PROCEEDINGS OF SCIENCE

Long Baselines in the Square Kilometre Array

Roshene McCool[1]
Square Kilometre Array Programme Development Office
Jodrell Bank Centre for Astrophysics,
3rd Floor, Alan Turing Bldg
The University of Manchester,
UK
E-mail: mccool@skatelescope.org

This paper will look at the data networking requirements and the challenges for data transmission over long baselines for the SKA and describe possible implementation strategies.

Science and Technology of Long Baseline Real-Time Interferometry:
The 8th International e-VLBI Workshop - EXPReS09
Madrid , Spain
June 22-26, 2009

---

[1] Speaker

1. Introduction

The Square Kilometre Array (SKA) is a programme dedicated to the construction of a radio interferometer with unprecedented sensitivity. It will be an extremely powerful survey telescope with the capability to follow up individual objects with high angular and time resolution. The telescope will be made up of thousands of elements and will be distributed across an area extending to thousands of kilometres. The array configuration will be made up of a number of dense cores, with dishes spread out to baselines of around 180 km. At long baselines (> 180 km) it is likely that dishes will be arranged in stations. These stations will contain approximately 18, wide-band single pixel feed dishes. The stations will be beamformed locally and data sent to a signal processing facility for correlation. Whilst long baseline receptors represent approximately only 25% of the collecting area, they will represent a disproportionate fraction of the cost of the data transmission by collecting area. This paper will discuss in detail the data networking requirements and challenges for these long baselines, describe possible implementation strategies and their associated cost drivers.

A number of implementation scenarios have been examined for data transport networks designed to take beamformed data, from dishes arranged in stations, back to a correlator at a signal processing facility, these are:

- A station solution. –The fibre connections in this scenario are built specifically for the SKA.
- A dark fibre solution. –The fibre connections in this scenario are implemented using existing leased dark fibre.
- A commercial bandwidth solution. –In this scenario the bandwidth from each station is provided in a turn-key contract with a bandwidth provider with an existing network.

2. Bit Rates for the SKA

The bit rate, R, of receptors in the SKA scales proportionally according to the relationship shown in Eqn. 1 [1]

$$R \propto \frac{BA_{tot}\Omega}{\lambda^2} \qquad \text{Eqn. 1}$$

Where B is the bandwidth of the receptors, Atot is the total collecting area of the array, Ω is the Field of View of the receptors and λ is the mean of the wavelengths of interest. This relationship holds, irrespective of receptor technology, whilst an additional, technology dependent, scaling factor determines the actual bit rate. The estimated bit rate for long baseline dishes is 160 Gbps per dish for an 8 GHz bandwidth and a 4 bit precision sampling. It is anticipated that the output of the beamformed station will also be 160 Gbps for an 8 GHz bandwidth.

It is clear from Eqn.1 that a reduction in bit rates, transported over the data transmission system, will have an impact on the science performance of the SKA as an instrument. This means that if the cost of the data transmission system imposes a limit on the bit rates, then it will, in turn, restrict the scientific capabilities of the instrument. Careful consideration is required in order to identify compromises between science goals and implementation costs. The goal will be to maximise cost reductions for the lowest impact on instrument performance.

3. Cost scaling for a station solution.

Dishes on long baselines in the SKA may be arranged into stations. These stations will contain a number of dishes, with an associated beamformer per station. Beamformed data will be transmitted back to a correlator. Beamforming stations on long baselines has the advantage, when compared to a dish only solution, that the number of transmission systems required is reduced by the number of dishes per station (in this case 18), with the cost savings that follow. However this cost saving comes with a compromise in the field of view and a restriction in the flexibility of the instrument. Figure 1. shows the 10 year costs of ownership for stations on long baselines, on one spiral arm, of a 5 arm configuration. Clearly trenching costs dominate.
The station solution leads to a much reduced data transmission overhead and offers opportunities to deliver network solutions using dark fibre or self-built networks. The additional cost overhead of transmitting more bandwidth is not a driving factor, when compared to the cost of cabling and transmission for a minimum bandwidth over long distances. For this reason it may be possible to offer multiple beams from the output of a beamformed station at long baselines.

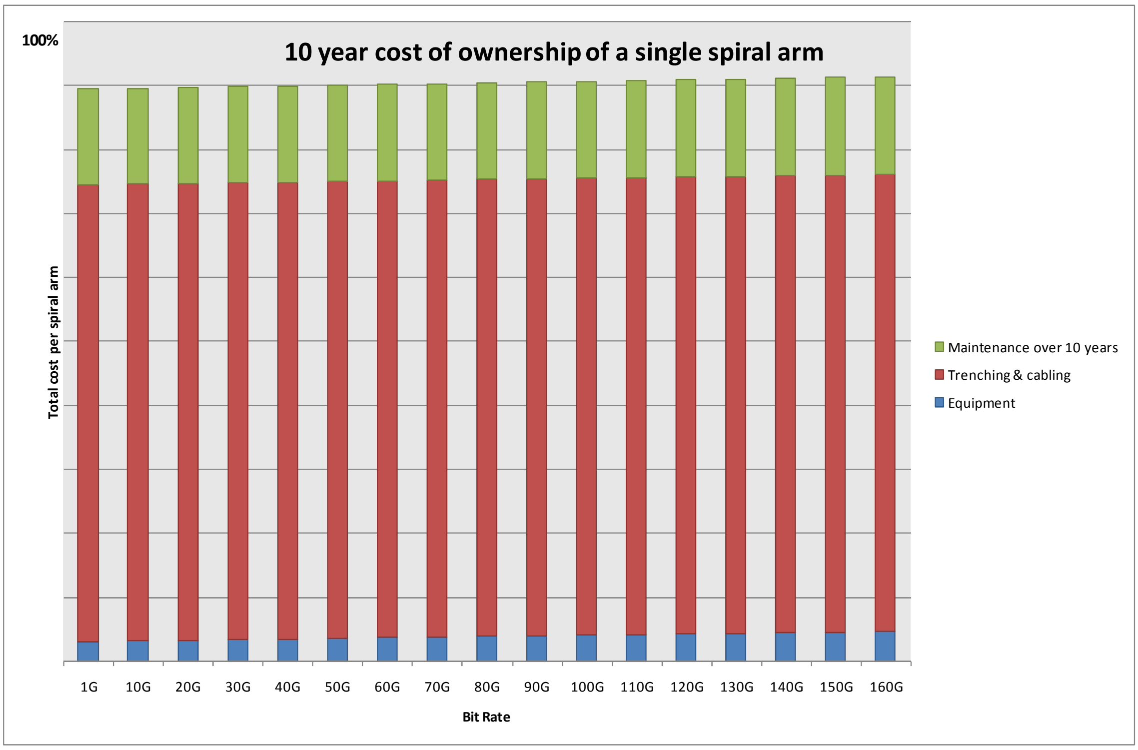

Figure 1. Relative share of the total 10 year cost of ownership of trenching, cabling and transmission equipment. Scale shows total costs in arbitrary units.

4. Cost scaling for a dark fibre solution.

The costs described in the previous sections are dominated by trenching costs. Evidently it will be desirable to reduce the cost of trenching. This can be done in one of three ways:

1. Reduce the cost/km of trenching
2. Reduce the distance of the longest baseline

3. Reduce the amount of trenching required

The practise of using existing fibre and 'lighting' it using a proprietary transmission system is known as leased dark fibre. The use of dark fibre can significantly reduce the trenching overhead for a new project. The ability to lease dark fibre is dependent on the availability of fibres in the site country and the location of the long baseline sites. The analysis assumes that some aggregation will take place along the dark fibre network and that data from 4 stations may be transmitted over a single dark fibre pair. This will be depend, heavily, on the network architecture on the final system. Figure 2 shows the 10 year cost of ownership of a dark fibre network. The dominant cost in this scenario is the length and the annual cost of the dark fibre lease.

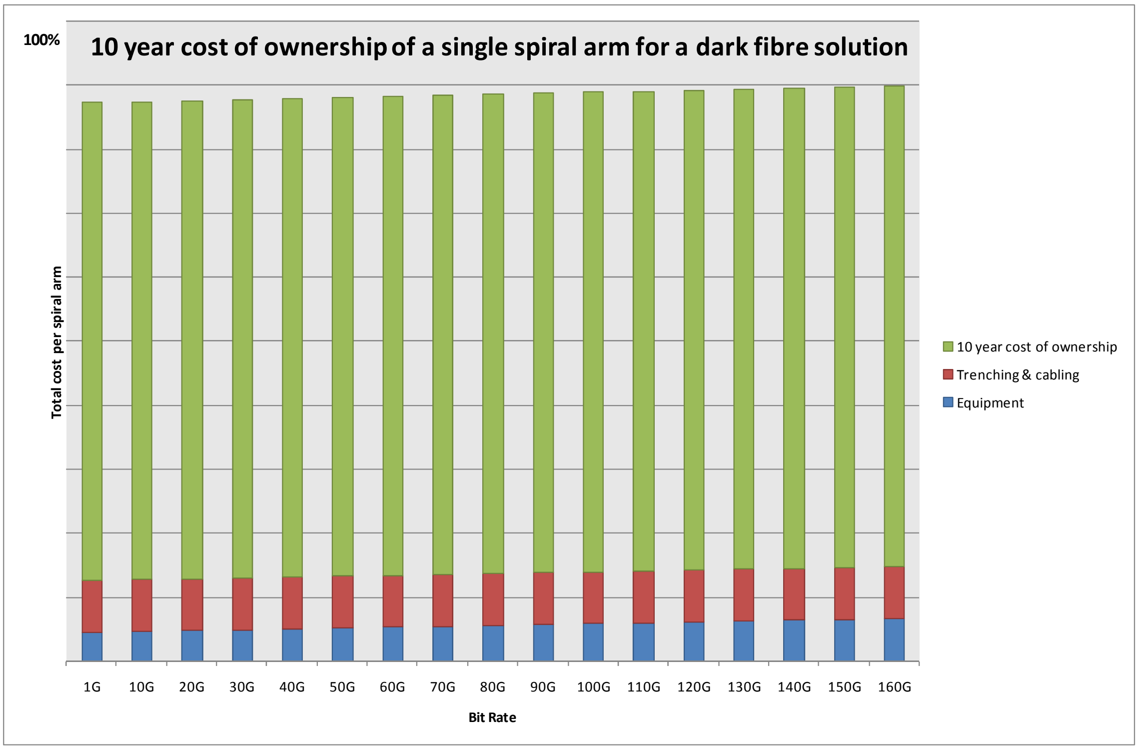

Figure 2. Relative share of the total 10 year cost of ownership of dark fibre lease and transmission equipment. Scale shows total costs in arbitrary units.

5. Cost scaling for a bandwidth solution.

In order to place the scale of the SKA data requirements in some perspective, one might compare SKA data rates to the total IP traffic generated globally in all areas of business and consumer use. Cisco [2] predicted that in 2008 the total, global IP traffic would be 10,803 Petabytes per month, or the equivalent of 35,725 Gbps. Taking the Cisco figures as a guide the SKA elements, operating at full bit rate, will generate the equivalent of the global IP traffic for 2008 with only 210 Wide Band Single Pixel Feed dishes.

The prospect of outsourcing the data communications aspects of the SKA is attractive. It transfers the risk and burden of implementation and operation to an external company.
The precedent for this type of solution for the SKA comes from the work done on e-VLBI over the past few years. In Europe the e-EVN operations have been very successful. Many

observatories now have local “last mile” fibre links to their telescope(s) and can access national research and education network (NREN) facilities to transmit data to the VLBI correlator at JIVE in the Netherlands. The e-EVN programme operates on specific observing windows. Generally an observation will take place for 2-3 days every 6 weeks.
The SKA operating model is different in a number of significant ways to the current e-EVN, eVLBI model.

1. The SKA will generate production traffic in an operation that is anticipated to work 24/7.
2. The traffic will be in quantities, as previously discussed, beyond those generated by the populace of the world.

The SKA requirements will swamp any network in existence and require operators to effectively construct a new network to serve the SKA alone. At the candidate sites themselves the local broadband infrastructure operates at several orders of magnitude below the estimates for SKA bit rate requirements. In order to provide a data transport network using current, commercial packet switched networks the SKA bit rate requirements will have to be significantly reduced. As illustrated in Eqn 1 this will have an impact on bandwidth, field of view or frequency coverage, and will possibly have to affect all three. It is true to note however, that commercial traffic is increasing at a rapid rate over time, whilst the SKA bandwidth requirements remain static. This means that as the networks expand with time they may eventually be large enough to cope with the additional capacity the SKA bandwidth would place on them. This is currently an unknown and we can only monitor this situation as the project progresses

6. Cost Drivers.

The cost estimates used in this analysis are based, wherever possible on available data. However at this stage in the project it is clear that the reliability of the data available is not at the level we need to achieve a costed system design for the SKA. It is useful therefore to highlight, for each of the solutions presented here, where the dominant costs are and therefore which costs will have the most impact on the design if changed.

Design Solution	**Dominant Cost drivers**
Station solution	• Trenching cost/km • Length of trenching *(depends upon baseline length)*
Dark Fibre solution	• Cost of the dark fibre/annum • Length of dark fibre required *(depends upon baseline length and number of stations)*
Bandwidth solution	• Cost per lambda or Gbps /annum • Bandwidth Requirements *(depends upon number of stations and bandwidth per station)*

Figure 3. Dominant costs for the various design solutions presented.

7. Conclusions

The dominant cost drivers for data transport networks from long baseline stations in the SKA are baseline length, the number of stations or dishes at long baselines, and to a lesser degree bandwidth. Requirements of the dominant cost drivers on long baselines will all impact implementation designs and costs. Science cases should be examined to identify science goals that will be impacted if the costs associated with long baselines require compromises in baseline length, station (or dish) count or bandwidth on long baselines.

The station solution leads to a reduced data transmission overhead when compared to a solution carrying data from every dish back to a correlator. It offers opportunities to deliver network solutions using dark fibre or self-built networks. The additional cost overhead of transmitting more bandwidth in a proprietary transmission system is not a driving factor, when compared to the cost of cabling and transmission for a minimum bandwidth over long distances. For this reason it may be possible to offer multiple beams from the output of a beamformed station at long baselines, depending on the bandwidths required per beam.

A dark fibre solution provides not only potentially the least cost, but a reduced operational overhead solution for providing data transport for the long baseline stations. For a practical implementation for the SKA, existing dark fibre must be available in the quantities and locations required. This has an impact on configuration studies, where stations in a carefully crafted configuration design may have to be moved in order to place them near a dark fibre network.

A bandwidth solution offers the least operational overhead of all the solutions described here, but also the solution with the most unknowns. The size of the SKA bandwidth requirements, far exceed the capacity of current commercial networks. A reduction in SKA bandwidths would result in considerable cut backs in bit rates and the corresponding reductions in bandwidth, field of view or frequency coverage. It is true to note however, that commercial traffic is increasing at a rapid rate over time, whilst the SKA bandwidth requirements remain static. This means that as the networks expand with time they may, given a rapid and sustained growth rate, eventually be large enough to cope with the additional capacity the SKA bandwidth would place on them. This is currently an unknown and we can only monitor this situation as the project progresses.

References

[1] P.E. Dewdney et al.,*The Square Kilometre Array*, Proceedings of the IEEE, Advances in Radio Telescopes, Vol. 97, Issue 8, August 2009.
[2] Cisco White Paper, *Visual Networking Index – Forecast and Methodology*, 2007–2012