

Stacking for Cosmic Magnetism with SKA surveys

Jeroen M. Stil^{*†}

Department of Physics & Astronomy, The University of Calgary

E-mail: jstil@ucalgary.ca

Ben W. Keller

Department of Physics & Astronomy, MacMaster University

E-mail: kellerbw@mcmaster.ca

Stacking polarized radio emission in SKA surveys provides statistical data on large samples that is not accessible otherwise due to limitations set by a combination of sensitivity, source statistics, and frequency averaging. Polarization is a special case because one obvious source of stacking targets is the Stokes I source catalog, possibly in combination with external catalogs, for example an SKA HI survey or a non-radio survey. We point out the significance of stacking polarization of subsamples selected by additional observable parameters. Applications of stacking polarization include but are not limited to obtaining in a statistical sense polarization information to the detection limit in total intensity, investigating depolarization as a function of cosmic time at consistent source-frame wavelengths, magnetic field properties in objects with a low radio luminosity such as dwarf and low-surface-brightness galaxies, and investigating potential correlations of observable parameters with the average magnetic field direction in a sample. We also point out the potential use of stacking in validating the polarization calibration of a survey. While stacking is flexible in terms of survey definition, we discuss optimal survey parameters for the science experiments presented, as well as computing and archiving requirements.

Advancing Astrophysics with the Square Kilometre Array

June 8-13, 2014

Giardini Naxos, Italy

^{*}Speaker.

[†]On behalf of the SKA Cosmic Magnetism Working Group

1. Why Stack Polarization?

One of the areas of strength for SKA on the evolution of galaxies and their magnetic fields is in the analysis of very large samples over a wide range of redshift. Stacking is a statistical analysis of the emission of a sources that are too faint to be detected individually in the survey, whose position is known from another survey. Stacking provides a flux density representative for a carefully selected sample of sources. The astrophysical interpretation depends on the selection criteria of the input catalog. A particularly useful application is to compare subsamples distinguished by an observable parameter, e.g. inclination, flux at another wavelength, galaxy type, or spectral index.

Stacking radio polarization Stil et al. (2014) offers three significant benefits. The first is that it allows investigation of the polarization of radio sources to the detection limit in total intensity, even though the polarized signal is usually only a few percent of the total flux density. The second is that application of stacking polarization as a function of flux density allows a uniform investigation of the polarization of radio sources without applying a detection threshold in polarized intensity. The third benefit of stacking polarization is that it provides an opportunity to study polarization of sources as a function of frequency with high sensitivity without the need to average over all observed frequencies. Faraday rotation and depolarization are highly dependent on wavelength at the source.

Figure 1 illustrates some of these advantages from stacking the NVSS survey (Condon et al. (1998)). The left panel in Figure 1 shows fractional polarization versus flux density for all NVSS sources. Each data point in this figure represents a stacking experiment and a sequence of Monte-Carlo realizations of the stack to correct for polarization bias and determine the error bars. A direct investigation of sources with detectable polarized flux limits this analysis to sources with $S_{1.4} \gtrsim 80$ mJy (Mesa et al. (2002), Tucci et al. (2004)). While deep surveys can detect polarization in fainter sources, detecting the gradual change in fractional polarization in Figure 1 requires sample sizes for which current deep fields are too small. The noise in the median image of the faintest bin is $1 \mu\text{Jy}$, sufficient to detect polarization in the faintest sources in the NVSS catalog without concern for confusion.

Splitting up the sample also by spectral index reveals how the dependence of polarization on flux density is mainly due to sources with intermediate spectral index (Figure 1, right). A difference between bright flat and steep spectrum sources was reported by Mesa et al. (2002); Tucci et al. (2004), but the trends in Figure 1 can only be studied through stacking at this time. Does the trend for sources with intermediate spectral index continue at lower flux density, and what does it mean? This example illustrates how subdivision of the input catalog - in this case by spectral index - provided additional information by revealing a correlation with an observable parameter. The flux density limit in the right panel of Figure 1 is set by uncertainty in the spectral index used to define the subsamples. One should also be careful when constructing the subsamples. For example, we subdivided the sample at the modal spectral index -0.75 so that noise broadening of the spectral index distribution of faint sources does not affect the number of steep-spectrum sources in the intermediate spectral index range.

This example illustrates an important point about polarization stacking: significant gains can be made using only the Stokes I source catalog as the input catalog. Such experiments would not suffer from confusion or source blending, and the interpretation of the stacked polarization is

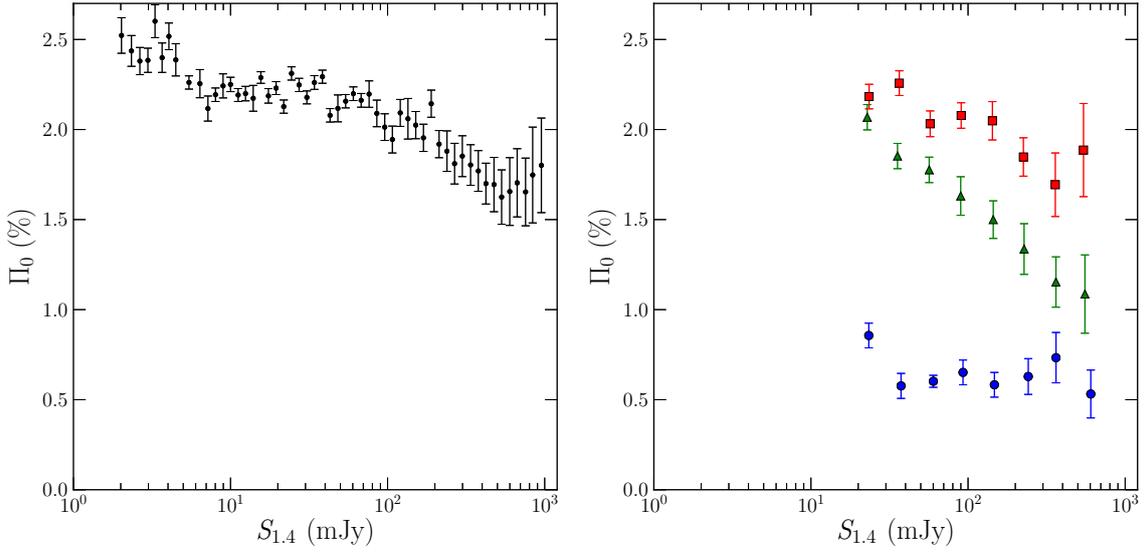


Figure 1: Fractional polarization of radio sources at 1.4 GHz as a function of flux density from stacking NVSS Condon et al. (1998) polarized intensity. Left: results from Stil et al. (2014) using all NVSS sources. Right: preliminary results stacking sources as a function of spectral index in the area of overlap between the WENSS (325 MHz) (Rengelink et al. (1997)) and the NVSS (1400 MHz). Blue circles: $\alpha > -0.3$, green triangles: $-0.75 < \alpha < -0.3$, red squares: $\alpha < -0.75$. The flux density range is limited on the high end because of the samples are smaller, and on the low end by the quality of the spectral indices of flat spectrum sources.

straightforward by subdividing the sample in ranges of total flux density. If additional information is available, possibly only for some part of the survey area, it becomes possible to do additional stacking experiments that provide further insight in the physics of the polarized emission by revealing correlations with observable parameters.

2. Stacking Applications for Cosmic Magnetism

We do not aim to give an exhaustive description of all possible stacking experiments that can be done with an SKA polarization survey. The few applications mentioned here serve to illustrate how this analysis can advance science with the SKA, supplementing other techniques.

2.1 Fractional polarization as a function of redshift

Faraday Synthesis averages the polarization across the a wide frequency range to optimize sensitivity and resolution in Faraday depth. Faraday rotation and depolarization in a source at redshift z occur at the sourceframe wavelength $\lambda_{\text{src}} = \lambda_{\text{obs}}/(1+z)$. For observations related to the evolution of cosmic magnetism it is preferable to compare for example depolarization in similar sources at the same λ_{src} , or equivalent wavelength range in the source frame. If not, interpretation of the results will be model dependent. For Faraday Synthesis, the ability to track the same range in λ_{src}^2 as a function of redshift requires a trade-off with sensitivity, or an additional survey at longer wavelength targeting high-redshift objects.

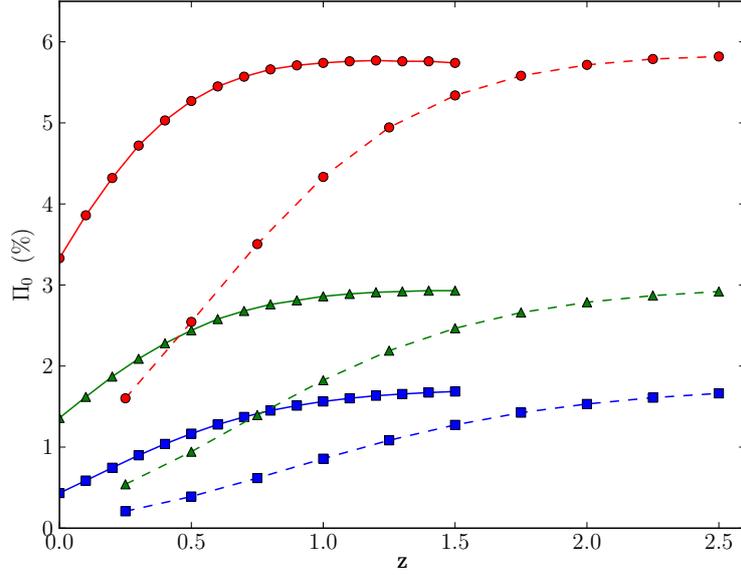


Figure 2: Fractional polarization derived from the peak of the Faraday depth spectrum from Faraday synthesis of models of the integrated polarization of galaxies in SKA-SUR band 3 (1.5 - 4.0 GHz; continuous curve) and SKA-MID band 2 (0.95 - 1.76 GHz; dashed curve). Galaxy models were evaluated for inclination 60° , and ratio of random to regular magnetic field strength 2.0 (red circles), 3.0 (green triangles), and 4.0 (blue squares).

Figure 2 shows fractional polarization from Faraday Synthesis of models of unresolved galaxies from ? in two bands: SKA-SUR band 3 and SKA-MID band 2. It is expected that star forming galaxies make up a majority of the faint radio source population observable with the SKA, but the discussion that follows holds qualitatively for any segment that is affected by depolarization local to the source. Figure 2 shows that the observed polarization increases significantly for same galaxy as a function of redshift as the observed frequency band shifts to shorter wavelengths in the source frame. Qualitatively, one would expect a passively evolving disk galaxy over time to deplete its interstellar medium and decrease its star formation rate per unit mass. We expect that the mean electron density is higher and that the magnetic field is more tangled in a high-redshift analogue of the Milky Way. The effect of galaxy evolution is therefore that depolarization at the same λ_{src} is stronger at higher redshift, a trend opposite to the one observed in Figure 2. As a consequence, interpretation of such data will be model dependent.

Stacking polarization averaged over narrow ranges of λ_{obs} (1 or more channels) for samples of galaxies as a function of redshift will allow us to derive fractional polarization as a function of λ_{em} to the extent allowed by the frequency range of the survey. SKA-SUR band 3 for lower redshifts and SKA-MID band 2 for higher redshifts (and higher sensitivity) are well positioned for this experiment. While Faraday synthesis of these galaxies provides a band-averaged fractional polarization, stacking will benefit from less depolarization at the short-wavelength end. The sample would necessarily require identification of optical counterparts with redshifts, as well as a criterion to reject galaxies with an AGN, for example through the infrared-radio correlation. It will require sample sizes of 10^5 per stack to derive the median polarization to the detection limit in total

intensity. In practice, the analysis will be limited by the size of the optical target catalogue, also because separating galaxies by inclination is important.

Stacking samples with known redshifts allows us to use the broad-band capability of the SKA to explore depolarization in a way that supplements information from Faraday Synthesis. This analysis requires cubes of Stokes I , Q , and U to be archived. In Section 3 we address the question what level of channel averaging can be done without compromising the analysis.

2.2 Probing deeper into the luminosity function

There are various reasons to investigate the polarization properties of less luminous AGN in comparison with higher luminosity AGN. The FRI/FRII morphological distinction is related to luminosity (?). Since part of the radio emission of these sources is beamed, polarization may also be related to (isotropic) luminosity through the orientation of the source axis. An important reason to probe the polarization of starforming galaxies with a lower radio luminosity is to probe the degree of ordering in magnetic fields in galaxy disks as a function of specific star formation rate and dynamics. These are important parameters in dynamo theory through the amount of shear and injection of turbulent energy through stellar feedback (see the chapter on nearby galaxies by R. Beck for a discussion).

Polarization of low-surface brightness galaxies and dwarf galaxies that are faint in the continuum extends our investigation of galactic magnetic fields into a different part of parameter space in terms of dynamics and star formation. Using an HI selected sample that includes subselection on the ratio of HI mass to stellar mass, shape of the line profile (dynamics, shear) and inclination we can address the question whether magnetic fields in quiescent galaxies with a low surface brightness are different from those in high-surface brightness galaxies with strong continuum emission. An HI selected sample has complete redshift information by definition, although this work is limited in redshift by the depth of the HI survey. The advantage of stacking polarization in this case is to probe magnetic field structure in objects with a low radio luminosity that have different properties that affect the galactic dynamo.

2.3 Stacking Stokes Q and U

At first sight it may not appear useful to consider Stokes Q and U individually for stacking. However, if a potential predictor for polarization angle can be identified, its correlation with polarization angle can be tested by stacking samples based on the value of the predictor, or rotated by a parallactic angle based on the predictor before stacking. A simple significance test can be constructed by randomizing the orientation of the sample. If the sample is also stacked in polarized intensity, then the median Q and U may be compared with the median polarization of the sample in order to model the degree of correlation with polarization angle. Possible predictors include optical polarization angle, optical minor axis for disk galaxies, or orientation of the jet.

If a predictor for polarization angle is identified, one can stack the sample as a function of frequency, to investigate if the predictor for polarization angle is equally valid at longer wavelengths. If the sample is generally Faraday thin at the highest frequencies, but experiences significant and diverse Faraday rotation at lower frequencies, this should be observable as the median stacked Q and U converge to zero as frequency decreases. Combining this with the behaviour of the median

fractional polarization in the same frequency range provides an interesting test for models of Faraday rotation in the sample. As before, extending this analysis to the faintest sources detectable in total intensity should be straightforward, depending on the availability of a predictor. Variation of resolution across the band should be considered. It can be avoided by integrating Q and U over the same aperture before stacking.

2.4 Stacking in relation to commissioning and data verification

Stacking requires a sound understanding of the data, including the statistics of the noise and systematics. A statistical analysis may reveal subtle features of the data that are difficult to identify otherwise. Examples are preferred polarization angles in the NVSS survey (Battye et al. (2008)), the effect of clean bias in stacking quasars in the FIRST survey (White et al. (2007)), and bandwidth depolarization in the NVSS (Stil & Taylor (2007)).

Direction-dependent quality control for polarization calibration is challenging because of the limited number of bright polarized sources. A stacking analysis of samples selected by their position in the field (the nearest beam or PAF element in a mosaic tile) can be used as part of the verification of polarization purity of a commissioning survey.

3. Optimal surveys for stacking experiments

Most surveys can be considered for stacking experiments. There are good reasons to stack wide surveys and deep fields. Wide surveys tend to be more effective because of the much larger number of target sources. However, deep surveys target areas of the sky for that are also covered by other deep surveys, allowing for a more sophisticated sample selection. Stacking a narrow deep survey may not provide a better measurement of the median polarization of a sample if the sample size is too small. Including sources near the edge of a survey where the noise is higher may be counter-productive. Sample size and noise are included in Monte-Carlo simulations that correct for polarization bias and estimate error bars. We do not anticipate stacking experiments on images that include zero-spacing data, but note that there may be applications for stacking positions in objects that are much larger than the beam.

Confusion is the most significant factor in deciding the suitability of a survey for a particular stacking experiment. For example, it is not possible to investigate the polarization of SDSS quasars by stacking the NVSS survey, because the density of the targets is so high that one is forced to stack well below the confusion limit of the NVSS. While confusion is a problem stacking total intensity, it is prohibitive for stacking polarization because of blending of sources with different polarization angle.

Higher angular resolution results in a lower tolerance for position errors in the input catalog, and some targets may be resolved. Targets selected from an X-ray or UV catalog have substantial position errors due to limited photon statistics, and the X-ray emission may be offset from the radio position. All of these complicate the interpretation of the stacking result as a representative flux density for the sample. These issues can be mitigated by integrating Stokes I , Q , and U over an aperture centered on the target position by the stacking software.

Channel averaging before stacking can increase the sensitivity before stacking and reduce the footprint of the stacking analysis in terms of data access. Channel averaging is the same as Faraday

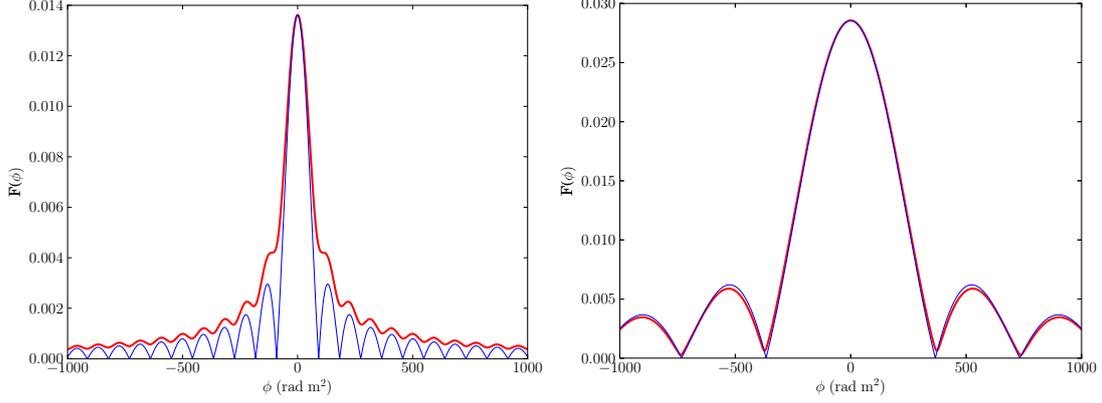


Figure 3: Model Faraday depth spectra (red) and associated RMTF (blue) of the same galaxy at two different redshifts, observed in SKA-SUR band 3. Left: $z = 0$, right: $z = 1$. At low redshift, the Faraday depth spectrum is clearly resolved, revealing Faraday rotation of diffuse emission in the disk. If the same galaxy at $z = 1$ is observed in the same frequency band, the resolution in Faraday depth is much reduced, allowing stacking of Faraday depth spectra.

synthesis under the assumption that Faraday rotation is negligible over the frequency range that is averaged. The question how much channel averaging can be done before stacking is therefore analogous to that of alignment errors in stacking: as long as the errors remain significantly smaller than the resolution of the survey, their effect is small. Potential issues with Faraday thick sources are no different from issues encountered when stacking HI line profiles of galaxies. If some of the targets experience significant Faraday rotation over the frequency range that is averaged, then these sources will appear depolarized after channel averaging Q and U . The difference between stacking channel averaged polarized intensity and channel averaged Q/U can be used as a test to decide if this is significant.

The resolution in Faraday depth in an object at redshift z using only data from wavelength $\lambda_{\text{obs},1}$ to $\lambda_{\text{obs},2}$ is (Brentjens & De Bruyn (2005))

$$\Delta\phi_z = \frac{2\sqrt{3}}{\lambda_{\text{obs},2}^2 - \lambda_{\text{obs},1}^2} (1+z)^2, \quad (3.1)$$

Figure 3 illustrates the effect of redshift on the Faraday depth spectrum of the identical objects observed in the same frequency range. For averaging channels one should always consider at least the Galactic foreground at $z = 0$ and the range of Faraday depths in the sample. At high Galactic latitude the foreground rotation may be estimated and removed with accuracy of a few rad m^{-2} . Compression of the λ^2 range and shift to shorter wavelength allows for a greater tolerance for the unknown Faraday depth of the same if a lower limit for the redshift of the sample is available. For a given uncertainty in Faraday depth of both foreground and the sample, Equation 3.1 can be inverted to derive the wavelength range that can be averaged for stacking polarization.

4. Archive and Computing Resources

4.1 Stacking Computational Resources

Generally, polarization stacking is computationally efficient, and amenable to parallelized implementation. The primary direct cost for median stacking is the selection of median values for each pixel. For stacked images m pixels across, this naturally means m^2 selections must be performed. Modern selection algorithms, such as *quickselect*, as used by Stil et al. (2014), have linear scaling to the number of sources in the average case. Thus, for a stack of N sources with $m \times m$ images, the computational complexity is $\mathcal{O}(m^2N)$, with the same spatial complexity (*quickselect* can be sorted in-place). However, the simple task of generating median values from a generated stack is not the only cost associated with stacking analysis. For example, if one wishes to use oversampled source alignment prior to generating a stack, each of N images must be interpolated to a higher resolution. This task can be accomplished in parallel, with each stacked image oversampled independently.

4.2 Bias Correction Computational Costs

As Stil et al. (2014) showed, bias correction for a polarized intensity stack is non-trivial. With Monte-Carlo realizations, a significant amount of CPU time is required to generate an estimate for the true stacked polarization. For a typical NVSS stack, roughly 5 CPU hours (on an AMD Opteron 2218 CPU) was required. However, little time has been spent optimizing our Monte-Carlo method, and this number can likely be reduced by an order of magnitude or more. Beyond this, the reduction for a large number of stacks containing many sources is extremely amenable to parallelization. A parallel implementation should be capable of scaling to many hundreds to thousands of CPUs.

4.3 Storage and Archive Requirements

While stacking can be used to push well beyond the sensitivity of a survey, a significant new application of this technique to wide-band radio data is to the statistics of samples as a function of frequency. These are not limited to stacking polarization (e.g. Stil et al. 2014). These applications depend on archiving image cubes, as the band-averaged images do not contain the necessary information, although some compromise can be made regarding averaging piecewise over channels. Stacking polarization invariably requires stacking total intensity in the same frequency range.

Working from existing survey images, stacking analysis requires negligible additional storage resources. With double-precision pixel values, each $m \times m$ stacked image requires a meagre $16m^2$ bytes per stacked sample. Stacking analysis on-the-fly (a stacking survey) offers the opportunity to greatly reduce the storage costs associated with storing massive amounts of survey data. By discarding data once a stack has been generated, the storage needed for generating a stack of N sources can be reduced by a factor of N . For large stacks, this can mean a reduction in storage requirements by many orders of magnitude.

References

Battye, R. A., Browne, I. W. A., & Jackson, N. 2009, MNRAS, 2008, 385, 274

- Brentjens, M. A., & De Bruyn, A. G. 2005, *A&A* 441, 1217
- Condon, J. J., Cotton, W. D., Greisen, E. W., Yin, Q. F., Perley, R. A., Taylor, G. B., & Broderick, J. J. 1998, *AJ*, 115, 1693
- Mesa, D., Baccigalupi, C., De Zotti, G., Gregorini, L., Mack, K.-H., Vigotti, M. & Klein, U. 2002, *A&A* 396, 463
- Rengelink, R. B., Tang, Y., De Bruyn, A. G., Miley, G. K., Bremer, M. N., Röttgering H. J. A., & Bremer, M. A. R. 1997, *A&AS*, 124, 259
- Stil, J. M., & Taylor, A. R. 2007, *ApJ*, 663, L21
- Stil, J. M., Keller, B. W., George, S. J. & Taylor, A. R. 2014, *ApJ*, 787, 99
- Tucci, M., Martínez-González, E., Toffolatti, L., González-Nuevo, J., & De Zotti, G. 2004, *MNRAS*, 349, 1267
- White, R. L., Helfand, D. J., Becker, R. H., Glikman, E. & De Vries, W. 2007, *ApJ*, 654, 99