al., 2010b) and the signal transport and networks high-level description (McCool, 2011b), $N_{\mathrm{b\,-\,RSt}} = 1$ and $\Delta\nu = 9\,\mathrm{GHz}$, hence $R_{\mathrm{RSt}} = 216\,\mathrm{Gbps}$. However, 9 GHz bandwidth was specified for a 3–12 GHz feed for the SKA_2 single-pixel feed dishes. More recently, the SKA_1 high-level system description (Dewdney et al., 2011a) specified an upper frequency of 10 GHz for the dishes. In which case, the maximum bandwidth is likely to be less than 9 GHz, unless a wideband single-pixel feed with 10:1 fractional bandwidth is used. I use $\Delta\nu = 9\,\mathrm{GHz}$ to maintain consistency with the site submission documents, which costed the long baseline networks for $R_{\mathrm{RSt}} = 216\,\mathrm{Gbps}$.

6.3.1.2 Maximum baseline and remote station geographical distribution

When combined with a distribution pattern, the maximum baseline length b_{max} determines the geographical distribution of the stations. In this analysis, I use a distribution pattern representative of the SKA site submissions, where the remote stations are distributed with exponentially increasing distance from the core (Millenaar et al., 2011). Figure 6.3 shows this distribution for the 'default' scenario of $N_{\mathrm{RSt}} = 25$ stations. The exponential distribution begins beyond 180 km from the core, and continues out to a distance $b_{\mathrm{max}} = 3000\,\mathrm{km}$, as detailed in Appendix G.1. Needless to say, the stations will not be distributed along a line as in Figure 6.3, but distributed across southern Africa.

The geographical distribution of the stations in relation to the SKA core largely determines the (u, v) coverage, and in turn, the angular resolution at a given frequency. This relationship means that b_{max} serves as a proxy for angular resolution, as indicated in Figure 6.2. An

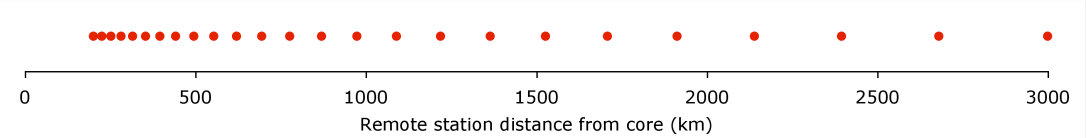


Figure 6.3: Baseline distribution of 25 remote stations. Each point marks the exponentially increasing distance between a remote station and the core, out to a distance $b_{\max} = 3000\,\mathrm{km}$.

actual geographical distribution gives a more accurate measure of angular resolution via the synthesised beam (the Fourier transform of the distribution of receptors in the (u, v) plane). However, combined with the exponential distribution pattern, the maximum baseline length $b_{\max}$ provides a sufficient approximation of angular resolution for the first-order trades in this analysis.

6.3.1.3 Number of remote stations

The number of remote stations affects sampling of the (u, v) plane at high angular resolution. Additional remote stations placed in the exponential distribution between 180 km and 3000 km increases the number of baselines, thereby improving the (u, v) coverage.

In this analysis, the total number of dishes placed at remote stations is held constant at 600; this approach ensures that point source image sensitivity remains unchanged. As the number of remote stations N_{RSt} increases, the number of dishes per station $N_{\mathrm{dish/RSt}}$ decreases: $N_{\mathrm{dish/RSt}} \approx 600/N_{\mathrm{RSt}}$.

The number of dishes per remote station can be varied without influencing the network costs. The data rate out of the remote station is a function of the number of remote station beams formed, and a single station beam is generally required for an observation, with up to three additional beams for calibration (Section 6.3.1). This is unlike the current architecture for SKA-low station processing hardware, which forms sufficient station beams to synthesise a specified processed FoV; the number of SKA-low beams depends on the station diameter.

6.3.2 Modelling framework

The modelling framework simplifies the exemplar network architecture as scalable blocks, representing the cost categories outlined earlier. The metrics described in Section 6.3.1 are the principal scalable parameters to drive the design; these parameters allow multi-dimensional performance trades to be easily investigated. I use SKACost (Section 3.5) to model the network architecture and carry out the trade studies.

The models are based on the exemplar network architecture described in Section 6.2.1. The modelling is independent of a specific configuration, because the remote stations are individually placed, following the exponentially increasing distance to each remote station described earlier. Each remote station data link can be costed separately, because the remote stations are dependent on carrier networks rather than purpose-built infrastructure. This approach is also different to previous analyses using spiral arms (e.g. McCool, 2010), where the costs are dominated by trenching and fibre costs.

The cost of each link is determined by creating scalable blocks based on the three cost categories outlined in Section 6.2.1, and assigning the major cost components in the DWDM network to one of these categories. This approach ensures that the model is representative of the costs incurred by an actual implementation. Prior to describing the blocks themselves, I summarise here the principal cost components for the exemplar DWDM network architecture; texts such as Kartalopoulos (2000) and Simmons (2008) describe DWDM networks in detail.

Figure 6.4 shows the major cost components for a single point-to-point link from a remote station to the central processing facility. The data transmitted from the remote station is carried by one or more DWDM channels, or wavelengths; many channels can be multiplexed onto a single optic fibre. A purpose-built fibre tail connects the remote station to the backbone network, which then carries the data to the central processing facility.

The principal components in Figure 6.4 are:

Transponder and multiplexer: A transponder transmits or receives the data at a particular wavelength; these wavelengths are multiplexed onto a single fibre. Present maximum transponder data rates are 10 Gbps or 40 Gbps, with 100 Gbps beginning to enter the market (Hansen, 2012).

Optical add-drop multiplexer (OADM): An OADM (with accompanying amplifiers) adds one or more channels from the fibre tail onto the backbone link, or drops one or more channels from the existing fibre onto the fibre tail.

Optical amplifier: Optical amplification is required beyond a distance of 80–100 km. An important characteristic of optical amplifiers is that they amplify all operational wavelengths, hence only a single optical amplifier per fibre is required .

Optical-electrical-optical (OEO) regeneration: Over longer distances, OEO signal regeneration is required, where the optical signal is converted to electrical and retransmitted as optical. OEO regeneration is expensive because transmission and reception components (transponders) are required for every channel. The transmission distance, or optical reach, of an un-regenerated signal depends on the individual network and may be up to several thousand kilometres for backbone connections. However, for regional networks, this distance is less: several hundred to a thousand kilometres (Simmons, 2008). Based on the location of population centres in southern Africa, I have estimated 480 km as a more likely optical section length used by the carrier networks.

Fibre pair Normally a number of fibres are bundled and laid together along a route. Deployment of the fibre includes costs such as trenching.

Hut The networking equipment is located in a telecommunications shelter which provides the infrastructure for the components, including power supply and active and/or passive cooling.

Table 6.4 summarises the principal cost components required for the self-build and carrier network scenarios outlined in Section 6.2. The active components require a power source, which is an operational cost to the system. The difference between the two scenarios are that the components in the carrier network scenario marked as shared leverage existing network

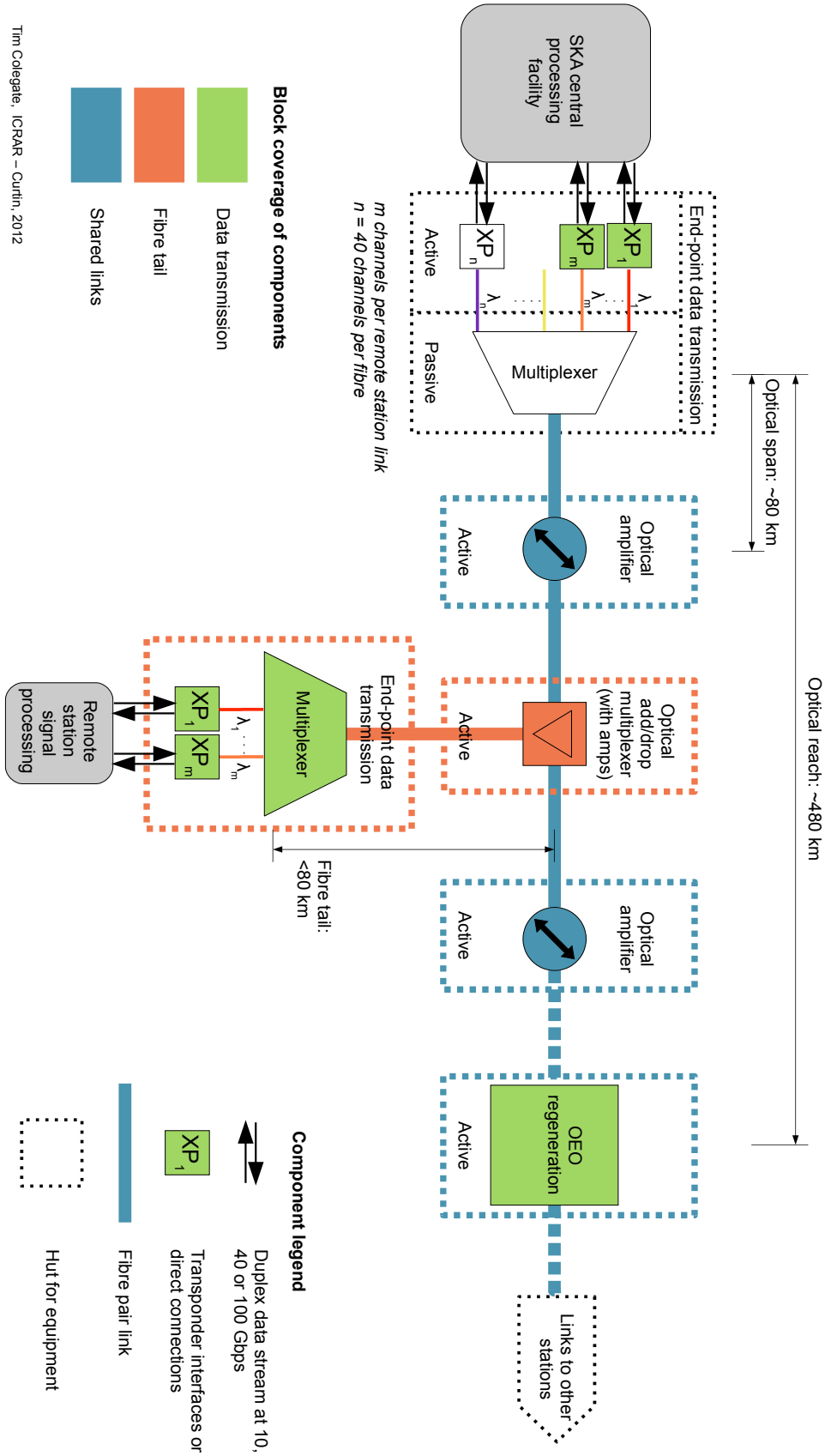


Figure 6.4: Block diagram of the major cost components for a point-to-point DWDM network architecture for the SKA, illustrating a single fibre tail to a remote station. Bi-directional components on one fibre pair are shown, colour-coded to match the block models (Section 6.3.2).

Table 6.4: Major cost components required for the network. Both scenarios require fibre tails.

Cost component	Active component	Unit	Scenario		Fibre tail[a]
			Self-build	Carrier network	
End-point data transmission					
↪transponder	Y	per channel	●	●	○
↪multiplexer[b]		per fibre	●	●	○
OEO regeneration					
↪transponder	Y	per channel	●	●	○
Optical amplifier[c]	Y	per fibre	●	◑	○
Optical add-drop multiplexer (OADM)[b]	Y	per fibre	○	○	●
Fibre			●	◑	●
Fibre deployment (including trenching)			●	◑	●
Hut			●	◑	◑

● Full cost ◑ Shared cost ○ Not applicable

[a] < 80 km length.
[b] Includes optical amplifiers.
[c] Includes dispersion compensation.

equipment and infrastructure, as described in Section 6.2.1. The models are based on those components marked in the carrier network column, plus the fibre tail components.

For each link, the components in Table 6.4 are placed into one of the three cost categories to form the scalable blocks (colour-coded in Figures 6.1 and 6.4). They are summarised as follows, and detailed in Appendix G:

Data transmission: Cost of transmitting astronomical data and monitor and control data. This block encompasses the full cost of transmission components (transponder and multiplexer) for each remote station, and additional components at OEO regeneration sites.

Fibre tail: Cost of connecting a remote station to the network. It is given as an average cost, and includes fibre and trenching cost estimates and the incremental components required to house and integrate the fibre-optical tails onto the carrier's network.

Shared links: Cost of carrying the data channels on the network. It approximates the *per channel* cost to the carrier network for all network traffic on a fibre cable.

The unit costs of the blocks are summarised in Table G.3; they are based capital costs. The cost coefficients for the parametric models are developed from a combination of reference class and bottom-up estimates, primarily using cost data from an industry study (Huelsermann et al., 2008) and the Australian SKA site submission (ANZSCC, 2011).

The network will also have operational costs, including maintenance costs for the fibre and components, and power costs for the active components in Table 6.4. Operations costs for optical networks are difficult to estimate as they are specific to the network architecture (Simmons,

2008). All three blocks contain components and infrastructure that will require maintenance. Therefore, a zeroth-order approximation is to assume that the operations cost is proportional to the capital costs. The operations cost for the long baseline network in the Australian site submission is of order 100–200 % of the capital cost when aggregated over 10 years of operation. A similar cost factor is likely to be applicable to the costs presented in this analysis.

Aside from the potential system–level cost drivers not included in this analysis (Section 6.2.2), there are some significant uncertainties which apply to both the capital and operational costs of the network. In particular:

- As commodity items, the components can be purchased relatively close to the construction date. With that date being 2018 or beyond (Table 1.1) the expected technical innovation of optic components introduces considerable uncertainty in the component price and availability, especially for the data transmission costs.
- The cost of the shared links will depend on the cost structure of the network carrier and the commercial terms of access, as mentioned in Section 6.2.1.
- The fibre tail costs for any particular remote station are likely to be affected by local geographical and political factors.

To account for uncertainties in the price of optical components, Appendix G develops low and high cost estimates for these components. For the shared links, I include a cost multiplier to account for different cost structures. Instead of attempting to anticipate variations to the fibre tail costs, I simply use an average cost.

While a final costing of the SKA long baselines is not possible at this point, the modelling framework itself is an important tool which can be refined. The first-order approximation of network costs, presented in the next section, enables trade-off analyses to guide future scientific studies and inform decision-making on scientific priorities and network design.

6.4 Results

I use the modelling framework to identify cost-driving design and performance parameters, and illustrate some of the scientific trades that are influenced by the network architecture. The basic steps in this section are:

- For a given beam–bandwidth product, determine the optimal transponder data rate (Section 6.4.1).
- Investigate the cost-driving effects of the three performance metrics (Section 6.4.2).
- Plot illustrative scenarios for a reduced-cost network implementation (Section 6.4.2).
- Repeat the process, to test the sensitivity of the results to an order of magnitude increase in the shared link costs (Section 6.4.3).

The scientific implications of the performance trades are discussed in Section 6.5.

6.4.1 Beam–bandwidth and transponder data rates

To eliminate transponder cost as a variable parameter in the later trades, I select the optimal transponder data rate for a particular beam–bandwidth. Figure 6.5 plots the network cost

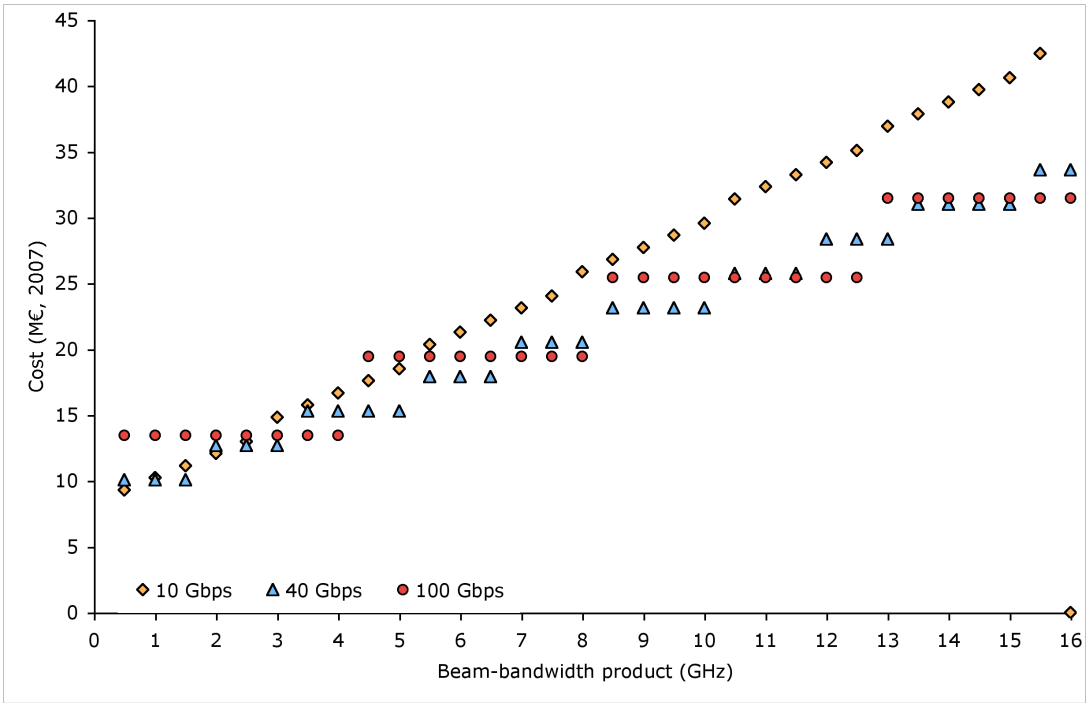


Figure 6.5: Network cost as function of beam–bandwidth product, at intervals of 0.5 GHz. Transponder speeds of 10, 40 and 100 G are shown; $N_{\mathrm{RSt}} = 25$ and $b_{\max} = 3000$ km. The zero-cost value for a beam–bandwidth of 16 GHz signifies that the maximum 40 channels per fibre has been exceeded.

as a function of beam–bandwidth, for astronomical data transmitted with 10, 40 and 100 G transponders (the monitor and control data always uses 10 G transponders, see Appendix G.2). The noticeable steps in cost for the 40 and 100 G transponders are due to the discrete transponder speeds.

The default beam–bandwidth of 9 GHz is a good example of the effect of these discrete transponder speeds on network utilisation and cost. In Figure 6.5, only the 10 G transponder is optimally utilised at every beam–bandwidth interval. For the 40 G transponder, up to ∼10 GHz beam–bandwidth is achievable for the same cost as 9 GHz, and for the 100 G transponder, up to ∼12.5 GHz is achievable. Alternatively, using a 100 G transponder for the 8 GHz beam–bandwidth results in a cost reduction of €3.7 M.

In the subsequent performance trades, I use the beam–bandwidth products and optimal transponder speeds shown in Table 6.5. For beam–bandwidth products less than 9 GHz, there is unlikely to be an effect on the broader SKA system design beyond that already considered in the high-level system description. I do not consider beam–bandwidth products higher than 9 GHz, because of the potential to increase design requirements and cost for other parts of the system not modelled here, such as the imaging processing.

Table 6.5: Values used for performance trades.

Beam–bandwidth product (GHz)	1.5	4	8
Transponder speed (Gbps)	40	100	100
Required number of duplex transmission channels per remote station[a]	2	2	3

[a] Includes a duplex channel for monitor and control.

6.4.2 Cost as a function of performance metrics

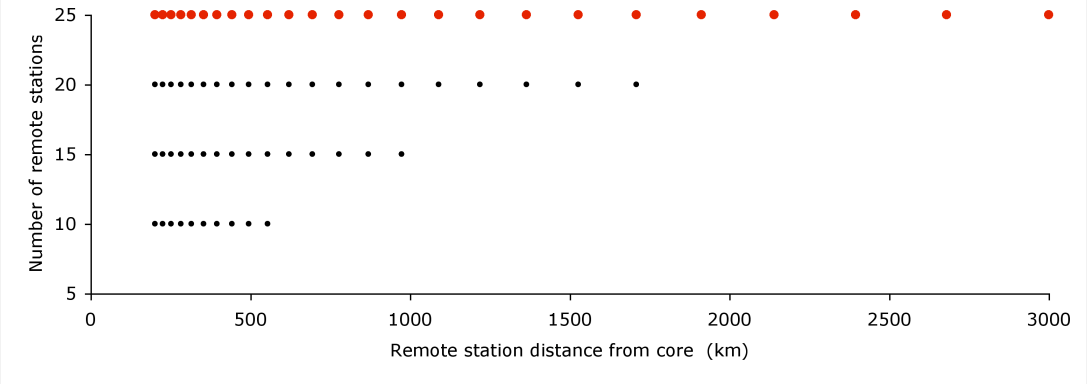
The first two performance–cost trades vary the number of remote stations, maximum baseline length, or both. Irrespective of the metrics varied, each of the trades maintains the exponential distribution of remote stations, illustrated in Figure 6.3 for the default exponential distribution for $N_{\mathrm{RSt}} = 25$ and $b_{\max} = 3000\,\mathrm{km}$. Figure 6.6 shows how the exponential distribution changes as the number of remote stations and maximum baseline length are varied. Figure 6.6a shows the first approach, where a subset of stations in the default distribution are used. Both the maximum baseline length and number of remote stations are reduced by incrementally eliminating the outermost station. Figure 6.6b shows the second approach, where the maximum baseline length is kept constant at 3000 km, and the distribution of the N_{RSt} remote stations is recalculated to maintain an exponentially increasing distance between stations. Figure 6.6c shows the third approach, where the number of remote stations is kept constant at $N_{\mathrm{RSt}} = 25$. As the maximum baseline length is varied, the distribution of the remote stations is recalculated to maintain an exponentially increasing distance between stations.

When the number of remote stations is varied in the first two approaches, the total effective area of the remote stations remains unchanged; as described in Section 6.3.1.3, the 600 dishes can be redistributed amongst the remaining remote stations.

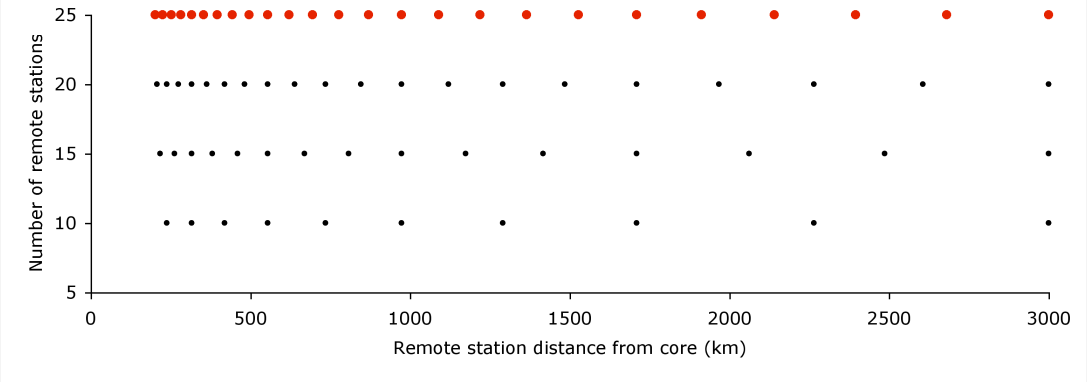
The first performance–cost trade is network cost as function of the number of remote stations N_{RSt}, shown in Figure 6.7a. This trade uses the distributions illustrated in Figures 6.6a and 6.6b. For the 'subset' distribution (open markers), both N_{RSt} and $b_{\max}$ reduces simultaneously. Therefore, as the number of stations is reduces, the network cost decreases faster than the constant maximum baseline length ($b_{\max} = 3000\,\mathrm{km}$) distribution (filled markers).

The second trade is network cost as function of maximum baseline length $b_{\max}$, shown in Figure 6.7b. Again, the subset of the $N_{\mathrm{RSt}} = 25$ and $b_{\max} = 3000\,\mathrm{km}$ distribution is used (Figure 6.6a), along with the distribution of a constant number of stations (Figure 6.6c). The data for the subset distribution is the same as plotted in Figure 6.7a (the exponentially increasing $b_{\max}$ causes the direction of the curve to change, compared to the linearly increasing N_{RSt}). Again, the network cost of the subset distribution (open markers) decreases faster than the $N_{\mathrm{RSt}} = 25$ distributions (filled markers).

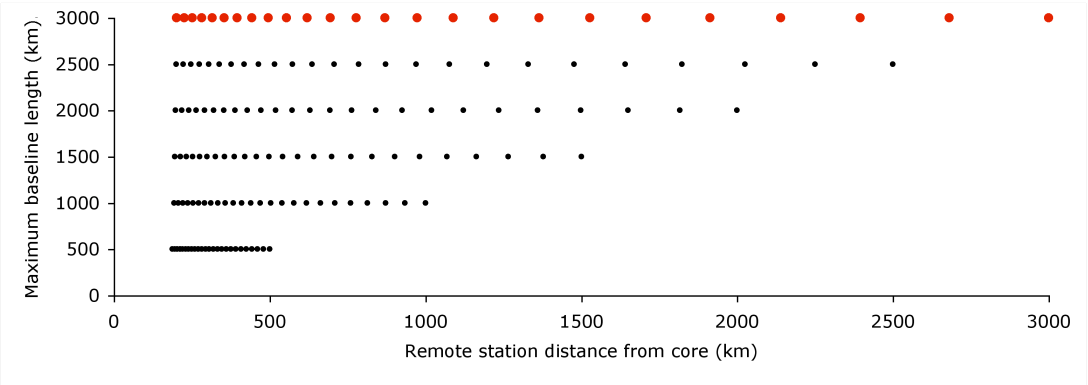
To illustrate the uncertainty of these results, Figure 6.8 re-plots the 8 GHz beam–bandwidth curve for constant maximum baseline length from Figure 6.7a. The black bars in Figure 6.8 show the minimum and maximum cost, using the low and high cost estimates in Appendix G (Table G.3). The cost of the fibre tails represents an additional uncertainty not plotted here;



(a) Station subset: both maximum baseline length and number of stations varied, using a subset of the $N_{\mathrm{RSt}} = 25$, $b_{\mathrm{max}} = 3000\,\mathrm{km}$ distribution.

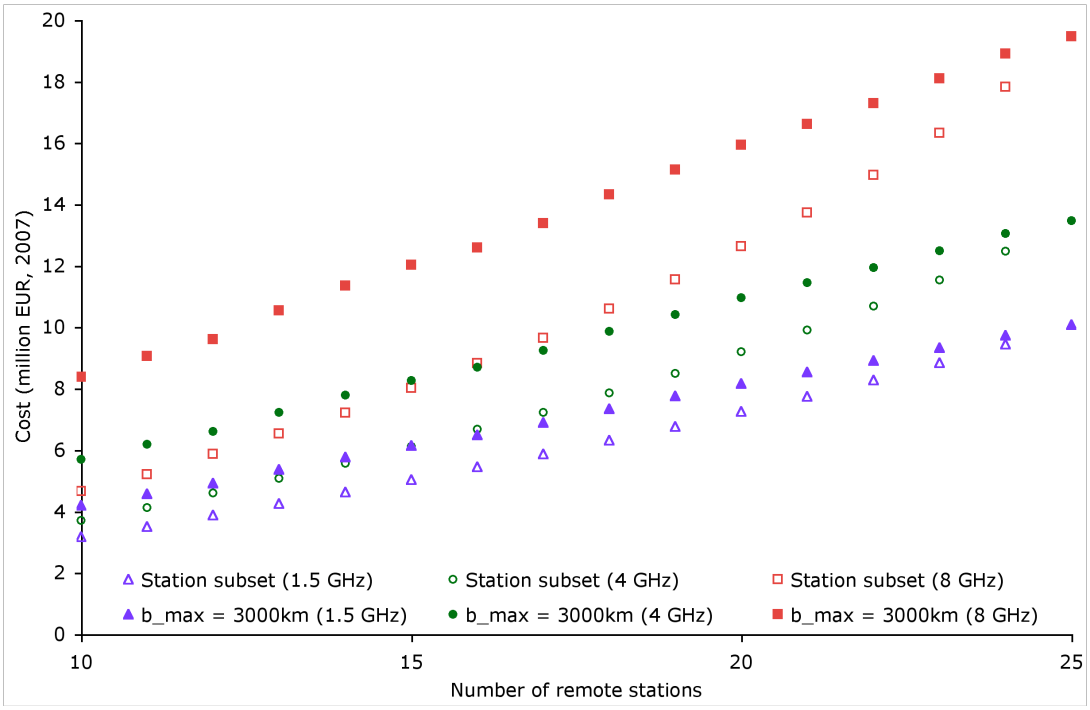


(b) $b_{\mathrm{max}} = 3000\,\mathrm{km}$: constant maximum baseline length and a variable number of remote stations.

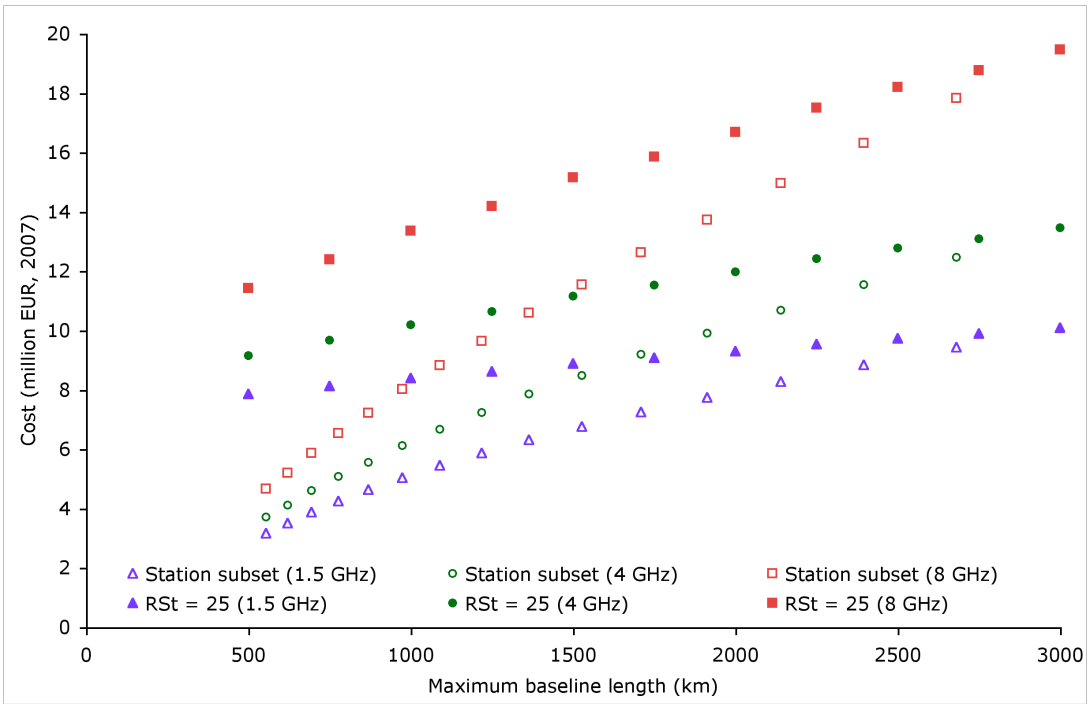


(c) $N_{\mathrm{RSt}} = 25$: constant number of stations and a variable maximum baseline length.

Figure 6.6: Baseline distribution of remote stations. Each point shows the distance between a remote station and the core. (a) and (b) show arrays where $N_{\mathrm{RSt}} = 10$, 15, 20 and 25. (c) shows arrays where $b_{\mathrm{max}} = 500$, 1000, 1500, 2000, 2500 and 3000 km. When $N_{\mathrm{RSt}} = 25$ and $b_{\mathrm{max}} = 3000\,\mathrm{km}$ (red points), all three methods have the same distribution.



(a) Cost as a function of the number of remote stations N_{RSt}, for constant maximum baseline length and subsets of the $N_{\mathrm{RSt}} = 25$, $b_{\mathrm{max}} = 3000\,\mathrm{km}$ distribution.



(b) Cost as a function of maximum baseline length b_{max} , for constant number of remote stations and subsets of the $N_{\mathrm{RSt}} = 25$, $b_{\mathrm{max}} = 3000\,\mathrm{km}$ distribution.

Figure 6.7: Network cost (capital cost) for beam–bandwidths of 1.5, 4 and 8 GHz. The optimal transponder data rate is used, as discussed in the text.

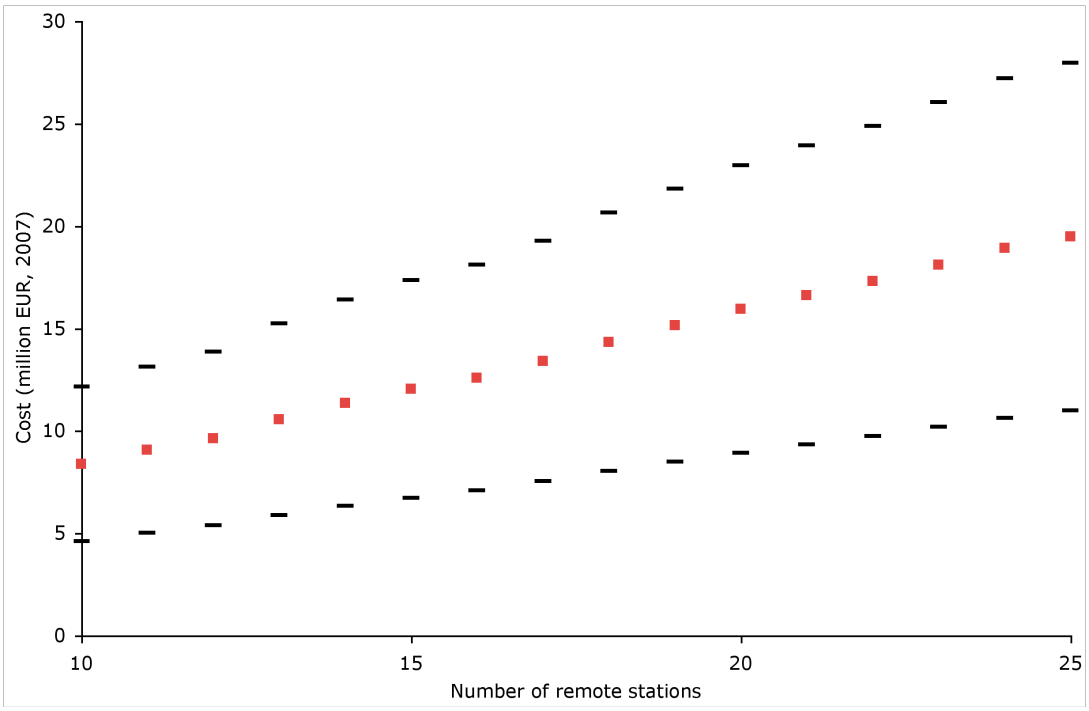


Figure 6.8: Network cost uncertainty as a function of the number of remote stations N_{RSt}, for constant maximum baseline length and 8 GHz beam–bandwidth. The black bars show minimum and maximum cost.

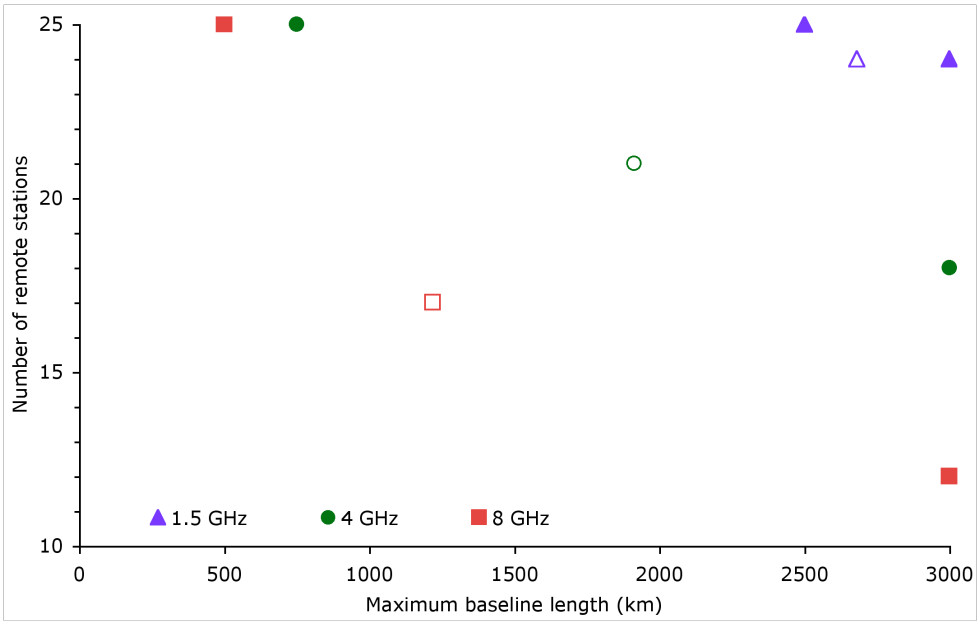


Figure 6.9: Long baseline design solutions of similar network cost, being ∼50 % of the $N_{\mathrm{RSt}} = 25$, $b_{\max} = 3000$ km and 8 GHz beam–bandwidth design solution. Each marker represents a design solution. The open markers are for subsets of the $N_{\mathrm{RSt}} = 25$, $b_{\max} = 3000$ km distribution, the filled markers are for $b_{\max} = 3000$ km or $N_{\mathrm{RSt}} = 25$.

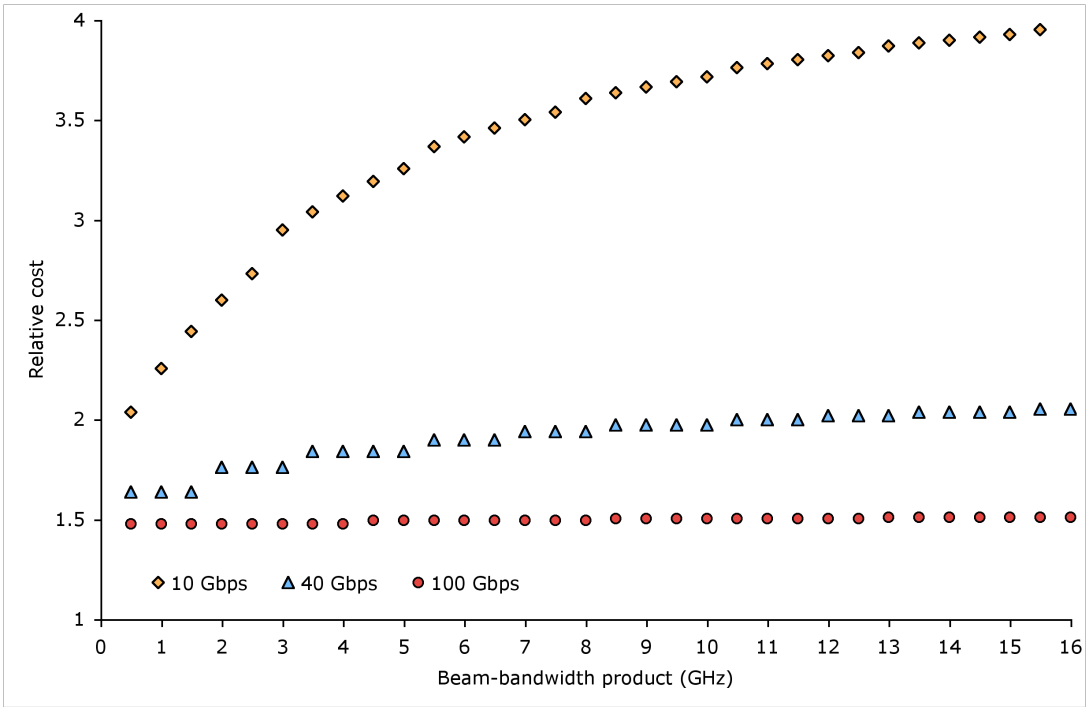


Figure 6.10: Increase in network cost as a function of beam–bandwidth product, for an order of magnitude increase in shared link cost ($K_{\text{shared - link}} = 10$). Cost is relative to network cost for $K_{\text{shared - link}} = 1$. Other details as per Figure 6.10.

as mentioned in Section 6.3.2, the fibre tail costs for any particular remote station are likely to be affected by local geographical and political factors.

Design solutions of similar cost form a useful basis for comparing cost-effectiveness. Each data point in Figures 6.7a and 6.7b represents a specific design solution for the SKA long baselines, where each design solution has slightly different performance characteristics, and a different cost for the network architecture. For $N_{\text{RSt}} = 25$, $b_{\max} = 3000\,\text{km}$ and 8 GHz beam–bandwidth, the network cost is approximately €19.5 M. A target cost reduction may be, for example, 50 % (∼€9.7 M). Figure 6.9 shows design solutions from Figure 6.7 that are closest to the target 50 % cost reduction. Each design solution in the plot represents a specific value of N_{RSt}, $b_{\max}$ and beam–bandwidth. The design solutions with a beam–bandwidth of 1.5 GHz are clustered close to the $b_{\max} = 3000\,\text{km}$, $N_{\text{RSt}} = 25$ point. To maintain the same 50 % cost reduction, higher beam–bandwidths require a trade with the number of remote stations, maximum baseline length, or both. For a beam–bandwidth of 1.5 GHz, $b_{\max} = 3000\,\text{km}$ and $N_{\text{RSt}} = 25$ (not plotted), the cost is ∼€10.1 M, or 52%.

Although the results in Figure 6.9 do not represent every potential design solution, they do give an indication of the performance trades required for a network architecture of similar cost.

6.4.3 Increased shared link costs

To assess the potential of the shared links as a cost-driver, Figure 6.10 compares the previously plotted network costs with an order of magnitude increase in shared link costs ($K_{\text{shared - link}} =$

10). For the 100 G transponder, the relative cost only slightly increases with beam–bandwidth; the increase is larger for the 40 G transponder. For the 10 G transponder, the higher shared link costs has a significant impact on the relative cost, even for the smaller beam–bandwidths. An increasing beam–bandwidth product requires many more channels, thereby accentuating this cost increase. Despite the 40 G transponder having a higher relative cost than the 100 G transponder, the transponder speed selected for each beam–bandwidth product listed in Table 6.5 remains optimal.

Figures 6.11a and 6.11b show the increased cost as a function of number of remote stations and maximum baseline length respectively. If the maximum baseline length of 3000 km is maintained (Figure 6.11a), a decrease in the number of remote stations does not change the relative cost. However, Figure 6.11b shows that the relative cost reduces for a smaller maximum baseline length, regardless of the number of remote stations.

6.5 Discussion

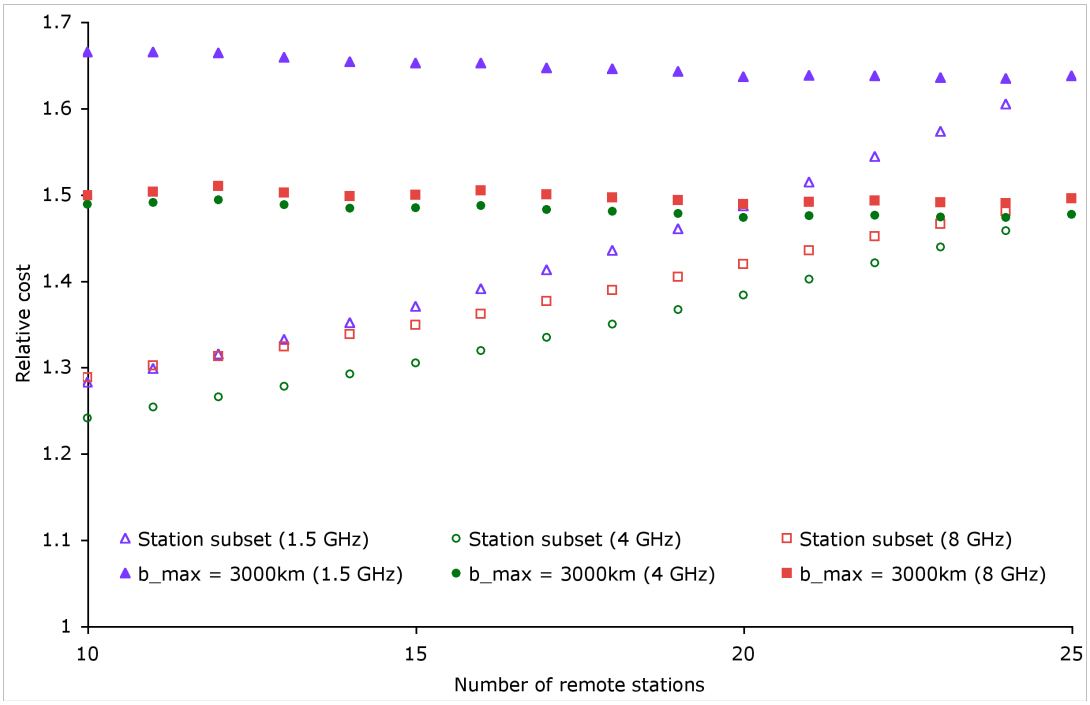
The representative implementation in the high-level system description was developed by specifying receptors to meet science requirements, then designing the subsequent sub-systems in the signal path to processes the signals from these receptor (Dewdney et al., 2010b). My results show that there are alternative approaches to specifying the design, and that the alternative approaches may result in more cost-effective designs.

In this section I discuss the high-level effects of the results on telescope performance (Section 6.5.1), the principal design and cost trades of the network architecture (Section 6.5.2) and the scope for further work (Section 6.5.3).

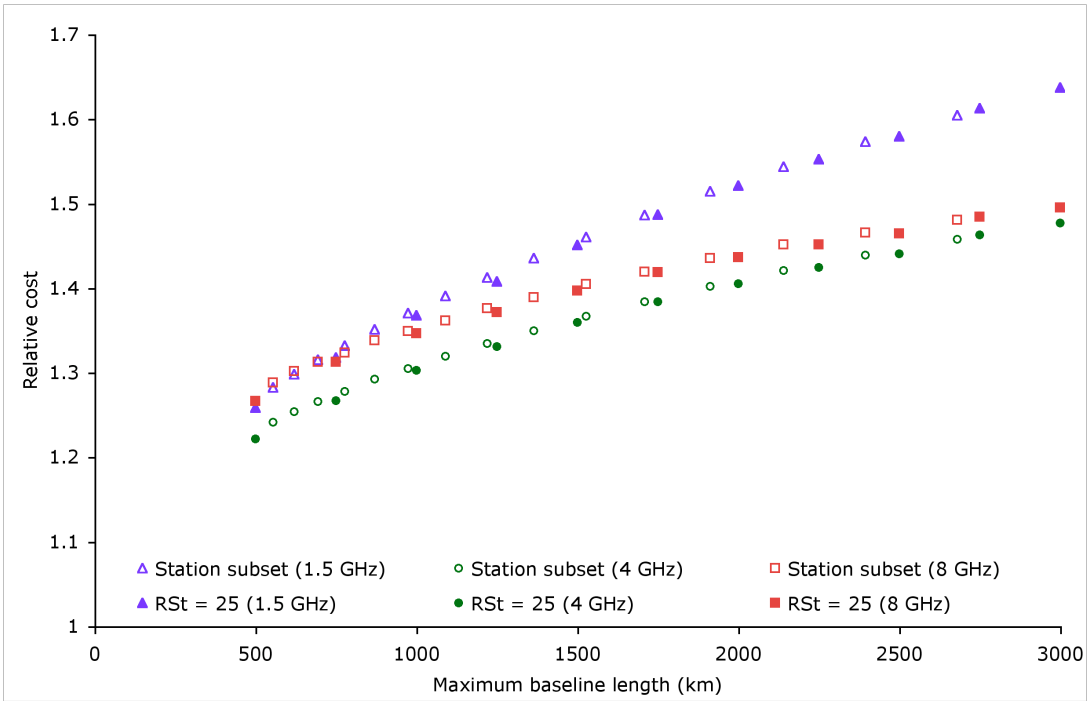
6.5.1 Performance and cost trades

The performance metrics (beam–bandwidth product, maximum baseline length and the number of remote stations) are cost-drivers for the long baseline data network, but also affect telescope performance. Figure 6.7 illustrated the basic scientific trades that can be made to reduce network costs, through varying these metrics. Figure 6.9 summarised these trades in terms of design solutions available for approximately half the cost of the network design for $N_{\mathrm{RSt}} = 25$, $b_{\max} = 3000\,\mathrm{km}$ and beam–bandwidth= 8 GHz. For example, lowering the beam–bandwidth to 1.5 GHz reduces cost by nearly half, without sacrificing angular resolution or (u, v) coverage. If a higher beam–bandwidth is required for sensitivity, then other trades to reduce cost involve a significant reduction of the maximum baseline length, number of remote stations, or both.

The modelling framework presented here enables the trades to be investigated; the subsequent task is for scientific analyses to determine which trades are acceptable. The high-level performance implications can be evaluated by considering the parameter space that is available. As outlined in Chapter 2 (Section 2.1.2), each telescope sub-system acts as a filter on parameter space, by restricting the signal arriving at the observer. The principal axes of parameter space for which the long baseline network design filters parameter space are discussed below.



(a) Relative network cost as a function of maximum baseline length $b_{\max}$. Other details as per Figure 6.7a.



(b) Relative network cost as function of the number of remote stations N_{RSt}. Other details as per Figure 6.7b.

Figure 6.11: Increase in network cost, for an order of magnitude increase in shared link cost ($K_{\mathrm{shared-link}} = 10$). Cost is relative to network cost for $K_{\mathrm{shared-link}} = 1$.

6.5.1.1 Processed field of view

For the long baselines, the station beamforming limits the processed FoV Ω_{proc}. With $N_{\text{b-0}}$ beams formed from a receptor diameter of D_0,

$$\Omega_{\text{proc}} \propto N_{\text{b-0}} D_0^{-2}. \tag{6.2}$$

Therefore, the processed FoV of a single beam from a remote station of diameter ∼350 m, is more than two orders of magnitude smaller than the primary beam of a 15 m dish. As mentioned in Section 6.3.1, this is not a problem for single-target observations. However, science goals containing a survey component, such as *resolving AGN and star formation in galaxies* in DRM_2 (SSWG, 2009), would require further analysis to ensure that sufficient processed FoV is available. In DRM_2, the required survey speed for the *resolving AGN and star formation in galaxies* experiment is $1000\,\text{m}^4\text{K}^{-2}\text{deg}^2$ at 1 GHz. For a sensitivity of $A/T = 10\,000\,\text{m}^2\text{K}^{-1}$, this equates to a required processed FoV larger than $10^{-5}\,\text{deg}^2$. The processed FoV of the 350 m station of approximately $2 \times 10^{-3}\,\text{deg}^2$ is therefore sufficient for this survey.

From a network design perspective, a smaller beam–bandwidth reduces the processed FoV by limiting the number of beams that can be formed for an observation of given bandwidth. This can be detrimental if multiple beams are required to provide independent FoVs for phase calibration and the full bandwidth is required for all beams. None of the trades in Section 6.4 limit the processed FoV *per se*, because either multiple beams can be formed or the data from all 24 dishes can be transported to the central processing facility. However, the reduced instantaneous bandwidth will affect sensitivity, as discussed next.

6.5.1.2 Sensitivity

Two measures of sensitivity are useful for high angular resolution radio astronomy. Image sensitivity measures the minimum detectable flux density of a source in the aperture synthesis image; the sensitivity increases with longer integration time. Baseline sensitivity gives the detectability of a source using a pair of receptors. Baseline sensitivity is important for calibration: phase referencing requires a calibrator source to be observed with short integrations to maintain coherence (Walker, 1999). Thus while the image sensitivity can be calculated for many hours of integration, the baseline must be sensitive enough to observe the calibrator on a timescale of minutes.

As described in Section 6.3.1.3, the dishes are distributed amongst the remote stations such that there are always 600 dishes on the long baselines, located at exponentially increasing distances from the core. The point source image sensitivity remains unchanged by this exponential distribution. However, the baseline sensitivity will vary. Fewer remote stations increases the number of dishes per station, hence the baseline sensitivity. These more sensitive baselines improve the probability of finding a suitable calibrator, as discussed in Godfrey et al. (2011).

The network design only influences the image sensitivity via processed bandwidth. For continuum observations, sensitivity $\propto \sqrt{\Delta\nu}$. Table 6.6 shows the reduction in sensitivity relative to 9 GHz beam–bandwidth, for a constant number of beams. Even the 1.5 GHz beam–bandwidth retains close to half the sensitivity of the 9 GHz beam–bandwidth. For 8 GHz beam–bandwidth, the loss of sensitivity is negligible.

Table 6.6: Sensitivity relative to 9 GHz beam–bandwidth product for a continuum observation (the number of beams remains constant).

Beam–bandwidth product (GHz)	Relative sensitivity
1.5	0.41
4	0.67
8	0.94

6.5.1.3 (u, v) coverage

Despite its influence on the long baseline network costs, the required (u, v) coverage is not well-specified for high angular resolution observations. Although various metrics have been developed to measure the scientific effectiveness of proposed SKA configurations (Millenaar & Bolton, 2011, and references therein), the scientific analyses have not used these metrics. For example, neither the SKA_2 Design Reference Mission (SSWG, 2009) or Godfrey et al. (2011) develop requirements for sampling the (u, v) plane; large-scale simulations are difficult. Furthermore, for a fixed number of dishes on long baselines, an improved (u, v) coverage through more remote stations will reduce the baseline sensitivity. Until these requirements and trade-offs are further quantified, the (u, v) coverage can only be considered here in broad terms.

For the exponential distribution, the number of remote stations is the primary driver of (u, v) coverage. For a given maximum baseline length, more remote stations increases the number of correlated baselines, thereby improving the sampling of the (u, v) plane. Figure 6.7a shows that the number of remote stations can be a significant cost driver for networks costs. The cost increase is approximately linear, but the rate of increase is higher for larger beam–bandwidths.

6.5.1.4 Angular resolution

The angular resolution is approximated by the maximum baseline length. Figure 6.7b shows that the maximum baseline length can be a significant cost driver for 8 GHz beam–bandwidth. For the lower beam–bandwidths, maximum baseline length is not a significant cost driver; the reason being that the the network costs of the lower beam–bandwidths are dominated by the fibre tails costs; these total €6.25 M for the 25 remote stations. For the subset distribution, where the number of remote stations is reduced, as well as the maximum baseline length, the maximum baseline length becomes a significant cost driver for all beam–bandwidths.

6.5.2 Design and cost trades

The design and cost trades considered in this chapter concern the optimisation of the transponder speed, and its relationship to the cost of the shared link. The problem of selecting an optimal transponder speed is not exclusive to the SKA data network; network carriers consider this problem within the broader context of network design and planning. For example, on long-distance DWDM links with high channel utilisation, the deployment of higher data rate transponders maximises the pre-existing investment in fibre cable, amplification and associated infrastructure; such deployment has the potential to reduce total system costs for the network

operators (see e.g. Batten, 2008; Huelsermann et al., 2008; Simmons, 2008; Weichenberg et al., 2009). Although the *per Gbps* transponder cost will reduce with technology improvements, it is likely that industry trends will determine which transponder speeds are commodity components at the time of hardware procurement for SKA_2. To generalise this problem, the modelling in this chapter assumes the same cost per Gbps for the 10 G, 40 G and 100 G transponders.

Because the transponder cost for a given beam–bandwidth is independent of the transponder speed, other trade-offs are more visible. A seemingly obvious, but nevertheless important result of this work is that optimising the beam–bandwidth for the network technology can significantly reduce costs. Section 6.4.1 shows that lowering the beam–bandwidth from 9 to 8 GHz reduces the cost by €3.7 M; the actual cost saving is will be at least double, once operational costs are factored in. The cost reduction is significant because for the 100 G transponder, the step for an additional transponder lies between the 8 and 9 GHz beam–bandwidth.

While the transponder speed is not a 'requirement', it is a design parameter that can significantly influence the network cost. The effect of this cost can be mitigated by matching the output data rate from the station to the transponder speed. In the site submissions, the networks for the long baselines of the SKA_2 dish array were costed for a data rate of 216 Gbps per station (9 GHz beam–bandwidth). The Australian submission specifically costed 100 G transponders for the long baseline stations (ANZSCC, 2011). Although the publicly available portion of the South African submission does not specify the technical details of the long baseline network, there remains the potential to reduce costs by limiting the data rate to below 200 Gbps per station. The change in cost also applies to the shared links, because an one less channel is required.

The cost of the shared links may depend on a variety of factors, as outlined in Section 6.2.1. The shared link multiplier accounts for uncertainties in the cost of sending data via the network carrier. Section 6.4.3 tests the sensitivity of the results to an order of magnitude increase in shared link costs, representing a higher estimate of network carrier costs. Lower transponder speeds and increased beam–bandwidth are more sensitive to higher shared link costs (Figure 6.10). This occurs because lower transponder speeds and increased beam–bandwidth use more channels to transmit the data. For the trades in maximum baseline length and number of remote stations (Figure 6.11), longer maximum baselines are more sensitive to the higher shared link cost, regardless of the number of remote stations. The 1.5 GHz beam–bandwidth with the 40 G transponder is most sensitive to the higher shared link cost. This is because the transponder costs are less than for 4 and 8 GHz beam–bandwidths, thus the shared link cost forms larger proportion of the network cost. Despite this, the 1.5 GHz beam–bandwidth remains the cheapest design.

6.5.3 Further work

This chapter provides a modelling framework for further trades, as more information becomes available. From a scientific perspective, more specific analyses of the minimum (u, v) coverage and baseline sensitivity requirements for high angular resolution observations will indicate how many remote stations are required. Also, given present data rate limitations, the beam–bandwidth product is a useful measure of both performance and cost. A more detailed analysis

of calibration and bandwidth requirements will indicate what beam–bandwidth product is acceptable in the initial years of operation for SKA long baselines.

From a technical perspective, the cost of the shared links is an area for further evaluation by SKA and industry network designers. The capability and extent of DWDM carrier networks in southern Africa will be a factor; as noted in the South African site submission, these networks will have expanded by 2020 (SKA SA, 2011).

The technological changes of the optical components will also affect the design and cost of SKA long baselines. Transmission data rates are likely to be the cheapest part of the network to upgrade, as these components improve following industry trends. Channel data rates of 400 Gbps, 1 Tbps and greater are being researched (Gringeri et al., 2012; Xia, 2011); it remains to be seen whether these data rates will require significant upgrades to fibre infrastructure, and what the implications are for the SKA. An upgrade to data rates will allow higher beam–bandwidths, where eventually transmission of the full bandwidth of each dish data stream may be possible, instead of just station beams. This will open up new scientific opportunities, such as imaging the full dish FoV.

6.6 Chapter summary

The modelling framework presented in this chapter is based on a exemplar data network for the long baselines of the SKA_2 dish array that maximises the use of infrastructure from existing carrier networks. The framework uses performance metrics as scalable parameters for both performance and cost trades of this data network. The performance metrics are: the number of remote stations; the maximum baseline length; and the beam–bandwidth product. From a cost perspective, these metrics drive the cost of the data transmission components, the shared link costs and the fibre tails required for each remote station. In terms of scientific performance, they can be used to assess how the data network restricts access to observational parameter space, in terms of processed field of view, sensitivity, (u, v) coverage and angular resolution.

Using the performance metrics as input parameters, I assess the cost-effectiveness of various network design solutions. This first-order analysis of network costs shows that reducing the beam–bandwidth product from 9 GHz to 8 GHz potentially saves millions of euros, by optimally using the data transmission components. An example science trade is that reducing the beam–bandwidth product to 1.5 GHz nearly halves the network cost, without reducing (u, v) coverage or angular resolution. If the data transmission costs are lower at the time of procurement for the SKA_2, these designs can relatively easily be 'upgraded' to larger beam–bandwidths.

Chapter 7

Conclusions

The work in this thesis investigated design challenges for high time resolution, low frequency and high angular resolution observations with the SKA. The results from each of these chosen case studies are significant in their own right, as presented in Chapters 4–6. In addition to these specific contributions, the cases studies exemplify the utility of parametric performance and cost modelling in achieving cost-effective design outcomes.

7.1 Parametric performance and cost modelling

To identify cost-effective design solutions in this thesis, I have utilised top-level measures of scientific performance within the context of a systems engineering process. Chapter 2 reviewed a set of metrics and figures of merit, such as sensitivity and survey speed, as top-level descriptors of telescope scientific performance. Chapter 3 placed these descriptors within the context of SKA systems modelling and trade studies.

Both the process of model development and the trade studies themselves show where cost-effective design solutions are likely to lie, or where further investigation of requirements, cost and performance will help to determine cost-effective solutions. For example, the low-cost fast transient processor for SKA_1 is a cost-effective design solution (Section 4.7.1). In contrast, the similar costs of the single and dual-band SKA_1-low implementations show that data gathered from further studies are needed to inform the choice between these design solutions (Section 5.8). Regardless of the specific results, the case study analyses were made possible by realistic performance and cost modelling frameworks.

7.2 Case study results and insights

The case studies in this thesis investigated alternative modelling frameworks and design solutions to those presented in the current high-level system descriptions and other Concept Design Review (CoDR) documents[1]. These case studies have provided insight into specific design challenges that otherwise may not have received sufficient attention. For each case, I employed the systems analysis framework detailed in Section 3.3.2 to analyse the requirements, develop a modelling framework and conduct trade studies. The key results and insights from these cases studies are re-iterated below.

7.2.1 Fast transients

The recent, potentially extragalactic, single-pulse detections with the Parkes multibeam system (Lorimer et al., 2007; Keane et al., 2011, 2012; M. Bailes et al., 2012, pers. comm.) show

[1] Principally the high-level system descriptions for SKA_2 (Dewdney et al., 2010b) and SKA_1 (Dewdney et al., 2011a), and the aperture array (Bij de Vaate et al., 2011) and signal processing (Turner et al., 2011) sub-system descriptions.

Table 7.1: Summary of systems analysis outcomes for fast transients (Chapter 4).

Requirements analysis

- Top-level requirements were to:
  - maximise exploration of the high time resolution Universe with searches for singly occurring or intermittent impulsive signals
  - enable low-cost fast transient searches for SKA_1.

Modelling framework

- Modelled search strategies that aim to maximise the number of verifiable events detected in a survey.
- Developed *event rate per beam* as a new metric to measure the cost-effectiveness of fast transient searches, and used *probability of intercept* to determine the total effectiveness of a survey.
- Evaluated fly's eye and incoherent and coherent signal combination modes for SKA_1 aperture arrays and low band dishes.

Trade study results

- Incoherent (phase insensitive) signal combination achieves the highest event rate per beam.
  - A commensal processing system using incoherent signal combination with coherent follow-up is a cost-effective fast transient search system for SKA_1.
  - Higher sensitivity targeted searches can explore those regions of high time resolution parameter space not accessible to the incoherent approach.
- Frequency-dependent effects have considerable impact on preferred receptor and frequency range. These frequency-dependent effects are:
  - direction (pulse broadening)
  - population characteristics (spectral index, spatial distribution of the objects)
  - field of view
  - sensitivity.
- The lower frequencies generally result in higher event rates, especially for steep spectrum sources and when scattering is lower.

that the high time resolution Universe is much more interesting at radio frequencies than was considered even a decade ago. Large, single-dish telescopes such as Parkes and Arecibo have traditionally been used in high time resolution surveys. However, the key to further detection and understanding of the phenomena producing these signals may be the improved field of view (FoV) and sensitivity afforded by radio telescope arrays, combined with their additional verification capabilities. Given the additional costs of retaining both FoV *and* sensitivity with radio telescope arrays, Chapter 4 investigated the system design implications of fast transient survey strategies for SKA_1 and radio telescope arrays in general, with the goal of identifying low-cost survey strategies capable of observing the large unexplored regions of high time resolution parameter space. Table 7.1 summarises the outcomes from Chapter 4 in the context of the systems analysis process.

Chapter 4 is the first comprehensive analysis of the scientific potential of fast transient searches with SKA_1, and demonstrates that SKA_1 in its own right will be a powerful detector for fast

transients, able to probe new and astronomically interesting parameter space. In particular, SKA_1-low is a highly capable fast transient instrument; this capability has not previously been quantified. The metrics used in Chapter 4 provide new insight into the cost-effectiveness of fast transient searches: event rate is an effective metric for estimating the probability of intercepting fast transient events, while the event rate per beam metric provides a first-order cost of intercepting the events. It also demonstrates how the exploration of new high time resolution parameter space is maximised by having flexible search modes that can be adjusted to suit the sky direction and expected source population.

7.2.2 Single and dual-band SKA_1-low

Recent engineering studies for SKA_1-low (e.g. Bij de Vaate et al., 2011) have assumed a single antenna element design (single-band implementation) to observe the 70–450 MHz frequency range. However, there is more design scope to optimise a dual-band implementation for the key low-frequency science drivers for SKA_1: epoch of re-ionisation (EoR) observations (centred at ∼100 MHz); and pulsar searching and timing, which is optimal at the top-end (∼450 MHz) of the SKA_1-low band (Huynh et al., 2012). Furthermore, a revised science case for EoR observations lowers the minimum frequency from 70 MHz to 54 MHz or less (Mellema et al., 2012), which would strongly favour a dual-band system to cover the 9 : 1 fractional bandwidth. Chapter 5 compared the cost-effectiveness of implementing low-frequency aperture arrays as a single-band or dual-band system, to test whether the cost of the dual-band system is double, as might naively be expected from an array with twice the number of elements. Table 7.2 summarises the Chapter 5 outcomes.

The key insight from Chapter 5 is that the cost of a single versus dual-band array is not significantly different *a priori*. The cost comparison depends on design and architectural choices, and detailed scientific requirements and other constraints. An example cost driver of the single-band array is the processing required to form beams at the higher frequencies ($\gtrsim$ 200 MHz). A 20 $\deg^2$ processed FoV is specified in the SKA_1 high-level system description (Dewdney et al., 2011a) to achieve the survey speed requirement for *Tracking Galaxy Evolution over Cosmic Time via HI Absorption* (Chapter 3 of DRM_1, (SSWG, 2012)). Options to reduce this FoV requirement include increasing the survey time beyond the nominal 2 years, increasing sensitivity or reducing the performance capability at these higher frequencies.

Detailed scientific requirements will also help narrow the design choices. In terms of the major science goals applicable to SKA_1-low, the analysis by Mellema et al. (2012) advances the scientific requirements for EoR observations in the context of overall SKA_1-low performance. A similar analysis is required for pulsar searching and timing. For example, Chapter 5 recognises the cost advantages of a more compact core for pulsar searches, but this is not accounted for in the recent analysis of pulsar searching and timing (Section 2.4 of Huynh et al., 2012), nor in a previous analysis for SKA_1 (Smits et al., 2011a).

If there are strong scientific requirements for a large effective area at the lowest frequencies and a large processed FoV at the higher frequencies, Chapter 5 shows that the dual-band SKA_1-low becomes increasingly favourable compared with the single-band implementation.

Table 7.2: Summary of systems analysis outcomes for SKA_1-low (Chapter 5).

Requirements analysis

- The principal SKA_1-low analysis used existing requirements from the SKA_1 Design Reference Mission (DRM_1).
- Supplementary analyses considered changes to the requirements, through modifiers such as a larger number of smaller-diameter stations and reduced processed FoV.

Modelling framework

- Developed scalable parametric models of the SKA_1-low station hardware and broader system costs; the station hardware costs used two different cost data sources developed with different methodologies.
- Used differential cost estimates to compare the cost of single and dual-band implementations of similar performance.
- Considered design alternatives in detail:
  - single and dual-band arrays
  - analogue tile and all-digital beamforming
  - smaller station diameter
  - reduced processed FoV, achieved with a lower beam–bandwidth product (the product of processed FoV and instantaneous bandwidth).

Trade study results

- Despite the dual-band array having twice the number of antenna elements, neither the representative single or dual-band implementations are cheaper *a priori*. Furthermore, the cost uncertainties are currently high, for both the station hardware and the broader comparative system costs.
- Design alternatives, such as RF tile beamforming and smaller station diameters, have potentially significant cost and power implications for the both the single and dual-band implementations.
  - These design alternatives cannot be considered in isolation from the single vs. dual-band comparison, or from each other.
  - Changed scientific requirements, such as reduced processed FoV at higher frequencies, can also impact cost.
- Less obvious, but potentially significant SKA_1-low cost drivers include central processing, antenna deployment and site preparation costs.

Table 7.3: Summary of systems analysis outcomes for long baselines (Chapter 6).

Requirements analysis

- Described scientific performance with simple metrics: beam–bandwidth product; maximum baseline length; and the number of remote stations.

Modelling framework

- Developed scalable parametric models of the exemplar network.
- Used the performance metrics as input parameters for both performance and cost trades.

Trade study results

- Significant cost efficiencies can be achieved by modifying the beam–bandwidth product to optimally use the data transmission components.
- While all three metrics are cost drivers, limiting the beam–bandwidth product reduces cost markedly, without significantly impacting the scientific performance.

7.2.3 Long baselines of the SKA_2 dish array

For the long baselines of the SKA_2 dish array, both scientific requirements and network design solutions have been previously developed in separate analyses[2]. However, the cross-connection between the science and the engineering design is low. Chapter 6 of this thesis developed a modelling framework to link telescope performance with network cost, using an exemplar data network for the SKA long baselines. The aim of the framework is to highlight the science–engineering trades, by informing network designers of the principal performance drivers, and by informing SKA planners and scientists of the principal cost drivers of the long baseline network. Table 7.3 summarises the Chapter 6 outcomes.

Chapter 6 showed that the long baselines need not be a major cost driver for the SKA; significant cost reductions are achieved by modifying the scientific requirements to better suit the available design solutions. This chapter highlights the importance of interaction between the scientific and engineering domains, and provides a modelling framework for further science-engineering trades.

7.3 Cost-effective SKA designs

So, what are the features of these case studies that lead to new results and insights for cost-effective SKA designs? Although the extent of the systems analysis varies between case studies, the parametric models are a key feature of the performance and cost trades. The scientific requirements and the engineered design are inextricably coupled; the models enable a cross-disciplinary approach to trade-offs, using information obtained from both the science and engineering domains to capture this coupling.

Top-level measures of performance are an important tool for estimating the scientific effectiveness of a given design solution. These performance metrics and figures of merit both summarise the technical requirements of the telescope, and also indicate the parameter space to which the

[2] e.g. Godfrey et al. (2011); SSWG (2009) for the science requirements and McCool (2010); ANZSCC (2011); SKA SA (2011) for the network design solutions.

telescope design is sensitive. When used in the parametric performance and cost modelling, the metrics and figures of merit make the trade studies a more tractable problem.

Trade studies require a suitable modelling framework, with the trade study scope and methodology chosen as realistic to the problem; the high-level system models in this thesis provide such a framework. These scalable parametric models allow investigation of the performance and cost impact for multiple design solutions, such as new design features for fast transients detection, a comparative analyses of the single vs. dual-band aperture arrays or design trades for the long baseline data network.

Importantly, selective definition of trade studies within the telescope system ensures that insights into cost-effective design solutions are available prior to the full requirements analysis. Although insufficient data presently limits performance–cost analyses of the full SKA system, the case studies show that first-order estimates of cost-effectiveness can be made for large subsystems. Such estimates require pragmatic choices to narrow the exploration space, combined with an understanding of both the scientific and engineering implications of the trades. With this approach, the results and insights are more rapidly available for the next iteration of the systems analysis.

7.4 Quantifying the exploration of the unknown

In addition to being employed in performance–cost analyses, the top-level measures of performance are also useful indicators of the capability of the SKA to access unexplored regions of parameter space; this access is key to making discoveries. Of course, non-technological factors also influence the discovery of new phenomena; an example being researchers with curiosity, persistence and an open mind (e.g. Kellermann et al., 2009). However, if the telescope simply cannot access particular regions of parameter space, no amount of serendipity will help discover the phenomena residing in those regions. Although the exploration of the unknown is a design philosophy for the SKA, a quantitative measure of the telescope's capability to search new parameter space is required for performance and cost trades.

The top-level performance metrics and figures of merit correspond to the axes of parameter space described in Harwit (1981), and can be translated into technical requirements for the telescope as part of the usual requirements analysis process (see Chapters 2 and 3). Any increase in the capability of the SKA to access new regions of parameter space requires an assessment of cost-effectiveness. The top-level performance metrics allows quantitative trade-offs to be made between cost and system effectiveness, thereby providing more robust information for subsequent design decisions.

Chapter 4 exemplifies both the development of top-level performance metrics, and the subsequent quantitative trade-offs. For the high time resolution Universe, the 'probability of intercept' measures the effectiveness of a survey in searching new parameter space; Figure 1.3 showed that the SKA will explore new regions of high time resolution parameter space. Meanwhile, the event rate per beam metric enables performance–cost trade-offs between search strategies; Chapter 4 quantified these trade-offs. Using these results, this thesis identified a cost-effective design solution for fast transient searches with SKA_1 that accesses the raw data (voltages) via

Table 7.4: Types and sources of data to improve the analyses in this thesis.

Fast transients
• New scientific results from current and planned fast transient searches (Table 4.1) will provide insight into prospective target source populations, and influence SKA fast transient search strategies.
• Detailed design solutions for the SKA signal processing sub-systems will enable full cost-comparisons between search strategies.
SKA_1-low
• More detailed scientific requirements, such as the white paper by Mellema et al. (2012) on observing the epoch of re-ionisation, will highlight the design solutions most suitable for further study.
• Refined cost information for central processing, antenna deployment and site preparation will improve the comparison between single and dual-band SKA_1-low implementations.
• The pathfinder and precursor telescopes such as LOFAR, the MWA and the Long Wavelength Array (LWA) will be important sources of data. For example, a comparison of the different processing architectures and deployment strategies will be help refine cost estimates for SKA_1-low.
Long baseline data network for the SKA_2 dish array
• A more detailed analysis of scientific requirements, including scientific simulations in areas such as proto-planetary disks and AGN jets (L. Godfrey, 2012, pers. comm.), will better inform the relative importance between beam–bandwidth product, maximum baseline length and the number of remote stations.
• e-VLBI arrays will continue to test and inform the ability of carrier networks to handle increasingly large data rates.
• ASKAP and the MWA will utilise a dedicated long distance (∼900 km), high data rate (8 Tbps current capacity) fibre link to the computing facility in Perth, WA (ANZSCC, 2011). LOFAR uses dedicated long distance links to connect the stations to the central processing facility (de Vos et al., 2009). Cost information regarding the operation of these links will be useful inputs for the SKA long baseline network design.

a spigot or transient buffer, enabling the verification of candidate events. Given that fast transients is not currently a design driver for SKA_1, such design solutions highlight significantly increased regions of high time resolution parameter space available to the SKA, for a relatively low impact on cost and signal processing architecture.

7.5 Future work

The case studies in this thesis show that first-order estimates of cost-effectiveness (maximising scientific capability and minimising cost) can be made for large sub-systems, even when there is insufficient performance and cost data for detailed trade-off analyses. For example, the performance analyses in Chapter 4 compared the SKA to other fast radio transient search instruments to show that SKA_1, and perhaps surprisingly, SKA_1-low, will be a powerful detector for fast transients. Meanwhile, the differential cost analysis of the single vs. dual-band SKA_1-low (Chapter 5) shows that *a priori*, neither is implementation is significantly cheaper.

Although the first-order results an important initial step, more data would evidently improve

the analyses. Table 7.4 summarises the types and sources of data to improve the analyses for the case studies in this thesis. An important source of the required data, in both the science and engineering domains, is the precursors, pathfinders and design studies. These projects will test key technologies and sub-systems, with the resulting verification data flowing into the SKA design process (Hall et al., 2008).

To make use of this data, the general modelling framework presented in this thesis is extensible, and the parametric models I developed are highly reusable. The best example is the SKA_1-low modelling, where the minimum required frequency of observation ($\sim$50 or 70 MHz) and the FoV requirements at the higher frequencies will strongly impact the design, including the choice between a single or dual-band SKA_1-low implementation. To extend the models from the SKA_1-low case study, additional information about the cost and performance of active antenna designs capable of observing the lower frequencies is required. Implementing these models in a modelling tool such as SKACost (Section 3.5) will enable further exploration of the cost and performance impact of evolving SKA_1-low requirements and iterated SKA_1-low designs.

During the current formal SKA system definition stage, trade studies such as those shown in this thesis will increasingly be required to support decision-making processes and select design solutions. An example application of the general modelling framework is assessing the cost and performance of mid-frequency design solutions for SKA_2 in the Advanced Instrumentation Program: wideband single-pixel feeds and phased array feeds for the SKA_2 dish array; and mid-frequency aperture arrays. These trade studies require a common modelling framework built on a uniform set of data, and tools such as SKACost to exercise the models. The high-level performance and cost modelling approach presented in this thesis lays the foundation for further trade-offs and optimisations in the pursuit of cost-effective designs for the SKA.

7.6 Final note

This thesis has shown that an analytical approach to requirements analysis and performance–cost modelling, combined with pragmatic choices to narrow the exploration space, yields new insights into cost-effective SKA designs. This approach conforms with, and contributes to, the SKA decision-making processes through systems analysis and trade studies. It also engages the science community by ensuring that the scientific impact of design decisions are more easily understood. Continuation of this cross-disciplinary approach will be essential to successfully integrate the forthcoming results from various verifications systems, including the pathfinder and precursor telescopes, over the next few years.

Appendices

Appendix A

List of publications

Refereed publications related to this thesis are as follows:

- A. Jiwani, **T. M. Colegate**, N. Razavi-Ghods, P. J. Hall, S. Padhi, and J. G. bij de Vaate, *Square Kilometre Array station configuration using a two-stage beamforming,* Publications of the Astronomical Society of Australia, 2012, in press (arXiv:1211.4932).
- S. J. Tingay, R. Goeke, J. D. Bowman, D. Emrich, S. M. Ord, D. A. Mitchell, M. F. Morales, T. Booler, B. Crosse, D. Pallot, A. Wicenec, W. Arcus, D. Barnes, G. Bernardi, F. Briggs, S. Burns, J. D. Bunton, R. J. Cappallo, **T. M. Colegate**, B. E. Corey, A. Deshpande, L. deSouza, B. M. Gaensler, L. J. Greenhill, P. J. Hall, B. J. Hazelton, D. Herne, J. N. Hewitt, M. Johnston-Hollitt, D. L. Kaplan, J. C. Kasper, B. B. Kincaid, R. Koenig, E. Kratzenberg, C. J. Lonsdale, M. J. Lynch, B. McKinley, S. R. McWhirter, E. Morgan, D. Oberoi, J. Pathikulangara, T. Prabu, R. A. Remillard, A. E. E. Rogers, A. Roshi, J. E. Salah, R. J. Sault, N. Udaya-Shankar, F. Schlagenhaufer, K. S. Srivani, J. Stevens, R. Subrahmanyan, S. Tremblay, R. B. Wayth, M. Waterson, R. L. Webster, A. R. Whitney, A. Williams, C. L. Williams, and J. S. B. Wyithe, *The Murchison Widefield Array: the Square Kilometre Array Precursor at low radio frequencies*, Publications of the Astronomical Society of Australia, 2012, in press (arXiv:1206.6945).
- **T. M. Colegate** and N. Clarke, *Searching for Fast Radio Transients with SKA Phase 1*, Publications of the Astronomical Society of Australia, vol. 28, pp. 299–316, Nov. 2011.
- J.-P. Macquart, M. Bailes, N. D. R. Bhat, G. C. Bower, J. D. Bunton, S. Chatterjee, **T. M. Colegate**, J. M. Cordes, L. D'Addario, A. Deller, R. Dodson, R. Fender, K. Haines, P. Hall, C. Harris, A. Hotan, S. Johnston, D. L. Jones, M. Keith, J. Y. Koay, T. J. W. Lazio, W. Majid, T. Murphy, R. Navarro, C. Phillips, P. Quinn, R. A. Preston, B. Stansby, I. Stairs, B. Stappers, L. Staveley-Smith, S. Tingay, D. Thompson, W. van Straten, K. Wagstaff, M. Warren, R. Wayth, and L. Wen, *The Commensal Real-time ASKAP Fast Transients (CRAFT) survey*, Publications of the Astronomical Society of Australia, vol. 27, pp. 272–282, June 2010.

Memos, technical documents and conference papers related to this thesis are as follows:

- P. Hall, **T. M. Colegate**, J.-P. Macquart, N. Clarke, S. Tingay, C. Trott, and R. Wayth, *Towards SKA studies of the radio transient universe*, in Resolving the Sky - Radio Interferometry: Past, Present and Future, (Manchester, UK), Apr. 2012, in press.
- **T. M. Colegate**, P. J. Hall, and A. W. Gunst, *Cost-effective aperture arrays for SKA Phase 1*: single or dual-band?, SKA Memo 140, 2012.
- I. H. Stairs, M. J. Keith, Z. Arzoumanian, W. Becker, A. Berndsen, A. Bouchard, N. D. R. Bhat, M. Burgay, D. J. Champion, S. Chatterjee, **T. M. Colegate**, J. M. Cordes, F. M. Crawford, R. Dodson, P. C. C. Freire, G. B. Hobbs, A. W. Hotan, S. Johnston, V. M. Kaspi, V. Kondratiev, M. Kramer, T. J. W. Lazio, W. Majid, R. N. Manchester, D. J. Nice,

A. Pellizzoni, A. Possenti, S. M. Ransom, N. Rea, R. Shannon, R. Smits, B. W. Stappers, D. F. Torres, A. G. J. van Leeuwen, W. van Straten, and P. Weltevrede, *Pulsars with the Australian Square Kilometre Array Pathfinder*, in Radio Pulsars: An astrophysical key to unlock the secrets of the Universe, vol. 1357 of AIP Conference Proceedings, pp. 335–340, Aug. 2011.

- R. McCool, P. Crosby, D. Hall, **T. M. Colegate**, R. Bolton, and D. Ford, "*SKA costing strategy (draft)*," System Delta CoDR document MGT-040.070.000-MP-001 Rev C, SPDO, 2010.
- D. Ford, R. Bolton, **T. M. Colegate**, P. Alexander, and P. Hall, *The SKA Costing and Design Tool*. SKA Memo 120, 2010.
- D. Ford, R. C. Bolton, **T. M. Colegate**, P. Alexander, and P. Hall, *The SKA cost/performance tool: A hierarchical SKA modelling tool*, in Widefield Science and Technology for the SKA: SKADS Conference, ed. S. A. Torchinsky, A. van Ardenne, A. van den Brink-Havinga, A. J. van Es, & A. J. Faulkner, Château de Limelette, Belgium, 2009.
- R. C. Bolton, P. Alexander, D. C. Ford, **T. M. Colegate**, and P. J. Hall, *System level design trade-offs with the SKA costing tool*, in Widefield Science and Technology for the SKA: SKADS Conference, ed. S. A. Torchinsky, A. van Ardenne, A. van den Brink-Havinga, A. J. van Es, & A. J. Faulkner, Château de Limelette, Belgium, 2009.
- **T. M. Colegate**. *Building mega-science: A systems engineering tool for the Square Kilometre Array*, The Tenth Postgraduate Electrical Engineering and Computing Symposium, Perth, Australia, Oct 2009.

Appendix B

Detecting a radio source

I derive here the system equivalent flux density and minimum detectable flux density as measures of radio telescope sensitivity (see Section 2.2.1).

The radiation delivered by the antenna to the telescope receiver is measured as power per unit bandwidth ($\mathrm{W\,Hz^{-1}}$):

$$P_\nu = \frac{1}{2} A_\mathrm{e} S_\nu, \tag{B.1}$$

where A_e ($\mathrm{m^2}$) is the effective area of the telescope and S_ν is the flux density (or more correctly, spectral flux density) of the source. It indicates the strength of a signal from a radio source and is commonly expressed in Jansky, where $1\,\mathrm{Jy} = 10^{-26}\,\mathrm{W\,m^{-2}\,Hz^{-1}}$. Random polarisation is assumed, hence half the average power is delivered to the singly polarised antenna. For an antenna of noise temperature T_ant (K), where 1 K equals $1.38 \times 10^{-23}\,\mathrm{W\,Hz^{-1}}$, the power level per unit bandwidth at the telescope receiver input can be measured as:

$$P = kT_\mathrm{ant}, \tag{B.2}$$

where $k = 1.38 \times 10^{-23}\,\mathrm{J\,K^{-1}}$ is Boltzmann's constant (Christiansen & Högbom, 1969; Rohlfs & Wilson, 2004).

A radio telescope detects the signal of the source above the Gaussian noise in the telescope system. This means that the noise in the system is a limiting factor in the ability of the telescope to detect the signal. Rohlfs & Wilson (2004) presents an 'intuitive' derivation of the noise limit, which is summarised here:

- The Fourier transform of a Gaussian distribution is Gaussian, and the width of these Fourier transform pairs is related such that

  $$\Delta t \Delta \nu = 1,$$

  where Δt and $\Delta \nu$ are the widths of the Gaussian distribution in the time and frequency domains respectively.
- From the Nyquist sampling theorem, samples are only independent if they are taken at time intervals at least $\Delta t = 1/\Delta\nu$ apart, where $\Delta\nu$ is now the receiver bandwidth.
- Over a total time τ, $N_\mathrm{t} = \tau/\Delta t = \tau\Delta\nu$ independent samples are taken and the rms error for a Gaussian is $1/\sqrt{N_\mathrm{t}}$ of a single sample. If the error of the single sample is given by T_sys (K), then the total rms error for an ideal receiver is

  $$\Delta T = \frac{T_\mathrm{sys}}{\sqrt{\tau\Delta\nu}}. \tag{B.3}$$

This is the smallest rms error achievable for a given system. T_sys is the system temperature of telescope, and includes noise from the source, sky and the receiver; ΔT is therefore larger for stronger sources.

The rms error is used to calculate the minimum detectable flux density of a point source using a single antenna telescope. From Equations B.1 and B.2, the contribution of the source to the noise temperature is

$$T_{\mathrm{ant}} = \frac{S_\nu A_{\mathrm{e}}}{2k}. \tag{B.4}$$

The change in flux density due to the source ΔS should cause the antenna temperature to change by ΔT_{ant}, so

$$\Delta T_{\mathrm{ant}} = \frac{\Delta S A_{\mathrm{e}}}{2k}. \tag{B.5}$$

Re-arranged, this gives the system equivalent flux density S_{sys}:

$$S_{\mathrm{sys}} = \frac{2kT_{\mathrm{sys}}}{A_{\mathrm{e}}}, \tag{B.6}$$

which is defined such that the flux density of a point source in the antenna primary beam would double the noise power compared to the system noise without the source (Thompson et al., 2001; Lorimer & Kramer, 2005).

Equating Equation B.5 with Equation B.3 leads to the rms variation in measured flux density for a randomly polarised source (Christiansen & Högbom, 1969):

$$\Delta S = \frac{2kT_{\mathrm{sys}}}{A_{\mathrm{e}}\sqrt{\Delta\nu\tau}}. \tag{B.7}$$

To ensure that the signal detected is not due to Gaussian statistics, a sigma value is used, such that the minimum detectable flux density is

$$S_{\mathrm{min}} = \frac{\sigma 2kT_{\mathrm{sys}}}{A_{\mathrm{e}}\sqrt{\Delta\nu\tau}}. \tag{B.8}$$

If each polarisation contributes equal power,

$$S_{\mathrm{min}} = \frac{\sigma 2kT_{\mathrm{sys}}}{A_{\mathrm{e}}\sqrt{N_{\mathrm{pol}}\Delta\nu\tau}}, \tag{B.9}$$

where N_{pol} is the number (1 or 2) of oppositely polarised signals detected (Thompson et al., 2001).

Appendix C

Survey metrics and figures of merit

This appendix derives the survey metrics and figures of merit discussed in Section 2.2.

C.1 Survey area and speed

The characteristics of a given survey, including $S_{\rm min}$, are described by a hypervolume in observational parameter space. The multiple observations of the particular class of signal described by the hypervolume can be represented by the total area of sky $\Omega_{\rm tot}$, and calculated as

$$\begin{aligned} \Omega_{\rm tot} &= N_{\rm tot}\Omega_{\rm proc} \\ &= \frac{T_{\rm tot}\Omega_{\rm proc}}{\tau}, \end{aligned} \tag{C.1}$$

where $N_{\rm tot}$ is the number of independent observations (pointings) in the survey. $T_{\rm tot}$ is the total observing time, $\Omega_{\rm proc}$ is the processed FoV and τ is the integration time for each pointing in the survey (equal to time resolution Δt). Equation C.1 does not describe the signal properties to which each pointing in the survey is sensitive; as mentioned, these are described by a hypervolume in the observational parameter space.

Some caveats apply to Equation C.1:

- For steady (time-continuous) sources, each pointing must observe a different patch of sky to be an independent observation.
- For time-varying sources, multiple observations of the same patch of sky can be independent observations if the time interval between observations is such that the properties of the source signal changes.

The most obvious methods to increase $\Omega_{\rm tot}$ is operate the survey for longer, thereby increasing $T_{\rm tot}$, or to increase the processed FoV. Also, for a steady source, the integration time for a single observation can be reduced by increasing telescope sensitivity. Re-arranging Equation 2.1,

$$\tau = \frac{4}{N_{\rm pol}\Delta\nu}\left(\frac{\sigma k T_{\rm sys}}{S_{\rm min}A_{\rm e}}\right)^2. \tag{C.2}$$

Therefore an observation of sensitivity $S_{\rm min}$ can be made more quickly if $T_{\rm sys}$ is decreased, effective area increased, the polarisations are summed and, for continuum radiation, bandwidth increased.

A related metric is areal survey speed; a measure of the processed FoV observed per unit time by the telescope. Equally, it is the total solid angle observed over the duration of the survey, and thus a re-arrangement of Equation C.1:

$$\begin{aligned} {\rm SS} &= \frac{\Omega_{\rm tot}}{T_{\rm tot}} \\ &= \frac{\Omega_{\rm proc}}{\tau}. \end{aligned} \tag{C.3}$$

With substitution of Equation C.2,

$$\mathrm{SS} = \frac{\Omega_{\mathrm{proc}} N_{\mathrm{pol}} \Delta\nu}{4} \left(\frac{S_{\mathrm{min}} A_{\mathrm{e}}}{\sigma k T_{\mathrm{sys}}} \right)^2 . \tag{C.4}$$

Thus SS describes the FoV (solid angle) per unit time that can be surveyed by the telescope, to a sensitivity limit S_{min} (Johnston & Gray, 2006). Equation C.4 is the basis for a survey speed figure of merit SSFoM (Cordes, 2009b):

$$\mathrm{SSFoM} = \Omega_{\mathrm{proc}} \Delta\nu \left(\frac{A_{\mathrm{e}}}{T_{\mathrm{sys}}} \right)^2 . \tag{C.5}$$

C.2 Survey volume

An alternative approach to measure the capability of a telescope to make multiple observations is to consider the number of sources the survey will detect. Following Cordes (2009a), for a homogeneously distributed population of sources in a Euclidean Universe, the number detections is

$$N_{\mathrm{det}} = n_{\mathrm{s}} V_{\mathrm{tot}}, \tag{C.6}$$

where n_{s} is the source number density (number per unit volume) and V_{tot} is the total volume of space observed in the survey; the latter given by

$$\begin{aligned} V_{\mathrm{tot}} &= N_{\mathrm{tot}} V_{\mathrm{proc}} \\ &= \frac{T_{\mathrm{tot}} V_{\mathrm{proc}}}{\tau}, \end{aligned} \tag{C.7}$$

where V_{proc} is the volume processed per pointing. For time-continuous and time-varying sources, the same caveats as for Equation C.1 apply.

The processed volume is a cone of solid angle Ω_{proc} (describing the fraction of the sky observed) and depth of r_{max}:

$$\begin{aligned} V_{\mathrm{proc}} &= \frac{4\pi}{3} \frac{\Omega_{\mathrm{proc}}}{\Omega_{\mathrm{sky}}} r_{\mathrm{max}}^3 \\ &= \frac{\Omega_{\mathrm{proc}}}{3} r_{\mathrm{max}}^3 . \end{aligned} \tag{C.8}$$

The depth r_{max} is the maximum distance to which a source can be detected. Assuming there is no loss of flux density due to intervening media, that distance is

$$r_{\mathrm{max}} = \sqrt{\frac{\mathcal{L}}{4\pi S_{\mathrm{min}}}}, \tag{C.9}$$

where $\mathcal{L}$ is the intrinsic luminosity of an isotropically radiating source.

Substituting S_{min} (Equation 2.1) and Equations C.7, C.8 and C.9 into Equation C.6 gives the number of sources of luminosity $\mathcal{L}$ detected in a survey:

$$N_{\mathrm{det}} = \frac{n_{\mathrm{s}} T_{\mathrm{tot}} \Omega_{\mathrm{proc}}}{3\tau^{1/4}} \left(\frac{\mathcal{L} A_{\mathrm{e}} \sqrt{\Delta\nu}}{8k\pi\sigma T_{\mathrm{sys}}} \right)^{3/2} . \tag{C.10}$$

As done for survey area Ω_{tot}, the detection rate (number of detections per unit time) is a

re-arrangement of Equation C.10:

$$\begin{aligned}\mathcal{R}_{\rm det} &= \frac{N_{\rm det}}{T_{\rm tot}} \\ &= \frac{n_{\rm s} V_{\rm proc}}{\tau} \\ &= \frac{n_{\rm s}\Omega_{\rm proc}}{3\tau^{1/4}} \left(\frac{\mathcal{L} A_{\rm e} \sqrt{\Delta\nu}}{8k\pi\sigma T_{\rm sys}} \right)^{3/2}.\end{aligned} \tag{C.11}$$

If the integration time τ remains constant, $\mathcal{R}_{\rm det}$ can be expressed as

$$\mathcal{R}_{\rm det} \propto \Omega_{\rm proc} \Delta\nu^{3/4} \left(\frac{A_{\rm e}}{T_{\rm sys}} \right)^{3/2}, \tag{C.12}$$

which is of the same form as Equation C.5, but with different exponents. If τ can be traded for sensitivity, substitution of Equations C.2 and C.9 into Equation C.11 gives:

$$\mathcal{R}_{\rm det} = \frac{n_{\rm s}\Omega_{\rm proc} A_{\rm e}^2 \Delta\nu}{48\pi^{3/2} r_{\rm max} \left(\sigma k T_{\rm sys}\right)^2} \mathcal{L}^2. \tag{C.13}$$

Thus the rate of detection is inversely proportional to the maximum survey depth ($\mathcal{R}_{\rm det} \propto r_{\rm max}^{-1}$). If depth $r_{\rm max}$ and source luminosity $\mathcal{L}$ are constant,

$$\mathcal{R}_{\rm det} \propto \Omega_{\rm proc} \Delta\nu \left(\frac{A_{\rm e}}{T_{\rm sys}} \right)^2. \tag{C.14}$$

Appendix D

SKACost trade-off examples

I show here example trade-offs made with SKACost (Section 3.5). Costs are generally limited to the capital cost of the hardware. The reader should refer to the original sources for assumptions, caveats and interpretation of these examples.

D.1 Cost optimisation within a fixed performance scope

Figure D.1 is an example of modelling cost as a function of one or more input parameters, while keeping the high-level requirements fixed. The plot is for an SKA composed of dishes with phased array feeds (PAFs), where cost is given as a function of dish diameter. The number of beams formed by the PAF is varied so that field of view (FoV) remains constant at 20 $\deg^2$. Because sensitivity (A/T) is also held fixed, survey speed figure of merit (SSFoM) remains constant. This plot shows a shallow cost minimum around a dish diameter of 15–20 m. It gives an indication how the correlator and computing costs begin to dominate at the smaller diameters and the dish antenna costs dominate at larger diameters.

D.2 Performance trade-offs within a fixed cost

An alternative approach is to only model telescope designs that meet a fixed cost. Figure D.2 shows an optimisation between SSFoM, A/T and FoV, for dishes equipped with either PAFs or wideband single-pixel feeds (WBSPFs), where the number of dishes is free to vary. In this example, the dishes with WBSPFs show an optimum dish diameter of 10–12 m to achieve both high SSFoM and A/T. For the dishes with PAFs, diameter is fixed at 15 m and the FoV formed by the PAF is varied (values of 6, 15 and 30 $\deg^2$ are labelled).

D.3 Architectural comparisons

Figure D.2 also shows that comparison between two or more design architectures is possible; in this case different receptor technologies are compared (PAFs and WBSPFs).

D.4 Sensitivity analyses

A sensitivity analysis involves varying one or more input or cost parameters or assumptions to explore the effects on system cost. This is a method to determine major cost drivers in the system. A sensitivity analysis also enables identification of areas of risk to the project, by testing specific 'what if' scenarios using different value ranges (NASA, 2008). This is the basis for the Monte Carlo analysis discussed in Section 3.6.3.

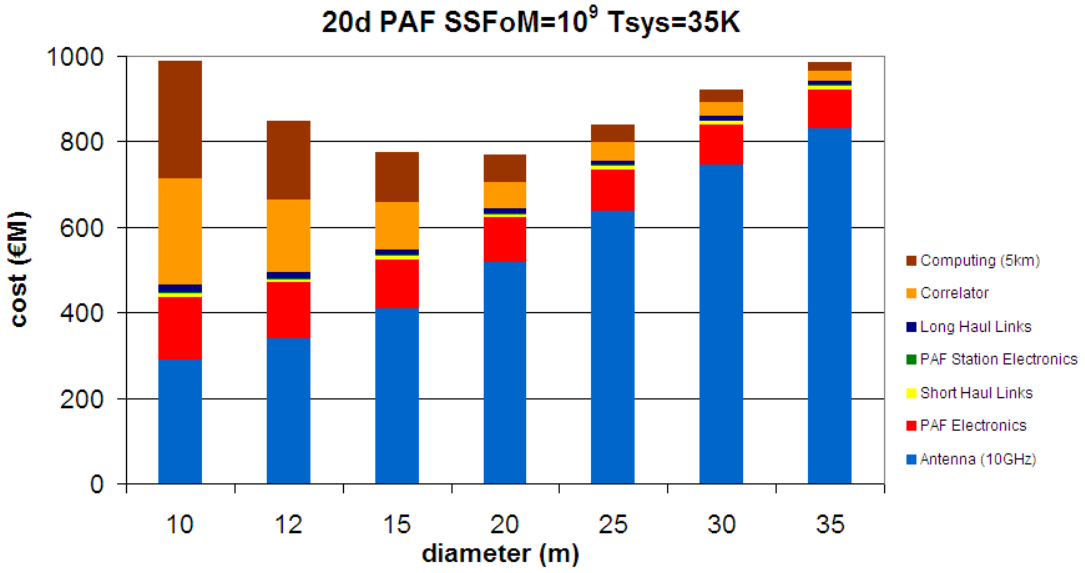


Figure D.1: Cost as a function of dish diameter. Source: Schilizzi et al. (2007), modelled using SKAcost (Memo 92) web interface.

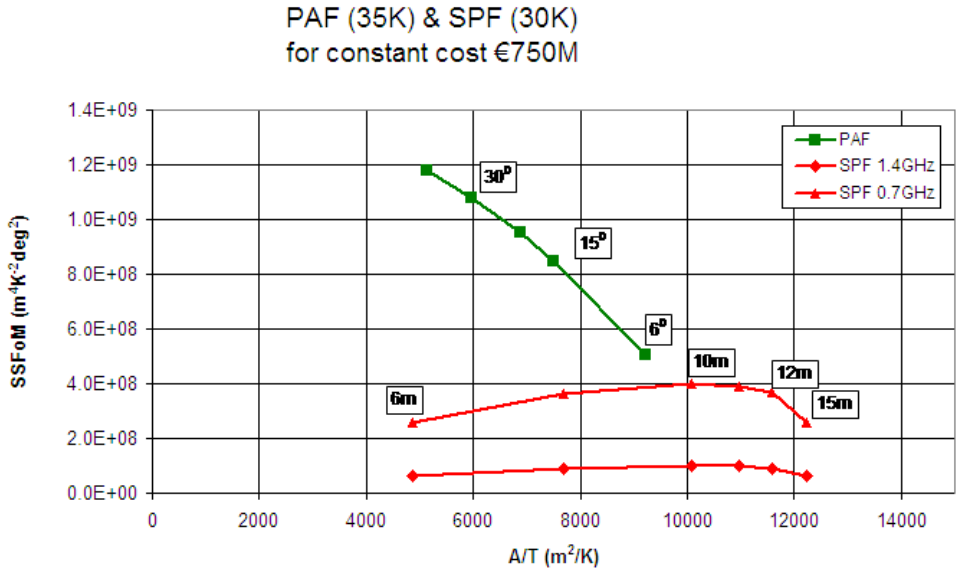


Figure D.2: SSFoM and A/T trade-off for dishes with WBSPFs (at 0.7 and 1.4 GHz) or PAFs, for fixed cost. Source: Schilizzi et al. (2007), modelled using SKAcost (Memo 92) web interface.

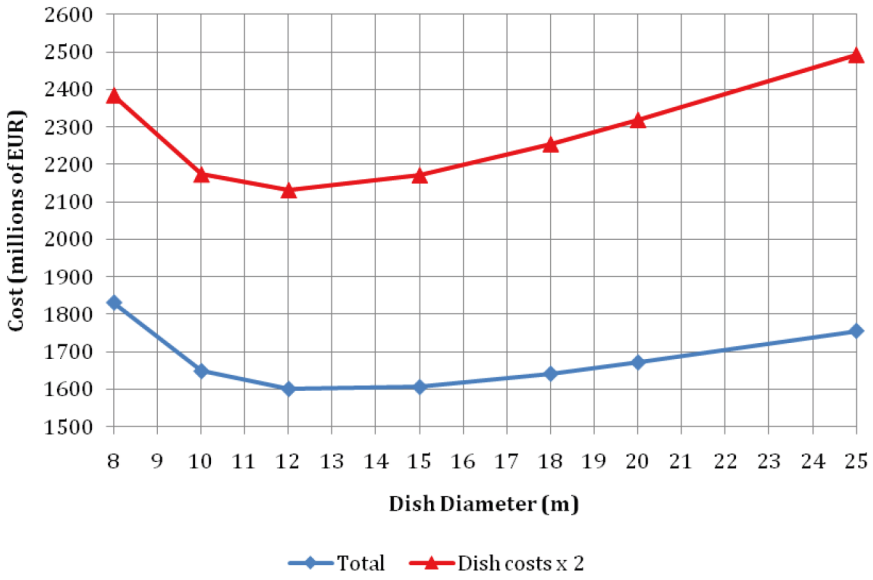


Figure D.3: Total cost as a function of dish diameter, for the default SKA cost estimate, and one where the dish cost is doubled. Source: Faulkner et al. (2010), modelled using SKACost (Memo 120) SKADS benchmark scenario.

D.4.1 Cost impact of a changed sub-system cost

Varying the cost of one or more sub-systems will show how much the cost of these systems drives the total cost. Figure D.3 shows two SKA cost estimates, the only difference between them being a factor of 2 difference in dish costs. This example shows that the cost curve maintains its broad minimum. It also shows that dish cost is a significant cost driver in this design.

D.4.2 Cost impact of changed technical requirements

Changing the requirements, hence telescope performance, will impact the system cost. An example is Figure D.4, which plots cost as a function of data rate output from an AA station. The data rate is a proxy for processed station FoV, and therefore SSFoM. This sort of performance–cost exploration does not usually have an optimal solution, but gives an indication of the incremental cost of certain requirements. For Figure D.4, the "...total cost slope amounts to around €23 million per Tbit/s data rate" (Faulkner et al., 2010). Although not indicated in Faulkner et al. (2010), this could be translated to a cost per deg^2 of processed station FoV.

The result of such investigations can feed back to the requirement analysis (Section 3.3.1).

D.4.3 Cost impact of changed schedule

Delay in the project schedule means that the products are built or purchased at later dates; this impacts cost, especially for digital components. Digital technology advancements creates a time-dependent axis in the SKA design space. Such technological advancements can be generalised by exponential laws; the most common being Moore's law, where the cost of an equivalent digital product halves every 1–2 years, or more often, the performance of an equivalently priced

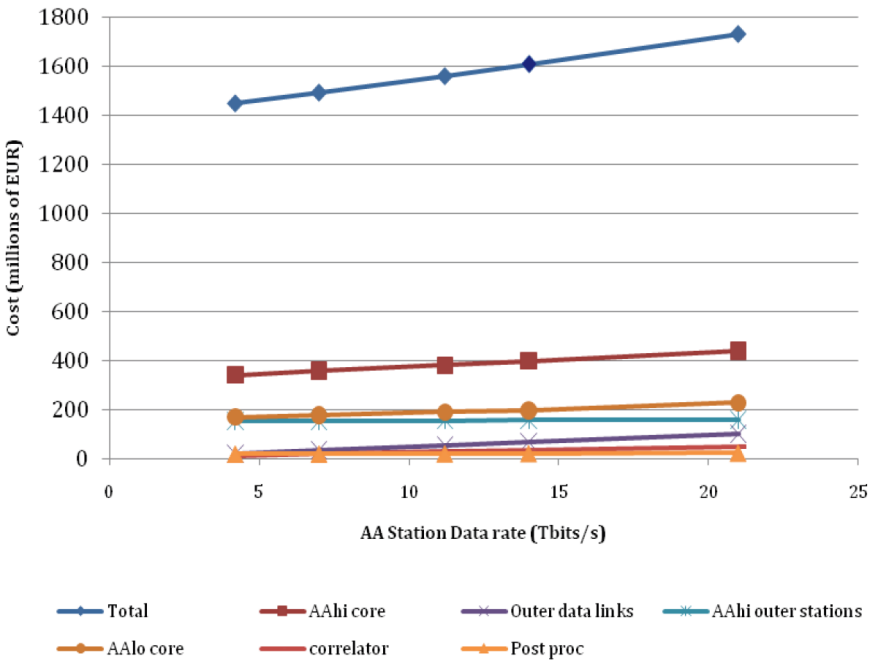


Figure D.4: Cost as as function of data rate output from an AA station. Source: Faulkner et al. (2010), modelled using SKACost (Memo 120) SKADS benchmark scenario.

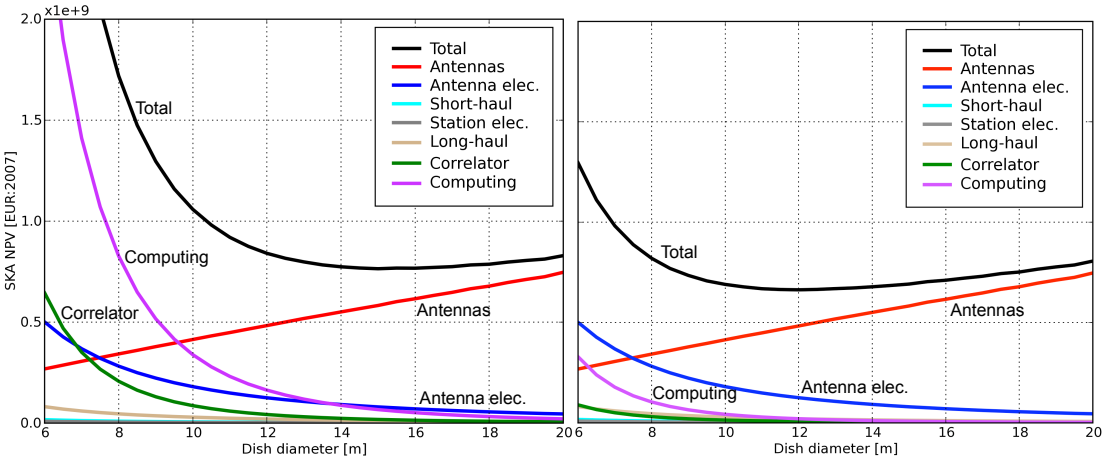


Figure D.5: Net present value cost as a function of dish diameter, for dishes with WBSPFs and constant effective area. Correlator and computer purchase year is 2015 (left) and 2020 (right). Source: Colegate (2009), modelled using SKAcost (Memo 92).

product doubles every 1–2 years (Mollick, 2006). The financial tools in SKACost allow for such scaling laws to be factored in. These scaling laws can have a large impact on optimisation curves, as shown in Figure D.5. In Figure D.5, Moore's law has a cost halving period of 2 years, so a 5 year delay in the purchase of the correlator and computing components results in the shallow minimum of dish diameter moving from around 12–20 m to 10–15 m.

Appendix E

Fast transient searches

This appendix provides additional details for fast transient searches (Chapter 4).

E.1 Signal combination techniques

The performance attributes of a radio telescope array depend on how the signals from the array receptors are combined. This in turn affects the detection rate for fast transients. The array receptors may be an antenna (such as a dish or dipole) or a phased group of antennas (stations). I term the single antenna primary beam or the phased station beam as the receptor beam, with FoV Ω_0; the receptor has effective area $A_{\mathrm{e\text{-}0}}$. The signals from these receptors may then be combined incoherently or coherently as discussed later in this section.

For beams from multiple receptors pointing at the same location on the sky, the signals detected may be combined incoherently or coherently. The following tables show how sensitivity (defined by effective area), FoV, beamformer processing cost and the number of data streams to be searched scale for different signal combination modes. The scaling equations assume that the polarisations are summed prior to searching and all receptors being combined are of equal diameter and sensitivity.

E.1.1 Incoherent combination

Incoherent combination of the receptor signals requires the signal from each receptor to be detected, a geometric delay applied and the signals summed. Figure E.1 shows the steps to incoherently combine signals and Table E.1 shows some performance attributes.

As long as the appropriate geometric delay is applied to the signal at each receptor, incoherently combined receptors do not need to be located close together. For the SKA, this means that while the core is being used for low angular resolution experiments, the mid and long baselines could be used for fast transient searches. Note that incoherent combination cannot account for the geometric delays within the beam, but away from the beam centre.

I assume that digitising, channelising and station beamforming are existing telescope functions and that these do not factor into the additional processing cost of incoherently combining the station beams. Square-law detection involves squaring and summing the real and imaginary components of each channel of each beam from each station. The averager integrates the power samples for each beam, and then equivalent beams from different receptors are summed together, resulting in a total of $N_{\mathrm{b\text{-}0}}$ incoherently combined data streams. The integration of the power samples greatly reduces the data rate of the incoherent beams, although it comes at a cost of a lower time resolution.

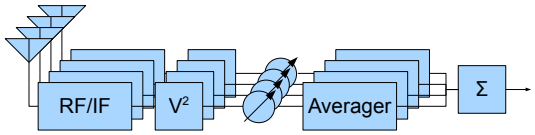


Figure E.1: Incoherent combination. For aperture arrays, station beamforming takes place prior to the V^2 block.

Table E.1: Incoherent combination attributes

Attribute	Scaling	Comment
Sensitivity (A_e)	$\sqrt{N_0}A_{e\text{-}0}$	N_0 receptors, each with effective area $A_{e\text{-}0}$.
FoV (Ω_{proc})	$N_{b\text{-}0}\Omega_0$	$N_{b\text{-}0}$ beams formed per station, each with FoV Ω_0[a].
Processing	$N_0 N_{b\text{-}0}$	
Data streams	$N_{b\text{-}0}$	

[a] $\Omega_0 = \frac{\pi}{4}\left(\frac{c\mathcal{K}_0}{\nu D_0}\right)^2$, where $\mathcal{K}_0$ is the dish illumination or station beam taper and D_0 is the diameter of the dish or station in metres.

Table E.2: Incoherently combined subarray attributes.

Attribute	Scaling	Comment
Sensitivity (A_e)	$\sqrt{N_{0/sa}}A_{e\text{-}0}$	$N_{0/sa}$ receptors per subarray.
FoV (Ω_{proc})	$N_{sa}N_{b\text{-}0}\Omega_0$	N_{sa} independently pointed subarrays.
Processing	$N_0 N_{b\text{-}0}$	
Data streams	$N_{sa}N_{b\text{-}0}$	

E.1.2 Independently pointed subarrays, incoherently combined

A further increase in FoV can be achieved by pointing subarrays of receptors in different directions and incoherently combining the signals from the receptors in the subarray. Table E.2 shows performance attributes for this mode. The processing cost for incoherently combining subarrays is approximately the same as it is for incoherently combining all stations because the same number of beams are square-law detected and summed across stations; the difference is that separate sums are maintained for each subarray. The approximate beamforming operations cost is therefore independent of the number of incoherently combined subarrays (N_{sa}).

The smallest subarray size is one receptor, meaning each receptor is pointing to a unique patch of sky — this is termed fly's eye. In reality, the minimum number of receptors in a subarray would be 3, to allow for a triggered buffer to be used to localise any detected signal.

E.1.3 Coherent combination—array beamforming

For coherent combination, each array beam is formed by the weighted sum of N_0 receptor beams pointing in the same direction. Geometric delays are applied to the signals from the receptors, which are then summed and detected. Like for AA stations, multiple beams can be formed. Figure E.2 shows the steps for coherent array beamforming, and Table E.3 shows the performance attributes. Cordes (2009a) discusses array beamforming in the context of the SKA.

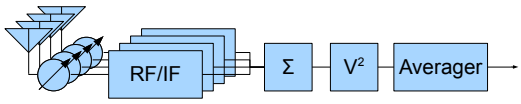


Figure E.2: Coherent combination. For aperture arrays, station beamforming takes place prior to the $\sum$ block.

Table E.3: Coherent combination attributes.

Attribute	Scaling	Comment
Sensitivity (A_e)	$N_0 A_{e\text{-}0}$	Coherent sum of N_0 receptors.
FoV (Ω_{proc})	$N_{b\text{-}arr}\Omega_{arr}$ [a]	$N_{b\text{-}arr}$ array beams formed.
Processing	$N_0 N_{b\text{-}arr}$	
Data streams	$N_{b\text{-}arr}$	
Data streams to pixelise full FoV	$N_{b\text{-}0} N_{pix}$	N_{pix} pixels to fill Ω_0 with array beams (see Equation 4.6).

[a] $\Omega_{arr} = \frac{\pi}{4}\left(\frac{c\mathcal{K}_{arr}}{\nu D_{arr}}\right)^2$, where $\mathcal{K}_{arr}$ is the array beam taper and D_{arr} is the longest baseline in the array.

Table E.4: Correlation beamforming attributes.

Attribute	Scaling	Comment
Sensitivity (A_e)	$N_0 A_{e\text{-}0}$	For large N_0.
FoV (Ω_{proc})	$N_{b\text{-}0}\Omega_0$	
Processing	1	Use existing correlator hardware.
Data streams	$N_{b\text{-}0} N_{pix}$	

E.1.4 Correlation beamforming—'fast imaging'

An alternative method to pixelise the full FoV is to create images from the correlator at a high time resolution. Although the major advantage is that the correlator hardware can be re-used, it should be noted that this means that the data must be able to be dumped from the correlator at these high rates which may place additional requirements on the correlator architecture. There is also a computational cost for gridding and imaging at high time resolution.

E.2 Event rate for a broadened pulse in a volume of sky

Appendix C.2 derived the detection rate $\mathcal{R}_{det}$ for a volume of objects, and a similar approach can be taken for intermittent sources, such as fast transients. The detected event rate is calculated for observable space. I assume extragalactic sources of intrinsic luminosity $\mathcal{L}\,\mathrm{Jy\,pc^2}$ are homogeneously distributed in a sphere of volume V with a nominal intrinsic event rate density of $\rho_i\,\mathrm{events\,s^{-1}\,pc^{-3}}$. The event rate for this sphere is given by $\rho_i V$:

$$\mathcal{R}_{sphere} = \rho_i \frac{4\pi}{3} r^3 \quad \mathrm{events\,s^{-1}}, \tag{E.1}$$

where r is the radius of the sphere in pc. Assuming that no energy is lost due to attenuation, all events within radius

$$r_{\max} = \sqrt{\frac{\mathcal{L}}{4\pi S_{\min}}} \tag{E.2}$$

are detectable.

The intrinsic flux density of a pulse differs from the observed flux density due to pulse broadening (or smearing) effects of the interstellar medium and of the detection system itself. However, following Deneva et al. (2009), if no energy is lost due to attenuation, then pulse 'area' is conserved such that $S_{\rm i}W_{\rm i} = SW$, where $W_{\rm i}$ is the intrinsic width of the pulse, S is the observed flux density and W is the width of the broadened pulse. If the post-detection integration time τ is equal to W, a telescope with minimum detectable flux density $S_{\min,\tau=W}$ can detect the broadened pulse to a maximum distance

$$r_{\max} = \sqrt{\frac{W_{\rm i}\mathcal{L}}{4\pi W S_{\min,\tau=W}}}. \tag{E.3}$$

A more general relationship which uses an integration time of $\tau = W_{\rm i}$ is

$$r_{\max} = \left(\frac{W_{\rm i}}{W}\right)^{1/4}\left(\frac{\mathcal{L}}{4\pi S_{\min}}\right)^{1/2}. \tag{E.4}$$

Pulse broadening factors are discussed further in Appendix E.3.

The nominal extragalactic population is observable out to $r_{\max}$ for the fraction of the sky observed ($\Omega_{\rm proc}/\Omega_{\rm sky}$) and the detected event rate is

$$\begin{aligned}\mathcal{R} &= \frac{4\pi}{3}\rho_i\frac{\Omega_{\rm proc}}{\Omega_{\rm sky}}r_{\max}^3 \\ &= \frac{1}{3}\rho_i\Omega_{\rm proc}\left(\frac{W_{\rm i}}{W}\right)^{3/4}\left(\frac{\mathcal{L}}{4\pi S_{\min}}\right)^{3/2},\end{aligned} \tag{E.5}$$

where $\Omega_{\rm proc}$, $S_{\min}$ and W are functions of how the signals are combined and processed by the telescope system.

Also, the energy from the impulsive event many not be isotropically radiated. If the event occurs via a pulsar-like radio beam, the pulse energy only fills some fraction $\Omega_{\rm beam}/4\pi$ of the celestial sphere. Because the beam geometry is not usually known, pulsar astronomers often refer to a pseudo-luminosity at a given frequency, such that $\mathcal{L} = Sr^2$, where S is the mean flux density (pulse intensity integrated over the pulse period), and r is the distance to the pulsar (Lorimer & Kramer, 2005). For simplicity, calculations in this thesis assume that all sources are isotropic radiators.

E.3 Pulse broadening and correction (dedispersion)

Cordes & McLaughlin (2003) model the broadening of a delta function pulse due to propagation through the interstellar medium and signal processing response times using the following approximation:

$$\Delta t = \sqrt{\Delta t_{\rm DM}^2 + \Delta t_{\delta\rm DM}^2 + \Delta t_{\Delta\nu}^2 + \tau_{\rm d}^2}, \tag{E.6}$$

where $\Delta t_{\rm DM}$ represents the broadening due to dispersion smearing; $\Delta t_{\delta \rm DM}$ is due to the error in the DM, δDM, used by the system's dedispersion signal processing; $\Delta t_{\Delta\nu}$ is the system's filter response time; and $\tau_{\rm d}$ is due to the multipath propagation effects of the medium. Expanding on this, the broadening of a pulse of intrinsic width $W_{\rm i}$ can be modelled as

$$W = \sqrt{W_i^2 + \Delta t_{\rm DM}^2 + \Delta t_{\delta \rm DM}^2 + \Delta t_{\Delta\nu}^2 + \tau_{\rm d}^2}. \tag{E.7}$$

The filter response time, $\Delta t_{\Delta\nu}$, is approximately equal to $\Delta\nu^{-1}$, where $\Delta\nu$ is the bandwidth of the filtered signal. It is important to note that for a fully coherent transient detection system where the receptor beams are coherently combined, coherently dedispersed and searched, $\Delta\nu$ represents the full signal bandwidth; whereas for an incoherent transient detection system in which the signal is channelised, detected and searched, $\Delta\nu$ represents the much smaller channel bandwidth. Consequently, the filter response component of pulse broadening, $\Delta t_{\Delta\nu}$, is significantly higher for incoherent transient detection systems.

Furthermore, while coherent dedispersion techniques can completely correct for dispersion smearing (given that the DM is known), incoherent dedispersion techniques can only correct for dispersion between the filter-bank channels; they cannot correct for dispersion within the channels. Intra-channel dispersion smearing can be reduced by choosing smaller channel bandwidths, but at the expense of larger filter response times. The optimum channel bandwidth for incoherent dedispersion occurs where the dispersion smearing within each channel equals the filter response time (Hankins & Rickett, 1975; Cordes & McLaughlin, 2003). This leads to a minimum pulse width after incoherent dedispersion of

$$W_{\rm inc} = \sqrt{W_i^2 + 2(\Delta t_{\rm DMmin})^2 + \Delta t_{\delta \rm DM}^2 + \tau_{\rm d}^2}, \tag{E.8}$$

where $\Delta t_{\rm DMmin} = \sqrt{8.3 \times 10^{15} DM\nu^{-3}}$. It should be noted that this optimum cannot be realised for all DMs, because it expects the channel bandwidth to be a function of the DM.

For coherent dedispersion the $\Delta t_{\rm DM}$ term is completely removed:

$$W_{\rm coh} = \sqrt{W_i^2 + \Delta t_{\Delta\nu}^2 + \Delta t_{\delta \rm DM}^2 + \tau_{\rm d}^2}, \tag{E.9}$$

where (as noted above) $\Delta\nu$ is the full processed bandwidth and $\Delta t_{\Delta\nu} \approx \Delta\nu^{-1}$.

E.4 SKA$_1$-low frequency dependence

For illustrative purposes, I plot the event rate $\mathcal{R}$ and a breakdown of its frequency-dependent components. Figure E.3 shows $\mathcal{R}$ for source luminosities with spectral indices $\xi = -1.6$ and 0, over a frequency range 70–450 MHz, at 1 MHz steps with processed bandwidth $\Delta\nu = 1$ MHz and normalised to $\mathcal{R} = 1$ at 70 MHz. The slope is steep: at 160 MHz, the event rate is 10% of the event rate at 70 MHz. At 450 MHz, the event rate is 0.016% of the 70 MHz event rate. A further breakdown of $S_{\rm min}$ is shown in Figure E.4.

E.5 Event rate as a function of frequency

The event rate is determined numerically by calculating the S/N ratio (SNR) for each of $N_{\rm ch}$ frequency channels of width $\Delta\nu_{\rm ch}$, weighting it by w, the amount of sky seen with that SNR

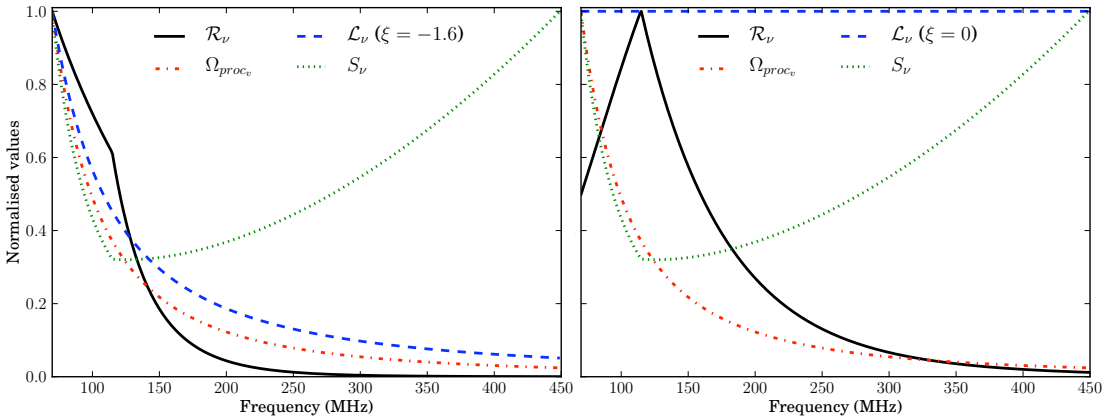


Figure E.3: Event rate $\mathcal{R}$ for $\xi = -1.6$ (left) and $\xi = 0$ (right), and breakdown of the frequency-dependent components comprising $\mathcal{R}$, normalised to the maximum value of each.

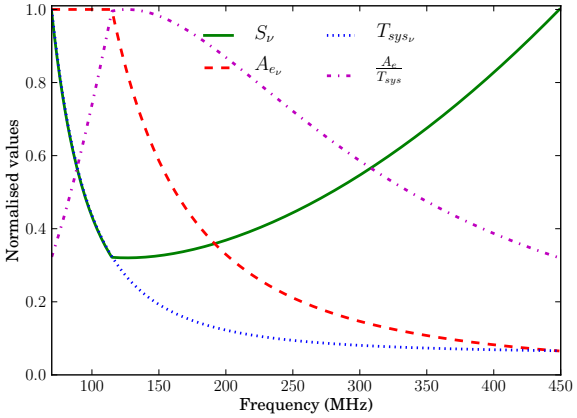


Figure E.4: $S_{\min}$ and breakdown of the frequency-dependent components comprising $S_{\min}$, normalised to the maximum value of each.

and then taking the root of the sum of the squares:

$$\mathcal{R}_{\Delta v} = \sqrt{\sum_i^{N_{\rm ch}} (\mathcal{C} w_i \mathrm{SNR}_i)^2}, \tag{E.10}$$

where $\mathcal{C}$ is a constant. Given $\mathrm{SNR} = \mathcal{L}/4\pi S$,

$$\mathcal{R}^{2/3} = \left(\frac{1}{3}\rho_i \Omega_{\rm proc}\right)^{2/3} \frac{\mathcal{L}}{4\pi S_{\min}}. \tag{E.11}$$

Setting $w = \Omega_{\rm proc}^{2/3}$ and $\mathcal{C} = (\frac{1}{3}\rho_i)^{2/3}$ so $\mathcal{R}^{2/3} = \mathcal{C} \times w \times \mathrm{SNR}$,

$$\begin{aligned} \mathcal{R}_{\Delta v} &= \left(\sum_i^{N_{\rm ch}} \mathcal{R}_i^{4/3}\right)^{3/4} \\ &= \frac{1}{3}\rho_i \left(\frac{\mathcal{L}_0}{\nu_0^\xi}\right)^{3/2} \left(\sum_i^{N_{\rm ch}} \left(\frac{\Omega_{{\rm proc},i}^{2/3} \nu_i^\xi}{S_{{\rm min},i}}\right)^2\right)^{3/4}, \end{aligned} \tag{E.12}$$

for a processed bandwidth of $\Delta\nu = N_{\rm ch}\Delta\nu_{\rm ch}$. Where there is no frequency dependence, the event rate for total bandwidth $\Delta\nu$ and unit bandwidth $\Delta\nu_{\rm ch} = 1$ becomes

$$\mathcal{R}_{\Delta\nu} = \frac{1}{3}\rho_i \Delta\nu^{3/4}\Omega_{\rm proc}\left(\frac{\mathcal{L}}{S_{\rm min}}\right)^{\frac{3}{2}} \tag{E.13}$$

as expected.

E.6 Giant pulse energy distributions

Following Sallmen et al. (1999), the power-law energy probability distribution function of a giant pulse, with a power-law index of $\alpha_{\rm GP}$, a low-energy cut-off $E_{\rm low}$ and no high-energy cut-off, can be given by

$$p(E) = KE^{\alpha_{\rm GP}}, \quad E > E_{\rm low}, \tag{E.14}$$

where

$$K = -\frac{1+\alpha_{\rm GP}}{E_{\rm low}^{1+\alpha_{\rm GP}}}, \quad \alpha_{\rm GP} < -1. \tag{E.15}$$

The complementary cumulative distribution function determines the fraction of detectable pulses:

$$\begin{aligned} {\rm P}(E > E_{\rm min}) &= \int_{E_{\rm min}}^{\infty} p(E)\, dE \\ &= \begin{cases} 1, & E_{\rm min} \leq E_{\rm low}, \\ (E_{\rm min}/E_{\rm low})^{1+\alpha_{\rm GP}}, & E_{\rm min} > E_{\rm low}, \end{cases} \end{aligned} \tag{E.16}$$

where

$$E_{\rm min} = \tau S_{\rm min} \tag{E.17}$$

is the minimum detectable pulse energy, which assumes that $W < \tau/2$, such the pulse's energy is measured by a single integration. Only pulses of energy $E > E_{\rm min}$ will be detectable, therefore more sensitive telescopes, with a lower $E_{\rm min}$, will see a larger fraction of events. When $E_{\rm low} > E_{\rm min}$, all pulses are detectable.

From Equation E.17, the pulse luminosity can be described in terms of pulse energy:

$$\mathcal{L} = \frac{4\pi E r^2}{\tau}, \tag{E.18}$$

where r is the distance to the source. The complementary cumulative distribution function can thus be re-written in terms of luminosity:

$${\rm P}(\mathcal{L} > \mathcal{L}_{\rm min}) = \begin{cases} 1, & \mathcal{L}_{\rm min} \leq \mathcal{L}_{\rm low}, \\ (\mathcal{L}_{\rm min}/\mathcal{L}_{\rm low})^{1+\alpha_{\rm GP}}, & \mathcal{L}_{\rm min} > \mathcal{L}_{\rm low}, \end{cases} \tag{E.19}$$

where the minimum detectable luminosity $\mathcal{L}_{\rm min}$ is

$$\mathcal{L}_{\rm min} = 4\pi S_{\rm min} r^2 \tag{E.20}$$

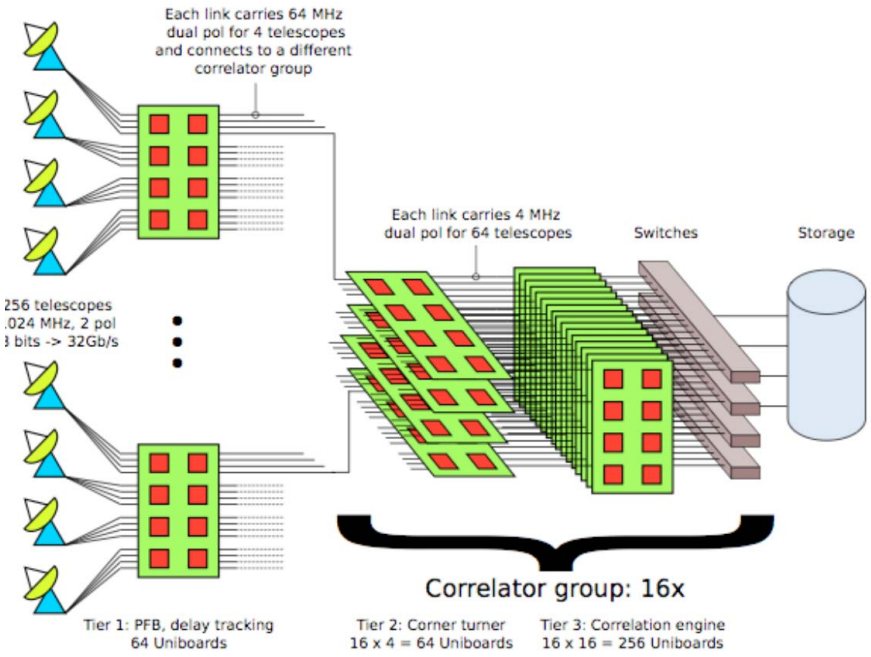


Figure E.5: Uniboard-based SKA_1 dish correlator architecture. Source: Szomoru et al. (2011)

and the low-luminosity cut-off $\mathcal{L}_{\mathrm{low}}$ for giant pulses from the Crab pulsar is

$$\mathcal{L}_{\mathrm{low}} = \frac{4\pi E_{\mathrm{low}} r_{\mathrm{Crab}}^2}{\tau}. \tag{E.21}$$

E.7 Low-cost SKA_1 transient search system details

The cost estimate for the transient search system described in Section 4.7.1 is modelled on a Uniboard-based dish correlator architecture (Figure E.5), which uses 64 Uniboards as filterbanks (each serving 4 dishes), 64 Uniboards for the corner-turn and and 256 Uniboards for the correlation. Szomoru et al. (2011) estimates a cost of €10 k per board for SKA_1 production quantities (excluding memory modules, cabling, switches, racks and power supplies).

E.7.1 Filterbank spigot

A spigot at the filterbank provides incoherent data (squared and integrated voltages), where 1 ms integration time reduces the data rate from 16 Gbps[1] to 16 Mbps[2]. In a CPU-based software FX correlator, the incoherent data can be provided for negligible cost (e.g. Deller et al., 2011). An FPGA-based filterbank may require extra components to output the data; I conservatively allow €1000 per filterbank Uniboard.

E.7.2 Transients processor

The pipeline in Figure 4.2 shows that for single-pixel feed dishes, signals are incoherently combined to produce a single data stream to be dedispersed and searched. However, dedispersion

[1] 2 polarisations × 2 Nyquist × 4 bits × 1 GHz bandwidth
[2] 2 polarisations × 2 Nyquist × 4 bits × 1000 channels / 1 ms integration time

of each data stream prior to incoherent combination and searching improves detection sensitivity and resilience to RFI (Thompson et al., 2011), but there are $N_{\rm dish}$ data streams to dedisperse. A reasonable first-order estimation based on V-FASTR (Wayth et al., 2011) and CRAFT (Macquart et al., 2010a) designs is €1000 per dedispersed and detected data stream, for a system that is capable of searching at 1 ms time resolution, over large DMs ($\gtrsim 4000\,{\rm pc\,cm^{-3}}$) and wide bandwidth ($\gtrsim 256$ channels of $\lesssim 1$ MHz).

E.7.3 Buffer spigot

A rolling voltage buffer is used for post-detection event localisation and very high time resolution pulse analysis. The buffer requires capacity to store the largest delay due to dispersion. This delay is a function of the largest dispersion measure (DM) to be trialled and the minimum and maximum frequencies (Lorimer & Kramer, 2005):

$$\Delta t_{\rm delay} = 4.15{\rm DM}\,(\nu_{\rm min}^{-2} - \nu_{\rm max}^{-2})\,{\rm ms}, \tag{E.22}$$

where $\nu_{\rm min}$ and $\nu_{\rm max}$ are in units of GHz. For example, at the lowest dish frequencies (0.45–0.7 GHz), $\Delta t_{\rm buffer} = 48$ s for a maximum DM of $4\,000\,{\rm pc\,cm^{-3}}$. For 1–2 GHz and the same maximum DM, $\Delta t_{\rm delay} = 12$ s. To account for read and write overheads and the time delay to detect a trigger signal, a 60 second buffer should suffice for the dish frequencies. This requires of order 30 GB RAM per dish[3], or approximately €250 at current prices. A conservative order-of-magnitude estimate is that the buffering and data dump requires an additional Uniboard per 4 dishes.

The transients buffer is significant fraction of the transient processing system cost. However, some filterbank designs incorporate a buffer (e.g. Bunton, 2010); these could potentially be expanded with additional RAM rather than installing additional hardware. Furthermore, a buffer does not necessarily need to be implemented for every dish data stream. The number of dish data streams to be buffered depends on the signal combination mode. The sensitivity of incoherent combination is a factor of $\sqrt{N_{\rm dish}} \approx 16$ less than coherent combination. Therefore, to achieve similar sensitivity for a coherent follow-up of a detection made in incoherent combination mode, the voltages from only $250/16 \approx 16$ dishes are required to be buffered; choosing dishes with longer baselines allows for higher angular resolution localisation.

Some buffering of the data is required regardless of signal combination mode. Although using the 1 km diameter core for coherent combination maximises $\mathcal{R}_{\rm beam^{-1}}$, it does not provide the full localisation capability of the interferometer and is more susceptible to local RFI. If the coherently combined core is used for searches, then data from the longer baselines could be buffered to improve follow-up localisation and imaging capabilities.

E.7.4 Storage

A single trigger of 250 buffers requires up to 7.5 TB of data. At €100 per TB of usable space, storage of 100 events is reasonable. The actual quantity of storage required will depend on the trigger rate and cost.

[3] 60 s × 2 polarisations × 2 Nyquist × 4 bits × 250 MHz bandwidth / 8 bits to a byte

Appendix F

Parametric models and costs for SKA-low

This appendix provides additional SKA-low details (Chapter 5).

F.1 Summary of assumptions

For clarity, the key assumptions made for this analysis are summarised here and page references listed.

- Performance-related assumptions:
  - signals are dual-polarisation ($N_{\mathrm{pol}} = 2$), or full-Stokes (p 105)
  - $20\,\mathrm{deg}^2$ processed field of view is observed concurrently across the 70–450 MHz band (p 105)
  - antenna element gain is the same for the single and dual-band (low and high) antenna elements to ensure that the first-order station A/T estimates are comparable (p 106)
  - the intra-station element layout is an irregular layout of uniform element distribution (p 105)
  - only one tile beam is formed for each tile (p 125).
- Cost-related assumptions:
  - the reference class LOFAR cost estimate is for 50 MHz output bandwidth with 2007 technology, so by 2016 newer technology will allow for the processing of the full 380 MHz bandwidth for the same cost (p 103)
  - the estimated costs do not represent the total cost of building the telescope, because costs are excluded when it can be shown that they remain constant between the single and dual-band implementations (p 104)
  - the low and high-band cores are separate (p 98)
  - low and high-band stations are co-located beyond the core, and the trenching and cables for the data transmission and power to these stations are shared (p 106)
  - infrastructure common to both stations, such as housing for the station processing node, is not shared (p 106).
- Signal processing assumptions:
  - a time to frequency domain transformation and cross-correlation 'FX' correlator is used (p 100)
  - the number of coarse frequency channels and their channel width is constant (p 196)
  - the station beamformer cost scaling is the same for both frequency and time domain beamforming (i.e. the coarse filterbank does not dominate the beamformer cost) (p 196)
  - the SKA_1-low correlator frequency resolution is defined by the HLSD ($\Delta\nu_{\mathrm{ch}} = 1\,\mathrm{kHz}$) rather than D_{st} (p 203)

– the central processing facility sub-systems operate on a 'per beam' basis (p 111)
– the imaging cost is dominated by the data buffer, rather than the processing (p 207).
– only the high-band core of the dual-band array is used in the non-imaging processor, and the processing for the AAs, not the dishes, dominate the cost (p 207).

F.2 Parametric models and costs for SKA_1-low stations

The parametric equations for the station sub-system blocks consist of cost coefficients representing fixed and variable units costs, C_{fix} and C_{var} respectively. The variable costs scale with one or more parameters. The total cost of a particular block in the system is the product of quantity and cost. Table F.1 summarises the quantity and the cost scaling of the blocks in this analysis (see Section 5.3). The detailed block descriptions below provide further justification and references for the cost scaling. Table F.2 lists the fixed and variable unit costs of each block.

For simplicity, this analysis assumes that first-stage coarse channel filterbank (CFB) is implemented at the station bunker, prior to the beamforming; its cost is implicitly included in the station beamformer cost (Appendix F.2.6). The second-stage fine channel filterbank (FFB) is located at the correlator (Appendix F.4.2.2). The all-digital beamforming architecture in Bij de Vaate et al. (2011) integrates the first-stage coarse channel filterbank (CFB) between the digitisation and the tile beamforming blocks shown in Figure 5.12 (p 124). Channels outside processed bandwidth $\Delta\nu = \nu_{\text{max}} - \nu_{\text{min}}$ are discarded after the coarse filterbank, hence are not transported to the station beamformer. Although the data transport requirements are reduced, there are additional infrastructure and power supply implications in having the filterbank at the digitiser, which are beyond the scope of this analysis.

The station cost estimates are for the sub-system hardware costs, where a sub-system is generally described by one of the scalable blocks. These cost estimates include 'sub-system infrastructure', such as the housing for the signal processing units. But the sub-system hardware costs themselves are not just the procurable components (the physical hardware); they also include costs such as non-recurring engineering, assembly and integration and testing for the sub-system. Faulkner et al. (2011) details the cost coverage for the AA CoDR estimates; in general, only the component costs are accounted for. Note that some components may implicitly include non-component hardware costs in the component purchase price. For example, the purchase price of a digitiser board would usually include assembly, integration and testing prior to delivery.

I derive the fixed and variable unit costs from these cost estimates. Recalling Equation 3.1 in Section 3.4.2.1: $C_{\text{block}} = C_{\text{fix}} + C_{\text{var}}\, x$; this parametric equation can be solved if C_{block}, x and the proportion of costs between C_{fix} and C_{var} are known. Say a cost estimate gives $C_{\text{block}} = €200$, for $x = 8$. Using a reasonable estimate of what proportion of cost is fixed (say 20%), the unit costs for the block can be determined: $C_{\text{fix}} = €40$ and $C_{\text{var}} = €20$. The cost of the block can then be determined for other values of x (within reasonable design limitations).

Table F.1: Summary of blocks and scaling for SKA$_1$-low sub-systems.

Block name	Quantity in SKA$_1$	Parametric equation	Block coverage
Active antenna element	$N_{\mathrm{st}} N_{\mathrm{e/st}}$	C_{fix}	mechanical element, LNA, gain and filter, housing and ground plane
Analogue (RF) tile beamformer[a]	$N_{\mathrm{st}} N_{\mathrm{tile/st}}$	$N_{\mathrm{b\,-\,tile}} N_{\mathrm{e/tile}} C_{\mathrm{var}}$	hardware
Element/tile–digitiser RF link	RF: $N_{\mathrm{st}} N_{\mathrm{tile/st}} N_{\mathrm{b\,-\,tile}}$ Dig: $N_{\mathrm{st}} N_{\mathrm{e/st}}$	RF: C_{fixRF} Dig: C_{fixDig}	RF cable
Digitiser	RF: $N_{\mathrm{st}} N_{\mathrm{tile/st}} N_{\mathrm{b\,-\,tile}}$ Dig: $N_{\mathrm{st}} N_{\mathrm{e/st}}$	C_{fix}	digitiser
Digitiser–bunker link [b]	Dig: $N_{\mathrm{st}} N_{\mathrm{e/st}}$	$C_{\mathrm{var1}} R_{\mathrm{dig}} + C_{\mathrm{var2}} \overline{L_{\mathrm{e-st}}}$	C_{var1}: electronics, cable connectors C_{var2}: fibre cable
Station beamformer[c]	N_{st}	RF: $\overline{N_{\mathrm{b\,-\,st}}} N_{\mathrm{tile/st}} C_{\mathrm{var}}$ Dig: $\overline{N_{\mathrm{b\,-\,st}}} N_{\mathrm{e/st}} C_{\mathrm{var}}$	coarse filterbank, station beamformer
Station infrastructure (bunker)	N_{st}	RF: $C_{\mathrm{fix}} + A_{\mathrm{g\,-\,st}} C_{\mathrm{var1}} + \overline{N_{\mathrm{b\,-\,st}}} N_{\mathrm{tile/st}} C_{\mathrm{var2}}$ Dig: $C_{\mathrm{fix}} + A_{\mathrm{g\,-\,st}} C_{\mathrm{var1}} + \overline{N_{\mathrm{b\,-\,st}}} N_{\mathrm{e/st}} C_{\mathrm{var2}}$	C_{fix}: building etc. C_{var1}: preparation, trenching etc. C_{var2}: environmental conditioning, rack space etc.
Station–CPF link transmission	N_{st}	$R_{\mathrm{st}} C_{\mathrm{var}}$	fibre transmission

[a] Optional block. If analogue tile beamforming is included, subsequent quantities and costs are denoted 'RF'. If not, the system is all-digital beamforming, denoted by 'Dig'.
[b] Optional block. Assumes no digital beamforming at the tile. See Section 5.7.3 for alternative architectures.
[c] Approximate cost scaling, see Appendix F.2.6.1.

F.2.1 Active antenna element

The active antenna element describes an integrated system, which includes the mechanical element, LNA, gain and filter, housing and ground plane. I only consider a consolidated unit cost for the active antenna element, because its constituent components are specified and costed for a particular design. For example, the mechanical antenna element is matched to the LNA to minimise the receiver noise across the frequency band (van Ardenne et al., 2009), and the ground plane is designed for a given antenna element. Additionally, the active antenna element design should minimise manufacturing, transportation, deployment and operations costs, as discussed in Faulkner et al. (2011).

Consolidating the active antenna element cost does conceal potential trade-offs within that design space; a trade-off of topical interest is the cost of providing a ground plane for the element. However, a simple analysis in Appendix F.2.1.2 finds that there is not currently the

Table F.2: SKA_1-low sub-system unit costs in € (2007).

Block name	Bottom-up SKADS cost estimate[a]	Reference class LOFAR cost estimate[b]	Block unit[c]
Active antenna element	C_{fix}: 75	C_{fix}: 172	per element
RF tile beamformer	C_{var}: 7.0 per output beam	C_{var}: same as low	per element
Element/tile–digitiser RF link	C_{fixRF}: 81 C_{fixDig}: 18	C_{fixRF}: same as low C_{fixDig}: same as low	per signal
Digitiser	C_{fix}: 12	C_{fix}: 80	per signal
Digitiser–bunker link	C_{var1}: $1\,\text{Gbps}^{-1}$ C_{var2}: $0.01\,\text{Gbps}^{-1}\text{m}^{-1}$	C_{var1}: same as low C_{var2}: same as low	per link
Station beamformer	C_{var} : 0.1 per input per output beam	C_{var} : 2.6 per input per output beam	per station
Station infrastructure	C_{fix}: 28 k C_{var1}: 0 C_{var2}: 0.06 per input per output beam	C_{fix}: 74 k C_{var1}: same as low C_{var2}: 2.6 per input per output beam	per station
Station–CPF link transmission	C_{var}: $100\,\text{Gbps}^{-1}$	C_{var}: same as low	per link

[a] All-digital beamforming, and technology advances. Most costs extrapolated from Table 3 of Faulkner et al. (2011).
[b] LOFAR estimate for 50 MHz output bandwidth and analogue (RF) tile beamforming. Most costs extrapolated from Table 4 of Faulkner et al. (2011). The availability of only one cost estimate is indicated by 'same as low'.
[c] All elements, beams, inputs and signals are dual polarisation.

justification for costing the ground plane separately to the rest of the active antenna element, because the cost of ground plane is more closely linked to the type of antenna element than the areal cost, and the ground plane cost is not significant in the broader SKA_1-low context, given the uncertainties of the present first-order analysis.

F.2.1.1 Active antenna element costs

The cost data sources in Faulkner et al. (2011) provide a consolidated cost for the active antenna element. The cost per dual polarisation active antenna element is €75 for the bottom-up SKADS estimate, the cost being taken from existing arrays. The reference class LOFAR estimate for the same system is €179 and is a direct transfer from the cost of the LOFAR high band (120–240 MHz) antennas. For the purposes of this analysis, the antenna element's RF beamformer cost of €7 per element is subtracted from this cost and listed separately in Table F.2, resulting in a reference class LOFAR element cost of €172.

A comparison of the representative single and dual-band implementations clearly requires es-

timation of the low and high-band active antenna element costs. Factors to consider are:

- The fractional bandwidth of both the low and high-band elements is approximately 2.5:1, as opposed to 6.5:1 for the single-band element.
- The low-band elements, with a maximum frequency $\nu_{\max} = 180\,\mathrm{MHz}$, will require less exacting manufacturing standards than the single and high-band elements.
- The high-band elements, with a minimum frequency $\nu_{\min} = 180\,\mathrm{MHz}$, are physically smaller than the single and low-band elements.
- The average inter-element spacing for the high-band array is 0.75 m, compared with 1.5 m for the single and low-band arrays.

To provide a first-order estimate of the influence of these factors on the low and high-band active antenna element costs, multipliers (discounts) are applied to the single-band element costs (€75 and €172 for the bottom-up SKADS and reference class LOFAR estimates respectively). The cost of the active antenna element is assumed to be split 2:1 between the physical components and the electronics (such as LNAs, filters). The cost of the low-band physical components is estimated at 80 % of the single-band cost. The high-band $\nu_{\min}$ is about 2.5 times the single-band $\nu_{\min}$, hence the cost of the physical components is estimated at 40 % of the single-band cost. The electronics for each band in the dual-band array are assumed to be only marginally cheaper, at 90 % of the single-band cost. The multipliers are thus calculated as

$$\begin{aligned} \text{Low-band} &= \frac{2 \times 0.8 + 0.9}{3} \\ &= 0.83 \end{aligned} \tag{F.1}$$

and

$$\begin{aligned} \text{High-band} &= \frac{2 \times 0.4 + 0.9}{3} \\ &= 0.57. \end{aligned} \tag{F.2}$$

These multipliers are first-order estimates; a detailed analysis, including the applicability to more directive antennas, is an important investigation for future SKA studies.

F.2.1.2 Ground plane costs

The bottom-up SKADS and reference class LOFAR active antenna element estimates include the cost of the ground plane. However, this section makes a simple analysis of ground plane costs to clarify its contribution to the consolidated active antenna element cost. The cost of a wire mesh ground plane will vary with the area of the mesh and the size of the openings in the mesh. One cost estimation method, applicable to both the single and dual-band implementations, is to calculate the total length of wire l_{wire} used to manufacture the mesh and assume that, for large quantities, the cost of the ground plane is linearly proportional to the cost of the wire. For a square mesh ground screen, $l_{\mathrm{wire}} = 2/l_{\mathrm{opening}}$ per m^2 of mesh, where l_{opening} is the opening width (distance between wires). To ensure a radio mirror surface, a rule-of-thumb is that $l_{\mathrm{opening}} \leq 0.1\lambda_{\min}$, where $\lambda_{\min}$ is the minimum wavelength to be observed. The ground plane area for an antenna element will depend on the inter-element spacing and the intra-station element layout. For an irregular layout of approximately uniform element distribution,

Table F.3: Ground plane specifications and cost, for the single and dual-band implementations.

Band	d_{avg} (m)	Antenna element ν_{max} (MHz)	Suitable antenna types	Ground plane opening dimensions (mm)[a]	Ground plane usable ν_{max} (MHz)[b]	Example cost per m^2 (€ 2007)	Example cost per antenna element (€ 2007)[c]
Single	1.5	450	all	50×50	600	6.0	13.5
			LPD	200×200[d]	150	1.5	3.4
Low	1.5	180	all	150×150	200	2.0	4.5
			LPD	200×200[d]	150	1.5	3.4
High	0.75	450	all	50×50	600	6.0	3.4
			LPD	100×100[d]	300	3.0	1.7

[a] Rounded to multiples of 25 mm.
[b] Calculated assuming $l_{\mathrm{opening}} \leq 0.1\lambda_{\mathrm{min}}$.
[c] For a ground plane of area d_{avg}^2. Wire (€0.15 m^{-1}) costs assumed to represent the ground plane cost.
[d] Sizes and usable ν_{max} are an example only and are not based on a particular antenna design.

the ground plane area for each element is approximately d_{avg}^2, where d_{avg} is the average inter-element spacing. Thus, the total length of wire per antenna element is $l_{\mathrm{wire}} = 2d_{\mathrm{avg}}^2/l_{\mathrm{opening}}$.

Table F.3 shows the ground plane specifications and cost for the representative single and dual-band implementations. A ground plane which reflects all frequencies in the band is shown, as is a ground plane suitable for antenna elements such as the more directional log-periodic dipole (LPD), where the higher frequency portion of the element uses the element structure as the ground plane (Braun & van Cappellen, 2006). In that case, the ground plane openings are sized only for the lower frequencies. Example ground plane costs (per m^2 and per antenna element) are given in Table F.3, based on the total length and cost of the wire in galvanised steel mesh. For mesh with 50×50 mm openings, as used for the MWA, $l_{\mathrm{wire}} = 40$ m per m^2 of mesh. Extrapolating actual MWA costs, this equates to a wire cost of €0.15 m^{-1}.

In Table F.3, the type of antenna element is a dominant factor in determining the ground plane costs. This effect can be seen in the cost per m^2 column, where there is a factor of 4 variation. This means that rather than being costed as a distinct component, the ground plane should be considered in conjunction with the antenna element design and included in the consolidated active antenna element cost. Fortunately, the ground plane cost is not significant in the broader context of SKA$_1$-low station hardware and system variable costs. Even the largest cost difference, €10.1 per element between the two single-band ground planes, only equates to a €113 k cost per station. In the context of station hardware costing at least €1.2 M (see Figure 5.11), this is less than 10 % of the hardware cost and is much smaller than the uncertainties described in Section 5.6.3.

An additional factor to consider is that the wire cost, which has been used as a proxy for ground plane cost, will depend on the wire diameter. The following practical requirements will influence the wire diameter:

- Rigidity: ensures a planar surface, within some level of tolerance.
- Deployability: is the ground plane to be deployed as sheets of mesh (larger diameter wire) or longer rolls of mesh (smaller diameter wire)?
- Durability: will the mesh entirely cover the station area, such that it needs to be durable enough to be walked on to enable hardware maintenance?

The wire diameter used for the MWA mesh sheets is approximately 3 mm, but if the requirements are less stringent and the wire diameter can be smaller, then the ground plane costs would further reduce.

F.2.2 Optional: RF tile beamformer

The RF tile beamformer is assumed to be a part of the active antenna element system. Only one tile beam is formed ($N_{\mathrm{b\,-\,tile}} = 1$) for the comparisons made in this document. Multiple independent FoVs require the formation of multiple tile beams.

F.2.3 Element/tile–digitiser RF links

The links listed in Faulkner et al. (2011) are CAT-7 for the all-digital system and co-axial cable for the RF tile beamformed system, although in principle, CAT-7 or co-axial cable could be used in either system. The co-axial cable for the RF tile beamformed system is more expensive because of the longer cable lengths required for that architecture.

F.2.4 Digitiser

The digitiser sample rate R_{sample} and number of bits $N_{\mathrm{bit\,-\,dig}}$ are fixed at 1 GS/s and 8 bits respectively. The digitiser over-samples the data by having a sample rate larger than the maximum frequency.

F.2.5 Optional: digitiser–bunker links

The data from the digitiser is transmitted over fibre to the station processing. This link assumes digitisation occurs at or near the tile or element and the station processing is near the centre of the station. Table 3 of Faulkner et al. (2011) describes a unit cost of €152 for the short optical fibre link. This has the capacity of 120 Gbps per fibre (12×10 Gbps channels). To fully utilise this capacity, a few digitised element or tile beam signals could be transmitted on each link. The cost is composed of transmit (at antenna) and receive (at station) units, connectors and the fibre. A simplistic cost breakdown is a cost of €120 for the electronics and connectors and €1.0 m^{-1} for the fibre, which approximates to $C_{\mathrm{var1}} = €1\,\mathrm{Gbps}^{-1}$ and $C_{\mathrm{var2}} = €0.01\,\mathrm{Gbps}^{-1}\mathrm{m}^{-1}$ respectively.

The parameter $\overline{L_{\mathrm{e\,-\,st}}}$ is the average link length between the element or tile and the station processing. A simplified calculation (average radius to a circle centre) applies to an irregular layout of uniform element distribution, such that $\overline{L_{\mathrm{e\,-\,st}}} \approx D_{\mathrm{st}}/3$.

The data rate out of the digitiser R_{dig} is given by

$$R_{\mathrm{dig}} = N_{\mathrm{pol}} N_{\mathrm{bit\,-\,dig}} R_{\mathrm{sample}} OH \tag{F.3}$$

where OH is the overhead, assumed to be 1.25 for digital encoding.

F.2.6 Station beamformer (including coarse channel filterbank)

Station beamforming is required for the SKA to reduce the number of inputs to the correlator. Digital beamforming can be done in the time domain, or in the frequency domain on the channelised signal. The computational cost of the frequency and time domain beamforming approaches is discussed in Barott et al. (2011) and Khlebnikov et al. (2010), where computation costs are expressed as a function of the number of beams, input antennas and frequency channels, and other costs to implement a time delay (where necessary) and the FFT. These latter costs are architecture specific, as is the cost scaling with the number of channels. Because any scalable digital signal processing description (e.g Bunton, 2010) requires a specific architecture with implicit assumptions, this first-order analysis keeps the number of coarse frequency channels and the channel width constant. Jones et al. (2011) discuss beamforming architectures in more detail.

The station beamformer cost is approximated using Equation F.6 below, and is calculated as cost per input per output beam. This cost is extrapolated from Faulkner et al. (2011); 11 264 inputs and 160 output beams (averaged over the band) is assumed to make the extrapolation. The bottom-up SKADS station beamformer cost already takes into account the processing discount from the two-stage beamforming. The reference class LOFAR cost in Table 4 of Bij de Vaate et al. (2011) is for only 50 MHz bandwidth, however it is also for 2007 technology. I assume that newer technology will allow for the beamforming of the full 380 MHz bandwidth for the same cost.

F.2.6.1 Computational cost of frequency and time domain beamforming

From Barott et al. (2011) and Khlebnikov et al. (2010), for a given architecture and number of channels, the frequency and time domain station beamformer processing load can be respectively simplified to

$$P_{\mathrm{BF}[\nu]} \propto N_{\mathrm{inputs}}(K_{\mathrm{CFB}} + K_{\mathrm{BF}[\nu]}N_{\mathrm{b\,-\,st}}) \tag{F.4}$$

and

$$P_{\mathrm{BF[t]}} \propto N_{\mathrm{b\,-\,st}}(K_{\mathrm{BF[t]}}N_{\mathrm{inputs}} + K_{\mathrm{CFB}}), \tag{F.5}$$

where N_{inputs} is the number of elements or tiles being beamformed, $N_{\mathrm{b\,-\,st}}$ is the number of station beams formed, the constant K_{CFB} is the coarse filterbank cost and $K_{\mathrm{BF}[\nu]}$ and $K_{\mathrm{BF[t]}}$ are the beamforming costs.

If the CFB cost does not dominate (i.e. $N_{\mathrm{b\,-\,st}} \gg K_{\mathrm{CFB}}/K_{\mathrm{BF}[\nu]}$ and $N_{\mathrm{inputs}} \gg K_{\mathrm{CFB}}/K_{\mathrm{BF[t]}}$), the cost scaling is the same for both frequency and time domain beamforming. The processing cost of the station beamformer is thus approximated by

$$P_{\mathrm{BF}} \propto N_{\mathrm{inputs}}N_{\mathrm{b\,-\,st}}. \tag{F.6}$$

In this analysis, N_{inputs} is either $N_{\mathrm{e/st}}$ or $N_{\mathrm{tile/st}}$ and $N_{\mathrm{b\,-\,st}}$ is $\overline{N_{\mathrm{b\,-\,st}}}$.

This approximation is used to derive the unit costs from the cost data sources. In deriving the station beamformer cost from the reference class LOFAR estimate, a factor of 16 fewer inputs (equal to the number of elements per tile) is used relative to the bottom-up SKADS station beamformer cost estimate, which is calculated from the aggregate of the first-stage and station processing costs. This results in a factor of 26 difference in station beam unit cost (Table F.2). The same scaling approximation is applied to 80 % of the station bunker cost (see Appendix F.2.7), hence the factor of 43 difference in the variable unit cost for that sub-system.

F.2.6.2 Hierarchical beamforming

Hierarchical digital beamforming reduces the data transport and processing load on the system. A simple example of this is shown here. Faulkner et al. (2010) presents a two-stage digital beamforming approach and this is reflected in Table 3 of Faulkner et al. (2011), where the first stage consists of a tile of 256 antenna elements as inputs ($N_{\mathrm{inputs\,-\,tile}} = 256$). There are 44 tiles in a station ($N_{\mathrm{tile/st}} = 44$), so one beam from each of the 44 tiles (all pointing in the same direction) are input into the second-stage (station) beamformer ($N_{\mathrm{inputs\,-\,stn}}$ =44). The total station processing load to form $N_{\mathrm{b\,-\,st}}$ station beams (from Equation F.6) is approximately:

$$P_{\mathrm{BF}} \propto N_{\mathrm{tile/st}} N_{\mathrm{inputs\,-\,tile}} N_{\mathrm{b\,-\,tile}} + N_{\mathrm{inputs\,-\,stn}} N_{\mathrm{b\,-\,st}}. \tag{F.7}$$

Given $N_{\mathrm{e/st}} = N_{\mathrm{tile/st}} N_{\mathrm{inputs\,-\,tile}}$ and $N_{\mathrm{inputs\,-\,stn}} = N_{\mathrm{tile/st}}$, this becomes

$$P_{\mathrm{BF}} \propto N_{\mathrm{e/st}} N_{\mathrm{b\,-\,tile}} + N_{\mathrm{tile/st}} N_{\mathrm{b\,-\,st}}. \tag{F.8}$$

For this example, $P_{\mathrm{BF}} \propto 11\,264 N_{\mathrm{b\,-\,tile}} + 44 N_{\mathrm{b\,-\,st}}$. In comparison, for a single stage of beamforming, where $N_{\mathrm{inputs}} = 11\,264$, $P_{\mathrm{BF}} \propto 11\,264 N_{\mathrm{b\,-\,st}}$. Hence when $N_{\mathrm{b\,-\,st}} > 1$, the two-stage beamforming reduces processing costs. However, caution should be used to ensure that the assumptions for Equation F.6 still hold.

F.2.6.3 Trading $N_{\mathrm{e/st}}$ for N_{st}

From Equation F.6, the total processing cost for N_{st} single-stage station beamformers is approximately

$$P_{\mathrm{BF-total}} \propto N_{\mathrm{st}} N_{\mathrm{e/st}} \overline{N_{\mathrm{b\,-\,st}}}. \tag{F.9}$$

Given $N_{\mathrm{st}} \propto 1/(N_{\mathrm{e/st}} F F_{\mathrm{st}} d_{\mathrm{avg}}^2)$ for constant $A_{\mathrm{e\,-\,arr}}$ (Equation F.51) and $\overline{N_{\mathrm{b\,-\,st}}} \propto N_{\mathrm{e/st}} d_{\mathrm{avg}}^2$ (Equation F.22),

$$P_{\mathrm{BF-total}} \propto \frac{N_{\mathrm{e/st}}}{F F_{\mathrm{st}}}, \tag{F.10}$$

where FF_{st} is the frequency-dependent station filling factor (Equation F.49). If d_{avg} and the antenna element gain G_{e} do not change, then the function FF_{st} remains constant and

$$P_{\mathrm{BF-total}} \propto N_{\mathrm{e/st}}. \tag{F.11}$$

As long as the assumptions for the approximation (Equation F.6) still hold, the scaling relationship can also be applied to hierarchical beamforming.

As shown in Appendix F.3.2, the ratio of the average number of beams between single and dual-band implementations $\overline{N_{\mathrm{b\,-\,st}}}(\mathrm{dual} : \mathrm{single})$ is independent of N_{st} for constant $A_{\mathrm{e\,-\,arr}}$. Because

$P_{\mathrm{BF-total}} \propto \overline{N_{\mathrm{b\,-\,st}}}$, the beamformer processing cost ratio between the single and dual-band implementations is independent of station diameter, although the absolute (euro) cost difference is less for the smaller stations.

F.2.7 Station infrastructure (bunker)

Within a station array, it is assumed that the antenna elements are closely packed, hence cables (power and fibre) would be laid as part of the station construction, rather than individual trenches being dug. Thus these costs scale with area. However, this areal infrastructure cost is difficult to estimate with much accuracy until after the site selection. There are also costs for a controlled environment housing at each station for the processing hardware. This cost would increase linearly with the amount of processing, although there will be a fixed cost for the housing. These costs are extrapolated from Faulkner et al. (2011) to obtain a cost per input signal per output beam and a fixed cost. I estimate the zeroth-order breakdown of costs between the variable and fixed costs to be 80% and 20% respectively.

F.2.8 Station–CPF link transmission

The output data from the station beamformer is transmitted over fibre to the central processing facility. This analysis only considers the transmission costs; the per metre trenching and cabling cost is ignored because the layout (configuration) of the SKA_1-low stations do not change significantly between representative systems. The layouts shown in Bolton et al. (2011) have a compact core and spiral arms. The spiral arms may require trenching for the data links, but for the compact core this can be absorbed into the areal infrastructure cost. For the HLSD, 97% of the network infrastructure costs are in the trench network (McCool, 2011a). This means that although a higher data rate may require more strands of fibre, the cost of the fibre cable is not significant, being less than 3%.

For the transmission costs, there are technology steps, where more expensive transmission equipment is required for longer distance. An estimate of dense wavelength division multiplexing (DWDM) transmission system costs per channel are derived from the SKA Design Studies costing work in SKACost (Bolton et al., 2009a). Although there is some variation with distance and the discrete cost steps of DWDM transmission, for this first order costing it is sufficient to estimate an average cost of €1 k per 10 Gbps channel for all links, or €100 per Gbps.

The total data rate out of a station beamformer is given by

$$R_{\mathrm{st}} = \overline{N_{\mathrm{b\,-\,st}}} N_{\mathrm{pol}} N_{\mathrm{bit\,-\,CFB}} \Delta\nu \times OS \times OH \times \mathrm{Nyq.}, \tag{F.12}$$

where $\overline{N_{\mathrm{b\,-\,st}}}$ is the average number of station beams formed over the band. The number of bits out of the coarse channel filter-bank $N_{\mathrm{bit\,-\,CFB}} = 4$, oversampling $OS = 1.1$, digital encoding overhead $OH = 1.25$ and the data is Nyquist sampled: Nyq. = 2.

F.2.9 Dual-band station costs

The dual-band implementation is costed as a separate a low-band (70–180 MHz) and high-band (180–450 MHz) array. For each station block (sub-system), its cost will either remain the same,

Table F.4: Cost multiplier estimates for the dual-band array, where each multiplier is a fraction of the single-band unit cost shown in Table F.2.

Block name	Low-band	High-band	Comment
Active antenna element[a]	C_{fix}: 0.83	C_{fix}: 0.57	See Appendix F.2.1.
Optional: RF tile beamformer	C_{var}: 1	C_{var}: 1	
Element/tile–digitiser RF link	C_{fixRF}: 1 C_{fixDig}: 1	C_{fixRF}: 0.3 C_{fixDig}: 0.3	Closer spacing means smaller tiles and stations.
Digitiser[a]	C_{fix}: 0.4	C_{fix}: 1	Sampling speed is set by ν_{max} of each band (0.4 GS/s for the low-band). Assumes that cost is linearly proportional to sampling speed.
Digitiser–bunker link	C_{var1}: 1 C_{var2}: 1	C_{var1}: 1 C_{var2}: 1	C_{var1}: Per unit of data. C_{var2}: Per unit of data per unit length.
Station beamformer[a]	C_{var} : 0.29	C_{var} : 0.71	Assumes cost is linearly proportional to processed bandwidth.
Station infrastructure[a]	C_{fix}: 1 C_{var1}: 1 C_{var2}: 0.29	C_{fix}: 1 C_{var1}: 1 C_{var2}: 0.71	Processing infrastructure is the same fraction as the station beamformer.
Station–CPF link transmission	C_{var}: 1	C_{var}: 1	Per unit of data.

[a] These values are for a dual-band split frequency of 180 MHz; they will vary for other split frequencies and overlapping bands.

or be some fraction of the single-band cost, depending on how each block is modelled. This 'cost multiplier' is then applied to the costs in Table F.2. Table F.4 shows the cost multipliers used for each block in the low and high-band arrays. The cost multipliers chosen are reasonable approximations; this is an area that requires expert attention to verify these numbers.

F.3 Constant FoV as a function of frequency

F.3.1 Single-band implementation

If a constant FoV as a function of frequency is a requirement, then determining the average number of beams required over the receptor bandwidth can simplify calculations (Alexander et al., 2009). The number of beams as a function of frequency can be given by

$$N_{\text{b-st}}(\nu) = N_{\text{b-st}}(\nu_0)\left(\frac{\nu}{\nu_0}\right)^2, \tag{F.13}$$

where $N_{\text{b-st}}(\nu_0)$ is the number of dual-polarisation beams required at frequency ν_0, calculated as

$$N_{\text{b-st}}(\nu_0) = \Omega_{\text{req}}/\Omega_{\text{st}}(\nu_0). \tag{F.14}$$

Integrating Equation F.13 over the processed bandwidth $\Delta\nu = \nu_{\max} - \nu_{\min}$ gives the number of beams of unit bandwidth:

$$\begin{aligned} N_{\mathrm{b-st\_Hz}} &= \int_{\nu_{\min}}^{\nu_{\max}} N_{\mathrm{b\,-\,st}}(\nu_0)\left(\frac{\nu}{\nu_0}\right)^2 dv \\ &= \frac{N_{\mathrm{b\,-\,st}}(\nu_0)}{\nu_0^2}\left[\frac{\nu^3}{3}\right]_{\nu_{\min}}^{\nu_{\max}} \\ &= \frac{N_{\mathrm{b\,-\,st}}(\nu_0)(\nu_{\max}^3 - \nu_{\min}^3)}{3\nu_0^2}. \end{aligned} \tag{F.15}$$

The average number of dual-polarisation beams over the band is

$$\overline{N_{\mathrm{b\,-\,st}}} = \frac{N_{\mathrm{b-st\_Hz}}}{\nu_{\max} - \nu_{\min}}. \tag{F.16}$$

Also, given $\Omega_{\mathrm{st}} \propto {D_{\mathrm{st}}}^{-2}$ and $N_{\mathrm{b\,-\,st}} \propto \Omega_{\mathrm{st}}^{-1}$, substitution into Equation F.16 shows that

$$\overline{N_{\mathrm{b\,-\,st}}} \propto D_{\mathrm{st}}^2. \tag{F.17}$$

An actual implementation requires that a discrete number of beams be formed at each frequency channel. This introduces some error, because enough beams need to be formed for every frequency channel such that the FoV requirement is always met; i.e. $N_{\mathrm{b\,-\,st}}(\nu)\Omega_{\mathrm{st}}(\nu) \geq \Omega_{\mathrm{req}}$. An actual calculation requires a summation to replace the integral, where the ceiling $\lceil N_{\mathrm{b\,-\,st}}(\nu) \rceil$ is taken for each frequency channel (strictly speaking, $N_{\mathrm{b\,-\,st}}(\nu)$ should to be calculated at the maximum frequency of each channel, not the centre frequency). Although the error is larger at the lower frequencies where Ω_{st} is larger, even at 70 MHz the error is $<$10 % for $\Omega_{\mathrm{req}} = 20\,\mathrm{deg}^2$ (higher error for smaller FoV).

F.3.2 Dual-band implementation

The dual-band implementation is more complex. Equation F.16 can be applied separately to each band in the dual-band array, resulting in the average number of beams in each of the low ($\overline{N_{\mathrm{b\,-\,st[L]}}}$) and high ($\overline{N_{\mathrm{b\,-\,st[H]}}}$) bands. The average number of beams over the full band is given by

$$\overline{N_{\mathrm{b\,-\,st[D]}}} = \frac{(\nu_{\mathrm{split}} - \nu_{\min})\overline{N_{\mathrm{b\,-\,st[L]}}} + (\nu_{\max} - \nu_{\mathrm{split}})\overline{N_{\mathrm{b\,-\,st[H]}}}}{\nu_{\max} - \nu_{\min}}, \tag{F.18}$$

where ν_{split} is the frequency split between the bands and $\nu_{\min}$ and $\nu_{\max}$ are the minimum and maximum frequencies of the dual-band implementation (i.e. 70 and 450 MHz). From Equation F.16,

$$\overline{N_{\mathrm{b\,-\,st[L]}}} = \frac{N_{\mathrm{b\,-\,st[L]}}(\nu_0)(\nu_{\mathrm{split}}^3 - \nu_{\min}^3)}{3\nu_0^2(\nu_{\mathrm{split}} - \nu_{\min})}, \tag{F.19}$$

and an equivalent substitution can be made for $\overline{N_{\mathrm{b\,-\,st[H]}}}$.

The ratio of the average number of beams across the band between the dual and single-band implementations, given by $\overline{N_{\mathrm{b\,-\,st}}}(\mathrm{dual} : \mathrm{single}) = \overline{N_{\mathrm{b\,-\,st[D]}}}/\overline{N_{\mathrm{b\,-\,st[S]}}}$, is a useful metric to compare the data rate from the station and also the processing costs in the central processing facility.

Substitution gives

$$\overline{N_{\text{b-st}}}(\text{dual}:\text{single}) = \frac{N_{\text{b-st[L]}}(\nu_0)(\nu^3_{\text{split}} - \nu^3_{\text{min}}) + N_{\text{b-st[H]}}(\nu_0)(\nu^3_{\text{max}} - \nu^3_{\text{split}})}{N_{\text{b-st[S]}}(\nu_0)(\nu^3_{\text{max}} - \nu^3_{\text{min}})}, \tag{F.20}$$

or

$$\overline{\text{N}_{\text{b-st}}}(\text{dual}:\text{single}) = \frac{\text{D}^2_{\text{st[L]}}(\nu^3_{\text{split}} - \nu^3_{\text{min}}) + \text{D}^2_{\text{st[H]}}(\nu^3_{\text{max}} - \nu^3_{\text{split}})}{\text{D}^2_{\text{st[S]}}(\nu^3_{\text{max}} - \nu^3_{\text{min}})}, \tag{F.21}$$

where the L, H and S sub-scripts indicate low, high and single-band arrays respectively. Thus for a constant FoV across the processed bandwidth $\Delta\nu$, the ratio depends on both the station diameter of each band and the frequency split between the low and high-band arrays.

F.3.3 Trading $N_{\text{e/st}}$ for N_{st}

As shown in Appendix F.5.3, the station diameter is both a function of the number of elements in the station and the average inter-element spacing. To make a distinction between these effects, Equations F.17 and F.21 requires the substitution of $D_{\text{st}} \propto \sqrt{N_{\text{e/st}}} d_{\text{avg}}$ (Equation F.53), such that

$$\overline{N_{\text{b-st}}} \propto N_{\text{e/st}} d^2_{\text{avg}} \tag{F.22}$$

and

$$\overline{N_{\text{b-st}}}(\text{dual}:\text{single}) = \frac{N_{\text{e/st[L]}} d^2_{\text{avg[L]}}(\nu^3_{\text{split}} - \nu^3_{\text{min}}) + N_{\text{e/st[H]}} d^2_{\text{avg[H]}}(\nu^3_{\text{max}} - \nu^3_{\text{split}})}{N_{\text{e/st[S]}} d^2_{\text{avg[S]}}(\nu^3_{\text{max}} - \nu^3_{\text{min}})}. \tag{F.23}$$

For the representative implementations considered in this analysis, $N_{\text{e/st[L]}} = N_{\text{e/st[H]}} = N_{\text{e/st[S]}}$, thus $N_{\text{e/st}}$ is constant as a function of frequency, resulting in

$$\overline{N_{\text{b-st}}}(\text{dual}:\text{single}) = \frac{d^2_{\text{avg[L]}}(\nu^3_{\text{split}} - \nu^3_{\text{min}}) + d^2_{\text{avg[H]}}(\nu^3_{\text{max}} - \nu^3_{\text{split}})}{d^2_{\text{avg[S]}}(\nu^3_{\text{max}} - \nu^3_{\text{min}})}. \tag{F.24}$$

So if D_{st} is traded for N_{st} (for fixed $A_{\text{e-arr}}$), this ratio still holds as long as the same trade is made for all bands. For example, if D_{st} is halved, then $N_{\text{e/st}}$ decreases by a factor of 4. But given $N_{\text{e/st[L]}} = N_{\text{e/st[H]}} = N_{\text{e/st[S]}}$, $\overline{N_{\text{b-st}}}(\text{dual}:\text{single})$ does not change.

F.4 Parametric models and costs for other SKA sub-systems

The presentation of the models and costs in this section are, in general, similar to Appendix F.2. However, only zeroth-order cost estimates are given, with the purpose of illustrating system-level costs (Section 5.5). Table F.5 summarises the quantity and the cost scaling of the blocks which are not SKA_1-low sub-systems, while the detailed block descriptions provide further justification and references for the cost scaling. Table F.6 lists the fixed and variable unit costs of each block.

F.4.1 Site-related costs

A comparison of single and dual-band implementations can be made by only considering the costs that vary between implementations, while excluding costs which remain fixed. For example, if an activity such as site preparation is being undertaken, the fixed costs of that activity

Table F.5: Summary of blocks and scaling for other sub-systems relevant to SKA_1-low.

Block name	Quantity in SKA_1	Parametric equation	Block coverage
Antenna element deployment	$N_{\mathrm{st}}N_{\mathrm{e/st}}$	C_{fix}	deployment
Site preparation	$N_{\mathrm{st}}N_{\mathrm{e/st}}$	$d^2_{\mathrm{avg}}C_{\mathrm{var}}$	site preparation
Correlator	1	$\overline{N_{\mathrm{b\,-\,st}}}N^{\alpha}_{\mathrm{st}}C_{\mathrm{var}}$; $1 < \alpha < 2$	see Appendix F.4.2.2
Correlator–computing data transport	1	$R_{\mathrm{corr\,-\,out}}C_{\mathrm{var}}$	fibre transmission
Imaging processor	1	$R_{\mathrm{corr\,-\,out}}C_{\mathrm{var}}$	
Non-imaging processor	1	see Appendix F.4.2.5	

could include contractor mobilisation and demobilisation, whereas the variable costs are one or more pro rata (such as per hour or per m^2) costs of undertaking the activity.

Some site-related costs are independent of whether the implementation is single or dual-band. The first-order array infrastructure costs discussed in the HLSD, such as road networks and power and fibre reticulation, will generally be independent of the number of stations or their exact location and are not considered here. In the core region, the spatial density of stations will be high enough that any infrastructure work can be incorporated into the site preparation costs. For the inner and mid region stations, located beyond the core, the stations are placed in groups (clusters) on spiral arms (Bolton et al., 2011). The array infrastructure requirements to connect cluster will be similar between the single and dual-band implementations, regardless of the number of stations at each cluster. For the central processing facility buildings, there may be some cost scaling with the amount of processing.

F.4.1.1 Antenna element deployment

Faulkner et al. (2011) costs deployment of the antenna elements at less than €50 per element. This excludes the deployment and connection of the rest of the infrastructure, such as tile or station processing nodes. To show the sensitivity of the single versus dual-band comparison to changes in this cost, two deployment costs are considered: €50 and €100 per element.

F.4.1.2 Site preparation cost

There is no published data on site preparation costs, and it is likely to be highly dependent on what activities the site preparation involves. For example, is it a simple land clearing activity, or are earthworks and trenching required? To show the sensitivity of the single versus dual-band comparison to changes in this cost, two areal site preparation costs are considered: €10 m^{-2} and €100 m^{-2}.

F.4.2 Central processing facility sub-systems

F.4.2.1 Correlator frequency resolution and integration time requirements

Two parameters which are are relevant to the correlator and image processor design are the correlator frequency resolution (channel width) $\Delta\nu_{\mathrm{ch}}$ and integration time Δt; these need to be

Table F.6: Unit costs for other sub-systems relevant to SKA_1-low (€2007).

Block name	Cost estimate A	Cost estimate B	Block unit
Antenna element deployment	$C_{\rm fix}$: 50	$C_{\rm fix}$: 100	per element
Site preparation	$C_{\rm var}$: $10\,{\rm m}^{-2}$	$C_{\rm var}$: $100\,{\rm m}^{-2}$	per element
Correlator	$C_{\rm var}$: 40 k per input beam	$C_{\rm var}$: same as A	per correlator
Correlator–computing data transport	$C_{\rm var}$: $200\,{\rm Gbps}^{-1}$	$C_{\rm var}$: same as A	
Imaging processor	20 M for single-band implementation	same as A	N/A
Non-imaging processor	30 M for single-band implementation	same as A	N/A

small enough to respectively keep the radial and circumferential smearing below some acceptable threshold (Thompson et al., 2001). The frequency resolution and integration time required is inversely proportional to antenna or station beamwidth (Turner et al., 2011). If the maximum baseline (distance between antenna pairs) and station beam taper $\mathcal{K}_{\rm st}$ are constant, then

$$\Delta\nu_{\rm ch} \propto D_{\rm st} \tag{F.25}$$

and

$$\Delta t \propto D_{\rm st}. \tag{F.26}$$

However, SKA_1-low has a more stringent requirement on frequency resolution which makes it independent of station size. The HLSD (Table 4) specifies a scientifically derived requirement of $\Delta\nu_{\rm ch} = 1\,{\rm kHz}$ for AAs, while Turner et al. (2011) calculates $\Delta\nu_{\rm ch} = 590\,{\rm kHz}$ to meet the 2% smearing requirement with a 180 m diameter station. Given the much more stringent specification in the HLSD, any change in $D_{\rm st}$ for SKA-low will not affect $\Delta\nu_{\rm ch}$.

F.4.2.2 Fine channelisation and correlation

The 'FX' correlator cost scalings can be understood by analysing the data streams flowing through the fine filterbank (channelisation) and cross-correlation sub-systems, described in Section 2.3.3. The station beamformer outputs coarsely channelised station beams, where each channel has width $\Delta\nu_{\rm ch\,-\,CFB}$. A single data stream from a station will contain one coarse channel from one station beam. Each data stream is input into a fine filterbank (FFB) and split into channels of smaller frequency resolution. The total processing cost of the FFBs is thus

$$P_{\rm FFB} = N_{\rm ch/CFB}\overline{N_{\rm b\,-\,st}}N_{\rm st}K_{\rm FFB}, \tag{F.27}$$

where $N_{\rm ch/CFB}$ is the number of channels per coarse filterbank ($N_{\rm ch/CFB} = \Delta\nu/\Delta\nu_{\rm ch\,-\,CFB}$) and $K_{\rm FFB}$ is the FFB processing cost for a single data stream.

The value of $K_{\rm FFB}$ is not easily determinable, because the filterbank processing architecture is designed so that, for each filterbank, the data flow, mathematical operations and memory

usage are optimised for some output frequency resolution (e.g. Bunton, 2010; Barott et al., 2011). However, K_{FFB} is constant in the present analysis, because the FFB output frequency resolution (equal to $\Delta\nu_{\mathrm{ch}}$) is fixed by the scientific requirements described above. Additionally, $\Delta\nu_{\mathrm{ch\,-\,CFB}}$ is held fixed in the present analysis; Bij de Vaate et al. (2011) specifies $\Delta\nu_{\mathrm{ch\,-\,CFB}} = 0.25\,\mathrm{MHz}$. Therefore a re-evaluation of Equation F.27 gives

$$P_{\mathrm{FFB}} \propto \Delta\nu\overline{N_{\mathrm{b\,-\,st}}}N_{\mathrm{st}}. \tag{F.28}$$

Once the data streams are split into fine channels, equivalent data streams (i.e. m^{th} fine channel of the n^{th} beam) from each antenna pair ($\sim N_{\mathrm{st}}^2/2$ pairs) are cross-correlated, at a correlation rate equal to the sample rate $\Delta\nu_{\mathrm{ch}}$ (Bunton, 2000). Because there are $N_{\mathrm{ch}} = \Delta\nu/\Delta\nu_{\mathrm{ch}}$ fine channels to be correlated, and $\Delta\nu_{\mathrm{ch}}$ is constant, the correlation processing cost is approximately

$$P_{\mathrm{X}} \propto \Delta\nu\overline{N_{\mathrm{b\,-\,st}}}N_{\mathrm{st}}^2. \tag{F.29}$$

Additional to the 'F' and 'X' computation hardware, there are also processing costs in the form of memory buffers and the inter-connects (corner turn) between the filterbanks and correlation devices (Turner et al., 2011). These costs are design dependent, and one or more of these costs may dominate the total correlator processing hardware cost. But it is reasonable to expect that the total cost will scale as

$$P_{\mathrm{corr}} \propto \Delta\nu\overline{N_{\mathrm{b\,-\,st}}}N_{\mathrm{st}}^{\alpha}, \tag{F.30}$$

where $1 < \alpha < 2$ depending on the design and technologies used. If N_{st} is held constant,

$$P_{\mathrm{corr}} \propto \Delta\nu\overline{N_{\mathrm{b\,-\,st}}}. \tag{F.31}$$

Dual-band implementation

The correlator processing for the dual-band implementation simply requires the substitution of $\overline{N_{\mathrm{b\,-\,st[D]}}}$ (Appendix F.3.2) into Equation F.31. The correlator processing ratio between the dual and single-band implementations is therefore

$$P_{\mathrm{corr}}(\mathrm{dual} : \mathrm{single}) = \overline{\mathrm{N_{b\,-\,st}}}(\mathrm{dual} : \mathrm{single}). \tag{F.32}$$

Trading $N_{\mathrm{e/st}}$ for N_{st}

For constant $A_{\mathrm{e\,-\,arr}}$, Equations F.22 and F.51 can be substituted into Equation F.30. Assuming d_{avg} and the antenna element gain G_{e} remain constant,

$$P_{\mathrm{corr}} \propto \Delta\nu N_{\mathrm{e/st}}^{1-\alpha}, \tag{F.33}$$

where $1 < \alpha < 2$.

Because N_{st} changes equally for both the single and dual-band implementations, Equation F.32 still holds true. Although the cost of the correlator processing increases with N_{st}, the relative cost is independent of the N_{st} vs. $N_{\mathrm{e/st}}$ trade-off.

Fine channelisation and correlation costs

There are a range of correlator cost estimates (<€1 M to nearly €100 M) for SKA_1 in Turner (2011), representing different architectures, technologies and options for flexibility. The correlation of SKA_1-low stations, rather than the dishes, represent most of the cost. The cost estimates generally only encompass the processing units and data inter-connects; doubling the cost accounts for all the accessory hardware required to support these processing units. Costs such as non-recurring engineering, which are generally not included in the estimates, can be considered a fixed cost. A zeroth-order estimate of the cost of the parts of the SKA_1-low correlator which scale as the number of input beams is €20 M; this equates to a GPU-class correlator. This estimate is specified for 480 station beams ($\Delta\nu = 380\,\mathrm{MHz}$), which is approximately €40 k per station beam.

F.4.2.3 Correlator–computing data transport

Each datum produced from the correlator (Equation F.29) is integrated for time Δt. Thus the data rate out of the correlator is:

$$R_{\mathrm{corr\,-\,out}} \propto \frac{N_{\mathrm{ch}}\overline{N_{\mathrm{b\,-\,st}}}N_{\mathrm{st}}^2}{\Delta t}. \tag{F.34}$$

For a fixed processed FoV Ω_{proc}, the data rate out of the correlator $R_{\mathrm{corr\,-\,out}}$ is independent of station diameter when the frequency resolution and integration time are set by the maximum smearing requirements (Equations F.25 and F.26). In that case,

$$\begin{aligned} R_{\mathrm{corr\,-\,out}} &\propto \frac{\Delta\nu\overline{N_{\mathrm{b\,-\,st}}}N_{\mathrm{st}}^2}{D_{\mathrm{st}}^2} \\ &\propto \Delta\nu\Omega_{\mathrm{proc}}N_{\mathrm{st}}^2, \end{aligned} \tag{F.35}$$

given $\Omega_{\mathrm{proc}} \propto \overline{N_{\mathrm{b\,-\,st}}}\Omega_{\mathrm{st}} \propto \overline{N_{\mathrm{b\,-\,st}}}D_{\mathrm{st}}^{-2}$. This result is independent of station diameter.

However, the maximum smearing requirements do not set the frequency resolution for SKA_1-low. The more stringent requirement on frequency resolution discussed earlier means that $\Delta\nu_{\mathrm{ch}}$, hence N_{ch}, does not vary with station diameter. Thus a re-evaluation of Equation F.34 gives

$$R_{\mathrm{corr\,-\,out}} \quad \propto \quad \frac{\Delta\nu\overline{N_{\mathrm{b\,-\,st}}}N_{\mathrm{st}}^2}{D_{\mathrm{st}}}. \tag{F.36}$$

This is consistent with the more detailed analysis in McCool (2011b).

Dual-band implementation

For the dual-band implementation, $R_{\mathrm{corr\,-\,out}}$ is calculated separately for the low and high band station diameters and summed to achieve a data rate for the full bandwidth. Thus

$$R_{\mathrm{corr\,-\,out}} \propto \left(\frac{(\nu_{\mathrm{split}} - \nu_{\mathrm{min}})\overline{N_{\mathrm{b\,-\,st[L]}}}N_{\mathrm{st[L]}}^2}{D_{\mathrm{st[L]}}} + \frac{(\nu_{\mathrm{max}} - \nu_{\mathrm{split}})\overline{N_{\mathrm{b\,-\,st[H]}}}N_{\mathrm{st[H]}}^2}{D_{\mathrm{st[H]}}} \right). \tag{F.37}$$

Given $N_{\mathrm{st[L]}} = N_{\mathrm{st[H]}} = N_{\mathrm{st[S]}}$ and $\overline{N_{\mathrm{b\,-\,st}}} \propto D_{\mathrm{st}}^2$ (Equation F.17), the ratio $R_{\mathrm{corr\,-\,out}}$(dual : single)

can be simplified to

$$R_{\text{corr - out}}(\text{dual} : \text{single}) = \frac{D_{\text{st[L]}}(\nu^3_{\text{split}} - \nu^3_{\text{min}}) + D_{\text{st[H]}}(\nu^3_{\text{max}} - \nu^3_{\text{split}})}{D_{\text{st[S]}}(\nu^3_{\text{max}} - \nu^3_{\text{min}})}. \tag{F.38}$$

Trading $N_{\text{e/st}}$ for N_{st}

For constant $A_{\text{e - arr}}$, Equations F.22, F.51 and F.53 can be substituted into Equation F.36. Assuming d_{avg} and the antenna element gain G_{e} remain constant,

$$R_{\text{corr - out}} \propto \frac{\Delta\nu}{N^{3/2}_{\text{e/st}}}. \tag{F.39}$$

To account for the trade between $N_{\text{e/st}}$ and N_{st} for fixed $A_{\text{e - arr}}$, as well as the different inter-element spacing between the low and high-band arrays, $D_{\text{st}} \propto \sqrt{N_{\text{e/st}}} d_{\text{avg}}$ (Equation F.53) can be substituted into Equation F.38. Because $N_{\text{e/st[S]}} = N_{\text{e/st[L]}} = N_{\text{e/st[H]}}$,

$$R_{\text{corr - out}}(\text{dual} : \text{single}) = \frac{d_{\text{avg[L]}}(\nu^3_{\text{split}} - \nu^3_{\text{min}}) + d_{\text{avg[H]}}(\nu^3_{\text{max}} - \nu^3_{\text{split}})}{d_{\text{avg[S]}}(\nu^3_{\text{max}} - \nu^3_{\text{min}})}. \tag{F.40}$$

Although the correlator output data rate increases with N_{st} (Equations F.36 and F.37), the relative data rate between the single and dual-band implementations is independent of the N_{st} vs. $N_{\text{e/st}}$ trade.

Correlator–computing data transport cost

For SKA$_1$-low, the HLSD lists an average data rate of 332×10^9 byte s^{-1} (2.66 Tbps) from the correlator to the computer; a factor of 1.25 encoding overhead brings this to 3.32 Tbps. Taking an approach similar to the station-CPF link (Appendix F.2.8), only the correlator–computing data transmission is costed; the trenching and cabling cost is ignored. Assuming that the computing is off-site at a nearby city or other suitable location, a cost-effective option is to multiplex the signals onto fibre using DWDM technology (see Section 6.3.2).

The DWDM transmission can be split into a data transmit–receive cost and a signal amplification cost. McCool (2010) costs the transmitter–receiver pair at €2 k per 10 Gbps channel and the optical amplifier and dispersion compensator at €10 k per 16-channel unit. The optical signal amplification is required every 80 km. If three amplification units are required, this results in a conservative estimate of €4 k per 10 Gbps channel, or €400 per Gbps for the data transmission. This is higher than the average cost of the station–CPF transmission, because of the greater distances. The correlator–computing transmission cost for the HLSD is then approximately €1.3 M.

F.4.2.4 Imaging

The SKA post-correlator processing requirements and algorithms, and their effect on computational cost, is an area of active research (e.g. Alexander, 2011). The 'imaging' sub-system encompasses the processing of visibilities from the correlator into imaged data products, as outlined in Alexander et al. (2009). Considered here are the 'gridding' operations on the visibility data which are the main computational cost (Cornwell, 2004; Alexander et al., 2009). Many

of the imaging operations act on the data from the correlator, thus buffering of these data is required (Faulkner et al., 2010). Assuming that the imaging cost is dominated by the data buffer, rather than the operations cost of the imaging algorithms, then the data rate out of the correlator (Appendix F.4.2.3) can be used as an indicator of cost (Alexander et al., 2009).

If the processing cost is considered, then there is a contribution from the data volume, but also a cost for correcting for non-coplanar baselines in wide-field images. For continuum imaging, which requires high dynamic range in the presence of confusing sources, Perley & Clark (2003) and Cornwell (2004) establish scaling relationships for an array of single-pixel feed dishes. Cornwell (2005) extends this analysis to multi-beam systems such as aperture arrays and dishes equipped with phased array feeds. The processing cost and its scaling as a function of dish or station diameter varies: the data rate out of the correlator is one factor; the other is the cost per visibility to correct for non-coplanar baselines. That cost depends on whether the correction can be done by separately imaging each independently pointed station beam, imaging the entire processed FoV at once, or by using some other algorithm. Modelling the cost of this processing and its scaling relationship with station diameter is beyond the scope of this work.

Imaging processor cost

The hardware implementation for SKA computing is currently ill-defined, but a simple estimate can be derived from the SKA budget. Garrett et al. (2010) budget €350 M for SKA_1-low capital investment, which includes a "significant element of contingency". I use €20 M as a zeroth-order estimate of the processing hardware and related infrastructure for the imaging aspect of the computing. Software is the other major computing cost, but estimating any cost difference between single and dual-band is beyond the scope of this work.

F.4.2.5 Non-imaging processing

The main sub-systems of the non-imaging processing are a central beamformer, and pulsar searching and timing on the beams formed, as described in Turner et al. (2011). Each beam is a phased or 'tied' array beam, formed using some or all the stations in the array as inputs to the beamformer. (In contrast, station beams are formed from the antenna elements or tiles in the station.)

The pulsar survey costs are assumed to be dominant, compared to the pulsar timing costs (Turner, 2011). Also, although the central beamformer may be a combined with the correlator sub-system (e.g. Turner et al., 2011), the processing for the pulsar survey is likely to be a significantly larger cost. The search for pulsars is conducted on a 'per beam' basis and can be computationally expensive, especially if searching for binary pulsars using acceleration searches. For example, Smits et al. (2009) calculate the number of computational operations required for the acceleration search with SKA Phase 2, for a fiducial set of search parameters; the number of operations for the acceleration search are two orders of magnitude greater than for the array beamforming. However, determining the algorithm for optimal processing loads and data rates for the SKA requires further investigation (Turner et al., 2011).

Regardless of the algorithm, the general processing trend can be analysed because the search is conducted on each beam. The number of array beams $N_{\mathrm{b\,-\,arr}}$ required to survey the sky

depends on the FoV of each beam. That in turn depends on the frequency of observation, and the diameter of the array from which the beams are formed: $N_{\mathrm{b\,-\,arr}} \propto (\nu D_{\mathrm{arr}})^2$. Although using more stations increases sensitivity, it also increases D_{arr}. To limit the computational requirements, Smits et al. (2011a) suggest that only stations in the 1 km diameter core be used to form the processed FoV.

The single-band implementation (Figure 5.4) has a 1 km diameter core. However, for the high band of my representative dual-band implementation, the smaller inter-element spacing means that the core is only 0.5 km in diameter (Figure 5.5). This results in the required FoV being met with factor of four fewer core array beams. The cost of beamforming is approximately linearly proportional to the number of beams (Appendix F.2.6.1) and the subsequent processing to search for pulsars is conducted on the per beam basis, hence the processing cost for the dual-band implementation is 25% of the single-band implementation. To realise the factor of 4 cost reduction, the 0.5 km diameter high-band core must be physically separate to the low-band core, as shown in Figure 5.5.

Turner et al. (2011) assume that the 35 SKA_1-low stations (180 m diameter) in the central 5 km of the array are used for pulsar searches. In that case, more compact stations will not necessarily decrease this 5 km diameter, hence the dual-band array provides no extra benefit. However, using all the stations within the 5 km is not cost-effective. As Table 4.3 (Chapter 4) shows, the cost-effectiveness of high time resolution searches is significantly reduced outside the densely packed core of the SKA_1-low array; the extra sensitivity gained from including more stations is insufficient to offset the many more array beams which must be formed to meet the required FoV.

Non-imaging processor cost

Turner (2011) provides a cost of €28 m for a pulsar search concept description by Knittel & Horneffer (2011). This cost is for processing hardware including server cases. Adding in racks and power distribution (which scale approximately linearly with processing), I round the cost up to €30 m. The concept description does not specify whether processing hardware searches the 1.25 deg^2 FoV or some subset of that. However, given the search is done on a per beam basis, this is not important because the relative cost applies regardless.

F.4.3 Power demand

The power demand estimate is sourced from the bottom-up SKADS power budget in Faulkner et al. (2011). The power budget is for the all-digital station architecture, but extrapolation to the RF tile beamforming architecture is done by including a power cost for the RF beamformer. The scaling relationships for power demand are shown in Table F.7 and the unit costs in Table F.8.

Table F.7: Summary of blocks and scaling for power demand in SKA_1-low station sub-systems (RF first-stage beamforming is optional).

Block name	Quantity in SKA_1	Power scaling	Block coverage
Active antenna element	$N_{\mathrm{st}}N_{\mathrm{e/st}}$	C_{fix}	LNA and antenna gain
Analogue (RF) tile beamformer[a]	$N_{\mathrm{st}}N_{\mathrm{tile/st}}$	$N_{\mathrm{b\,-\,tile}}C_{\mathrm{var}}$	beamformer
Digitiser	RF: $N_{\mathrm{st}}N_{\mathrm{tile/st}}N_{\mathrm{b\,-\,tile}}$ Dig: $N_{\mathrm{st}}N_{\mathrm{e/st}}$	$C_{\mathrm{fix}}+R_{\mathrm{sample}}C_{\mathrm{var}}$	C_{fix}: analogue signal conditioning, clock distribution C_{var}: analogue to digital converter
Digitiser–bunker link[b]	Dig: $N_{\mathrm{st}}N_{\mathrm{e/st}}$	$C_{\mathrm{var}}R_{\mathrm{dig}}$	copper communication from digitiser and fibre transmission electronics
Station beamformer[c]	N_{st}	RF: $\overline{N_{\mathrm{b\,-\,st}}}N_{\mathrm{tile/st}}C_{\mathrm{var}}$ Dig: $\overline{N_{\mathrm{b\,-\,st}}}N_{\mathrm{e/st}}C_{\mathrm{var}}$	digital processing, inter-connections and control
Station–CPF link transmission	N_{st}	$R_{\mathrm{st}}C_{\mathrm{var}}$	fibre transmission

[a] Optional block. If analogue tile beamforming is included, subsequent quantities and costs are denoted 'RF'. If not, the system is all-digital beamforming, denoted by 'Dig'.
[b] Optional block. Assumes no digital beamforming at the tile. Alternative architectures are discussed in Section 5.7.3.
[c] Approximate cost scaling, see Appendix F.2.6.1.

Table F.8: Unit costs for the power demand of SKA_1-low station sub-systems.

Block name	Power demand estimate (mW)	Block unit[a]
Active antenna element	C_{fix}: 180	per element
RF tile beamformer	C_{var}: 400 per output beam	per tile
Digitiser	C_{fix}: 180 C_{var}: 200 per GS/s	per signal
Digitiser–bunker link transmission	C_{var}: 12.4 per Gbps	per link
Station beamformer[b]	C_{var} : 18.6 per input per output beam	per station
Station–CPF link transmission	C_{var}: Not available	per link

[a] All elements, beams, inputs and signals are dual polarisation.
[b] Dual-band station beamformer power demand is 29% and 71% of this value, for the low and high-bands respectively.

F.5 Station performance considerations

F.5.1 Sensitivity requirements and inter-element spacing

Telescope sensitivity is a key requirement on the system. The required sensitivity is usually derived from the required minimum detectable flux density and the telescope time available for each pointing (patch of sky being observed). An exception is some time-domain astronomy where further integration of a single pointing does not increase sensitivity. Sensitivity is generally given as the metric $A/T = A_{\mathrm{e\,-\,arr}}/T_{\mathrm{sys}}$, where $A_{\mathrm{e\,-\,arr}}$ is the effective area of the telescope array and T_{sys} is the system temperature. For the electronically steered aperture arrays, significant variations in A/T are caused by:

- inter-element spacing
- scan angle
- strong sources in the sidelobes.

Additionally, the effect of these parameters on sensitivity is frequency dependent. The effect of inter-element spacing is the most relevant to this analysis, and the basic trends are considered here. While the other parameters are also important, a complete analysis is beyond the scope of this document.

The inter-element spacing defines the frequency at which the antenna elements transition from 'dense' to 'sparse'. There is no single definition for when an array is dense or sparse. The broad definition used in this work is that an array is dense when the aperture is fully sampled, such that effective area is approximately constant with frequency: $A_{\mathrm{e\,-\,arr}}(\nu) = \text{constant}$. When the array is sparse, the effective area of each isolated element contributes to the array effective area, hence $A_{\mathrm{e\,-\,arr}} \propto \lambda^2$. There is also a transition region between dense and sparse, which occurs at an inter-element spacing of $0.5 - 1.5\lambda$ for dipole-like antennas (Braun & van Cappellen, 2006).

To give an indication of how the inter-element spacing affects the representative systems, Figure F.1 plots A/T as a function of frequency, using the first-order analysis of the problem in Nijboer et al. (2009) and the HLSD. The array effective area is given by

$$A_{\mathrm{e\,-\,arr}} = \begin{cases} N_{\mathrm{st}} \frac{\pi}{4} D_{\mathrm{st}}^2 & \nu < \nu_{\mathrm{transition}} \text{ (dense)} \\ N_{\mathrm{st}} N_{\mathrm{e/st}} \frac{\lambda^2}{3} & \nu > \nu_{\mathrm{transition}} \text{ (sparse)}, \end{cases} \tag{F.41}$$

where $\nu_{\mathrm{transition}}$ is the frequency of the dense–sparse transition. The system temperature is approximated by the sum of the receiver and sky noise temperatures:

$$T_{\mathrm{sys}} = 150 + 60\lambda^{2.55}. \tag{F.42}$$

The 'always sparse' curve in Figure F.1 reflects the isolated antenna element case, where $\nu > \nu_{\mathrm{transition}}$ is always true. The frequency at which this curve peaks depends somewhat on the receiver and sky noise models. At the higher frequencies, the arrays are sparse for all four spacing values and A/T is independent of the inter-element spacing. At the lower frequencies, the inter-element spacing defines the discontinuity. This is the transition frequency; the point at which the aperture becomes fully sampled and A/T begins to drop below the always sparse curve.

Figure F.1 applies to the single-band implementation and the low (70–180 MHz) band of the dual-band implementation. The sensitivity for the high (180–450 MHz) band is plotted in Figure F.2. In this figure, the discontinuities exist for the 0.5 m and 0.75 m inter-element spacing. The 1 m spacing transition frequency is lower than 180 MHz, so the A/T curve is the same as the always sparse curve. The representative single-band A/T is also equal to the always sparse curve, due to its 1.5 m inter-element spacing. Although the canonical dual-band array is designed for comparable performance to the single-band array, the 0.75 m spacing means the dual-band array has less sensitivity between 180 and 230 MHz.

In reality, the layout of the tile elements within a station will affect the form of the discontinuity and the slope of the curve at frequencies below this discontinuity. Intra-station layouts such as golden ratio spiral (GRS)[1], fractal patterns and irregular arrays will have a distribution of inter-element spacing, i.e. a minimum ($d_{\min}$), maximum ($d_{\max}$) and average (d_{avg}) spacing. The station will be dense once the frequency is low enough that those elements with spacing of $d_{\max}$ become dense. The station will be sparse once the frequency is high enough that those elements with spacing of $d_{\min}$ become sparse. Between these frequencies, this distribution of inter-element spacing causes the station to be 'semi-sparse', and neither case in Equation F.41 is applicable. Such layouts can broaden the discontinuity at the dense-sparse transition and reduce the slope of the A/T curve, as can be seen in Figures 22 to 27 of Bij de Vaate et al. (2011).

Although this simple analysis gives some indication of how A/T changes with inter-element spacing, there is scope for more detailed investigation. The station effective area changes as a function of zenith angle θ and azimuth angle ϕ, due to a changing beam pattern. The beam pattern itself, pointed at a particular direction (θ, ϕ), is dependent on the intra-station layout pattern (this also defines the station beam FoV). For example, van Cappellen et al. (2006) compares regular and irregular layouts of uniform aperture distribution for sparse AA stations. Strong astronomical sources in the sidelobes of these station beams will also greatly influence T_{sys} (hence A/T), as shown in Wijnholds & van Cappellen (2011). Also, spatial tapering would increase the A/T at lower frequencies. But the extent of this increase, and the related frequency-dependent effect on beam pattern (hence Ω_{st}) requires further investigation.

The station effective area calculations also depend on the gain or directivity of the antenna element. For example, log-periodic, conical spiral and Vivaldi elements are discussed in Bij de Vaate et al. (2011). These have higher directivity (at $\theta = 0$) than the proposed element in the HLSD. However, the directivity as a function of scan angle depends on the antenna design. Thus the station directivity (hence $A_{\text{e-st}}$) must be considered down to the maximum scan angle (zenith angle) $\theta_{\max}$, which Bij de Vaate et al. (2011) specifies as $\theta_{\max} = 45°$. Appendix F.5.4 details further work to refine station sensitivity estimates.

F.5.2 Filling factor and station calibration

The ability to calibrate the SKA_1-low telescope also has system-wide implications. To achieve the desired performance, the instrumental response of the telescope needs to be accurately

[1] GRS is a form of spatial taper, with the density of elements reducing with increased radius from the centre.

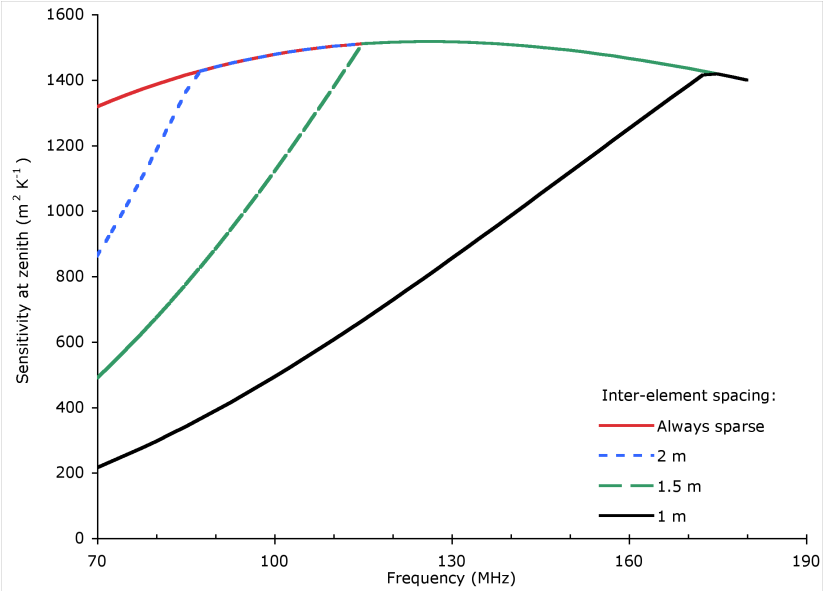


Figure F.1: Approximate SKA$_1$-low sensitivity (A/T) at zenith as a function of frequency (70–180 MHz) and inter-element spacing, using Equations F.41 and F.42.

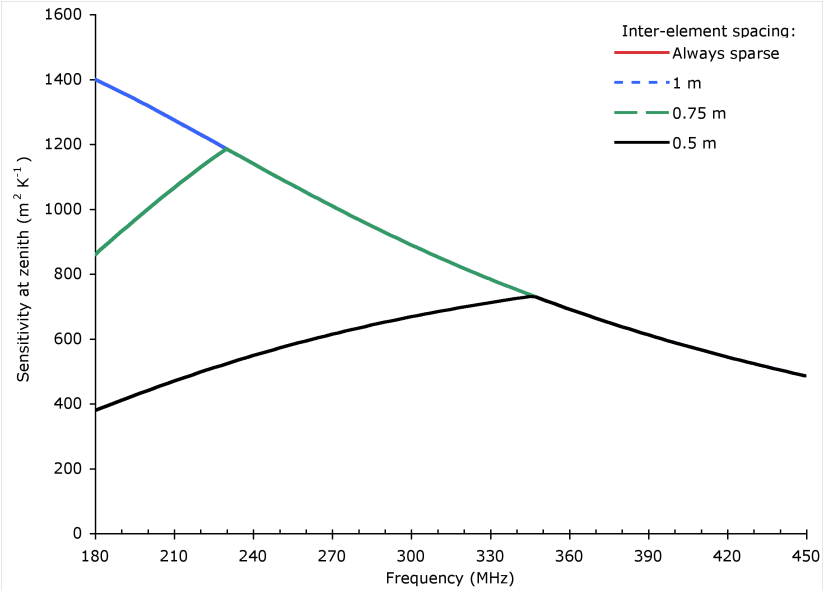


Figure F.2: Approximate SKA$_1$-low sensitivity (A/T) at zenith as a function of frequency (180–450 MHz) and inter-element spacing, using Equations F.41 and F.42.

characterised via calibration (e.g. Wijnholds et al., 2010) and calibrating the station beams is one aspect of this. Wijnholds et al. (2011b) discusses station calibration and determines that 3–5 calibration sources need to be detectable in the station beam, assuming that the station size meets the requirements outlined in Wijnholds et al. (2011a). A metric which is related to the number of detectable sources is the station filling factor FF_{st}, given by

$$FF_{\mathrm{st}} = \frac{A_{\mathrm{e-st}}}{A_{\mathrm{g-st}}}, \tag{F.43}$$

where $A_{\mathrm{e-st}}$ is the station effective area, and $A_{\mathrm{g-st}}$ is the geometrical (physical) area occupied by the station. For a station with a uniform taper (aperture distribution), a fractional bandwidth of 20 % and an antenna and receiver noise of 50 K, Wijnholds et al. (2011b) require a filling factor of 0.2–0.4 at 400 MHz, although more recently this has been revised to 0.1 (Wijnholds, 2012).

The competing effects of array sparseness to maintain sensitivity at low frequencies and array density for station calibration at high frequencies is problematic. Wijnholds & Bregman (2011) propose that a dual-band implementation should be used, unless a solution can be found by either optimising the intra-station layout of the single band array to meet both requirements, or using multiple beams at the higher frequencies to detect the required calibration sources. The dual-band array has more flexibility to adjust the high-band design to meet the filling factor requirement.

This problem is reflected in the single-band implementation considered in this analysis, where $FF_{\mathrm{st}} = 0.08$ at 400 MHz for a zenith pointing. For the representative dual-band implementation, $FF_{\mathrm{st}} = 0.33$ at the same frequency and pointing. Although in this case, the filling factor is too low for the single-band implementation, the required and calculated filling factors will vary, depending on the intra-station layout and element gain of the single-band design. Determining whether single-band solution can meet the filling factor requirements is a current work in progress within the SKA aperture array community.

F.5.3 Trading $N_{\mathrm{e/st}}$ for N_{st} and the relationship with station diameter

For constant array effective area $A_{\mathrm{e-arr}}$, there is a trade-off between aperture array station diameter D_{st} and the number of stations N_{st}, where $N_{\mathrm{st}} \propto 1/D_{\mathrm{st}}^2$. This is analogous to previous investigations trading dish diameter for the number of dishes (e.g. Chippendale et al., 2007; Schilizzi et al., 2007). However, this does not completely describe the independent variables in the trade-off. A parabolic dish (with single-pixel feed) can be thought of as a densely sampled aperture, with the mechanical structure performing the beamforming. If the aperture array is dense, such that the aperture is fully sampled (see Appendix F.5.1), then $N_{\mathrm{st}} \propto 1/D_{\mathrm{st}}^2$. But when the antenna elements are sufficiently spaced such that the AA is sparse over some or all frequencies, then D_{st} is a function of two independent parameters, $N_{\mathrm{e/st}}$ and the average inter-element spacing d_{avg}, as shown below.

Array effective area is given in Equation F.41, but it can be more generally described by

$$A_{\mathrm{e-arr}}(\nu) = N_{\mathrm{st}} A_{\mathrm{e-st}}(\nu), \tag{F.44}$$

where $A_{\mathrm{e-st}}$ is the station effective area at some frequency ν. Assuming an irregular intra-station element layout with an approximately uniform element distribution,

$$A_{\mathrm{e-st}}(\nu) \approx N_{\mathrm{e/st}} A_{\mathrm{e-e}}(\nu), \tag{F.45}$$

where $A_{\mathrm{e-e}}(\nu)$ is the antenna element effective area. For an isolated antenna element,

$$A_{\mathrm{e-e}}(\mathrm{isolated}) = \frac{G_{\mathrm{e}}\lambda^2}{4\pi}, \tag{F.46}$$

where G_{e} is the antenna gain. However, in the presence of neighbour elements, $A_{\mathrm{e-e}}$ cannot be greater than the available physical area (Nijboer et al., 2009), thus

$$A_{\mathrm{e-e}} \approx \min\left\{\frac{\mathrm{G_e}\lambda^2}{4\pi}, \,\mathrm{available\,geometric\,area}\right\}, \tag{F.47}$$

where, for a uniform intra-station element distribution, the available geometric area is approximated by d_{avg}^2. Restating Equation F.47 in terms of the frequency-dependent station filling factor (Equation F.43) gives

$$A_{\mathrm{e-e}} = d_{\mathrm{avg}}^2 FF_{\mathrm{st}}, \tag{F.48}$$

where

$$FF_{\mathrm{st}} \approx \min\left\{\frac{\mathrm{G_e}\lambda^2}{4\pi \mathrm{d_{avg}^2}}, \,1\right\}. \tag{F.49}$$

As expected, $FF_{\mathrm{st}} \leq 1$; a filling factor of 1 signifies that the array is dense.

In its full form, Equation F.44 is thus

$$A_{\mathrm{e-arr}}(\nu) \approx N_{\mathrm{st}} N_{\mathrm{e/st}} d_{\mathrm{avg}}^2 FF_{\mathrm{st}}(\nu). \tag{F.50}$$

When the function $A_{\mathrm{e-arr}}$ is held constant,

$$N_{\mathrm{st}} \propto \frac{1}{N_{\mathrm{e/st}} FF_{\mathrm{st}} d_{\mathrm{avg}}^2}. \tag{F.51}$$

If d_{avg} and the function FF_{st} also remain constant (i.e. G_{e} does not vary in FF_{st}), then

$$N_{\mathrm{st}} \propto \frac{1}{N_{\mathrm{e/st}}}. \tag{F.52}$$

This is not applicable for intra-station layouts with non-uniformly distributed antenna elements (see Appendix F.5.1), because the approximation for $A_{\mathrm{e-st}}$ (Equation F.45) does not hold.

The relationship with station diameter can also be determined, assuming that the array is dense at the lowest frequency. Equation F.47 shows that $A_{\mathrm{e-e}} \lesssim d_{\mathrm{avg}}^2$, therefore the maximum station effective area is $A_{\mathrm{e-st}}(\mathrm{max}) \approx N_{\mathrm{e/st}} d_{\mathrm{avg}}^2$. From geometry, $A_{\mathrm{e-st}}(\mathrm{max}) \propto D_{\mathrm{st}}^2$, hence

$$D_{\mathrm{st}} \propto \sqrt{N_{\mathrm{e/st}}} d_{\mathrm{avg}}. \tag{F.53}$$

This relationship is important, because it shows that station diameter is not an independent parameter, but is influenced by both the number of elements in the station and the average spacing between the elements.

The veracity of Equation F.51 can be confirmed by considering the extreme cases. If the array

is completely dense, such that $FF_{\rm st} = 1$, then

$$N_{\rm st} \propto 1/D_{\rm st}^2. \tag{F.54}$$

If the array is completely sparse, then

$$FF_{\rm st} = \frac{G_{\rm e}\lambda^2}{4\pi d_{\rm avg}^2} \tag{F.55}$$

and

$$N_{\rm st} \propto \frac{1}{N_{\rm e/st} A_{\rm e\,-\,e}({\rm isolated})}. \tag{F.56}$$

F.5.4 Further work to refine station performance metrics

Matching the top-level science requirements to the telescope performance requires an understanding of the telescope's A/T and processed FoV performance. The A/T performance of a higher-frequency dish-based aperture synthesis telescope is well-understood (e.g. Crane & Napier, 1989). The A/T performance of an aperture synthesis telescope composed of aperture array stations, in the lower frequency, sky noise dominated regime, is more complex (e.g Wijnholds & van Cappellen, 2011).

Considering the factors affecting sensitivity discussed in Appendix F.5.1, the requirements and performance metrics can be specified as $A/T(\theta, \phi, \nu)$ and $\Omega_{\rm req}(\theta, \phi, \nu)$, or at least a minimum A/T and $\Omega_{\rm req}$ as a function of frequency over some range of (θ, ϕ). Some simple rules-of-thumb are used in this analysis, but more accurate estimates could be obtained through array layout simulation software, such as Xarray[2] and OSKAR (Dulwich et al., 2009).

For each layout, simulation of a small set of interdependent input parameters and performance metrics would be useful inputs for the parametric analysis. They are:

- input parameters
  - antenna element pattern
  - intra-station layout and inter-element spacing
  - number of elements per station
  - number of elements per tile (if used)
- performance metrics
  - effective area as a function of frequency and direction
  - station beam FoV as a function of frequency and direction.

A layout will also have advantages and disadvantages which cannot be captured in the parametric analysis (e.g. station beam pattern, sidelobes and calibration); these must be considered separately.

Improved accuracy will also be obtained if the simulations take into account effects such as

- mutual coupling between elements

[2] http://sites.google.com/site/xarraytool/

Table F.9: SKA_1-low station details for the half-diameter station example.

	Single-band	Dual-band	
		Low band	High band
Diameter	90 m	90 m	45 m
Number of elements per station	2 800	2 800	2 800
Number of stations	200	200	200
Average spacing between elements	1.5 m	0.75 m	1.5 m

- the non-ideal gain of the antenna element, as a function of ν, θ and ϕ (this parametric analysis assumes ideal gain)
- LNA response
- the effect on gain due to the analogue beamformer, as a function of ν, θ and ϕ.

F.6 Smaller station diameter example

I present a simple comparative example showing the effect of smaller stations, where the diameter of every single-band, low-band or high-band station is halved. The system details for the single and dual-band implementations of this half-diameter example are listed in Table F.9.

F.6.1 Station hardware costs

Costing the smaller station diameters requires consideration of the location of the station processing node. As in the dual-band implementation, I assume that each smaller station continues to have its own station beamformer processing node—the bunker. The only change to the cost model (Appendix F.2) is to halve the unit cost of the tile–digitiser RF link to €40 per link, due to the smaller station diameter. For the all-digital architecture, the digitiser–bunker link cost equation already has a distance-dependent term, so remains the same. Also, the dual-band cost multipliers (Appendix F.2.9) do not change.

Figure F.3 plots the hardware cost of all the SKA_1-low stations in the half-diameter example alongside that of the full-sized stations previously shown in Figure 5.6. Because the number of elements per station differ by a factor of 4, the cost of all stations is a more useful comparison than the cost per station. The total number of elements in the array remain constant, hence the total cost of the active antenna elements, RF links (for the all-digital architecture) and digitisers do not vary. Compared to the full-sized stations, the smaller average distance to the bunker reduces the total cost of the RF links in the RF tile beamforming architecture and the digitiser–bunker links in the all-digital architecture.

The average number of station beams is varied to maintain the required FoV. For this reason, the total processing cost of N_{st} beamformers (Equation F.11, p 197) can be approximated by

$$P_{\mathrm{BF-total}} \propto N_{\mathrm{e/st}}, \tag{F.57}$$

which is a factor of 4 reduction for the half-diameter stations. But despite the decrease in intra-

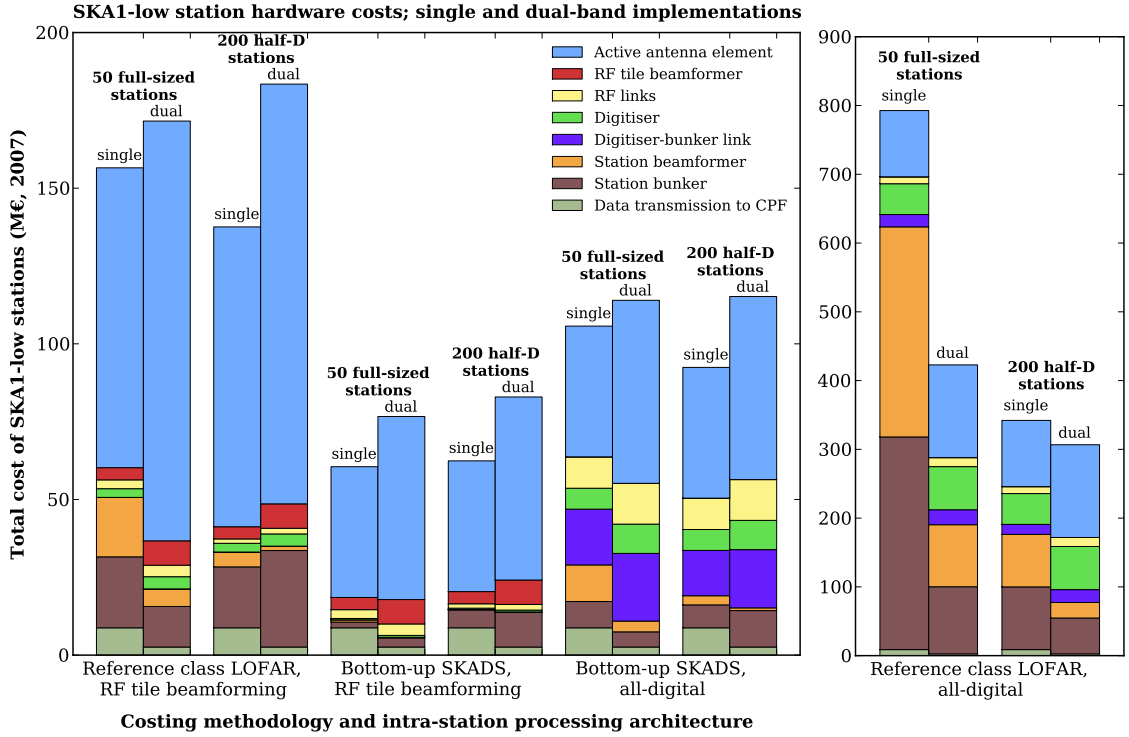


Figure F.3: SKA_1-low station hardware cost, for 50 full-sized stations (per band) and 200 stations of half the diameter, hence a quarter of the number of elements per station. Other details as per Figure 5.6.

station signal transport and beamformer processing costs, the station bunker puts upwards cost pressure on the total cost. Although a part of the station bunker cost varies with station beamformer processing, the fixed-cost portion of each bunker begins to become significant with the smaller stations. Because there are more bunkers (one for each station), the total fixed cost portion of the bunker is higher. This trend is further exacerbated in the dual-band implementation, where twice as many bunkers are used (one for each low-band and high-band station).

In the present analysis, the fixed cost portion of the bunker is set at 20 % of the source data cost estimates (bottom-up SKADS and reference class LOFAR); the other 80 % is assumed to scale linearly with the amount of station processing. Although the fixed cost is a zeroth-order estimate, its purpose is to recognise that there may be inefficiencies in providing environmental conditioning (cooling), RFI shielding and power to many smaller controlled environments (the bunkers), as well as higher construction, testing and deployment costs. Potential solutions such as shared processing nodes are discussed in Appendix F.6.3.

F.6.2 System implications of trading $N_{e/st}$ for N_{st}

As in the single and dual-band comparison (Section 5.5), there are other costs that vary with the N_{st} vs. $N_{e/st}$ trade-off. The only costs in Section 5.5 which vary are the correlator and imaging processor costs. The constant $A_{e\text{-}arr}$ means that the total area occupied by the array

and the total number of antenna elements in the array do not change, thus the site preparation and antenna element deployment costs do not change. Of course, there will still be twice as many antenna elements in the dual-band implementation and it will occupy an extra 25 % area, as shown in Table 5.6. The non-imaging processor cost does not change, because it depends on the diameter of the core, which remains the same.

The correlator and imaging processor costs do increase as the number of elements per station are traded for more stations, as derived in Appendix F.4.2. These increases apply equally to both the single and dual-band implementations. The correlator processing (Equation F.33) is

$$P_{\mathrm{corr}} \propto N_{\mathrm{e/st}}^{1-\alpha}, \tag{F.58}$$

where $1 < \alpha < 2$ depending on the correlator design and technologies used. The correlator output data rate, and hence imaging processor cost (Equation F.39) is

$$R_{\mathrm{corr\,-\,out}} \propto \frac{1}{N_{\mathrm{e/st}}^{3/2}}. \tag{F.59}$$

For the half-diameter station example, where there are a quarter of the number of elements per station, the correlator cost increases by up to 400 %. Although the number of correlations increases by N_{st}^2 (i.e a factor of 16), the number of beams required to produce the $20\,\mathrm{deg}^2$ processed FoV is reduced by a factor of 4 for the half-diameter stations, resulting in only a factor of 4 increase in correlation cost. For the post-correlation imaging processor, the cost increases by 800 %.

Figure F.4 plots the variable costs, for the zeroth-order cost estimates used in Section 5.5. A value of $\alpha = 2$ is used for the correlator processing cost. The correlator and imaging processor costs are significant for the half-diameter example, more so for the single-band implementation. Although the cost for the reference class LOFAR, all-digital beamforming scenario remains significantly larger than the others, the cost of single-band implementation of that scenario has reduced considerably.

The principal cost differences arise from whether the processing is distributed amongst the station beamformers or centralised at the correlator and image processor, however the magnitude of the cost differences depend on the implementation. The power demand, which scales with the amount of processing, also follows the same trend. The power demand of the station is not plotted, because it should be considered in context with the power demand of the correlator and imaging processor, which is beyond the scope of the present analysis.

F.6.3 Station diameter and shared processing nodes

To control the bunker cost, which is higher due to the increased number of stations, shared station nodes may be the preferable option for some architectures. Conceptually, this leads to two different sized stations: a ‘logical’ and ‘physical’ station, to borrow from software engineering terminology (e.g. Kruchten, 1995). The logical (functional) station is the beamformed station, where the beams are input to the correlator. The physical (infrastructure) station has one node (bunker) shared between a number of the beamformed stations. In the centrally condensed part of the array, where there is a ‘sea’ of antenna elements, the logical stations could

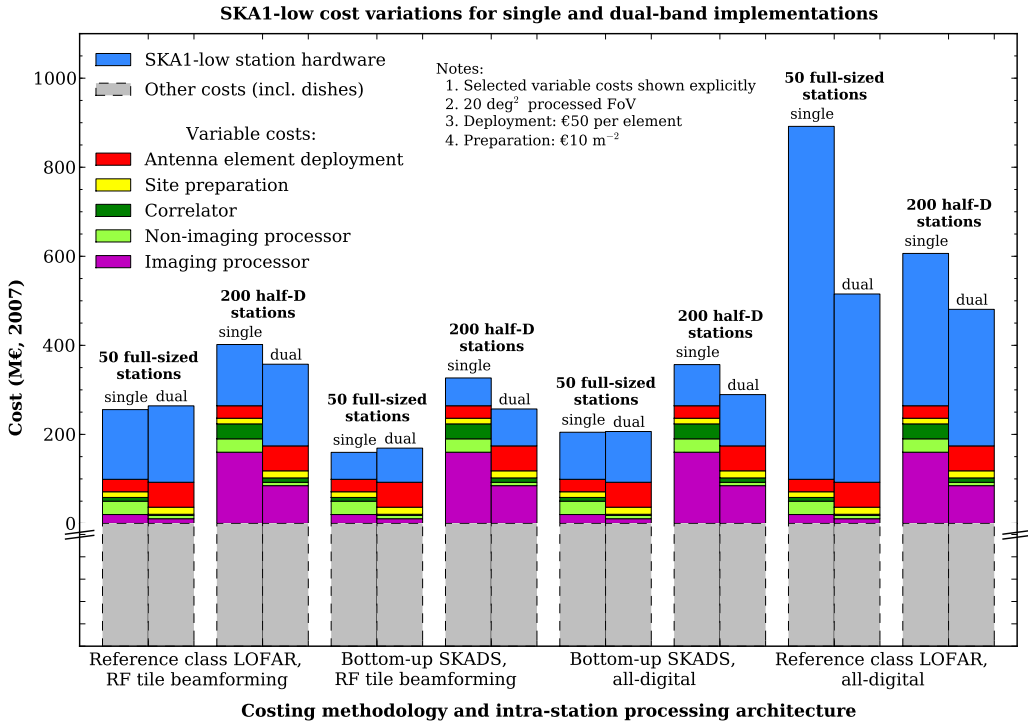


Figure F.4: Comparison of significant variable costs (excluding power) for 50 full-sized stations (per band) and 200 stations of half the diameter. The correlator processing cost is assumed to be scale as $1/N_{\mathrm{e/st}}$. Other details as per Figure 5.7.

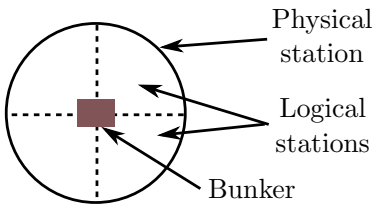


Figure F.5: Schematic of logical stations within a physical station, and the shared processing node (bunker).

be re-configurable in diameter, to suit the scientific application (Alexander, 2011).

As shown in Figure F.5, an obvious alternative in the half-diameter station example is to have 4 logical stations per physical station, which would mean the physical station size is similar to the full-diameter stations. The would be viable if the cost reduction of a single larger bunker outweighs the higher intra-station link costs. The link costs are higher due to the larger average distance to the bunker; inefficiencies and losses could be introduced in the longer RF and power cables, or alternatively, more expensive cables or digital transmission equipment may be required. A higher ratio of logical to physical stations may be cost-effective, especially for the high band of the dual-band array, which is already smaller in diameter.

The required geographical layout of stations, which determines (u, v) coverage, is another trade-off consideration. To maintain comparable infrastructure costs, the half-diameter station example assumes that the 4 smaller stations are located adjacent to each other, as will already

be the case in the centrally condensed part of the array. But for stations placed outside of the centrally condensed part of the array, 'clustering' or grouping of antennas is proposed as a method to reduce the infrastructure costs (Bolton et al., 2011). Beyond the 5 km 'inner' region of SKA_1, the HLSD describes clusters of receptors, containing 5 dishes and a single SKA_1-low station. Each physical station could be divided into 4 adjacently located smaller logical stations. However, if the purpose of reducing station diameter (thereby increasing the number of stations) is to improve (u, v) coverage through more diverse placement of stations, then such a strategy will not be particularly useful. Estimating the extra infrastructure cost to separate the stations with larger distances is beyond the scope of this analysis.

F.7 Reduced fixed beam–bandwidth product

If the aperture array system has a fixed processing capacity, then the processed FoV and processed bandwidth are tradable quantities. There are a few contemporary examples of this approach to aperture array system design:

- The trade-offs in the SKADS aperture array designs are based on the capacity to transmit data from the station to the central processing facility being a primary limitation (Bolton et al., 2009a).
- The LOFAR station processing is reconfigurable such that processed bandwidth can be traded for station beams, thus maintaining a set data rate from the station (de Vos et al., 2009).
- The MWA design has 220 MHz of sampled bandwidth available, but at any one time it only transports 30.72 MHz of this bandwidth to the central processing hardware (Tingay et al., 2012).
- The LWA design constrains the bandwidth for multiple station beams, due to limitations in transporting data from the stations to the correlator (Ellingson et al., 2009).

The amount of FoV that can be processed over some bandwidth can be constrained by fixing the beam–bandwidth product, $\overline{N_{\mathrm{b\text{-}st}}}\Delta\nu$, where $\overline{N_{\mathrm{b\text{-}st}}}$ is the average number of beams formed over the bandwidth $\Delta\nu$. Assuming that the data consists of many channelised beams (Section 2.3.3), the processing capacity of the station beamformer, station–CPF data transmission and the central processing sub-systems can then be determined by $\overline{N_{\mathrm{b\text{-}st}}}\Delta\nu$.

F.7.1 Strawman details

An alternative analysis by Faulkner[3] of the SKA_1 Design Reference Mission (DRM_1) version 2.0 (SSWG, 2011) considers the beam–bandwidth product required for individual DRM_1 chapters. I do not examine the details of such an analysis, but simply consider a strawman example where the largest beam–bandwidth product is defined by requirement to observe 20 deg^2 processed FoV over the 70–180 MHz band. However, I note that a frequency split at 200 MHz (hence a 70–200 MHz low band) for the dual-band implementation would provide a better fit to the science relevant to SKA_1-low in the current DRM_1 (Chapters 2–5); three of those science chapters specify either a maximum or minimum frequency of 200 MHz.

[3] AA-low system & station architecture for SKA_1, 18 January 2012

For this strawman, $20\,\mathrm{deg}^2$ FoV (70–180 MHz) equates to $\overline{N_{\mathrm{b\text{-}st}}} = 44$ over the band; this applies to both the single and dual-band implementations, because the station diameters are equal between 70 and 180 MHz. The beam–bandwidth product is thus 4.8 GHz. In contrast, to achieve $20\,\mathrm{deg}^2$ FoV (70–450 MHz) for the single and dual-band representative implementations, the beam-bandwidth capacity is 80 GHz and 24 GHz respectively.

The same 4.8 GHz of beam–bandwidth product can be applied to the dual-band array, although the strawman requires further design decisions to be assumed. I maintain the original assumption of separate low and high-band cores (p 98), hence the same 4.8 GHz beam–bandwidth requirement applies to each station in each core. The stations beyond the core are co-located (p 106), but I now assume that the station node hardware (beamformer, bunker and station–CPF data transmission) is shared between each pair of low and high-band stations; this is conceptually similar to the LOFAR station design (Gunst & Bentum, 2010). A different design of the dual-band strawman could have a single core populated with both low and high-band stations, with shared processing. However, this would result in a larger core, which will have consequences on the science applications, such as low surface brightness density and non-imaging processing observations.

Because the station beamforming computational capacity and the maximum rate of station–CPF data transmission is defined by the beam–bandwidth product, observations with other processed FoV and bandwidth combinations are possible, although the observations cannot be concurrent. For example, an observation over the 180–450 MHz frequency range with the single-band array results in a processed FoV of $\Omega_{\mathrm{proc}} = 1.3\,\mathrm{deg}^2$. The same observation with the dual-band results in a processed FoV four times larger, due to the smaller diameter of the high-band station.

F.7.2 Station hardware costs

Figure F.6 shows the station hardware costs for the fixed 4.8 GHz beam–bandwidth product. For all scenarios, the single-band implementation is cheaper than the dual-band, at approximately 70 % of the dual-band cost. The station beamformer and station–CPF data transmission costs are now insignificant. This is because the beam–bandwidth capacity of the station node hardware in Figure F.6 is reduced by a factor of approximately 17 and 5, compared to the representative single and dual-band implementations respectively. The station bunker cost does not reduce to the same extent as the station beamformer, due to the fixed cost portion of the bunker (see Appendix F.6.1). Meanwhile, the cost of the station hardware sub-systems located in the signal path prior to the station beamformer remain the same; they are independent of changes to the beam–bandwidth product. The net result is a reduction in station hardware cost for each scenario, compared to when processed FoV of $20\,\mathrm{deg}^2$ over 70–450 MHz is required.

The previous results, calculated for the representative implementations, show that the dual-band implementation puts downward pressure on the station beamformer and station-CPF transmission costs, which counteracts, to varying extent, the increase in the cost of other station hardware sub-systems in the dual-band implementation. But for this strawman, which is limited by the beam–bandwidth product, the downward pressure is insignificant. Thus in all the scenarios, the cost of the dual-band station hardware is higher. The increase in cost not

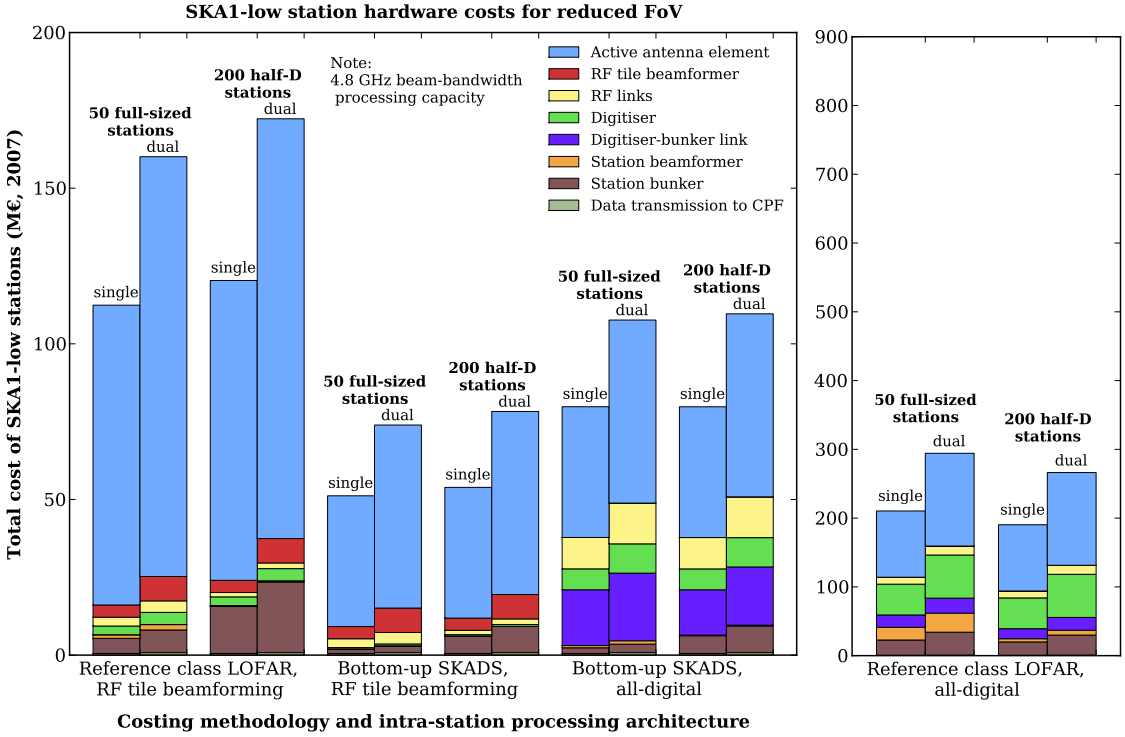


Figure F.6: SKA_1-low station hardware cost for a beam–bandwidth product limited to 4.8 GHz, which is equivalent to a processed FoV of 20 deg^2 over a 70–180 MHz band. Other details as per Figure F.3.

only depends on the cost data source and intra-station architecture, but given the dominance of the active antenna element costs, the cost multipliers (Appendix F.2.9) used to determine the active antenna element costs for the dual-band implementation will have a strong influence.

F.7.3 System implications

A fixed beam–bandwidth product also has implications for the downstream processing. The processing capacity of the correlator and imaging processor sub-systems scales linearly with the beam–bandwidth product (see Appendix F.4.2). The cost of these sub-systems is thus 4.8/80 = 6.0% of the single-band costs listed in Table F.6 (p 203). Because the processing capacity of these sub-systems is now defined by the same beam–bandwidth product for each implementation, the cost does not change between the single and dual-band implementation. However, the cost scaling effects due N_{st} vs. $N_{e/st}$ trade (Appendix F.6.2) still apply: more stations of smaller diameter increase the correlator and imaging processor costs.

Figure F.7 shows the significant variable costs for SKA_1-low. For all scenarios in Figure F.7, the single-band implementation is cheaper than the dual-band, at approximately 80 % of the dual-band cost. The station hardware, correlator and imaging processor costs are smaller than when a processed FoV of 20 deg^2 over 70–450 MHz is required (Figure F.7); the latter two are no longer significant costs. The antenna element deployment and site preparation costs do not change, because the number and location of the antenna elements is independent of changes to

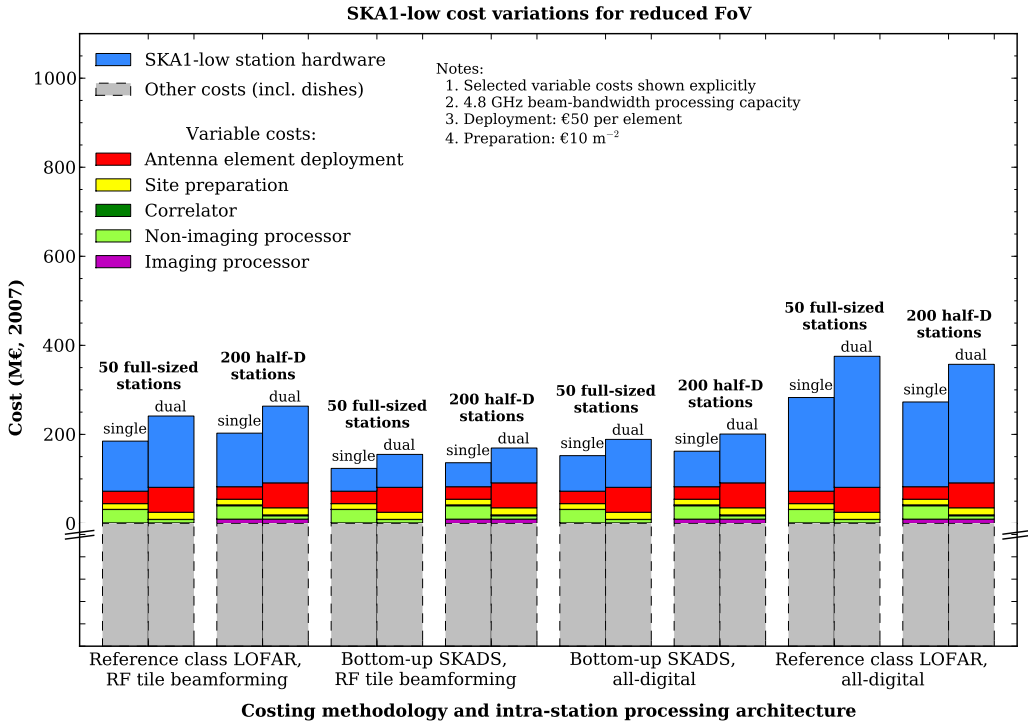


Figure F.7: Comparison of significant variable costs (excluding power) for a beam–bandwidth product limited to 4.8 GHz. Other details as per Figure F.4.

the beam–bandwidth product. The non-imaging processor also remains unchanged, because the strawman presented in this section still has a separate high-band core, with half the diameter of the single-band core.

Appendix G

Parametric models and costs for SKA long baselines

This appendix provides additional details for the SKA long baseline data network (Chapter 6). I model the data network with a small number of scalable blocks. Although the approach is similar to the SKA_1-low modelling (Appendix F), I use a simple hierarchical block structure here. The equations in the blocks use the input parameters from Tables 6.1 and G.1; the blocks themselves are summarised in Table G.2 and described in the sub-sections below.

Although dense wavelength division multiplexing networking equipment can be considered as common off-the-shelf products, the present commercial environment limits cost data availability. For example, actual costs are usually confidential and placing the vendor-specific products into common categories is difficult (Gunkel et al., 2006). Using data collected from a European research project, Huelsermann et al. (2008) presents normalised costs of DWDM components, developed as reference class estimates from equipment vendors and network operators. I use these costs as my primary data source. I convert normalised costs to euros using a low and high cost estimate, where a normalised unit cost in Huelsermann et al. (2008) is equal to €1000 (low) and €5000 (high). The cost of components in the low estimate are consistent, to a first-order approximation, with the costs used in McCool (2010) and the SKA Design Studies costing work in SKACost (Bolton et al., 2009a). Costs are summarised in Table G.3.

G.1 Remote links

A data link for each of the N_{RSt} remote stations is costed. The cost of each link, $C_{\text{remote link}}$, is distance-dependent and calculated independently.

Table G.1: Summary of network design parameters.

Design parameters	Values modelled	Comment
Channel speed R_{ch}	10, 40, 100 Gbps	Determines transponder cost and fibre utilisation.
Shared link cost multiplier $K_{\text{shared - link}}$	1, 10	See shared links block.
Number of channels per fibre $N_{\text{optic - ch}}$	40	
OEO distance d_{optic}	480 km	Average reach of optical long-haul links, after which electrical regeneration is required.
Routing overhead OH_{routing}	1.5	Overhead for sub-optimal routing.
Span distance	80 km	Distance between optical amplifiers.

Table G.2: Summary of blocks and scaling for SKA long baseline data network.

Block name	Parametric equation	Block coverage
Remote links	$N_{\mathrm{RSt}} C_{\text{remote link}}$	
↪Remote link	$C_{\text{data transmission}} + C_{\text{fibre tail}} + C_{\text{shared links}}$	
↪Data transmission	$C_{\mathrm{fix}} + 2N_{\text{optic-sect}}(N_{\text{data-ch}} C_{\mathrm{var}} + C_{\mathrm{var[10\,G]}})$	C_{fix}: multiplexer[a] C_{var}: transponder $C_{\mathrm{var[10\,G]}}$: transponder for monitor and control
↪Fibre tail	C_{fix}	C_{fix}: fibre tail
↪Shared links	$N_{\mathrm{span}}(N_{\text{data-ch}} + 1) K_{\text{shared-link}} C_{\mathrm{var}} / N_{\mathrm{ch/fibre}}$	C_{var}: optical amplifier, dispersion compensation, multiplexer, incremental component housing, fibre cable

[a] Multiplexer at remote station. Other multiplexers are included in the shared links.

Table G.3: Long baseline data network costs in € (2007). Mean cost is calculated from the low and high costs, assuming a top-hat (uniform) probability distribution function.

Block name	Low	High	Mean	Block unit
Data transmission	C_{fix}: 4 k $C_{\mathrm{var[10\,G]}}$: 1 k $C_{\mathrm{var[40\,G]}}$: 4 k $C_{\mathrm{var[100\,G]}}$: 10 k	C_{fix}: 20 k $C_{\mathrm{var[10\,G]}}$: 5 k $C_{\mathrm{var[40\,G]}}$: 20 k $C_{\mathrm{var[100\,G]}}$: 50 k	C_{fix}: 12 k $C_{\mathrm{var[10\,G]}}$: 3 k $C_{\mathrm{var[40\,G]}}$: 12 k $C_{\mathrm{var[100\,G]}}$: 30 k	per remote station
Fibre tail	C_{fix}: 250 k	C_{fix}: 250 k	C_{fix}: 250 k	per remote station
Shared links	C_{var}: 18 k per fibre pair per span	C_{var}: 38 k per fibre pair per span	C_{var}: 28 k per fibre pair per span	per remote station

The model is representative of the layouts developed for the SKA site submissions, where the 180–3000 km region was divided into 25 logarithmically spaced annuli and a remote station placed within each annulus (Millenaar et al., 2011); the distance d_{RSt} between the core and a remote station is independent of geographical location. Following the models in SKACost by Bolton et al. (2009a), I calculate d_{RSt} for the i^{th} remote station as

$$d_{\mathrm{RSt}}(i) = d_{\mathrm{start}} r^i, \tag{G.1}$$

where the growth rate r is

$$r = \left(\frac{d_{\mathrm{end}}}{d_{\mathrm{start}}} \right)^{1/N_{\mathrm{RSt}}}. \tag{G.2}$$

The actual link length depends on the fibre optic routes available. To account for the routed link distance being longer than the straight-line distance, I include a 50 % overhead: $OH_{\mathrm{routing}} = 1.5$.

The cost of the DWDM terminal and aggregate link blocks vary as a function of the data rate

from the remote station. The number of channels required to carry the astronomical data of rate R_{RSt} from the remote station is

$$N_{\mathrm{data\,-\,ch}} = \lceil R_{\mathrm{RSt}}/R_{\mathrm{ch}} \rceil \,, \tag{G.3}$$

where R_{ch} is the channel speed of the transponder, and

$$R_{\mathrm{RSt}} = N_{\mathrm{b\,-\,RSt}} \Delta\nu N_{\mathrm{pol}} N_{\mathrm{bit\,-\,CFB}} \times OS \times OH \times \mathrm{Nyq.}, \tag{G.4}$$

where $N_{\mathrm{b\,-\,RSt}}$ is the number of remote station beams formed and $\Delta\nu$ is the processed bandwidth. From the signal transport and networks high-level description (McCool, 2011b), the number of polarisations $N_{\mathrm{pol}} = 2$, the number of bits per sample out of the coarse channel filter-bank $N_{\mathrm{bit\,-\,CFB}} = 4$, oversampling $OS = 1.2$, digital encoding overhead $OH = 1.25$ (any frame overheads are excluded) and the data is Nyquist sampled: Nyq. = 2.

Even though the astronomical data traffic is unidirectional, the channels represents duplex (bidirectional) traffic, as this is the most common network implementation (Simmons, 2008). Duplex components are costed in Huelsermann et al. (2008); the duplex components generally encompass two unidirectional devices, each acting on a single fibre in the fibre pair. Although potentially cheaper, a unidirectional device may not be available as a common off-the-shelf component, nor compatible with the carrier network, and therefore is not considered here.

For each remote link, an additional duplex channel is costed to allow for monitor and control traffic.

G.2 Data transmission

This block costs the digital transmission of signals as channels (wavelengths) on a single fibre using dense wavelength division multiplexing (DWDM) technology. The principal transmission components are located at each end of an optical section, consisting a multiplexer (or demultiplexer) and associated signal amplification for each fibre, and a transponder for each channel being multiplexed onto the fibre.

The cost of the multiplexer at the remote station terminal (C_{fix}) is included in this block. The cost of the multiplexer at the central processing facility terminal is shared amongst the channels from a number of remote stations, and is therefore not significant.

The total number of optical sections $N_{\mathrm{optic\,-\,sect}}$ per remote link, hence number of transponders required, depends on the link length d_{RSt}:

$$N_{\mathrm{optic\,-\,sect}} = \left\lceil \frac{d_{\mathrm{RSt}} OH_{\mathrm{routing}}}{d_{\mathrm{optic}}} \right\rceil , \tag{G.5}$$

where $OH_{\mathrm{routing}} = 1.5$ is the overhead due to sub-optimal routing mentioned previously and d_{optic} is the 'optical reach', the length of the optical section. I use $d_{\mathrm{optic}} = 480\,\mathrm{km}$ as likely optical reach for the carrier networks in southern Africa.

For each of the $N_{\mathrm{data\,-\,ch}}$ channels transmitted from the remote station, a transponder of channel speed R_{ch} is costed for each of the two end-points. In addition, a 10 G transponder at each end-point is costed for monitor and control traffic.

When the signal is regenerated through conversion from optical to electrical to optical (OEO), a similar set of transponders are required, but placed back-to-back. Two transponders are costed for each channel of data.

The parametric cost equation is thus

$$C_{\text{data transmission}} = C_{\text{fix}} + 2N_{\text{optic-sect}}(N_{\text{data-ch}}C_{\text{var}} + 2C_{\text{var[10 G]}}). \tag{G.6}$$

The derived costs are €1 k (low) and €5 k (high) for a 10 G for long-haul transponder ($C_{\text{var[10 G]}}$), and approximately €4 k (low) and €20 k (high) for a 40 channel, long-haul multiplexer that includes amplifiers (C_{fix}).

The higher speed transponders are not significantly cheaper on a cost per Gbps basis. For example, in Huelsermann et al. (2008), the 40 G transponder cost is 3.75 times the cost of the 10 G transponder. Simmons (2008) estimates a factor of 2.5–3 cost increase once the technology is mature, and a 1.5–2 increase for the 40 G to 100 G transponder. For simplicity, I use a factor of 4 increase for the 40 G transponder and 10 for the 100 G transponder, resulting in the same unit cost (€/Gbps) for all three channel speeds. Therefore, the primary benefit of the higher transponder speeds in this model is that the limited number of channels available on the carrier fibre network are better utilised.

G.3 Fibre tail

The fibre tail, or 'last mile', describes the fibre and trenching from the carrier fibre link to the remote station. It also includes the optical add-drop multiplexer (OADM) to connect specific DWDM channels to or from the carrier network. Although the cost of a fibre tail is site specific, ANZSCC (2011) costs the 25 remote station fibre tails to the nearest carrier fibre at €5 892 500. That costing includes only the incremental components to house and integrate the fibre-optical tails onto the carrier's network, which is consistent with the assumptions in the present analysis. I therefore use an approximate average cost of €250 k per tail. The remote stations are assumed to be located such that the fibre tails are less than 80 km from the carrier network connection, and therefore do not require additional amplification.

G.4 Shared links

Once the data from the remote stations are added to the carrier's fibre cables, the cost of carrying the data is no longer exclusive to that remote link, but is shared between all network traffic on that fibre cable. As a first-order approximation, I estimate the *per fibre* costs, and I apportion these costs to each channel transmitted on the fibre, to give a *per channel* cost. The principal per fibre costs are distance based: additional to the fibre cable itself, an optical amplifier, dispersion compensation and a hut (telecommunications shelter) is required for every 80–100 km span. Trenching costs are excluded, because the network infrastructure is assumed to already exist (see Section 6.2).

The shared link block averages these per fibre costs over $N_{\text{optic-ch}}$ channels, being the number of channels available per fibre. If the costs are equally apportioned to each channel, the cost to

each remote link is

$$C_{\text{shared links}} = N_{\text{span}}(N_{\text{data-ch}} + 1)K_{\text{shared-link}}C_{\text{var}}/N_{\text{optic-ch}}, \tag{G.7}$$

where

$$N_{\text{span}} = \left\lceil \frac{d_{\text{RSt}}OH_{\text{routing}}}{80\,\text{km}} \right\rceil \tag{G.8}$$

is the number of 80 km spans along each link and $K_{\text{shared-link}}$ is a cost multiplier to account for higher shared link costs, as outlined in Section 6.2. It also accounts for under-utilised links, as not every link will use all $N_{\text{optic-ch}}$ available channels.

Even for a cost multiplier of $K_{\text{shared-link}} = 1$, the actual cost of the shared link depends on the detailed design, and involves trade-offs between the cost and utilisation of the network components and the cost and availability of fibre cables and infrastructure. For the purposes of this approximation, $N_{\text{optic-ch}} = 40$. Although there are DWDM systems that carry 80 or 120 channels per fibre (Simmons, 2008), these are generally utilised on the longer-distance 'backbone' fibre links, that are likely to be less relevant to this work.

The cost coefficient C_{var} for this block is a bottom-up summation of reference class cost data. From the models in Huelsermann et al. (2008), the cost of the duplex optical amplifier and dispersion compensation components is ∼€2.5 k (low) and ∼€12.5 k (high) per span. A 24 core fibre optic cable costed in ANZSCC (2011) equates to ∼€80 per fibre pair per km. Doubling these values allows for costs associated with infrastructure and deployment, such as space for the optical amplifiers and termination of fibre cables. The resulting cost estimates are $C_{\text{var}} \approx$ €18 k (low) and $C_{\text{var}} \approx$ €38 k (high) per fibre pair per 80 km span.

Bibliography

Alexander, P. 2011, AA Processing Requirements for SKA1, AAVP 2011, Dwingeloo, the Netherlands, 12-16 December, http://www.astron.nl/aavp2011

Alexander, P., Bregman, J. D., & Faulkner, A. J. 2009, in Widefield Science and Technology for the SKA: SKADS Conference, ed. S. A. Torchinsky, A. van Ardenne, A. van den Brink-Havinga, A. J. van Es, & A. J. Faulkner, Château de Limelette, Belgium

Alexander, P., Faulkner, A. J., Grainge, K., & Lazio, J. 2012, Analysis of requirements derived from the DRM, Software and Computing CoDR document WP2-050.020.010-RR-001 Rev E, SPDO, http://wiki.skatelescope.org/bin/view/SoftwareComputing

Alexander, P., & Hall, P. 2010, Design issues and implementation challenges, AAVP 2010, Cambridge, UK, 8-10 December, http://www.mrao.cam.ac.uk/projects/aavp/index.html

Alexander, P., et al. 2007, SKADS benchmark scenario design and costing, SKA Memo 93

ANZSCC. 2011, Request for Information from the Candidate SKA Sites, Site documentation, Australia–New Zealand SKA Coordination Committee, http://www.skatelescope.org/the-location/site-documentation

Bagchi, M., Nieves, A. C., & McLaughlin, M. 2012, Monthly Notices of the RAS, 425, 2501

Balanis, C. A. 2005, Antenna theory, 3rd edn. (New Jersey, USA: John Wiley & Sons)

Barott, W. C., Milgrome, O., Wright, M., MacMahon, D., Kilsdonk, T., Backus, P., & Dexter, M. 2011, Radio Science, 46, RS1016

Barr, E. 2011, in American Institute of Physics Conference Series, Vol. 1357, American Institute of Physics Conference Series, ed. M. Burgay, N. D'Amico, P. Esposito, A. Pellizzoni, & A. Possenti, 52–53

Batten, J. 2008, in Topical Workshop on Electronics for Particle Physics, ed. E. Dho & F. Vasey, Naxos, Greece, 53–60

Bhat, N. D. R. 2011, Bulletin of the Astronomical Society of India, 39, 353

Bhat, N. D. R., Cordes, J. M., Camilo, F., Nice, D. J., & Lorimer, D. R. 2004, Astrophysical Journal, 605, 759

Bhat, N. D. R., Cordes, J. M., Chatterjee, S., & Lazio, T. J. W. 2005, Radio Science, 40, 5

Bhat, N. D. R., Cordes, J. M., Cox, P. J., Deneva, J. S., Hankins, T. H., Lazio, T. J. W., & McLaughlin, M. A. 2011, Astrophysical Journal, 732, 14

Bhat, N. D. R., Tingay, S. J., & Knight, H. S. 2008, Astrophysical Journal, 676, 1200

Bhat, N. D. R., et al. 2007, The Astrophysical Journal, 665, 618

Bij de Vaate, J.-G., et al. 2011, AA Concept Descriptions, AA CoDR document WP2-010.020.010-TD-001 Rev E, SPDO, http://wiki.skatelescope.org/bin/view/ApertureArrays/AACoDR

Binggeli, B., Sandage, A., & Tammann, G. A. 1985, Astronomical Journal, 90, 1681

Bird, T., & Cortes-Medellin, G. 2003, in IEEE Antennas and Propagation Society International Symposium, 116–119

Bolton, R. C., Alexander, P., Ford, D. C., Colegate, T. M., & Hall, P. J. 2009a, in Widefield Science and Technology for the SKA: SKADS Conference, ed. S. A. Torchinsky, A. van Ardenne, A. van den Brink-Havinga, A. J. van Es, & A. J. Faulkner, Château de Limelette, Belgium

Bolton, R. C., Millenaar, R., & Harris, G. D. 2011, SKA configurations design, System Delta CoDR document WP3-050.020.000-R-002 Rev A, SPDO, http://wiki.skatelescope.org/bin/view/SystemEngineering/SysDeltaCoDRDocs

Bolton, R. C., Scaife, A., Grigorescu, G., Millenaar, R., & Lobanov, A. 2009b, in Widefield Science and Technology for the SKA: SKADS Conference, ed. S. A. Torchinsky, A. van Ardenne, A. van den Brink-Havinga, A. J. van Es, & A. J. Faulkner, Château de Limelette, Belgium

Bolton, R. C., et al. 2009c, SKADS benchmark scenario design and costing 2 (The SKA Phase 2 AA scenario), SKA Memo 111

Bradt, H. 2003, Astronomy Methods: A Physical Approach to Astronomical Observations (Cambridge, UK: Cambridge University Press)

Braun, R. 1996, in The Westerbork Observatory, Continuing Adventure in Radio Astronomy, ed. E. Raimond & R. Genee, Netherlands Foundation for Research in Astronomy (Kluwer Academic Publishers), 167

Braun, R., & van Cappellen, W. 2006, Aperture Arrays for the SKA: Dense or Sparse?, SKA Memo 87

Bregman, J. D. 2000, in Proc. SPIE Astronomical Telescopes and Instrumentation, Vol. 4015, Radio Telescopes, ed. H. R. Butcher, 19–32

Bregman, J. D. 2004, Experimental Astronomy, 17, 365

Bunton, J. D. 2000, An Improved FX Correlator, ALMA Memo 342, http://www.alma.nrao.edu/memos/html-memos/alma342/memo342.pdf

—. 2003a, Figure of Merit for SKA survey speed, SKA Memo 40

—. 2003b, Multi-resolution FX Correlator, ALMA Memo 447, http://www.alma.nrao.edu/memos/html-memos/alma447/memo447.pdf

—. 2010, Strawman SKA Correlator, SKA Memo 126

Bunton, J. D., & Hay, S. G. 2004, Experimental Astronomy, 17, 381

Bunton, J. D., & Hay, S. G. 2010, in International Conference on Electromagnetics in Advanced Applications (ICEAA), 728 –730

Burke-Spolaor, S., & Bailes, M. 2010, MNRAS, 402, 855

Burke-Spolaor, S., Bailes, M., Ekers, R., Macquart, J., & Crawford, III, F. 2011, Astrophysical Journal, 727, 18

Camilo, F., Ransom, S. M., Halpern, J. P., Reynolds, J., Helfand, D. J., Zimmerman, N., & Sarkissian, J. 2006, Nature, 442, 892

Carilli, C. L., & Rawlings, S. 2004, New Astronomy Reviews, 48, 979

Casani, J., Ballhaus, W., Dorfman, S., Gallagher, D., Illingworth, G., Klineberg, J., & Schurr, D. 2010, James Webb Space Telescope (JWST) Independent Comprehensive Review Panel (ICRP) Final Report, Tech. rep., NASA, http://www.nasa.gov/pdf/499224main_JWST-ICRP_Report-FINAL.pdf

Chippendale, A. P., Colegate, T. M., & O'Sullivan, J. D. 2007, SKAcost: a tool for SKA cost and performance estimation, SKA Memo 92

Christiansen, W. N., & Högbom, J. A. 1969, Radiotelescopes, 1st edn. (Cambridge: Cambridge University Press)

Clarke, N. L., D'Addario, L., Navarro, R., Cheng, T.-H., & Trinh, J. 2011, An Architecture for Incoherent Dedispersion, CRAFT Memo 6, https://pm.atnf.csiro.au/askap/projects/sup-craft

Cloete, K. 2011, Strategy to proceed to the next phase, System Delta CoDR document WP2-005.010.030-PLA-002 Rev A, SPDO, http://wiki.skatelescope.org/bin/view/SystemEngineering/SysDeltaCoDRDocs

—. 2012, Updated summary of the results of the Preconstruction Phase Stage 1 Expression of Interest (EoI), Tech. Rep. MGT-001.005.015-MR-003 Rev 2, SKA Organisation, http://www.skatelescope.org/the-project/eoi-preconstruction-2/eoi-request/

Colegate, T. M. 2009, in The Tenth Postgraduate Electrical Engineering and Computing Symposium, Perth, Australia

Colegate, T. M., & Clarke, N. 2011, Publications of the Astron. Soc. of Australia, 28, 299

Colegate, T. M., Hall, P. J., & Gunst, A. W. 2012, Cost-effective aperture arrays for SKA Phase 1: single or dual-band?, SKA Memo 140

Conway, J. E., Cornwell, T. J., & Wilkinson, P. N. 1990, Monthly Notices of the RAS, 246, 490

Cordes, J. M. 2009a, The SKA as a Radio Synoptic Survey Telescope: Widefield Surveys for Transients, Pulsars and ETI, 14th edn., SKA Memo 97

—. 2009b, Survey metrics, 1st edn., SKA Memo 109

Cordes, J. M., Bhat, N. D. R., Hankins, T. H., McLaughlin, M. A., & Kern, J. 2004a, Astrophysical Journal, 612, 375

Cordes, J. M., Kramer, M., Lazio, T. J. W., Stappers, B. W., Backer, D. C., & Johnston, S. 2004b, New Astronomy Review, 48, 1413

Cordes, J. M., & Lazio, T. J. W. 2002, arXiv:astro-ph/0207156

Cordes, J. M., Lazio, T. J. W., & McLaughlin, M. A. 2004, New Astronomy Reviews, 48, 1459

Cordes, J. M., & McLaughlin, M. A. 2003, The Astrophysical Journal, 596, 1142

Cordes, J. M., & SWG. 2006, Discovery and Understanding with the SKA, SKA Memo 85

Cordes, J. M., et al. 2006, Astrophysical Journal, 637, 446

Cornwell, T. J. 2004, Experimental Astronomy, 17, 329

—. 2005, SKA computing costs for a generic telescope model, SKA Memo 64

—. 2012, Visibility processing, Software and Computing CoDR document WP2-050.020.010-SR-001 Rev C, SPDO, http://wiki.skatelescope.org/bin/view/SoftwareComputing

Cornwell, T. J., Golap, K., & Bhatnagar, S. 2008, IEEE Journal of Selected Topics in Signal Processing, 2, 647

Cornwell, T. J., & Perley, R. A. 1992, Astronomy and Astrophysics, 261, 353

Cornwell, T. J., Voronkov, M. A., & Humphreys, B. 2012, Wide field imaging for the Square Kilometre Array, submitted (arXiv:1207.5861)

Craeye, C., & González-Ovejero, D. 2011, Radio Science, 46

Crane, P. C., & Napier, P. J. 1989, in Astronomical Society of the Pacific Conference Series, Vol. 6, Synthesis Imaging in Radio Astronomy, ed. R. A. Perley, F. R. Schwab, & A. H. Bridle, 139–165

Crosby, P. 2012a, International Journal of Information Technology Project Management, 3, 1

—. 2012b, PhD thesis, Curtin University

D'Addario, L. 2010, ASKAP Surveys for Transients: Which Observing Mode is Best?, SKA Memo 123

de Vos, M., Gunst, A. W., & Nijboer, R. 2009, IEEE Proceedings, 97, 1431

DeBoer, D. R., et al. 2009, IEEE Proceedings, 97, 1507

Deller, A. T., et al. 2011, Publications of the ASP, 123, 275

Deneva, J. S., et al. 2009, Astrophysical Journal, 703, 2259

Dewdney, P. E. 2010, SKA Science–Technology Trade-Off Process, System CoDR document WP2-005.010.030-MP-004 Rev 1.2, SPDO, http://wiki.skatelescope.org/bin/view/SystemEngineering

Dewdney, P. E., Bij de Vaate, J.-G., Cloete, K., Gunst, A. W., Hall, D., McCool, R., Roddis, N., & Turner, W. 2010a, SKA Phase 1: Preliminary System Description, SKA Memo 130

Dewdney, P. E., Hall, D., McCool, R., Roddis, N., Turner, W., Gunst, A., University of Cambridge, & CSIRO. 2011a, SKA1: High level system description, System Delta CoDR document WP2-005.030.010-TD-002 Rev A, SPDO, http://wiki.skatelescope.org/bin/view/SystemEngineering/SysDeltaCoDRDocs

Dewdney, P. E., Hall, P. J., Schilizzi, R. T., & Lazio, T. J. L. W. 2009, Proceedings of the IEEE, 97, 1482

Dewdney, P. E., Lonsdale, C., Roovers, R., & Williams, R. 2011b, SKA Aperture Arrays Concept Design Review Panel Report, AA CoDR document, SPDO, http://wiki.skatelescope.org/bin/view/ApertureArrays/AACoDR

Dewdney, P. E., et al. 2010b, High-level SKA system description, System CoDR document WP2-005.030.010-TD-001 Rev A, SPDO, http://wiki.skatelescope.org/bin/view/SystemEngineering

Djorgovski, S. G., Mahabal, A. A., Drake, A. J., Graham, M. J., & Donalek, C. 2012, Planets, Stars and Stellar Systems, Vol. 2, Sky Surveys, ed. H. Bond (Springer Verlag, in press (arXiv1203.5111))

DRAO, ed. 1991, The Radio Schmidt telescope: proceedings of a workshop held at Penticton, 1989 October 11-12 (Dominion Radio Astrophysical Observatory, Herzberg Institute of Astrophysics, National Research Council of Canada)

Duffy, A. R., Meyer, M. J., Staveley-Smith, L., Bernyk, M., Croton, D. J., Koribalski, B. S., Gerstmann, D., & Westerlund, S. 2012, Monthly Notices of the RAS, 426, 3385

Dulwich, F., Mort, B. J., Salvini, S., Zarb Adami, K., & Jones, M. E. 2009, in Widefield Science and Technology for the SKA: SKADS Conference, ed. S. A. Torchinsky, A. van Ardenne, A. van den Brink-Havinga, A. J. van Es, & A. J. Faulkner, Château de Limelette, Belgium

Duren, R. M. 2006, in IEEE Aerospace Conference, Big Sky, Montana, 13

Ekers, R. D. 2003, in Astronomical Society of the Pacific Conference Series, Vol. 295, Astronomical Data Analysis Software and Systems XII, ed. H. E. Payne, R. I. Jedrzejewski, & R. N. Hook, 125

Ekers, R. D. 2009, in "Accelerating the Rate of Astronomical Discovery IAU GA", PoS(sps5)007, Rio de Janeiro, Brazil

Ellingson, S. W., Clarke, T. E., Cohen, A., Craig, J., Kassim, N. E., Pihlstrom, Y., Rickard, L. J., & Taylor, G. B. 2009, IEEE Proceedings, 97, 1421

Faulkner, A. J., Bij de Vaate, J. G., & van Ardenne, A. 2011, AA deployment and operation, AA CoDR document WP2-010.020.010-TR-001 Rev F, SPDO, http://wiki.skatelescope.org/bin/view/ApertureArrays/AACoDR

Faulkner, A. J., et al. 2010, Aperture Arrays for the SKA: the SKADS White Paper, SKA Memo 122

Fender, R. P., & Bell, M. E. 2011, Bulletin of the Astronomical Society of India, 39, 315

Flyvbjerg, B. 2008, European Planning Studies, 16, 3

Flyvbjerg, B., Bruzelius, N., & Rothengatter, W. 2003, Megaprojects and risk: an anatomy of ambition (Cambridge University Press)

Flyvbjerg, B., & COWI. 2004, Procedures for Dealing with Optimism Bias in Transport Planning: Guidance Document, Tech. Rep. 58924, The British Department for Transport

Flyvbjerg, B., Holm, M. S., & Buhl, S. 2002, Journal of the American Planning Association, 68, 279

Ford, D., Bolton, R. C., Colegate, T., Alexander, P., & Hall, P. 2009, in Widefield Science and Technology for the SKA: SKADS Conference, ed. S. A. Torchinsky, A. van Ardenne, A. van den Brink-Havinga, A. J. van Es, & A. J. Faulkner, Château de Limelette, Belgium

Ford, D., Bolton, R. C., Colegate, T., Alexander, P., & Hall, P. 2010, The SKA Costing and Design Tool, SKA Memo 120

Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006, Physics Reports, 433, 181

Gaensler, B. M. 2004, Key Science Projects for the SKA, SKA Memo 44

GAO. 2009, GAO Cost Estimating and Assessment Guide, Tech. Rep. GAO-09-3SP, Government Accountability Office, http://www.gao.gov/new.items/d093sp.pdf

—. 2011, NASA: Assessments of Selected Large-Scale Projects, Tech. Rep. GAO-11-239SP, Government Accountability Office, http://www.gao.gov/products/GAO-12-207SP

Garrett, M. A. 1999, in Perspectives on Radio Astronomy - Science with Large Antenna Arrays, ed. M. P. van Haarlem, 139

Garrett, M. A., Cordes, J. M., Deboer, D. R., Jonas, J. L., Rawlings, S., & Schilizzi, R. T. 2010, A Concept Design for SKA Phase 1 (SKA1), SKA Memo 125

Godfrey, L., Bignall, H., & Tingay, S. 2011, Very High Angular Resolution Science with the SKA, SKA Memo 135

Graham, R. L., Lubachevsky, B. D., Nurmela, K. J., & Ostergard, P. R. J. 1998, Discrete Mathematics, 181, 139

Greenhill, L. J., & Bernardi, G. 2012, arXiv:1201.1700

Gringeri, S., Basch, E. B., & Xia, T. J. 2012, IEEE Communications Magazine, 50, S21

Gunkel, M., et al. 2006, in PS '06. International Conference on Photonics in Switching

Gunst, A. W. 2007, LOFAR Architectural Design Document of the Astronomical Applications, Tech. Rep. LOFAR-ASTRON-ADD-006 Rev 5, ASTRON, http://www.lofar.org/wiki/doku.php?id=public:documents:lofar_documents

Gunst, A. W., & Bentum, M. J. 2010, in Phased Array Systems and Technology (ARRAY), 2010 IEEE International Symposium on, 632–639

Hall, P. J. 2004a, SKA Engineering Report, International SKA Conference 2004, Penticton, Canada, 18-22 July, http://www2.skatelescope.org/pages/conf_pentincton.htm

—. 2004b, Experimental Astronomy, 17, 5

Hall, P. J. 2009, in Widefield Science and Technology for the SKA: SKADS Conference, ed. S. A. Torchinsky, A. van Ardenne, A. van den Brink-Havinga, A. J. van Es, & A. J. Faulkner, Château de Limelette, Belgium, 405–406

Hall, P. J. 2011, in Proceedings of the XXXth General Assembly of International Union of Radio Science (URSI), Istanbul

Hall, P. J., Beresford, R. J., Chippendale, A. P., Ferris, D., Jackson, C. A., James, G., & Wieringa, M. 2002, Eyes on the Sky: A Refracting Concentrator Approach to the SKA, SKA Memo 22

Hall, P. J., Colegate, T., Macquart, J.-P., Clarke, N., Tingay, S., Trott, C., & Wayth, R. 2012, in Resolving the Sky - Radio Interferometry: Past, Present and Future, Manchester, UK, in press

Hall, P. J., Schilizzi, R. T., Dewdney, P. E., & Lazio, T. J. W. 2008, The Radio Science Bulletin, 326

Hankins, T. H., Kern, J. S., Weatherall, J. C., & Eilek, J. A. 2003, Nature, 422, 141

Hankins, T. H., & Rickett, B. J. 1975, in Methods in Computational Physics, ed. B. Alder, S. Fernbach, & M. Rotenberg, Vol. 14, 55–129

Hansen, P. 2012, The case for coherent-transponder subsystems, Editorial guide, Lightwave, http://www.lightwaveonline.com/editorial-guides/2012/05/what-new-in-100g.whitepaperpdf.render.pdf

Harwit, M. 1981, Cosmic Discovery (Basic Books)

Harwit, M. 1998, in IAU Symposium, Vol. 179, New Horizons from Multi-Wavelength Sky Surveys, ed. B. J. McLean, D. A. Golombek, J. J. E. Hayes, & H. E. Payne

—. 2003a, Physics Today, 56

—. 2003b, Astrophysical Journal, 597, 1266

Hayman, D. B. 2011, PhD thesis, Macquarie University

Hessels, J. W. T., Stappers, B. W., van Leeuwen, J., & Transients Key Science Project, L. 2009, in The Low-Frequency Radio Universe, ed. D. J. Saikia, D. Green, Y. Gupta, & T. Venturi

Hewish, A., Bell, S. J., Pilkington, J. D. H., Scott, P. F., & Collins, R. A. 1968, Nature, 217, 709

Holdaway, M. A. 1999, in Astronomical Society of the Pacific Conference Series, Vol. 180, Synthesis Imaging in Radio Astronomy II, ed. G. B. Taylor, C. L. Carilli, & R. A. Perley, 401

Horiuchi, S., Chippendale, A., & Hall, P. 2004, SKA system definition and costing: a first approach, SKA Memo 57

Howard, R. 2012, JWST Program Office, February 2012 APS meeting, http://science.nasa.gov/media/medialibrary/2012/02/23/Howard.pdf

Huelsermann, R., Gunkel, M., Meusburger, C., & Schupke, D. A. 2008, Journal of Optical Networking, 7, 814

Huynh, M., et al. 2012, Is There an Optimum Frequency Range for SKA1-lo? Question 1 of the Magnificent Memoranda II, SKA Memo 141

IEEE. 1993, IEEE Standard Definitions of Terms for Antennas, Tech. Rep. IEEE Std 145-1993, Institute of Electrical and Electronics Engineers

—. 2005, IEEE Standard for Application and Management of the Systems Engineering Process, Tech. Rep. IEEE Std 1220-2005, Institute of Electrical and Electronics Engineers

INCOSE. 2010, Systems Engineering Handbook, Tech. Rep. INCOSE-TP-2003-002-03.2, International Council on Systems Engineering

ISO/IEC. 2008, Systems and software engineering - System life cycle processes, Tech. Rep. ISO/IEC 15288:2008(E), International Organization for Standardization and International Electrotechnical Commission

ISPA. 2008, Parametric Estimating Handbook, Tech. rep., International Society of Parametric Analysts, http://www.ispa-cost.org/newbook.htm

ISPO. 2006, Reference design for the SKA, SKA Memo 69

—. 2007, SKA Infrastructure Development, SKA Memo 96

James, C. W., Ekers, R. D., Álvarez-Muñiz, J., Bray, J. D., McFadden, R. A., Phillips, C. J., Protheroe, R. J., & Roberts, P. 2010, Physical Review D, 81, 042003

Johnston, S., & Gray, A. 2006, Surveys with the xNTD and CLAR, SKA Memo 72

Johnston, S., van Straten, W., Kramer, M., & Bailes, M. 2001, Astrophysical Journal, Letters, 549, L101

Johnston, S., et al. 2008, Experimental Astronomy, 22, 151

Jones, D. L. 2004, SKA Science Requirements: Version 2, SKA Memo 45

Jones, M., Zarb Adami, K., Salvini, S., Faulkner, A., Khlebnikov, V., & Shenton, C. 2011, SKA station beamformer concept description, Signal processing CoDR document WP2-040.120.010-TD-001, SPDO, http://wiki.skatelescope.org/bin/view/SignalProcessing

Juswardy, B., Bij de Vaate, J. G., Schlagenhaufer, F., Padhi, S., & Hall, P. J. 2011, Towards Robust Solar-powered SKA-low Front-end, The Path to SKA-low Workshop, Perth, Western Australia, 6-9 September, http://ict.icrar.org/store/Presentations/skalow

Kaplan, D. L., Chatterjee, S., Gaensler, B. M., & Anderson, J. 2008, Astrophysical Journal, 677, 1201

Kaplan, D. L., Esposito, P., Chatterjee, S., Possenti, A., McLaughlin, M. A., Camilo, F., Chakrabarty, D., & Slane, P. O. 2009, Monthly Notices of the RAS, 400, 1445

Kartalopoulos, S. V. 2000, Introduction to DWDM Technology: data in a rainbow (IEEE Press)

Karuppusamy, R., Stappers, B. W., & van Straten, W. 2010, Astronomy and Astrophysics, 515, A36

Keane, E. F., Kramer, M., Lyne, A. G., Stappers, B. W., & McLaughlin, M. A. 2011, Monthly Notices of the RAS, 415, 3065

Keane, E. F., Ludovici, D. A., Eatough, R. P., Kramer, M., Lyne, A. G., McLaughlin, M. A., & Stappers, B. W. 2010, MNRAS, 401, 1057

Keane, E. F., Stappers, B. W., Kramer, M., & Lyne, A. G. 2012, Monthly Notices of the RAS, 425, L71

Keith, M. J., et al. 2010, Monthly Notices of the RAS, 409, 619

Kellermann, K. I., Cordes, J. M., Ekers, R. D., Lazio, J., & Wilkinson, P. 2009, in "Accelerating the Rate of Astronomical Discovery IAU GA", PoS(sps5)005, Rio de Janeiro, Brazil

Kellermann, K. I., et al. 2006, Report of the SKA Operations Working Group, SKA Memo 84

Khlebnikov, V. A., Zarb Adami, K., Armstrong, R. P., & Jones, M. E. 2010, in IEEE International Symposium on Phased Array Systems and Technology (ARRAY), 911–916

Kijak, J., Lewandowski, W., Maron, O., Gupta, Y., & Jessner, A. 2011, Astronomy and Astrophysics, 531, A16

Knight, H. S., Bailes, M., Manchester, R. N., Ord, S. M., & Jacoby, B. A. 2006, Astrophysical Journal, 640, 941

Knittel, G., & Horneffer, A. 2011, A scalable computer architecture for on-line pulsar search on the SKA, Signal processing CoDR document WP2-040.130.010-TD-002 Rev 1, SPDO, http://wiki.skatelescope.org/bin/view/SignalProcessing

Koay, J. Y. 2012, PhD thesis, Curtin University, submitted

Kramer, M., Backer, D. C., Cordes, J. M., Lazio, T. J. W., Stappers, B. W., & Johnston, S. 2004, New Astronomy Review, 48, 993

Kraus, J. D. 1986, Radio Astronomy, 2nd edn. (Cygnus-Quasar Book)

Kruchten, P. B. 1995, IEEE Software, 12, 42

Lal, D. V., Lobanov, A. P., & Jiménez-Monferrer, S. 2009, Array configuration studies for the Square Kilometre Array—Implementation of figures of merit based on spatial dynamic range, SKA Memo 107

Lobanov, A. P. 2012, in The Square Kilometre Array: Paving the way for the new 21st century radio astronomy paradigm, ed. D. Barbosa, S. Anton, L. Gurvits, & D. Maia, Astrophysics and Space Science Proceedings (Springer Berlin Heidelberg), 75–83

Lonsdale, C. J., & Cappallo, R. J. 1999, in Perspectives on Radio Astronomy: Technologies for Large Antenna Arrays, ed. A. B. Smolders & M. P. van Haarlem, 243

Lonsdale, C. J., et al. 2009, IEEE Proceedings, 97, 1497

Lorimer, D. R., Bailes, M., McLaughlin, M. A., Narkevic, D. J., & Crawford, F. 2007, Science, 318, 777

Lorimer, D. R., & Kramer, M. 2005, Handbook of Pulsar Astronomy (Cambridge University Press)

Lorimer, D. R., Yates, J. A., Lyne, A. G., & Gould, D. M. 1995, Monthly Notices of the RAS, 273, 411

Lundgren, S. C., Cordes, J. M., Ulmer, M., Matz, S. M., Lomatch, S., Foster, R. S., & Hankins, T. 1995, Astrophysical Journal, 453, 433

Maan, Y., et al. 2012, Astrophysical Journal, submitted (arXiv:1210.2573)

Macquart, J.-P. 2011, Astrophysical Journal, 734, 20

Macquart, J.-P., Hall, P. J., & Clarke, N. 2010a, in "International SKA Forum 2010 Science Meeting" PoS(ISKAF2010)03, Assen, the Netherlands

Macquart, J.-P., et al. 2010b, Publications of the Astron. Soc. of Australia, 27, 272

Mailloux, R. J. 1995, Phased array antenna handbook, 2nd edn. (Artech House)

Majid, W. A., Naudet, C. J., Lowe, S. T., & Kuiper, T. B. H. 2011, Astrophysical Journal, 741, 53

Malofeev, V. M., Gil, J. A., Jessner, A., Malov, I. F., Seiradakis, J. H., Sieber, W., & Wielebinski, R. 1994, Astronomy and Astrophysics, 285, 201

Manchester, R. N., et al. 2001, MNRAS, 328, 17

McCool, R. 2010, Data Transmission Cost Scaling for Long Baselines in the SKA, SKA Memo 119

—. 2011a, Network infrastructure concept description, Signal Transport and Network CoDR document WP2-030.080.000-TD.001 Rev 1, SPDO, http://wiki.skatelescope.org/bin/view/STaN/STaNCoDR

—. 2011b, STaN high level description, Signal Transport and Network CoDR document WP2-030.030.030-TD-001, SPDO, http://wiki.skatelescope.org/bin/view/STaN/STaNCoDR

McCool, R., Crosby, P., Hall, D., Colegate, T., Bolton, R., & Ford, D. 2010, SKA costing strategy (draft), System Delta CoDR document MGT-040.070.000-MP-001 Rev C, SPDO, http://wiki.skatelescope.org/bin/view/SKACosting/WebHome

McLaughlin, M. A., & Cordes, J. M. 2003, Astrophysical Journal, 596, 982

McLaughlin, M. A., et al. 2006, Nature, 439, 817

Mei, S., et al. 2007, Astrophysical Journal, 655, 144

Mellema, G., et al. 2012, arXiv:1210.0197

Millenaar, R. P., & Bolton, R. C. 2011, Figures of merit for SKA configuration analysis, Site documentation WP3-050.020.000-TR-001 Rev C, SPDO, http://www.skatelescope.org/the-location/site-documentation

Millenaar, R. P., Bolton, R. C., & Lazio, J. 2011, Array configurations for candidate SKA sites: design and analysis, Site documentation WP3-050.020.010-TR-001 Rev C, SPDO, http://www.skatelescope.org/the-location/site-documentation

Miyoshi, M., Moran, J., Herrnstein, J., Greenhill, L., Nakai, N., Diamond, P., & Inoue, M. 1995, Nature, 373, 127

Mollick, E. 2006, IEEE Annals of the History of Computing, 28, 62

Moran, J. M. 1993, Nature, 361, 17

Nakai, N., Inoue, M., & Miyoshi, M. 1993, Nature, 361, 45

NASA. 2007, NASA Systems Engineering Handbook, Tech. Rep. NASA/SP-2007-6105 Rev 1, NASA, http://hdl.handle.net/2060/20080008301

—. 2008, 2008 NASA Cost Estimating Handbook, Tech. rep., NASA, http://www.nasa.gov/pdf/263676main_2008-NASA-Cost-Handbook-FINAL_v6.pdf

Nijboer, R. J., Pandey-Pommier, M., & de Bruyn, A. G. 2009, LOFAR imaging capabilities and system sensitivity, SKA Memo 113

Noordam, J. E., Braun, R., & de Bruyn, A. G. 1991, EURO16: Proposal for an Array of 16 Low Cost 100 Meter Radio Telescopes, NFRA Note 585, ASTRON Netherlands Institute for Radio Astronomy, https://www.astron.nl/documents/Notes/ASTRON-NOTE-585.pdf

Noordam, J. E., & Smirnov, O. M. 2010, Astronomy and Astrophysics, 524, A61

Perini, F. 2011, Antenna network for AA-lo: Concept description, Signal Transport and Network CoDR document WP2-030.050.010-TD-002 Rev 1, SPDO, http://wiki.skatelescope.org/bin/view/STaN/STaNCoDR

Perley, R., & Clark, B. 2003, Scaling Relations for Interferometric Post-Processing, EVLA Memo 63, http://www.aoc.nrao.edu/evla/geninfo/memoseries/evlamemo63.pdf

Popov, M. V., & Stappers, B. 2007, Astronomy and Astrophysics, 470, 1003

Prasad, P., & Wijnholds, S. J. 2012, in 'New Windows on Transients across the Universe': A Discussion meeting of the Royal Society, London, UK

Reichhardt, T. 2006, Nature, 440, 140

Rohlfs, K., & Wilson, T. L. 2004, Tools of Radio Astronomy, 4th edn. (Springer)

Rubio-Herrera, E., Stappers, B. W., Hessels, J. T. W., & Braun, R. 2012, Monthly Notices of the RAS, in press (arXiv:1210.4660)

Sallmen, S., Backer, D. C., Hankins, T. H., Moffett, D., & Lundgren, S. 1999, Astrophysical Journal, 517, 460

Schilizzi, R. T., et al. 2007, Preliminary Specifications for the Square Kilometre Array, SKA Memo 100

Schilizzi, R. T., et al. 2011, Project execution plan: pre-construction phase for the Square Kilometre Array (SKA), Tech. Rep. MGT-001.005.005-MP-001 Rev K, SPDO, http://www.skatelescope.org/publications

Siemion, A. P. V., et al. 2012, Astrophysical Journal, 744, 109

Simmons, J. M. 2008, Optical Network Design and Planning, ed. B. Mukherjee, Optical Networks (Springer)

SKA Drafting Groups. 2012, Pre-construction phase, stage 1 work breakdown structure and statement of work, Tech. Rep. MGT-001.005.010-WBS-001 Rev H Draft, SKA Organisation, http://wiki.skatelescope.org/bin/view/Management/Stage1WBSandSOW

SKA SA. 2011, South African Response to the SSG Request for Information, Site documentation, SKA South Africa, http://www.skatelescope.org/the-location/site-documentation

Smits, R., Kramer, M., Stappers, B., Lorimer, D. R., Cordes, J., & Faulkner, A. 2009, Astronomy and Astrophysics, 493, 1161

Smits, R., Stappers, B., Kramer, M., & Karastergiou, A. 2011a, Pulsar survey with SKA Phase 1, Signal processing CoDR document WP2-040.030.010-TD-003 Rev 1, SPDO, http://wiki.skatelescope.org/bin/view/SignalProcessing

Smits, R., Tingay, S. J., Wex, N., Kramer, M., & Stappers, B. 2011b, Astronomy and Astrophysics, 528, A108

SOWG. 2012, Report of the SKA Siting Options Working Group, Tech. rep., SKA Organisation, http://www.skatelescope.org/the-location/site-documentation

SSWG. 2009, The Square Kilometre Array Design Reference Mission: SKA-mid and SKA-lo, Tech. Rep. V1.0, SPDO, http://www.skatelescope.org/publications

—. 2011, The Square Kilometre Array Design Reference Mission: SKA Phase 1 (draft), Tech. Rep. SCI-020.010.020-DRM-002 Rev 2.0, SPDO

—. 2012, The Square Kilometre Array Design Reference Mission: SKA Phase 1, Tech. Rep. SCI-020.010.020-DRM-002 Rev 3A, SPDO, http://www.skatelescope.org/publications

Staelin, D. H., & Reifenstein, E. C. 1968, Science, 162, 1481

Stappers, B. W., et al. 2011, Astronomy and Astrophysics, 530

Staveley-Smith, L., et al. 1996, Publications of the Astronomical Society of Australia, 13

Stevenson, T. J. 2011a, Phase 1 System Requirements Specification (SyRS), SKA System Delta CoDR, Manchester, UK, 23-25 February, http://www.skatelescope.org/public/2011-02_System_delta_CoDR_Documents

—. 2011b, SKA Phase 1 system requirements specification, System Delta CoDR document WP2-005.030.000-SRS-002 Rev B, SPDO, http://wiki.skatelescope.org/bin/view/SystemEngineering/SysDeltaCoDRDocs

—. 2011c, System engineering management plan, System Delta CoDR document WP2-005.010.030-MP-001 Rev F, SPDO, http://wiki.skatelescope.org/bin/view/SystemEngineering/SysDeltaCoDRDocs

Swarup, G. 1991, Current Science, 60, 106

Szomoru, A., Boven, P., Hargreaves, J., Pirruccio, S., Pogrebenko, S., Gunst, A., & Schoonderbeek, G. 2011, A Uniboard-based Phase 1 SKA correlator and beamformer concept description, Signal processing CoDR document WP2-040.070.010-TD-001 Rev 1, SPDO, http://wiki.skatelescope.org/bin/view/SignalProcessing

Taylor, A. R., & Braun, R., eds. 1999, Science with the Square Kilometer Array

Taylor, G. B., Carilli, C. L., & Perley, R. A., eds. 1999, Astronomical Society of the Pacific Conference Series, Vol. 180, Synthesis Imaging in Radio Astronomy II

Thidé, B., et al. 2007, Physical Review Letters, 99, 087701

Thompson, A. R., Moran, J. M., & Swenson Jr, G. W. 2001, Interferometry and Synthesis in Radio Astronomy, 2nd edn. (New York: Wiley)

Thompson, D. R., Wagstaff, K. L., Brisken, W. F., Deller, A. T., Majid, W. A., Tingay, S. J., & Wayth, R. B. 2011, Astrophysical Journal, 735, 98

Tingay, S. J., et al. 2012, Publications of the Astron. Soc. of Australia, in press (arXiv:1206.6945)

Turner, W. 2011, SKA signal processing costs, Signal processing CoDR document WP2-040.030.020-TD-001 Rev 1, SPDO, http://wiki.skatelescope.org/bin/view/SignalProcessing

Turner, W., Faulkner, A. J., Stappers, B. W., Ransom, S. M., Webster, R. L., Eatough, R. P., & Kramer, M. 2011, High-level SKA signal processing description, Signal processing CoDR document WP2-040.030.010-TD-001 Rev 1, SPDO, http://wiki.skatelescope.org/bin/view/SignalProcessing

van Ardenne, A., Bregman, J. D., van Cappellen, W. A., Kant, G. W., & Bij de Vaate, J. G. 2009, IEEE Proceedings, 97, 1531

van Cappellen, W. A., Wijnholds, S. J., & Bregman, J. D. 2006, in Proceedings of the 3rd European Radar Conference (EuRAD), 76–79

van Es, A., van Ardenne, A., Faulkner, A., Gunst, A., & bij de Vaate, J.-G. 2011, Aperture Array Risk Register, AA CoDR document WP2-010.020.010-RE-001 Rev C, SPDO, http://wiki.skatelescope.org/bin/view/ApertureArrays/AACoDR

van Langevelde, H. J. 2010, in "International SKA Forum 2010 Science Meeting" PoS(ISKAF2010)083, Assen, the Netherlands

van Leeuwen, J., & Stappers, B. W. 2010, Astronomy and Astrophysics, 509

Veidt, B. 2006, Focal-Plane Array Architectures: Horn Clusters vs. Phased-Array Techniques, SKA Memo 71

Walker, R. C. 1999, in Astronomical Society of the Pacific Conference Series, Vol. 180, Synthesis Imaging in Radio Astronomy II, ed. G. B. Taylor, C. L. Carilli, & R. A. Perley, 433

Wayth, R. B., Brisken, W. F., Deller, A. T., Majid, W. A., Thompson, D. R., Tingay, S. J., & Wagstaff, K. L. 2011, Astrophysical Journal, 735, 97

Wayth, R. B., Tingay, S. J., Deller, A. T., Brisken, W. F., Thompson, D. R., Wagstaff, K. L., & Majid, W. A. 2012, Astrophysical Journal, Letters, 753, L36

Weichenberg, G., Chan, V., & Medard, M. 2009, IEEE/OSA Journal of Optical Communications and Networking, 1, B81

Weinreb, S., & D'Addario, L. 2001, The SKA Cost Equation, SKA Memo 1

Welch, J., et al. 2009, IEEE Proceedings, 97, 1438

Wijnholds, S., van der Tol, S., Nijboer, R., & van der Veen, A. 2010, IEEE Signal Processing Magazine, 27, 30

Wijnholds, S. J. 2012, System Requirements Derived from Calibratability, AA Calibration and Calibratability Meeting 2012, Amsterdam, Netherlands, 12–13 July, http://www.astron.nl/AACal2012

Wijnholds, S. J., & Bregman, J. D. 2011, AA-lo station aspects illustrated by a fractal hexagonal structure, Tech. Rep. ASTRON-RP-468 Rev 1.1, ASTRON, http://wiki.skatelescope.org/bin/view/ApertureArrays/SystemDesign

Wijnholds, S. J., Bregman, J. D., & van Ardenne, A. 2011a, Radio Science, 46

Wijnholds, S. J., Nijboer, R., Grainge, K., & Bregman, J. D. 2011b, in Proceedings of the XXXth General Assembly of International Union of Radio Science (URSI), Istanbul

Wijnholds, S. J., & van Cappellen, W. A. 2011, IEEE Transactions on Antennas and Propagation, 59, 1981

Wilkinson, P. 2007, in "From Planets to Dark Energy: the Modern Radio Universe" PoS(MRU)144, Manchester, UK

Wilkinson, P. N. 1991, in Astronomical Society of the Pacific Conference Series, Vol. 19, IAU Colloq. 131: Radio Interferometry. Theory, Techniques, and Applications, ed. T. J. Cornwell & R. A. Perley, 428–432

Wilkinson, P. N., Kellermann, K. I., Ekers, R. D., Cordes, J. M., & Lazio, T. J. W. 2004, New Astronomy Reviews, 48, 1551

Xia, T. J. 2011, Optical Fiber Technology, 17, 328

Zwicky, F. 1948, The Observatory, 68, 121

Zwicky, F. 1957, Morphological astronomy (Berlin: Springer)

Author acronyms:

IEEE Institute of Electrical and Electronics Engineers

IEWG International [SKA] Engineering Working Group

ISPO International SKA Project Office (now Office of the SKA Organisation)

SPDO SKA Program Development Office (now Office of the SKA Organisation)

SOWG [SKA] Siting Options Working Group

SSWG SKA Science Working Group.

The SKA Memo series is available from http://www.skatelescope.org/publications.