Getting the Best from the Pulsar Folding Algorithm

Peter East

Abstract

The signal-to-noise ratio (SNR) prediction of the pulsar radiometer equation is achieved by matching the folding algorithm bandwidth to that required to just pass the pulsar pulse spectrum. The basic folding algorithm has SNR limitations but still has its uses. This article investigates folding theory and applies the recommended process to measurements made on the Vela and B0329 pulsars covering SNR extremes and finally interprets the results.

Introduction

The folding algorithm is used for enhancing the visibility of a pulsar by synchronously adding a number of pulsar periods along the data record. Simply explained, the pulsar signal integrates linearly whilst the background noise integrates as the square root, so improving the detection signal-to-noise ratio (SNR). As with most digital signal processing tasks, there are several ways of implementing this algorithm. A typical amateur method, simply divides the pulsar period into a number of time bins and sums each consecutive period synchronously. The final graphical indication of both the pulse shape and noise depend upon the number of time bins set in the pulsar period; many time bins improve the pulse detail but in the simplest case can degrade the SNR. This basic process can be improved as implied by the pulsar radiometer equation and the aim of this document is to show how to achieve both optimum SNR and folded time resolution simultaneously. Firstly the pulsar radiometer equation is reviewed to understand what should be achieved.

The Pulsar Radiometer Equation

Equation 1 is the pulsar radiometer equation derived by Lorimer[(1)] which describes the folding algorithm performance in generating an observable signal-to-noise ratio (SNR) for a single received polarization,

$$SNR = \frac{S_p A_e \sqrt{t_{\text{int}} \Delta f}}{2\beta k_b T_{sys}} \sqrt{\frac{P-W}{W}} \qquad (1)$$

where, S_p is the pulsar mean flux at the observation frequency, as listed in the ATNF data base in Janskys (1 Jansky = 10^{-26} W/Hz/m^2).
A_e is the receiving antenna effective collecting area in m^2.
t_{int} is the total data acquisition time in seconds.
Δf is the receiver RF bandwidth in Hz.
P is the predicted accurate pulsar period in ms.
W is the pulsar pulse half-height width in ms.
the factor 2 halves the pulsar flux to conform with the flux with a single polarization.
β is a modifying factor to account for digitization losses for coarser digital increments.
k_b is Boltzmann's constant(= 1.380649×10^{-23} W/^{o}K/Hz).
T_{sys} is the antenna terminal system noise temperature in ^{o}K.

This equation actually describes the best possible SNR and to achieve this optimum value in practice requires some careful attention.
Since the raw data is sampled, to ensure some timing resolution the pulsar folding period is normally divided into a number of bins N for display. This effectively factors the integration time, so that it seems logical that the value N should appear in the Equation 1 as a t_{int} divisor. Also, the pulsar flux S_p is catalogued as a mean value but for SNR calculation we are more

interested in the peak value, so it might be thought that S_p should really be increased by the pulse peak-to-mean factor $(P\text{-}W)/W$, rather than its square root.
However, if the number of bins chosen, $N = (P\text{-}W)/W$, then Equation 1 appears to be satisfied. Indeed, this is the number of bins usually recommended for the basic folding method to obtain the best indicated SNR; but this may not always be the case. In addition, pulsar duty cycles $(P\text{-}W):W$ vary over quite a wide range and as a consequence the timing resolution of the basic method may be very poor for pulsars with low duty cycles.

Information in Recorded Data
The recorded data is assumed to comprise multi-level digital in-phase and quadrature (I/Q) samples sampled at a rate $1/t_c$ (= Δf). An analogue representation of the digital square-law detection process following a software defined radio (SDR) is shown in Figure 1.
The data sample amplitudes ($I^2 + Q^2$) contain both DC and AC components; both of these DC and AC components are proportional to T_{sys} (powers ≈ $kT_{sys}\,\Delta f$).
After low-pass filtering (LPF), often simply by data series averaging, the AC component power is reduced to $kT_{sys}\sqrt{\Delta f\,\Delta lpf}$, where Δlpf ($\rightarrow 1/t_{int}$) is the effective low-pass filter cut-off frequency. For block averaging of N samples, $\Delta f \rightarrow 1/Nt_c$.
In the presence of a pulsar pulse T_{sys} increases by T_p - the pulsar peak equivalent temperature. The pulsar DC contribution sits on the system noise DC component and it is the difference between these components that is of interest.

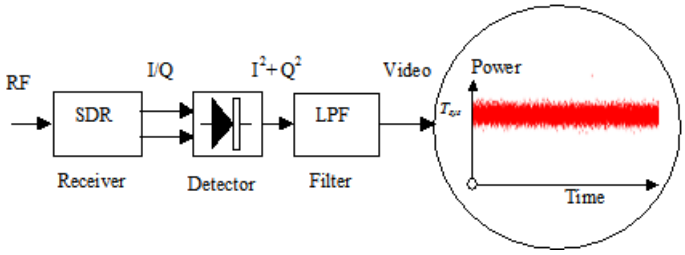


Figure 1. Raw Data Analogue-equivalent Detection Process.

With a large antenna and a strong pulsar, pulses may be visible on the noise section in Figure 1. This is a 30 minute section of real B0329 system data downsampled to 1ksps. The noise rms level here is approximately equivalent to 2°K and the expected pulsar pulse amplitude for this system is 0.05°K - a fair task for the folding algorithm.
Describing the detected system AC noise signal by $v_n(t)$ and the pulsar DC level by $v_p(t)$ - the pulsar AC component can usually be ignored compared to T_{sys} - then the equivalent pulsar input SNR is, $SNR_{in} = \overline{v_p^2(t)} / \overline{v_n^2(t)} = T_p/T_{sys}$.

Basic Bin Folding Analysis
In preparation for folding, the data recording for the duration T is notionally divided into R sections exactly equal to the pulsar period P so that, if present, the pulsar pulse will appear at the same relative position 't' amongst the random system noise as depicted in Figure 2.

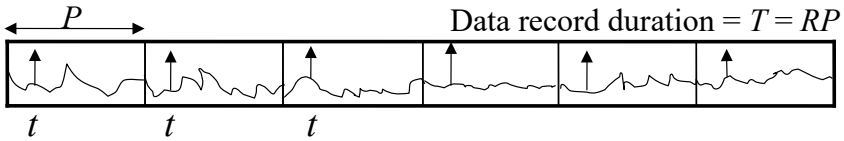


Figure 2. Data Record T Divided into R Digitized Blocks Equal in Time to the Pulsar Period P.

In folding, each period block of data is divided into *N* bins and the *R* periods added synchronously together.
After synchronously adding/folding, RP/Nt_c of the original data samples on average fall in each bin.
In a bin filled with a pulse, the summed DC detected voltage increases linearly to $\overline{v_p(t)}.RP/Nt_c$.
Once all RP/t_c data samples are folded in, the AC rms noise voltage increases by power to $\sqrt{\overline{v_n^2(t)}.RP/Nt_c}$ in all bins.
The indicated output voltage SNR_o is the ratio of the in-bin pulsar DC and the system noise AC terms from noise-only bins, or,

$$SNR_o = \frac{\overline{v_p(t)}RP/Nt_c}{\sqrt{\overline{v_n^2(t)}}\sqrt{RP/Nt_c}} = \frac{T_p}{T_{sys}}.\sqrt{\frac{RP}{Nt_c}}$$

The peak equivalent pulsar temperature for one received polarization, based on the radiometer equation definitions above is, $T_p = \frac{S_p}{2}\frac{P-W}{W}.\frac{A_e}{k_b}$ and since, $1/t_c = \Delta f$ (for complex I/Q samples), $t_{int} = RP$. With these substitutions, a basic version of the pulsar radiometer equation is derived for when the pulsar lies centrally in a bin,

$$SNR_o \approx \frac{S_p A_e \sqrt{\frac{t_{\text{int}}\Delta f}{N}}}{2k_b T_{sys}}\frac{P-W}{W} \qquad (2)$$

When, $N = (P-W)/W$, this equation can be rewritten as suggested earlier,

$$SNR_o \approx \frac{S_p A_e \sqrt{t_{\text{int}}\Delta f}}{2k_b T_{sys}}\sqrt{\frac{P-W}{W}}$$

and now this basic pulsar radiometer equation for the folding algorithm agrees with the official Equation 1 form.

Improving the Basic Folding Algorithm for Large Bin Numbers
As described, the basic folding algorithm of Equation 2 matches the performance of the pulsar radiometer Equation 1 at one value of $N = (P-W)/W$; for larger bin numbers the output SNR is reduced by the square-rooting function, $\sqrt{N}$.
Now, P/N ms represents the time resolution of the folded result, which has a corresponding noise frequency bandwidth of N/P kHz. So the larger bin numbers are associated with increased noise frequencies whilst the pulsar signal integration is unaffected and this is the mechanism that reduces the SNR. When $N = (P-W)/W$ the time resolution and bandwidth are just sufficient to pass the pulse so by restricting or filtering the noise bandwidth of large *N* folded results to this value, the matched SNR can be recovered.

Bin Folding with Filtering

Imagine again that the data record is divided into *R* periods and that these are added at the sample rate. In this case, the folded pulsar-to-noise SNR, increases by *R*/√*R* = √*R*. Passing this result through a low-pass filter sufficient to just pass the pulse, the filter cut-off frequency = 1/*W*, and the SNR becomes, $\frac{T_p}{T_{sys}}\sqrt{R}\sqrt{\frac{W}{t_c}}$. Now since, $\sqrt{t_{\text{int}}\Delta f} = \sqrt{RP/t_c}$, substituting for t_c, we get,

$$SNR_o = \frac{T_p}{T_{sys}}\sqrt{R}\sqrt{t_{\text{int}}\Delta f}\sqrt{\frac{W}{RP}}$$

The peak equivalent pulsar temperature is again, $T_p = \frac{S_p}{2}\frac{P-W}{W}.\frac{A_e}{k_b}$

Substituting for T_p we get the official radiometer equation form showing that the predicted SNR is no longer dependant upon the number of folding bins.

$$SNR_o = \frac{S_p A_e \sqrt{t_{\text{int}}\Delta f}}{2 k_b T_{sys}}\sqrt{\frac{P-W}{W}}$$

This last statement is not quite true, as for low bin numbers, the bin time window can get wider than the pulse, so the pulse contribution to the bin SNR reduces producing the low bin number tail off in SNR, as shown later by the green curve in Figure 3. The theoretically predicted curve with bin size taken into account is explored in the Appendix 1.

Matched Filtering Approach

Band limiting can be accomplished in a number of ways, but noting that the required filtering operation is equivalent to convolving the folded result with a similarly shaped clean pulse target, it can be accomplished digitally by the process,

1) Fourier transform the folded output.
2) Fourier transform a target pulse shape closely matching the pulsar pulse.
3) Determine the product of 1) and the normalized modulus of 2).
4) Inverse Fourier transform the product to produce the filtered result.

By this means, the signal-to-noise ratio of the resulting folded data is improved without affecting the delay shape or fidelity of the pulse. Improved pulse detail with the penalty of some SNR reduction is possible by increasing the filter cut-off by reducing the width of the pulse target. A MathCad code outline is given in Appendix 2.

Practical Folding Results

Vela Pulsar Example SNR = 55:1, P/W ≈ 43

Figure 3 compares the folding results as a function of number of folding bins for the basic folding method and the enhanced post-filtering scheme, using some Vela pulsar data collected by G. Gancio on the Argentine Institute's 30m dish. The plots indicate the SNR resulting as a function of the number of bins dividing the pulsar period. All integer bin numbers are represented from 10 to 800.

The red dots indicate the SNR for various period bin numbers using the basic folding method clearly showing the √*N* reduction in SNR. The blue dots indicate the enhanced performance with spectral band limiting on the same data set. Both methods demonstrate similarity as the number of bins reduces below the optimum value of *N* due to due to the data

bandwidth now being controlled by the bin number rather than the pulse bandwidth. The wide uncertainty occurs due to the pulse power being shared between adjacent bins; best performance being achieved when the pulsar position lies centrally in a bin.
The green and magenta curves are the predicted values assuming Gaussian statistics background noise and a Gaussian-shaped pulsar pulse of equal half-height to the expected pulsar, arriving at a bin center, as derived in the Appendix 1.

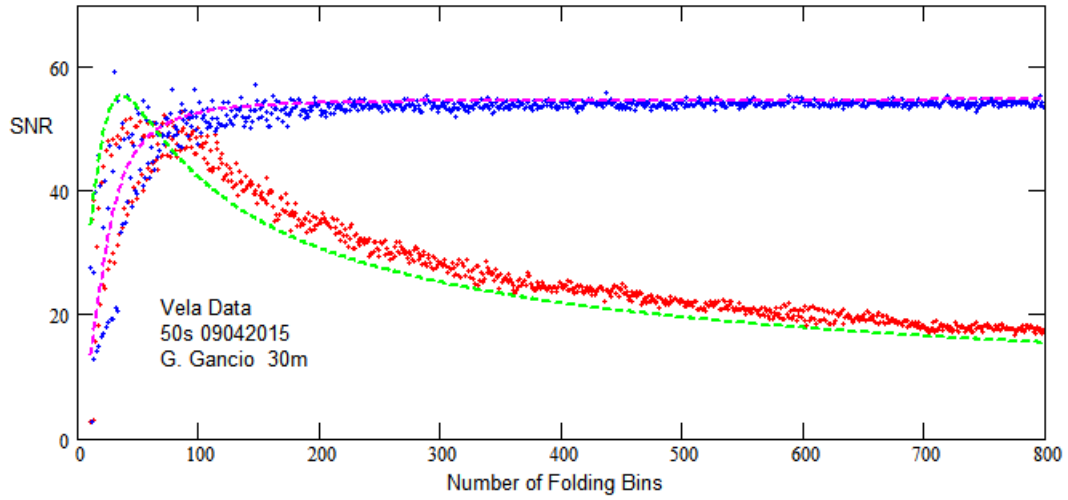


Figure 3. Vela Pulsar Data SNR as a Function of Number of Folding Bins (red - basic; blue - matched).

For the basic folding method, there is an evident peak SNR at around the bin value, 42 ≈ (*P-W*)/*W*) = (89.3-2.1)/2.1 - the Vela pulsar period, and half-height width parameters. The uncertainties at the higher bin numbers are due to the data quantization and the odd/even bin numbers varying the pulsar phase across bin boundaries. The plots highlight the limitations of the basic algorithm implementation. The simple approach only offers modest SNR performance over a small range of bin numbers between 40 and 100. Even within this range, care must be taken to ensure the pulsar phase is such that the peak lies centrally in a bin. Over this region, there is little difference between the folding methods since the bandwidth-limiting becomes dominated by the lower bin number averaging reducing the effective bandwidth (*N/P*). Very low bin numbers are to be avoided, due to the serious roll-off as the bin duration exceeds the pulse width allowing in more noise. Matched-filtering really comes into its own at large bin numbers as shown by the blue dots plot in Figure 3; above 250 bins in this example, the SNR uncertainty is minimized.
Figure 4 plots the relative offset error/period ratio indicating the position of the pulsar pulse peak within the period data block. The red plot has been offset slightly to improve visibility where there is data overlap.

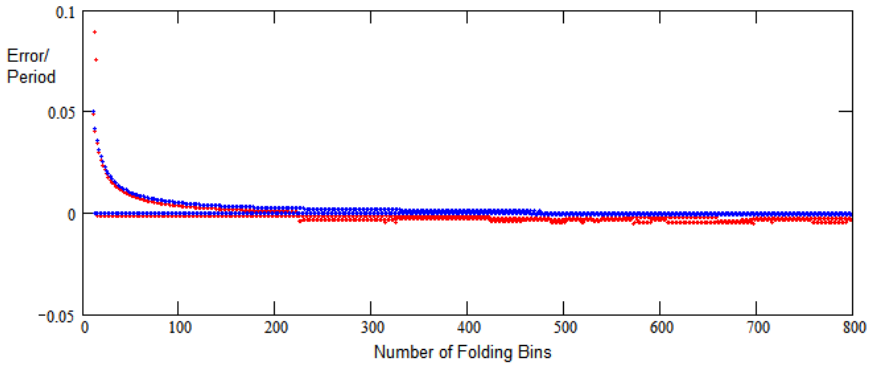
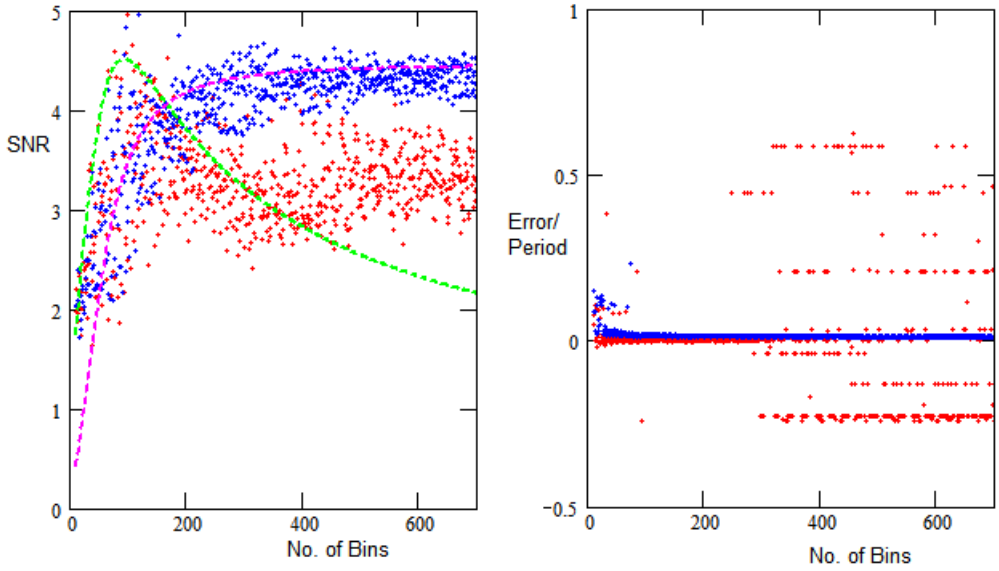


Figure 4. Vela Pulsar Pulse Bin Position Error v Number of Folding Bins, *N.*

The red, basic folding scheme shows increasing bin pulsar peak error uncertainty at higher bin numbers due to the lower SNR. Below about 150 bins, both schemes illustrate the effect of pulse energy being split between adjacent bins, mainly driven by odd/even bin number values. For example if the pulsar phase lies in the center of a period, it would be naturally either split by or fill a bin. No attempt was made to control the pulsar phase in this instance.

B0329 Pulsar Example SNR = 4.4, P/W ≈ 108

Figure 5(a) compares the two folding methods on much lower SNR data. The performance below 100 bins for both methods is again similar with the matched filter approach slightly superior. Here the 'optimum' bin number is ≈108. Above this bin number range the matched filter method dominates, but exhibits some wide SNR variation due to the increased noise level as a result of the low intrinsic SNR. Note that the basic method (red dots) does exhibit some tendency to respond to the predicted curve peak (green plot) at low bin numbers. Up to about 300 bins, the basic method median is showing some consistency with the expected *√N* SNR roll-off. This is confirmed in Figure 5(b) where the pulsar peak bin position error is very small. The matched-filter method shows much less peak uncertainty.



(a) (red - basic fold; blue - matched-filter fold) **(b)**

Figure 5. 4.4:1 SNR, B0329 Pulsar Data Result.

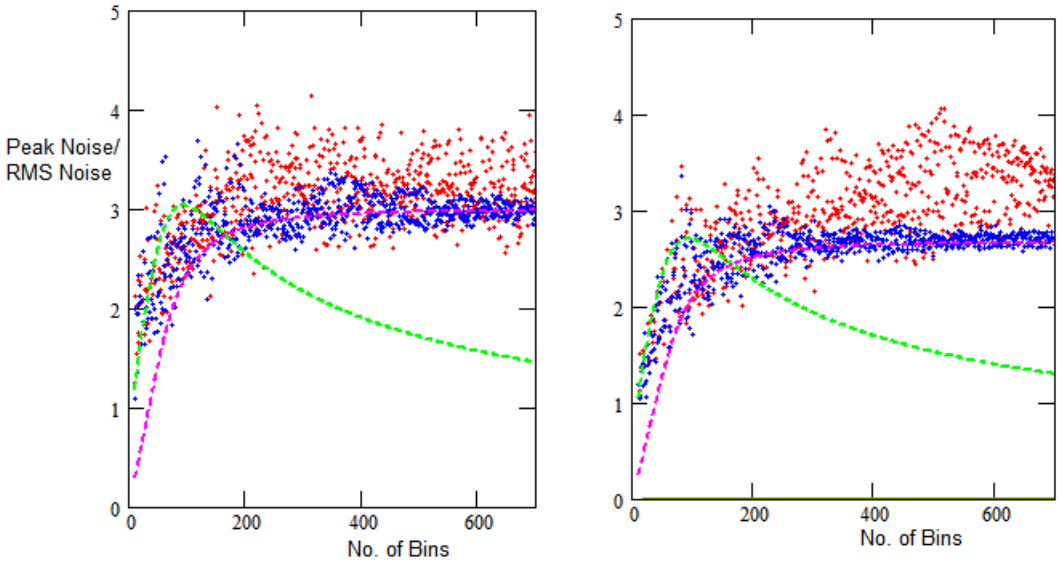
Above 300 bins however, the basic method red dots deviate from the theoretical curve and appear to maintain their dispersion range but level off in SNR at just over 3:1. The bin errors in Figure 5(b) indicate a number of consistent error values. Basically, what is happening here is the noise level is rising above the pulsar level and digitization noise peaks are capturing the peak detection function. Some actual folding plots are presented in Appendix 3 to support this conclusion.

Effects of RFI and Low SNR

In general the folding algorithm can cope quite well with low-level RFI, slightly raising the noise floor and similarly reducing the achievable SNR. Transient RFI and spectral lines should always be removed prior to folding.

In general, noise alone does not track the expected pulsar present theoretical curves. The algorithm response to just noise is investigated in Figure 6. This plots the recorded data but with the period adjusted such that the pulsar pulse is spread out, effectively reducing the pulsar amplitude below the background noise rms level without significantly changing the

folded noise statistics. Changing the B0329 search period by $4W/T \times 10^6$ parts per million, reduces the folded amplitude by a factor of 4; making it about equal to the noise rms level[2].



(a) 55:1 SNR Vela Data **(b)** 4.4:1 B0329 Data

Figure 6. Noise-only Folding (red - basic folds; blue - matched-filter folds).

It is evident from these entirely different data sets that there is a lot of similarity in the noise-only responses showing that the regular pulsar pulse has been effectively removed. What now is being indicated as SNR is actually, peak noise-to-rms noise ratio. Noise peak-to-rms noise ratio of between 2.5:1 to 3.0:1 is expected for Gaussian noise. Deviation from this range at low bin numbers, is probably due to the noise statistics varying from Gaussian to Binomial. Considering the data medians, there is little evidence of the data following the green curve or its predicted peak and there is no sign of the basic fold data (red dots) undergoing any reduction in SNR. Indeed over this region the SNR data is significantly different to the low real pulsar SNR plot of Figure 5(a). The reduced SNR uncertainty shown by the blue dot matched fold algorithm is due to the noise bandwidth reduction of the latter algorithm with increasing fold bin numbers and its more effective removal of quantization noise spikes (as evidenced by the increased noise spikes in the plots of Appendix 3).
The differences with pulsar present (Figure 5(a) - red dots) and absence (Figure 6(b) - red dots) for the basic algorithm are highlighted in Figure 7 (Figure 6(b) red dot noise-only data is now plotted as blue dots).

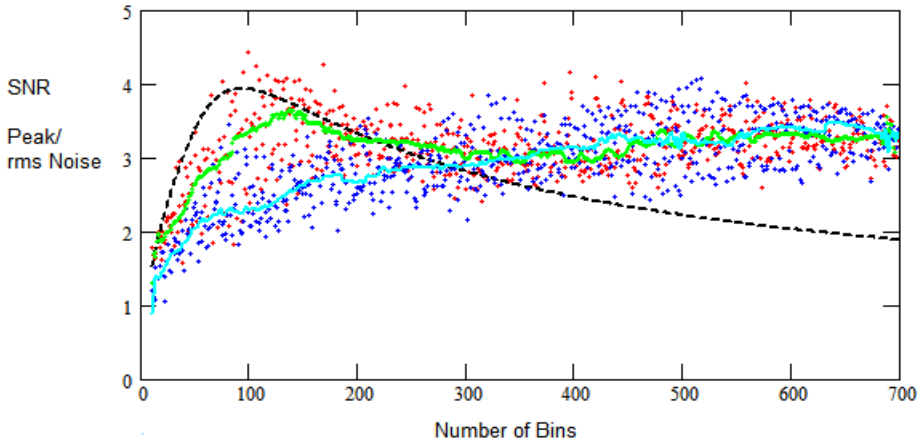


Figure 7. Basic Folding Only**:** red - 4.4 SNR B0329 pulsar; blue - noise; black- theory; green - pulsar data median; cyan - noise data median.

The red dots now show the actual SNR reported as a function of bin number with a 4.4:1 SNR pulsar pulse present and the blue dots peak noise-to-rms noise reported for the same data for noise only. As noted, to simulate noise only, the fold period is adjusted to spread the pulse out so reducing the pulsar amplitude below the background noise rms level. Figure 7 shows that above 300 bins the statistics of both data sets are very similar. But below this bin number the pulse-present median data is somewhat larger. The differences are highlighted in the green and cyan curves which plot the computed data medians of the two data sets. These obvious median differences which follow the predicted curve suggest it might be useful as an added indicator for recognizing that a pulsar pulse train is actually present in the data.

Conclusions

This article has examined the folding algorithm and shown that the limitations of the basic folding concept can be overcome by matched filtering. Matched filtering operates on the basic fold algorithm output to limit the fold data spectrum to just pass the integrated pulsar pulse spectrum. This procedure is implied by the published pulsar radiometer equation. A working folding SNR theoretical performance has been derived for both methods and are shown to prove reasonable matches to the two high and low SNR data examples tested.
The surprise bonus in doing this study was the unexpected capability of the basic folding method to show up the possible presence of a low level SNR pulsar, from plotting the data medians as seen in Figure 7. In addition, by indicating noise peaks in the correct range for Gaussian noise, it can be concluded that RFI is within acceptable limits.
Other articles on recognizing and validating low SNR pulsar detections can be found by following the link in Reference 3. Reference 4 links to the multiple bin folding software developed for this article.

References

1. D. Lorimer, M. Kramer. Handbook of Pulsar Astronomy. Cambridge University Press, 2005.
2. East PW. An Analytical Method of Recognizing Pulsars at Moderate SNR. Journal of the Society of Amateur Radio Astronomers. November-December 2018. p17.
3. Pulsar Astronomy Website. http://www.y1pwe.co.uk/RAProgs/Pulsars.html
4. Multifold Software. http://www.y1pwe.co.uk/RAProgs/MultiFold.zip

Appendix 1. Folding Integration Analysis

As shown previously, the indicated SNR is the ratio of the pulsar DC and the system noise AC terms, or

$$SNR = \frac{T_p}{T_{sys}} \cdot \sqrt{\frac{RP}{Nt_c}}$$

This equation is valid only when the pulsar pulse lies centrally in and fills a bin.
To take account of the pulse shape and effect on filling or not filling a bin, we need to integrate the samples within the bin where the pulse shape lies centrally.
If *p(t,W)* describes the pulse, width *W*, shape, then the SNR contribution of the pulsar power in the bin is calculated using the factor, $\frac{N}{P}\int_{-\frac{P}{2N}}^{\frac{P}{2N}} p(t,W)dt$ and the SNR for all bin numbers with no data filtering is given by,

$$SNR = \frac{T_p}{T_{sys}} \sqrt{\frac{RP}{Nt_c}} \frac{N}{P} \int_{-\frac{P}{2N}}^{\frac{P}{2N}} p(t,W)dt$$

As before, $1/t_c = \Delta f$, $t_{int} = PR$, this equation can be rewritten so that the predicted SNR for the basic folding algorithm now becomes...

$$SNR_{pred} = \frac{S_p A_e \sqrt{t_{\text{int}} \Delta f}}{2 k_b T_{sys}} \frac{P - W}{W} \frac{\sqrt{N}}{P} \int_{-\frac{P}{2N}}^{\frac{P}{2N}} p(t, W) dt$$

Assuming the pulse shape *pt,W*) takes a Gaussian form, $p(t, W) = e^{-4\ln(2)\left(\frac{t}{W}\right)^2}$, using this approximation, calculation results in the green SNR curve in Figure 3; the tail-off at higher bin numbers is as expected, with no extra filtering or integration save provided by the basic folding algorithm.

The analysis of the matched filtered version is similar and using the same scaling factor is plotted as the magenta SNR curve in Figure 3. This relationship is set for equal response when *N* = (*P-W)/W*.

The predicted SNR for the enhanced folding algorithm now becomes...

$$SNR_{pred} = \frac{S_p A_e \sqrt{t_{\text{int}} \Delta f}}{2 k_b T_{sys}} \sqrt{\frac{P - W}{W}} \frac{N}{P} \int_{-\frac{P}{2N}}^{\frac{P}{2N}} p(t, W) dt$$

Digital integration of the term $\frac{N}{P} \int_{-\frac{P}{2N}}^{\frac{P}{2n}} p(t, W) dt$, where *p*(*t,W*) is the preferred Gaussian function $p(t, W) = e^{-4\ln(2)\left(\frac{t}{W}\right)^2}$. Computation can be made faster and easier using a numerical approximation function. A useful substitution in the present instance is Bell's approximation,

$$\frac{N}{P} \int_{-\frac{P}{2N}}^{\frac{P}{2N}} p(t, W) dt \;\; -> \;\; \frac{\sqrt{\pi}\, 0.6 WN}{P} \sqrt{1 - e^{-\left(\frac{P}{\sqrt{\pi}\, 0.6\, WN}\right)^2}}$$

Appendix 2. Matched Fold Filtering

Figure A2 lists the MathCad code to apply matched filtering to the data to optimize detection of a pulsar in the final data fold. The displayed signal-to-noise ratio is then no longer dependent upon the number of fold bins.

$$\begin{aligned}
&\text{M_Fil}(P, W, \text{Fold_Dat}, N) := \\
&\quad \text{for } n \in 0..N-1 \\
&\quad\quad \text{dat}_n \leftarrow \text{Fold_Dat}_n \\
&\quad\quad \text{gau}_n \leftarrow \exp\left[-4\cdot\ln(2)\cdot\left(\frac{n - \frac{N}{2}}{W\cdot\frac{N}{P}}\right)^2\right] \\
&\quad \text{fdat} \leftarrow \text{cfft}(\text{dat}) \\
&\quad \text{fgau} \leftarrow \text{cfft}(\text{gau}) \\
&\quad \text{for } n \in 0..N-1 \\
&\quad\quad \text{mfdat}_n \leftarrow \text{fdat}_n \cdot \left|\frac{\text{fgau}_n}{\text{fgau}_0}\right| \\
&\quad \text{tdat} \leftarrow \text{icfft}(\text{mfdat}) \\
&\quad \text{Re}(\text{tdat})
\end{aligned}$$

Figure A2. Matched Pulse Filtering MathCad Code.

Appendix 3. B0329 Low SNR Folding Results

Figure A3 illustrates the folding results with various bin numbers. Red plots are the basic fold results and blue the matched filtered results. Note the increase in noise for the basic method and the gradual SNR improvement and increased resolution with matched filtering. Near identical performance is observed for 108 bins, as at and below this number the folded spectrum is controlled more and more by increased bin averaging. Data has not been bin-centered.

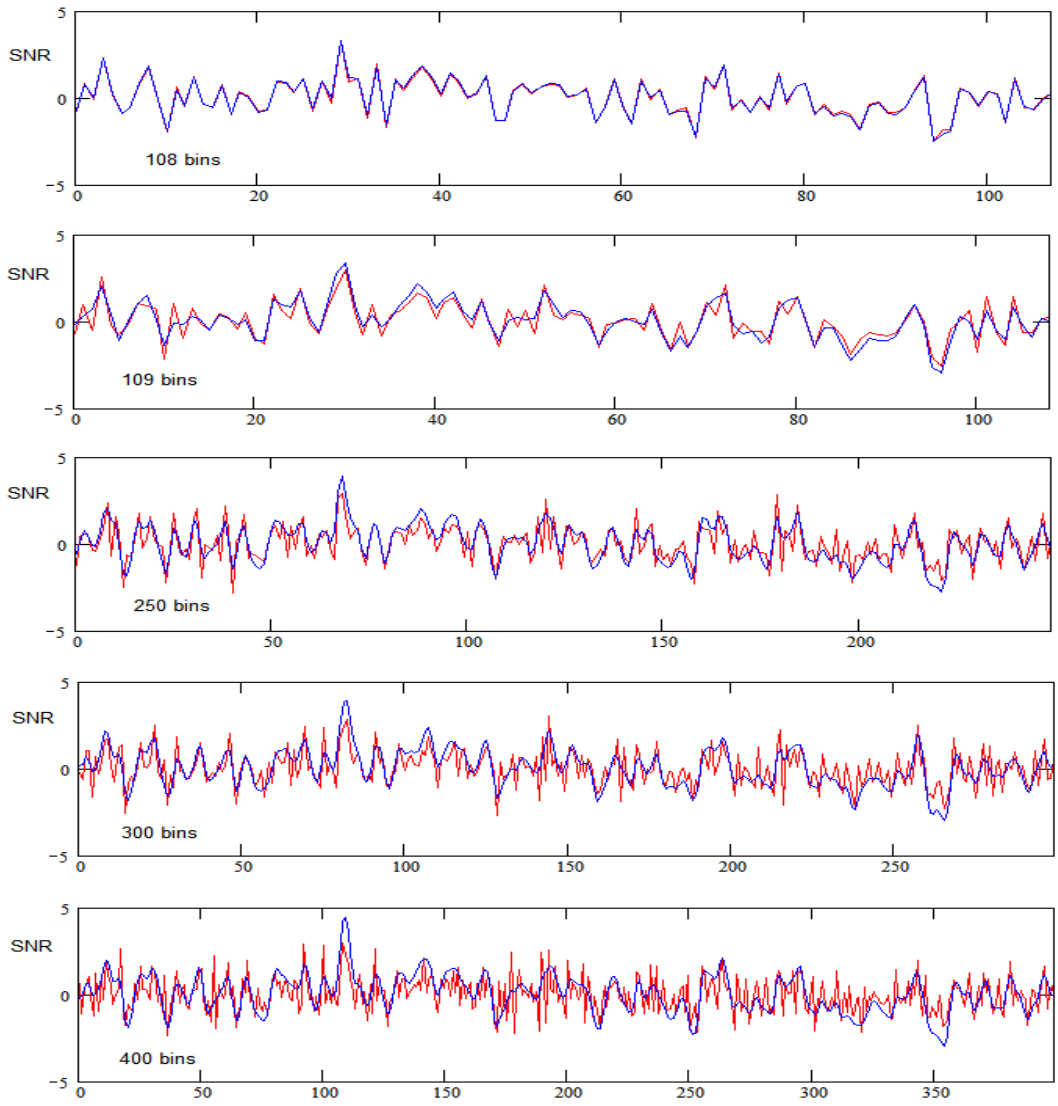


Figure A2. B0329 Comparison of Fold Methods Responses for Various Bin Numbers.
red - basic fold method; blue - matched folding method

Peter East, *pe@y1pwe.co.uk* is retired from a career in radar and electronic warfare system design. He has authored a book on Microwave System Design Tools, is a member of the British Astronomical Association since the early '70s and joined SARA in 2013. He has had a lifelong interest in radio astronomy; presently active in amateur detection of pulsars using SDRs, and researching low SNR pulsar recognition. He encourages free information exchange in the amateur community and is keen to widen interest in radio astronomy generally.

He maintains an active RA website at *http://www.y1pwe.co.uk*

P W East August 2020