EVLA memo 77

EVLA and SKA computing costs for wide field imaging (Revised)[1]

T.J. Cornwell, NRAO[2]

tcornwel@nrao.edu

Abstract: *I investigate the problem of high dynamic range continuum imaging in the presence of confusing sources, using scaling arguments and simulations. I derive a quantified cost equation for the computer hardware needed to support such observations for the EVLA and the SKA. This cost has two main components – from the data volume, scaling as* D^{-6}*, and from the non-coplanar baselines effect, scaling as* D^{-2}*, for a total scaling of* D^{-8}*. A factor of two in antenna diameter thus corresponds to 12 years of Moore's law cost reduction in computing hardware. For a SKA built with 12.5m antennas observing with 1 arcsecond at 1.4GHz, I find the 2015 computing cost to be about $5B. For 25m antennas, the cost is about 256 times lower: $20M.*

This new cost equation differs from that of Perley and Clark, which has scaling as D^{-6}*. This is because I find that the excellent Fourier plane coverage of the small antenna design does not significantly improve the convergence rate of the Clean algorithm, which is already excellent in this regime.*

1. Introduction

Perley and Clark (2003) have recently derived a cost equation for synthesis arrays that includes the computing costs to counteract the non-coplanar baselines aberration. One conclusion from their work is that the cost equation should include a cubic term in the number of antennas. Consequently the minimum cost antenna diameter for fixed collecting area is increased over that derived while ignoring the costs of non-coplanar baselines. To determine how much the diameter is increased, the actual scaling coefficient must be known. In this document, I estimate the scaling relationships using analysis of the processing algorithms and large simulations performed in AIPS++.

---

[1] *An earlier version was published as EVLA memo 76. This version has been revised extensively, mainly in section 2 to improve the scaling arguments. Cost numbers have changed, some up, some down.*

[2] *The National Radio Astronomy Observatory is operated by Associated Universities, Inc., under cooperative agreement with the National Science Foundation.*

2. Scaling behavior

Perley and Clark analyzed the time taken to clean an image afflicted by non-coplanar baselines smearing using the facet based algorithms (see Cornwell, Golap, and Bhatnagar, 2004 for more on the taxonomy of wide field imaging algorithms). The w projection algorithm in AIPS++ outperforms the facet-based algorithms in AIPS, and AIPS++ by about an order of magnitude (Cornwell, Golap, and Bhatnagar, 2003, 2004), and so I choose to use it for these simulations.

Calculation of the work required to make a dirty image using w projection is straightforward but for clarity I first consider the case where the non-coplanar baselines effect can be ignored. Let there be N antennas of diameter D on baselines up to B. The number of channels and the integration time both scale as B/D. The number of baselines goes as N^2 and so the data rate goes as N^2B^2/D^2. For a constant collecting area, this is B^2/D^6. Only one image is made. The working in gridding nearly always dominates over that for the Fast Fourier Transform. The data are gridded onto the *(u,v,w=0)* plane using a convolution function of fixed size, typically 7 by 7 or 9 by 9 pixels. Hence the number of operations required to grid the data goes as the data rate N^2B^2/D^2.

If the non-coplanar baselines effect is important, either w projection or facet-based imaging must be used. For the former, one image is made and the area of the gridding function in pixels goes as $\lambda B/D^2$, and for the latter, the gridding function is constant but the number of images goes as $\lambda B/D^2$. Hence the number of operations required to grid the data goes asymptotically as $\lambda N^2B^3/D^4$, which for a constant collecting area goes as $\lambda B^3/D^8$.

For the cleaning, there are two main operations – the minor cycle clean and the major cycle calculation of the residuals. The latter nearly always dominates, and so the total cost goes as the number of major cycle times the gridding cost. The number of cycles is driven by the maximum exterior sidelobe (exterior to the beam patch in the Clark minor cycle). Each major cycle lowers the noise floor by roughly the sidelobe level, and so the dynamic range achieved after N_c major cycles is roughly:

$$\Lambda \sim \left(\frac{1}{2p_b}\right)^{N_c}$$

The factor of 2 occurs because for stability the clean in a major cycle typically terminates somewhat above the maximum sidelobe. The number of major cycles is:

$$N_c \sim -\frac{log(\Lambda)}{log(2p_b)}$$

I show in Appendix A the familiar result that the rms sidelobe level goes as the inverse of the square root of the number of visibility samples. If the field of view is held fixed, then the number of visibilities goes as D^{-4} and the sidelobe level as D^2, which is the scaling assumed by Perley and Clark. However, the field of view must expand as the antenna diameter decreases. Taking this into account, the number of visibilities goes as D^{-6}, and the rms sidelobe as D^3.

If the peak sidelobe level is 5 times the rms, the number of major cycles is:

$$N_c \sim \frac{log(\Lambda)}{log(\sqrt{N_s}/50)}$$

This excellent scaling behavior holds as long as the work in the minor cycle can be ignored. For imaging of complex sources, this may not be true.

Note also that the source spectral index enters only via the logarithm of the dynamic range. Hence that source of wavelength dependency can be ignored.

Taking all of these factors into account, I obtain scalings as shown in Table 1. This result differs from that of Perley and Clark principally because of the cleaning behavior. There is surprisingly little gain in speed for improvements in sidelobe level. I take special note that this is mostly due to the excellent characteristics of the Clark Clean algorithm.

I agree with Perley and Clark that there is an enormous penalty to using small antennas. This analysis is somewhat unfair to small antennas in that the benefits of excellent Fourier coverage for imaging complex sources are not incorporated but it does represent well the cost of removing confusing sources. This may change the numbers by 2 or 4 but will not compensate for the power of eight scaling.

Table 1 Expected scaling of imaging time with antenna diameter. This includes the costs of gridding, Fourier transform, and deconvolution. *C* is a constant.

	Number of antennas	**Time and frequency sampling**	**Non-coplanar baselines**	**Cleaning**	**Total**
General	N^2	$\frac{B^2}{D^2}$	$\frac{\lambda B}{D^2}$	$\frac{log(\Lambda)}{log(2\sqrt{N_s})}$	$\frac{N^2 B^3 \lambda}{D^4} \frac{log(\Lambda)}{log(2\sqrt{N_s})}$
Fixed collecting area	D^{-4}	$\frac{B^2}{D^2}$	$\frac{\lambda B}{D^2}$	$\frac{log(\Lambda)}{C - 3log(2D)}$	$D^{-8} \frac{log(\Lambda)}{C - 3log(2D)}$

2. Simulations

There are many subtle points to get right in the above analysis. Hence simulation is essential to check the scaling, and to determine the scaling coefficient.

Simulation of SKA observing on current computers is barely possible. I therefore choose to simulate only a short period of observing: 50s of time spread over 3000s of hour angle. The maximum baseline length was chosen to be only 10km. The antenna locations were chosen using a random process designed to give approximately Gaussian Fourier plane coverage. I performed three sets of simulations, with the same baselines and antennas but with wavelengths separated by factors of ten to separate out the influence of the non-coplanar baselines effect (this does not mean that I expect that a 12.5m antenna would be used to observe at 2.1m – just that it's convenient to scale the simulations in such a way).

These simulations have been constructed to scale appropriately with antenna diameter – hence as the antenna gets smaller, the integration time and narrow the bandwidth decrease linearly. This means that the data volume does indeed scale as D^{-6}.

The simulations were performed using the AIPS++ (version 1.9, build 549) `simulator` and `qimager` tools, running on a Dell 650 Workstation (dual processor Xeon 3.06Ghz processors, 3GB memory, Redhat Linux 7.2, special large memory kernel). The SPEC2000 floating point benchmark (CFP2000) is 13.8. Only one processor was used.

Table 2 Details of simulation

Total collecting area	Equivalent to 1600 12m antennas within 10km
Antenna diameter	12.5, 15, 17.5, 20, 22.5, 25, 27.5, 30, 32.5, 35, 37.5, 40m
Number of antennas	Set by antenna diameter to achieve fixed collecting area
Array configuration	Random antenna locations
Frequency	14GHz, 500MHz bandwidth; 1.4GHz, 50MHz bandwidth; 140MHz, 5MHz bandwidth.
Observing pattern	50s at transit, integration time 10s, scaling as antenna diameter, with gaps of 600s
Number of spectral channels	8 channels maximum, scaling inversely with antenna diameter
Array latitude	34 deg N
Source declination	45 deg
Source details	250 point sources per primary beam with source count index –0.7. Peak strength = 1Jy (but two sources may be in same pixel).
Antenna illumination pattern	Unblocked, uniformly illuminated
Synthesis imaging details	0.15, 1.5, 15 arcsec pixels, uniform weighting, with 0.6, 6, 60 arcsec taper. The image size scales inversely as antenna diameter. For the AIPS++ FFT, the image size must chosen to be the next largest composite of 2, 3, and 5.
Number of w planes in w projection algorithm	128

Clean details	Cotton-Schwab algorithm, loop gain 0.1, maximum 100,000 iterations, stopping threshold 0.1mJy for 40m scaling as $1/\sqrt{N_s}$
Resolution	0.6, 6, and 60 arcsec

The quantitative simulation results are given in Table 3. Some notes:

- The Fresnel number is $\lambda B / D^2$.
- The width of the w projection gridding function is determined from the numerically calculated form, and is closely related to the Fresnel number.
- Times shown are wall clock.
- The image properties shown are the minimum (affected by cleaning errors around bright sources), and the median absolute deviation from the median (a robust statistic showing the off source error level.
- I was not able to complete the most time-consuming case – 12.5m antennas observing at 2.1m. My estimate is that it would have taken about 4 days.

Table 3 Simulation results

Wavelength	**2.1**	**m**																
Antenna	*Fresnel*	*GCF*				*Sizes...*				*Times to ...*			*Clean*			*PSF*		*Image properties*	
diameter	*number*	*width*	*Ant*	*Int*	*Chan*	*Sources*	*Image*	*Vis*	*MS*				*Threshold*	*Comps*	*Cycles*	*min*	*Outer*	*minimum*	*robust*
m		*pixels*					*pixels*	*records*	*GB*	*construct*	*predict*	*clean*	*mJy*						
15.0	93.3	129	1111	8	13	694	4320	64126920	3.917	750.3	8865.5	73767.2	1.9	16516	4	-0.0015	0.0015	-2.97E-05	2.99E-06
17.5	68.6	128	816	7	11	510	3600	25604040	1.612	224.0	1848.4	15858.8	1.2	14517	4	-0.0011	0.0011	-7.66E-05	2.61E-06
20.0	52.5	109	625	6	10	390	3200	11700000	0.691	85.1	574.5	4924.7	0.8	11791	4	-0.0013	0.0015	-7.87E-05	2.29E-06
22.5	41.5	90	493	5	8	308	2880	4851120	0.297	43.2	183.4	1582.2	0.5	10951	4	-0.0017	0.0023	-4.19E-04	2.97E-06
25.0	33.6	71	400	5	8	250	2500	3192000	0.196	28.8	85.1	738.5	0.4	10457	4	-0.0023	0.0030	-7.74E-05	2.75E-06
27.5	27.8	66	330	4	7	206	2304	1519980	0.096	14.5	41.4	358.4	0.3	9357	4	-0.0033	0.0059	-3.39E-05	2.34E-06
30.0	23.3	54	277	4	6	173	2160	917424	0.06	10.1	22.3	202.7	0.2	7533	4	-0.0052	0.0077	-2.65E-05	1.79E-06
32.5	19.9	48	236	3	6	147	1944	499140	0.035	5.8	12.2	134.6	0.2	7187	5	-0.0097	0.0157	-5.80E-05	2.74E-06
35.0	17.1	39	204	3	5	127	1800	310590	0.022	4.2	5.8	71.2	0.1	9524	5	-0.0150	0.0149	-6.47E-05	4.79E-06
37.5	14.9	39	177	3	5	111	1728	233640	0.017	3.5	4.7	69.0	0.1	10556	6	-0.0193	0.0180	-8.40E-05	5.08E-06
40.0	13.1	32	156	3	5	97	1600	181350	0.014	2.8	5.4	73.6	0.1	8218	6	-0.0238	0.0241	-9.03E-05	3.63E-06
Wavelength	**0.21**	**m**																	
Antenna	*Fresnel*	*GCF*				*Sizes...*				*Times to ...*			*Clean*			*PSF*		*Image properties*	
diameter	*number*	*width*	*Ant*	*Int*	*Chan*	*Sources*	*Image*	*Vis*	*MS*				*Threshold*	*Comps*	*Cycles*	*min*	*Outer*	*minimum*	*robust*
m		*pixels*					*pixels*	*records*	*GB*	*construct*	*predict*	*clean*	*mJy*						
12.5	13.4	32	1600	10	16	1000	5000	2.05E+08	11.46	2288.3	2547.9	21671.3	3.4	20631	4	-0.0003	0.0003	-2.83E-05	3.43E-06
15.0	9.3	25	1111	8	13	694	4320	64126920	3.917	677.0	660.9	5678.9	1.9	16353	4	-0.0005	0.0007	-2.28E-05	2.72E-06
17.5	6.9	19	816	7	11	510	3600	25604040	1.612	203.7	217.6	2007.5	1.2	14270	4	-0.0007	0.0011	-4.01E-05	2.37E-06
20.0	5.3	16	625	6	10	390	3200	11700000	0.691	85.9	95.3	867.6	0.8	11619	4	-0.0011	0.0018	-8.83E-05	2.15E-06
22.5	4.1	14	493	5	8	308	2880	4851120	0.297	43.5	44.5	416.7	0.5	10468	4	-0.0019	0.0034	-4.09E-04	2.42E-06
25.0	3.4	11	400	5	8	250	2500	3192000	0.196	29.0	26.4	254.0	0.4	10196	4	-0.0021	0.0038	-1.99E-04	2.54E-06
27.5	2.8	11	330	4	7	206	2304	1519980	0.096	14.4	17.4	217.4	0.3	8099	5	-0.0048	0.0072	-2.60E-05	1.68E-06
30.0	2.3	10	277	4	6	173	2160	917424	0.06	10.1	12.4	123.7	0.2	7576	4	-0.0057	0.0078	-5.60E-05	1.76E-06
32.5	2.0	9	236	3	6	147	1944	499140	0.035	5.8	8.4	113.7	0.2	7376	5	-0.0097	0.0169	-5.51E-05	2.99E-06
35.0	1.7	8	204	3	5	127	1800	310590	0.022	4.2	3.6	59.7	0.1	10777	5	-0.0150	0.0152	-1.06E-04	5.40E-06
37.5	1.5	8	177	3	5	111	1728	233640	0.017	3.4	3.3	65.5	0.1	12111	6	-0.0193	0.0171	-8.50E-05	5.80E-06
40.0	1.3	7	156	3	5	97	1600	181350	0.014	2.9	4.6	67.7	0.1	7847	6	-0.0237	0.0243	-4.46E-05	3.31E-06
Wavelength	**0.021**	**m**																	
Antenna	*Fresnel*	*GCF*				*Sizes...*				*Times to ...*			*Clean*			*PSF*		*Image properties*	
diameter	*number*	*width*	*Ant*	*Int*	*Chan*	*Sources*	*Image*	*Vis*	*MS*				*Threshold*	*Comps*	*Cycles*	*min*	*Outer*	*minimum*	*robust*
m		*pixels*					*pixels*	*records*	*GB*	*construct*	*predict*	*clean*	*mJy*						
12.5	1.3	7	1600	10	16	1000	5000	2.05E+08	3.202	2372.8	1195.5	10339.1	3.4	20621	4	-0.0003	0.0003	-2.75E-05	3.38E-06
15.0	0.9	6	1111	8	13	694	4320	64126920	3.917	727.9	380.6	3354.4	1.9	16333	4	-0.0004	0.0006	-2.11E-05	2.70E-06
17.5	0.7	6	816	7	11	510	3600	25604040	1.612	201.2	141.3	1374.2	1.2	14209	4	-0.0007	0.0011	-6.48E-05	2.37E-06
20.0	0.5	5	625	6	10	390	3200	11700000	0.691	87.2	71.0	659.9	0.8	11632	4	-0.0012	0.0018	-4.33E-05	2.15E-06
22.5	0.4	5	493	5	8	308	2880	4851120	0.297	43.1	36.6	351.1	0.5	10697	4	-0.0021	0.0032	-1.06E-04	2.60E-06
25.0	0.3	5	400	5	8	250	2500	3192000	0.196	29.6	23.0	222.2	0.4	10169	4	-0.0021	0.0040	-6.26E-05	2.53E-06
27.5	0.3	5	330	4	7	206	2304	1519980	0.096	14.7	15.8	201.3	0.3	8051	5	-0.0044	0.0072	-2.58E-05	1.69E-06
30.0	0.2	5	277	4	6	173	2160	917424	0.06	10.1	12.8	118.7	0.2	7597	4	-0.0055	0.0078	-3.00E-05	1.75E-06
32.5	0.2	4	236	3	6	147	1944	499140	0.035	6.2	8.3	111.5	0.2	7744	5	-0.0097	0.0168	-3.33E-05	3.28E-06
35.0	0.2	4	204	3	5	127	1800	310590	0.022	4.2	3.5	58.4	0.1	10727	5	-0.0150	0.0151	-1.16E-04	5.36E-06
37.5	0.1	4	177	3	5	111	1728	233640	0.017	3.5	3.0	64.3	0.1	11896	6	-0.0193	0.0171	-1.12E-04	5.74E-06
40.0	0.1	4	156	3	5	97	1600	181350	0.014	5.6	4.2	68.3	0.1	7633	6	-0.0237	0.0235	-2.68E-05	3.27E-06

3. The scaling laws

The rms sidelobe scales with the cube of the antenna diameter as expected (Figure 1), for both natural and uniform weighting. These are very low by the usual standards in radio synthesis but there is no qualitative change in behavior for small antennas.

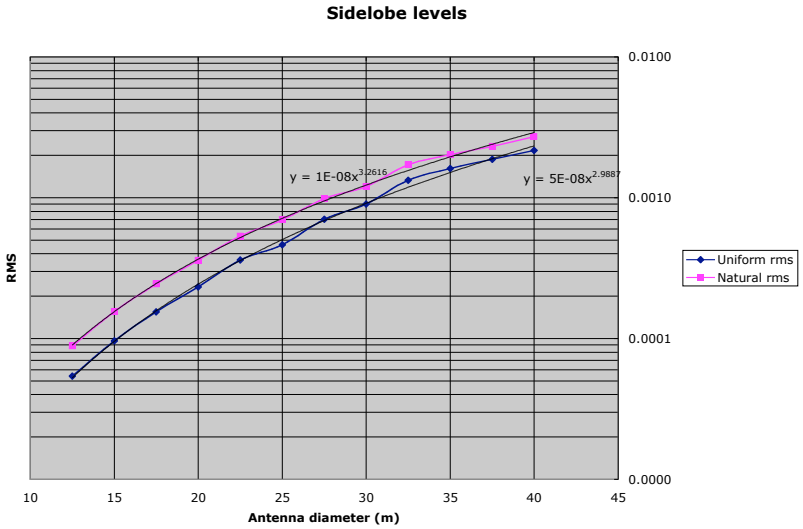

Figure 1 Sidelobe levels as a function of antenna diameter, showing scaling as the cube for both uniform and natural weighting.

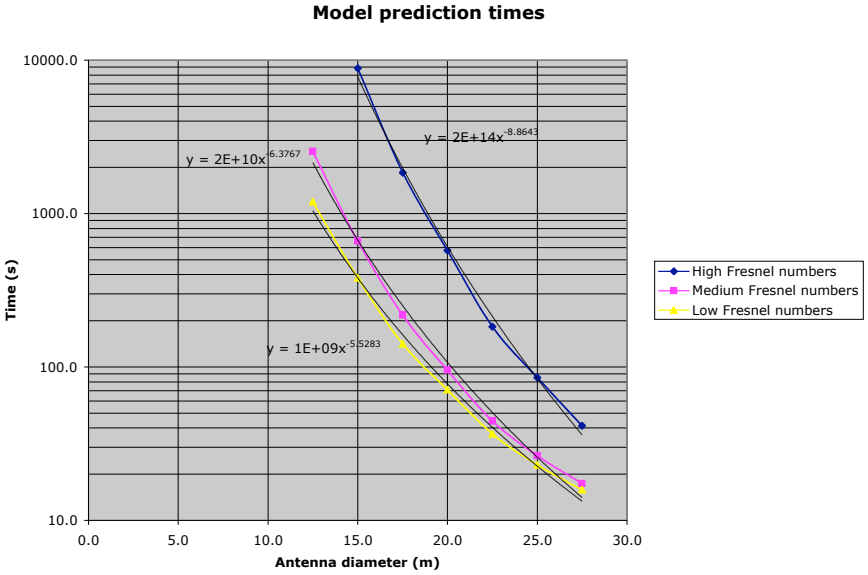

Figure 2 Model prediction times as a function of antenna diameter for low (<<1), medium (~1), and high (>>1) Fresnel numbers. Antennas > 27.5m have been excluded.

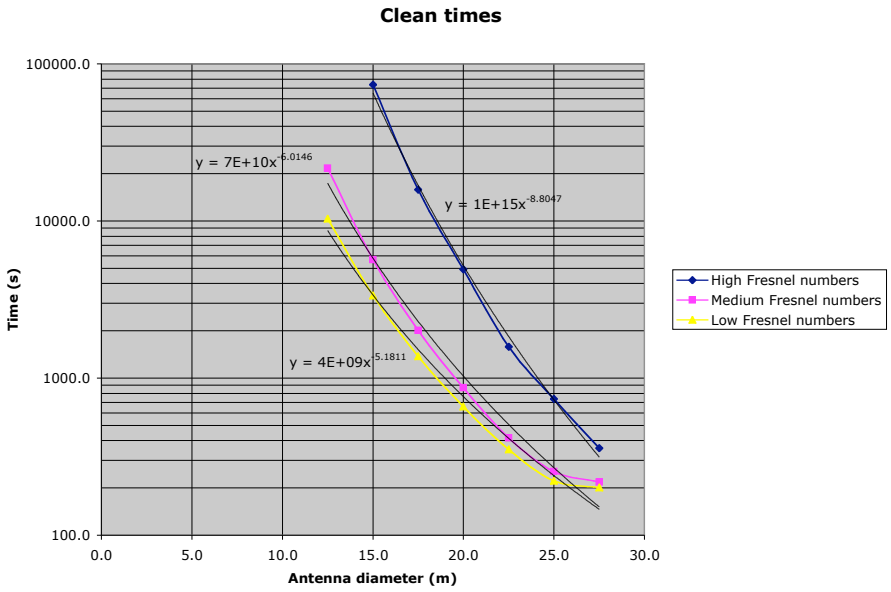

Figure 3 Clean times as a function of antenna diameter.

I find in the simulations that the scaling index for imaging time with antenna diameter varies with the Fresnel number as shown in figures 2 and 3 and as summarized in table 4.

Table 4 Observed scaling index for cleaning time as a function of antenna diameter.

Fresnel number	Model prediction	Clean
~ 10 - 1000	~ -8.9	~ -8.8
~ 1 - 10	~ -6.4	~ -6.0
~ 0.1 - 1	~ -5.5	~ -5.2

The scaling is steeper than –8 at the extreme ends but I believe this is most probably due to an onset of moderate paging. Hence for the high Fresnel number case, the scaling power can be taken to be –8.

Assuming Moore's Law for the cost of processing, the scaling law is:

$$C_{SKA} \sim C_{12.5m}\left(\frac{0.1}{\eta}\right)\left(\frac{f}{0.5}\right)^{2}\left(\frac{B}{5km}\right)^{3}\left(\frac{N}{1600}\right)^{2}\left(\frac{D}{12.5m}\right)^{-4}\left(\frac{\lambda}{0.2m}\right)\left(\frac{\Delta\nu}{500MHz}\right)2^{\frac{2(2010-t)}{3}}$$

Or, for a constant collecting area:

$$C_{SKA} \sim C_{12.5m}\left(\frac{0.1}{\eta}\right)\left(\frac{f}{0.5}\right)^2\left(\frac{B}{5km}\right)^3\left(\frac{D}{12.5m}\right)^{-8}\left(\frac{\lambda}{0.2m}\right)\left(\frac{\Delta\nu}{500MHz}\right)2^{\frac{2(2010-t)}{3}}$$

The filling factor f is the fraction of collecting area within the baseline B. For the SKA, the scientific specification on collecting area is 50% within 5km. The efficiency of processing, η, is both very important and as yet unknown. It includes, for example, the cost of correcting for source spectral effects, and antenna primary beams, and the efficiency of parallel processing. A reasonable value for this efficiency is about 10%.

For a 17.5m antenna design, the ratio between observing time and real time in our simulation is roughly 3000, so the efficiency is about 0.03% (for 25m, the ratio is ~ 100, efficiency is ~ 1%). The computer used in the simulations cost about $8000 in 2003. Solving, I find that the coefficient $C_{12.5m}$ is about $7M.

Since the antenna size for the EVLA has been chosen, I write the EVLA cost equation as:

$$C_{EVLA} \sim C_A\left(\frac{0.1}{\eta}\right)f^2\left(\frac{B}{35km}\right)^3\left(\frac{\lambda}{0.2m}\right)\left(\frac{\Delta\nu}{500MHz}\right)2^{\frac{2(2010-t)}{3}}$$

Scaling appropriately from C_{SKA}, I find that C_A is $170K.

The number of operations required per data point can be estimated by scaling by the CPU clock rate. The curves shown in figure 4 reach a minimum at about 20,000 floating pointing operations per data point. This should be taken as correct in order of magnitude only but it does reflect the scale of processing per data point.

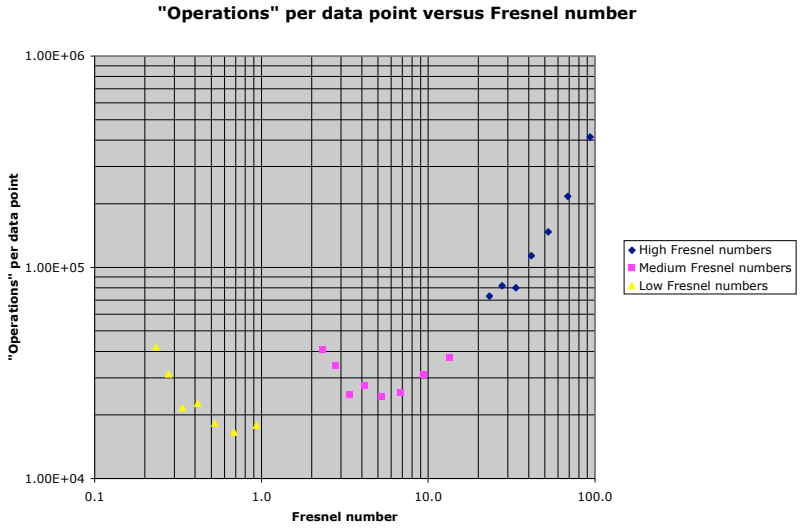

Figure 4 Operations per data point versus Fresnel number, calculated by scaling time by clock frequency, dividing by number of data points. The left end of each curve is biased upwards by constant cost terms.

4. Implications for the EVLA

For the EVLA A configuration (baselines up to 35km), the cost of computing hardware required for wide-field processing is $170K in 2010, $17K in 2020. This is quite modest and not dissimilar to previous estimates. In phase II, the EVLA will have baselines up to 350km baselines, and the costs would be $17M (2015), and $1.7M (2025). Such observations would be fairly rare and so the actual required duty cycle would be low.

Algorithm improvements help. The advent of w projection brings the cost down by about an order of magnitude, which is equivalent to a decade of Moore's law gains. Poor symmetry and stability of *e.g.* primary beams and pointing will hurt a lot by decreasing the efficiency (see *e.g.* Cornwell, 2003).

In addition, there remains a lot of software development to be done. It is clear that parallel processing using tens or hundreds of processors will be required to handle EVLA data. There has been relatively little work on parallelization of synthesis imaging algorithms.

Finally, operational models of the EVLA will affect the cost estimates. If the most demanding observations occur infrequently and turnaround can be a few days or weeks (as is now often the case) then the computing costs can be reduced proportionately.

5. Implications for the SKA

The canonical case of SKA imaging with 12.5m antennas on the 5km baselines at 20cm would require only $12M in 2010, and $1.2M in 2020. However, for the more interesting case of the 35km baselines, the costs rise to $5.4B and $0.5B. Increasing the antenna diameter to 25m brings the costs down to $21M and $2M. For 350km baselines, the cost increases to $21B and $2B, even with 25m antennas!

A key point is that the scaling behavior is very dramatic, as the cube of the baseline and the inverse eighth power of the antenna diameter. In comparison, the effects of more bandwidth and longer wavelength are linear. Thus the SKA computing budget will be determined by the scientific emphasis placed on baselines in the range of 10km and longer.

The primary conclusion is that computing hardware is a major cost driver for the SKA, and much more attention is required before the concept cost estimates can be viewed as accurate. In addition, simulations should start to include the non-coplanar baselines effect, so as to raise awareness of the importance of the effect for SKA. In the specific case of the LNSD concept, the cost minimization with respect to antenna diameter should be repeated with these more accurate computing costs included.

6. Possible remedies

Are there ways to avoid this large cost penalty for small antennas?

- *Invent an algorithm for non-coplanar baselines with better scaling:* This is a good idea, of course. The newest algorithm, w projection, is much faster than the old algorithm (faceted transforms), but has fundamentally the same D^{-8} scaling behavior, arising from the data volume, D^{-6}, and the physics of Fresnel diffraction, D^{-2}. More algorithm development is probably needed, but we shouldn't cash in breakthroughs not yet made.
- *Reweight the data taking account of the superb Fourier plane coverage of LNSD:* This has been done in these simulations. Tapered uniform weighting brings the sidelobes down by some factor but the basic scaling with the inverse square root of the number of visibility sample still applies. In any event, I have shown that the number of major cycles is only weakly determined by the sidelobe level.
- *Average data per facet:* It has been suggested (Lonsdale, private communication) that in a faceted transform, the data can be averaged in time and frequency before gridding, thus removing a factor of D^{-2}. While this is correct, it must be done appropriately for each facet, and therefore the factor is immediately lost. There may be a slightly lower scaling coefficient, bringing faceted processing closer to w projection in speed.
- *Ignore the sources outside the minimum field of view needed for science*: Rely on averaging in time and frequency to suppress the sidelobes from these sources. Experience at the VLA is that this approach does not work well. However, it may be more effective given the superior Fourier plane coverage of the SKA.
- *Form stations from clusters of antennas:* This would help a lot and certainly seems necessary on the longest SKA baselines. Whether it is acceptable for the shorter, 35km, spacings needs more study. It would undermine the strength of LNSD – the superb Fourier plane coverage.
- *Only observe snapshots:* For snapshots, the effect of non-coplanar baselines is less. At the dynamic range required, the integration time would have to be very short (~minutes or less).
- *Only do hard cases infrequently:* The VLA followed this path in the early eighties when spectral line observing was almost impossible. As technology improves, the duty cycle can be changed. This requires continuing investment in computing, and deferred gratification.
- *Mandate an efficiency of 100%:* The costs scale inversely as the efficiency of processing, η. We could mandate a "one-shot" policy. This seems to be counter-productive – why build a $1B telescope and then not reduce the data correctly? Efficiency of 100% is unlikely, anyway, since self-calibration will almost always be needed.
- *Build special purpose hardware for imaging:* This is quite plausible and should be investigated. The best approach would probably be to build a special (digital) processor to do the w-projection part of the imaging, and keep the rest of the processing in general purpose computers. Most of the work in w projection arises

from convolving the measured visibilities onto a regular grid, using a convolution function that varies in width as $\sim\sqrt{w}$ pixels to reflect the Fresnel diffraction effect (Cornwell, Golap, and Bhatnagar, 2003, 2004). The half-width of the convolution kernel is given in Table 3. For the most difficult cases, this can be in the range 100-300. This type of processing is also well suited to Graphical Processing Units (GPUs), which are now a commodity item, and some investigation of their use would be worthwhile.

Moving to very large antennas might seem the best way to solve this problem. However, the simulations with 40m antennas were marginally stable, tending to diverge for fewer hour angles. This should be understood in more detail before concluding that very large antennas are acceptable.

7. Summary

I find that for the specific problem of imaging in the presence of confusing sources, the use of small antennas comes at huge computing cost, as the inverse eighth power of the antenna diameter or the fourth power of the number of antennas. The cost has two main components – from the data volume, scaling as D^{-6}, and from the non-coplanar baselines effect, scaling as D^{-2}.

Continued algorithm research in this area is vital. We should investigate deploying existing algorithms on parallel machines, and possibly GPUs and special purpose hardware.

Finding a way to avoid this cost should be a high priority for SKA concepts that use relatively small antennas, such as the LNSD and the Aperture Array, as should be justifying the use of such small antennas.

Acknowledgements

I thank Sanjay Bhatnagar for help thinking about sidelobe levels, and Rick Perley for comments on an early version of this memo.

References

Briggs, D.S., (1995), "High fidelity deconvolution of moderately resolved sources", Ph.D. thesis, New Mexico Tech, http://www.nrao.edu/~tcornwel/danthesis.pdf.

Cornwell, T.J., 2003, EVLA memo 62.

Cornwell, T.J., Golap, K., and Bhatnagar, S., 2003, EVLA memo 67.

Cornwell, T.J., Golap, K., and Bhatnagar, S., 2004, *submitted*.

Perley, R.A., and Clark, B.G., 2003, EVLA memo 63. Appendix A: Scaling relations for sidelobe levels

Appendix A

For the scaling relations, we need approximate relations for the typical sidelobe level outside the center region of the PSF. This number determines how deeply any one major cycle of the clean algorithm can go.

The PSF is the Fourier transform of the weights attached to the visibility samples. For a gridded transform, the weight per grid cell is a product of the sum of the weights in that cell, optionally divided by some uniform weighting correction, and optionally multiplied by a taper function. The weighting correction is chosen to minimize the noise level (natural weighting), the sidelobe level (uniform weighting), or some compromise (robust weighting). See Briggs (1995) for more details.

We can use a random model for the distribution of Fourier plane samples. The PSF is simply a linear combination of the N_g grid weights. Thus the variance per grid cell adds:

$$\sigma_{psf}^2 = N_g \left(\sigma_w^2 - \langle w \rangle^2 \right)$$

We will limit our considerations to the case where all samples have the same intrinsic weight (before gridding). Normalizing the PSF then amounts to dividing by the total number of samples N_s.

For **natural weighting**, we can use a Poisson model of mean N_s / N_g. The variance about the mean is N_s / N_g and so the rms sidelobe level is:

$$\sigma_{psf,nat} = \sqrt{\frac{1}{N_s}}$$

For **uniform weighting**, the weight per grid cell is either 0 or 1. If the number of empty cells is N_e, then the rms sidelobe level is:

$$\sigma_{psf,uni} = \frac{\sqrt{N_g - N_e}}{N_g - N_e}$$

We can apply the Poisson model again, but this time we are interested in the number of empty cells given that I distributed N_s samples at random. The probability that a given cell is empty is:

$$e^{-\frac{N_s}{N_g}}$$

Hence the rms sidelobe goes as:

$$\sigma_{psf,uni} = \frac{1}{\sqrt{N_g\left(1 - e^{-\frac{N_s}{N_g}}\right)}}$$

For small average sample density much less than 1, this can be approximated by the natural weighting result. For high density, the number of grid points plays the role of the number of samples.