THE UNIVERSITY OF WESTERN AUSTRALIA
Achieving International Excellence

Resolution and confusion in ASKAP HI Surveys

Lister Staveley-Smith (UWA)

Date	Version	Revision
13 September 2008	0.1	Initial draft

Abstract

In order to resolve a galaxy detected in an HI survey from its neighbours, and in order to allow accurate optical/IR identification, good resolution is required. However, in order to detect a galaxy in the first place, there needs to be a reasonable match between the array resolution and the angular diameter of the galaxy. Too much resolution will make the galaxy undetectable. By considering simulated galaxy surveys, and neglecting any evolution of clustering strength $\xi(r)$ or HI mass function $\phi(M)$, I examine the range of suitable baseline lengths for shallow (all-sky) and deep (pencil-beam) ASKAP HI surveys. For the strawman ASKAP design and shallow survey (Johnston et al. 2007), the optimal range for the outer baselines is 500–1700 m, corresponding to a z=0 resolution of 0.5–1.7 arcmin ('natural' weighting). For the strawman deep survey, the optimal range for the outer baselines is 2000–8000 m, corresponding to a z=0 resolution of 8–30 arcsec ('natural' weighting). Such considerations confirm a 1.7 to 2.0 km array as a suitable compromise for the outer baseline lengths for a Gaussian ASKAP array.

1. Introduction

The Australian SKA Pathfinder (ASKAP), to be located at the Murchison Radio-astronomy Observatory (MRO) in Western Australia, will be one of a new generation of radio telescopes designed to pave the way for the SKA. ASKAP will trial new technology, a new world-class site, and will allow extensive new radio surveys. At its operational frequency of around 1 GHz, ASKAP will also be one of the world's best survey telescopes until the SKA is built (Johnston et al. 2007). One of its main scientific outputs will be extensive surveys of galaxies in the 21cm line of neutral hydrogen. In a previous memo (Staveley-Smith 2006), I used estimations of the HI sizes of distant galaxies and angular correlation functions from optical/IR surveys to suggest an optimum resolution of around 30 arcsec at z=0.2 for ASKAP HI surveys. In this memo, I further explore the question of optimum resolution by using HIPASS correlation functions (Meyer et al. 2007) to estimate angular and spectroscopic confusion levels as a function of redshift, and examine the effect of various array configurations on the detectability of galaxies in simulated deep and all-sky surveys.

2. Simulated Surveys

For both the all-sky (2π) and deep (30 deg^2) surveys, I assume integration times of 1 yr, and assume that the ASKAP system parameters (area, T_{sys}, aperture efficiency and field-of-view) are as given in the strawman specifications of Johnston et al. (2007). Galaxies with HI masses between 10^6 and 10^{11} $M_\odot$ are initially populated randomly in comoving space (Ω_Λ=0.7, Ω_m=0.3 H_o=75 km s^{-1} Mpc^{-1}) with a mass spectrum approximated by a Schechter function (ϕ=0.006 Mpc^{-3}, M^*=6.3×10^9 $M_\odot$, α=-1.37; Zwaan et al. 2005). The mass function is assumed not to evolve. Corresponding redshifts and luminosity distances are then computed for each galaxy, and a velocity width is assigned in such a way as to approximately reproduce the HIPASS mass-linewidth relation. No corrections for profile shape are applied. Galaxies are only 'detected' if their frequency-smoothed spectra are more than 5 times the rms noise.

3. Resolution

The effect of resolution on the visibility of galaxies at 21cm depends in detail on a number of factors including: the three-dimensional structure of the galaxy in the spatial and Doppler velocity domain; the spatial frequency coverage of the interferometer array; and the nature of the galaxy detection algorithm. For this study, I use the very simple approximation that, as the redshift-dependent synthesised beamwidth, θ decreases, the S/N ratio of a galaxy detection is correspondingly degraded by $\theta/\sqrt{(\theta^2+\mu^2)}$, where μ is the apparent galaxy diameter in similar units. This approximation is based on the fact that the average flux per beam area decreases in proportion to beam area once the galaxy is resolved, but that useful galaxy finders are able to regain S/N by *incoherent* addition of independent beam areas through techniques such as wavelet decomposition.

In order to compute galaxy angular sizes, I assume an approximately constant HI rest-frame surface density of 1 $M_\odot$ pc^{-2} (Meyer et al. 2004) with a modest dispersion. Staveley-Smith et al. (1998), Broeils & van Woerden (1994), and others, have also shown that constant surface density is a good approximation for HI-rich galaxies. Dispersion and cosmological dimming both cause apparent surface density to vary from the mean value. Variable Doppler width causes further dispersion in the ratio of flux density and beam area.

For a low-resolution 2π 'all-sky' survey, around 6×10^5 simulated galaxies are detected at low redshift. As the resolution is increased, the number of detections decreases as shown in Figure 1. For maximum baselines below 1000 m, the percentage of galaxies detected remains above 80% of the low-resolution value. Above 2000 m, the resolution is too great to detect large numbers of low-redshift galaxies.

For the strawman 'deep' survey, the typical redshifts are higher and the typical galaxy diameters are smaller. Therefore there is little change in the number of objects (around 7×10^4) detected as a function of baseline length (Figure 1).

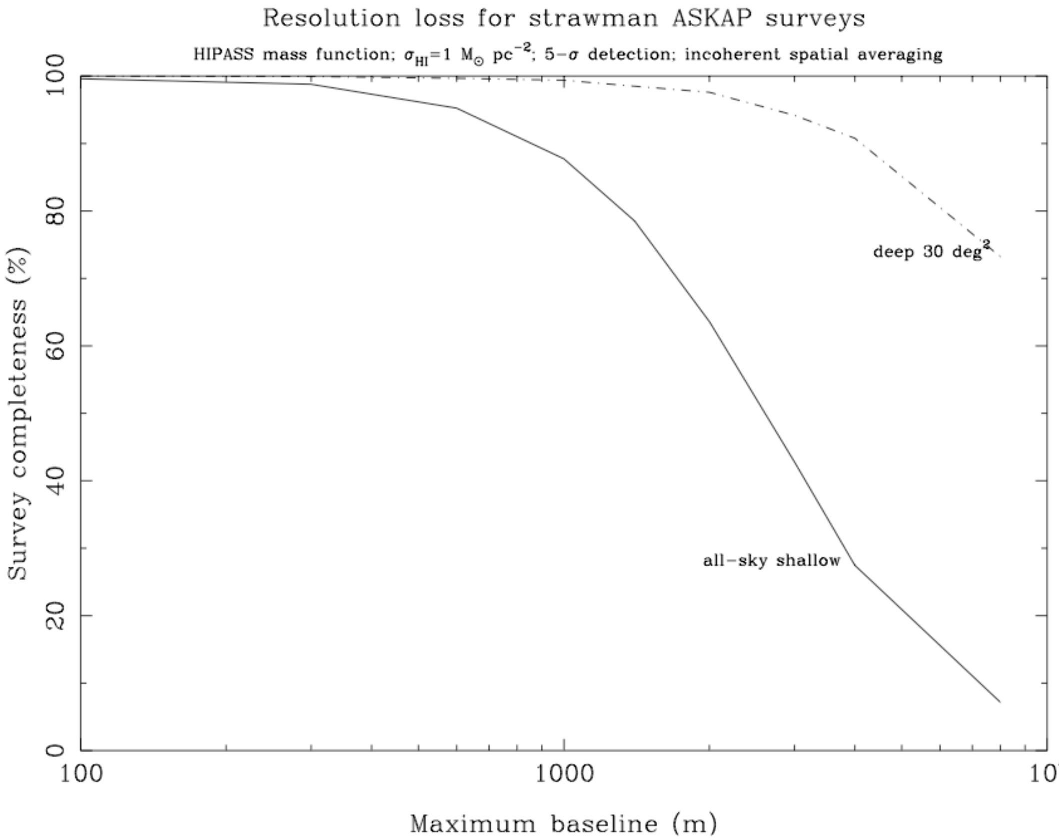

Figure 1: The simulated number of galaxies detected in (a) an all-sky shallow survey, and (b) a deep 30 deg^2 survey, as a fraction of the number detected at low angular resolution, as a function of maximum ASKAP baseline length. The survey parameters are described in the text. The strawman 30-dish ASKAP option is assumed (Johnston et al. 2007), but the size of the array is allowed to vary from a maximum baseline of 100 m to 10 km. For the natural Gaussian antenna distributions described in Staveley-Smith (2006) and modelled in Gupta, Johnston & Feain (2008), the relation between resolution θ and maximum baseline B is approximately $\theta = 1.4\lambda/B$.

4. Confusion

Around a given galaxy, the density of neighbours is not random, but is normally given by the product of the HI mass function $\phi(M)$ (or the optical luminosity function $\psi(L)$), and the galaxy-galaxy correlation function $\xi(r)$, (assuming no mass-dependence in the latter). The Schechter form of the mass function is given by:

$$\phi(M_{HI})dM_{HI} = \phi^* \left(\frac{M_{HI}}{M^*}\right)^{\alpha} \exp\left(-\frac{M_{HI}}{M^*}\right) d\left(\frac{M_{HI}}{M^*}\right).$$

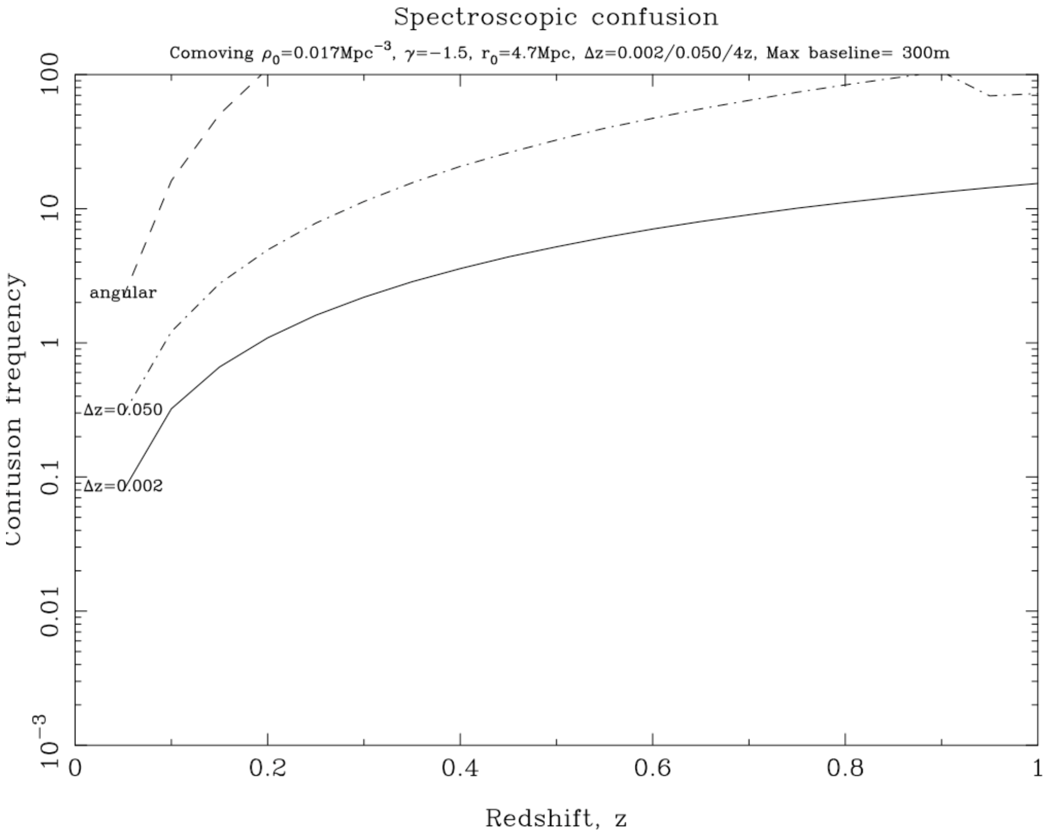

Figure 2: The average frequency of confusing galaxies, within the synthesised beam of a 300 m ASKAP configuration, as a function of redshift. The synthesised beam is assumed to be 1.4λ/B in diameter, where λ is the redshifted wavelength of the 21cm line and B is the maximum baseline. Spectroscopic confusion frequencies within a cylinder of total length Δz=0.002 (corresponding to twice the Doppler width of a typical M* galaxy; solid line) and Δz=0.05 (a typical accuracy for photometric redshifts, Hildebrandt et al. 2008; dot-dashed line) are plotted. The angular confusion frequency (where the correlation function is integrated in a cone out to redshift 4z) is also plotted (dashed line). The mean galaxy density is assumed to be 0.017 Mpc^{-3}, corresponding to 0.1M*, and above, galaxies. For M* galaxies, the confusion frequency is reduced by a factor of 15. Correlation function parameters from Meyer et al. (2007) are used.

The number of galaxies with HI masses larger than M_{HI} is given by:

$$n_0\left(M > M_{HI}\right) = \phi^* \, \Gamma\left(1+\alpha, \frac{M_{HI}}{M^*}\right),$$

and the total HI mass density above M_{HI} is given by:

$$\int_{M_{HI}}^{\infty} M_{HI}\phi(M_{HI})dM_{HI} = M^*\phi^* \, \Gamma\left(2+\alpha, \frac{M_{HI}}{M^*}\right).$$

For HIPASS mass function parameters already described (Zwaan et al 2005), the density of galaxies greater than M* is 0.0011 Mpc^{-3}, and the density of galaxies greater than 0.1M* is 0.017 Mpc^{-3}. The HI mass density in galaxies more massive than M*, as a fraction of the total HI mass density, is 21%; for galaxies more massive than 0.1M* the fraction is 75%. Since galaxies with mass greater than

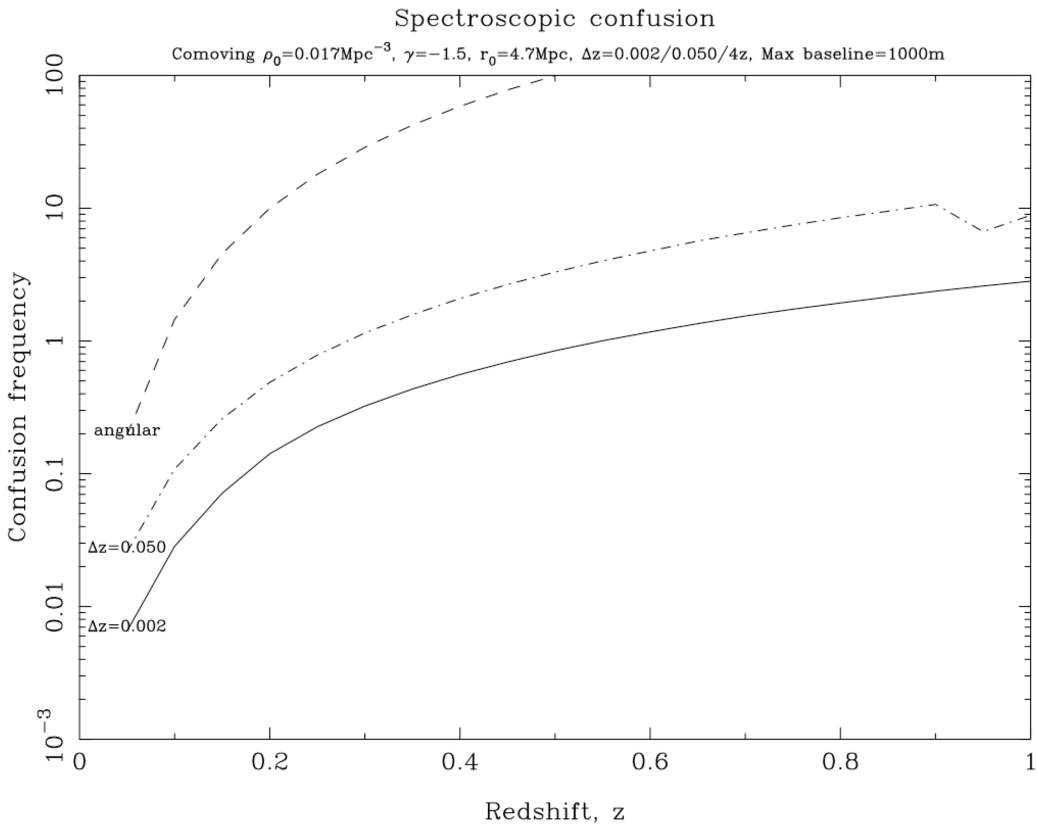

Figure 3: The average frequency of confusing galaxies, within the synthesised beam of a 1000 m ASKAP configuration, as a function of redshift. Other details as for Figure 2.

$0.1M^*$ appear to trace the bulk of the HI (and optical light) in the Universe, I will use the density of these objects as the measure of average confusion level.

The galaxy-galaxy correlation function, which gives the density of galaxies as a function of (comoving) distance, r from a given galaxy, is normally approximated in the non-linear regime by:

$$\rho(r) = n_0\left(1+\left(\frac{r}{r_0}\right)^{\gamma}\right),$$

where r_0 is the correlation length. The average number of 'confusing' galaxies in a cylinder of comoving line-of-sight depth, β (Mpc) and transverse comoving radius, κ (Mpc) is therefore given by:

$$n(\beta,\kappa) = \int\int 2\pi\kappa\rho(r)d\kappa d\beta.$$

For $r^2=\beta^2+\kappa^2$, the solution is:

$$n(\beta,\kappa) = \pi\beta\kappa^2\left(1-\frac{2}{\gamma-2}\left(\frac{r_0}{\kappa}\right)^{\gamma}{}_2F_1\left(\frac{1}{2},\frac{\gamma}{2}-1;\frac{3}{2};-\frac{\beta^2}{\kappa^2}\right)\right),$$

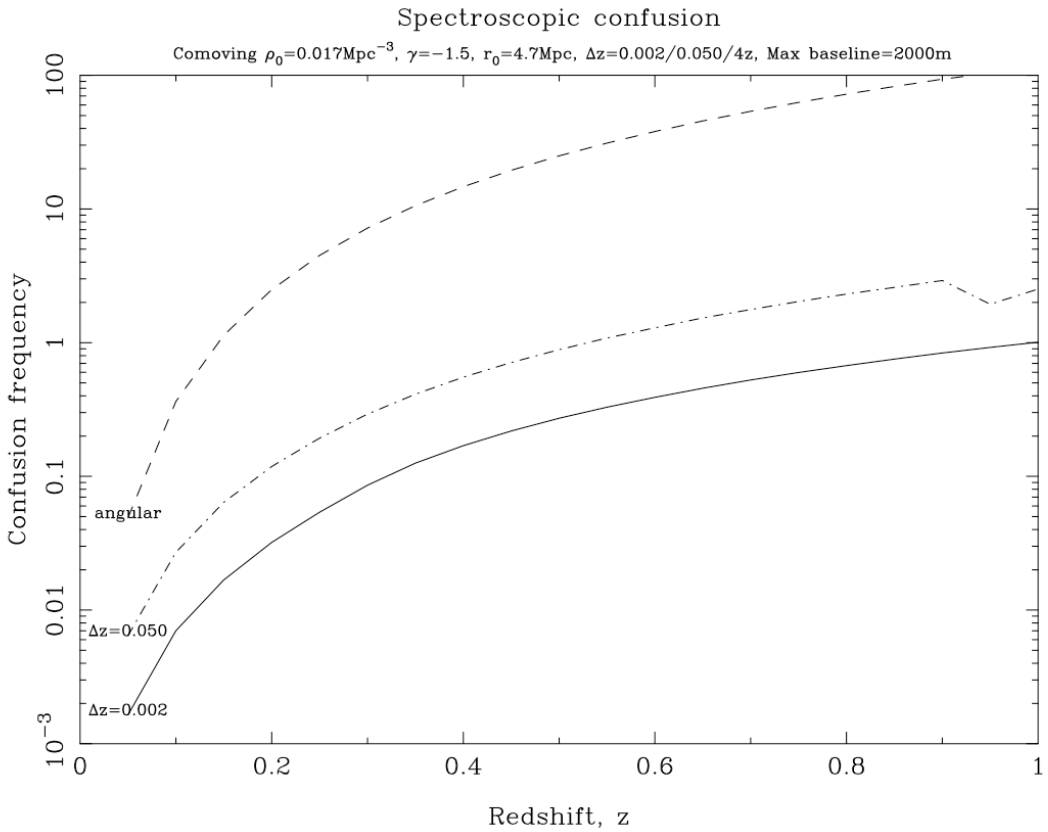

Figure 4: The average frequency of confusing galaxies, within the synthesised beam of a 2000 m ASKAP configuration, as a function of redshift. Other details as for Figure 2.

where $_2F_1$ is a hypergeometric function. I use the HIPASS correlation function parameters $\gamma = -1.5$, r_0=4.7 Mpc (Meyer et al. 2007, H_0=75 km s^{-1} Mpc^{-1}) and assume no evolution – i.e. no change of these parameters as a function of redshift when expressed in comoving space. The integral was evaluated numerically using a 2D trapezoidal Riemann sum down to minimum comoving radius of 0.1 Mpc, representing the limit of knowledge of the HI correlation function (Meyer et al. 2007).

The cylinder diameter 2κ is set by the synthesised beam of the array at redshift z. In principle, κ may be reduced if the positional accuracy for the centroid of the detected galaxy is substantially better than the beam size. For a typical survey detection threshold of 5-σ, the rms positional accuracy will be ~20% of the beam size, so an error circle of the order of around half the beam size may be slightly more appropriate. The relationship between the synthesised beam for Gaussian ASKAP configurations is approximated by 1.4λ/B, where λ=0.211(1+z) m and B is the outer baseline length (m).

The cylinder depth is set by the accuracy of the available redshifts for the confusing population. For example, a candidate optical counterpart at z=0.110±0.001 will be rejected as the counterpart of an HI detection at z=0.1000±0.0001. A lower limit to the cylinder depth of Δz≈0.002 is set by the intrinsic Doppler width of pairs of galaxies. This is probably an appropriate value to use for 'spectroscopic' confusion. In cases where dense optical spectroscopy is unavailable, it is possible that photometric redshifts, with Δz≈0.05 (Hildebrandt et al. 2008), can be relied on. For each of these cases, the confusion frequency (i.e. the average number of galaxies that could be potential counterparts to an HI-detected object) is shown in Figures 2-4 for putative 300-m, 1000-m and 2000-m ASKAP configurations. Also shown are the angular confusion rates.

At a redshift of z=0.05, which is close to the median redshift for a 1-yr shallow ASKAP HI survey, the angular confusion rate is below 30% for the 1000 and 2000 m configurations, and the spectroscopic confusion rate is below 3% for $\Delta z=0.05$ and below. In other words, even a 1000-m configuration results in acceptably small levels of confusion if photometric redshifts are available. At low redshifts, the principle benefit of photo-z's arises from the ability to discriminate against background objects. The 300-m configuration has an angular confusion rate in excess of unity at z=0.05, and is lower than 10% only when full spectroscopic redshifts are available for all potentially confusing galaxies (there are around 16 million galaxies per hemisphere with HI masses in excess of $0.1M^*$ at redshifts below 0.2, and therefore a similar number of potentially confusing optical counterparts).

At a redshift of 0.2, close to the median redshifts of the deepest ASKAP surveys, confusion is more serious. The 300-m configuration (Figure 2) is spectroscopically confused (the corollary being that it is also self-confused with at least one companion galaxy expected per beam area per 600 km s^{-1} rest-frame velocity interval). The 1000-m configuration (Figure 3) requires full photometric redshift coverage ($\Delta z=0.05$) to push the confusion probability below ~50%. Even the corresponding 2000-m (Figure 4) confusion frequency is still 10%, and the angular confusion rate is above unity.

5. Summary

The maximum acceptable baseline length is set by the desire for good surface brightness sensitivity. Once this sensitivity begins to approach the mean surface brightness of galaxies in HI, galaxy numbers begin to decrease markedly in simulated surveys. They appear to do so more at the higher redshifts due to the combined effects of cosmological dimming and higher intrinsic Doppler widths appropriate to higher-mass galaxies. For the 1-yr all-sky and deep surveys described, the upper limits are approximately 1.7 and 8 km, respectively. At these baseline lengths, around a quarter of the sample is resolved out compared with lower-resolution arrays. The only way to recover 'lost' galaxies is to increase integration time.

The minimum acceptable baseline length is set by the desire to: (a) resolve companion galaxies from each other; and (b) correctly identify counterparts at other wavelengths, particularly in the optical/IR regime. These considerations are more complex and depend on the nature of other available data. Dense photometric and, especially, spectroscopic data can be very helpful in aiding identification in surveys done with poor angular resolution. Without redshift data, baselines of around 1 km are required to reduce the confusion to barely acceptable levels (20%) at the low redshifts (z~0.05) typical for the all-sky survey. With photo-z data, maximum baselines below 1 km, but above 300 m, are required –

probably around 500 m. For deep surveys, photo-z's are not only more accurate as a fraction of observed redshift, they are also more likely to be available over the smaller fields required. In this case, the maximum baseline needs to be around 2 km, or greater.

The optimal range for the maximum ASKAP baseline is therefore 500 – 1700 m for a nominal 1-yr shallow survey, and 2000 – 8000 m for a nominal 1-yr 30 deg^2 deep survey. It is suggested that a compromise Gaussian configuration of maximum baseline 1700—2000 m is more suitable than a hybrid configuration.

References

Broeils A.H., van Woerden, H. 1994, A&AS, 107, 129

Gupta, N., Johnston, S., Feain, I. 2008, ATNF SKA memo series no. 16

Hildebrandt, H., Wolf, C., Benitez, N. 2008, arXiv:0801.297

Meyer, M.J. et al. 2004, MNRAS, 350, 1195

Meyer, M.J. et al. 2007, ApJ, 654, 702

Staveley-Smith, L. 2006, ATNF SKA Memo Series no. 6

Staveley-Smith, L. et al. 1998, AJ, 116, 2717