

Data Rate and Processing Load Considerations for the LNSD SKA design

Colin Lonsdale, April 29 2003

Symbols:

N_s = # of stations	f = areal filling factor of stations
N_b = # of station beams	D = diameter of array
n_a = # of antennas per station	d = diameter of station
n_t = Total # of antennas = $N_s n_a$	d_a = diameter of antenna
B_v = Bandwidth (Hz)	θ = station FoV diameter $\sim \frac{\lambda}{d}$
S = sample rate ($\sim 2B$ samples/sec)	n_{res} = resolution elements across θ $\left(\sim \frac{D}{d} \right)$
C_{out} = correlator output data rate (vis/sec)	τ_{int} = correlator accumulation period
	Δv = correlator spectral resolution
	v = observing frequency

Field of View

To cover the full FoV of a single antenna $\sim \frac{\lambda}{d_a}$ we require a number of beams N_b

$$N_b \sim \left(\frac{d}{d_a} \right)^2, \text{ and since } d \sim d_a \left(\frac{n_a}{f} \right)^{1/2}$$
$$N_b \sim \frac{n_a}{f} \quad - (1)$$

Note that there is a special case: $f=1.0$ for $n_a=1$

Correlator dump rate + output data rate considerations.

For full antenna FoV:

$$C_{out} = N_{baselines} \times N_{freq\ channels} \times N_{integrations/sec} \times N_b$$
$$= \frac{N_s^2}{2} \frac{B_v}{\Delta v} \frac{1}{\tau_{int}} N_b$$
$$= \left(\frac{n_t}{n_a} \right)^2 \frac{n_a}{f} \frac{B_v}{2\Delta v \tau_{int}} \quad - (2)$$

For adequate sampling of visibility structure to retain FoV with a small loss of sensitivity at the edge, the approximate requirement on time integration on the longest baselines is:

$$\tau_{\text{int}} \sim \frac{4000}{n_{\text{res}}} \text{sec} = 4000 \frac{d_a}{D} \left(\frac{n_a}{f} \right)^{\frac{1}{2}}$$

This follows because a source m resolution elements away from the phase center will produce a maximum fringe rate of one turn in $\sim 4/m$ hours ($\sim 24/2\pi m$ hours), and we require 3 to 4 points per turn of phase.

A corresponding requirement on spectral averaging is:

$$\Delta \nu \sim \frac{\nu}{4} \frac{d}{D} = \frac{\nu}{4} \frac{d_a}{D} \left(\frac{n_a}{f} \right)^{\frac{1}{2}}$$

Yielding:

$$\begin{aligned} C_{\text{out}} &= \frac{n_t^2}{n_a f} \frac{B_\nu}{8000 \left(\frac{d_a}{D} \frac{\nu}{4} \frac{d_a}{D} \frac{n_a}{f} \right)} \\ &= \frac{1}{2000} \frac{n_t^2}{n_a^2} \frac{B_\nu}{\nu} \left(\frac{D}{d_a} \right)^2 \end{aligned} \quad - (3)$$

for $n_t = 5000$, $n_a = 1$, $\frac{B_\nu}{\nu} = 0.3$, $D = 1200\text{km}$, $d_a = 12\text{m}$

$$C_{\text{out}} = 2.5 \times 10^7 \times 0.3 \times 10^{10} \times 5 \times 10^{-4} = 3.75 \times 10^{13} \text{ vis / sec}$$

1 visibility = 4 pols \times (real, imag, weight) \times bytes/quantity \sim 50 bytes

Total output data rate for this example **~ 2 petabytes/sec**

The corresponding number for the inner 35km with 2500 antennas is 8×10^9 vis/sec = 400 Gbytes/sec. Expanding the area of single-antenna correlation with full FoV to 100km and ~ 3000 antennas increases this to 4.7 Tbytes/sec.

Consider stations of 13 antennas each. For $n_a=13$ and $f=0.2$, $N_b \sim 65$. By imaging the full FoV of a 12m antenna at resolutions corresponding to 1200km, the data rate drops to 11 terabytes/sec from 2 petabytes/sec. If only a single beam is fully imaged, the rate drops further to 170 Gbytes/sec.

These dataflows cannot be stored, and must be processed in real time. For imaging, this would consist of application of calibration solutions, followed by gridding into a (u,v,w) volume. The larger data rates discussed above do not appear feasible in the foreseeable future.

Conclusion: Correlator output data rate is a real constraint for the LNSD concept. It requires one or more of:

- Beamforming at stations with $n_a \gg 1$
- Reduction of FoV at high resolution
- Reduction of bandwidth at high resolution

Correlation and beamforming processing load

Fundamental operation is complex multiply and accumulate: CMAC

Correlation:

$$\begin{aligned} CMAC / \text{sec} &= \frac{4N_s^2}{2} SN_b \quad (\text{factor of 4 for polarimetry}) \\ &= \frac{4n_t^2}{2n_a^2} \frac{Sn_a}{f} \\ &= \frac{2n_t^2 S}{n_a f} \end{aligned}$$

Station beamforming:

$$\begin{aligned} CMACs / \text{sec} &= 2N_s n_a SN_b \\ &= \frac{2n_t n_a S}{f} \end{aligned}$$

Total processing load L

$$L(CMACs / \text{sec}) = \frac{2S}{f} \left(\frac{n_t^2}{n_a} + n_t n_a \right) \quad - (4)$$

$$\text{minimized when } \frac{2S}{f} \left(-\frac{n_t^2}{n_a^2} + n_t \right) = 0$$

$$\text{minimum} \rightarrow n_a = n_t^{1/2}$$

for $S = 10^9$, $n_t = 5000$, $n_a = 1$, $f = 1.0$ (special case)

$$L = 2 \times 10^9 (2.5 \times 10^7 + 5000) = 5 \times 10^{16} \text{ CMAC} / \text{sec}$$

This corresponds to **hundreds of peta-ops** (likely problematic) for 500 MHz of sky bandwidth. To perform full antenna-antenna correlations just for the inner 50% of the collecting area, this number is reduced by a factor of 4.

If we set $n_a = 70$ to roughly minimize, and set $f=0.2$:

$$L = 10^{10} \left(\frac{2.5 \times 10^7}{70} + 5000 \times 70 \right) \sim 7 \times 10^{15} \text{ CMAC/sec}$$

corresponding to a **few tens of peta-ops**, which may be feasible in special-purpose hardware. These two estimates bracket the processing load generated by full FoV requirements. A mixture of single antennas and stations, with $n_a < 70$, will be somewhere in between.

Conclusion: On SKA timescales, beamforming + correlation processing loads for LNSD are of concern, but are not showstoppers. Situation can be eased by:

- Some degree of station beamforming (savings according to equation 4)
- Reduction of FoV for high bandwidth (= high frequency) observations in stationized architecture by reduction in N_b (this cannot be done for full cross-correlation architecture).

Data transport Bandwidth

Total bandwidth B_t (bits/sec), b = bits/sample

$$\begin{aligned} B_t &= S b N_s N_b \\ &= S b \frac{n_t n_a}{n_a f} = \frac{S b n_t}{f} \end{aligned} \quad - (5)$$

For $n_a = 1$, special case $f = 1.0$

For $S = 10^9$, $b = 8$ (4 bits/polarization), $n_t = 5000$, $f = 0.2$

$$\begin{aligned} B_t &= 2 \times 10^{14} \text{ bits/sec} \\ &= 200 \text{ Tbits/sec} \end{aligned}$$

This is the array aggregate for full FoV at 500 MHz sky bandwidth, and is independent of n_a , except for the special case of $n_a = 1$, when this drops to 40 Tbits/sec. Clearly, if one wishes to preserve the full FoV, the data from each antenna should be transported to the central facility, and any beamforming should be done there instead of at the station.

Conclusion: To avoid a requirement for tens of Tbit/sec over 100s to 1000s of km, we must restrict FoV at high resolution, or restrict bandwidth at high resolution.