The High Time Resolution Universe survey – XIV. Discovery of 23 pulsars through GPU-accelerated reprocessing

V. Morello [1,2]★ E. D. Barr [3,2] S. Cooper,[1] M. Bailes,[2,4] S. Bates,[1] N. D. R. Bhat,[5]
M. Burgay [6] S. Burke-Spolaor,[7,8] A. D. Cameron [3] D. J. Champion,[3] R. P. Eatough,[3]
C. M. L. Flynn,[2,4] A. Jameson,[2,4] S. Johnston [9] M. J. Keith,[1] E. F. Keane [10]
M. Kramer,[1,3] L. Levin,[1] C. Ng [11] E. Petroff [12] A. Possenti,[6,13] B. W. Stappers,[1]
W. van Straten [14] and C. Tiburzi [15]

[1] *Jodrell Bank Centre for Astrophysics, School of Physics and Astronomy, The University of Manchester, M13 9PL, UK*
[2] *Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Mail H39, PO Box 218, VIC 3122, Australia*
[3] *Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, D-53121 Bonn, Germany*
[4] *ARC Center of Excellence for Gravitational Wave Discovery (OzGrav), Swinburne University of Technology, Mail H11, PO Box 218, VIC 3122, Australia*
[5] *International Centre for Radio Astronomy Research, Curtin University, Bentley, WA 6102, Australia*
[6] *INAF - Osservatorio Astronomico di Cagliari, Via della Scienza 5, I-09047 Selargius (CA), Italy*
[7] *Center for Gravitational Waves and Cosmology, West Virginia University, Chestnut Ridge Research Building, Morgantown, WV 26505, USA*
[8] *Department of Physics and Astronomy, West Virginia University, P.O. Box 6315, Morgantown, WV 26506, USA*
[9] *CSIRO Astronomy & Space Science, Australia Telescope National Facility, PO Box 76, Epping, NSW 1710, Australia*
[10] *SKA Organisation, Jodrell Bank Observatory, SK11 9DL, UK*
[11] *Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, BC V6T 1Z1, Canada*
[12] *Anton Pannekoek Institute for Astronomy, University of Amsterdam, P.O. Box 94249, NL-1090 GE Amsterdam, the Netherlands*
[13] *Università di Cagliari, Dip. di Fisica, S.P. Monserrato-Sestu Km 0,700, I-09042 Monserrato (CA) - Italy*
[14] *Institute for Radio Astronomy & Space Research, Auckland University of Technology, Private Bag 92006, Auckland 1142, New Zealand*
[15] *Fakultät fur Physik, Universität Bielefeld, Postfach 100131, D-33501 Bielefeld, Germany*

ABSTRACT
We have performed a new search for radio pulsars in archival data of the intermediate and high Galactic latitude parts of the Southern High Time Resolution Universe pulsar survey. This is the first time the entire dataset has been searched for binary pulsars, an achievement enabled by GPU-accelerated dedispersion and periodicity search codes nearly 50 times faster than the previously used pipeline. Candidate selection was handled entirely by a Machine Learning algorithm, allowing for the assessment of 17.6 million candidates in a few person-days. We have also introduced an outlier detection algorithm for efficient radio-frequency interference (RFI) mitigation on folded data, a new approach that enabled the discovery of pulsars previously masked by RFI. We discuss implications for future searches, particularly the importance of expanding work on RFI mitigation to improve survey completeness. In total, we discovered 23 previously unknown sources, including 6 millisecond pulsars and at least 4 pulsars in binary systems. We also found an elusive but credible redback candidate that we have yet to confirm.

Key words: methods: data analysis – pulsars: general.

1 INTRODUCTION

The Southern High Time Resolution Universe (HTRU) project (see Keith et al. 2010, for a complete description) is an extensive survey of the Southern Sky for pulsars and fast transients performed with the Parkes multibeam receiver (Staveley-Smith et al. 1996) between 2008 and 2013. The survey area was partitioned into three sections: a low, intermediate, and high galactic latitude component referred to as lowlat, medlat, and hilat, with on-sky integration times of 70, 9, and 4.5 min, respectively. Lowlat was designed to discover relativistic binaries in the Galactic plane, medlat to find new bright millisecond pulsars worthy of being included in pulsar timing array experiments, and hilat to improve the pulsar census at high galactic

latitudes and explore an undersearched portion of the sky for radio transient events such as fast radio bursts (FRBs; Lorimer et al. 2007; Thornton et al. 2013).

Prior to this work, the medlat portion of the survey had already been entirely searched for periodic pulsar signals once (Bates et al. 2012; Levin 2012), while 32 per cent of hilat had been processed (Thornton 2013). Due to limited computing resources, none of these searches were sensitive to pulsars with high orbital acceleration; yet with more than 80 per cent of the known millisecond pulsar population found in binary systems, running an extensive acceleration search (Johnston & Kulkarni 1991; Ransom, Eikenberry & Middleditch 2002) would ensure the fulfilment of the initial science goals of the survey.

This, however, comes with two major obstacles: high data rates and the ubiquitous presence of radio-frequency interference (RFI). The HTRU survey is a prime target to develop and test new software with the goal of demonstrating quasi-real time processing with minimal human intervention in candidate classification, as will be required on upcoming telescopes such as CHIME (Ng 2017) or the SKA (Keane et al. 2015). Running new analyses of archival data with refined tools can also lead to a significant number of new discoveries even in a previously searched portion of the parameter space, as demonstrated on the Parkes Multibeam Pulsar Survey (PMPS; Manchester et al. 2001) by several successful reprocessings (e.g. Keith et al. 2009; Keane et al. 2010; Eatough et al. 2013; Knispel et al. 2013).

Here, we present a new search of the intermediate and high latitude parts of the HTRU survey. In Section 2, we delve into the details of our processing pipeline. We also developed a new efficient RFI mitigation algorithm based on outlier detection, of which we release a fast python implementation with this paper. In Section 3, we briefly present the resulting 23 new pulsar discoveries; full timing solutions will be provided in upcoming publications. We also report a credible accelerating MSP candidate that we failed to confirm despite extensive radio re-observations, and invite the wider community to pursue the effort. In Section 4, we pursue a comparison of our new search pipeline with the one previously used on the same data, and finally discuss in Section 5 the potential implications for future pulsar searches.

2 METHODS AND DATA ANALYSIS

Below we describe the pulsar parameter space searched and the three major operations of the search pipeline: Fourier search, candidate folding, candidate classification along with the RFI mitigation methods employed. The detailed sequence of processing steps is summarized in Fig. 1.

2.1 Search parameters

The first processing passes of the med- and hilat portions of the HTRU survey used SIGPROC-based processing pipelines (Keith et al. 2010). While these pipelines were clearly effective, discovering 122 new pulsars, resource-constraints limited these analyses to pulsars without significant acceleration. The development of the GPU-accelerated PEASOUP[1] search code removed these constraints allowing for both regions to be searched out to moderate accelerations. The PEASOUP pipeline does not offer improved sensitivity for slow or unaccelerated pulsars and as such we selected a parameter

space to search that focused on the discovery of binary MSPs. Table 1 shows the parameters for the searches conducted on medlat and hilat. These parameters are explained throughout this section. For medlat we selected a maximum DM to search of 400 pc cm^{-3}. This corresponds to ~1 ms of dispersive smearing within the 391 kHz frequency channels of the BPSR backend (described in section 3.2 of Keith et al. 2010). For hilat, we limited ourselves to twice the maximum line-of-sight DM according to the NE2001 Galactic free electron density model (Cordes & Lazio 2002) with a cap of 1000 pc cm^{-3}.

In both cases trial DMs to search were generated using the method explained at length in section 2.3 of Levin (2012), which we briefly outline here. A pulse of intrinsic width W_{int} is broadened to a total width W_{tot} by three numerical effects, namely sampling, smearing due to intra-channel dispersion, and de-dispersion at an incorrect DM. Each have their associated time-scales, respectively τ_{samp}, τ_{smear}, and $\tau_{\Delta DM}$. These effects add in quadrature so that we have

$$W_{tot}^2 = W_{int}^2 + \tau_{samp}^2 + \tau_{smear}^2 + \tau_{\Delta DM}^2. \tag{1}$$

The first three terms in the sum are predetermined by the observing set-up and by the DM of the source in the case of τ_{smear}. Only $\tau_{\Delta DM}$ can be reduced by adopting a narrower DM trial spacing. We can rewrite the above as

$$W_{tot}^2 = W_{eff}^2 + \tau_{\Delta DM}^2, \tag{2}$$

where W_{eff} is the effective pulse width once all unavoidable broadening effects have been taken into account. The idea is to choose the spacing from one DM trial to the next so that $\tau_{\Delta DM}$ never exceeds some small fraction ε of W_{eff}, which can be written

$$W_{tot} \leq (1 + \epsilon)W_{eff}. \tag{3}$$

The constraint above is usually expressed in terms of a so-called *DM tolerance parameter*, defined as $DM_{tol} = 1 + \epsilon$. $DM_{tol} = 1.1$ in all our searches. W_{int} also constitutes a free parameter and represents the minimum intrinsic pulse width one expects in the data.

Due to an oversight,[2] approximately 3 per cent of the pointings were searched out to DM values less than the NE2001 model predictions for their lines of sight. These pointings are predominantly located close to the Galactic plane where there exists deeper coverage from lowlat (Ng et al. 2015) but also the PMPS (Manchester et al. 2001) and SUPERB surveys (Keane et al. 2018). For both medlat and hilat, we selected an acceleration range of $|a| < 50$ m s^{-2}. Calculating the maximum line-of-sight acceleration (l.o.s.a.) for all the binary pulsars in version 1.58 of the ATNF Pulsar Catalogue (Manchester et al. 2005), we find that this range contains ~97 per cent of the known pulsar population with most of the remaining pulsars belonging to more massive double neutron star systems less likely to be found far from the Galactic plane. As with the DM trials, the acceleration step size was determined based on an acceleration tolerance parameter a_{tol} of 1.1.

2.2 Search code overview

PEASOUP implements a time-domain resampling acceleration search (Johnston & Kulkarni 1991) on NVIDIA GPUs. The pipeline includes dedispersion through the DEDISP library (Barsdell et al. 2012), dereddening (low-frequency noise removal) in the Fourier domain,

---

[1] https://github.com/ewanbarr/peasoup

[2] Beam coordinate offsets from the centre beam were occasionally applied in degrees instead of radians, before passing the result as input to the NE2001 model.
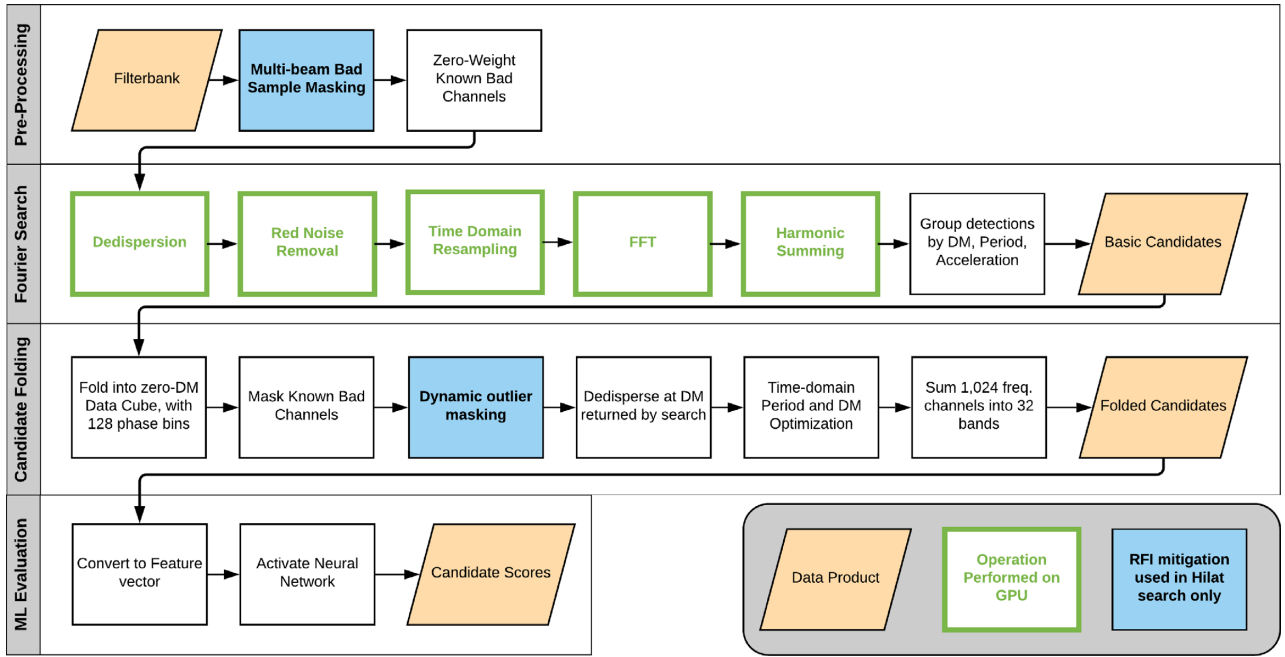
Figure 1. Detailed overview of the full processing pipeline. Results from the processing of medlat encouraged the use of extra RFI mitigation layers on the subsequent hilat search, including a new outlier masking algorithm described in Section 2.4.

Table 1. Search parameters for both medlat and hilat searches. W_{int} and DM_{tol} control the spacing of consecutive DM trials (details in Section 2.1). Likewise, W_{int} and a_{tol} control that of the acceleration trials. f_5 and f_{25} are parameters of the dereddening of the Fourier spectrum (see Section 2.2).

Parameter	Medlat	Hilat
DM_{max} (pc cm^{-3})	400	See the text
$N_{DMtrials}$	1347	See the text
DM_{tol}	1.1	1.1
W_{int} (μs)	64	40
$|a_{max}|$ (m s^{-2})	50	50
δa (m s^{-2})	1.46	3.66
a_{tol}	1.1	1.1
$N_{acctrials}$	71	30
$N_{harmonics}$	16	16
f_5 (Hz)	0.05	0.05
f_{25} (Hz)	0.5	0.5
f_{min} (Hz)	0.1	0.1
f_{max} (Hz)	1 100	1 100
N_{beams}	95 940	358 644
Fraction processed	100%	81%

resampling, FFT, harmonic summing up to the 16th harmonic, peak detection and optional time series folding.

To save computing time, PEASOUP performs dereddening only once per DM trial, producing a time series with reduced low-frequency noise that is subsequently searched at all trial accelerations. The dereddening method consists of taking the FFT of the dedispersed time series and scaling the complex-valued Fourier coefficients by an appropriate real-valued, frequency-dependent factor (following section 3.1 of Ransom et al. 2002), before inverse transforming the corrected Fourier amplitudes back to the time domain. The scaling factor is calculated so that all bins of the corrected power spectrum follow a chi-squared distribution with 2 degrees of

freedom, i.e.

$$B_i = A_i \sqrt{\frac{2 \ln 2}{\underset{i-m \leq k \leq i+m}{\text{median}} \{|A_k|^2\}}}, \qquad (4)$$

where A_i are the complex Fourier amplitudes of the original dedispersed time series, B_i represent the dereddened amplitudes, the denominator term is a robust estimate of the Fourier power around bin number i and m is half the size of a running median window. The factor $2 \ln 2$ is the expected median value of a chi-squared distribution with 2 degrees of freedom. An underlying assumption here is that the expected value of the power spectrum is constant across the median window; if such is the case, then the B_i can be expected to have normally distributed real and imaginary components each with zero mean and unit variance, like the Fourier transform of Gaussian white noise.

While equation (4) corresponds to scaling by a running median of width $w = 2m + 1$, in practice PEASOUP breaks the spectrum into blocks of $w = 5$ bins, takes the median of each block and then interpolates the output to all frequencies to obtain an *approximate* running median, which is much better suited to parallel architectures. The process can be iterated multiple times (median of medians) to obtain a local Fourier power estimate over windows of width 5^n. Since large median windows at low frequency tend to underestimate the red noise power, we define two parameters f_5 and f_{25}, that correspond to the frequencies at which we switch from a median smoothing window of 5–25 spectral bins and 25–125, respectively. Note that since we deredden only once per DM trial, the dereddening applied at zero acceleration must also be valid for higher accelerations. Thus, the selection of the f_5 and f_{25} parameters is important as the size of the smoothing window should be at least twice the expected number of Fourier bins drifted by an accelerated pulsar over the course of an observation. As our smoothing window size is capped at 125 bins, we start to see reductions in sensitivity for signals above 1294 Hz at the maximum acceleration of 50 m

s^{-2}. The reduction in signal-to-noise ratio (S/N) due to this effect is a function of the true S/N of the given signal. We find empirically that signals with true S/N $\lesssim$ 12 are unaffected, and stronger ones are never reduced to below S/N $\approx$12.

Two basic levels of RFI rejection were used in both searches. First, 158 frequency channels (15.4 per cent of the band) known to be contaminated by narrow-band RFI were ignored during dedispersion. Secondly, a list of Fourier spectral bins known to contain high-occupancy periodic RFI were zero-weighted such that they would not affect the acceleration search. These bins amounted to no more than 0.1 per cent of the Fourier spectrum. As with dereddening, this process was performed once per DM trial. Finally, for the hilat search one further level of RFI rejection was included. Here, the multibeam RFI detection system of Kocz et al. (2012) was used to identify time samples affected by near-field impulsive RFI, which were replaced by representative noise samples drawn from the distribution of surrounding valid data.

2.3 Processing

The data were processed on the Green II (G2) supercomputer at Swinburne University of Technology.[3] G2 is composed of a 3.4-petabyte lustre file system and two computing clusters, gSTAR and SwinSTAR. They host, respectively, 102 NVIDIA Tesla C2070s (two per node) and 64 NVIDIA Tesla K10 accelerators (one per node, with two GPUs). Processing of the medlat portion of the survey used both gSTAR and SwinSTAR, while we restricted processing of the hilat portion of the survey to SwinSTAR due to reduced availability of the gSTAR cluster. PEASOUP is multiGPU capable and therefore nodes were assigned processing jobs that used both GPUs.

The use of GPU accelerators allowed us to greatly reduce the overall processing time required. The initial processing of medlat, without acceleration trials, consumed ~700 000 CPU-h spread over the 3-yr period during which the survey was being observed. Had all the data been available when the initial processing started, it would have taken a minimum of 8 months on the hardware available at the time (Levin 2012). In contrast, the accelerated reprocessing of medlat presented here took only 3 months (~15 000 GPU-h).

Performance logs taken from the hilat reprocessing show median execution times per DM trial of 43 ms for dedispersion and 128 ms for the acceleration search. Including overheads, we find a median execution time of 182 ms per DM trial for the full search (excluding candidate folding, see Section 2.4). The performance demonstrated here was a driver for the development of the real-time acceleration search deployed as part of the SUPERB survey (Keane et al. 2018). Detailed performance comparisons between the old and new pipelines are presented in Section 4.1.

As can be seen in Table 1, the total number of beams processed (N_{beams}) does not match the numbers presented in table 1 of Keith et al. (2010). In the case of medlat, we exceed the total number of beams in the survey due to the inclusion of re-observations of RFI-affected pointings and confirmation observations for newly discovered pulsars. In the case of hilat we were limited to the data available on the G2 file system at the time of processing. This constituted 81 per cent of the full survey.

2.4 Improved candidate folding

For each candidate identified by the FFT search, the original data need to be phase-coherently folded at the candidate period, in order to obtain more accurate parameters and additional diagnostic information. It is standard practice to partition the input data in equal-sized time segments dubbed *sub-integrations*, and then for each sub-integration, fold every channel individually. The folded output is therefore a 3D array with frequency, time, and phase axes. We refer to it as a *data cube* below. The optimal candidate parameters can then be found via a fine grid search over a range of trial dispersion measures and periods centred around the best candidate solution reported by the FFT search.

2.4.1 The need for additional RFI mitigation

For our search of medlat, we used the DSPSR (van Straten & Bailes 2011) and PSRCHIVE software packages (Hotan, van Straten & Manchester 2004) to perform, respectively, candidate folding and time-domain candidate parameter optimization. Our processing strategy was to put the entire burden of RFI rejection on the final candidate classification algorithm (described in Section 2.5). However, we subsequently found that this approach was sub-optimal; for observations affected by strong interference, the optimization process may not converge to the true pulsar parameters. It is only guaranteed to report the parameters that maximize a statistical test for the presence of a pulse in the corresponding integrated profile. Occasionally, RFI distorts this test to an extent such that the final candidate plot is impossible to identify as a pulsar even by a trained eye. The accuracy of the candidate classification stage, regardless of it being human or algorithmic, is dependent on the quality of the candidate plots produced.

For the hilat search, we therefore decided to develop our own software package to perform candidate folding and optimization called CUBR (see chapter 4 of Morello 2016),[4] which was also integrated into the SUPERB pulsar search pipeline (Keane et al. 2018). Its main feature is that it removes interference in the folded data cube before the optimization process. For this, it uses a simple outlier rejection algorithm to identify and suppress data that significantly deviates from the dominant distribution. In what follows, A is the data cube and i, j, and k indices along the frequency, time, and phase axes, respectively. A_{ijk} represents a single data point while A_{ij} denotes the pulse profile for channel index i and sub-integration j. Before we apply any RFI mitigation, the dispersion measure of the folded data is set to zero, and we subtract from every profile A_{ij} its own median value. It should be noted that RFI comes in many forms; however, by contrast with pulsar emission, it rarely exhibits any dispersion and often occurs in either a narrow band of frequencies or in a short interval of time. These properties guided the design of the method below, which we demonstrate on a folded observation of a known pulsar affected by interference (Fig. 2, left-hand panel) to serve as evidence of its effectiveness.

2.4.2 Step 1: removal of bad frequency and time intervals

Here, we flag abnormal profiles A_{ij} based on three numerical profile features:
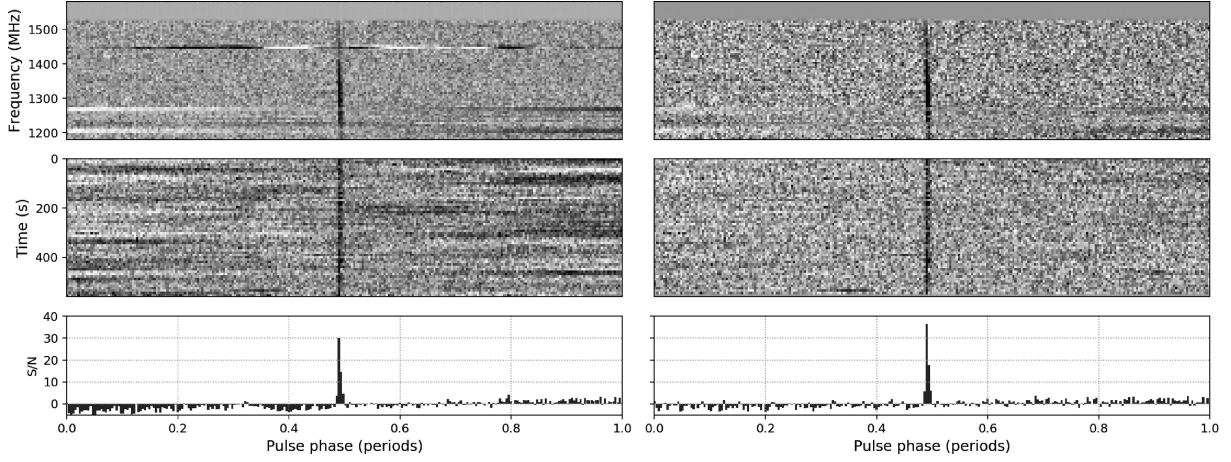
(i) Standard deviation.

Figure 2. A folded 9-min observation of PSR J0736–6304 ($P = 4.86$ s, DM $= 19.4$ pc cm^{-3}) before and after the RFI mitigation algorithms described in Section 2.4 have been applied (left and right columns, respectively). Top: frequency-phase plots, where we have grouped and summed the original 1024 frequency channels into 64 sub-bands for readability. Centre: time-phase plots. Bottom: integrated pulse profiles over the entire observation. In both cases, integrated profiles have been normalized by a factor $s\sqrt{N}$, where N is the number of unmasked profiles in the cube, and s the standard deviation of unmasked data. The top 150 frequency channels are permanently occupied by telecommunication signals and masked by an RF filter (Keith et al. 2010). The peak S/N of the pulse increases noticeably from 31 to 37. Improving the general quality of candidate plots reduces the probability of genuine pulsars being improperly rejected in the classification stage of a blind search.

(ii) Difference between maximum and minimum value (peak-to-peak difference).

(iii) Absolute value of the second bin of the profile's Fourier transform. This helps identify strong low-frequency noise and signals with a sine-wave shape.

Every profile is associated to a point in this 3D feature space (Fig. 3), which reveals a number of outliers. To identify them, we apply Tukey's rule (Tukey 1977; Chandola, Banerjee & Kumar 2009) to each feature: a value is considered anomalous if it falls out of the interval $[Q_1 - qR, Q_3 + qR]$, where Q_1 is the 25th percentile of the distribution, Q_3 the 75th percentile, and $R = Q_3 - Q_1$ is called the interquartile range. q is a free parameter that can be accurately mapped to a false rejection probability if most of the data follow a normal distribution. Tukey's original recommendation is $q = 1.5$; we set $q = 2.0$ in our processing following tests on a sample of RFI contaminated pulsar observations. Q_1, Q_3, and R are not distorted by the presence of outliers, which make them a good choice as opposed to the mean or the standard deviation. A profile that stands out as anomalous with respect to any of the three features above is flagged as bad (Fig. 4) and zero-weighted before further processing.

2.4.3 Step 2: removal of zero-DM interference pulses

Next, we identify short duration bright interference bursts that frequently appear in Parkes data. These are not dispersed, so they are more easily identified in the so-called sub-integrations plot, i.e. the sum of the data cube along the frequency axis that we denote I_{jk}:

$$I_{jk} = \sum_{i=1}^{n_c} A_{ijk},\qquad(5)$$

where n_c denotes the number of frequency channels, 1024 in the case of HTRU data. A short interference burst manifests itself in a single sub-integration as a small interval of phase bins with an abnormally high value. We may therefore identify them by separately applying Tukey's rule to every column of I, where a given column corresponds

to a constant phase. For a bad data point (j_0, k_0) thus flagged in I, we replace the associated data $A_{ij_0k_0}$ across the frequency dimension of the cube according to

$$A_{ij_0k_0} = \frac{1}{n_c}\operatorname*{median}_{1 \le j \le n_t}\{I_{jk_0}\},\qquad(6)$$

where n_t is the number of sub-integrations, which was set to 32 in our processing. The replacement is performed for every i from 1 to n_c. The overall effect of the application of the cleaning algorithms described here can be seen in the right-hand panel of Fig. 2. Once RFI mitigation has been applied, we proceed with a standard optimization stage and produce a candidate plot to be evaluated in the final classification stage.

2.4.4 Possible limitations

The upside of our proposed method lies in that it requires no explicit manual parametrization of RFI, having only a single free parameter q defined above. Instead, interference is defined as a small sub-set of the data cube that significantly differs from the rest, a flexible approach that should in principle work in any RFI environment. We do however make a few implicit assumptions about what the properties of a pulsar signal should be: broad-band, dispersed, with no pulses vastly brighter than the others. Should any of these assumptions not hold, there is a risk of erasing a significant part of a pulsar's signal, but that situation should occur infrequently. For the first step to remove a pulse of a genuine pulsar source in a given channel and sub-integration, it should at least be distinguishable from white noise; however, its S/N can be expected to be $\sqrt{n_c \times n_t} \simeq 181$ times lower than the S/N of the pulse integrated over the whole band and observation length. The second step can potentially remove sub-integration samples containing bright individual pulses of an intermittent or nulling pulsar with low DM. Such sources should be more easily detected in a single pulse search however, which are now routinely performed on most pulsar surveys including HTRU (Burke-Spolaor et al. 2011).
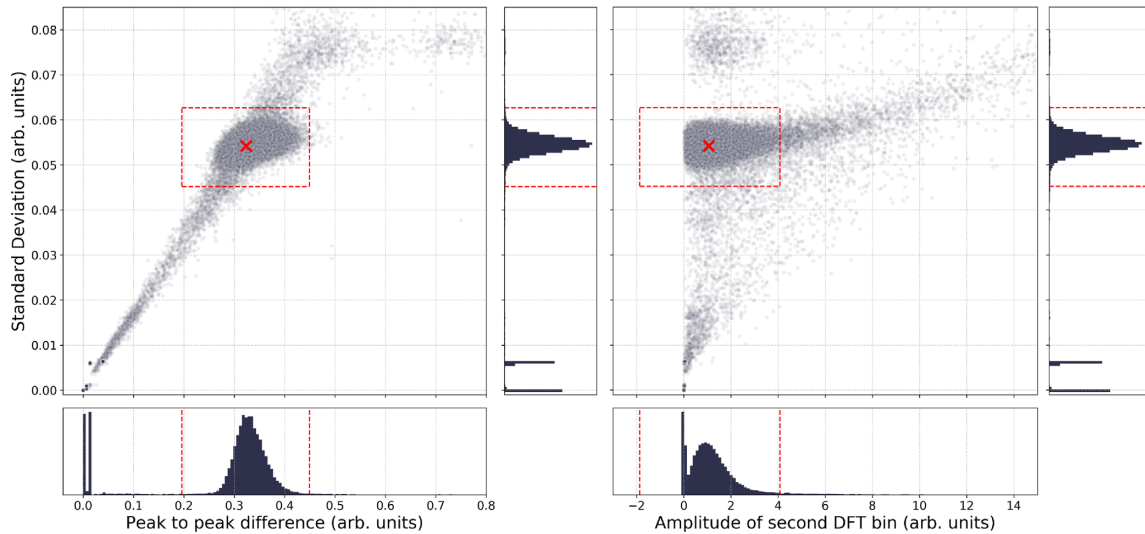
Figure 3. Visualization of the outlier rejection algorithm based on Tukey's rule, here applied to the pulsar observation shown in the left-hand panel of Fig. 2. For each profile in the data cube, we compute three features (see Section 2.4), and project all profiles in the resulting 3D space of which we show two 2D slices in the scatter plots above. For each dimension of the feature space, we determine an acceptable value range based on the median and inter-quartile range of the whole distribution; their histograms are displayed along scatter plot axes. The resulting decision boundary in feature space is a 3D box whose 2D projections are visible as dashed lines in the scatter plots above.
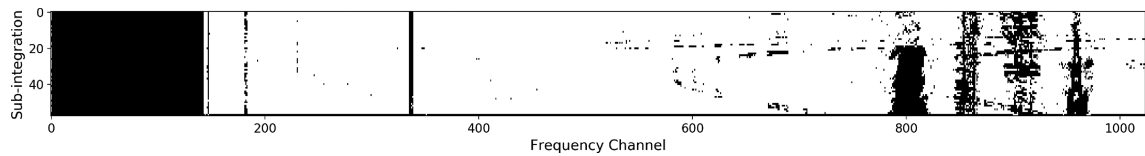

Figure 4. Profile mask determined from the analysis shown in Fig. 3. Each pixel corresponds to a pulse profile, which is part of a given frequency channel (*X*-axis) and sub-integration (*Y*-axis). Outliers appear in black and the corresponding data are zero-weighted before further processing. Bad frequency channels are successfully flagged in this example. Arc-like patterns are visible, revealing narrow-band interference sources whose frequencies change over a time-scale of ≈10 s.

2.4.5 Code release

A fast, standalone PYTHON implementation of the two RFI mitigation algorithms described above is made publicly available.[5] It comes with an interface to the PSRFITS format, via the PYTHON bindings of the PSRCHIVE package. We also note that it can be readily applied to perform RFI excision on pulsar timing data.

2.5 Candidate classification

Most large scale pulsar searches of the past two decades have produced candidate numbers large enough that visual inspection of the entire search output is not a reasonable option (Lyon et al. 2016). Our re-processing was no exception and left us with 17.6 million candidates to sort through. Furthermore, medlat had been entirely searched once before, making the number of expected discoveries lower than on fresh data and the cost effectiveness of extensive visual inspection worse. Developing an accurate Machine Learning (ML) classifier was therefore a requirement to make our search successful.

Our classification algorithm SPINN consists of a simple two-layer artificial neural network (ANN). Its first version (SPINN v1 hereafter) has previously been described at length in Morello et al.

(2014), and was used to evaluate candidates from the medlat portion of the survey. The changes subsequently made to the folding software and candidate format prior to running the hilat search (Section 2.4) were also an opportunity to make adjustments to the classification stage of the pipeline, resulting in an improved SPINN v2. In this section, we will limit ourselves to a description of the design guidelines we followed for SPINN v2 and an evaluation of its classification accuracy and speed. Further details and an in-depth explanation of ANNs can be found in chapter 3 of Morello (2016).

2.5.1 Candidate features

In any ML problem, each data instance must be converted into a set of descriptive numerical properties called features, that are then fed as an input vector to the ML algorithm. The choice of features plays a critical role in the algorithm's success. For the design of features, we followed the general guidelines below:

(i) Keep the feature count below 10. An important constraint encountered in pulsar candidate classification is the low number of available real pulsar training examples, on order of 1000 at best for a given instrument. Intuitively, these can only sparsely sample a 10D feature space, leaving vast volumes where the classification function to be learned is undefined, but where unseen real pulsar examples may nonetheless lie. This is an aspect of the so-called *curse of dimensionality* (e.g. Hastie, Tibshirani & Friedman 2009). Some

---

[5]https://github.com/v-morello/clfd

Table 2. List of candidate features used as an input to SPINN v2, used on the hilat search. The formulas used to compute these features can be found in chapter 3 of Morello (2016).

No.	Name	Short description
1	snr	S/N of the integrated profile
2	deq	Equivalent duty cycle of the integrated profile
3	cider	Confidence index in true source DM being higher than 2.0 pc cm^{-3}
4	cscore	Level of presence of candidate period in other receiver beams during the same observation
5	actime	Average correlation coefficient between integrated profile and individual sub-integrations
6	acfreq	Average correlation coefficient between integrated profile and individual sub-bands
7	zratio	Ratio between profile S/N at best DM and at DM=0

Table 3. A confusion matrix tracks the four possible outcomes of binary classification on a dataset where the true class labels (noted + and −) are known.

	Predicted +	Predicted −
Actual +	True positive (TP)	False negative (FN)
Actual −	False positive (FP)	True negative (TN)

learning algorithms such as decision trees and forests (Breiman 2001) are intrinsically robust to high dimensional data because they operate only on select subsets of the feature space, but such is not the case of ANNs.

(ii) Choose only features with high relevance, i.e. highly correlated to the class labels (but not necessarily in a linear sense). This is in line with properly spending the limited feature budget we imposed above.

(iii) Prefer monotonic features. A monotonic feature is such that an increase (or decrease) in its value always corresponds to an increased probability of positive data class membership. When using only monotonic features, both the positive and negative data classes tend to lie in distinct corners of the feature space, which means that the optimal decision boundary between both classes has a simple parametrization. We found monotonic features to facilitate the training process of our ANN and help to increase the accuracy of its predictions.

The list of features we use as inputs to SPINN v2 can be found in Table 2. We relied heavily on domain knowledge to design them, but we note that more systematic approaches can be adopted. Given a large feature set on a learning problem where high dimensionality is undesirable, a nearly optimal subset of useful features can be determined using so-called feature selection algorithms. Section 5 of Lyon et al. (2016) covers in detail the topic of feature selection in the context of pulsar candidate classification.

2.5.2 Classifier evaluation

A classifier's performance is measured by deploying it on a test sample of labelled data that was *not* seen during training. The total number of successful or failed predictions on each data class, positive and negative, can be conveniently tracked in a so-called confusion matrix (Table 3) from which several figures of merit can be derived. In the case of pulsar candidate classification, there are two simultaneous goals: maximize discovery potential while significantly reducing the number of false positive candidates to be

viewed by human operators. An inevitable trade-off exists between the two, that can only be fully captured by measuring the two types of error rates:

$$FNR = \frac{FN}{FN + TP}$$
$$FPR = \frac{FP}{FP + TN}, \tag{7}$$

where FNR is the false negative rate, here the fraction of pulsars missed, and FPR the false positive rate or fraction of spurious candidates incorrectly reported as pulsars. In place of FNR we may also use *Recall*, defined as the fraction of positives properly identified:

$$Recall = \frac{TP}{FN + TP} = 1 - FNR. \tag{8}$$

We trained SPINN v2 on a sample of labelled candidates from medlat and tested it on all the candidates produced by searching the first year of observations of the recent SUPERB survey. SUPERB also uses the Parkes multibeam receiver with the same backend, similar integration times on the sky, and the raw data were processed with the same infrastructure described in Sections 2.2 and 2.4. Most importantly, SUPERB covers a region of the sky that does not overlap with medlat by design, ensuring the fairness of the test. The test set contained 139 known pulsar observations and 1 418 598 negatives, whose class labels were determined by a combination of cross-referencing with the ATNF pulsar catalogue (Manchester et al. 2005) and visual inspection for all potential positives. All known pulsars were thus identified; a few undiscovered pulsars may have remained in the set of negatives, but the potential impact on the measured error rates is negligible. Every test candidate was assigned a score between 0 and 1 by SPINN v2, where a higher score indicates a higher likelihood of being a pulsar. This allowed us to evaluate classification error rates as a function of score threshold. The results can be seen in fig. 8 of Keane et al. (2018). The balance between recall and false positive rate is left to the user; for example, one can expect to find 99 per cent of known pulsars while still dividing the size of the metaphorical haystack of spurious candidates by a factor of nearly 1000.

As for classification speed, one instance of SPINN v2 runs on a single CPU and scores approximately 180 candidates per second on an Intel Xeon E5 2.2GHz CPU. This takes into account the time required to read the candidate from disc, compute the features, activate the neural network to obtain a score and store the results. Any classification problem can be trivially parallelized by deploying multiple identical classifier instances. The G2 supercomputer allowed us to run 64 SPINN v2 instances at a time, which means that the entire output of the pulsar searches described here (17.6 million candidates) could be processed in less than 30 min. We decided for both searches to visually inspect candidates with a score above a threshold corresponding to 99 per cent recall; this returned less than 0.1 per cent of the total output of the search (Table 4). The detailed visual inspection of these high-scoring candidates from the hilat portion of the survey required less than 30 person-hours.

3 DISCOVERIES

We have found a total of 23 new, confirmed pulsars whose discovery parameters are presented in Table 5. Here, we only provide positions of the discovery beam and spin periods to six significant figures, leaving full timing solutions to future publications (Keith et al., in preparation; Barr et al., in preparation). The observation IDs in which these pulsars were found and the date of their initial

Table 4. Candidate classification and inspection summary. In both searches, candidates were selected for visual inspection down to a score threshold corresponding to a recall of 99%.

Search	Candidates	Classifier	Recall	Candidates inspected	Fraction inspected	Discoveries
Medlat	4 375 642	SPINN v1	99%	4 909	0.11%	8
Hilat	13 240 927	SPINN v2	99%	11 849	0.09%	14

Table 5. Confirmed pulsar discoveries. Where no DM-dependent distance estimation is given, the pulsar's DM lies beyond the limits of the Galaxy according to that model, which likely points at model flaws for some lines of sight. Some of these pulsars, although not formally published, have been independently found by other survey teams.

PSR J	R.A. (hh:mm:ss)	Decl. (dd:mm:ss)	l ($°$)	b ($°$)	P_0 (s)	DM (pc cm^{-3})	S/N	D_{NE2001} (kpc)	D_{YMW16} (kpc)	Indep. discovery
J0125−23[a]	01:25:10	−23:27:26	189.1	−81.5	0.003 675 86	10	33	0.4	0.9	G
J0636−3044[b]	06:37:09	−30:50:42	239.7	−16.3	0.003 945 76	16	11	1.0	0.7	
J0753−0816[b]	07:53:33	−08:19:35	227.6	9.8	2.093 62	38	15	1.9	1.7	
J0839−66	08:39:37	−66:34:52	281.4	−14.9	0.449 339	83	15	3.2	0.4	
J0951−71	09:51:24	−71:19:01	289.5	−13.3	0.212 378	77	16	2.5	1.7	
J1000+08	10:00:38	+08:19:58	229.7	45.5	0.440 372	21	16	0.9	1.7	L
J1403−0314[b]	14:03:41	−03:15:28	335.6	55.0	0.362 634	31	21	−	−	
J1517−32[a]	15:17:01	−32:26:05	335.4	21.1	0.064 401 9	26	36	1.0	0.9	G
J1558−67	15:58:21	−67:35:59	319.3	−10.9	0.267 268	105	10	2.7	5.1	
J1654−26	16:54:39	−26:36:18	355.5	10.6	1.623 73	129	18	3.6	11.5	
J1703−18	17:03:26	−18:48:54	3.1	13.6	1.270 24	46	12	1.4	1.4	
J1705−1903[c]	17:05:27	−19:08:04	3.1	13.0	0.002 480 22	58	16	1.7	2.4	
J1708+02	17:08:39	+01:48:06	22.3	23.6	0.410 772	29	24	1.3	1.5	
J1754+0032[c]	17:54:52	+00:36:33	26.9	12.8	0.004 410 80	70	10	2.4	3.7	
J1804−2858[c]	18:04:19	−28:58:31	2.0	−3.6	0.001 492 68	232	11	4.9	8.8	
J1842−27	18:42:02	−27:58:19	6.7	−10.5	0.815 269	62	11	1.9	2.6	
J1843−40	18:43:03	−40:33:51	355.0	−15.8	0.324 187	66	16	1.8	3.9	
J1921−05	19:21:04	−05:23:08	31.5	−9.0	2.227 59	98	16	3.2	5.7	G
J1940+0239[b]	19:40:33	+02:36:41	40.9	−9.7	1.232 24	90	26	3.5	5.7	G
J1942+0147[b]	19:42:27	+01:50:11	40.5	−10.5	1.405 04	151	15	7.0	−	
J1947−18	19:47:31	−18:58:12	21.5	−20.6	0.002 603 23	25	16	1.0	1.1	
J2228−65	22:28:18	−65:11:47	323.5	−45.8	2.745 98	36	31	1.8	−	S
J2354−22[a]	23:54:26	−22:51:53	48.1	−76.4	0.557 996	8	13	0.4	0.9	G

Notes. [a]Date of identification in other survey predates ours.
[b]Timing solutions will be presented in Keith et al. (in preparation).
[c]Timing solutions will be presented in Barr et al. (in preparation).
G: Green Bank Northern Cap Celestial survey, http://astro.phys.wvu.edu/GBNCC/.
S: SUPERB survey, https://sites.google.com/site/publicsuperb/discoveries.
L: LOTAAS survey, http://www.astron.nl/lotaas/.

identification as candidates worthy of confirmation can be found in Table 6. While none of the pulsars reported here have been previously published formally in the literature, we acknowledge that five have been independently found by other survey teams and listed on their official websites. Of the 23 confirmed discoveries, 6 are millisecond pulsars (MSPs) and another one (PSR J1517 − 32) appears to be mildly recycled. At least four are part of binary systems: PSRs J0636−3044, J1517 − 32, J1705−1903, and J1754 + 0032.

PSR J1705−1903 is a millisecond pulsar with a spin period of 2.48 ms and a very low mass companion (0.047 M$_\odot$ median mass) indicative of a black widow. It undergoes eclipses for about 15 per cent of its orbit at 1400 MHz and has an extremely narrow pulse width of $\approx$40 μs. Its usefulness as a timing array pulsar might be hindered by the presence of long-term timing noise likely caused by a non-uniform density of ablated companion material dispersed in its orbital plane (e.g. Fruchter et al. 1988). However, its narrow pulse width gives it excellent short-term timing properties which can be leveraged to accurately measure DM and scattering varia-

tions throughout its orbit; this can in turn be used to characterize the eclipses and the ablation process the companion is undergoing, following work previously done on other similar black widow systems such as PSR J2051−0827 (Stappers et al. 2001), PSR J1544 + 4937 (Bhattacharyya et al. 2013), or PSR J1810 + 1744 (Polzin et al. 2018).

PSR J1804−2858 has a spin frequency of 670 Hz, the third highest currently known. Despite being isolated, it shows a positive $\dot{P}$ that is likely caused by acceleration in the local Galactic potential; the possibility of a multiyear eccentric orbit around a massive companion cannot be ruled out however, as was recently demonstrated in the case of PSR J2032 + 4127 (Lyne et al. 2015). PSR J1804−2858 has the second highest DM/P ratio of any known pulsar, behind that of the globular cluster pulsar PSR J1748-2446ad (Hessels et al. 2006). The highest DM/P a survey finds has previously been put forward as a measure of survey depth (e.g. Lazarus 2013), which in the case of MSPs is limited by so-called DM smearing: the artificial pulse widening caused by the uncorrected dispersion delay Δt

Table 6. Pointing and beam number for each of the newly discovered pulsars. Pointings are uniquely identified by their start UTC date and time. Discovery date refers to the first visual identification as a candidate worthy of re-observation.

PSR	Survey	Observation UTC (Y-m-d H:M:S)	Beam	Discovery date (Y-m-d)
J0125–23	H	2010-12-31 07:13:55	6	2015-05-07
J0636–3044	H	2011-06-26 00:01:54	13	2015-05-07
J0753–0816	H	2013-01-24 10:57:05	3	2015-06-11
J0839–66	H	2012-03-26 09:31:39	10	2015-05-28
J0951–71	M	2008-12-10 20:37:06	1	2014-01-07
J1000+08	H	2013-12-28 17:40:54	3	2015-05-28
J1403–0314	H	2013-12-30 22:21:55	3	2015-05-28
J1517–32	H	2013-04-05 12:48:51	8	2015-05-07
J1558–67	M	2010-01-21 22:22:30	5	2014-01-07
J1654–26	M	2008-11-23 23:33:40	1	2014-01-07
J1703–18	M	2009-08-21 10:26:40	8	2014-02-23
J1705–1903	M	2010-04-21 16:56:27	10	2013-10-14
J1708+02	H	2013-12-05 23:27:05	4	2015-05-28
J1754+0032	M	2009-08-19 08:37:33	12	2014-01-16
J1804–2858	M	2009-03-08 01:45:33	13	2013-11-02
J1842–27	M	2008-11-24 06:06:51	8	2014-02-19
J1843–40	H	2011-12-08 07:23:15	3	2015-05-28
J1921–05	H	2012-03-31-00:30:01	12	2015-06-11
J1940+0239	H	2013-05-01 17:13:28	7	2015-06-11
J1942+0147	H	2013-10-17 10:30:42	6	2015-06-11
J1947–18	H	2013-07-12 18:01:34	4	2015-05-07
J2228–65	H	2012-01-29 05:18:35	1	2015-06-11
J2354–22	H	2009-04-16 21:19:11	13	2015-05-28

within a single frequency channel, given by

$$\Delta t = 8.3 \times 10^3 \left(\frac{DM}{pc\,cm^{-3}} \right) \left(\frac{\Delta\nu}{MHz} \right) \left(\frac{\nu}{MHz} \right)^{-3} s, \qquad (9)$$

where DM is the dispersion measure of the source, $\Delta\nu$ the width of a frequency channel and ν the central observing frequency. For the HTRU survey observing set-up $\Delta t = 285\,\mu s$ at the pulsar's $DM = 232\,pc\,cm^{-3}$, compared to 2.2 ms for the previous generation of highly successful Parkes surveys (e.g. Manchester et al. 2001). The increased frequency resolution of HTRU made this discovery possible, and pursuing this trend of finer channelization on the next generation of telescopes will push the MSP detectability horizon further into the Galactic plane, unless limited by interstellar scattering.

Finally, one MSP candidate deserves mention. J1618–36 was found in the medlat portion of the survey with a period of 5.78 ms and an unusually high acceleration (16.3 ms^{-2}) for a millisecond pulsar. The parameters of its original detection can be found in Table 7, and the candidate plot in Fig. 5. We note that the positive sign on the acceleration means that the source is accelerating away from the observer, and would correspond to the near side of an orbit. The detection is not due to random noise fluctuations (S/N = 14), and has all the properties that indicate a pulsar nature; it was detected in a single beam, is broad-band with a very significantly non-zero dispersion measure and all candidates reported with a similar score by our classifier are genuine pulsars. Its period also lies within an RFI-quiet range. We have unfortunately failed to detect it again after spending a cumulated 6 h of Parkes time over the years 2015–2017, mostly in 9- and 18-min integrations.

To constrain the possible nature of J1618–36, we need to place its acceleration at the time of detection in a broader context. We

Table 7. Original detection parameters of the highly accelerating MSP *candidate* J1618–36. The FWHM of a beam of the Parkes multibeam receiver is 14 arcmin.

Parameter	Value
α (J2000)	16h18m32.1s
δ (J2000)	−36d03m42s
l	343.11°
b	10.18°
MJD	55065.240992
P_{bary}	5.778747 ms
DM	104 pc cm^{-3}
Acc	16.3 ms^{-2}
S/N	14
D_{NE2001}	2.7 kpc

calculated[6] the maximum l.o.s.a. of all binary pulsars for which orbital parameters are available in the ATNF pulsar catalogue. Defining MSPs as pulsars with spin periods below 20 ms, we find that the only known MSP to reach a maximum l.o.s.a. higher than 16.3 ms^{-2} is PSR J2215+5135 (Hessels et al. 2011), a redback in a 0.17-d orbit around a main-sequence companion star with a median mass of 0.24 M$_\odot$. The top six millisecond pulsars ranked by maximum acceleration all have companions with ≥ 0.1 M$_\odot$ and orbital periods shorter than a day. Combined with the sequence of non-detections, this tentatively suggests that J1618–36 is a redback eclipsed for the vast majority of its orbit, or a new specimen of transitional millisecond pulsar (see e.g. Campana & Di Salvo 2018, and references therein). If such is the case, it should be detectable as a low-mass X-ray binary when in an accreting state. There are two faint ROSAT X-ray sources within one Parkes beam error radius
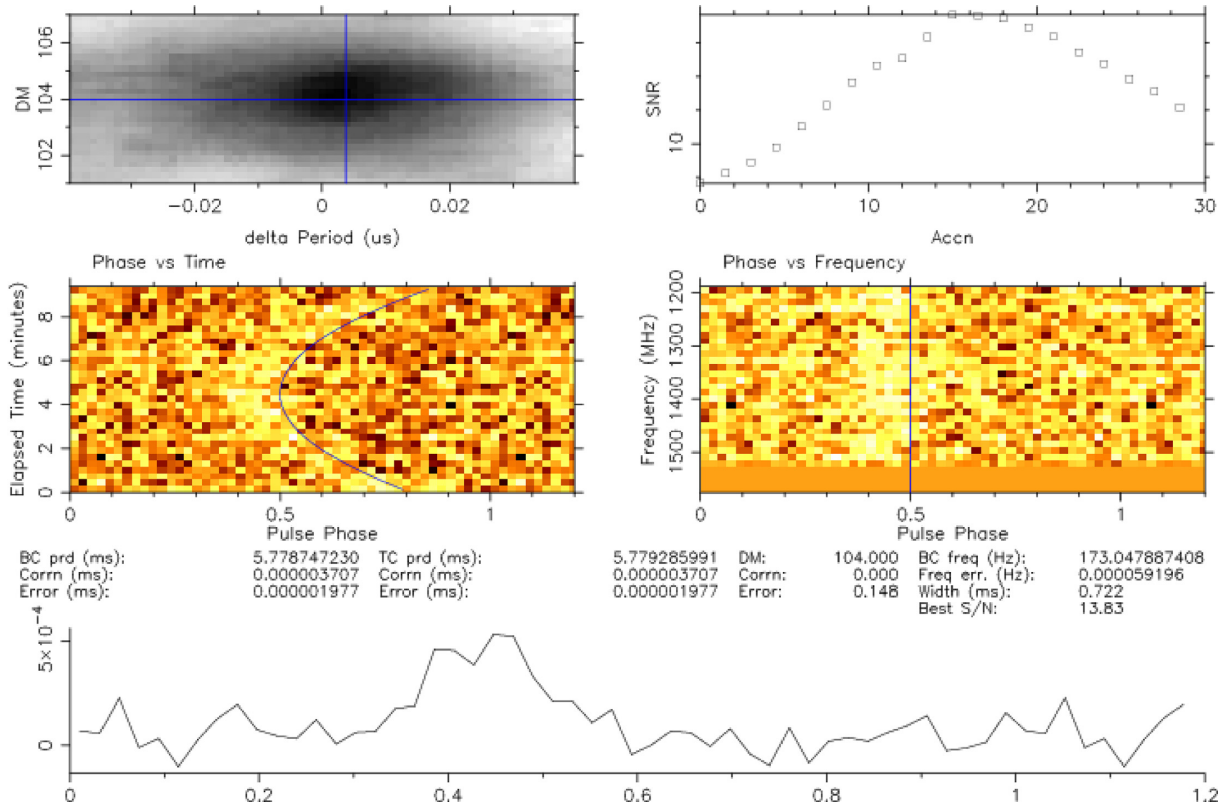
---

[6]https://github.com/ewanbarr/pyorbit

Figure 5. Candidate plot for J1618–36, generated with PSRCHIVE. The phase versus time plot (centre left) was generated without correcting for acceleration to make the residual quadratic phase drift visible; the overplotted thin blue line shows the best-fitting quadratic phase model. The upper right plot shows folded S/N as a function of trial acceleration, giving a best-fitting value between 15 and 18 m s^{-2} consistent with the 16.3 m s^{-2} reported by the Fourier search.

around the discovery position, namely 1RXS J161835.3–360329 and 1RXS J161820.8–360306 whose angular separation from the candidate position are 41 and 142 arcsec, respectively. Both sources, if placed at a distance of 2.7 kpc, would have 0.5–10 keV X-ray fluxes of $\approx 10^{33}$ erg s^{-1}, which would be broadly consistent with a transitional MSP hypothesis. We were granted a 2 ks Swift observation (Observation ID: 00033792002, taken on 2015 May 28) to put this idea to the test. If the flux of the source matched that described above then a 46-count source would be detected, but unfortunately no significant source was visible in the field. In spite of the elusiveness of J1618–36, we encourage further confirmation attempts and checking of any archival data available due to the very high quality of the initial radio detection, which remains very unlikely to be explained by radio frequency interference or noise.

4 COMPARISON WITH PREVIOUS SEARCHES OF THE SAME DATA

Pulsar searches can be evaluated using three major criteria, namely speed, completeness, and human effort involved in candidate classification. In this section, we compare the previously used SIGPROC-based pipeline and candidate classification tools with the new pipeline presented in Section 2. We will mostly focus on the case of hilat, the survey section on which we used our new RFI mitigation algorithm (see Section 2.4).

4.1 Speed

The speed evaluation was focused on estimating how long it would have taken to perform a complete acceleration search of hilat with the old pipeline, using the same search parameters we selected (Table 1) and the same computing cluster. The old pipeline was purely CPU-based. It performed brute-force dedispersion with the dedisperse_all routine and periodicity search with seek, both of which are part of SIGPROC. Candidate summary plots were produced with the DSPSR (van Straten & Bailes 2011) and PSRCHIVE (Hotan et al. 2004) software packages, using dspsr for folding, paz for masking a static list of bad channels, and finally pdmp for candidate optimization. For the sake of this comparison, the programs that could only use a single CPU core (seek and pdmp) were parallelized using the multiprocessing PYTHON library, with negligible overhead.

We measured the execution times of every major pipeline step on G2 (see Section 2.3), allocating to each of them three Intel Xeon E5-2660 CPU cores and one Tesla K10 accelerator to process a single 270-s hilat beam. This resource allocation simulated optimal cluster usage conditions in a large scale search, since users are simultaneously limited to a maximum of 64 concurrent jobs, 64 GPUs, and 192 CPU cores. Based on the average number of DM trials in our own hilat processing, we performed the benchmark with 650 DM trials, 30 acceleration trials, and 50 candidates folded per beam. The benchmark results are shown in Table 8. We note that under the 'Search' operation are included red noise removal, time domain re-sampling, FFT, and harmonic summing. We then scaled those results by the total number of beams in hilat ($N_{\mathrm{beams}} =$

Table 8. Speed comparison between major steps of the old and new processing pipelines on a single hilat beam with a 270-s integration time. Benchmarks were performed on the Green II supercomputer, assuming an average of 650 DM trials, 30 acceleration trials, and 50 folded candidates produced per beam. See 4.1 for details. It is interesting to note that folded candidate production could now become the main bottleneck of pulsar searching in some use cases at least.

	Operation	Old pipeline (s)	New pipeline (s)	Speed-up factor
Periodicity search	Dedispersion	380	28	14
	Search	5577	83	67
	Sub-total	5957	111	54
Candidate production	Folding	142	53	2.7
	RFI mitigation	5	40	0.12
	Optimization	97	65	1.5
	Sub-total	243	158	1.5
TOTAL		6200	269	23.1

Table 9. Estimated resource consumption of a complete hilat acceleration search with both the old and new pipelines, based on the run times given in Table 8.

	Old pipeline	New pipeline
CPU-hours	2 300 000	58 000
GPU-hours	0	41 000

443287) to evaluate the total cost of a complete search of the survey (Table 9). Although faster single-beam run times could be obtained for the old pipeline by allocating it more CPUs, this would translate into more total CPU-hours used in a full-scale search.

From these results, it is clear that using the GPU-based PEASOUP pipeline made an acceleration search of HTRU possible with the specific resources we had at our disposal, running Fourier searching between 1.5 and 2 orders of magnitude faster than on CPUs. We note however that contrary to PEASOUP, seek was not designed to efficiently process multiple acceleration trials concurrently. This large speed-up factor also created an interesting situation where folding candidates takes more time than a wide-range acceleration search.

4.2 Output candidate quality and search completeness

An in-depth output comparison between the old and new pipelines on the hilat survey section has been performed in chapter 5 of Cooper (2017), of which we briefly summarize the results relevant to this section. This study was limited to slow pulsars with periods longer than 100 ms and our discussion will carry the same restrictions. Of interest is the survey area processed by both pipelines, called the overlap region hereafter and which contains 10 427 pointings, amounting to 31 per cent of the survey and nearly all the data that had been searched prior to this work. It was found that 42 known pulsars were detected by either pipeline in this overlap region; successful detection means that an associated candidate plot has been produced and is identifiable as a pulsar by eye. All 42 were detected by the new pipeline versus 38 for the old one. Furthermore, three of the new discoveries lying in the overlap region were not detected by the old pipeline either: PSR J0753–0816, PSR J1942 + 0147 and PSR J2228–65. The parameters of all seven pulsars are summarized in Table 10. In at least two cases (PSR J0536–7543 and PSR J2048–1616), the old pipeline produced a candidate plot where the pulsar could not be identified due to the presence of RFI.

With small sample caveats in mind, the fact that three slow pulsar discoveries were not detected by the old pipeline over 10 427 pointings can be taken to be very significant: linearly scaled to the amount of data we searched (27 588 pointings), this would amount to an expected eight pulsars missed. Considering that the current slow pulsar yield of hilat stands at 22 including the 11 new ones reported here (Table 5), it means that changes in RFI mitigation schemes and other low-level processing steps have the potential to increase the slow pulsar yield by *tens of percent*.

This validates the efficiency of the RFI mitigation algorithm presented in Section 2.4, in real large-scale search conditions: no pulsars were lost due to its use and extra discoveries were enabled. But more importantly, this comparison underlines the large impact that various data manipulations (RFI mitigation being an important one, but not the only one) in the processing chain can have on search completeness before candidates even have a chance to reach the classification stage. We discuss this further in the next section.

4.3 Candidate selection accuracy and human effort

Two methods have been previously used to deal with candidate selection in HTRU med- and hilat searches: interactive candidate selection software such as JREAPER (Keith et al. 2009), and another ANN (Bates et al. 2012). With the first method, users project a batch of candidates on scatter plots where the axes are typically period and S/N ratio, or period and DM; from there they make a reduced selection of candidate plots to view in detail, exploiting tendencies of RFI candidates to cluster in narrow ranges of periods or around a dispersion measure of zero (Levin 2012). Due to its subjectiveness, the accuracy of this pre-selection process cannot be measured, precluding any comparison with ML algorithms. In terms of human effort, the total time spent viewing candidates to complete the initial medlat search using interactive software was approximately 100 person-hours.

The ANN of Bates et al. (2012) provides a more interesting point of discussion. It was both trained and evaluated on medlat candidates produced with the previous SIGPROC pipeline, and was found to have a false positive rate (FPR) of 0.3 per cent but a recall limited to 85 per cent (70 per cent on millisecond pulsars). Such an FPR is low enough to make the inspection of the entire survey output easily manageable by a single person, but a recall rate of 85 per cent offers a more limited discovery potential. We managed to significantly improve on the accuracy figures above, mainly due to the larger amount of training data we had at our dis-

Table 10. Parameters of the seven pulsars *not* detected by the old pipeline in hilat. Four were previously known, three are part of the new discoveries (Table 5) and are marked with a star symbol. Their candidate plots were either not produced, or not recognizable. $D_{\rm beam}$ is the beam offset from the true source position in arcmin, and the S/N is that reported by the new pipeline.

PSR J	P_0 (s)	DM (pc cm^{-3})	S_{1400} (mJy)	$D_{\rm beam}$ (arcmin)	S/N
J0536–7543	1.246	18	13.0	12.3	33
J0753–0816*	2.093	38	–	–	15
J1332–3032	0.650	15	0.30	7.1	17
J1714–1054	0.696	51	–	7.0	18
J1942+0147*	1.405	151	–	–	15
J2048–1616	1.962	12	13.0	9.0	366
J2228–65*	2.746	36	–	–	31

posal. Bates et al. (2012) trained their classifier soon after survey observations first started, on a limited data sample containing 70 candidates of the positive (pulsar) class. We were fortunate to have access to the whole survey and more than 1000 such data instances, including discoveries from previous searches. They also used more candidate features (22 in total) in spite of having less training examples, which may have had a negative impact on accuracy (Section 2.5).

5 DISCUSSION AND CONCLUSIONS

In this paper, we have described a pipeline with the two main functional elements that will be needed for future searches: a fast search code that leverages many-core architectures and an accurate candidate classification ML algorithm. We also devised and applied a new RFI mitigation method based on a simple outlier detection algorithm, which operates on folded data and has just one free parameter. The pipeline was used to search the intermediate and high latitude regions of the HTRU survey up to an acceleration of $50\,{\rm m\,s^{-2}}$ and 23 new pulsars were discovered, including 6 MSPs and at least 4 in binary systems. The new search described here is shown to exceed the performance of previous searches on the same data both in terms of speed and completeness, and requires a much lower amount of human involvement in the candidate selection process. The improved completeness is evidenced by the fact that the majority of the new discoveries are isolated, slow pulsars that lie in a previously searched area of the parameter space.

The acceleration search of medlat and hilat was made possible by moving the time-domain re-sampling and Fourier search operations to the GPU, as they were by far the most time consuming in the previous CPU-based pipeline. The resulting search code PEASOUP performs those tasks on order of 50 times faster (Table 8), making candidate folding the new bottleneck. Speeding up candidate folding will likely become a requirement in the future, considering that, for example the folding of up to 1000 candidates per beam is currently being considered for SKA pulsar searches (Levin et al. 2017). GPUs are also a promising solution here. The folding of high time resolution spectra into a small number of phase bins is equivalent to histogram computation, and optimizing this operation on GPUs is an area of active research in computer science (e.g. Nugteren et al. 2011; Gómez-Luna et al. 2013). Large speed-ups over CPU implementations have already been demonstrated.

The most important lesson learned from the hilat search is the large impact of RFI mitigation on search completeness, and the need to invest more effort into it. Since their introduction less than

a decade ago (Eatough et al. 2010), candidate classification algorithms have greatly improved in accuracy, almost enough to be used on surveys envisioned with the 1500-beam SKA1-Mid (Keane 2017). SPINN v2 would still have made the sifting of a 1500-beam version of HTRU med- and hilat doable within a few person-days, if a recall of 90 per cent can be considered acceptable (fig. 5.1 of Morello 2016). Lower error rates have since been achieved using ensemble classification (Yao, Xin & Guo 2016). While this is welcome news, investing further effort in pushing classifiers to higher levels of accuracy is not the most cost-effective avenue if more than 10 per cent of potential discoveries can be lost in earlier stages of the processing chain.

We have pointed out the large potential impact of RFI mitigation on search completeness, but major concerns have also been raised about red noise removal schemes negatively affecting the detectability of pulsar signals (Lazarus et al. 2015; van Heerden, Karastergiou & Roberts 2017). Other processing steps such as harmonic summing, clustering of Fourier detections into candidates, or the time-domain candidate optimization process also have their own free parameters and the values chosen for these parameters impact search completeness in a way that is not quantitatively known. If one is ever to evaluate the recall and false positive rate of a pipeline *as a whole* then it must be extensively tested on data generated by injecting simulated pulsar signals in randomly selected survey observations. Setting up such an evaluation framework may be computationally challenging but it would allow one to submit all internal pipeline parameters to an optimization process, similar to the training of an ML algorithm where its internal weights are fit to a sample of training data. And with a wide variety of generated pulsar signals available, candidate classifiers could be trained with nearly unlimited amounts of data, and the search completeness could be accurately known as a function of many physical pulsar parameters. Similar testing frameworks are already employed on the LIGO and Virgo gravitational wave detectors (e.g. Abbott et al. 2016; Biwer et al. 2017) to characterize detections caused by non-Gaussian noise, and one is planned to monitor the sensitivity of the CHIME FRB search (The CHIME/FRB Collaboration et al. 2018). Artificial signal injection has also been used as part of the PALFA survey to better evaluate the sensitivity of periodicity searches (Lazarus et al. 2015; Parent et al. 2018). As we move towards massively multibeam systems and real-time pulsar searches where the raw data are immediately discarded, with the additional requirement of automating the discovery process, pulsar searching pipelines are in danger of becoming completely opaque unless their whole processing chain is tested thoroughly.

ACKNOWLEDGEMENTS

REFERENCES

Abbott B. P. et al., 2016, Class. Quantum Gravity, 33, 134001

Barsdell B. R., Bailes M., Barnes D. G., Fluke C. J., 2012, MNRAS, 422, 379

Bates S. D. et al., 2012, MNRAS, 427, 1052

Bhattacharyya B. et al., 2013, ApJ, 773, L12

Biwer C. et al., 2017, Phys. Rev. D, 95, 062002

Breiman L., 2001, Mach. Learn., 45, 5

Burke-Spolaor S. et al., 2011, MNRAS, 416, 2465

Campana S., Di Salvo T., 2018, preprint (arXiv:1804.03422)

Chandola V., Banerjee A., Kumar V., 2009, ACM Comput. Surv., 41, 15:1

Cooper S., 2017, PhD thesis, The University of Manchester

Cordes J. M., Lazio T. J. W., 2002, preprint (arXiv:astro-ph/0207156)

Eatough R. P., Molkenthin N., Kramer M., Noutsos A., Keith M. J., Stappers B. W., Lyne A. G., 2010, MNRAS, 407, 2443

Eatough R. P., Kramer M., Lyne A. G., Keith M. J., 2013, MNRAS, 431, 292

Fruchter A. S., Gunn J. E., Lauer T. R., Dressler A., 1988, Nature, 334, 686

Gómez-Luna J., González-Linares J., Ignacio Benavides J., Guil N., 2013, Mach. Vis. Appl., 24, 899

Hastie T., Tibshirani R., Friedman J., 2009, The elements of statistical learning: data mining, inference and prediction, 2 edn. Springer-Verlag, Berlin

Hessels J. W. T. et al., 2011, in Burgay M., D'Amico N., Esposito P., Pellizzoni A., Possenti A., eds, AIP Conf. Proc. Vol. 1357, Radio Pulsars: an Astrophysical Key to Unlock the Secrets of the Universe. Am. Inst. Phys., New York, p. 40

Hessels J. W. T., Ransom S. M., Stairs I. H., Freire P. C. C., Kaspi V. M., Camilo F., 2006, Science, 311, 1901

Hotan A. W., van Straten W., Manchester R. N., 2004, PASA, 21, 302

Johnston H. M., Kulkarni S. R., 1991, ApJ, 368, 504

Keane E. et al., 2015, Advancing Astrophysics with the Square Kilometre Array (AASKA14), 40, Available at: http://adsabs.harvard.edu/abs/2015aska.confE..40K

Keane E. F. et al., 2018, MNRAS, 473, 116

Keane E. F., 2017, Proc. IAU Symp. 337, Pulsar Astrophysics the Next Fifty Years. Macclesfield, UK, p. 158

Keane E. F., Ludovici D. A., Eatough R. P., Kramer M., Lyne A. G., McLaughlin M. A., Stappers B. W., 2010, MNRAS, 401, 1057

Keith M. J. et al., 2010, MNRAS, 409, 619

Keith M. J., Eatough R. P., Lyne A. G., Kramer M., Possenti A., Camilo F., Manchester R. N., 2009, MNRAS, 395, 837

Knispel B. et al., 2013, ApJ, 774, 93

Kocz J., Bailes M., Barnes D., Burke-Spolaor S., Levin L., 2012, MNRAS, 420, 271

Lazarus P. et al., 2015, ApJ, 812, 81

Lazarus P., 2013, in van Leeuwen J., ed., Proc. IAU Symp. 291, Neutron Stars and Pulsars: Challenges and Opportunities after 80 years. Beijing, China, p. 35

Levin L., 2012, PhD thesis, Swinburne University of Technology, Available at: http://hdl.handle.net/1959.3/312190

Levin L. et al., 2017, Proc. IAU Symp. 337, Pulsar Astrophysics the Next Fifty Years. Macclesfield, UK, p. 171

Lorimer D. R., Bailes M., McLaughlin M. A., Narkevic D. J., Crawford F., 2007, Science, 318, 777

Lyne A. G., Stappers B. W., Keith M. J., Ray P. S., Kerr M., Camilo F., Johnson T. J., 2015, MNRAS, 451, 581

Lyon R. J., , Stappers B. W., Cooper S., Brooke J. M., Knowles J. D., 2016, MNRAS, 459, 1104

Manchester R. N. et al., 2001, MNRAS, 328, 17

Manchester R. N., Hobbs G. B., Teoh A., Hobbs M., 2005, AJ, 129, 1993

Morello V., 2016, Master's thesis, Swinburne University of Technology, Available at: http://hdl.handle.net/1959.3/434704

Morello V., Barr E. D., Bailes M., Flynn C. M., Keane E. F., van Straten W., 2014, MNRAS, 443, 1651

Ng C. et al., 2015, MNRAS, 450, 2922

Ng C., 2017, Proc. IAU Symp. 337, Pulsar Astrophysics the Next Fifty Years. Macclesfield, UK, p. 179

Nugteren C., van den Braak G.-J., Corporaal H., Mesman B., 2011, Proc. Fourth Workshop on General Purpose Processing on Graphics Processing Units. GPGPU-4. ACM, New York, NY, USA, p. 1:1

Parent E. et al., 2018, ApJ, 861, 44

Polzin E. J. et al., 2018, MNRAS, 476, 1968

Ransom S. M., Eikenberry S. S., Middleditch J., 2002, AJ, 124, 1788

Stappers B. W., Bailes M., Lyne A. G., Camilo F., Manchester R. N., Sandhu J. S., Toscano M., Bell J. F., 2001, MNRAS, 321, 576

Staveley-Smith L. et al., 1996, PASA, 13, 243

CHIME/FRB Collaboration et al., 2018, ApJ, 863, 16

Thornton D., 2013, PhD thesis, The University of Manchester

Thornton D. et al., 2013, Science, 341, 53

Tukey J. W., 1977, Exploratory Data Analysis. Addison-Wesley, Boston, USA

van Heerden E., Karastergiou A., Roberts S. J., 2017, MNRAS, 467, 1661

van Straten W., Bailes M., 2011, PASA, 28, 1

Yao Y., Xin X., Guo P., 2016, 2016 12th International Conference on Computational Intelligence and Security (CIS). Wuxi, China, p. 120

This paper has been typeset from a TeX/LaTeX file prepared by the author.