

Towards Real-Time Impulsive RFI Mitigation for Radio Telescopes

Kaushal D. Buch*, Shruti Bhatporia, Yashwant Gupta, Swapnil Nalawade, Aditya Chowdhury,
Kishor Naik, Kshitij Aggarwal and B. Ajithkumar

Giant Metrewave Radio Telescope, National Centre for Radio Astrophysics
TIFR, Pune, Maharashtra 411007, India

*kdbuch@gmrt.ncra.tifr.res.in

Received 2016 August 15; Accepted 2016 November 29; Published 2017 January 13

Radio Frequency Interference (RFI) is a growing concern for contemporary radio telescopes. This paper describes techniques for real-time threshold-based detection and filtering of broadband and narrowband RFI for the correlator and beamformer chains of a telescope back-end, with specific applications to the upgraded Giant Meterwave Radio Telescope (uGMRT). The Median Absolute Deviation (MAD) estimator is used for robust estimation of dispersion of the received signal in temporal and spectral domains. Results from the tests carried out for the GMRT wide-band backend (GWB) using this technique show 10 dB improvement in the signal-to-noise ratio. MAD-based estimation and filtering was also found to be useful for filtering beamformer data. The RFI filtering technique demonstrated in this paper will find applications in other radio telescopes as well as receivers for digital communication and passive radiometry.

Keywords: RFI, median absolute deviation, radio telescope, GMRT.

1. Introduction

Radio telescopes are wide-band radio receivers used to detect and analyze faint radio emissions from celestial sources. Typical radio telescope receivers have almost 50 dB higher sensitivity than their terrestrial-communication counterparts. This increases their susceptibility towards man-made Radio Frequency Interference (RFI) which impairs the detection of weak radio sources and transient events (Edwards, 2014). Thus, mitigation of RFI is an important aspect in the design and operation of radio telescopes (Dewdney *et al.*, 2009; Ellingson, 2004). In terms of its frequency-domain characteristics, RFI be categorised as broadband or narrowband. In the time-domain, broadband RFI manifests as short duration, impulsive bursts of interfering signal, typically from spark-like events (e.g. lightning discharges, power line discharges, spark ignition discharges etc); whereas narrowband RFI is due to long duration spectrally confined

sources such as transmissions from communication systems — terrestrial or otherwise.

A radio telescope receiver gets a combination of three signals at its input — system noise, sky noise and RFI. System noise is a combination of sky background noise and receiver noise, both of which are zero-mean uncorrelated Gaussian distributed random signals. In general, RFI has a non-Gaussian distribution (Fridman & Baan, 2001) and gets added to the overall receiver noise. Thus, a generalized time-domain model for a signal received by a radio telescope in the presence of RFI can be given as (Fridman & Baan, 2001),

$$x(t) = s(t) + n(t) + r(t), \quad (1)$$

where, $s(t)$ is the contribution from the astronomical source, $n(t)$ is the system noise and $r(t)$ is the RFI. The distribution of signal with impulsive RFI is heavy-tailed Gaussian distribution (Fridman, 2008). Sources of broadband RFI are sparking and corona discharges on high-power transmission lines which occur at submultiples of the power-line frequency (Swarup, 2008) with a time duration

* Corresponding author.

ranging from 5 ns to 200 ns (Langat, 2011). The main sources of narrowband RFI are emissions from TV broadcast transmitters, cellphone base station transmitters, mobile communication devices and satellites.

A class of RFI mitigation techniques, known as RFI excision, is used for mitigating impulsive time-domain or frequency-domain RFI by determining signal dispersion and robust thresholding (ITU-R, 2013). Moment-based estimation using Standard Deviation (STD) and Kurtosis has also been used for RFI excision in radio telescopes (ITU-R, 2013), (Niamsuwan et al., 2004) and passive microwave radiometers (Guner et al., 2007). For RFI having a heavy-tailed Gaussian distribution, Median Absolute Deviation (MAD) has been proposed as a robust rank-based statistical estimator (Fridman, 2008). MAD-based excision is also shown to be effective in radio astronomy applications (Fridman, 2008; Shankar & Pandey, 2006) as well as in other signal processing areas like time-series analysis and image processing (Crnojevic et al., 2004). However, the authors have not come across any account of MAD-based real-time implementation in a wide-band radio telescope backend.

This paper describes techniques for real-time broadband and narrowband RFI detection and excision (filtering) using MAD estimator. The detection of RFI is followed by a nonlinear filtering operation wherein the RFI is replaced by a constant value or sample from noise source. An optimized implementation is developed which enables resource-efficient real-time filtering of RFI in a wide-band system. The effect of real-time RFI filtering on the sensitivity and signal correlation are analyzed for different configurations of a radio telescope backend. RFI filtering using the proposed technique has been implemented in real-time and tested with the wide-band receiver chain at the Giant Meter-wave Radio Telescope (GMRT) (Swarup et al., 1991). The broadband and narrowband RFI filtering results show an improvement of 10 dB in signal-to-noise ratio and a similar improvement in the cross-correlation power spectrum, thus improving the receiver sensitivity.

2. Radio Telescope Signal Processing Backend

A radio telescope consists of either a single antenna or an array of antennas with a sensitive radio receiver(s) for processing weak astronomical signal.

A typical digital backend of a radio telescope receiver processes baseband signals downconverted from radio frequency signals obtained from individual antenna elements. The primary objective of an array radio telescope receiver is to compute correlations between pairs of antennas also known as visibilities. Towards that end, in a typical digital backend of a radio telescope (Fig. 1), the signals are time-aligned through delay correction at sample and subsample level. After Fourier transformation, the auto-and cross-correlation spectra are generated. For N antenna inputs, the MAC block produces $N(N - 1)/2$ time-integrated cross-correlation and N auto-correlation spectra. Alternatively, received signals may be combined to produce single or multiple beam outputs. The signals in the beamformer can either be combined after phase alignment to produce Phased-Array (PA) beam or without phase alignment to produce Incoherent Array (IA) beam. The temporally-integrated signal from the correlator and beamformer signal is recorded for further offline processing. The beamformer mode is generally used for pulsar observations. To get the final pulsar profile, the beam output undergoes the process of de-dispersion and folding. De-dispersion is carried out to remove the temporal smearing that the pulse undergoes while it passes through the inter-stellar medium. Folding is used to improve the signal-to-noise ratio by adding successive pulses.

2.1. GMRT Wide-band Backend

The GMRT Wide-band Backend (GWB) is a wide-band digital signal processing system (Ajithkumar et al., 2013) being developed as part of the Upgraded GMRT (uGMRT). The real-time processing bandwidth of 400 MHz is achieved using a combination of contemporary Field Programmable Gate Array (FPGA) and Graphics Processing Unit (GPU) (Ajithkumar et al., 2013). GWB is built using ROACH-1 boards^(a) which digitize the data and provide it through 10 GbE links to the GPU cluster. The data from high-speed 8-bit, 1 Gbps ADC (e2V AT84AD001B) is provided to ROACH-1 board which consists of Xilinx Virtex-5 (XC5VSX95T-1FF1136) FPGA. Nvidia K20 GPUs are used for real-time signal processing in the GWB and are hosted by server-class machines having multi-core CPUs. The number of FFT channels supported by the GWB range from 2,048 to 16,384.

^a<https://casper.berkeley.edu/wiki/ROACH>.

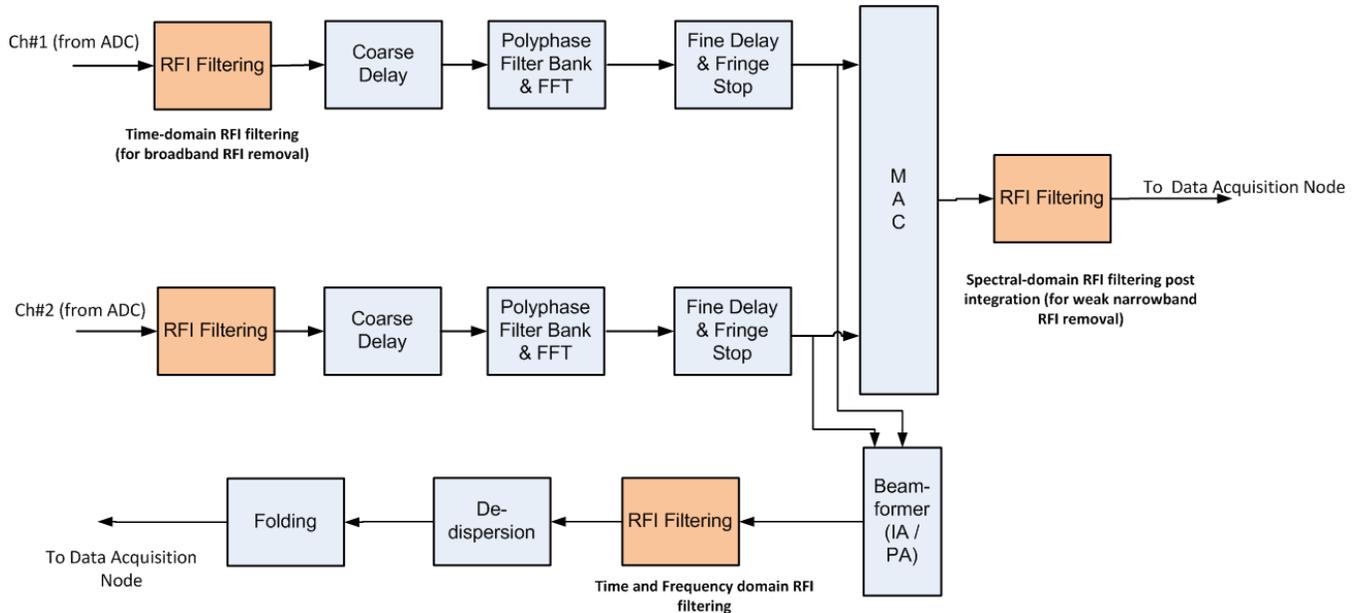


Fig. 1. Block diagram of the digital backend for the uGMRT, illustrating the points where real-time RFI filtering can be carried out.

GWB supports a correlator output and two simultaneous beam outputs for a 32-antenna dual-polarization input. The data processed by the GPU cluster is sent to three recording nodes, one for correlator and one each for the two beam outputs. The shortest integration times for GWB system are 671 ms and 1.3 ms for the correlator mode and the beamformer mode, respectively.

Real-time broadband RFI filtering is carried out on Nyquist-sampled time-series, on FPGA. Correlation and beamforming operations in the GWB are carried out in the frequency-domain on spectral channels having a 100 kHz resolution. Narrowband RFI filtering is carried out (post-Fourier transformation) on the CPU-GPU cluster. A data acquisition system receives the processed data from this cluster and stores it for offline analysis. In order to avoid potential clipping of useful astronomical data, RFI filtering is carried out before the data undergoes the process of de-dispersion and folding.

3. Robust Signal Processing Techniques for Real-Time RFI Excision

A conventional measure of the signal dispersion is its STD. In case of a radio telescope signal affected by strong RFI, the STD estimator is biased by the RFI. Median-based estimators provide a robust estimate of the dispersion (scale parameter) in presence of RFI (Fridman, 2008). One such estimator is

MAD which has a breakdown point of 50%, which indicates that its estimate of the dispersion is unaffected till the number of outliers in a data set is less than 50%. Here, MAD is proposed for estimating signal dispersion for symmetric Gaussian distributed data with heavy tails. When the underlying distribution is Gaussian, the robust STD can be calculated from the MAD value using a scaling factor of 1.4826 (Rousseeuw & Croux, 1993). We extend the process to compute robust threshold around the median of the distribution which is followed by nonlinear filtering of RFI. This type of filtering is also called decision-based nonlinear filtering (Astola & Kuosmanen, 1997) or Hampel filtering (Pearson *et al.*, 2016).

3.1. MAD-based RFI Excision

Real-time MAD-based RFI detection and excision proposed in this paper is explained through the following steps:

- (1) Consider a window of W input data samples $X = [x_1 x_2 \dots x_W]^T$. Let $M(\cdot)$ represent the median operator such that the median of the input X is $M(X) = \mathcal{M}$. Now, let $Y = [y_1 y_2 \dots y_W]^T$, where $Y_i = x_i - \mathcal{M} \quad \forall i \in 1, 2, \dots, W$ which is median subtracted from the input data and the difference converted to absolute values. MAD is obtained by performing median operation on Y , i.e. $\text{MAD} = M(Y)$. The

resultant MAD value is scaled by a factor of 1.4826 (for Gaussian distribution) to get robust STD (σ_R).

- (2) Assuming Gaussian distributed data at the input, the detection thresholds are computed as multiples of σ_R on either side of the median value \mathcal{M} . The thresholds are computed on consecutive windows of W samples. The upper and lower thresholds are denoted by τ_U and τ_L , respectively, and are

$$\begin{aligned}\tau_U &= M(X) + n_U \times (1.4826 \times M(Y)), \\ \tau_L &= M(X) - n_L \times (1.4826 \times M(Y)),\end{aligned}\quad (2)$$

where n_U and n_L are the multiplying factors for the upper and lower thresholds, respectively. If the distribution is symmetric, $n_U = n_L$.

- (3) A comparison between the input data and threshold values provides a one-bit detection flag where ‘1’ indicates the detection of RFI and ‘0’ indicates otherwise.
- (4) The samples detected as RFI can be either replaced by the threshold value (clipping) or by a user-defined value. The choice of replacement value can be modified in real-time.

In case of clipping, the RFI detected samples are held at their respective upper and lower threshold values (τ_U and τ_L). Alternatively, the values can be replaced by the median, a constant value, or noise sample (having statistical properties like mean and STD similar to the underlying distribution as estimated using median and MAD). The constant value could be any number within the range of numbers supported by the bit-precision. For example, in case of 8-bit signed number system, the constant value can be any integer between -128 and $+127$. In Eq. (3), the element z_i in the output vector $z = [z_1 z_2 \dots z_W]^T$ is replaced by K if the input is outside the robust threshold.

$$\begin{aligned}z_i &= K \quad \text{for } x_i \geq \tau_U \text{ or } x_i \leq \tau_L \\ &= x_i \quad \text{for } \tau_L < x_i < \tau_U.\end{aligned}\quad (3)$$

The window size for MAD estimation is determined by the worst-case duration of RFI and the sampling rate of the system. In order to get an unbiased estimate using MAD, less than 50% of the samples in a data window should be affected by RFI. The window size thus depends on the ratio of the duration of RFI (T_R) to the sampling interval (T_S), i.e. $2(T_R/T_S)$ samples. The factor of 2 is due to the 50%

breakdown point of the MAD estimator. For the technique described in this paper, the window size remains constant once the above parameters are computed and the design is implemented.

The values of n_U and n_L in Eq. (2) are chosen in such a way that the astronomical data does not get detected and filtered as RFI. If these values are lower, the filtering operation would remove much of the useful data and if they are too high, most of the data would pass through without filtering. We find that for most of the receiver system parameters and RFI conditions, values of n_U and n_L around 3 are optimal. However, this is a tunable parameter and in the implementation described in this paper, it can be changed on-the-fly to account for the dynamic changes.

3.2. Performance of MAD-based detection

The performance of MAD-based detection is evaluated through simulation. The MAD estimator was subjected to varying levels of impulsive RFI. For this purpose, impulsive RFI above 3σ was generated and added to samples of noise having zero mean and fixed STD. RFI was added to the noise at random instances in time in a block of 1,024 samples. As the percentage of impulsive RFI increases above 50%, the MAD estimator gets biased. The bias is positive due to larger dispersion and hence the obtained threshold is higher than the expected threshold. This causes the filter to not detect (under-filter) the RFI samples. Hence, the difference of total number of RFI samples and detected samples indicates the efficiency of the filter. These values are computed for different fractions of impulsive RFI in the noise sample and is an average of 100 blocks each having 1,024 samples. The difference values range from 0 indicating all RFI samples are successfully detected to 1,024 which means that none of the RFI samples are detected.

The behavior of MAD-based threshold and the difference for different values of n_U and n_L in comparison with the conventional STD estimator-based threshold (3σ) is shown in Fig. 2. STD estimator gets biased even for a small percentage of impulsive RFI. On the other hand, MAD estimator performs reasonably well up to 50% impulsive RFI. There is a steep rise in the difference value for more than 50% impulsive RFI because of the breakdown point of MAD estimator. The best performance in this case is observed for 3σ detection threshold because, in

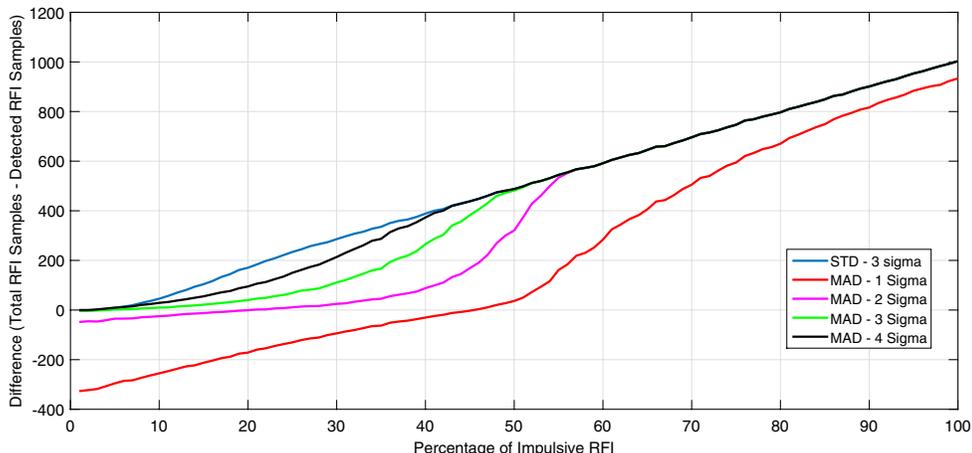


Fig. 2. Simulation showing performance of MAD-based impulsive RFI detection.

this simulation, the samples of impulsive RFI are added above 3σ value of the underlying distribution of noise.

The number of difference values for n_U and n_L equal to 4 increases gradually with the increase in percentage of impulsive RFI. This shows that as the threshold increases, there are more number of samples which are missed by the detector. The missed samples are the ones between 3σ and 4σ . For values of n_U and n_L equal to 2, the samples which are not RFI are also detected leading to a negative value of the difference. This simulation shows that MAD-based detection performs better in terms of filtering impulsive RFI as compared to the STD-based estimator. Also, the values of n_U and n_L in the detection threshold needs to be chosen based on the strength of RFI and the system parameters, to achieve optimal filtering.

3.3. Quantitative metric for analyzing RFI filtering

The performance of RFI filtering is quantified by a ratio of mean-to-the Root Mean Square (RMS) value of the signal. This ratio is a normalized quantity and is robust to system gain fluctuations in the RF and analog receiver subsystems. The mean-to-RMS ratio (μ/σ) for a radio telescope receiver is given as $\sqrt{B \times T}$, where μ , σ , B , T are the sample mean, sample RMS, bandwidth and integration time of the system, respectively.

The effect of RFI on mean-to-RMS ratio, comparison with the theoretical limit and its use as quantitative metric is illustrated through a time-series plot of a single spectral channel in Fig. 3

corresponding to 651 MHz radio frequency (RF) (200 MHz bandwidth, 2,048 spectral channels and 1.31 ms integration time) having instances of broadband RFI. The filtering is carried out in time-domain on Nyquist-sampled digital time-series at 3σ threshold and the detected samples are replaced by zero. The results can be compared with the theoretical limits that can be achieved for a Gaussian random signal without RFI, which, for a single spectral channel and the system specifications used in Fig. 3 can be given by $\sqrt{(B \times T)}$ which is $((200 \times 10^6)/2,048) \times (1.31 \times 10^{-3})^{0.5} = 11.31$. Mean-to-RMS ratio is computed for 2,000 samples around regions 1 and 2. It can be seen that the mean-to-RMS ratio reduces in case of strong instances of RFI. The mean-to-RMS ratio around region 1, before and after filtering, is 3.27 and 11.06, respectively, and that of noise (unfiltered) around region 2 is 10.99.

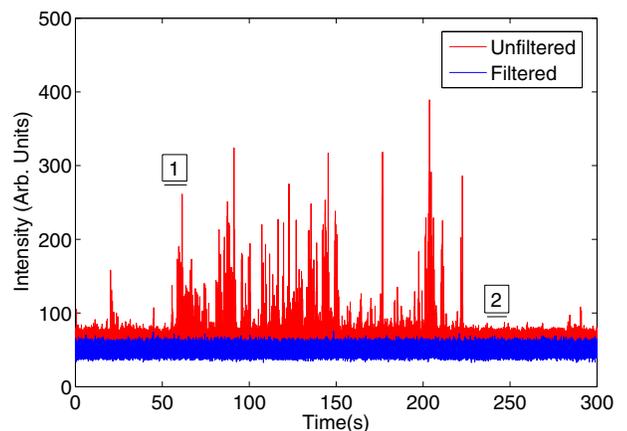


Fig. 3. Mean-to-RMS ratio for broadband RFI filtering.

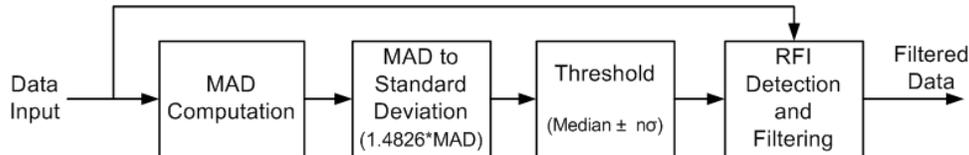


Fig. 4. Block diagram of MAD-based RFI filtering.

The improvement metric (I) in dB derived from the mean-to-RMS ratio is

$$I = 10 \log (MR_F / MR_U) \text{ dB}, \quad (4)$$

where MR_U and MR_F are the mean-to-RMS ratio for the unfiltered and filtered signal, respectively. If mean-to-RMS for filtered and unfiltered is the same, the improvement is 0 dB. If the mean-to-RMS of the filtered signal is higher than that of the unfiltered signal, the improvement (in dB) is positive.

4. Broadband RFI Filtering in the GWB

Real-time broadband RFI filtering is carried out on FPGA for 60 inputs (30 antennas, dual-polarization) from the GMRT antennas, each having 400 MHz bandwidth. For processing Nyquist-sampled time-series at 8-bit precision, in real-time, hardware parallelism available on the FPGA is utilized. Figure 4 shows the block diagram of MAD-based RFI filtering.

Here, the first step is to compute MAD, which requires two successive median computations. Each median computation requires sorting the input data set which is a resource intensive operation. Several optimized algorithms with varying computational complexities are available for median computation. Their performance is governed by the input data precision and the window size (Juhola et al., 1991). A few examples of optimized sorting algorithms are quick sort, heap sort and k -selection method. As this application requires a moderate window size and operates on fixed point data, it was found that the histogram-based approach is computationally optimal (Juhola et al., 1991). Also, the histogram method of median calculation is amenable to real-time implementation (Fahmy et al., 2009). We optimize the histogram method to meet the real-time requirements within the constraints of available hardware resources on Xilinx Virtex-5 SX95T FPGA. The design^(b) is developed and tested on

ROACH-1 board and is available as open-source module under the Collaboration for Astronomical Signal Processing and Electronics Research (CASPER)^(c) library.

4.1. Real-time architecture

An optimized implementation enables median computation of W samples following an initial latency of W clock cycles. There are two median computation blocks which operate simultaneously in order to compute MAD on contiguous data windows. To begin with, the binning of data is carried out to get a cumulative histogram. As part of the optimized implementation, this cumulative histogram is not stored, instead, a median value search is carried out. This is done by locating the data value for which the cumulative frequency crosses half the total number of samples in the window. The median so obtained is subtracted from the input data and the difference is converted to absolute values which are then used by the second median computation block to get the MAD value. Both the median computation blocks are architecturally similar except that the first one operates on a range of N -bit signed data, i.e. $-2^{\frac{N}{2}}$ to $2^{\frac{N}{2}} - 1$, whereas the second one operates on N -bit unsigned data, i.e. 0 to $2^N - 1$.

The median and MAD values are then used to compute the robust threshold as per Eq. (2). Subsequently, the input data is compared with the threshold. The detection flag is set to 1 if the sample is outside the threshold. The filtering operation uses a data multiplexer to replace the flagged sample with one of the replacement options chosen by the user. For replacement by noise, samples having standard normal distribution are generated on FPGA and shifted and scaled to conform to the mean and STD obtained from the median (\mathcal{M}) and robust STD (σ_R), respectively. An FPGA-based digital noise source is used to generate Gaussian

^bhttps://casper.berkeley.edu/wiki/Impulsive_RFI_Excision:_CASPER_Library_Block.

^c<https://casper.berkeley.edu/>.

distributed noise samples noise with programmable mean and STD. The detailed architecture of the digital noise source and its properties are described in [Buch et al. \(2014\)](#).

4.2. Results from antenna and emulator tests

The effect of RFI filtering carried out in the time-domain is observed by recording data in the beamformer and correlator mode of the GWB simultaneously. Since broadband RFI affects all the spectral channels equally, the analysis is carried out on a single spectral channel output for a pair of antennas. Three copies of each signal are created in order to get a simultaneous comparative analysis of the two filtered and one unfiltered output. Two different replacement options are applied for both the antennas and one copy from each antenna passes without filtering. Detection and filtering are carried out on consecutive windows in time-domain before computing the Fourier transform. The sampling rate is 800 MHz and each window contains 4,096 data samples with 8-bit fixed-point precision. The processed data has an integration time of 1.3 ms and 671 ms for the beamformer and correlator mode, respectively. This setup is common to all the three examples mentioned in this section.

4.2.1. Tests using antenna signals

The temporal behavior of a single spectral channel with unfiltered and filtered data corresponding to 636 MHz radio frequency is shown in [Fig. 5](#). In this experimental setup, two antennas, say, A1 and A2 are connected to the input of the GWB system and observe the astronomical source 3C147. The results show a comparative analysis between unfiltered and filtered data processed simultaneously through the GWB. The first subplot is the beamformer mode intensity data for filtered (blue) and unfiltered (red) options. The X-axis shows the time at which the data was recorded. RFI is filtered at 3σ threshold in time-domain and the detected samples are replaced by *zero*. The negative values of intensity are caused by saturation of beam output indicating the presence of strong RFI. The second subplot shows improvement in the mean-to-RMS ratio which is calculated as shown in [Sec. 3.3](#) over 1,024 samples of the beamformer data. A maximum of 10 dB improvement is observed in this case.

The third and fourth subplots are time aligned with the first subplot and show amplitude and phase of the cross-correlation for antenna A2 with respect to antenna A1. Generally, for short baselines, the cross-correlation amplitude increases in the event of correlated RFI (shown in red). RFI filtering reduces the amplitude of the cross-correlation towards its desired values. The phase of the cross-correlation fluctuates in the event of RFI as seen in the fourth subplot. Under normal conditions, the phase for a point source should not have large fluctuations especially when observed for a short duration. The phase of the filtered signal is maintained at the same value as it was in the absence of RFI. The effect of replacement with *zero* (shown in blue) and *threshold* (shown in green) value are overlaid in the third and fourth subplots. In this case, replacement with *zero* provides better performance over replacement by *threshold*.

Another example of filtering is shown in [Fig. 6](#). In this case, moderate level of RFI is persistent for a longer duration. It is not as strong and distinct as it was in the previous example. Filtering is carried out at 3σ threshold and replacement options are *threshold* (shown in green) and *noise* (shown in magenta). Subplot 2 shows uniform improvement in mean-to-RMS ratio by 5 dB. Improvement is also observed in the magnitude and phase of the cross-correlation outputs (subplots 3 and 4).

4.2.2. Tests using RFI emulator

This is a controlled test carried out to probe the behavior of the RFI filtering technique for a known input signal. RFI emulator is an analog instrument developed at the GMRT to generate RFI with desired duration and power level in order to test the filtering system. It generates a broadband noise floor and adds periodic broadband RFI signal or pulsed noise with a specific duty cycle and having average power greater than the noise floor. The duty cycle is programmable. The output from the emulator can be fed as baseband signal to the GWB system. For the example shown in [Fig. 7](#), the *on* period of the RFI is $16\ \mu\text{s}$ and *total* period is 2 s. The tests have been carried out with the GWB and RFI filtering configuration similar to those for the antenna signals described earlier in this section. [Figure 7](#) shows the effect of filtering on the signal generated using emulator. As can be seen in the first subplot, the unfiltered time-series shows periodic occurrences of RFI (pulsed-noise) which are filtered out at 3σ

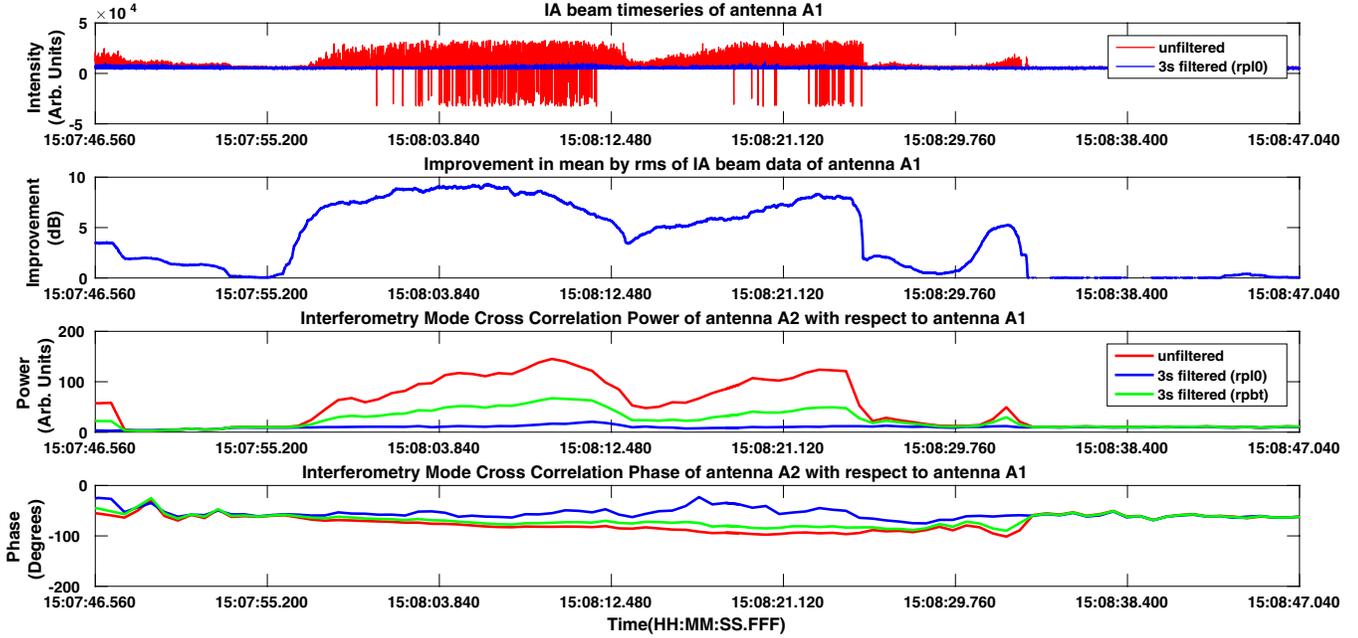


Fig. 5. RFI filtering in GWB correlator and beamformer mode outputs simultaneously for antenna signals.

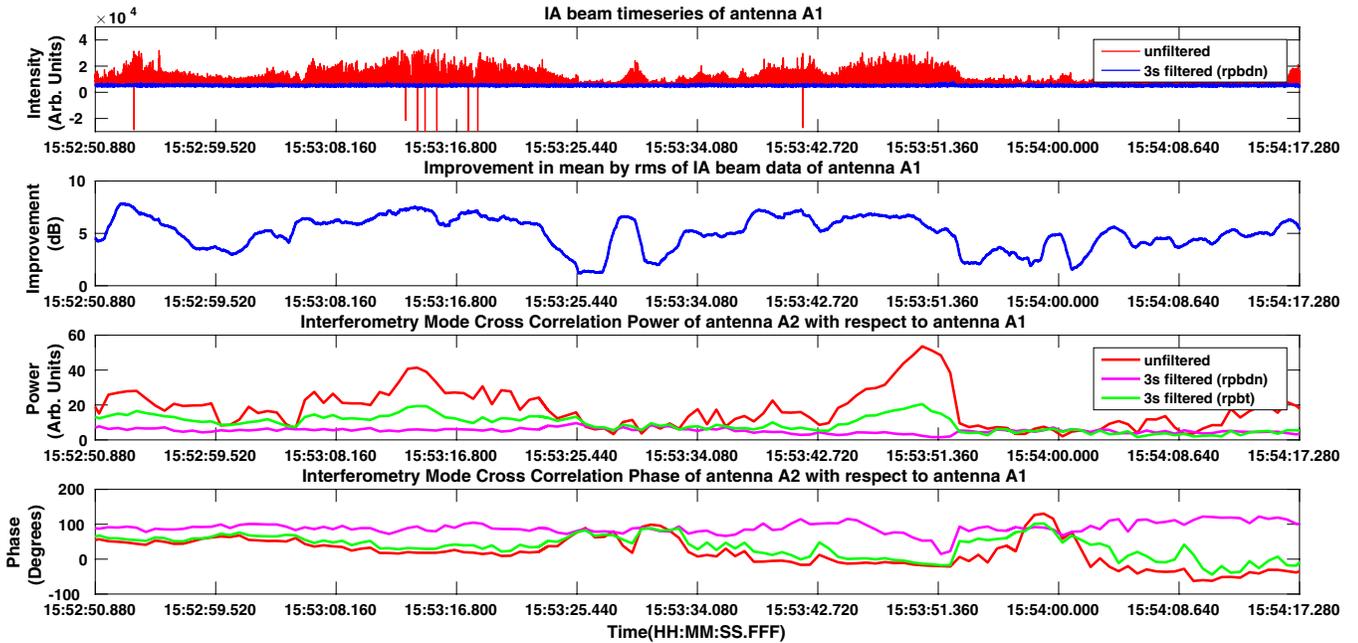


Fig. 6. RFI filtering simultaneously in GWB correlator and beamformer mode outputs for antenna signals.

threshold and replaced by *zero*. The corresponding improvement (in dB) is shown in the second subplot. Since the output time resolution is 671 ms which is much larger compared to the ON-time of RFI, and since the emulator has a low duty cycle, the power gets averaged at the correlator output. Thus, the third and fourth subplots show an

averaged effect in the magnitude and phase of the cross-correlation. The average value of cross-correlation magnitude increases due to addition of pulsed-noise (RFI). Filtering at 3σ and replacement by *zero* brings this value to the actual value of cross-correlation magnitude which is equal to that observed in absence of RFI.

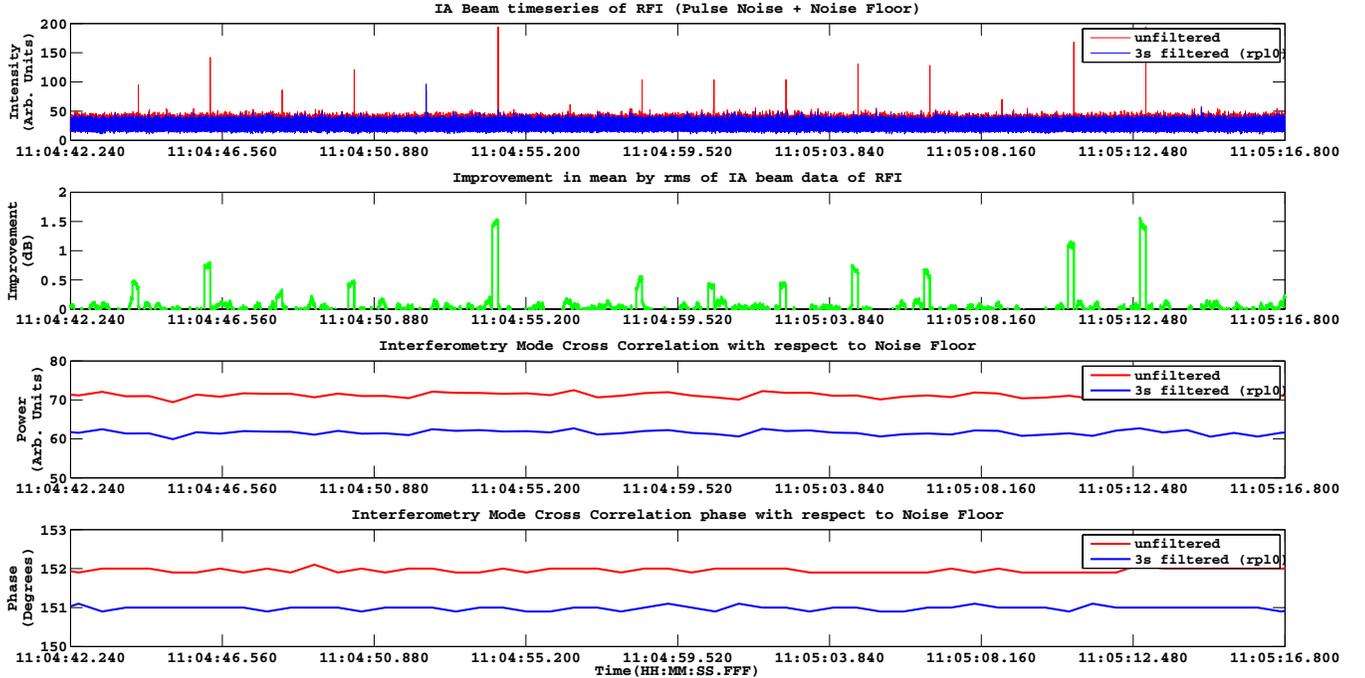


Fig. 7. RFI filtering simultaneously in GWB correlator and beamformer mode outputs for RFI emulator signals.

5. Narrowband RFI Filtering in the GWB

Detection of narrowband RFI in the GWB is carried out in the post-FFT domain (Fig. 1) using MAD-based estimation (of dispersion) across the spectrum. Based on the relative strength of the narrowband RFI as compared to the total receiver noise, the detection can be carried out either on the real and imaginary parts of the spectrum or on temporally integrated spectrum. For example, narrowband RFI buried within the system noise can be detected only after temporal integration of consecutive spectra. The filtering threshold is calculated by computing median and MAD across spectral channels for each spectrum. As a proof-of-concept, this scheme is implemented on recorded data, however, a real-time implementation is being developed.

5.1. Spectral normalization

If there are large variations in the power level across the frequency band, the median and MAD estimators for computing the filtering threshold get affected. The estimation is better when there is equal power in all the spectral channels. Hence, a compensation scheme equalizes the power level across the spectral channels before computing the robust threshold. Of the various schemes for spectral

equalization, we propose a technique where the entire band is normalized to unity. This highlights the locations in the spectrum that are corrupted by RFI making them easier to identify. The normalization process uses median-based estimator as the antenna signal may have RFI. The steps involved in normalizing spectrum are:

- (i) For the m th spectral channel, $m \in 1, 2, \dots, N_{ch}$, the median spectrum (\mathcal{M}_m) is obtained by taking median (M) of over a range of time-stamps $T = t_0, t_1, \dots, t_n$. Thus,

$$\mathcal{M}_m = M(S_{m,T}). \quad (5)$$

- (ii) This median spectrum would still contain temporally persistent RFI. Thus, a running median filter is applied across the median spectrum. This step smoothens the median spectrum and reduces any spectrally impulsive RFI present in it. The value of each sample in the median spectrum is replaced by the median of the values in the window around that sample. This spectrum can be called smoothed spectrum (\mathcal{M}_S) as it is impulse free. This spectrum can now be used for the normalization process. For a median filter having a window of $2w + 1$ spectral channels, the r th channel of the smoothed median spectrum can

be calculated as

$$\mathcal{M}_S(r) = M(\mathcal{M}_S(r-w), \dots, \mathcal{M}_S(r), \dots, \mathcal{M}_S(r+w)). \quad (6)$$

- (iii) Steps (i) and (ii) are used to get a spectrum for the normalization process even in the presence of RFI. During the filtering process, median and MAD estimates are computed for each individual spectrum. The unfiltered spectrum is divided by the smooth spectrum to get normalised data. For the i th spectral channel, $\mathcal{S}(i)$ is calculated from incoming spectrum $S(i)$ for each time-stamp t as

$$\mathcal{S}(i, t) = \frac{S(i, t)}{\mathcal{M}_S(i)}. \quad (7)$$

Filtering is carried out on this spectrum as shown in Sec. 3.1. It has to be ensured in the design that the values of \mathcal{M}_S do not become zero or vanishingly small.

5.2. Narrowband RFI filtering on recorded data

Figure 8 shows filtering on the GMRT data affected by narrowband RFI in the L -band (1,100–1,450 MHz). MAD-based filtering is carried out across 2,048 spectral channels and values above 3.5σ threshold are replaced by the robust *threshold*. Normalization of the spectrum has been carried out

before filtering. The median spectrum is computed over 1,024 time-stamps and the smooth spectrum over a window size of 20 spectral channels. Figure 8 shows 5,000 s of data in a time–frequency plot (with and without filtering). The flagged data shown below the unfiltered spectrum corresponding to the spectral channels that are detected as RFI and replaced by the threshold value. In this case, the filtered spectrum shows significant post-filtering improvement (15 dB maximum).

To illustrate the effect of normalization of spectrum on the filtering, the spectrum shown in Fig. 8 is filtered with normalization. Figure 9 shows that the amount of data flagged is lesser as compared to that seen in Fig. 8. The amount of narrowband RFI detected without normalization is lesser due to power variation across the spectrum. This shows that the ability to detect RFI from spectrum having variations in the power level is affected in absence of the normalization process. The effect is more pronounced towards the edge of the spectrum where there is a roll-off in the power level.

6. RFI Filtering in the Beamformer Chain of the GWB

The beam mode of a telescope is used primarily for pulsar observations. Section 6.1 describes RFI mitigation for the GWB beamformer data. Sections 6.2

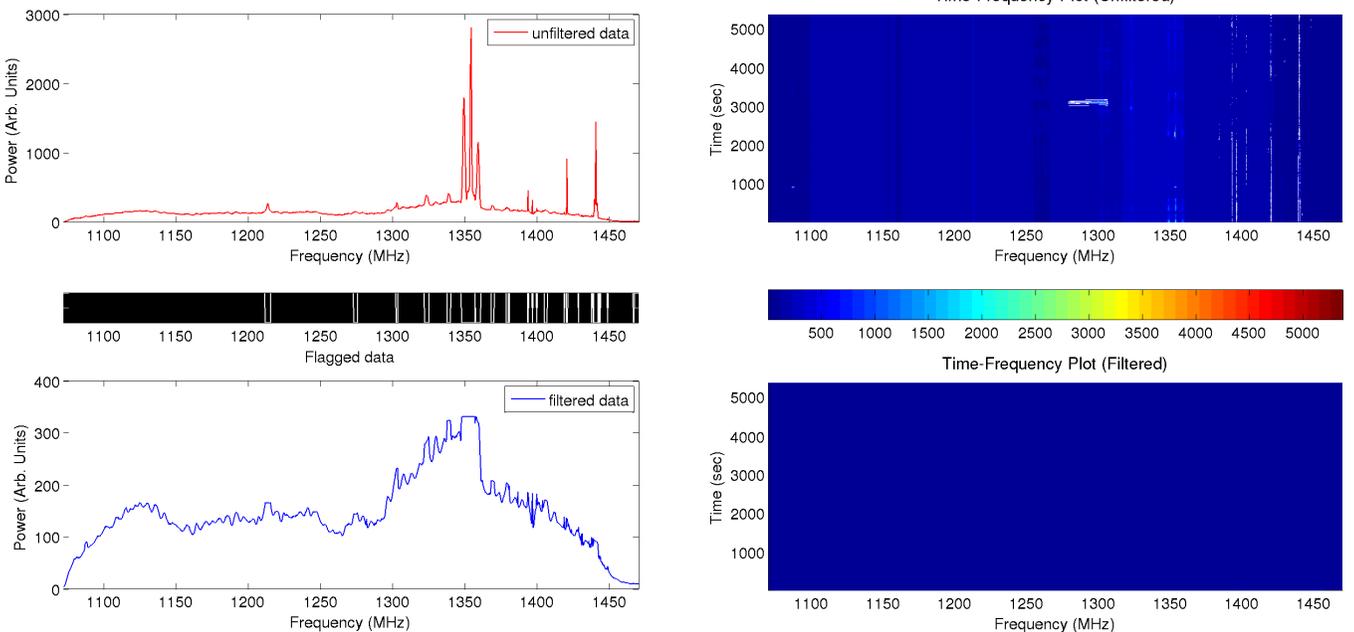


Fig. 8. Narrowband RFI filtering at 3.5σ threshold (with smoothing and normalization).

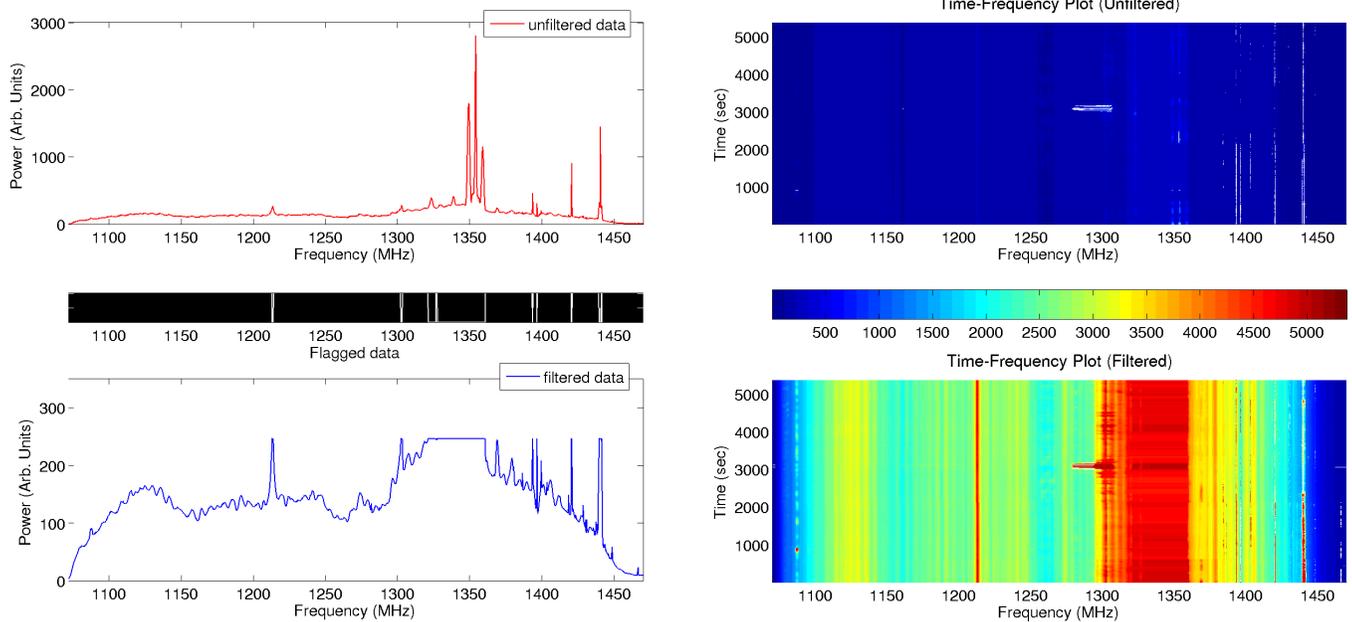


Fig. 9. Narrowband RFI filtering at 3.5σ threshold (without smoothing and normalization).

and 6.3 describe the real-time implementation and some test results from GMRT pulsar observations.

6.1. Description

The entire length of beamformer output data is divided into blocks. Each block has a few seconds of data. The size of the block is chosen based on the time-scales on which the RFI at GMRT persists. It has been observed that the spectral RFI switches ON and OFF on a few seconds time-scale. The software utility which has been developed performs the following steps of pulsar signal processing and filtering RFI on each block:

- (1) Filter narrowband RFI: Search and flag narrowband RFI from a normalized spectrum (as discussed in Sec. 5).
- (2) Filter broadband impulsive RFI: Collapse all spectral channels and use MAD-based detection and filtering (Sec. 3.1) to find (and flag) outliers from this time-series. There is an advantage of detecting RFI from this time-series; because for a large number of pulsars, owing to the pulse getting smeared (due to dispersion in the interstellar medium), the signal is buried under the noise floor and hence would not be picked up as an outlier.
- (3) The time and frequency flags generated in the above steps are kept track of, and the data

samples corresponding to the flags are masked as the data undergoes the process of de-dispersion.

- (4) The de-dispersed time-series generated in the above step undergoes a signal processing operation called folding wherein successive blocks are added and averaged to improve the signal-to-noise ratio. The de-dispersed and folded time-series is recorded on to a hard disk.

6.2. Implementation

The utility is developed in C++ and can run on any general-purpose, multi-core CPU. It uses OPENMP to manifest parallelism and takes an approach that ensures scalability, and optimum utilization, of the available resources. The division of tasks between threads is illustrated in Fig. 10. Tasks are performed for k blocks at a given point in time. All I/O operations for each of the k blocks happen serially on a single thread to avoid any bottleneck arising out of simultaneous disk or memory access. The other tasks are divided into three teams, each of which perform a certain set of operations on the k blocks, in parallel. Thus the tool uses $2 + 3k$ number of threads for the operation, where k can be optimally chosen at run-time by the user. With five threads ($k = 1$), the utility can perform all of the above operations, write out the filtered as well as unfiltered time-series (corrected for dispersion) to

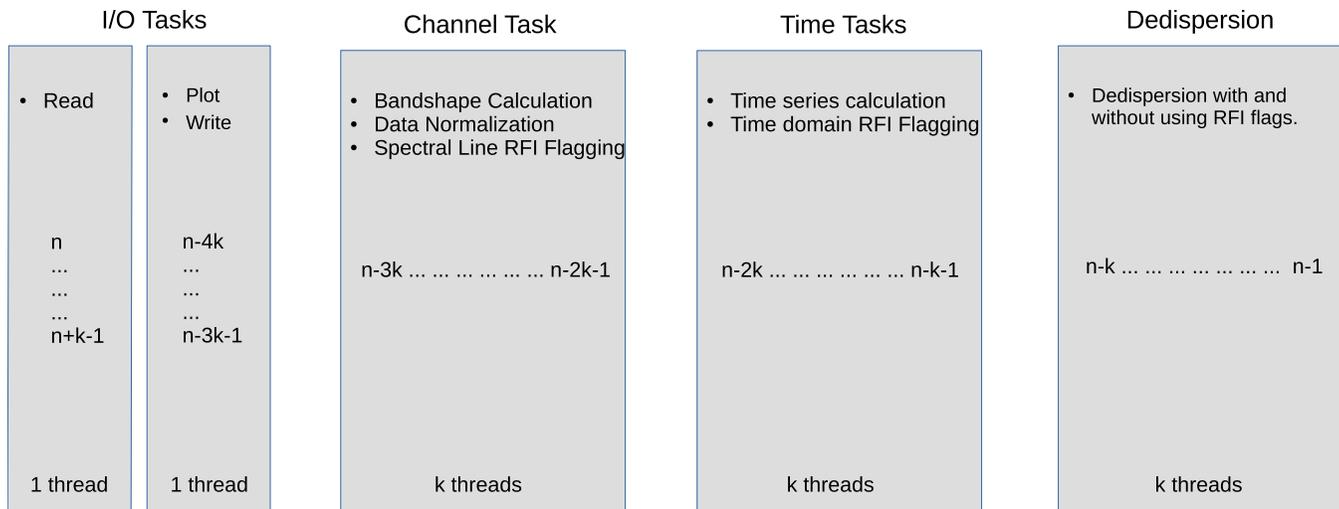


Fig. 10. Division of tasks among various threads in the beamformer RFI mitigation software.

the disk and fold the time-series in real-time at a sampling interval of $40 \mu\text{s}$ with 2,048 spectral channels.

6.3. Results

The utility has shown improvement in the quality of pulsar data across various parameters. This is illustrated by two examples of filtering on different pulsars observed from the GMRT. The observation parameters of these pulsars are listed in Table 1.

The parameters used to quantify the improvement achieved due to RFI filtering: (1) the ratio of mean-to-RMS of each block after filtering as compared to the unfiltered value (Eq. (4)) and (2) the ratio of the on-pulse signal to the off-pulse noise in the folded profile. The filtered and unfiltered profiles, and the improvement in mean-to-RMS ratio for B0138+59 in Fig. 11 and that for J1807-0847 are shown in Fig. 12. Filtering improves the pulse signal-to-noise ratio by a factor of 2 for B0138+59 pulsar and by a factor of 10 for J1807-0847 pulsar as seen in Table 2.

Table 1. Observation parameters of pulsars during beamformer RFI filter testing.

Pulsar	RF band (MHz)	Sampling interval (μs)	Channels	Observation duration (min)
J1807-0847	1260–1460	1310.72	2,048	10
B0138+59	200–300	81.92	2,048	5

7. Summary and Discussion

7.1. Summary

Real-time techniques for impulsive RFI mitigation for broadband and narrowband RFI were proposed along with results from implementation on GWB. Broadband RFI filtering showed significant improvement (10–12 dB) in the signal-to-noise ratio and cross-correlation performance. MAD-based spectral domain RFI filtering showed up to 15 dB improvement for spectral channels affected by RFI. MAD-based estimation for real-time RFI filtering was also found to be effective in improving the signal-to-noise ratio by a factor of 10 during pulsar observations carried out using GWB. These real-time RFI mitigation techniques will find applications in the existing radio telescopes and the upcoming ones like the Square Kilometer Array (SKA).

7.2. Discussion of future work

The information regarding the samples that were removed or altered during the RFI filtering operation needs to be conveyed to the user along with the final output of the backend system. In the case of GWB, the filtering is carried out at various locations. The information of the filtered samples is generated at different rates depending on the location of the RFI filtering block in the signal processing chain. Hence, it is a challenging problem to collate the information of samples flagged at different rates for the correlator and beamformer modes of the GWB. This information needs to pass through the

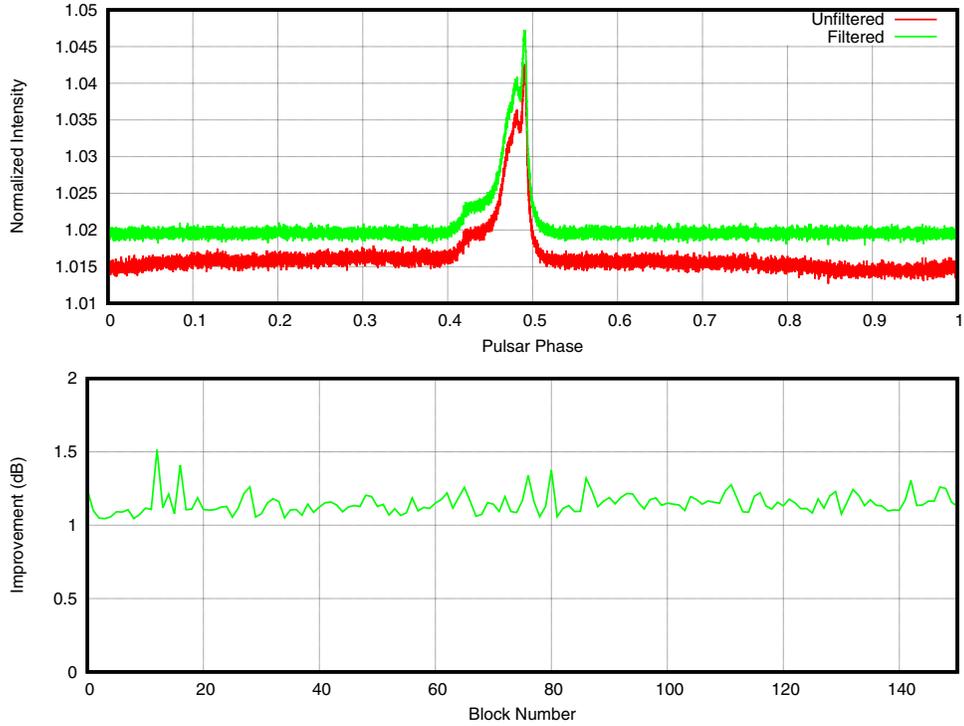


Fig. 11. Effect of RFI filtering on folded profile (subplot 1) and blockwise improvement in mean-to-RMS ratio (subplot 2) for PSR B0138+59.

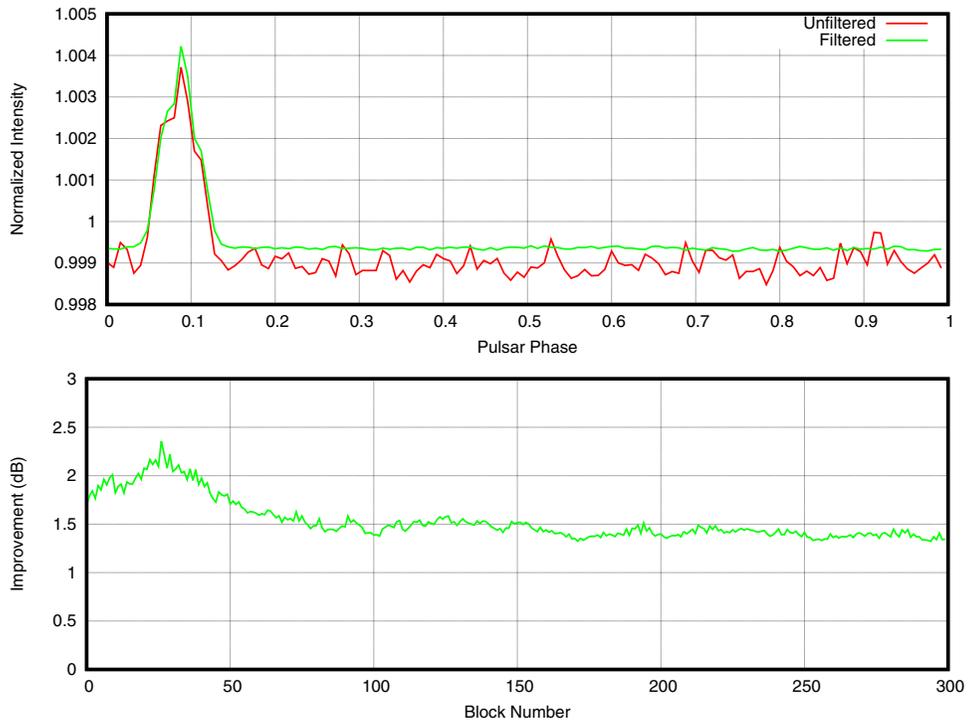


Fig. 12. Effect of RFI filtering on folded profile (subplot 1) and blockwise improvement in mean-to-RMS ratio (subplot 2) for PSR J1807-0847.

Table 2. Improvement in pulse signal-to-noise ratio (Arb. units) due to RFI filtering.

	Pulse signal to off-pulse noise	
	Unfiltered	Filtered
J1807-0847	25	116
B0138+59	376	647

signal processing chain and come out as a number indicating the fraction of original data that was detected as RFI for a given block and spectral channel. A technique is being developed for the same.

The characterization of the effect of real-time RFI filtering using the technique shown in the previous sections is carried out through antenna tests. The parameters measured are the effect of different thresholds and replacement options on the auto-correlation and cross-correlation spectrum (amplitude and phase) of the output. Also, tests are being carried out to determine the effect of RFI filtering on the astronomical image. These tests will also help in fine-tuning the parameters of the technique like the estimation window size, optimum filtering threshold, and the choice of replacement options under varying RFI conditions.

The algorithm for narrowband RFI filtering currently works on recorded correlator data. This algorithm would be optimized and ported on the CPU–GPU cluster of the GWB for real-time narrowband RFI filtering. To achieve real-time performance, the data would be read directly from the shared memory of the compute nodes processing the correlator data. The challenge is to perform this operation in real-time for 60 inputs (30 antennas, dual polarization) on a time-scale of few hundred milliseconds.

As a next level of RFI mitigation strategy for the uGMRT, there is a plan to develop techniques which result in minimal loss of astronomical data and which work effectively even for RFI buried within the system noise. The techniques under consideration are adaptive cancellation using reference antenna, spatial nulling and adaptive spatial filtering.

Acknowledgments

The authors would like to thank Sanjay Kudale for his help in data analysis. The authors would also like to thank the members of the backend group and control room at GMRT.

References

- Ajithkumar, B., Choudhari, S., Buch, K., Muley, M., Shelton, G., Reddy, S., Kudale, S., Roy, J. & Gupta, Y. [2013] Next generation digital backends for the GMRT, in *IOP Conf. Series: Materials Science and Engineering* (IOP Publishing), p. 012024.
- Astola, J. & Kuosmanen, P. [1997] *Fundamentals of Nonlinear Digital Filtering*, Vol. 8 (CRC Press, Boca Raton, NY).
- Buch, K. D., Gupta, Y. & B. A. Kumar, [2014] Variable correlation digital noise source on FPGA—A versatile tool for debugging radio telescope backends *J. Astron. Instrum.* **3**, 1450007.
- Crnojevic, V., Senk, V. & Trpovski, Z. [2004] Advanced impulse detection based on pixel-wise MAD, *IEEE Signal Process. Lett.* **11**, 589.
- Dewdney, P. E., Hall, P. J., Schilizzi, R. T. & Lazio, T. J. L. [2009] The square kilometre array, *Proc. IEEE* **97**, 1482.
- Edwards, J. [2014] Signal processing: On the edge of Astronomy’s new frontier special reports, *IEEE Signal Process. Mag.* **31**, 10.
- Ellingson, S. W. [2004] RFI mitigation and the SKA, *Exp. Astron.* **17**, 261.
- Fahmy, S. A., Cheung, P. Y. & Luk, W. [2009] High-throughput one-dimensional median and weighted median filters on FPGA, *IET Comput. Digit. Tech.* **3**, 384.
- Fridman, P. [2008] Statistically stable estimates of variance in radio-astronomy observations as tools for radio-frequency interference mitigation, *A J* **135**, 1810.
- Fridman, P. & Baan, W. [2001] RFI mitigation in radio astronomy, *A & A* **378**, 327.
- Guner, B., Johnson, J. T. & Niamsuwan, N. [2007] Time and frequency blanking for radio-frequency interference mitigation in microwave radiometry, *IEEE Trans. Geosci. Remote Sens.* **45**, 3672.
- ITU-R [2013] Techniques for mitigation of radio frequency interference in radio astronomy, RA.2126-1, RA Series, International Telecommunication Union.
- Juhola, M., Katajainen, J. & Raita, T. [1991] Time and frequency blanking for radio-frequency interference mitigation in microwave radiometry, *IEEE Trans. Signal Process.* **39**, 204.
- Langat, P. K. [2011] Power-line sparking noise characterisation in the SKA environment, PhD thesis, Stellenbosch University, Stellenbosch, South Africa.
- Niamsuwan, N., Johnson, J. T. & Ellingson, S. W. [2004] Examination of a simple pulse blanking technique for RFI mitigation, *Radio Sci* **40** RS5S03.
- Pearson, R. K., Neuvo, Y., Astola, J. & Gabbouj, M. [2016] Generalized Hampel filters, *EURASIP J. Adv. Signal Process.*, **2016**(1), 87.
- Rousseeuw, P. J. & Croux, C. [1993] Alternatives to median absolute deviation, *J. Am. Stat. Assoc.* **88**, 424.
- Swarup, G. [2008] Power-line radio frequency interference at the GMRT, NCRA-TIFR Technical Report.
- Swarup, G., Ananthakrishnan, S., Kapahi, V., Rao, A., Subrahmanya, C. & Kulkarni, V. [1991] The giant metre-wave radio telescope, *Curr. Sci.* **60**, 95.
- Shankar, N. U. & Pandey, V. [2006] Hierarchical RFI mitigation system at the Mauritius Radio Telescope, in *IAU Special Session* **6**, 22 August 2006, Prague, Czech Republic.