A Matched Filter Technique for Slow Radio Transient Detection and First Demonstration with the Murchison Widefield Array

L. Feng[1], R. Vaulin[1,2], J. N. Hewitt[1], R. Remillard[1], D. L. Kaplan[3], Tara Murphy[4,5], N. Kudryavtseva[6],
P. Hancock[5,7], G. Bernardi[8], J. D. Bowman[9], F. Briggs[10], R. J. Cappallo[11], A. A. Deshpande[12], B. M. Gaensler[4,5,13],
L. J. Greenhill[14], B. J. Hazelton[15], M. Johnston-Hollitt[16], C. J. Lonsdale[11], S. R. McWhirter[11], D. A. Mitchell[5,17],
M. F. Morales[15], E. Morgan[1], D. Oberoi[18], S. M. Ord[5,17], T. Prabu[12], N. Udaya Shankar[12], K. S. Srivani[12],
R. Subrahmanyan[5,12], S. J. Tingay[5,7,19], R. B. Wayth[5,7], R. L. Webster[5,20], A. Williams[7], and C. L. Williams[1]
[1] MIT Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2] Sqrrl Data Inc., 125 Cambridge Park Drive, Suite 401, Cambridge, MA 02140, USA
[3] Department of Physics, University of Wisconsin–Milwaukee, Milwaukee, WI 53201, USA
[4] Sydney Institute for Astronomy, School of Physics, The University of Sydney, NSW 2006, Australia
[5] ARC Centre of Excellence for All-sky Astrophysics (CAASTRO),
Building A28, School of Physics, The University of Sydney, NSW, 2006, Australia
[6] Institute of Cybernetics at Tallinn University of Technology, Akadeemia tee 21, 12618 Tallinn, Estonia
[7] International Centre for Radio Astronomy Research, Curtin University, Bentley, WA 6102, Australia
[8] Department of Physics and Electronics, Rhodes University, PO Box 94, Grahamstown 6140, South Africa
[9] School of Earth and Space Exploration, Arizona State University, Tempe, AZ 85287, USA
[10] Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT 2611, Australia
[11] MIT Haystack Observatory, Westford, MA 01886, USA
[12] Raman Research Institute, Bangalore 560080, India
[13] Dunlap Institute for Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, ON M5S 3H4, Canada
[14] Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA
[15] Department of Physics, University of Washington, Seattle, WA 98195, USA
[16] School of Chemical & Physical Sciences, Victoria University of Wellington, Wellington 6140, New Zealand
[17] CSIRO Astronomy and Space Science (CASS), PO Box 76, Epping, NSW 1710, Australia
[18] National Centre for Radio Astrophysics, Tata Institute for Fundamental Research, Pune 411007, India
[19] Osservatorio di Radio Astronomia, Istituto Nazionale di Astrofisica, I-40123 Bologna, Italy
[20] School of Physics, The University of Melbourne, Parkville, VIC 3010, Australia

Abstract

Many astronomical sources produce transient phenomena at radio frequencies, but the transient sky at low frequencies (<300 MHz) remains relatively unexplored. Blind surveys with new wide-field radio instruments are setting increasingly stringent limits on the transient surface density on various timescales. Although many of these instruments are limited by classical confusion noise from an ensemble of faint, unresolved sources, one can in principle detect transients below the classical confusion limit to the extent that the classical confusion noise is independent of time. We develop a technique for detecting radio transients that is based on temporal matched filters applied directly to time series of images, rather than relying on source-finding algorithms applied to individual images. This technique has well-defined statistical properties and is applicable to variable and transient searches for both confusion-limited and non-confusion-limited instruments. Using the Murchison Widefield Array as an example, we demonstrate that the technique works well on real data despite the presence of classical confusion noise, sidelobe confusion noise, and other systematic errors. We searched for transients lasting between 2 minutes and 3 months. We found no transients and set improved upper limits on the transient surface density at 182 MHz for flux densities between ∼20 and 200 mJy, providing the best limits to date for hour- and month-long transients.

Key words: methods: data analysis – radiation mechanisms: non-thermal – radio continuum: general – stars: variables: general – techniques: interferometric

1. Introduction

Many astrophysical systems and physical processes give rise to radio variability and transient phenomena. These range from propagation effects, e.g., extreme scattering events seen in active galactic nuclei (Fiedler et al. 1987), to magnetic activity in ultracool dwarfs (Hallinan et al. 2007) and magnetars (Gaensler et al. 2005), to explosive events such as supernovae (Soderberg et al. 2010) and gamma-ray bursts (Frail et al. 1997). In addition, there are transient sources of the unknown nature, such as the Galactic center radio transients (Hyman et al. 2005), fast radio bursts (Thornton et al. 2013), and sources with no known counterparts at other wavelengths (Thyagarajan et al. 2011), as well as hypothesized sources of

radio emission, such as exoplanets (Lazio et al. 2004), orphan afterglows (Levinson et al. 2002), and gravitational-wave counterparts from compact binary coalescence (Nakar & Piran 2011).

Despite the numerous observed and predicted radio transient sources, few transients have been detected at radio frequencies <300 MHz. However, recent blind surveys with new wide-field radio interferometers, such as the Murchison Widefield Array[21] (MWA; Lonsdale et al. 2009; Tingay et al. 2013) and the Low-Frequency Array[22] (LOFAR; van Haarlem et al. 2013), are setting increasingly stringent limits on the

---

[21] http://mwatelescope.org/
[22] http://www.lofar.org/

surface density of radio transients at different sensitivities and over a wide range of timescales. In particular, Bell et al. (2014) reported a limit of $<7.5 \times 10^{-5} \deg^{-2}$ above 5.5 Jy at 154 MHz for minute-to-hour transients, Carbone et al. (2016) reported a limit of $<1.0 \times 10^{-3} \deg^{-2}$ above 500 mJy at 152 MHz for minute-to-hour transients, Cendes et al. (2014) reported a limit of $<2.2 \times 10^{-2} \deg^{-2}$ above 500 mJy at 149 MHz for 11-minute transients, Rowlinson et al. (2016) reported a limit of $<6.4 \times 10^{-7} \deg^{-2}$ above 210 mJy at 182 MHz for 30 s transients up to $<6.6 \times 10^{-3} \deg^{-2}$ for 1 yr transients, Polisensky et al. (2016) reported a range of limits ($\sim 10^{-4} \deg^{-2}$ above 100 mJy) for 10-minute and 6 hr transients at 340 MHz, and Murphy et al. (2017) reported a limit of $<1.8 \times 10^{-4} \deg^{-2}$ above 100 mJy between 150 and 200 MHz for transients that lasted 1–3 yr. Jaeger et al. (2012) reported a transient detection that corresponded to a surface density of 0.12 $\deg^{-2}$ above 2.1 mJy over 12 hr at 325 MHz, while Stewart et al. (2016) reported a transient detection that implied a surface density of $1.5 \times 10^{-5} \deg^{-2}$ above 7.9 Jy for $\sim$10-minute transients at 60 MHz.

These surveys, searching for slow transients that are on the timescale of a snapshot or longer (generally $\gtrsim$1 minute), are usually limited by the image rms noise and the detection threshold of a source-finding algorithm. One component of the rms noise is the classical confusion noise σ_c, which arises from a background of faint, unresolved sources (Condon 1974). This is a spatial noise in the image that depends on the source distribution in the sky and the instrument resolution:

$$\sigma_c^2 = \Omega_b \int_0^{S_c} S^2 \frac{dn}{dS} dS, \qquad (1)$$

where Ω_b is the solid angle of the synthesized beam, S is the source flux density, and dn/dS is the differential number density of sources (Thyagarajan et al. 2013). The upper limit of integration S_c is the flux density of a source detected at a particular signal-to-noise ratio $q = S_c/\sigma_c$; usually S_c, referred to as the confusion limit, is determined iteratively until $q = 5$.

As the instrument resolution improves and more sources are resolved, the classical confusion noise decreases. For a modest angular resolution like that of the MWA at $\sim 2'$, however, the classical confusion noise can become a limiting factor in the image rms noise because the classical confusion noise does not decrease with longer integration times. This, in turn, limits the ability of source-finding algorithms to identify faint sources in an image. The theoretical estimate of the confusion limit for the MWA is $\sim$6 mJy at 150 MHz and $5'$ angular resolution (Tingay et al. 2013), while $P(D)$ analysis, using different source count estimates, suggests that the classical confusion noise for the MWA at 154 MHz is $\sim$1.7 mJy beam^{-1} at $2''3$ angular resolution (Franzen et al. 2016).

The classical confusion noise is largely independent of time, however, so it is, in principle, not a limit for detecting fainter but varying brightness (Condon 1974). A simple method for detecting brightness variations below the classical confusion noise is image subtraction. For example, one can subtract images taken at the same local sidereal time to remove both classical confusion noise and sidelobe confusion noise, the latter of which is due to residual synthesized beam sidelobes. For many surveys, however, the images are taken at different local sidereal times, and sidelobe confusion noise can dominate. In this case, image subtraction can be prone to artifacts. Even without image subtraction, CLEAN artifacts

impact radio transient searches. For instance, Frail et al. (2012) found that many transient candidates reported by Bower et al. (2007) were in fact artifacts or the result of calibration errors, and those that were not identified as such were detections at lower significance. A technique that could measure or account for the distribution of artifacts would be able to identify astrophysical transient candidates more reliably.

As many radio transients are expected to be faint ($\lesssim$1 mJy; see Metzger et al. 2015), we seek a method that not only takes advantage of the time-independent nature of the classical confusion noise to detect transients without relying on source-finding algorithms that could otherwise limit the sensitivity of the search, but also has well-defined statistical properties that take into account the distribution of artifacts.

We adapted the well-known matched filter technique (Helmstrom 1968) to detect radio transients in the presence of classical confusion noise by drawing on the experience of the LIGO community (Laser Interferometer Gravitational-Wave Observatory), which developed techniques to detect gravitational-wave signals in the presence of non-Gaussian noise (e.g., Allen et al. 2012; Biswas et al. 2012). This technique operates on the image pixel level, has well-defined statistical properties, and is applicable to variable and transient searches for both confusion-limited and non-confusion-limited instruments. We applied this technique to search for slow transients in the MWA data. In Section 2, we describe the mathematical framework for our radio transient detection technique and derive its statistical properties. In Section 3, we describe the MWA data reduction procedure. In Section 4, we discuss the performance of this technique on real MWA data to demonstrate its potential for sensitive transient searches. In Section 5, we describe the transient search analysis using this technique, discuss the results of our search, and set an improved upper limit on the transient surface density at 182 MHz, the best to date for hour- and month-long transients. In Section 6, we conclude that our technique is capable of detecting faint transients and discuss areas of improvement, as well as future work.

2. Theory

The presence of classical confusion noise limits the ability of source-finding algorithms to identify sources with flux densities near or below the confusion limit, but the source population that contributes to the classical confusion noise is independent of time unless they are genuine transient or variable events. Thus, a transient detection technique that searches for brightness variations on top of a constant signal without relying on a source-finding algorithm is needed to detect transient signals approaching the confusion limit. In this section we describe a technique that identifies transients in individual image pixels despite the classical confusion noise.

Adapted from matched filter techniques, which have been used in engineering applications (Helmstrom 1968) and gravitational-wave astronomy (Allen et al. 2012; Biswas et al. 2012), this technique searches for brightness variations on top of a constant signal in individual pixels without using source-finding algorithms. We derive a new transient detection statistic from this technique, discuss its statistical properties, and relate it to the sensitivity of a radio transient search. Although our formalism is derived for transient detection in the image domain, it is similar to the formalism for source detection in the visibility domain (Trott et al. 2011).

To determine whether or not there is a transient signal in a particular image pixel, we compare two hypotheses: the transient is absent (the null hypothesis H_0), and the transient is present (the alternative hypothesis H_1). Usually, H_0 only includes random noise, but in this case, we add a constant background[23] to H_0 to represent the time-independent brightness contribution from confusion sources (or other steady sources) because we are only interested in the change in brightness over time:

$$H_0 : x_i = c + \sigma_i, \tag{2}$$

$$H_1 : x_i = c + Af_i + \sigma_i. \tag{3}$$

For a fixed pixel, x_i is the measured brightness in the ith snapshot ($i = 1, 2, ..., N$), c is the constant background that follows the distribution characterized by σ_c, σ_i is the rms noise (thermal, sidelobe confusion, and other random errors) measured for each snapshot, and Af_i is the transient signal, where A is the overall amplitude for a light-curve template $f \equiv \{f_1, f_2, ..., f_N\}$. Given the two hypotheses and the data $x = \{x_1, x_2, ..., x_N\}$, we compute the ratio of the likelihood functions known as the Bayes factor or the likelihood ratio as part of hypothesis testing (Neyman & Pearson 1933; Kass & Raftery 1995):

$$\Lambda(x) = \frac{p(x \mid H_1)}{p(x \mid H_0)}, \tag{4}$$

where $p(x \mid H_i)$ is the probability of observing x given that H_i is true for $i = (0, 1)$.

To derive an analytical result, we assume that the image noise follows a Gaussian distribution with $\mu = 0$ and $\sigma_{im} = \{\sigma_1, \sigma_2, ..., \sigma_N\}$, but we show later that our transient analysis does not rely on this assumption. For the MWA data, this noise is a combination of thermal noise, (residual) sidelobe confusion noise, and other random errors. The likelihood functions are thus

$$p(x \mid H_0) = \int \mathcal{N}_0 \exp\left(-\sum_{i=1}^{N} \frac{b_i^2 (x_i - c)^2}{2\sigma_i^2}\right) p(c \mid H_0) dc, \tag{5}$$

$$p(x \mid H_1) = \int \mathcal{N}_1 \exp\left(-\sum_{i=1}^{N} \frac{b_i^2 (x_i - c - Af_i)^2}{2\sigma_i^2}\right) \\ \times p(c \mid H_1) p(A \mid H_1) dcdA. \tag{6}$$

$\mathcal{N}_0$ and $\mathcal{N}_1$ are the normalization factors for a multivariate normal distribution. b_i is the value of the ith primary beam for the given pixel. $p(c \mid H_0)$, $p(c \mid H_1)$, and $p(A \mid H_1)$ are the probability distributions of c and A given the respective hypotheses. We assume flat priors; in other words, we assume that c and A are independent and uniformly distributed, and we estimate them by using a least-squares approach (Sivia & Skilling 2006). To do that, we solve Equations (5) and (6) by approximating the integral with the value at its extremum, i.e., $p(x \mid H_j) \approx \text{const} \times \exp(-\chi_{min}^2/2)$, where χ_{min}^2 is the solution to $\nabla \chi^2 = 0$. Specifically, for Equation (5), where

$$\chi_0^2 \equiv \sum_{i=1}^{N} \frac{b_i^2 (x_i - c)^2}{\sigma_i^2}, \tag{7}$$

its extremum is computed by solving

$$\left.\frac{d\chi_0^2}{dc}\right|_{c=c_0} = \sum_{i=1}^{N} \left.\frac{-2b_i^2 (x_i - c)}{\sigma_i^2}\right|_{c=c_0} = 0; \tag{8}$$

and for Equation (6), where

$$\chi_1^2 \equiv \sum_{i=1}^{N} \frac{b_i^2 (x_i - c - Af_i)^2}{\sigma_i^2}, \tag{9}$$

its extrema are computed by solving the two equations

$$\left.\frac{\partial \chi_1^2}{\partial c}\right|_{c=c_1} = \sum_{i=1}^{N} \left.\frac{-2b_i^2 (x_i - c - Af_i)}{\sigma_i^2}\right|_{c=c_1} = 0 \quad \text{and} \tag{10}$$

$$\left.\frac{\partial \chi_1^2}{\partial A}\right|_{A=A_1} = \sum_{i=1}^{N} \left.\frac{-2f_i b_i^2 (x_i - c - Af_i)}{\sigma_i^2}\right|_{A=A_1} = 0. \tag{11}$$

We note that the choice of using flat priors in this derivation allows us to obtain analytical solutions and simplifies the computation process to demonstrate the technique. The choice of priors will affect the outcome at low signal amplitudes, but accounting for this effect is beyond the scope of this work.

For H_0, the solution $c = c_0$ is an estimate of the constant background level, including the contribution from confusion sources, and it is given by the weighted average of the data:

$$c_0 = \langle x \rangle \equiv \frac{\sum b_i^2 x_i / \sigma_i^2}{\sum b_i^2 / \sigma_i^2}. \tag{12}$$

For H_1, the solution $c = c_1$ is an estimate of the constant background level in the presence of a transient signal; c_1 becomes c_0 in the absence of the transient signal. The other solution $A = A_1$, which is unitless, is the amplitude of the transient signal given the predefined template f, which has brightness units:

$$c_1 = \langle x \rangle - A_1 \langle f \rangle, \tag{13}$$

$$A_1 = \frac{(x, f - \langle f \rangle)}{(f - \langle f \rangle, f - \langle f \rangle)}. \tag{14}$$

The notation $\langle x \rangle$ represents the weighted average of x as in Equation (12), and $(x, f) \equiv \sum b_i^2 (x_i f_i)/\sigma_i^2$ denotes the "weighted" inner product between x and f. Note that an equivalent way of writing $(x, f - \langle f \rangle)$ is $(x - \langle x \rangle, f)$, which we interpret as how well the light-curve template matches the data after the constant component is subtracted. However, $(f - \langle f \rangle)$ is simpler and computationally less intensive to calculate since f is predefined and identical for all pixels, so we write Equation (14) in its given form.

Having determined the best-fit values for c and A, we substitute Equations (12)–(14) into the approximations of Equations (5) and (6), which then lets us solve for Equation (4). As $\Lambda(x)$ is essentially a ratio of exponents normalized by a constant, we rewrite Equation (4) in the form $\Lambda(x) \approx \text{const} \times \exp[\rho^2/2\sigma_\rho^2]$ and define the new quantities ρ to be the "detection statistic" and σ_ρ to be the standard deviation of the ρ distribution:

$$\rho = (x, f - \langle f \rangle) = \sum_{i=0}^{N} \frac{b_i^2}{\sigma_i^2} x_i (f_i - \langle f \rangle), \tag{15}$$
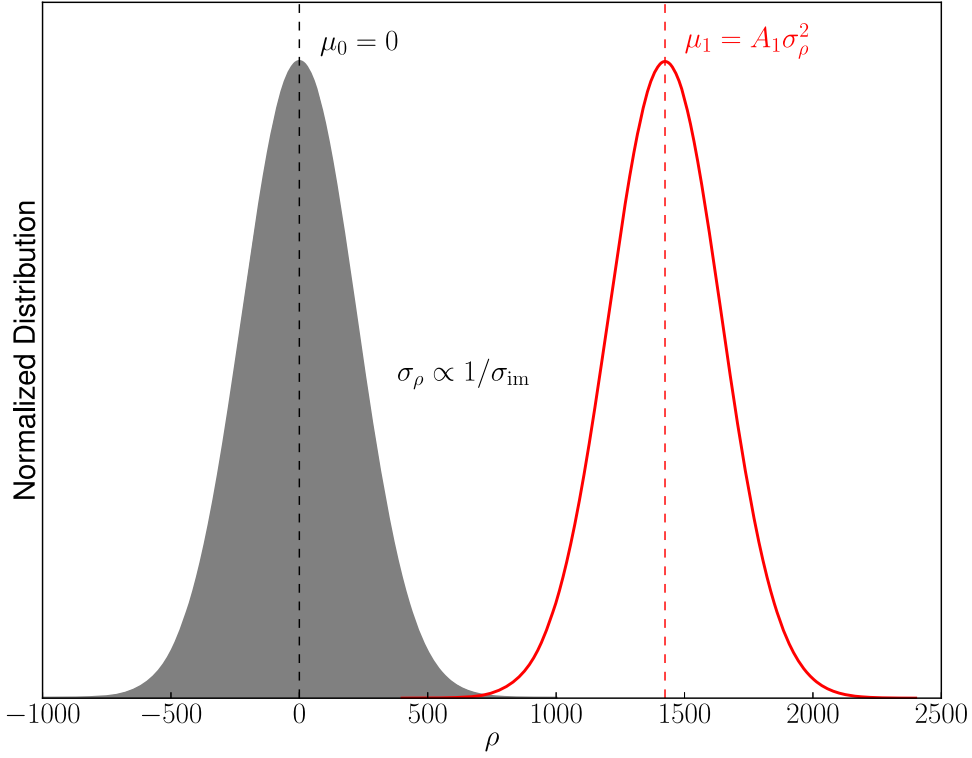
Figure 1. Theoretical distributions of the transient detection statistic ρ: background (gray filled) and signal (red line) with arbitrary image noise $\sigma_{\rm im}$ and signal amplitude A_1. As $\sigma_{\rm im}$ decreases, the widths of the distributions increase according to $\sigma_\rho \propto 1/\sigma_{\rm im}$, but the mean of the signal distribution increases faster according to $\mu_1 \propto 1/\sigma_{\rm im}^2$, so the sensitivity improves as there is better separation between background and signal. Similarly, the brighter the signal (larger A_1), the better the sensitivity.

$$\sigma_\rho = \sqrt{(\boldsymbol{f} - \langle \boldsymbol{f} \rangle, \boldsymbol{f} - \langle \boldsymbol{f} \rangle)} = \sqrt{\sum_{i=0}^{N} \frac{b_i^2}{\sigma_i^2}(f_i - \langle \boldsymbol{f} \rangle)^2}. \quad (16)$$

Here ρ is a modified version of the matched filter (Helmstrom [1968]) and determines the likelihood that $\boldsymbol{x}$ contains the signal $A_1 \boldsymbol{f}$. In other words, it determines which hypothesis is favored (transient absent or present) and by how much. As ρ is a linear superposition of Gaussian random variables, its distribution is also Gaussian with width σ_ρ. When the transient signal is absent, the mean of ρ is $\mu_0 = 0$. When the transient signal is present, the mean of ρ is shifted by the signal and becomes $\mu_1 = A_1(\boldsymbol{f} - \langle \boldsymbol{f} \rangle, \boldsymbol{f} - \langle \boldsymbol{f} \rangle) = A_1 \sigma_\rho^2$. This is illustrated in Figure 1.

Parameters ρ and σ_ρ are also related to the more familiar quantity $\sigma_{\rm im}$ through the weighted inner product: $\rho \propto A_1/\sigma_{\rm im}^2$ and $\sigma_\rho \propto 1/\sigma_{\rm im}$. As $\sigma_{\rm im}$ decreases, both ρ (or μ_1) and σ_ρ increase, but ρ (or μ_1) increases faster than σ_ρ, which leads to a better separation of background and signal (see Figure 1) and hence an improved sensitivity. Similarly, the brighter the transient (larger A_1), the better the sensitivity.

So far we considered ρ for a single template. However, in a real search, we use a bank of templates and maximize ρ over the various parameters that characterize these templates, such as start times and durations, so the practical statistic is

$$\tilde{\rho} = \max[(\rho/\sigma_\rho); (t_0, t_{\rm dur})], \quad (17)$$

where we calculate ρ/σ_ρ for every template and all pixels, and then, for each pixel, we save the maximum value and the template that provided this value. Using the ratio ρ/σ_ρ ensures that all the pixels are drawn from the same standard normal distribution. Note that $\tilde{\rho}$ is the maximum of a Gaussian random variable, so its distribution is no longer Gaussian. The trials factor arising from the bank of templates is included in the $\tilde{\rho}$ distribution through the maximization procedure, unlike the ρ distribution for a single template, where there is no need to correct for the trials factor. We handle the non-Gaussianity of the $\tilde{\rho}$ distribution in Section 4.

There are two more steps before transient identification: characterize the detection significance (reliability) and measure the detection efficiency (completeness). As we describe these steps in detail in Section 4, we summarize them here. To characterize the detection significance, one designates a small part of the image as the playground region, where there are assumed to be no transients; corrects the cumulative distribution of $\tilde{\rho}$ in the playground region by the trials factor, i.e., the number of synthesized beams in the search region compared to the number in the playground region; and extrapolates the tail of this corrected distribution to a tolerable probability of false alarm. The extrapolated distribution determines the significance of any detection during the actual search. To measure the efficiency, one applies the detection threshold to the injected transients and computes the fraction that is recovered. This procedure handles non-Gaussianity in the data, such as artifacts or sidelobe confusion noise, as well as the effects of maximization, thus making it a powerful technique. The pipeline implementation is described in Appendix A.

3. Data Reduction

The MWA is a low-frequency radio interferometer consisting of 128 aperture arrays known as "tiles" located at the Murchison Radio-astronomy Observatory in the Murchison

Table 1
Summary of Observations Used in Our Analysis

Date	Time Range (UT)	Number of Snapshots
2013 Sep 02	16:08:08–17:39:36	46
2013 Sep 04	16:00:16–18:32:48	74
2013 Sep 06	15:52:22–18:24:54	72
2013 Sep 09	15:40:39–18:13:03	75
2013 Sep 11	15:32:47–18:05:11	75
2013 Sep 13	15:24:55–17:57:19	72
2013 Sep 17	15:09:11–17:41:35	76
2013 Sep 19	15:01:19–17:31:43	71
2013 Sep 30	14:17:59–16:50:31	76
2013 Oct 02	15:07:03–16:42:39	47
2013 Oct 04	14:02:15–16:34:47	76
2013 Oct 08	13:46:31–16:19:03	76
2013 Oct 10	13:39:43–16:12:15	76
2013 Oct 23	12:48:39–15:21:03	76
2013 Oct 25	12:40:47–15:13:11	76
2013 Oct 29	12:25:03–14:57:27	75
2013 Nov 18	11:18:31–13:38:55	70
2013 Nov 29	11:30:41–12:56:09	42

Note. Each snapshot is centered on the EOR0 field, or (R.A., decl.) = $(0°, -27°)$, and integrated over 112 s at 182.4 MHz.

Shire of Western Australia; for a complete description of the instrument, see Tingay et al. (2013). The data for this paper were taken according to the commensal MWA observing proposals[24] G0009 ("Epoch of Reionisation," EOR) and G0005 ("Search for Variable and Transient Sources in the EOR Fields with the MWA") for Semester 2013-B. These observations were done using the "point-and-drift" strategy, where the primary beam pointing (beamformer setting) changed every 20–30 minutes to track the field after it drifted across the field of view of the instrument.

We used 1251 snapshot observations of the EOR0 field, which was centered on (R.A., decl.) = $(0°, -27°)$, taken on 18 nights between 2013 September 2 and November 30. We included only snapshots taken when the field center was <20° from the zenith, corresponding to ~2 hr of observation each night, and excluded data from 2013 October 15 because of known ionospheric activity (Loi et al. 2015). Each snapshot is a multifrequency synthesis image integrated over 112 s with a bandwidth of 30.72 MHz centered on 182.40 MHz. Table 1 lists a summary of the observations.

3.1. Preprocessing

Raw interferometric data were converted into the UVFITS format (Greisen 2012) by the Cotter MWA preprocessing pipeline (Offringa et al. 2015). During this process, Cotter used AOFlagger (Offringa et al. 2010, 2012) to flag radio-frequency interference, frequency channels affected by bandpass aliasing (240 kHz on each edge of a 1.28 MHz coarse channel), and known bad tiles, which might vary from night to night depending on the state of the instrument. To decrease the file size, Cotter also averaged the data to 1 s time resolution and 80 kHz frequency resolution.

3.2. Calibration

We developed a data reduction pipeline based on the Common Astronomy Software Applications (CASA) package[25] (v4.1.0; McMullin et al. 2007) and the wide-field imager WSCLEAN (Offringa et al. 2014).

We built a point-source sky model for each snapshot observation, using the 11 brightest point sources in the field after primary beam attenuation according to the MWA Commissioning Survey Catalog (Hurley-Walker et al. 2014). We generated the model visibilities and the calibration solutions in CASA, using in particular the tools `componentlist`, `ft`, `bandpass`, and `gencal`. Then we performed one iteration of phase and amplitude self-calibration with WSCLEAN.

Finally, we averaged over many observations on the same night the calibration solutions generated in the previous step to produce a single calibration solution for this night. This was done in two steps: (1) we selected a list of observations for which the AEGEAN source finder (v951; Hancock et al. 2012) detected at least 1500 sources in each 112 s snapshot, as we found this to be a practical indicator of image quality; and (2) we averaged the calibration amplitude and phase solutions for each tile, polarization, and frequency channel for the selected observations, while ignoring the highest and lowest 10% of the data.

This procedure provided stable and smooth calibration solutions, which were not expected a priori to vary significantly over time and frequency. It also gave more reliable estimates of the source flux densities.

3.3. Imaging

After we applied the average calibration solutions to each snapshot, we generated multifrequency synthesis images over 30.72 MHz bandwidth in the instrumental XX and YY polarizations, setting WSCLEAN to use uniform weighting that gave a synthesized beam of ~2′, a pixel size of 0″.5, and an image size of 4096 × 4096 pixels, which corresponded to a field of view of 34° × 34°. Figure 2 shows an example snapshot of the EOR0 field.

3.4. Primary Beam Correction

We adopted an empirical approach for primary beam correction. While there are theoretical primary beam models for the MWA (Sutinjo et al. 2015), they are less accurate at frequencies $\gtrsim$180 MHz. We found that the source light curves extracted from images corrected by the theoretical primary beam model at 182 MHz showed systematic trends of ~4% flux change per hour that depended on the right ascension of the source. To alleviate this effect, we measured the empirical primary beam value at each source location by comparing the observed flux of the source to its catalog flux, which we assumed to be true and static, and fitted a smoothing spline to the measured beam values. This procedure reduced the systematic trends of the light curves to $\lesssim$2% flux change per hour and removed most of the systematic variations in the light curves. A similar approach to empirical primary beam correction was done by Thyagarajan et al. (2011) for the Very Large Array; for a more detailed discussion on the empirical primary beam we used, see Appendix B.

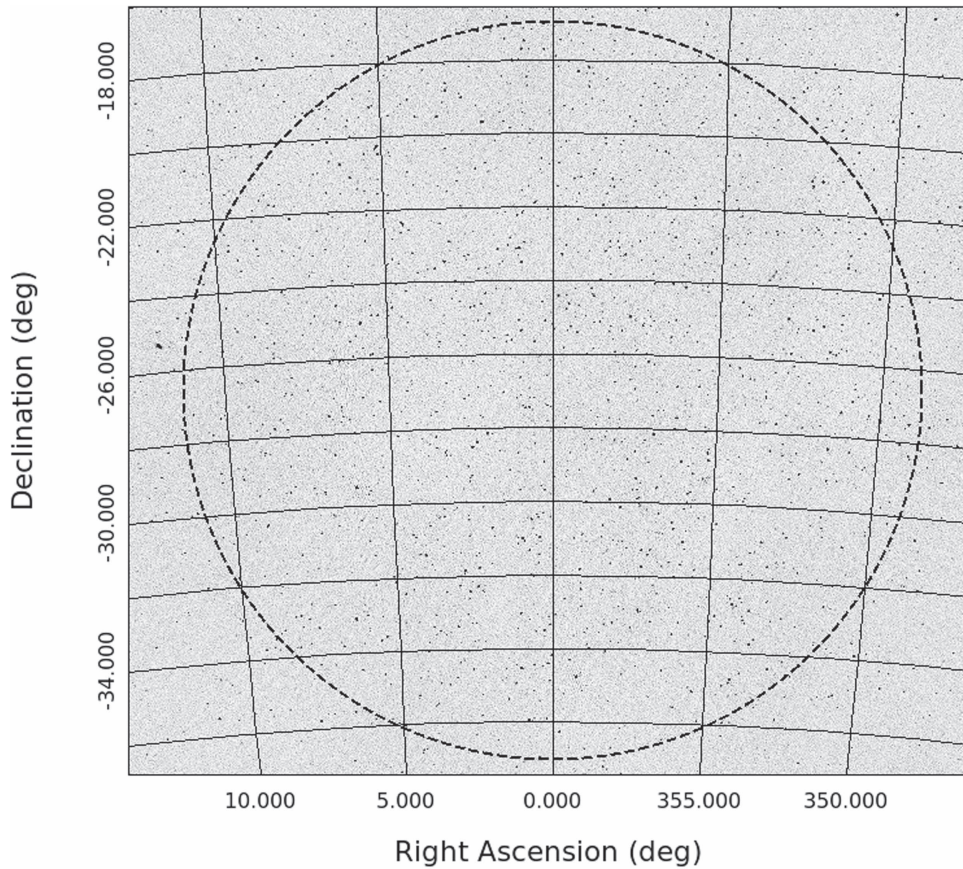[24] http://mwatelescope.org/astronomers/

[25] http://casa.nrao.edu/

Figure 2. Example snapshot of the EOR0 field. It was integrated over 112 s and cleaned in the XX polarization, plotted with the J2000 coordinate grid and a squared color scale. The dashed circle has a radius of 10°, which is approximately the outer boundary of the field included in our analysis.

4. Demonstration

In this section we demonstrate how a transient search with our matched filter technique proceeds and that it performs well on real data. Before we identify possible transient signals according to the matched filter detection statistic, we need to characterize the background distribution of $\tilde{\rho} \equiv \max(\rho/\sigma_\rho)$. Using $\tilde{\rho}$ instead of ρ ensures that all pixels are drawn from the same distribution. Characterizing the background determines the significance of any detection and provides a threshold $\tilde{\rho}^*$ for us to compare the performance of different searches and to measure the search efficiency.

We used a small area ($\sim$10%) labeled as the "playground region" in the images for the background characterization. In this region, we assumed that there were no transient events; existing limits on the rate of radio transients at low frequencies suggest that these events are relatively rare (e.g., Jaeger et al. 2012; Rowlinson et al. 2016), so our assumption should be valid. Significant transients would appear as a tail in the $\tilde{\rho}$ distribution that we would examine further.

To demonstrate our transient detection technique, we present the results for one light-curve template: the top hat with a duration of 15 days and a brightness of 1 Jy beam^{-1}, searched over different start times t_0. In principle, we could have treated the observation time of every 112 s snapshot as a unique t_0, but since computation time scaled with the number of search parameters, we shifted t_0 by $\sim$10% of the duration where there were data, which in this case corresponded to the start of every night of the observation. The distribution of $\tilde{\rho}$ for the entire playground region is presented in Figure 3.

As evident in Figure 3(a), there is a significant tail in the $\tilde{\rho}$ distribution. This is due to residual primary beam effects, as the pixels in the tail are located in the regions where the primary beam is poorly modeled, or they contain bright sources ($>$100 mJy). We masked the pixels beyond $\sim$10° from the phase center and excluded the source pixels by creating a 20×20 pixel square mask centered on the (R.A., decl.) of each source. After masking, we managed to remove the tail in the distribution as illustrated in Figure 3(b). The distribution after masking is what we work with and refer to as the "background distribution."

The background distribution of $\tilde{\rho}$ lets us derive the probability of false alarm P_{FA} (equivalent to reliability) for different values of $\tilde{\rho}$. Here P_{FA} is the probability that our experiment contains a false positive. Every experiment has its own observation timescale and sky coverage, which need to be factored accordingly. We start with a quantity closely related to P_{FA}: the number of background detections $N(\geqslant\tilde{\rho})$ above a particular threshold. We scale $N(\geqslant\tilde{\rho})$ as derived from the playground region to the search region according to the number of synthesized beams for the different sky areas, thus accounting for the trials factor correctly. The trials factor from the number of templates is already accounted for in the background distribution when we maximized over different templates. Then we fit a smooth function to the tail of $N(\geqslant\tilde{\rho})$ to extrapolate the background rate for larger values of $\tilde{\rho}$ where we might not have measurements, as the playground region contains much less data than the search region. Non-Gaussianity and nonthermal components of the image noise,
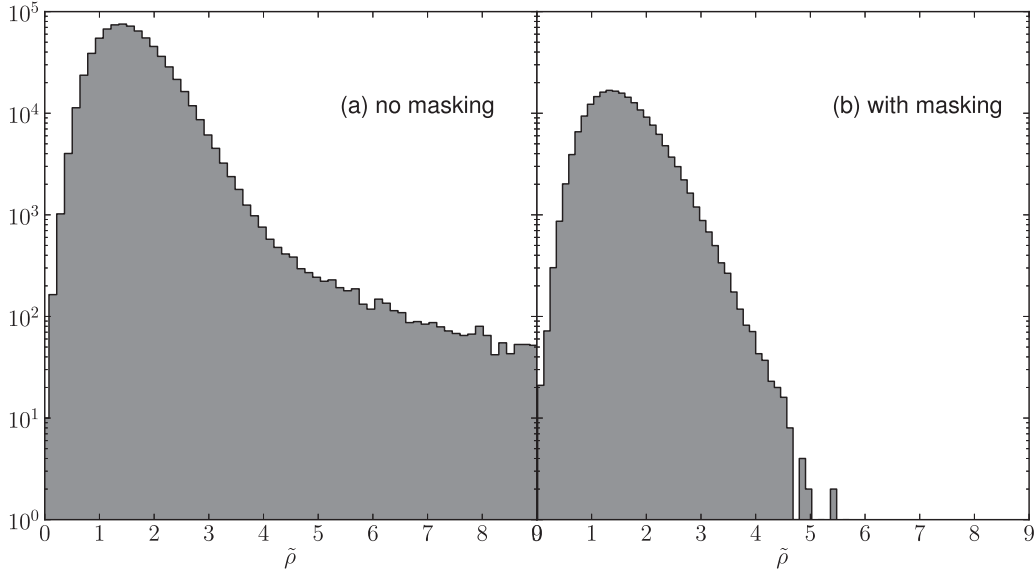
Figure 3. (a) Distribution of $\tilde{\rho}$ for all the pixels in the playground region, with the x-axis limited to $0 \leqslant \tilde{\rho} \leqslant 9$. (b) Same distribution of $\tilde{\rho}$ but without the "bad" pixels, i.e., the pixels in the region where the primary beam is poorly modeled and the pixels that contain bright sources affected by the primary beam systematics. The tail in the background distribution, which extends to $\tilde{\rho} \sim 140$ in the left panel, is removed by masking those bad pixels.
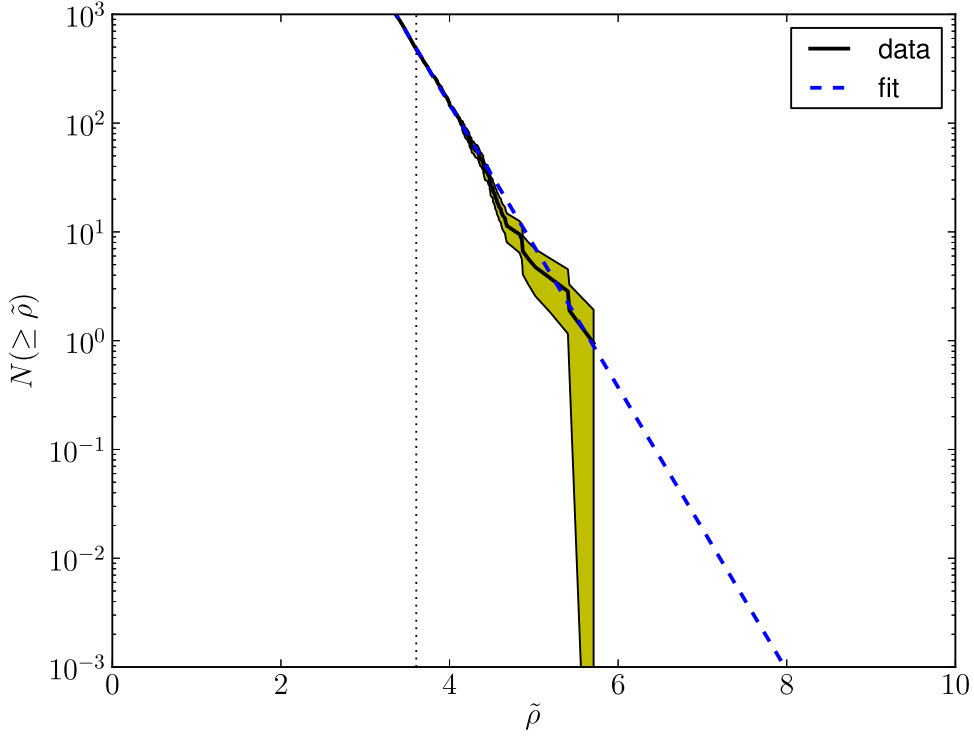
Figure 4. Cumulative background distribution of $\tilde{\rho}$ (black solid line). The shaded area is the error region as computed from $\sqrt{N}$. Only the tail of the distribution (500 data points to the right of the vertical dotted line) was used for fitting. The fit (blue dashed line) is an exponential function: $f_N(\tilde{\rho}) = \hat{N} \exp(-\tilde{\rho}/\hat{\rho}) = 2.38 \times 10^7 \exp(-\tilde{\rho}/0.33)$. The reduced χ^2 for the fit is 0.25. The fit was then used to compute the false alarm probability P_{FA}; see text for the details of the calculation. At $P_{\mathrm{FA}} = 10^{-3}$, the transient detection threshold is $\tilde{\rho}^* = 7.98$.

such as sidelobe confusion and image artifacts, are modeled in this empirical fit. We determine the fit parameters by minimizing the negative log-likelihood for a Poisson distribution, where $N(\geqslant \tilde{\rho})$ is treated as the Poisson mean. As illustrated in Figure 4, an exponential function of the form $f_N(\tilde{\rho}) = \hat{N} \exp(-\tilde{\rho}/\hat{\rho})$, where $\hat{N}$ and $\hat{\rho}$ are the fit parameters, fits the tail of $N(\geqslant \tilde{\rho})$ well within the errors $\sigma_N \equiv \sqrt{N}$. The probability for $N(\geqslant \tilde{\rho})$ to be nonzero, assuming a Poisson

distribution, is $P(N > 0) = 1 - P(N = 0) = 1 - e^{-N}$. We take $P_{\mathrm{FA}} = P(N > 0)$, and since we require $P_{\mathrm{FA}} \ll 1$, $P_{\mathrm{FA}} = P(N > 0) \approx N(\geqslant \tilde{\rho})$. Having determined $f_N(\tilde{\rho})$, we choose a tolerable value of P_{FA} (for example, $P_{\mathrm{FA}} = 10^{-3}$) and solve for $\tilde{\rho}^*$ such that $P_{\mathrm{FA}} = f_N(\tilde{\rho}^*) = N(\geqslant \tilde{\rho}^*)$. For our exponential function, the threshold is of the form

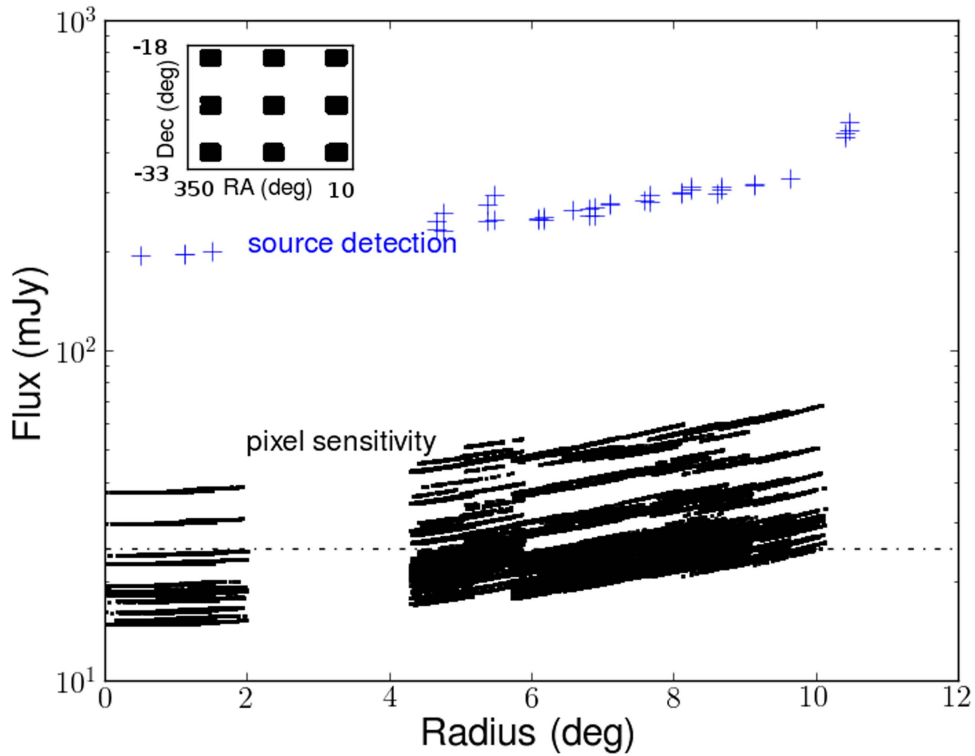$$\tilde{\rho}^* = \hat{\rho}\left(\log \hat{N} - \log P_{\mathrm{FA}}\right). \qquad (18)$$

Figure 5. Flux density sensitivity comparison between the matched filter transient detection technique (using a 15-day top-hat template) and a source detection technique for the same set of 2-minute snapshots. Each black dot is $A^* = (\tilde{\rho}^*/\sigma_\rho) = (7.98/\sigma_\rho)$ of a pixel, and each blue plus sign is $6\sigma_{\mathrm{im}}$ (an example source detection threshold) at a particular radius from the EOR0 field center. The dot-dashed line is the median flux density sensitivity 25.0 mJy quoted in the text for the matched filter technique. The matched filter technique achieves a better sensitivity than a source-finding algorithm for the same set of images without the need to produce a deeper integration image. The sensitivity decreases away from the phase center because of increased noise, but the significance remains the same. The gap between 2° and 4° is because of the discontinuous playground region sampling; see the inset, where black rectangles mark the playground regions, each of which consists of several smaller 86×86 pixel patches. σ_{im} is calculated independently for each patch, so the stripes in A^* correspond to the patches where the noise properties are different.

The threshold $\tilde{\rho}^*$ depends on how much of the distribution tail is used in the fit. For Figure 4, we fitted the tail with 500 points, which gave $\tilde{\rho}^* = 7.98$ at $P_{\mathrm{FA}} = 10^{-3}$. When we fitted 100 points, we obtained $\tilde{\rho}^* = 7.37$, whereas when we fitted 50 points, we obtained $\tilde{\rho}^* = 7.26$. This dependence implies that a single threshold value is not very reliable for identifying transient events near the threshold, so in the actual search we opted for the "loudest event statistic" (Brady et al. 2004; Biswas et al. 2009), which we describe in more detail in Section 5. Nevertheless, the threshold is useful for comparing how well different searches perform, for providing an estimate of the brightness sensitivity, and for verifying the recovery of injected transients.

The flux density sensitivity of the search can be calculated according to $A^* = \tilde{\rho}^*/\sigma_\rho$ (see Section 2; note that $\tilde{\rho}^*$ already contains a factor of σ_ρ, unlike ρ). Strictly speaking, A^* is unitless, so one needs to multiply by the template f to convert it into brightness units, but we drop f for a simpler notation as we choose f to have units of $1\,\mathrm{mJy\,beam}^{-1}$. Because σ_ρ depends on the primary beam, as presented in Equations (12) and (16), pixels closer to the edge of the primary beam have different values of σ_ρ compared to the pixels closer to the center of the primary beam. This implies that we have nonuniform brightness sensitivity across the image, where we are more sensitive to fainter transient sources toward the center of the primary beam. This is illustrated in Figure 5, where we also show an example detection threshold $6\sigma_{\mathrm{im}}$ for a source finder.

To quantify the flux density sensitivity in a single number, we computed A^* separately for each pixel, and then we computed the median value of A^* integrated over 1 beam. Since the distribution of σ_ρ and hence A^* is skewed, the median is a better estimator of the flux density sensitivity than the mean. However, we stress that this value only provides an estimate of how well the search might perform and should not be interpreted as a strict threshold because a single value for the flux density sensitivity is a convenience and cannot capture the complexity of the data.

It is $\tilde{\rho}$ that matters in this technique. Since our technique uses $\tilde{\rho}$ and not flux density as a metric to identify transient sources, it is capable of detecting sources fainter than the median flux density sensitivity at the *same significance* (reliability) but a lower efficiency (completeness), making it potentially more powerful than the techniques that apply a more stringent flux density threshold. For our clean images and a 15-day top-hat template, $P_{\mathrm{FA}} = 10^{-3}$ corresponds to $\tilde{\rho}^* = 7.98$ and a median flux density sensitivity of 25.0 mJy; note that fitting 50 points instead of 500 gives $\tilde{\rho}^* = 7.26$, which corresponds to a median flux density sensitivity of 22.7 mJy, a value not very different from 25.0 mJy.

5. Analysis

We ran three separate blind transient searches on the MWA data. Each search used a different set of light-curve templates and thus was sensitive to transients on a different timescale.
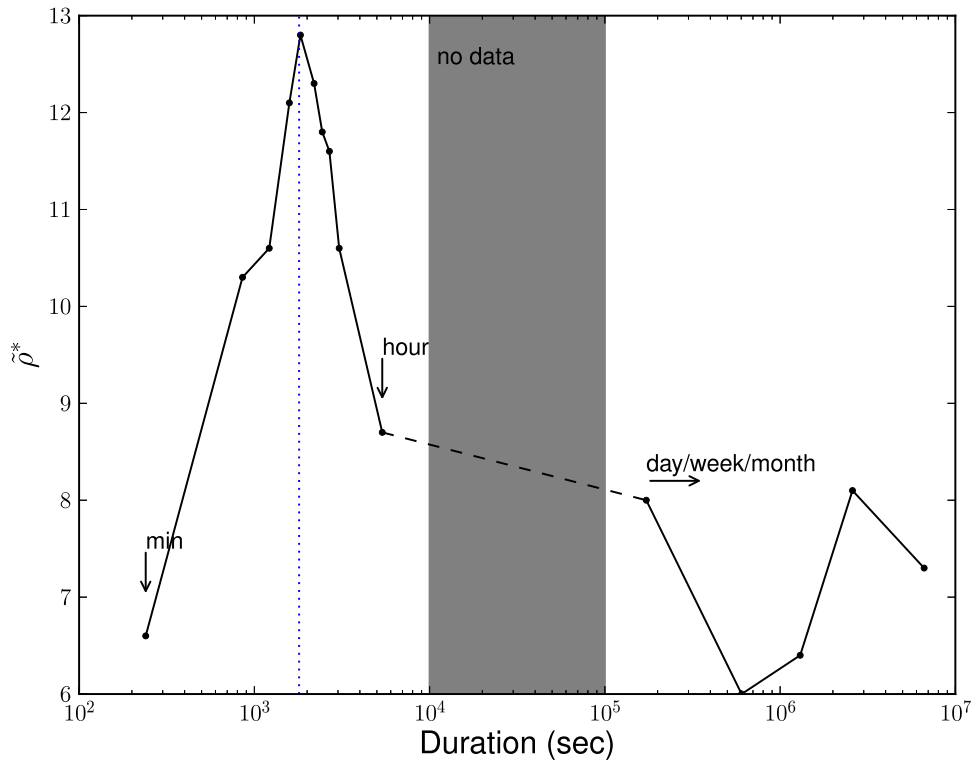
Figure 6. Expected thresholds $\tilde{\rho}^*$ for transients on different timescales, characterized by the matched filter statistic, which depends on the image noise, the light-curve template, and how the observations were taken. Smaller values of $\tilde{\rho}^*$ correspond to better sensitivities. The vertical blue dotted line marks the timescale for consecutive changes in the primary beam pointing; as we had poor sensitivity on this timescale, we excluded it from our search. The vertical gray region marks the approximate timescales that correspond to the gaps between observations where we had no data. Based on this plot, we divided our transient search into three parts: minute, hour, and day-to-month.

5.1. Defining the Searches

The timescales spanned by the data ranged from 2 minutes to 3 months, but we did not have uniform sensitivity over all possible timescales. Some timescales could not be probed because of gaps in the observations, while other particular timescales suffered from systematic effects, such as those due to the periodic change in the primary beam pointing.

In order to determine which searches to run and which templates to use, we did a test run on the playground region to compare the expected transient thresholds for various time-scales. Figure 6 shows the expected thresholds from the test run; we sampled roughly logarithmically between 2 minutes and 3 months but also sampled every ~5 minutes between 15 and 50 minutes. There is a drastic increase in threshold around the 30-minute timescale, which corresponds to the time between consecutive changes in the primary beam pointing. Hence, we excluded that timescale from our search and defined the following three searches: minute, hour, and day-to-month. Despite the variation in $\tilde{\rho}^*$ values for the day-to-month templates, the median flux density sensitivities were about the same, so we grouped them together. The specific templates we used for the searches are listed in Table 2. Because of the primary beam systematics, we avoided templates with durations between 15 and 60 minutes. The 4-minute template is capable of recovering transients with durations <15 minutes, as the difference between ρ computed from the 4-minute template and ρ computed from the perfectly matched template is <10%. As each night consisted of ~2 hr of observation, we chose the 1.5 hr template that is halfway between 1 and 2 hr for the second search. Finally, for the last search, we considered all

Table 2
Summary of Our Searches

Search	Template	Duration	$\tilde{\rho}^*$
1: minute	top hat	4 minutes	7.6
2: hour	top hat	1.5 hr	7.9
3: day-to-month	top hat	2 days, 4 days, 7 days, 9 days	8.0
		11 days, 15 days, 17 days, 28 days	
		30 days, 32 days, 36 days, 38 days	
		51 days, 53 days, 57 days, 77 days,	
		88 days	

Note. $\tilde{\rho}^*$ is the threshold with $P_{FA} = 10^{-3}$.

possible durations sampled by the data, excluding the gaps, and separated by at least 1 day.

Each search consisted of three parts: (1) characterizing the background to determine the significance of any detection, (2) measuring the efficiency at which the transient events were detected, and (3) identifying the transient candidates if there were any.

5.2. Characterizing the Background

We characterized the background distribution of $\tilde{\rho}$ by running the pipeline on the playground region. This region was ~10% of the image, which we assumed contained no transients. First, we divided the inner ~13° of the image, or 3096×3096 pixels, into 86×86 pixel squares. This was because the pipeline ran on one CPU core allocated 3 GB of RAM at any given time and would encounter memory issues if it processed more than ~100×100 pixels each with 1251

brightness measurements. Because of primary beam systematics, we only searched for transients in the inner ~10°, or 2096×2096 pixels, of the image. Then we chose the playground region to be nine 172×172 pixel patches divided into rows of three across the image. This choice sampled uniformly across the image to capture any spatial noise variation.

As mentioned in Section 4, we masked certain pixels to remove the tail in the background $\tilde{\rho}$ distribution. One can consider the source mask as an "auxillary channel" for the search, where events that occur within a certain area of a (bright) source are vetoed. Brighter sources have larger sidelobes and hence a larger "veto" area. To define the veto area and refine the background characterization, we performed an empirical fit to determine the size of the source mask as a function of source flux density. We picked six sources with flux densities above 5 Jy, measured their sidelobe contamination areas in the unmasked transient sky map for the hour search, and fitted a straight line through the two points with the steepest slope to obtain the most conservative relationship between the size of the source mask and the source flux density:

$$\Delta n_{\mathrm{pix}} = 6.6S + 0.5, \qquad (19)$$

where Δn_{pix} is the number of pixels to mask on each side of the source and S is the value of the source flux density. If the fit returned $\Delta n_{\mathrm{pix}} < 10$, however, we set $\Delta n_{\mathrm{pix}} = 10$ to give a 20×20 pixel mask region. This is a conservative value to account for the size of the synthesized beam. We also manually flagged 31 double sources that were misidentified as single sources by AEGEAN. We note that by masking known, bright sources >100 mJy, we were no longer able to search for light-curve variations in these sources although it would not affect our search for fainter transients. This was not ideal for a transient search, but it was necessary given the primary beam systematics in our data; with improvement on primary beam modeling and calibration techniques, it is conceivable that one does not need to mask those pixels.

We then divided the playground region into two parts, A and B, by choosing alternating 86×86 pixel squares. Playground A was used to characterize the background distribution of $\tilde{\rho}$ and to set a threshold $\tilde{\rho}^*$ by extrapolating the tail of the distribution to our choice of false-alarm probability P_{FA} (the reliability of detected candidates), while playground B was used to verify that extrapolation and the trials factor normalization. The extrapolation was done by fitting an exponential function to the tail of the cumulative distribution of $\tilde{\rho}$, as described in Section 4. For all three of our searches, we chose $P_{\mathrm{FA}} = 10^{-3}$ for $\tilde{\rho}^*$, which meant that the probability of detecting a false positive in each search (experiment) is $\leqslant 10^{-3}$. If the distribution is Gaussian, which it is not, this probability corresponds to a significance of 3.3σ. Figure 7 shows the extrapolation and verification for the three searches. While the normalization was set to the number of synthesized beams (independent pixels) in playground B for the verification process, the threshold was determined after normalizing the background distribution to the number of synthesized beams in the search region. The values of $\tilde{\rho}^*$ for the three searches are listed in Table 2.

We mention that another possible way to determine the background distribution is to shuffle the images in time and then use the entire image instead of defining a playground region. This avoids the need to extrapolate the tail of the $\tilde{\rho}$ distribution, but we caution that random shuffling could break any temporal correlations in the systematic errors and change the noise distribution. We did not do this in our analysis. Future work is necessary to determine the timescale on which to shuffle the images that would preserve the true noise distribution.

5.3. Measuring the Efficiency

The efficiency (completeness) of each search characterizes the fraction of real transients successfully recovered and plays a role in the upper-limit calculation of the transient surface density. We determined the efficiency by running the same search on the injection region.

The injection region consisted of 10^4 random pixels sampled over the entire image. For each pixel, we simulated a transient light curve by drawing randomly from uniform distributions of brightness, start times, durations, and other parameters that define the transient shape. The values for these parameters were identified by their pixel number and recorded in an injection file independent of and unknown to the search pipeline. Rather than simulating measurement uncertainties, we corrupted these simulated transient light curves by injecting them into the actual data, i.e., added the brightness of the simulated transient to the observed light curve in the corresponding pixel at the corresponding times. This preserved the noise properties of the real data. These light curves, with injected transient signals, were then processed through the pipeline in the exact same manner as the transient search itself. In other words, the pipeline did not know whether or not the light curves it was processing contained injections. The injection run remained blind in this manner and would not have biased the results. The pipeline returned a list of transient candidates that passed the same criteria applied to the search region. We identified these candidates by their pixel number and compared the transient parameters as recovered by the pipeline, which had been corrupted by noise in the real data, to their "true" values stored in the injection file. We note that gaps in the observed data made comparisons of start times and durations tricky, as an injected transient might have a "true" start time that occurred during a gap in the observation. We corrected for this by identifying the time of the closest observed snapshot before the "true" start (or end) time and compared the "effective" start times and durations since a real transient that starts before an observation is indistinguishable from a transient that starts at the beginning of an observation.

For all three searches, we injected transients with the top-hat profile, sampling uniformly in transient duration, start time, and brightness (amplitude). For the day-to-month search, we also injected transients with the fast rise and exponential decay (FRED) profile to mimic what real transients might look like, e.g., radio flares from X-ray binaries (Lo 2014), sampling uniformly in characteristic rise and decay times, start time, and peak flux. All injection parameters are listed in Table 3.

After running the search on the injection region, we applied a cut on $\tilde{\rho}$, choosing only the events with $\tilde{\rho} \geqslant \tilde{\rho}^*$. These were the recovered transient events. By computing the ratio of the number of recovered events to the total number of injected events, we determined the efficiency as a function of flux density for each search, as shown in Figure 8. Since sensitivity improves with lower image noise or longer integration time, the
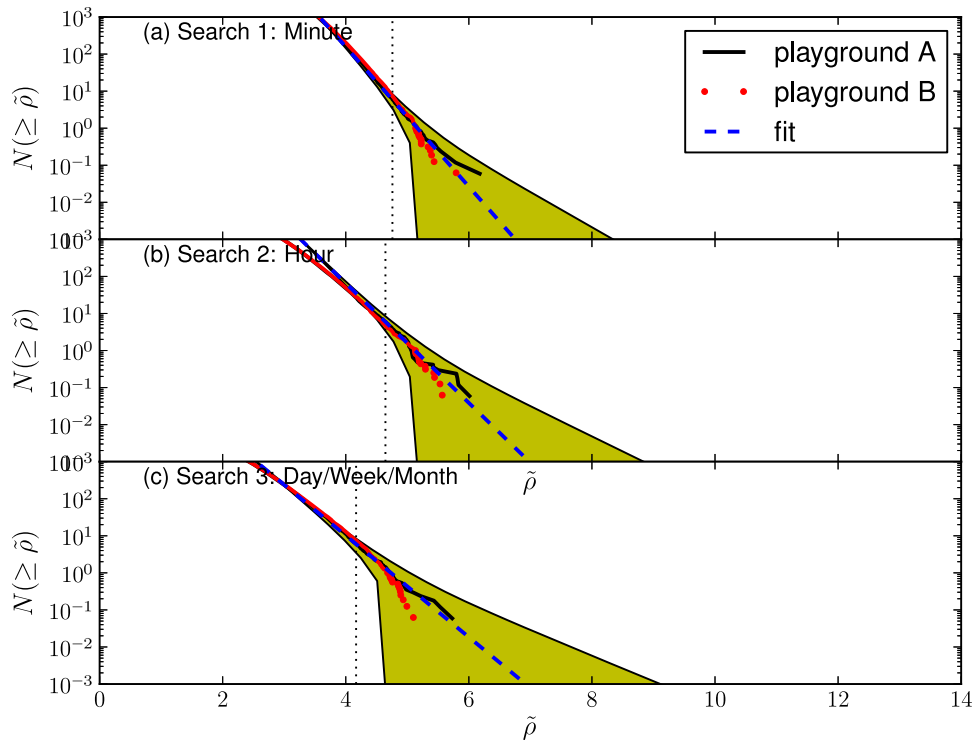
10

Figure 7. Cumulative background distribution of $\tilde{\rho}$ for the playground region for the three searches. The blue dashed line is an exponential fit to the tail of the playground A distribution, where the tail consists of 100 points to the right of the vertical dotted line. The yellow shaded region is the $\sqrt{N}$ error region for the fit, showing that the distributions from playground A and playground B agree within the error bars. All curves are normalized to the number of synthesized beams in playground B. The fit determines $\tilde{\rho}^*$ at a predetermined $P_{FA} = N(\geqslant \tilde{\rho})$ for $N(\geqslant \tilde{\rho}) \ll 1$ when it is normalized to the number of synthesized beams in the search region. See text for further discussion.

Table 3
Summary of Our Injection Runs

Injection	Template	Duration	Peak Flux
1: minute	top hat	2–12 minutes	<1300 mJy
2: hour	top hat	1–2 hr	<450 mJy
3: day-to-month	top hat	1–90 days	<150 mJy
	FRED	$\tau_1 = 1$–2 days, $\tau_2 = 30$–40 days	<160 mJy

Note. We sampled uniformly in durations and peak fluxes, as well as start times (not listed) that spanned the entire 3 months of observation. For the FRED profile, τ_1 is the characteristic rise time, and τ_2 is the characteristic decay time.

day-to-month search was able to recover fainter transient sources at a higher efficiency than the minute or hour search.

Since we also knew the true injection parameters, we checked the accuracy at which the pipeline recovered these parameters, as shown in Figures 9 and 10. The pipeline was able to recover injected parameters fairly accurately ($\lesssim 10\%$–30%, 1σ errors), even when the search was run with top-hat templates on injected transients with the FRED profile. This demonstrates that we are capable of detecting real transients in the data, if they are present, despite using simple top-hat templates for the search. This is because a top-hat template will always have significant overlap with a transient signal that rises and then falls on a similar timescale, regardless of the exact shape of the transient signal (see Equation (15)). However, the significance of a detection from a template that does not match exactly will be less than that from an exact match, as shown from the lowered efficiency and accuracy of recovering FRED

profiles using top-hat templates. On the other hand, if there is a detection, one can rerun the search with better-matched templates to measure the transient properties more accurately.

5.4. Identifying the Candidates

Finally, we ran the pipeline on the search region, which contained $\sim 3 \times 10^6$ image pixels or $\sim 2 \times 10^5$ synthesized beams, where one synthesized beam contained ~ 16 image pixels. We applied the cut $\tilde{\rho} \geqslant \tilde{\rho}^*$ on individual pixels to identify transient candidates. If multiple pixels passed the cut, we grouped the adjacent ones together and considered them as one candidate because the image was oversampled. Since we computed $\tilde{\rho}$ for every pixel, we could produce a "transient" sky map, as shown in Figure 11, where each pixel contains the corresponding $\tilde{\rho}$ instead of brightness. This map visualizes the variability on a particular timescale across the image, and a transient candidate would stand out as a source.

We found no transient candidates. The properties of the loudest events are listed in Table 4. The $\tilde{\rho}$ distributions from the search region agree very well with the background expectation, as shown in Figure 12.

5.5. Limits

As we did not detect any transient candidates, we placed an upper limit on the transient surface density. We based our upper-limit calculation on the "loudest event statistic" as derived by Brady et al. (2004) and Biswas et al. (2009), which meant that we used the largest observed $\tilde{\rho}$ instead of the search threshold $\tilde{\rho}^*$ to determine our search efficiency. This formalism does not rely on the threshold or the extrapolation described in
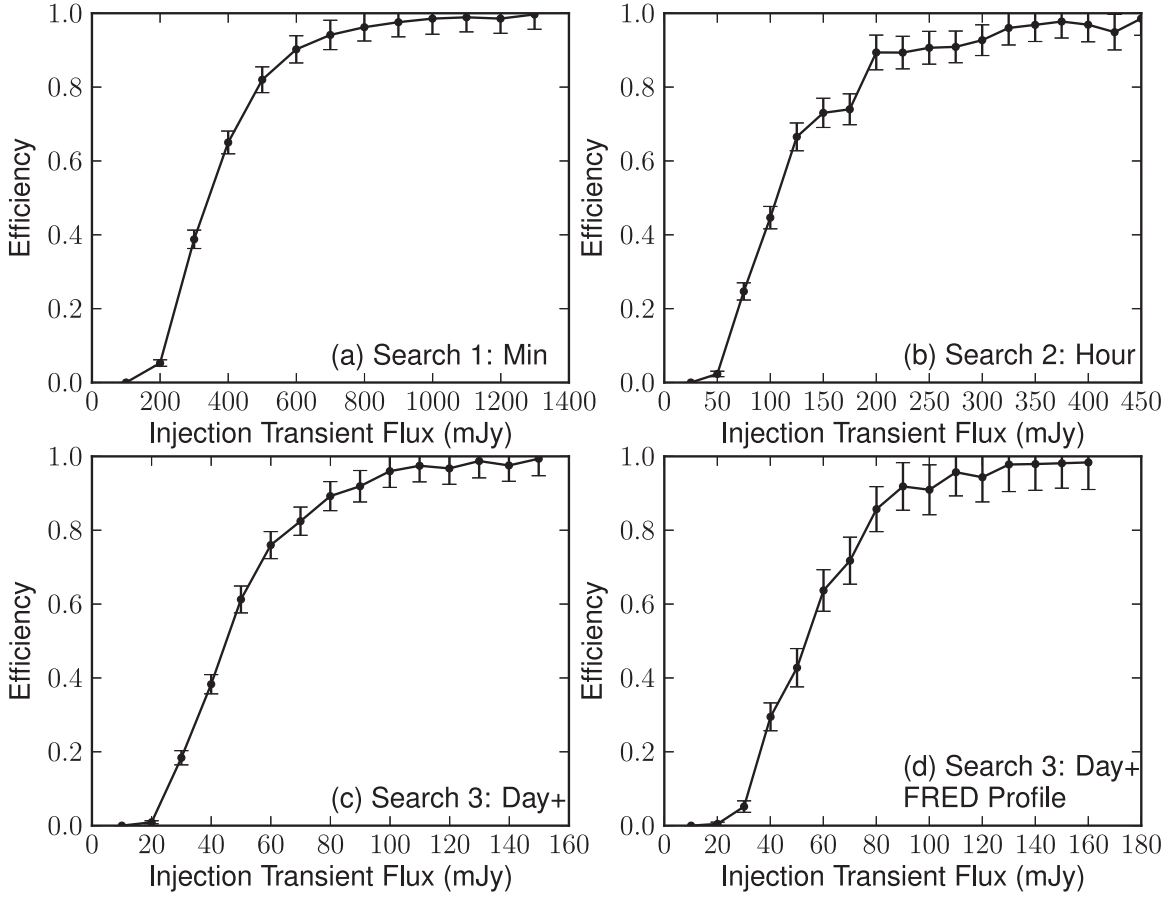
Figure 8. Efficiency (completeness) of recovered transients for the three searches. The efficiency increases at lower flux densities as the searched duration increases because longer durations imply longer integration times and lower image noise. Panel (d) shows the efficiency for when the transients were injected with the FRED profile, which is qualitatively different from the top-hat templates used in the search. It is not drastically different from the efficiency for when the transients were injected with the top-hat profile as in panel (c), demonstrating that the top-hat template is capable of recovering transients that are not top hat in shape.

Section 4, but on the brightest event actually observed in the data, making it independent of a threshold that could be set differently, and is very similar to the two-epoch equivalent snapshot rate introduced by Bower et al. (2007), except that it takes into account the search efficiency. We note that our upper-limit formulation is only applicable to the case of no detections since we detected no transient events, although the loudest event statistic in its full formulation can be used in the case of a detection (see Brady et al. 2004; Biswas et al. 2009).

The probability that we detect no events above $\tilde{\rho}$, assuming that the number of astrophysical transient events with a certain flux density is described by a Poisson distribution, is

$$P(\tilde{\rho}) = e^{-\mu \epsilon (\tilde{\rho})}, \qquad (20)$$

where $\mu = \Sigma \Omega N_e$ is the Poisson mean, Σ is the transient surface density, Ω is the area of each searched image, N_e is the number of epochs or independent time samples, and $\epsilon (\tilde{\rho})$ is the search efficiency as a function of flux density evaluated at $\tilde{\rho}$. The upper limit on Σ at a particular confidence level p is then determined by $P(\tilde{\rho}_m) = 1 - p$ or

$$\Sigma_p = -\frac{\ln(1 - p)}{\Omega N_e \epsilon (\tilde{\rho}_m)}, \qquad (21)$$

where $\tilde{\rho}_m = \max(\tilde{\rho})$ is the loudest event statistic.

We computed the upper limit at 95% confidence level separately for each of our three searches, as they probed

different timescales that corresponded to different astrophysical sources or processes. We determined Ω by multiplying the number of pixels searched and the area of each pixel, which gave us $\Omega = 186 \ \mathrm{deg}^2$. We determined N_e by dividing the whole observation period, i.e., the time between the first and the last snapshot, by the transient duration, or the timescale of the search; if there were gaps in the observation that were as long as the transient duration, we subtracted the number of gaps from N_e. For our three searches, we have 625, 28, and 3 epochs, respectively.

Figure 13 shows our results and compares them to the published results on the transient surface density between 150 and 330 MHz. Our technique allows us to explore a larger phase space more efficiently. Despite using images each with a 2-minute integration time, we achieved sensitivities equivalent to longer integration times for the longer-duration transient searches. Although our results do not set more stringent limits at the same flux densities compared to Rowlinson et al. (2016), also an MWA result, their analysis covered a bigger sky area (452 deg^2) and a longer observation period ($\sim$80 hr integration time spanning 1 yr). If we naively scaled Ω and N_e to match theirs, our limits would be comparable, e.g., $<5.6 \times 10^{-6}$ deg^{-2} (ours at 100% efficiency) compared to $<6.6 \times 10^{-6}$ deg^{-2} for 4-minute transients; we gained in the sense that our technique allows us to probe different timescales without the need to produce deeper integration images to achieve better sensitivities. The limits on these timescales will improve
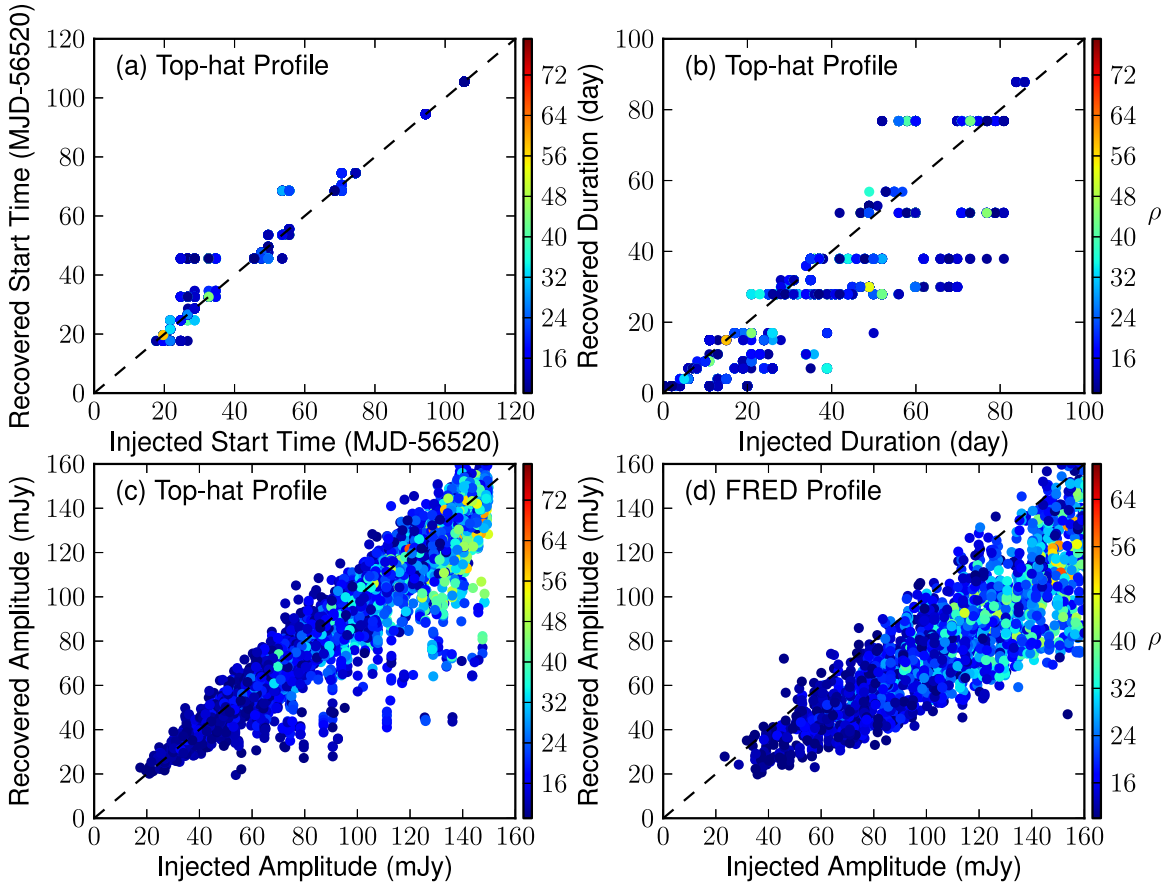
Search 3: Day/Week/Month

Figure 9. Recovery of transient parameters. The pipeline is able to recover injected transient parameters fairly accurately; see also Figure 10. There is a slightly bigger spread in the recovered durations, but that is because the search used a limited set of durations compared to the uniform sampling of injected durations. The flux densities recovered for the FRED profile injections are systematically lower than the injected peak fluxes, and that is a result of the qualitative difference between the injected and the searched light-curve profiles; the FRED profile is sharper than the top-hat profile, so the recovered flux density is smeared out. Although we only show this for the day-to-month search, the pipeline performs similarly for the other two searches.

simply by adding more data, pushing toward lower and lower transient surface densities at the same flux density sensitivities. Pushing toward fainter flux densities would require better calibration techniques or primary beam modeling to decrease the image noise. Even with our current data, we reported improved limits and the best to date at 182 MHz for flux densities between ∼20 and 200 mJy for hour- and month-long transients.

Our limits are also consistent with the reported detections of radio transients. The transient reported by Jaeger et al. (2012) was much fainter than the sensitivity we could achieve with our data even though it occurred on a timescale that we probed (∼day); if we assume a typical spectral index of −0.7, the source they detected at 2.1 mJy at 325 MHz would be 3.2 mJy at 182 MHz, which is an order of magnitude fainter than our best flux density sensitivity at ∼20 mJy. While our limit for the day-to-month search appears to overlap with the transient detection reported by Hyman et al. (2009), our results are still consistent with a nondetection. Their transient lasted ∼6 months, which is longer than the total observation time of our data. Furthermore, their transient was detected near the Galactic center, where it is plausible that the transient population and hence transient rate might be different from the extragalactic transient population, which we observed. If the transient population is similar, however, but with more

data, we should also begin to detect such transients. Likewise, with more data, we should be able to detect the transient population reported by Stewart et al. (2016), although their transient search was conducted at 60 MHz, a lower frequency that might be probing different physical processes.

6. Conclusion

We developed a transient detection technique based on matched filters to search for transients in the presence of classical confusion noise. It searches for the light-curve template that best matches the brightness variation above a constant signal in an individual pixel. The criterion for identifying transient candidates is set by the transient detection statistic ρ, which follows a well-defined distribution characterized by σ_ρ. The empirical background distribution of $\tilde{\rho} \equiv \max(\rho/\sigma_\rho)$ determines the probability of false alarm P_{FA} (reliability), which characterizes the significance of any detection and can be used to establish a threshold $\tilde{\rho}^*$ to compare the performances of different searches and verify the recovery of injected transients.

For every pixel, $\tilde{\rho}^*$ can be converted to a flux density sensitivity according to $A^* = \tilde{\rho}^*/\sigma_\rho$. As different pixels have different noise properties, A^* varies across the image, but the significance of the detection remains the same. The median flux density sensitivity, as computed from all the pixels, provides an
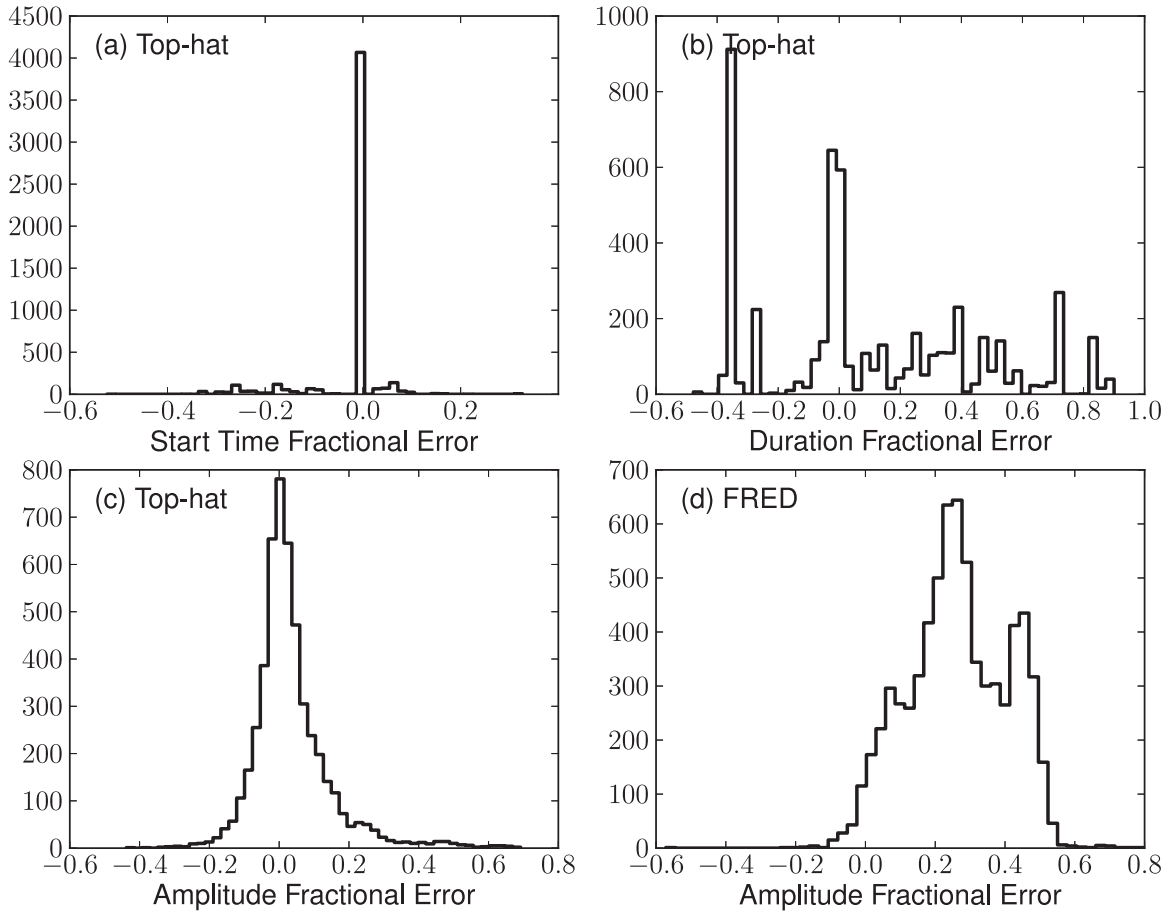
Search 3: Day/Week/Month

Figure 10. Accuracy of recovered transient parameters, corresponding to the panels in Figure 9. The accuracy of each light-curve parameter p is quantified by its fractional error, $(p_{inj} - p_{rec})/p_{inj}$, except for the start time, which is characterized by the start time difference relative to the injected duration, $(t_{0,inj} - t_{0,rec})/(t_{dur,inj})$. The means of the distributions are (a) -1.9%, (b) 10.2%, (c) 3.3%, and (d) 26.3%; the standard deviations are (a) 7.8%, (b) 35.5%, (c) 11.6%, and (d) 14.3%.

estimate of the flux density sensitivity of the search but is not a strict threshold below which nothing is detectable. The efficiency (completeness) is determined by recovering injected transients with light-curve parameters drawn from known or expected astrophysical distributions.

We applied our technique for the first time to real data and demonstrated that our technique performs well despite the presence of residual sidelobe confusion noise and calibration errors. We performed an example search, using the MWA, for transients that have light curves resembling top hats with a duration of 15 days. For this type of transient, our technique is capable of detecting transients with fluxes $\sim$25.0 mJy at $P_{FA} \leqslant 10^{-3}$ for the experiment. As our technique identifies transient candidates by applying a cut on $\tilde{\rho}$ and not flux, it remains sensitive to fainter transient sources at the same significance level but a lower efficiency. This is in contrast to flux-limited source detection techniques.

The ability to detect fainter transients in the presence of classical confusion noise increases the transient parameter space that a particular instrument can explore. As calibration techniques and primary beam modeling improve, one can push the limits of an instrument even further to study astrophysical transient sources that might be missed by source-finding algorithms. Our technique is also applicable to non-confusion-

limited instruments and provides a way to study fainter source variations.

We used this technique to search for transients in 3 months of MWA data. We ran three separate blind searches, using top-hat templates, to probe transients on different timescales: minute, hour, and day-to-month. For each search, we first characterized the background distribution of $\tilde{\rho}$, which allowed us to set the threshold $\tilde{\rho}^*$ above which events were considered to be transient candidates. This threshold corresponded to a false-alarm probability of 10^{-3}, i.e., the probability that a candidate is a false positive (reliability) in the entire search is $<10^{-3}$. Then we characterized the efficiency of each search, or completeness, by running transient injections. The injections also demonstrated that we were able to recover transient properties accurately, even if the light-curve profile of the injected transient differed qualitatively from the light-curve template used in the search.

We found no transient candidates. Thus, we set an upper limit on the transient surface density for each of our searches. We took into account the search efficiency in our upper-limit calculation; thus, we were able to push to fainter fluxes than would otherwise be available. We reported improved limits at fluxes between $\sim$20 and 200 mJy for hour- and month-long transients, the best to date at 182 MHz. This is consistent with
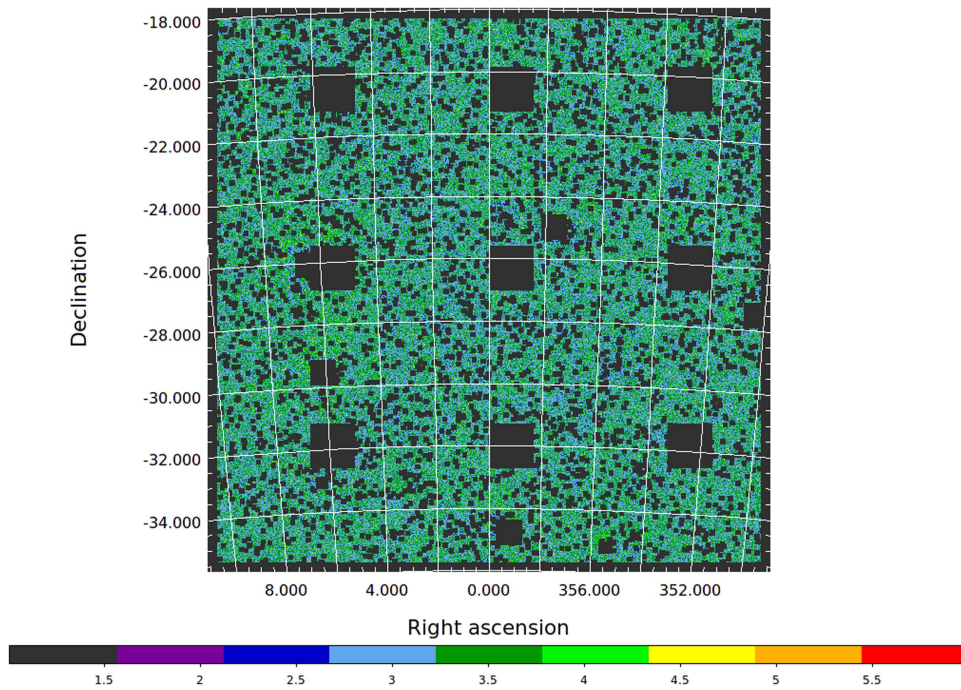
Figure 11. Example map of the transient sky, where the value of each pixel is $\tilde{\rho}$ instead of brightness. This is for the minute search, so it shows the variability across the image on the minute timescale. The largest nine black patches make up the playground region and are not part of the map. The other smaller black squares are masked pixels, regions excluded from the search because they surround radio sources with flux densities >100 mJy.

Table 4
Summary of the Loudest (Brightest) Events

Search	$\tilde{\rho}$	R.A. (deg)	Decl. (deg)	Amp (mJy)	Duration	Start Time (MJD)
Minute	6.1	355.16	−18.34	220.8	4 minutes	56,539.72
Hour	6.6	353.60	−32.41	58.3	1.5 hr	56,614.52
Month	7.0	351.64	−25.86	18.7	30 days	56,537.67

reported transient detections in the literature, and it will easily improve with more data.

Appendix A
Pipeline Implementation

We describe the *Simetra* pipeline, a `python` implementation of the matched filter transient detection technique. The pipeline determines the transient detection statistic and the parameter values of the light-curve template that best match the observed light curve for every image pixel. One can also use this pipeline to inject transient light curves with known parameters into the data before running the transient search; this determines the search efficiency. The pipeline is illustrated in Figure 14, and the code is available online.[27]

---

[26] http://www.astropy.org
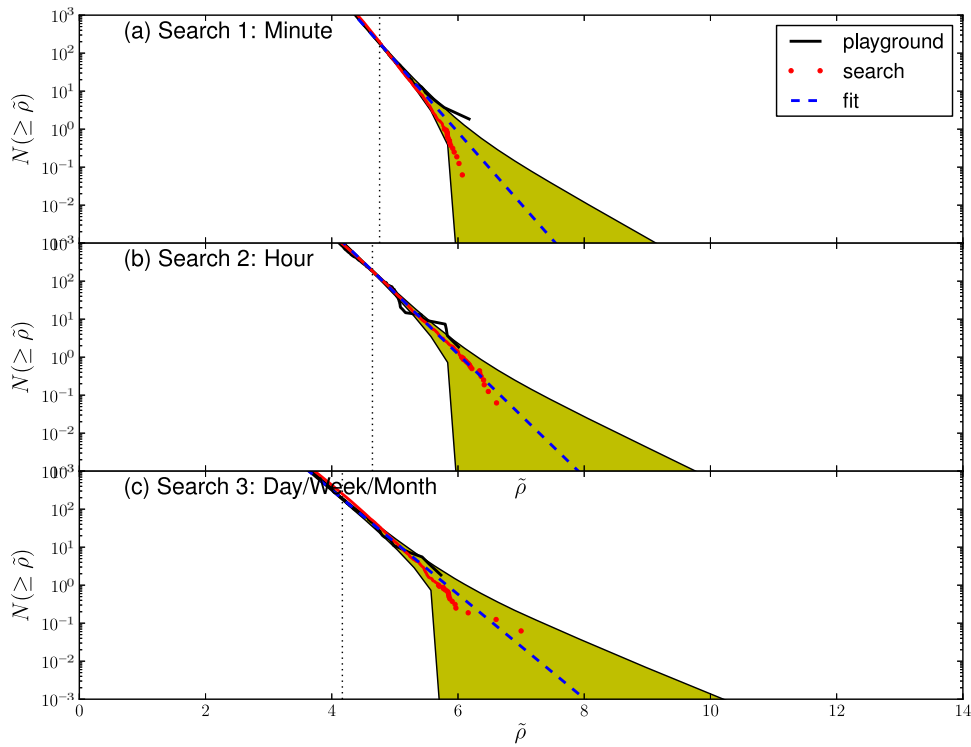[27] https://github.com/lufeng5001/simetra

Figure 12. Cumulative distributions of $\tilde{\rho}$ for the three searches. This is similar to Figure 7, but instead of comparing two playground regions, we compare the search and the playground regions. The tail to the right of the vertical dotted line was used for the fit. The searches are consistent with the background expectation and returned no transient candidates. See Table 4 for a summary of the loudest (brightest) events.

The pipeline takes, as input, a list of sky images, the corresponding primary beam images, and a choice of light-curve template. At the moment of writing, two template choices, the top hat and the power law, are available, but other choices are easy to implement as the code is designed to be as modular as possible. The pipeline reads the input FITS images and converts the flux density and primary beam values for every pixel into the time series x and b, i.e., light curves for every pixel. Out of memory consideration, the pipeline only loads a subset of the image pixels each time. The pipeline then estimates the noise σ_{im} in the subimages by calculating the median absolute deviation and then converting that into the standard deviation to account for outliers in a robust manner.

Here the user has the option to inject transients. The user chooses the type and the number of transient light-curve templates to inject and specifies the range and the distribution of the template parameters. Given these parameters, the pipeline generates the light curves on the fly and injects them into the data before it runs the transient search. We decided to inject transients directly into the pixel light curves instead of the visibilities because of computation concerns.

The transient search is the matched filter calculation. First, the pipeline generates a phase space of template parameters, which can be different from the injection parameters depending on the user's choice. Then it iterates over every set of parameters, including all possible transient start times, generates the corresponding light-curve template f, and calculates ρ and σ_ρ according to Equations (15) and (16). Gaps in the data are handled properly by sampling f at the existing image time stamps and do not pose a problem. Finally, the pipeline outputs a FITS table that contains the most significant (ρ, σ_ρ) and the corresponding template parameters for every pixel. As the amplitude can be determined by $A_l = \rho/\sigma_\rho^2$, it is not stored in the output file.

The pipeline is highly parallelizable and not limited by computation, as the matched filter calculation for each pixel is independent. I/O, i.e., converting the image FITS files into light curves stored as `numpy npz` files, is the bottleneck, but it only needs to be done once. Afterward, the maximization over different start times is the slowest step and scales with the number of time samples. For $\mathcal{O}(10^3)$ time samples, the computation time per 86×86 pixels, which was the smallest image unit we processed, for the entire pipeline was $\mathcal{O}(\mathrm{min})$. We ran this process on 1 CPU core with 3 GB of allocated RAM on a computing cluster that comprised 14 computers with 99 GB of RAM and 24 CPU cores per machine, each core operating at a speed of 60 MFLOPS. Running the pipeline for the full search of this paper, for example, can be completed in <1 week with 36 CPU cores.

Appendix B
Empirical Primary Beam

The primary beam model establishes the flux scale during calibration and after imaging, so there have been many efforts to measure and model the primary beam of the MWA. For example, there is a project to map the beam pattern with an octocopter and a transmitter, while another used ORBCOMM satellites to measure the beam pattern of an MWA tile at 137 MHz (Neben et al. 2015). However, extrapolating this result to other MWA frequencies is not straightforward, so we need to rely on antenna modeling.

The best available model for the MWA primary beam is that developed by Sutinjo et al. (2015). It improved on the previous model, which treated each antenna element as a Hertzian dipole, by incorporating mutual coupling between the elements
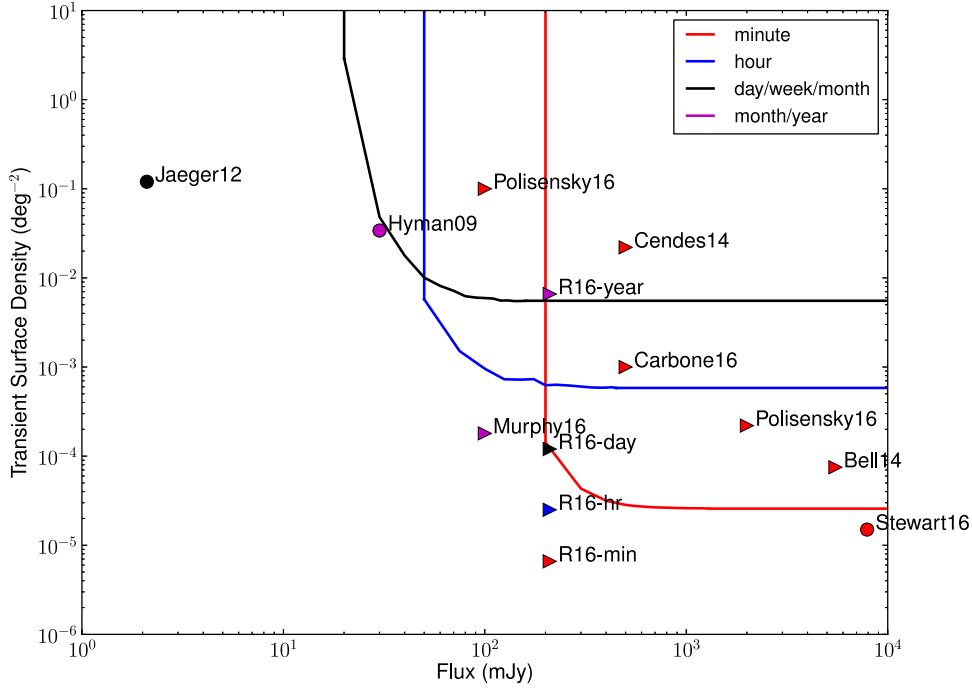
Figure 13. Upper limits on the transient surface density at 95% confidence level from our analysis (solid lines) and other published results (symbols) between 150 and 340 MHz. Note that each symbol should be interpreted as the vertex of an L-shaped upper-limit line, which defines an exclusion region to the upper right. R16 denotes Rowlinson et al. (2016). Jaeger12 (Jaeger et al. 2012), Hyman09 (Hyman et al. 2009), and Stewart16 (Stewart et al. 2016) reported detections; the rest reported upper limits (Bell et al. 2014; Cendes et al. 2014; Carbone et al. 2016; Polisensky et al. 2016; Rowlinson et al. 2016; Murphy et al. 2017). We only plotted two points for Polisensky16: the limit at the lowest flux density, and the best limit reported for 10-minute transients. Timescale is color-coded. Note that the detection reported by Stewart et al. (2016) is at 60 MHz.

and using an average embedded element pattern. This decreased the amount of instrumental Stokes leakage that was more prominent at frequencies $\gtrsim$180 MHz. However, the model beam assumed that all tiles were identical and unchanging over time, whereas in reality the MWA site is not perfectly flat and different tiles have different malfunctioning dipoles or beamforming errors (Neben et al. 2016). In fact, we found that the source light curves extracted from the images corrected by the model beam at 182 MHz showed systematic trends of ~4% flux change per hour that depended on the right ascension of the source. As a result, we decided to measure and model the primary beam empirically to remove these errors, which more severely affected the higher-frequency observations ($\gtrsim$180 MHz), as the Hertzian dipole approximation is less accurate than it is at $\lesssim$150 MHz, where the MWA is designed to operate best.

If we assume that we know the true flux densities of the sources, we can determine the empirical primary beam according to the following relationship:

$$b_{\mathrm{emp,pol}}(\alpha, \delta) = \frac{S_{\mathrm{meas,pol}}(\alpha, \delta)}{S_{\mathrm{ref,pol}}(\alpha, \delta)}, \qquad (22)$$

where $b_{\mathrm{emp,pol}}$ is the empirically measured primary beam for a particular instrumental polarization (XX or YY), $S_{\mathrm{meas,pol}}$ is the measured XX or YY flux density of a source with equatorial coordinates (α, δ), and $S_{\mathrm{ref,pol}}$ is the reference catalog ("true") flux density of the same source. Similar analyses were done for the Very Large Array (Thyagarajan et al. 2011).

To measure the empirical beam, we used a fixed subset of sources instead of the entire ensemble detected in individual XX and YY snapshots. We did this for two reasons: (1) we

wanted to ensure that the sources we used to measure the empirical beam have reliable flux density measurements, and (2) we wanted to avoid overfitting when we fitted a smooth function to the measured data points.

For self-consistency, we used the MWA Commissioning Survey Catalog (Hurley-Walker et al. 2014) as the reference catalog, which we also used for calibration. This avoided issues that could arise if we used source catalogs from other instruments, which might have different angular resolutions, frequency bands, sky coverage, and so on. We assumed the sources to be unpolarized and used the same catalog flux densities for both X and Y polarizations.

When we selected the subset of sources, we filtered out the sources close to the null of the primary beam ($>13°$ from the phase center, corresponding to $\lesssim$0.3 of the primary beam gain). We also filtered out the sources that appeared to have unreliable flux density measurements, which we determined by comparing the catalog flux densities >1 Jy to the mean flux densities that we measured: if the mean flux density that we measured was 3σ away from the catalog flux density, we removed it from the final catalog that we used to measure the empirical beam. As we could not measure a mean flux density without applying the primary beam correction to the images, we did this step iteratively by first using the model beam and then refining it with the empirical beam once. After this process, the reference catalog contained 245 sources. The agreement between the catalog flux densities and the measured flux densities ensured that the fitting procedure used reliable data.

For the fitting procedure, we assumed that the primary beam was smooth and fitted a smoothing spline to the empirical beam measurements on a snapshot-by-snapshot basis. We used
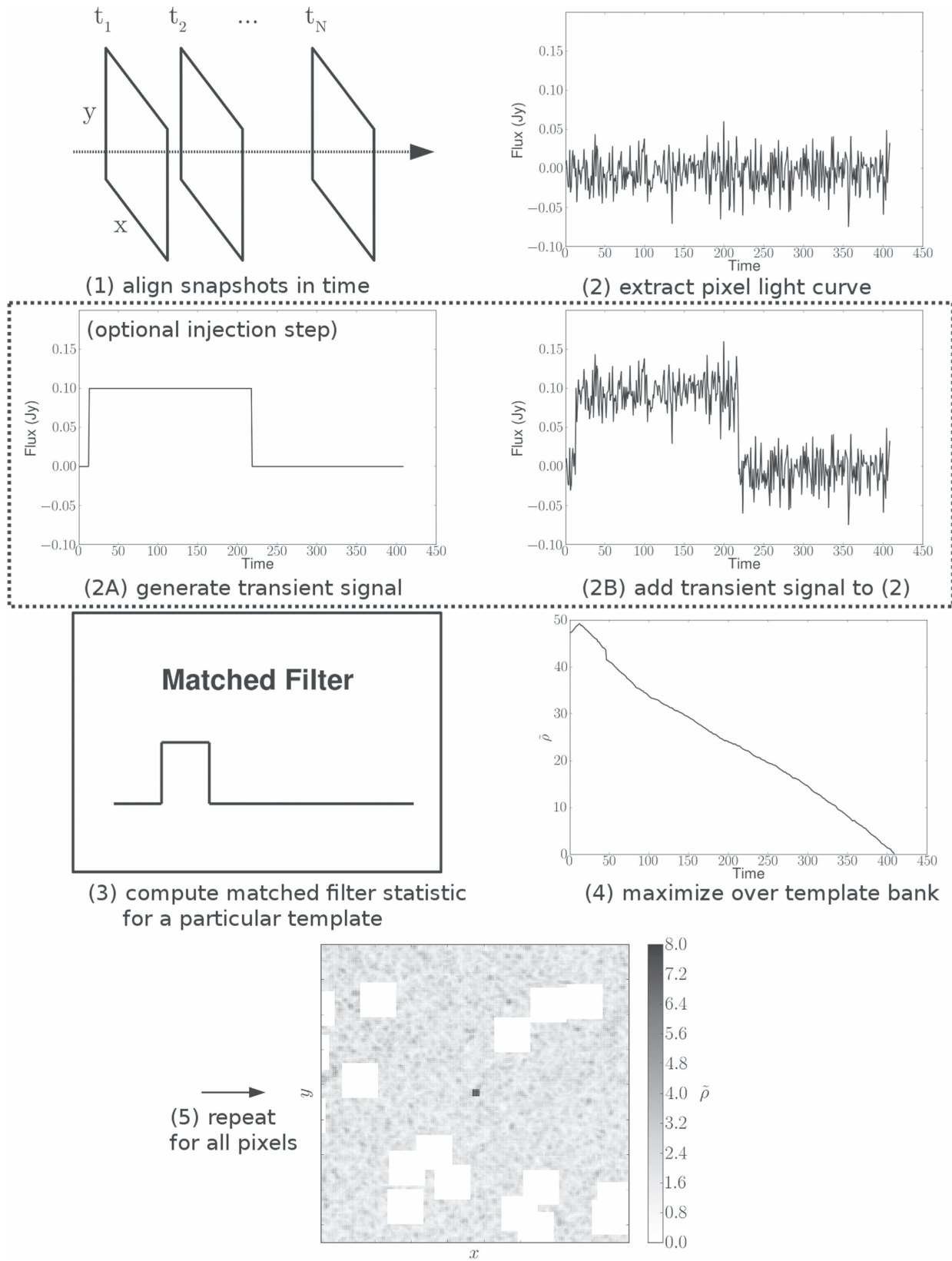
Figure 14. Illustration of the *Simetra* pipeline. The color scale of $\tilde{\rho}$ has been cut off at 8. The white boxes are the locations of masked bright sources. See text for details.

biquartic splines $(k_x, k_y) = (4, 4)$ as they provided a better fit (lower residuals) than the default bicubic splines, whereas biquintic splines did not improve the fit significantly.

Fitting a spline function $s(x_i, y_i)$ to a set of data z_i that have measurement errors involves a trade between the smoothness of the spline and the goodness of the fit. The algorithm used by

the `SmoothBivariateSpline` routine in `scipy`,[28] which we used for this analysis, determines the smoothest spline given the constraint that the goodness of fit is less than the smoothing factor S:

$$\sum_{i=1}^{m} w_i [z_i - s(x_i, y_i)]^2 \leqslant S, \qquad (23)$$

where z_i is the empirical beam measurement for each source, (x_i, y_i) is the equatorial coordinate of the source, and w_i is the weight of each measurement (Dierckx 1981). We chose S to be the number of sources that entered the fit (245), which was the default value, as it gave a satisfactory fit. We set w_i to be the catalog flux density for each source, because the flux density errors derived from AEGEAN appeared to be too large to provide reliable inverse variance weights and did, in fact, make the fit worse. The unreliable errors reported by AEGEAN are a known issue and have since been fixed.

We also verified that the fitting function was reliable. Instead of deriving an empirical primary beam based on flux density measurements in clean XX and YY images before primary beam correction, we derived an empirical "correction factor" for the Stokes I images corrected by the model beam, using the same fitting procedure. Both procedures gave the same results, thus demonstrating that the fitting function was reliable. The empirical fitted beam removed most of the systematic errors in the light curves. It reduced the light-curve slopes from $\lesssim 4\%$ to $\lesssim 2\%$ flux density change per hour and recovers (by design) a more accurate source flux density.

References

Allen, B., Anderson, W. G., Brady, P. R., Brown, D. A., & Creighton, J. D. E. 2012, PhRvD, 85, 122006
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Bell, M. E., Murphy, T., Kaplan, D. L., et al. 2014, MNRAS, 438, 352
Biswas, R., Brady, P. R., Burguet-Castell, J., et al. 2012, PhRvD, 85, 122008
Biswas, R., Brady, P. R., Creighton, J. D. E., & Fairhurst, S. 2009, CQGra, 26, 175009
Bower, G. C., Saul, D., Bloom, J. S., et al. 2007, ApJ, 666, 346
Brady, P. R., Creighton, J. D. E., & Wiseman, A. G. 2004, CQGra, 21, S1775
Carbone, D., van der Horst, A. J., Wijers, R. A. M. J., et al. 2016, MNRAS, 459, 3161
Cendes, Y., Wijers, R. A. M. J., Swinbank, J. D., et al. 2014, arXiv:1412.3986
Condon, J. J. 1974, ApJ, 188, 279
Dierckx, P. 1981, IJNA, 1, 267
Fiedler, R. L., Dennison, B., Johnston, K. J., & Hewish, A. 1987, Natur, 326, 675
Frail, D. A., Kulkarni, S. R., Nicastro, L., Feroci, M., & Taylor, G. B. 1997, Natur, 389, 261
Frail, D. A., Kulkarni, S. R., Ofek, E. O., Bower, G. C., & Nakar, E. 2012, ApJ, 747, 70
Franzen, T. M. O., Jackson, C. A., Offringa, A. R., et al. 2016, MNRAS, 459, 3314
Gaensler, B. M., Kouveliotou, C., Gelfand, J. D., et al. 2005, Natur, 434, 1104
Greisen, E. W. 2012, AIPS Memo 117, http://www.aips.nrao.edu/aipsmemo.html
Hallinan, G., Bourke, S., Lane, C., et al. 2007, ApJL, 663, L25
Hancock, P. J., Murphy, T., Gaensler, B. M., Hopkins, A., & Curran, J. R. 2012, MNRAS, 422, 1812
Helmstrom, C. W. 1968, Statistical Theory of Signal Detection (2nd ed.; London: Pergamon Press)
Hurley-Walker, N., Morgan, J., Wayth, R. B., et al. 2014, PASA, 31, 45
Hyman, S. D., Lazio, T. J. W., Kassim, N. E., et al. 2005, Natur, 434, 50
Hyman, S. D., Wijnands, R., Lazio, T. J. W., et al. 2009, ApJ, 696, 280
Jaeger, T. R., Hyman, S. D., Kassim, N. E., & Lazio, T. J. W. 2012, AJ, 143, 96
Kass, R. E., & Raftery, A. E. 1995, J. Am. Stat. Assoc., 90, 773
Lazio, W. T. J., Farrell, W. M., Dietrick, J., et al. 2004, ApJ, 612, 511
Levinson, A., Ofek, E. O., Waxman, E., & Gal-Yam, A. 2002, ApJ, 576, 923
Lo, K. K. Y. 2014, PhD thesis, Univ. Sydney, School of Physics
Loi, S. T., Murphy, T., Cairns, I. H., et al. 2015, GeoRL, 42, 3707
Lonsdale, C. J., Cappallo, R. J., Morales, M. F., et al. 2009, IEEEP, 97, 1497
McMullin, J. P., Waters, B., Schiebel, D., Young, W., & Golap, K. 2007, in ASP Conf. Ser. 376, Astronomical Data Analysis Software and Systems XVI, ed. R. A. Shaw, F. Hill, & D. J. Bell (San Francisco, CA: ASP), 127
Metzger, B. D., Williams, P. K. G., & Berger, E. 2015, ApJ, 806, 224
Murphy, T., Kaplan, D. L., Croft, S., et al. 2017, MNRAS, 466, 1944
Nakar, E., & Piran, T. 2011, Natur, 478, 82
Neben, A. R., Bradley, R. F., Hewitt, J. N., et al. 2015, RaSc, 50, 614
Neben, A. R., Hewitt, J. N., Bradley, R. F., et al. 2016, ApJ, 820, 44
Neyman, J., & Pearson, E. S. 1933, RSPTA, 231, 289
Offringa, A. R., de Bruyn, A. G., Biehl, M., et al. 2010, MNRAS, 405, 155
Offringa, A. R., McKinley, B., Hurley-Walker, N., et al. 2014, MNRAS, 444, 606
Offringa, A. R., van de Gronde, J. J., & Roerdink, J. B. T. M. 2012, A&A, 539, A95
Offringa, A. R., Wayth, R. B., Hurley-Walker, N., et al. 2015, PASA, 32, 8
Polisensky, E., Lane, W. M., Hyman, S. D., et al. 2016, ApJ, 832, 60
Rowlinson, A., Bell, M. E., Murphy, T., et al. 2016, MNRAS, 458, 3506
Sivia, D. S., & Skilling, J. 2006, Data Analysis: A Bayesian Tutorial (2nd ed.; Oxford: Oxford Univ. Press)
Soderberg, A. M., Chakraborti, S., Pignata, G., et al. 2010, Natur, 463, 513
Stewart, A. J., Fender, R. P., Broderick, J. W., et al. 2016, MNRAS, 456, 2321
Sutinjo, A., O'Sullivan, J., Lenc, E., et al. 2015, RaSc, 50, 52
Thornton, D., Stappers, B., Bailes, M., et al. 2013, Sci, 341, 53
Thyagarajan, N., Helfand, D. J., White, R. L., & Becker, R. H. 2011, ApJ, 742, 49
Thyagarajan, N., Udaya Shankar, N., Subrahmanyan, R., et al. 2013, ApJ, 776, 6
Tingay, S. J., Goeke, R., Bowman, J. D., et al. 2013, PASA, 30, 7
Trott, C. M., Wayth, R. B., Macquart, J.-P. R., & Tingay, S. J. 2011, ApJ, 731, 81
van Haarlem, M. P., Wise, M. W., Gunst, A. W., et al. 2013, A&A, 556, A2

---

[28] http://www.scipy.org/