Powerline Interference Rejection for GMRT Pulsar Data

A thesis submitted in partial fulfillment of the requirements
for the award of the degree of

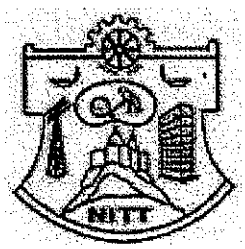
Master of Technology

in

Communication Systems

by

M.R. Visweshwar Ram



Department of Electronics and Communication Engineering

National Institute of Technology

(*Deemed University*)

Tiruchirappalli - 620 015.

June - 2004

Abstract

Pulsars are rapidly rotating neutron stars that emit synchrotron radiation at radio wavelengths continuously, but the radiation appears pulsed because of the misalignment of the magnetic axis and the axis of rotation. The main goal of observing pulsars is to understand the physics of neutron stars. Pulsed radiation is of immense experimental value to radio astronomers as an excellent probe of the interstellar medium[4]. Binary pulsars, besides, are excellent laboratories for testing gravitational theory and act as sources of gravitational waves.

Any radio telescope like the GMRT has its own peculiar problem of powerline interference. For the GMRT, it is 50 Hz, and there are radio telescopes around the world; some of them in countries that have a 60 Hz powerline frequency. Moreover, the signatures of the powerline interference are unique to the observing station due to the differing nature of the local electricity distribution schemes. However, the exact frequency is unimportant; inasmuch the problems due to the powerline interference are a serious impediment to extracting meaningful information from pulsar observations. To obtain valid results from pulsar radio data, it is important to have as clean a record of data as possible. The GMRT pulsar data (as also pulsar data from other radio telescopes) is corrupted by the powerline sinusoid and its harmonics upto a considerable order. This substantially distorts the single pulse shape. Moreover, presence of powerline harmonics in the data can deteriorate the probability of detection of new, unknown, pulsars.

An attempt is made to quantify the powerline interference and therefore suggest suitable parameters for its mitigation, especially when the frequency of the interference is not stationary in time. Three methods are discussed, each progressively more sophisticated. First, an ideal notch filter is devised and the data is filtered in blocks, the size of which is fixed from preliminary experiments. Next, an algorithm is devised, wherein instead of zeros, the notch is replaced by random noise, the power of which is equal to the noise from the electronics. It is found that these methods of filtering do not deal with the time-variant frequency of the powerline harmonics. The shortcomings of these methods motivate a better method, which allows tracking the frequencies in time from one block to the next.

The tracking filter starts with an *apriori* set of frequencies around which to look for the interfering lobes. The number of bins on either side of the prior to search for is programmable and is under the control of the user. In the first

powerline hunt, the local peak is picked and is updated as the prior for the next block of data. From the peak, a descent is made to lower-valued bins on either side upto a threshold value of $\lambda\sigma$, where it stops. In the process those bins are now replaced with random noise of power equal to that from the electronics. The whole procedure is repeated for each block of data. The same procedure is straightforwardly extended to the single channel raw data, where the algorithm now runs over blocks of data for each channel one-by-one. The filtered data is then de-dispersed. Results are presented with comparisons to the same before filtering. Finally, the entire algorithm is fit into the pulsar data acquisition chain of the GMRT to filter out the interference in real-time, as and when the pulsar is observed. It is shown that this improves the probability of detection of unknown pulsars during a targeted or an all-sky search.

Contents

List of Figures

List of Tables

Chapter 1

Signal Processing in Radio Astronomy

Radio Astronomy is impossible without signal processing. All signals collected are essentially electrical, and as such are not immediately tangible to the human intellect. Complex signal processing at all levels of the data chain of a radio telescope render possible the results obtained, be it a radio map of a distant galaxy, or the kinematics of a fast-spinning neutron-star. Radio telescopes across the world are becoming increasingly sensitive and ambitious in the conquest of the edge of the universe.

1.1 Large Synthesis Arrays

A radio telescope is a passive radio receiver. They can either be multi-element arrays or single-element receivers. With single-element receivers radio maps can be made although with limitations, but can continue to study pulsars, emission spectra and even the ionosphere. But to make high resolution radio maps of radio sources, an array is required. Such arrays are called synthesis radio telescopes. Synthesis arrays typically consist of many antennas spread over a large area, so that it effectively acts as a single antenna of a very large collecting area. Because of the very large spread of the antennas, the signal from the same source reaches the different antennas at different instants, depending on the location of the source with respect to the antennas. If the antennas are tracking the source, these delays are slowly changing with time as well. But a fixed delay in time

could translate into different phase delays for different frequencies of observation. Thus the phase delay between the signal reaching the different antennas varies continuously as the antennas are tracking the source as well with the frequency of observation even if antennas are pointing to a fixed co-ordinate in the sky. This continuous phase information is also part of the data in radio-map making. This is called *Interferometry*[7][13]. In interferometry, the signals from all the antennas are collected, and signals from each pair of antennas are correlated. The correlated data thus obtained is inverse-Fouriér transformed to get the radio map. The correlated data as such is called the visibility. Mathematically, the visibility function of a source is sampled spatially using the radio telescope. A Fouriér transformation yields the radio-image of the observed source.

1.2 Giant Metrewave Radio Telescope

1.2.1 Introduction

The *National Centre for Radio Astrophysics*(NCRA) has set up a unique facility for radio astronomical research using the metrewavelengths range of the radio spectrum, known as the *Giant Metrewave Radio Telescope*(GMRT). It is located at a site about 80 km north of Pune. GMRT consists of 30 fully steerable gigantic parabolic dishes of 45 m diameter each spread over distances of upto 25 km. GMRT is one of the most challenging experimental programmes in basic sciences undertaken by Indian scientists and engineers. Now fully operational, it is the world's most powerful radio telescope operating in the frequency range of about 50 to 1500 MHz.

The metre wavelength part of the radio spectrum has been particularly chosen for study with GMRT because man-made radio interference is considerably lower in this part of the spectrum in India. Although there are many outstanding astrophysics problems which are best studied at metre wavelengths, there has, so far, been no large facility anywhere in the world to exploit this part of the spectrum for astrophysical research.

The number and configuration of the dishes was optimized to meet the principal astrophysical objectives which require sensitivity at high angular resolution as well as ability to image radio emission from diffuse extended regions. Fourteen of the thirty dishes are located more or less randomly in a compact central array

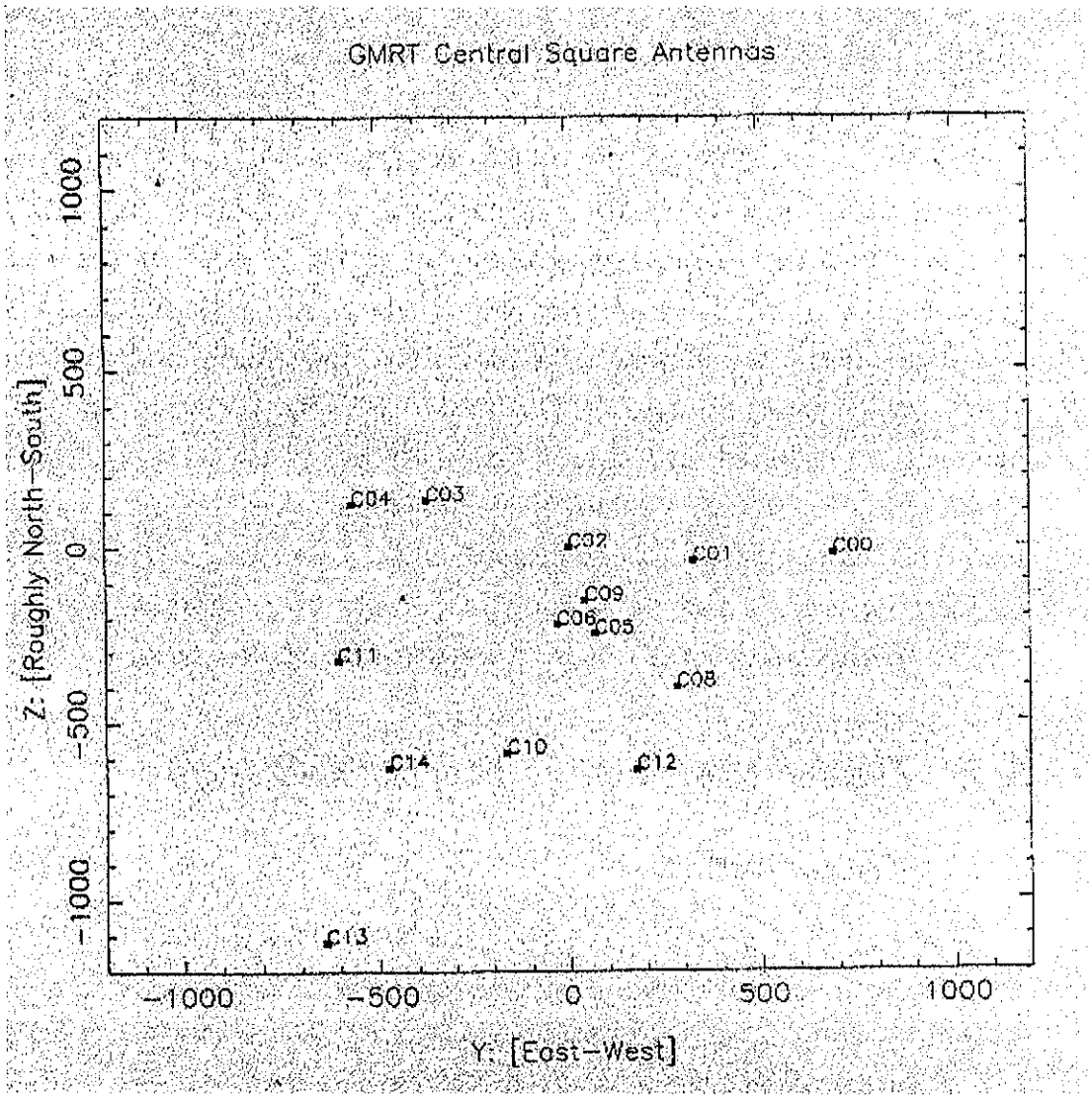


Figure 1.1: GMRT Central Square

in a region of about 1 sq km.

The remaining sixteen dishes are spread out along the 3 arms of an approximately 'Y'-shaped configuration over a much larger region, with the longest interferometric baseline of about 25 km.

The multiplication or correlation of radio signals from all the 435 possible pairs of antennas or interferometers over several hours thus enables radio images of celestial objects to be synthesized with a resolution equivalent to that obtainable with a single gigantic dish 25 kilometre in diameter! The array operates in six frequency bands centred around 50, 153, 233, 325, 610 and 1420 MHz. All these feeds provide dual polarization outputs. In some configurations, dual-frequency observations are also possible.

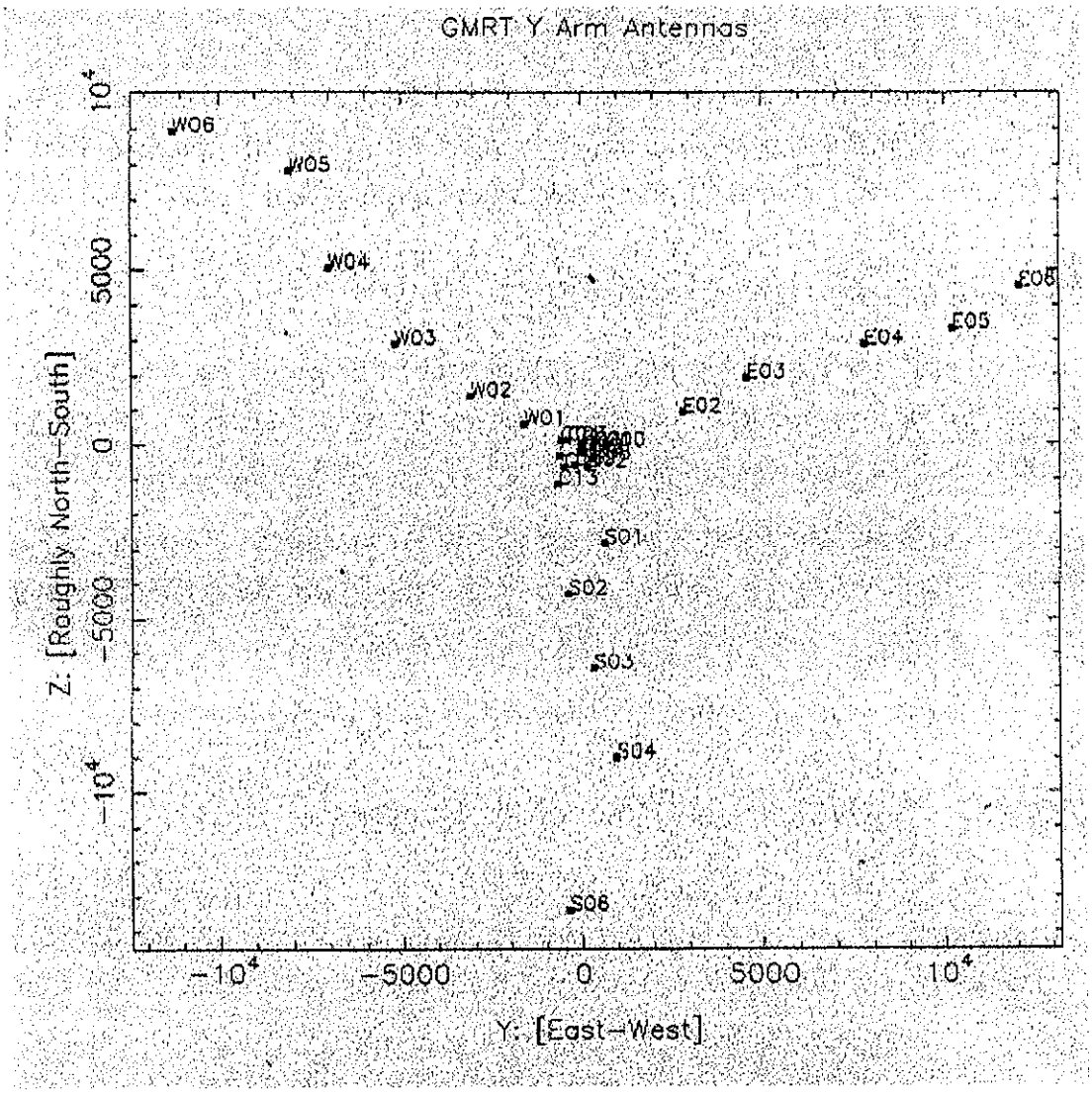


Figure 1.2: GMRT array configuration

The highest angular resolution achievable ranges from about 60 arcsec at the lowest frequencies to about 2 arcsec at 1.4 GHz.

The large size of the parabolic dishes implies that GMRT has over three times the collecting area of the Very Large Array (VLA) in New Mexico, USA which consists of 27 antennas of 25 m diameter and is presently the world's largest aperture synthesis telescope operating at centimetre wavelengths. At 327 MHz, GMRT is about 8 times more sensitive than VLA because of the larger collecting area, higher efficiency of the antennas and a substantially wider usable bandwidth because of the low level of man-made radio interference in India.

1.2.2 Goals of the GMRT

Although GMRT is a very versatile instrument for investigating a variety of radio astrophysical problems ranging from our nearby Solar system to the edge of the observable universe, two of its most important astrophysical objectives are:

- to detect the highly redshifted 21 *cm spectral line of* neutral Hydrogen *expected from protoclusters or protogalaxies before they condensed to form galaxies in the early phase of the universe and*
- *to search for and study rapidly-rotating* pulsars *in our galaxy*

1.2.3 GMRT Pulsar Receiver System

A radio telescope of the scale of the GMRT has a justifiably sophisticated electronic receiver system. There are two separate data chains, one for the interferometric observations and the other for pulsar observations. A brief sketch of the pulsar receiver system is given here. For a more detailed and complete technical explanation, refer to [3].

Each antenna has two orthogonally placed dipoles for the two polarizations. Each polarization produces a bandwidth of 32 MHz for the two sidebands together, or 16 MHz for each sideband. Since there are 30 such antennas, a total of 120 analog signal channels are produced. Data from each channel is sampled at Nyquist rate of 32 Mc/s and digitized to four bits per sample. The signals then pass through the delay units to compensate for the delays, and then Fouriér transformed in real-time once over $16\mu s$ to provide a continuous stream of 256 complex spectral channels. The first block of the pulsar receiver is the GMRT

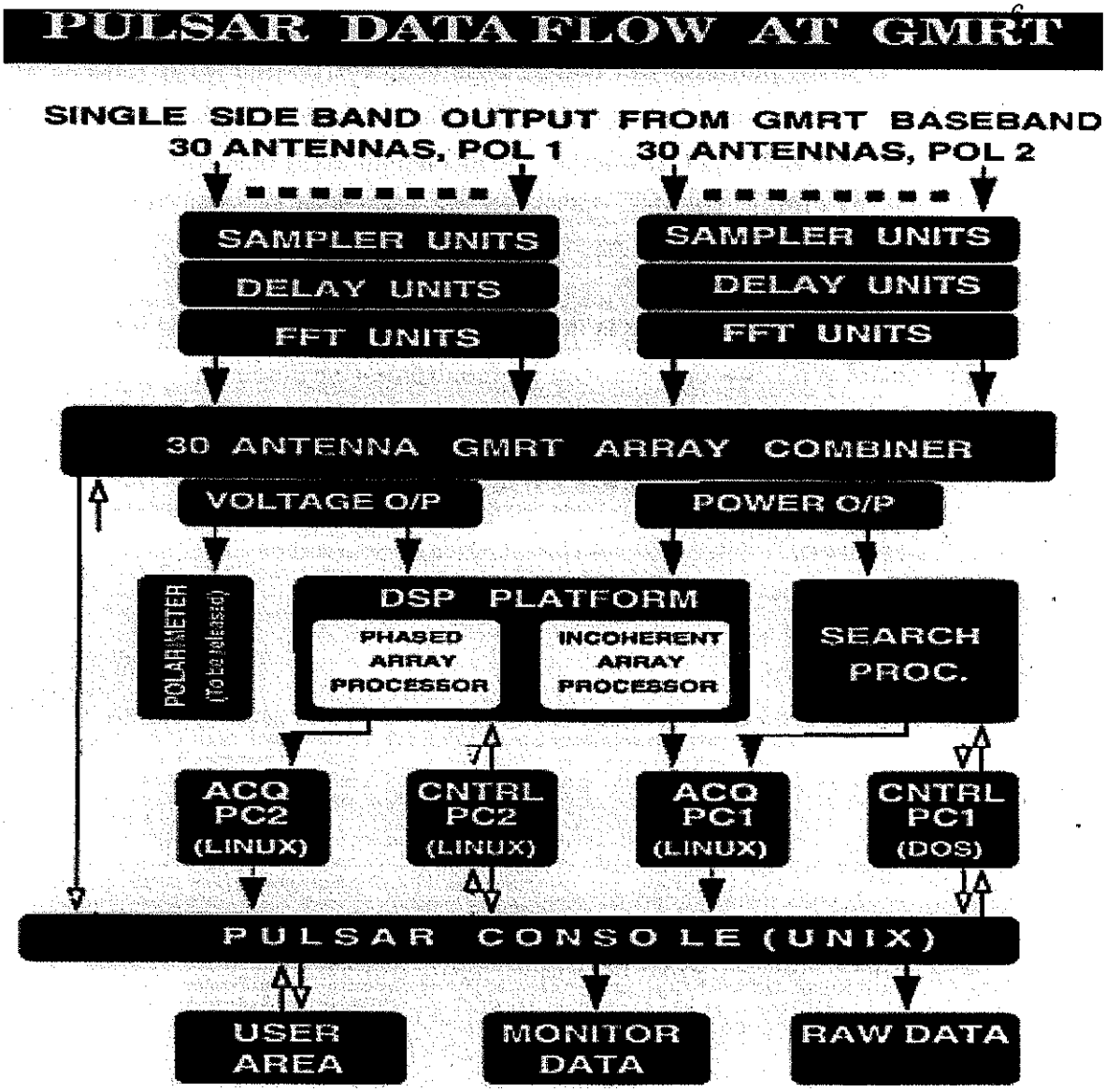


Figure 1.3: GMRT Pulsar Receiver System

Array Combiner(GAC). It combines the signals from a maximum of 30 antennas for both incoherent and coherent phased array modes of operations. Currently there are two pulsar back-ends for pulsar observations.

1. The search processor which preintegrates the data to the desired sampling rate, subtracts long-term mean for each frequency channel and provides 1-bit samples to the data acquisition computer(sub-system of GAC rack).

2. The DSP platform which provides full multi-bit data at fast sampling rates.

The DSP platform can handle both coherent phased array data as well as incoherent phased array data. The coherent array observations are recorded on the PC ACQPC2, whereas the incoherent array observations are recorded on ACQPC1. Besides, the search processor also accesses PC ACQPC1 for further processing.

The control and monitoring is done via a Unix console, which is interfaced with the control room systems and can be monitored continuously.

Chapter 2

The Physics of Pulsars

Pulsars were discovered serendipitously by Prof. Hewish and his Ph.D student Jocelyn Bell. When they were trying to study interplanetary scintillations, they observed extremely regular radio pulses with a period of 1.33 s appearing four minutes earlier every day. The four-minute advance led them to conclude that the signals were indeed of celestial origin, and not terrestrial RFI.

2.1 What is a pulsar?

When the initial mass of a star exceeds the *Chandrasekhar limit*[14], it may end up exploding as *supernova*. In the middle of the mantle of expanding gas is a tiny(in the astronomical sense) remnant of extremely dense material only a few tens of kilometers in diameter. This remnant is chiefly made of neutrons, and is hence called a neutron star. The neutron star may be 10^{14} times as dense as water. It has extemely powerful magnetic fields of the order of 10^{12} *Gauss* or more. They rotate very rapidly, and this happens because angular momentum is always conserved. Hence when the star shrinks in diameter, it spins up to conserve angular momentum.

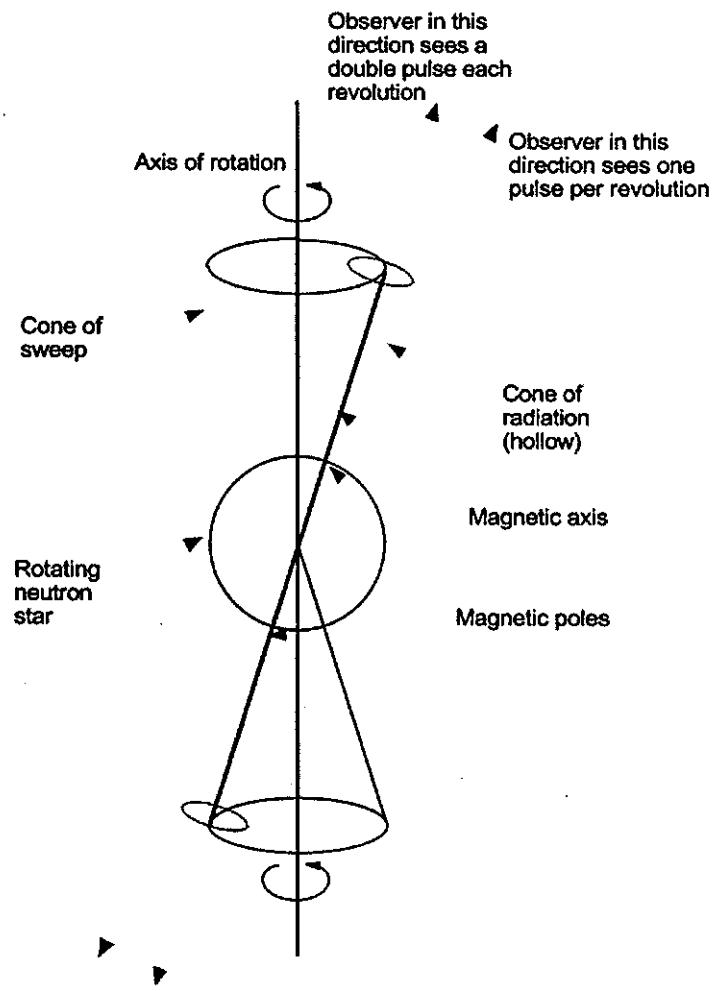


Figure 2.1: A Pulsar

2.2 Why are pulsar signals periodic?

The radiation from a neutron star comes from a very small portion of the surface of the star, and the star is also rotating at the same time. This way one can readily explain a pulsating radio source. What kind of a star can emits radiation in only one direction? Only if the radio beam is along one direction and if the star is also rotating can one have a lighthouse effect. Such a thing can happen in a neutron star, as figure 2.1 readily explains.

The exact mechanism for radiation from a pulsar is still a mystery, although at present a few models are available that explain the observed phenomena. One such model is briefly discussed in the next section. The radiation, whatever its originating mechanism, is indeed continuous, but the pulsed nature of the

observed radiation is due to the fact that the axis of rotation and the magnetic axis are misaligned. The radio beam from the pulsar sweeps through the observer at a rate equal to its spinning frequency. Thus, if the *Crab pulsar* is emitting radio pulses every 33 ms, it only means that it is spinning on its axis once every 33 ms. Astronomers study pulsar signals in an attempt to solve the mystery of the radiation mechanism as well, besides studying the physics of the interior of the neutron star.

2.3 Gold's model of pulsars

It was T. Gold who first suggested a plausible mechanism by which the observed radio pulses with clock-like precision in periods could be radiated by a rotating neutron star. Many other theories have since been proposed by astronomers to explain the mechanism of radiation by pulsars.

In the intense magnetic field of a pulsar, any charged particle that may escape from its surface is constrained to move only along the magnetic lines of force. The assemblage of particles is thus whirled around with the angular velocity of the neurtron star. In this manner, a co-rotating magnetosphere (consisting of *frozen-in* plasma) is formed around the star. The tangential velocity of these whirling particles gradually increases as they move further from the stellar surface. At a certain distance, depending upon the speed of rotation of the neutron star, the velocity of the particles approaches that of light.. The circle described by particles at this distance is called the *velocity of light* circle. The relativistic plasma beam near this circle radiates radio waves perpendicular to the beam, but at the velocity of light circle where the particles have attained the speed of light, they will break away from the magnetosphere and flow out into the surrounding regions of space.

It is deduced that the radius r of the velocity of light circle for the *Crab pulsar* is about 1600 km. It can be easily calculated from the simple formula

$$r = \frac{c}{\omega} = \frac{cP}{2\pi} \tag{2.1}$$

P being period and c being velocity of light. Near this circle the charged particles move along the magnetic lines of force in helical paths producing *sychrotron radiation.* According to Gold, magnetic activity will eject plasma from *only one or at most a few* places on the surface of the neutron star. The emitting particles are confined within a narrow cone. In fact, particles are confined in a narrow

pitch angle $\theta = \tau/P$ in radian measure, where τ is the duration of the pulse and P is the pulsar period. Since the pulsars rotate with very high velocities, the radiation will be *strongly beamed* in a narrow cone in the forward tangential direction sweeping a particular region periodically, so that to a distant observer the pulsar behaves like a *lighthouse* signalling radio waves.

A modification of Gold's theory has been suggested by P. Goldreich and W.H. Julian in respect of the mechanism for particle acceleration to high velocities. They suggest that a rotating magnet will, due to unipolar induction, generate an electric field E in its vicinity which can cause charge separation and accelerate particles away from the surface to form a radiating magnetosphere. This theory predicts that acceleration of particles to energy as high as $10^{18}\ eV$ is possible in this way.

Other models for pulsed radio emission have also been suggested. For a brief discussion of various models, refer to [6].

Chapter 3

Radio Study of Pulsars

Pulsars emit strong electromagnetic radiation. It is not established that pulsars radiate via synchrotron radiation - this is probably not the most accepted mechanism; rather, a variant of this - curvature radiation - is thought to be more relevant. A substantial portion of this radiation is emitted at radio wavelengths. A few pulsars are known to emit at other wavelengths, but most pulsars emit at radio wavelengths. Therefore radio observation of such sources should reveal their nature, and this has been the main reason why pulsars are studied through the radio window of the electromagnetic spectrum. The main goal of the study of pulsars is to understand the physics of the neutron star.

3.1 The Interstellar Medium

Radio pulses travel in the interstellar medium with the group velocity v_g, given by

$$v_g = c\sqrt{1 - \frac{\omega_p^2}{\omega^2}} \tag{3.1}$$

In an ionised gas the group velocity is related to the phase velocity v_p by

$$v_g.v_p = c^2 \tag{3.2}$$

The phase velocity at wavelength λ is obtained from the refractive index μ_p as

$$\mu_p = \sqrt{1 + \frac{Nr_0\lambda^2}{\pi}} \tag{3.3}$$

where $r_0 = e^2/mc^2$ is the classical radius of the electron, and N is the electron density. Hence for small N

$$v_g = \left(1 - \frac{Nr_0\lambda^2}{2\pi}\right) c \tag{3.4}$$

and the delay t in travel time over a distance L, compared with free space, is

$$t = \frac{Nr_0c\nu^{-2}}{2\pi} L \tag{3.5}$$

$$= 1.345 \times 10^{-3}\nu^{-2}NL \text{ s Hz}^{-2}\text{cm}^{-2}.$$

It is convenient to characterize the transmission path by the product NL, which in customary astrohpysical practice is measured in units pc[1] cm^{-3}. The product is then known as dispersion measure DM. Observers commonly quote radio frequencies in megahertz, so that the delay in seconds becomes

$$t = 4.15 DM \nu_{MHz}^{-2} \tag{3.6}$$

The frequency dependence of this delay has a very important effect on observations of radio pulses. A short broad-band pulse will arrive first at high frequencies and later at the lower frequencies, traversing the spectrum at the rate

$$\dot{\nu} = -\frac{\nu_{MHz}^3}{8.3 \times 10^3 DM} \text{ MHz s}^{-1} \tag{3.7}$$

The overall effect is that the pulse is smeared out if the signals from the different bands are added without proper correction for dispersion. Now, a pulse with length τ occupies instantaneously a bandwidth

$$B_i = \frac{\nu_{MHz}^3}{8.3 \times 10^3 DM} \tau \text{ MHz} \tag{3.8}$$

A receiver with a smaller bandwidth than B_i will not receive the whole of the pulse energy, and the sensitivity will be reduced. If the reciever has a larger bandwidth B_r the pulse will be detected for an extended time, while the frequency dispersion takes the pulse energy across the receiver band.

[1]pc - parsec

1 parsec = 3.2593 light years

3.2 Incoherent De-dispersion

The effect of dispersion is to reduce the sensitivity of a pulsar observation with fixed receiver bandwidth B_r to pulsars with high dispersion. One way to getting around this problem is to have a large enough bandwitdh to accommodate pulsars with the highest expectecable dispersion measures, and dividing the entire bandwidth into narrow bands each with bandwidth of atleast B_i and using separate receivers on each band. The output from receivers across each band can then be added with appropriate delays so that the pulse components are superposed. This process is called incoeherent de-dispersion. The GMRT pulsar system splits the 16 MHz bandwidth into 256 channels, so that the effective bandwidth of a channel is

$$B_{single\ channel} = \frac{16\ \text{MHz}}{256} = 62.5\ \text{KHz} \tag{3.9}$$

The GMRT system allows flexibility in selecting channels for pulsar observation. If there is, say, a particular channel which shows symptoms of harbouring RFI, that channel can be excluded from the observation by simply setting the gain of that channel to *zero*.

To perform the operation of de-dispersion, one must know *apriori* the *DM* of the pulsar being observed. If the pulses from the individual channels are added at *zero DM*, the added pulse will be smeared out, the breadth of smearing depending upon the *DM* itself. However, if the pulses are added with proper delays corresponding to the correct *DM*, the pulse shape is preserved to a great extent, of course under the assumption that the dispersion within each of the single channels is indeed negligible. Essentially, dispersion affects signal the pulse energy even within the single band, but such efects are neglected. In cases where such an approximation is not tolerable, the technique of coherent de-dispersion is employed. This makes a full correction for the continuous dispersion across the observing band. Though de-dispersion can be achieved at the hardware level, it is easier done in software. The highest frequency channel can be taken as a reference channel as it will be the one in which the pulse arrives earliest. For all other channels, the delay from the reference channel is calculated from 3.5, where ν must be instead substituted by the centre frequency of the channel under consideration. The time delay thus calculated is converted to number of bins, which depends on the *sampling frequency*. Thereafter, the signal from the channel is shifted by those many bins, and this whole process is continued for all channels. The amount

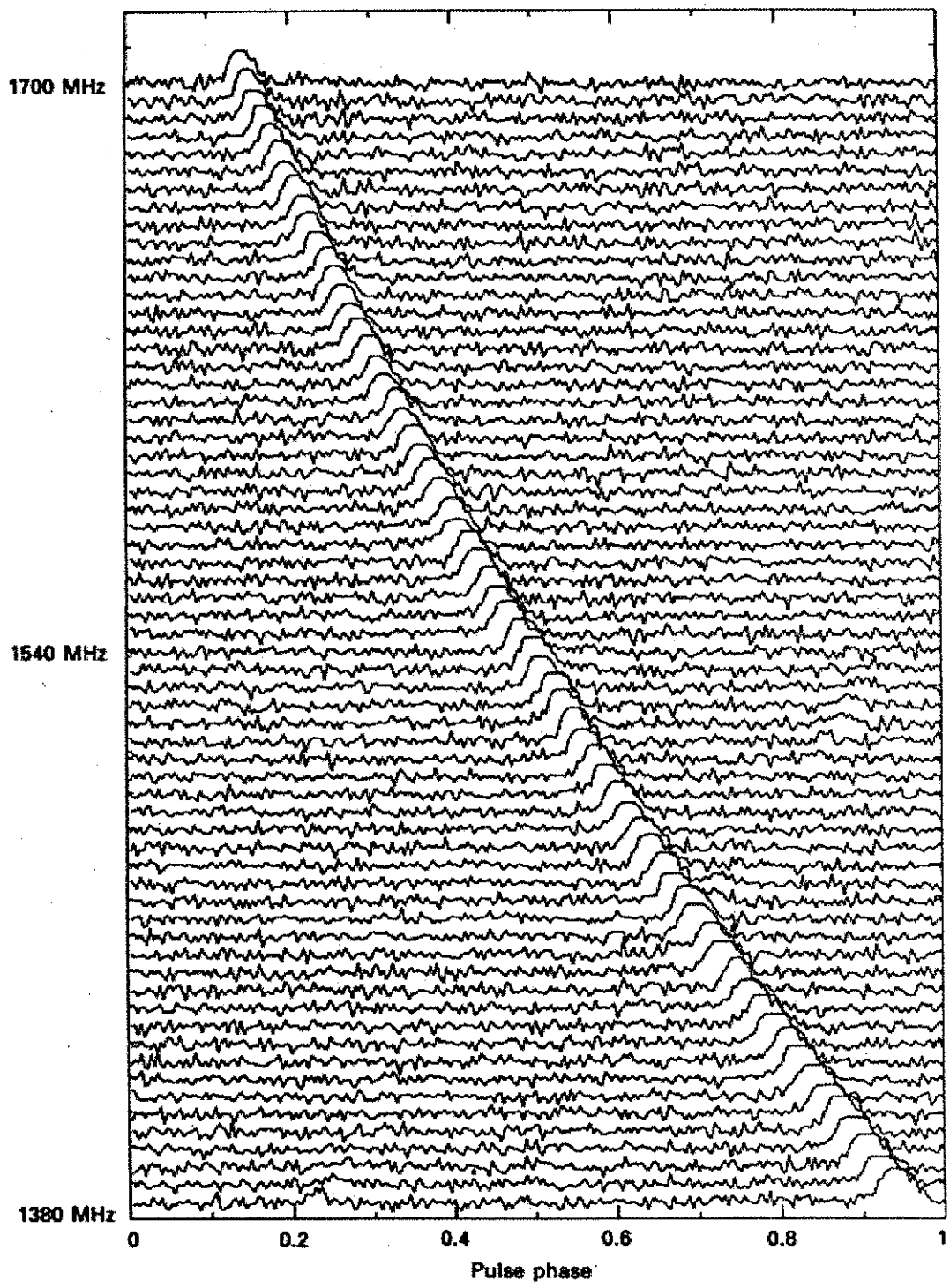


Figure 3.1: Dispersion of a pulse across frequencies

by which to shift the signal in a particular channel depends on the frequency of observation(which fixes the centre frequency of that particular channel as well), and the *DM*.

3.3 Phased Arrays for Pulsar Observations

Pulsars are fascinating compact objects that are known to emit extremely strong electromagentic radiation. They are powered by their inherently strong magnetic fields, the strengths of which are unimaginably high. However, because of their compactness and also because the effective area from which this radiation originates is very small, typically of the order of 100 km across, the fraction of energy that reaches the radio telescope several light years away is insignificantly small. Moreover the dispersive effects of the ISM come into play. To overcome these stumbling blocks, one needs a large collecting area. Building very large single-dish telescopes is prohibitive; and it can be circumvented by having a multidish radio telescope like GMRT. With an instrument like GMRT, the effective collecting area is increased manifold; typically for the GMRT it is equivalent to operating a single dish of diameter 25 km. The signal from a point source reaches the different antennas at different instants in time, the inter-element delay being a function of the geometry of the array, the location of the source in the sky and the frequency of operation. Besides, the different path lengths of the signal paths from the various antennas to the radio receiver have be to taken into consideration. Thus if the signals from the different antennas have to be added coherently, these delays must be corrected for. Different delays are now added to each antenna output at the radio receiver to offset the inherent delays, and this mode of operation of the radio telescope is called *phased array* operation.

In the GMRT phased array mode, pulsar observations can be carried out in two different ways: *(i)*coherent phased array mode and *(ii)*ncoherent phased array mode. In the incoherent phased array mode, the signal from each antenna is put through a square-law detector and the output from these is added in the *GMRT Array Combiner*(GAC) to obtain the net signal. In coherent phased array mode, the voltage signal from each antenna is added and the summed output is put through a square-law detector to obtain the final power signal. For an array of N antennas, the incoherently phased array gives a sensitivity of $\sqrt{N}$ times that

of a single antenna, while the coherent array gives a sensitivity of N times that of a single antenna. The incoherent array has an effective beam that is the same as that of the single dish, whereas the coherent array has a beam width that is much narrower than that of a single antenna dish being $\approx\lambda/D$, where D is the largest spacing between antennas in the array. The coherent phased array mode is ideally suited for observations of known pulsars. The incoherent phased array mode is most useful for large scale pulsar search observations, where the aim is to cover a maximum area of the sky in a given time at a given level of sensitivity[2].

Chapter 4

Pulsar Time Series and Spectra

Pulsar signals are inherently periodic in nature, so that if the m^{th} pulse has arrived and the period P is known, then arrival time of the $(m+n)^{th}$ pulse can be calculated precisely. However, one doesn't see a regular pulse train satisfying the well-known relation

$$x_{m+n}(t) = x_m(t); \qquad n = 0, 1, 2, \ldots \tag{4.1}$$

because the amplitude of the pulses appear to vary randomly with time. At radio wavelengths scintillation is encountered in many different circumstances, because there are many kinds of radio transmission paths which contain the necessary phase irregularities[4]. The effects of the interstellar medium may be thought of either as refraction of diffraction; in more general terms the waves are scattered, giving rise to an angular spread of waves and to fluctuations in wave amplitude. A significant part of this fluctutation is also thought to be intrinsic to the pulsar emission mechanism. However, scattering in the ISM is always a small angle scattering i.e. the direction of the intensity is redistributed in angle to some extent, but most of it does make it to the observer – there is no real loss of signal strength due to scattering.

4.1 Powerline Interference

Pulsar signals received on earth are inherently weak for the following reasons

- Pulsars are compact objects and the entire radiation comes from a very small area near the surface of the neutron star, typically 100 km across.

- They are scattered by the ISM and only a small portion of the beam sweeps across the telescope These factors combined with the instrumental limitations of an electronic receiver system render difficult the detection of very weak pulsars. But there are other contributing factors that degrade the probability of detection of pulsars. One of them is the *powerline interference.*

Powerline interference is an extremely undesirable form of interference, whereby the pulsar signal literally rides on the powerline wave and its harmonics. It worsens the quality of data to so serious an extent that the single pulse shape is no longer what it should like, but is invariably polluted with the signatures of the 50 Hz and its harmonics. The cumulative pulse shape is not affected much because the repeated superposition of pulsar pulses in phase annuls the 50 Hz effect becuase of its random phase from pulse to pulse. But single pulse proile is like the heartbeat of a neutron star; radio astronomers studying pulsars are now the cardiologists diagnosing the electrocardiogram of a patient. Most of the physics of the neutron stars can be revealed by the studying single pulse profiles, and radio astronomers have a strong tendency to spend a good deal of time viewing the single pulse profiles one-by-one. However, it is desirable to have a clean pulse profile that is inherently free from powerline interference.

4.1.1 Dirty time series data

Many data files with significant amounts of powerline interference were analysed. All of them invariably contain signatures of 50 Hz and its harmonics upto a considerable order. An example time series is given below, which gives an indication of the effect of the 50 Hz in resembling the periodicity of the real pulsar emission. It therefore becomes important to filter out the powerline interference if one is to get clean single pulse profiles.

4.1.2 Dirty spectra

The powerline interference may not be really obvious as such from the time series data unless it is farily strong, but it leaves a distinct signature in the frequency domain. Therefore, a Fouriér transformation of the time series should reveal the

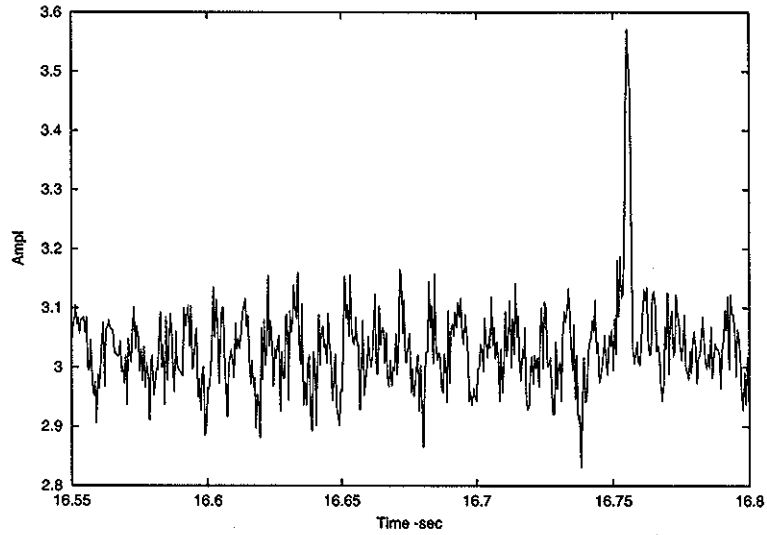


Figure 4.1: PSR b0943+10 time series with powerline interference

presence of these signatures. It is found that the powerline frequency is fairly strong and stands out from the pulsar harmonics.

However by cumulative spectral averaging, one can as well make out other interesting features of the powerline interference. The powerline frequency is not stationary at 50 Hz. The frequency is centered at 50 Hz, but wanders slowly from the mean 50 Hz. Thus, the whole data is segmented into several blocks within each of which the interference frequency is stationary, and the spectra of all these blocks are averaged. This reveals that the frequency of powerline interference has indeed changed, because what is seen is not a single line but a significantly broad lobe.

Ideally, one would want to see a spectrum that looks clean around 50 Hz, 100 Hz, 150 Hz etc. And filtering should be done in a manner that does not alter the spectral shape at any of these frequencies drastically. Moreover filtering pulsar data that has the pulsar fundamental close to 50 Hz can be very difficult, lest we should lose the pulsar signal itself.

Some examples of pulsar spectra for a single block of 1M data samples and the cumulative averaged spectra over the entire observational data are shown below.

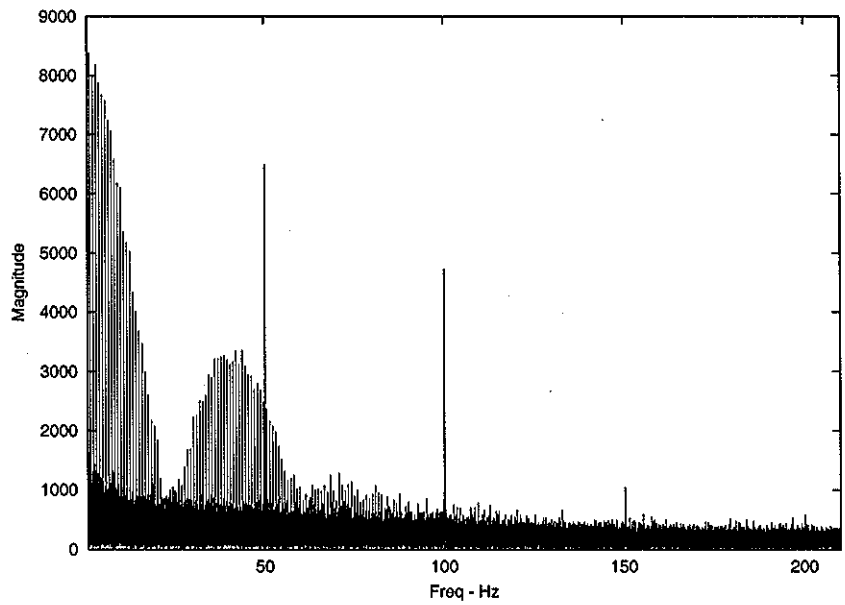


Figure 4.2: Spectrum of PSR b1133+16 for 1M data samples

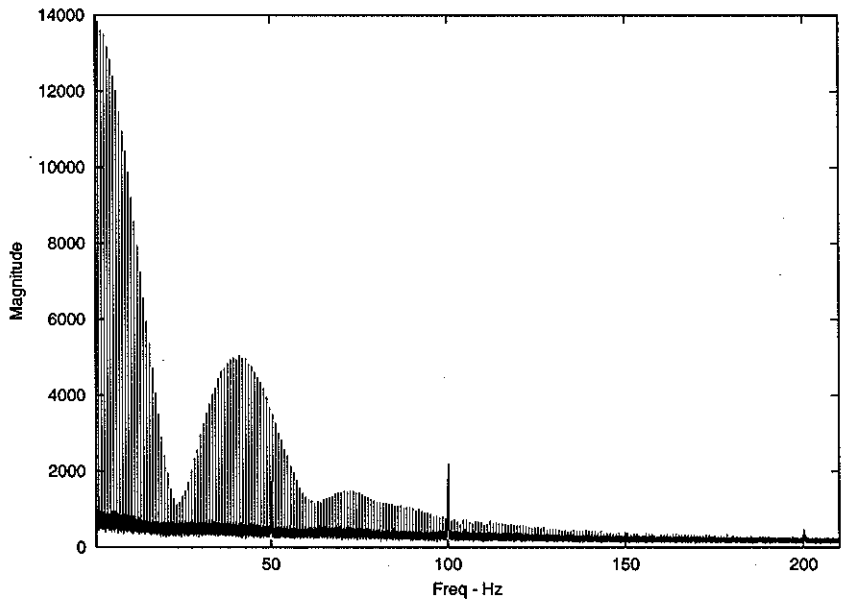


Figure 4.3: Cumulative averaged spectrum of PSR b1133+16

4.1.3 Why is 50 Hz undesirable?

Powerline frequency interference is extremely undesirable in data analysis where extremely high precision and time resolution are crucial to robust interpretation

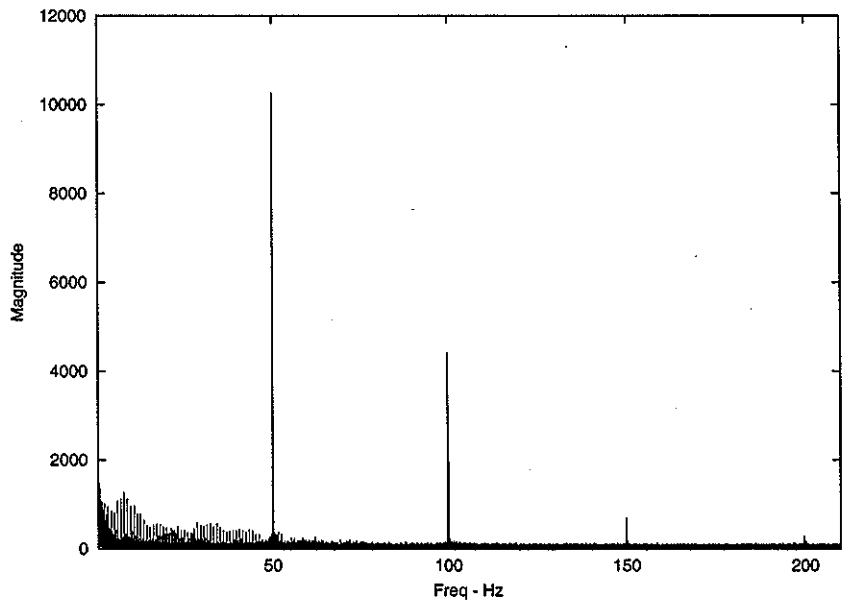


Figure 4.4: Spectrum of PSR b0943+10 for 1M data samples

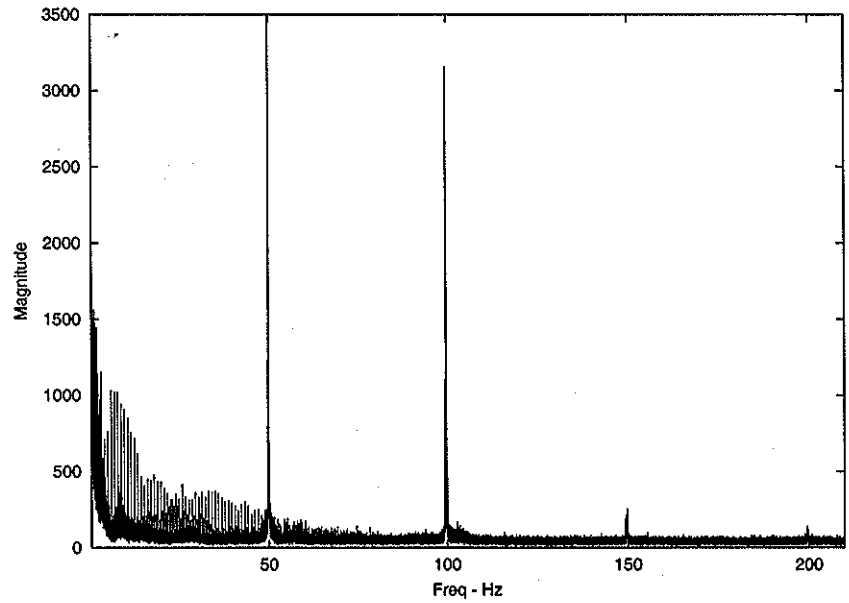


Figure 4.5: Cumulative averaged spectrum of PSR b0943+10

and doing meaningful science. The 50 Hz can at times be so strong that single pulse profiles look no different from the powerline spikes. However, this can be overcome in the cumulative averaged profile because of the random phases with

which the powerline interference signals add. Nevertheless, astronomers will be much happier to have clean single pulse profiles so that they can study effectively the pulse emission scenario in a pulsar.

50 Hz is also undesirable in a slightly more sophisticated manner. The pulsar emission profile has fine sub-pulses drifting in phase cyclically every few pulses.[1] From the scientific point of view, these fine structures can tell a lot about the physics of pulsars. Infact, astronomers look very carefully at these sub-pulses, how they drift across the pulsar period, their phase relations, etc. to understand the physics better. But the powerline interference can sometimes be so strong that the sub-pulses are not clearly resolved and becomes very difficult to tell one from another. It is now clear why powerline interference is undesirable.

Another avenue of pulsar research where 50 Hz and harmonics could be a serious impediment to doing meaningful science is discovering new pulsars. More than 1500 pulsars are now known. Nevertheless, pulsar astronomers always fancy discovering new pulsars and do a complete study of an object whose dynamics are not yet known.

Pulsars are searched for in two ways; one is by a targeted search and the other is an all-sky search. Though the two are different methods of search, one looks for signatures of periodicity in the data collected. There are two ways of doing this. Periodicity can be searched for in the time domain, which is difficult if the signals are very weak and also depends heavily on the telescope sensitivity. The other way to look for pulsars is to transform the collected data to the Fouriér domain and look for signatures of periodicity. This method yields better results than the former because the collected data can be sectioned into many blocks of reasonable length and cumulative spectra can be obtained. However, the second method works well only with targeted search where the telescope is always pointed at a fixed source in the sky. For an all-sky search, the methods to look for pulsar signatures are rather sophisticated and require supercomputing power.

The presence of powerline interference in such observational data seriously deteriorates the probability of detection of pulsar signals. The complete derivation substantiating this statement is given in section 7.1 on page 47.

[1] Refer pages 36 and 37 for plot of sub-pulse drifting

4.2 Preliminary Experiments

A good way to start thinking about a method for filtering out powerline interference is to study available data. These interfering harmonics must be rejected doing minimum damage to the actual pulsar harmonics around the powerline frequencies, lest the removal of pulsar harmonics should distort the pulse shape. Therefore the nature of the interfering harmonics demands a close study, in order for us to be able to quantify and characterize their effect on the required signals - namely the pulsar signals.

4.2.1 Broadband interference

Many data files were studied and they corresponded to different frequencies of observation. Notably, observations at 225 MHz, 325 MHz and 610 MHz were studied. And all of them invariably showed considerable powerline interference. But it was observed that the strength of the interfering powerline signals also depends on the frequency of observation. As a general rule, the strength decreases as the frequency of observation is chosen higher. But however, the interfering 50 Hz is present across a very broad range of frequencies, from as low as 225 MHz to atleast as high as 610 MHz in our case, but actually more. The interference is broadband in nature, though sometimes it is also observed that it leaks in through a particular 62.5 KHz channel that is polluted with a strong RFI.

However, one has to resort to a *zero DM* de-dispersion to get a correct spectral index of the 50 Hz interference. Otherwise, the actual non-zero DM spreads out the 50 Hz sinusoids, the extent of which depends on the frequency of observation. For high DM, it is quite possible that more of the powerline harmonics is left at the higher frequencies, and we may obtain a laterally inverted spectral index. This can be overcome by studying the single channel data itself or time series de-dispersed with zero DM, corresponding to simultaneous multi-frequency observations.

4.2.2 Time scales of powerline frequency variability

The motivation for the project stems from the fact that these powerline frequencies do not stay at a fixed frequency, but slowly undulate about the mean 50 Hz. The same applies to its harmonics. Therefore, the most important study con-

ducted towards understanding the interference was to quantify the time scales over which these frequencies stay reasonably stationary.

The criteria to decide on the number of samples to FFT was to estimate the time scale over which the movement of the powerline frequency was minimal, at the same time bearing in mind to have sufficient bin resolution. It would be better to keep the FFT length as small as possible, because the smaller the FFT length, the fewer the time samples considered and lesser is the movement of the 50 Hz line. However bin resolution of the spectrum should not be sacrificed for stationarity of the 50 Hz line. Finer bin resolution is required to effectively filter out only the powerline interference and effectively retain as much of the pulsar harmonics in the neighbourhood, if any, as possible.

After many trials, it was decided to fix the time scale of stationarity as 8/16 seconds, which correspond to an FFT length of 16384/32768 for 32 integrations(corresponding roughly to 2 KHz sampling rate). This should allow us to track any movement of the 50 Hz line effectively, as well as achieve sufficient bin resolution. The following table must help.

Table 4.1: Optimal FFT lengths for different sampling rates

No. of integrations	Sampling rate	Signal duration	FFT length
32	2 Khz	16s	32768
16	4 KHz	16s	65536
8	8 KHz	16s	131072
4	16 KHz	16s	262144

Chapter 5

Filtering de-dispersed data

Now that the time scale over which the powerline frequency remains reasonably stationary is known, the data can be segmented into blocks of that duration for filtering. Three methods of filtering are devised, each more effective and elegant than the previous.

5.1 The Brute Force Method

This is a very simple and straightforward approach to removing interfering signals. The algorithm is as follows:

- The data is sectioned into blocks of length corresponding to 16 seconds.
- Each block is Fouriér transformed and magnitude spectrum is obtained
- The frequencies 50 Hz, 100 Hz,etc., are located and a few neighbouring bins on either side of each frequency are replaced by zeros.
- The block is inverse Fouriér transformed to get back the time series.

A close look at the algorithm and some reflection reveals that sectioning the data into blocks is meaningless, as it is a fixed notching process. However, since it is computationally more efficient to Fouriér transform multiple blocks of smaller lengths of data than a single large block, sectioning is resorted to.

There are two disadvantages to filtering data in this manner.

- The Paley-Wiener criterion states that that for a filter to be causal and realizable, it should satisfy the following inequality:

$$\int_{-\infty}^{\infty} |\ln H(\omega)|\, d\omega < \infty$$

This integral does not converge for an ideal filter like the one above which has zeros over a finite band of frequencies.

- An abrupt change in one domain manifests as an infinite ringing in the other domain for Fouriér transform pairs. So zero-padding is not an effective way of filtering undesirable interference lines since the ringing introduced in the time series is as bad as the interference itself. This may not be very obvious in the time series but theoretically this is true; and every care must be taken to avoid such an ad-hoc method of filtering.

- It has been seen that the powerline frequency varies about the mean 50 Hz, and so do the harmonics about their respective means. So when these move away from the notches defined for zeroing, they are not at all rejected. This is as bad as not filtering the data at all, since portions of the time series are still polluted with the powerline interference.

But the advantages as such are not too few. If we can make the notches sufficiently broad – broad enough to take into account the leftmost and rightmost extents of the powerline frequencies throughout the data, then we can be sure that there is no remnant of these interfering sinusoids in the filtered data. This is computationally a very simple approach. But making the notches so broad may result in some of the pulsar harmonics to be rejected as well, a prospect that is nightmarish enough for the pulsar researcher. Sacrificing a few Fouriér coefficients may result in the pulse shape getting distorted, which though may not be severe, may be quite sufficient to do misleading physics.

5.2 A Better Method

A slightly better method of filtering is now described. The algorithm is almost the same, except for replacing interference lines with zeros. Simple though it may appear superficially, some explanation is proper at this point. The spectrum of the pulsar data series exhibits three distinct features.

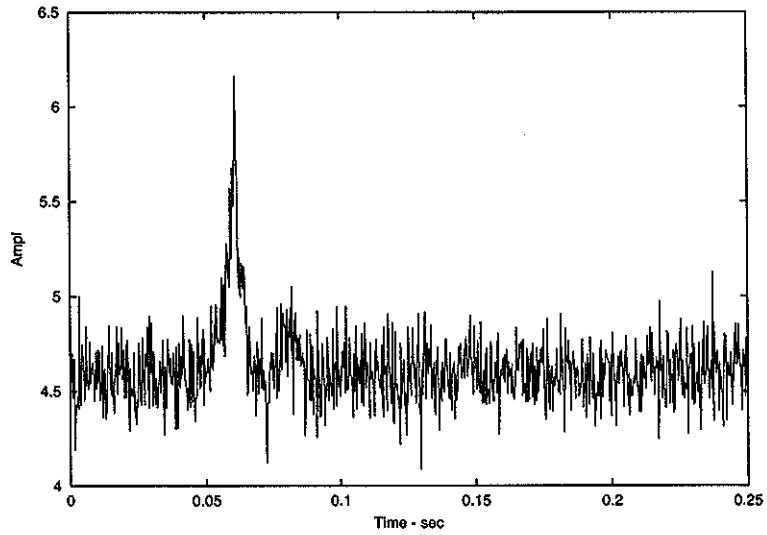


Figure 5.1: b1133+16 time series before filtering

1. The spectral envelope of the actual pulses.
2. The floor noise which can be considered almost white, except for the slightly higher power level at the red portion of the spectrum. This contribution comes chiefly from the slow gain variation of the electronics. The floor noise itself comes from the electronics, and is hence Gaussian in nature. Hence it manifests as a Gaussian noise in the frequency domain as well, since a Gaussian random variable remains Gaussian under any linear transformation.
3. Interference signatures if any.

Therefore when replacing the interfernce lobes with some interpolating values, care must be taken that the statistics of these interpolating samples be commensurate with those in the neighbourhood. So, an estimate of the statistics is obtained from a few bins that are sufficiently proximal. The complete algorithm is given below for clarity.

- The data is sectioned into blocks of length corresponding to 16 seconds.
- Each block is Fouriér transformed and magnitude spectrum is obtained, which is used only for the computation of the statistics.
- The frequencies 50 Hz, 100 Hz,etc., are located. Bins from 5 Hz to 10 Hz to the right of each powerline harmonic, including the fundamental, are

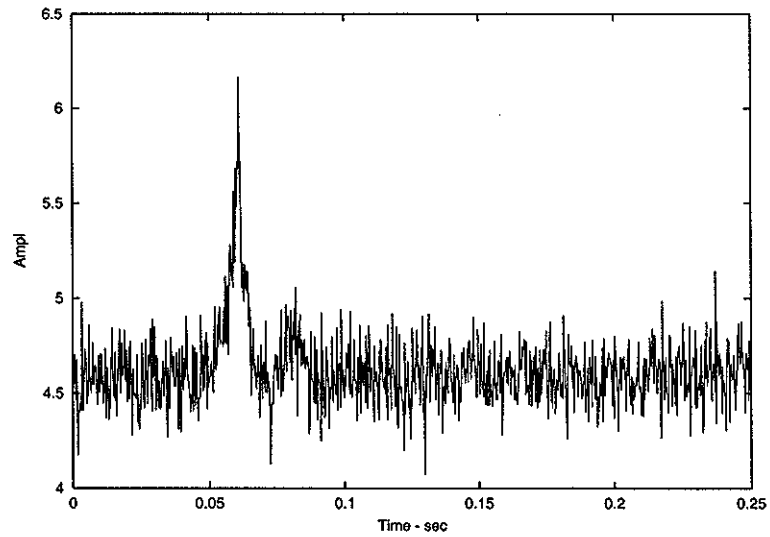


Figure 5.2: b1133+16 time series after filtering

considered and the sample mean and sample variance of these samples are calculated.

- A pair of uniformly distributed random variables is generated using a *linear congruential generator* implemented as a C subroutine. This pair of uniform deviates is tranformed into mormal(Gaussian deviates using the *Box-Muller* method,

$$y_1 = \sqrt{-2\ln x_1}\cos 2\pi x_2 \tag{5.1}$$

$$y_2 = \sqrt{-2\ln x_1}\sin 2\pi x_2 \tag{5.2}$$

- The pair of random variables thus generated replace the real and imaginary parts of the Fouriér transform values at the frequency bins defined *apriori.*

- The block is inverse Fouriér transformed to get back the time series.

This method scores better than the previous one in terms of theoretical feasibility and practical effectiveness. The discontinuity introduced in the spectrum associated with the previous method has been overcome by interpolating the bins containing the interference lobes. This effectively filters out the powerline interference without causing undesirable ringing effects in the time series. However, this method requires slightly more computational effort to estimate the floor noise statistics and generation of random numbers – a worthwhile effort which is of no concern in an offline solution to a so serious a problem as powerline interference.

But the most outstanding issue still remains to be addressed. How do we filter out the interference when it does not fall into the designated bins as it frequently happens? This method does not take care of movement of the interference lobes; hence data blocks containing powerline harmonics that have moved away from the designated bins are not filtered at all.

5.3 Variability of Powerline Frequencies

The powerline frequencies exhibit slow variability, and their characteristics have been studied for quite a few data files. There appears to be a strong correlation between the strengths of the interfering lines and the frequency of observation. For low DM pulsars, the powerline interference is stronger at the lower frequencies of observation. Typically, at around 325 MHz, it appears very strong, getting progressively weaker at higher frequencies. However, the reverse appears to be true for high DM pulsars.

Since most of the results are quantitative, they are given in the form of plots.

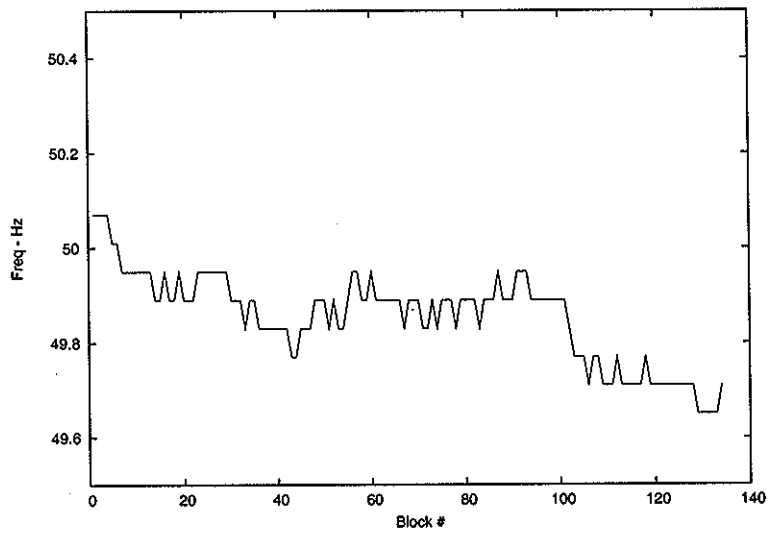


Figure 5.3: Variability of 50 Hz in b0943+10 data

There are two features that are striking about the trajectory of the 50 Hz interference line.

- There is a trend which shows a slow variation of the frequency over a very

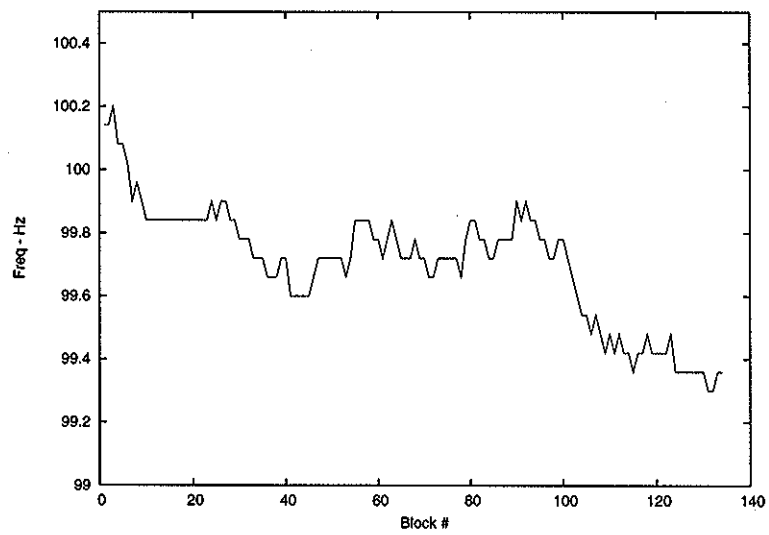


Figure 5.4: Variability of 100 Hz in b0943+10 data

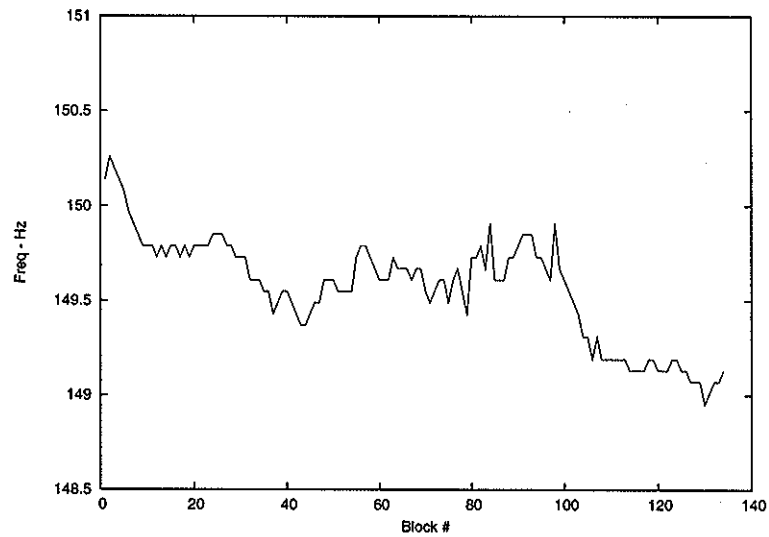


Figure 5.5: Variability of 150 Hz in b0943+10 data

large interval of time. This trend will be apparent if we try to fit a running median to the curve.

- The rapid variations that manifest from block to block - which is a more local phenomenon. This comes from the finite bin resolution of the FFT. The frequency is actually varying continuously given by the trend, but the finite bin resolution - 0.05965 Hz/bin in this case, and sectioning the data into blocks result in granulation.

The same can be said for the higher harmonics as well, though the granulation will be more rapid. Though the general trend will look almost the same, it will be scaled for the harmonics - the scaling factor being equal to the order of the harmonic.

5.4 The Tracking Filter

We now therefore have a very strong case for a filter that follows faithfully the movements of the powerline frequencies and replaces them with noise. Though not adaptive in the strict sense of the word, the filtering algorithm that is proposed tries to adapt the notches in accordance with the movement of the harmonics about their current locations.

The general structure of the filtering routine as a whole remains unchanged. The quintessential *tracking algorithm*, which is the most important addition to the previous algorithm is explained in all its detail below.

1. The powerline frequencies are located in the Fouriér magnitude spectrum for the first block using the apriori values defined in an input parameter file.

2. A local search is made around each powerline harmonic that is marked for filtering, for a few bins on either side and the peak magnitude is found out. The peak magnitude and its bin number are both collected in different variables.

3. Starting from the peak bin and moving to the right, the value of the magnitude is compared with a threshold. If the value exceeds the threshold, the real and imaginary parts of the spectral magnitude in that bin are replaced by Gaussian noise. This is continued as long as the magnitude at that bin is above the threshold. Once the threshold is reached, this procedure is stopped.

4. This is carried out on the other side of the peak bin as well and the whole process is repeated for all blocks of data from steps 2 through 5

5. The bin number corresponding to the peak is made the central bin around which to hunt for the peak for the next block of data. The peak magnitude is gathered in a file for statistical study.

Step 5 always ensures that the peak bin remains within the few bins around which to hunt in every subsequent block. Thus an effective way to track the powerline frequencies has been devised. This is chosen as the standard method to filter out the interference because of its obvious superiority over the previous two filtering methods.

- There is no abrupt discontinuity in the spectrum as seen in the discussion of the first method.
- There is no remnant of the powerline interference because all of it has been removed by the tracking them and replacing them with the interpolating Gaussian noise.

The improvement in the qualilty of the data after filtering is apparent from the plots given on page 35[1]. However, it is difficult to qualitatively obtain a measure of the improvement.

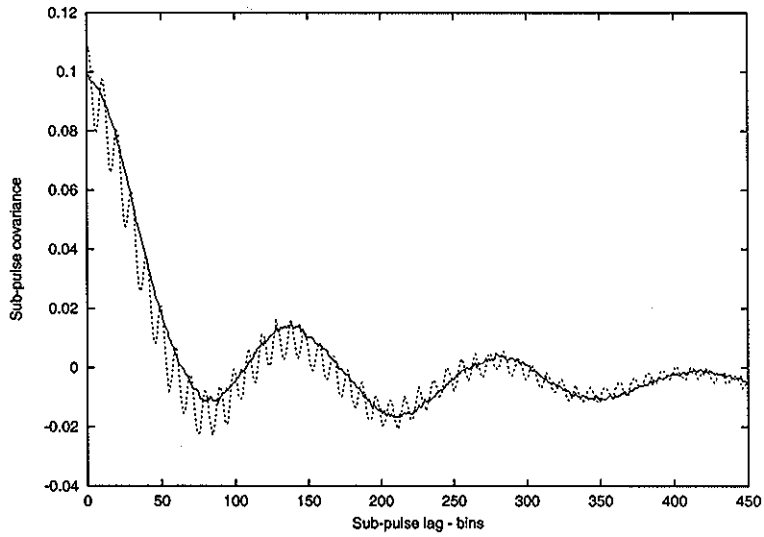


Figure 5.6: Sub-pulse covariance before and after filtering

The sub-pulse evolution is clearer after filtering out the powerline interference and this should be obvious from the two plots [2] given on pages 36 and 37 for the sub-pulse evolution of pulsar p0826-34.

But an unambiguous test of the effectiveness of the filtering method just described is the sub-pulse covariance plot. It was earlier discussed in section 4.1.3

[1] Courtesy: Y. Gupta

[2] Courtesy: B. Bhattacharyya

on page 21 that pulsar astronomers study the sub-pulse structures to understand the emission mechanism of a pulsar. Any presence of 50 Hz and its companion harmonics will severely degrade the estimate of the sub-pulse covariance. The sub-pulse covariance plot for a single pulse is shown before and after filtering.[3]

Before filtering, the covariance plot has ripples caused by self-convolution of the 50 Hz sinusoids. However, the absence of interference should result in no such features, and the sub-pulse covariance plot after filtering is indeed devoid of them.

[3]Courtesy: B.Bhattacharyya

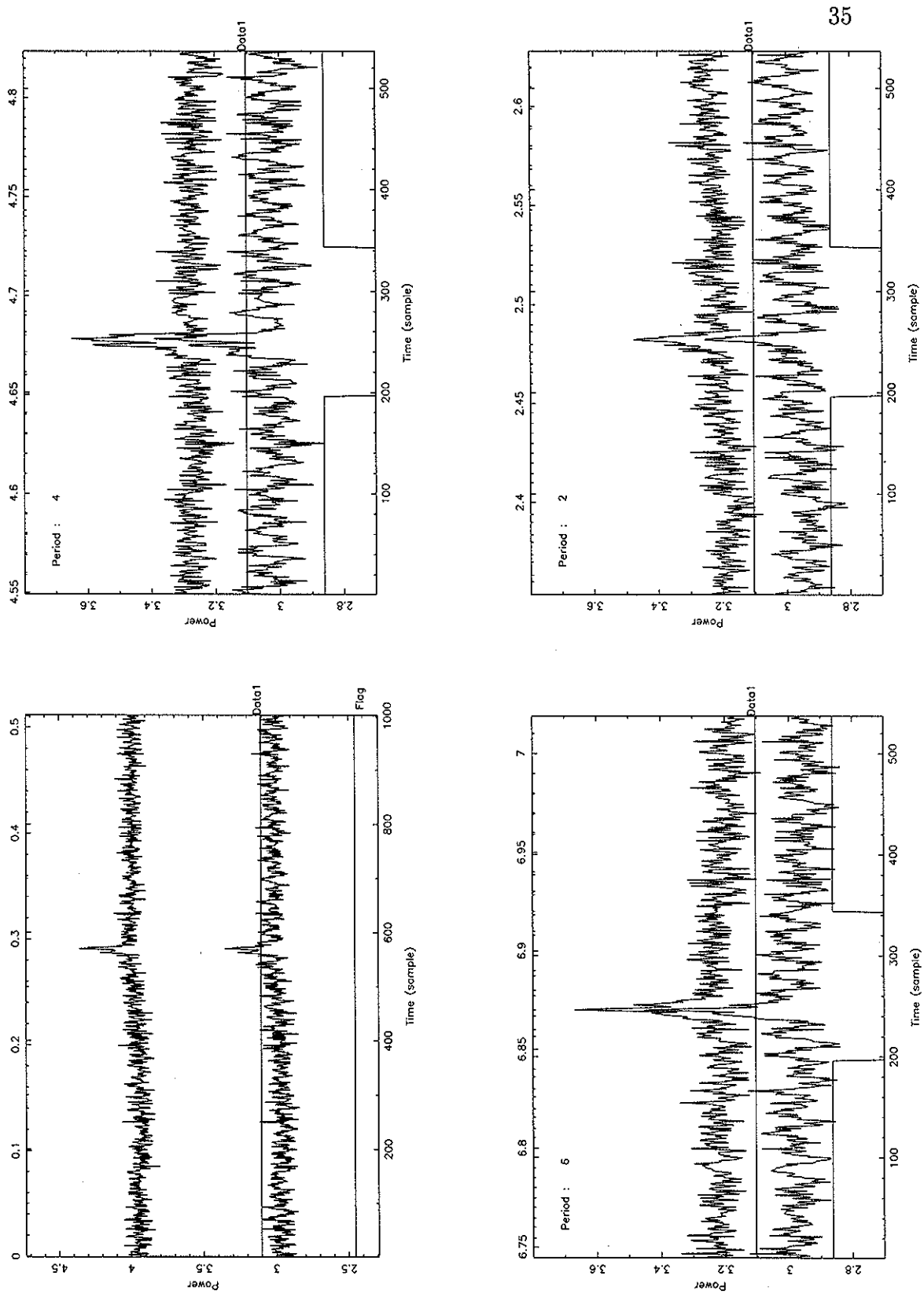


Figure 5.7: Single pulse profiles for b0943+10 before and after filtering

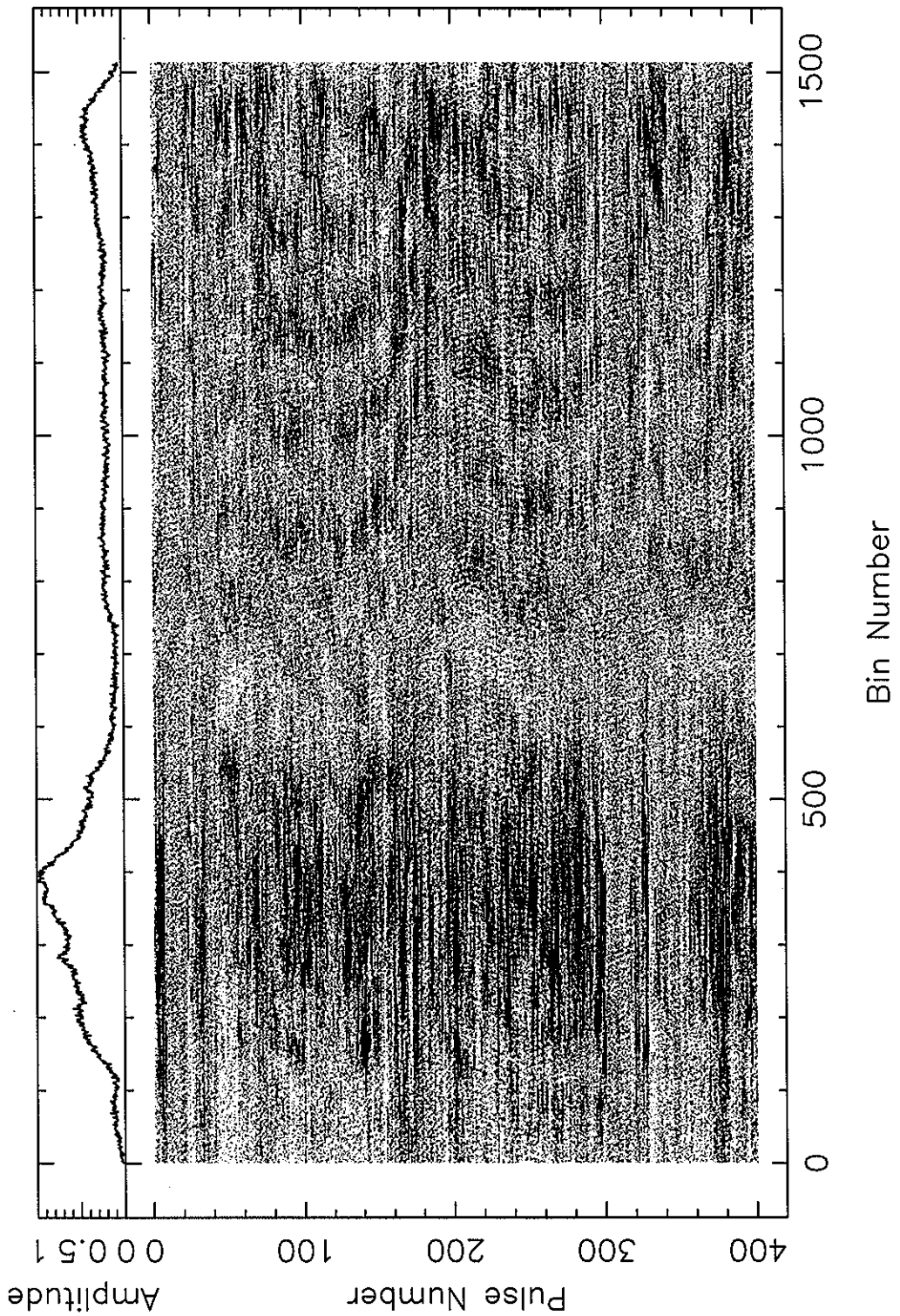


Figure 5.8: Subpulse evolution profile before filtering

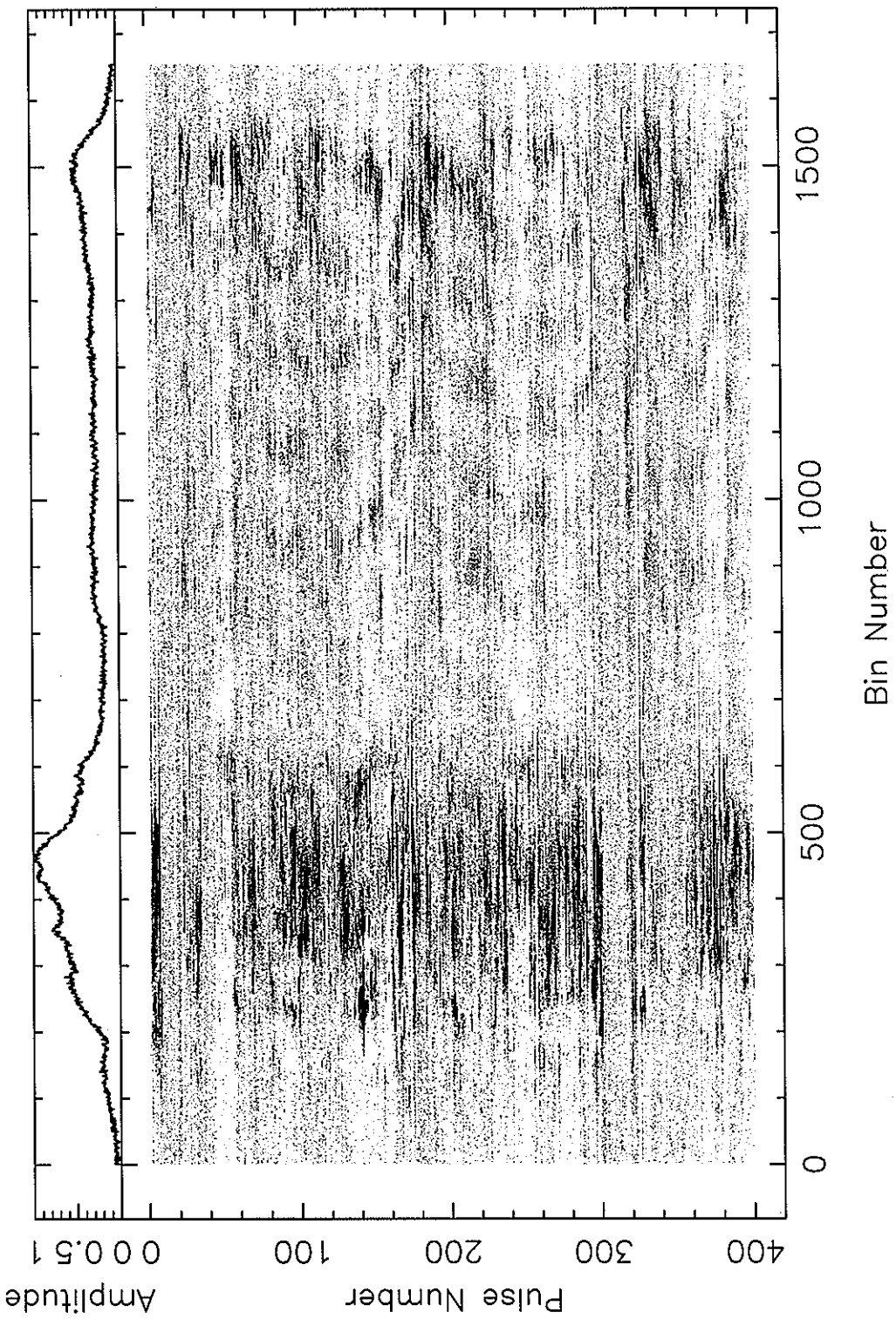


Figure 5.9: Subpulse evolution profile after filtering

Chapter 6

Single Channel Filtering

Filtering becomes more effective if it is done earlier in the data processing chain. Filtering de-dispersed data was found to be effective. But filtering prior to de-dispersion can be better for the following reasons

- De-dispersion adds signals from all channels without taking into account the contribution of individual channels to the powerline sinuoids of the final de-dispersed data. As a consequence, even an RFI channel modulated by 50 Hz is considered for de-dispersion, unless it is flagged. At the end, this channel contributes substantially to the total powerline interference in the de-dispersed data. Instead, if the interference is rejected at the single channel level itself, de-dispersion later could give better results.

- Filtering before de-dispersion is better for the other reason that the pulse power in each channel is lower than the pulse power in the de-dispersed data. But the powerline sinusoid and its harmonics are appreciably stronger than in the de-dispersed data because *de-dispersion* has *de-dispersed* the data but has *dispersed* the powerline sinusoids from all the channels. Therefore, it is easier to locate and track the powerline sinusoids in the single channel data format since the powerline interference is very strong. This is particularly effective if we are dealing with a very strong pulsar(like b1133+16), in which case it is very difficult to track the powerline sinusoids properly because the pulsar harmonics in the neighbourhood are quite strong and we could wrongly conclude one as a powerline harmonic instead.

- The powerline sinusoids across the channels is highly correlated, and as a

result the gaussian noise riding on the powerline interference is also correlated. The central limit theorem applies only to independent random variables. However, if the individual channels are devoid of 50 Hz interference, the noise across the channels is uncorrelated. If now the data is dedispersed, we can expect a $sqrtN$ improvement in the SNR, N begin the number of channels.

6.1 Filtering - the method and its results

The main algorithm for filtering single-channel data is exactly the same as the improved and final version of filtering algorithm in place for the de-dispersed data. Though, the program required certain modifiations to take into account the raw data format(as the single channel data is called), and accounting for the GPS bit in the two different array modes, namely IA and PA. The following changes have to be noted for the raw data case:

1. The data type is *short int*, and it is stored inverted.
2. The data is available sample by sample, where 256 channels are available one after the other in serial fashion for each sample
3. All samples across all channels have a GPS bit, the location of which is **MSB** for IA mode and **LSB** for PA mode

An outer loop in the program runs over blocks of data, each block of data being 32768/65536 samples(depending on the sampling rate) with 256 channels each. The inner loop runs over all the channels, each channel consisting of 32768/65536 samples. On each such block of data the filtering algorithm operates on and writes the output in the same format as the input raw data. There is an intermediate inversion and GPS bit-masking followed by typecasting to float so as to enable Fouriér transformation(since it requires floating point numbers to represent numbers between 0 and 1). The operations are repeated in the reverse order, except bit-masking(as it makes no more sense) to retain the data format in the output file. The number of data samples to consider in a block and the number of harmonics to filter out are read as input from the input parameter file.

The raw file as such is of little value as the power in a single narrow channel is hardly sufficient to show pulses for most pulsars. With de-dispersion comes

the improvement, and the effect of filtering on the raw data can be tested by de-dispersing the original raw data file and the filtered raw data file separately, and comparing the single pulse profiles and their signal-to-noise ratios.

6.2 Statistics from Single-channel data

The statistics of the interfering frequencies must give us a very useful insight into their behaviour. The filtering program gathers some statistics in a file with the same name as the input file, but with an extension *notch.history* in the same directory as the input raw file. The various plots are given in the following few pages so that we may get a more complete understanding of the workings of the interfering powerline frequencies.

6.2.1 Variability of powerline frequencies

The following plot corresponds to the movement of the 50 Hz line in the frequency axis, for all the 256 channels block by block. The cross-section viewed from the channel axis gives a picture almost similar to figure 5.3 on 30. The trend for the

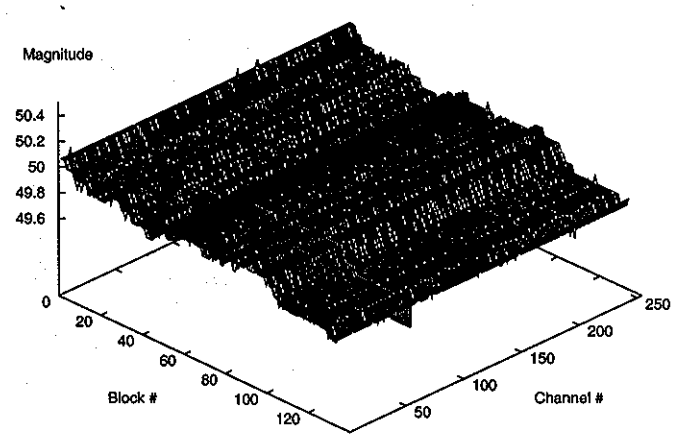


Figure 6.1: 50 Hz variability across 256 channels

harmonics is similar except for the more local variations. But the general trend of the curve in figure 5.3 is preserved across all the channels.

6.2.2 Strenghts of powerline frequencies

The strengths of the interfering sinusoids also varies with time. The contour plots for the strengths for the frequencies 50 Hz, 100 Hz, 150 Hz and 200 Hz across all 256 channels are given below, as a function of the block count. Some striking features are noticeable from the contour plot. Notice the strong trajectory at channels 41 and 42. Besides, there is a diagonally sweeping pattern very prominent from blocks 20 to 50, about 30 channels across. This could be an RFI source modulated by the powerline sinusoid, whose centre frequency is gradually changing. There are also some irregular clumps that could be RFI modulated by 50 Hz, briefly very strong at those areas.

The pattern looks similar but clearer in the frequencies 100 Hz, 150 Hz and 200 Hz. The diagonal trajectories are better pronounced. All the plots have a very strong and stable trajectory at channels 41 and 42. But is this an RFI source radiating at a fixed radio frequency, modulated by 50 Hz and its harmonics? The only way to ascertain this is to get an averaged bandshape for the whole data and compare the same with these plots.

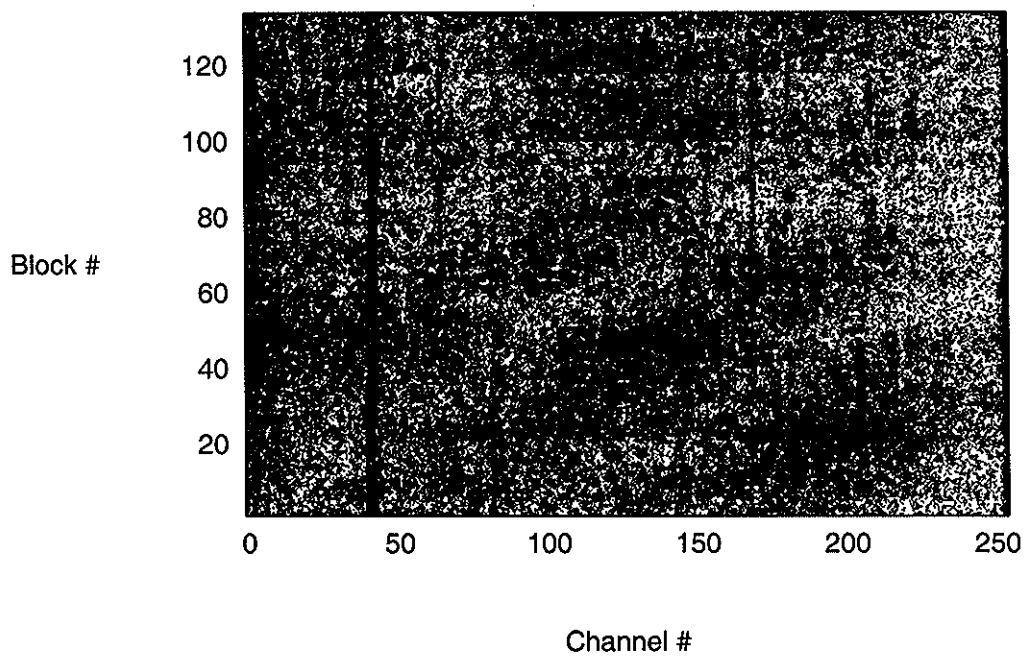


Figure 6.2: 50 Hz trajectories across 256 channels

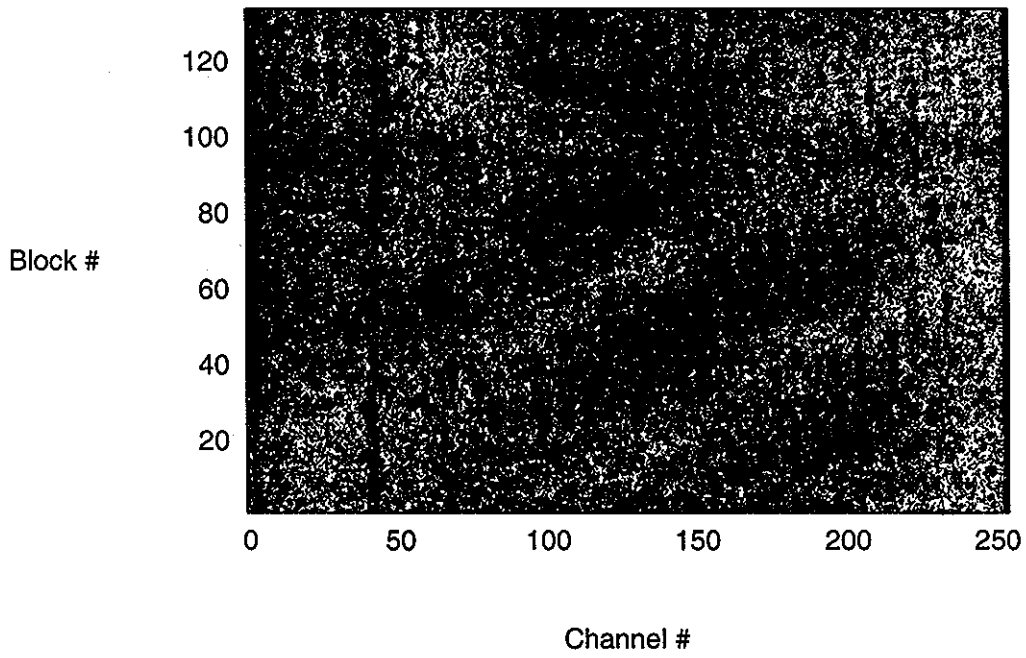


Figure 6.3: 100 Hz trajectories across 256 channels

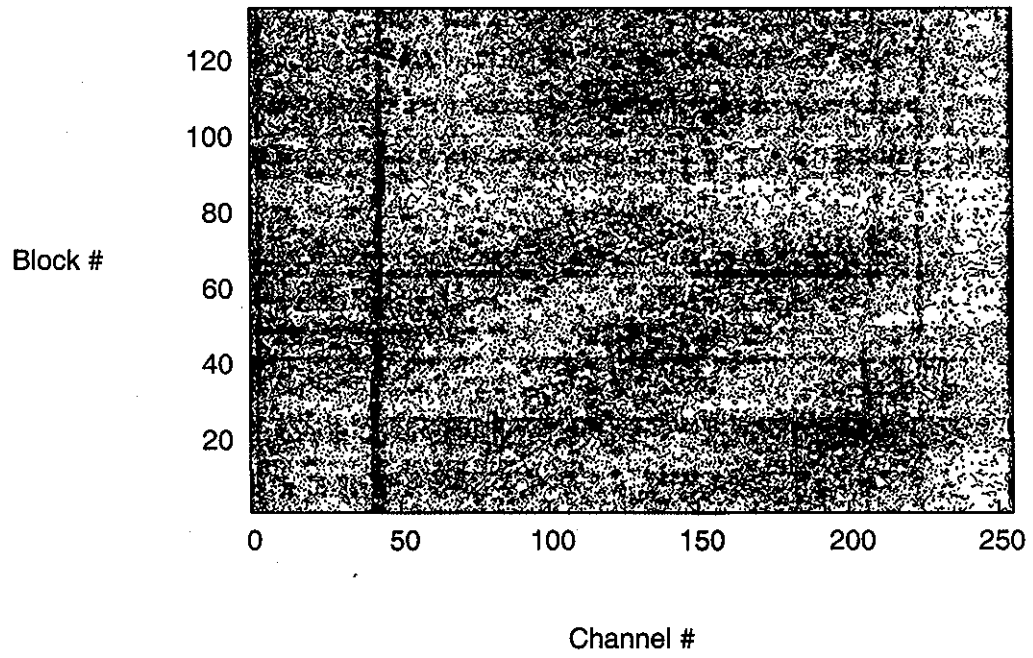


Figure 6.4: 150 Hz trajectories across 256 channels

6.3 The entry point of 50 Hz

It is now established that pulsar observational data obtained with the GMRT is polluted with powerline interference. But what is the mechanism by which this

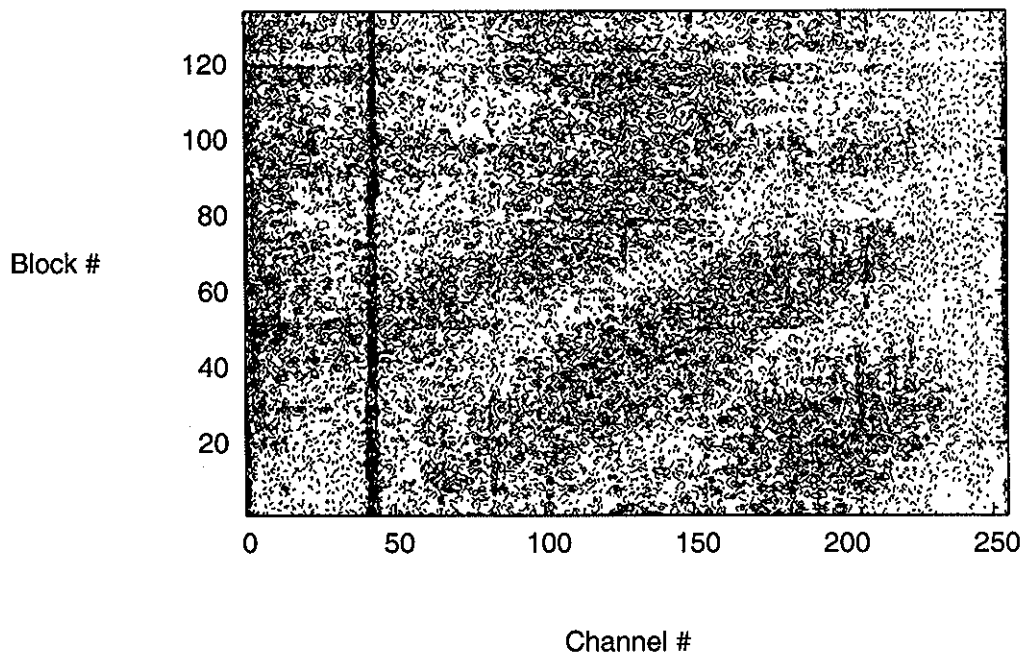


Figure 6.5: 200 Hz trajectories across 256 channels

interference enters the data? Is there a correlation between the averaged bandshape of this data and the powerline trajectories shown in the plots above? This is an interesting question, and some inferences can indeed been drawn from such a comparison. Figures 6.6 and 6.7 give the grand bandshape plots[1] of the b0943+10 raw data before and after filtering. Since the raw data is in a very convenient format that allows us to deal with the data sample-by-sample for all channels, the grand bandshape is simply obtained by averaging all the samples taken 256 channels wide. Comparing the grand bandshapes before and after filtering, and considering the powerline trajectories in figures 6.2 to 6.5, the following inferences can be drawn about the powerline interference.

- The grand bandshapes before and after filtering look very similar, although they are imperceptibly different. This shows that filtering has very little effect on the bandshape of the raw data. The bandshape should not be confused with the spectrum of the pulsar data.

- Corresponding to channels 41 and 42 in figures 6.2 through 6.5, there is a small spike in the same channels in the grand bandshape plot. This

[1] Courtesy: B.Bhattacharyya

means that the powerline interference could actually be modulating the RFI in channels 41 and 42, although the spike is not very strong in the grand bandshape plot. This means that it is a low power RFI, but strongly modulated by 50Hz and its harmonics.

- There are other strong spikes in the grand bandshape plots that do not show up in the contour plots of the powerline trajectories. These could be RFI channels that are not modulated by the powerline sinusoids.

- Since filtering has little modified the grand bandshape plot, it means that the powerline interference is a broadband phenomenon. This is substantiated by the fact that all channels showed significant powerline frequency and its harmonics in the Fouriér domain. The plots in figures 6.2 through 6.5 also show broadband powerline interference apart from the strong conspicuous patterns.

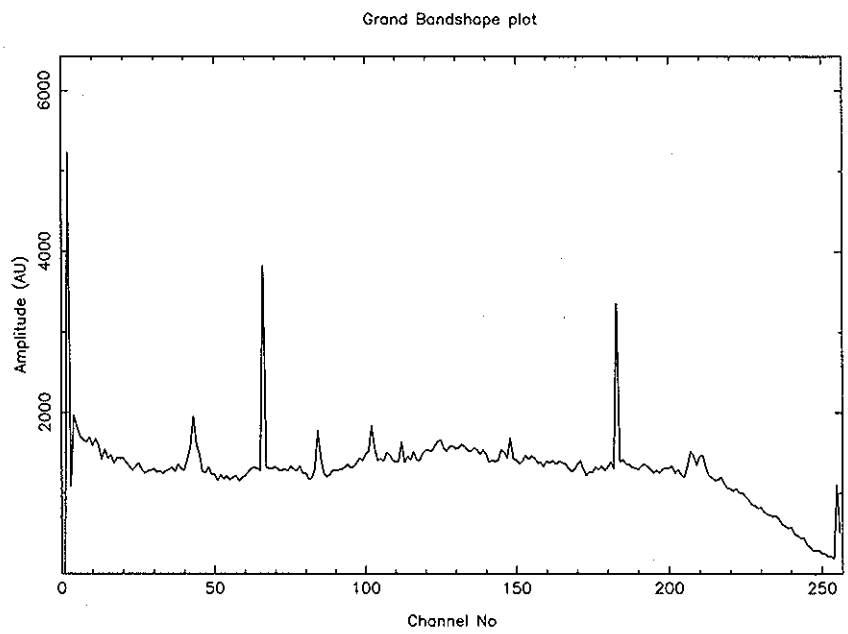


Figure 6.6: Grand bandshape for b0943+10 data before filtering

The broadband nature of the powerline interference can be explained by its origin itself. The high tension lines carrying electrical power carry a sinusoid of frequency 50 Hz. Since the voltages carried by these lines are of the order of a hundred kilovolts, there is a breakdown of the surrounding air every sinusoidal cycle. This breakdown causes braodband electromagnetic radiation; and if this

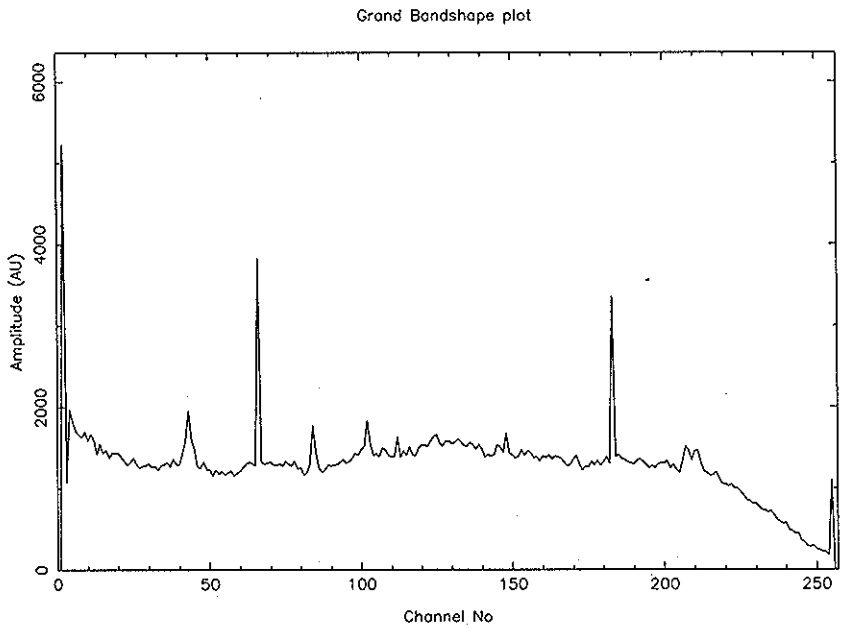


Figure 6.7: Grand bandshape for b0943+10 data after filtering

happens at both the peaks of the sinusoidal cycle, then 100 Hz is the observed data. This phenomenon has been occasionally observed. Besides, the higher harmonics also result.

Many other data files were observed, for different pulsars and at different frequencies of observation. It was observed that higher frequencies of observations behave better than the lower frequencies in terms of strength of powerline interference. As an example, the following plot gives the 50 Hz trajectory for the pulsar b0031 observation at 610 MHz. Unlike the lower frequencies, there is not much broadband 50 Hz power in the data. There are occasional stripes, which could be some sources of RFI modulated by 50 Hz, that were switched on during that period.

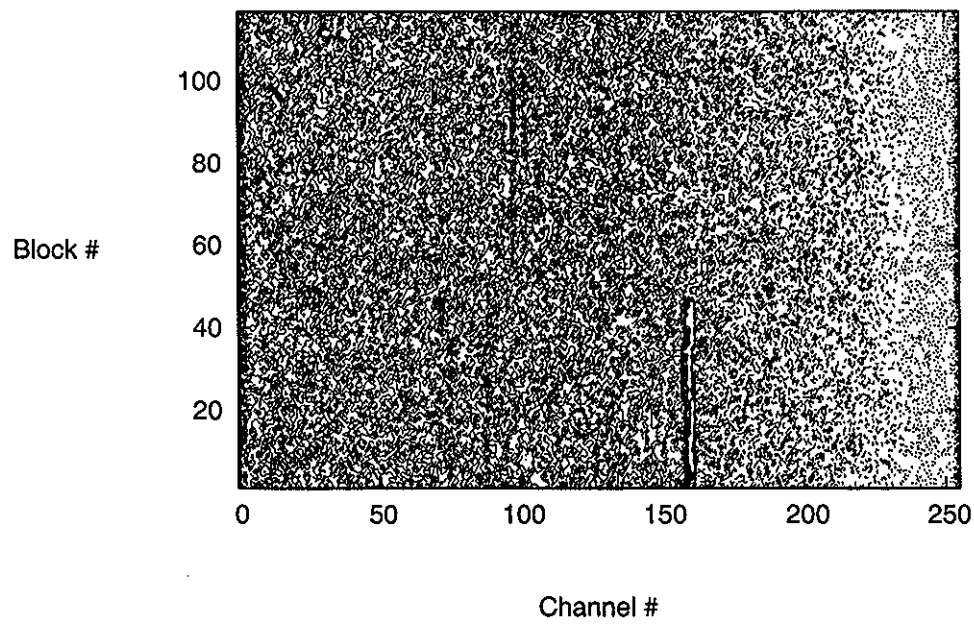


Figure 6.8: 50 Hz trajectory for pulsar b0031 at 610 MHz

Chapter 7

Real-Time Filtering

Powerline interference can be filtered offline and good results have been achieved. However, it would be good to strike the problem at the root and reject the interference at the earliest opportunity in the data flow chain. This amounts to filtering the interference real-time even before it is recorded for analysis. The motivation for such a filtering scheme, the requirements on the data acquisition system in such a case and the effectiveness of this scheme are the contents of this chapter.

7.1 Motivation for Real-Time Filtering

7.1.1 Some mathematics

Inorder to deal qualitatively with the problem in hand, we need a well-defined model. At the outset, it is defined as follows:

$$H_1 : r(t) = s(t) + n(t) \tag{7.1}$$

$$H_0 : r(t) = n(t) \tag{7.2}$$

Hypothesis H_0 is called the null hypothesis and has a prior probability P_0; H_1 is called the alternative hypothesis with a prior P_1.

We decide on hypothesis H_1 if a signal is detected or the hypothesis H_0 if not, depending on the likelihood ratio exceeding the threshold τ or not.

$$L[r(t)] = \frac{p_1[r(t)]P_1}{p_0[r(t)]P_0} \gtrless \tau \tag{7.3}$$

p_1 is the conditional p.d.f. of hypothesis H_1 and p_0 is the conditional p.d.f. of the hypothesis H_0. Since the additive noise is from the front-end electronics of the radio telescope, it is natural to proceed further under the Gaussian framework. Hence the noise p.d.f. is given by

$$p\left[n(t)\right] = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{n^2(t)}{2\sigma^2}} \tag{7.4}$$

assuming zero-mean noise. Henceforth, explicit mention of t will be suppressed for notational simplicity. Moreover, the ratio of priors is taken across the equality sign; consequently $\tau = P_0\tau/P_1$.

$$L\left[r\right] = \frac{p\left[s\left(t\right)\right]}{p\left[r\left(t\right)\right]} = \frac{p_1\left[r-s\right]}{p_0\left[r\right]} \tag{7.5}$$

now therefore becomes

$$L\left[r\right] = \frac{e^{-\frac{1}{2}\frac{(r-s)^2}{\sigma^2}}}{e^{-\frac{1}{2}\frac{r^2}{\sigma^2}}} = e^{\frac{r^2-(r-s)^2}{2\sigma^2}} \tag{7.6}$$

$$= e^{\frac{1}{2}\left[\frac{2rs-s^2}{\sigma^2}\right]} = e^{\frac{1}{2}\left[\frac{s(2r-s)}{\sigma^2}\right]} \tag{7.7}$$

Since the natural logarithm is a monotone increasing function,

$$\ln\left[L[r]\right] = \frac{rs - s/2}{\sigma^2} > \ln\tau = \tau' \tag{7.8}$$

$$= s\left(r - \frac{s}{2}\right) > \sigma^2 \ln\tau = \tau' \tag{7.9}$$

Therefore, a sample from the observed data is taken and compared with a threshold i.e.,

$$r > \frac{\sigma^2\tau'}{s} + \frac{s}{2} \tag{7.10}$$

for hypothesis H_1 to be favoured. The probability of detection of the pulsar signal is evaluated as

$$\int_r^\infty p_1\left(r\left(t\right)\right) dr(t) \tag{7.11}$$

Simple though it may appear superficially, a careful and closer study of the mathematical analysis will reveal practical difficulties in retrospect, which are enumerated below:

1. It requires prior knowledge of the signal $s(t)$, its period and instantaneous values.

2. It also requires prior estimate of the statistics of the noise $n(t)$, though this is not a difficult or exacting requirement.

Despite these difficulties, some generalizations can be introduced at this juncture that can lead to considerable simplification of the atmosphere for further progress.

If $r(t)$ is the instantaneous time sample of a thermal noise process, then it can be hazarded without loss of generality that it is a sample from a noise process which generates independent and identically distributed random variables. Under such an assumption, the dependence of the threshold on time can be dropped. 7.10 becomes independent of time and 7.11 can be rewritten as

$$\int_r^\infty p_1(r)\,dr \tag{7.12}$$

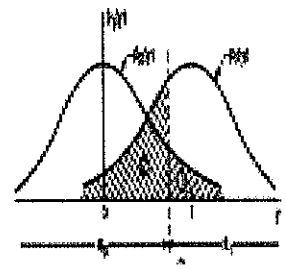


Figure 7.1: The binary hypothesis testing problem

For a picture like this, the threshold can always be expressed as $p\sigma$, where p is a positive real number corresponding to the point of intersection of the two conditional p.d.fs for a simple binary hypothesis testing case with equal priors. When the priors are not equal as in the present case, $p\sigma$ is not the abscissa of the point of intersection[9].

7.1.2 One-bit encoding for pulsar search

Pulsar search is carried out by looking for signatures of periodicities in the data. However, when search data is recorded, it is quite uneconomical to record data in the full 16 bit format. It can be represented by as few bits as possible, but this is possible only if there is no d.c. bias in the signal. D.C. bias contributes to the highest encoded bit in the data; hence removing the mean will be a significant

step towards more economical data storage. If the data is now encoded to 1 bits after subtracting the running mean from a few samples of the data in a buffer, the event to look for is the regular and periodic occurence of a string of '1's, because all positive excursions above zero will be encoded as '1' and all negative excursions below zero will be encoded as '0'.

However, after bit-packing, the calculation of the probability density function for the encoded data becomes extremely diffcult analytically due to the fact that one-bitting is akin to a hard-limiter(or signum function for the mathematically inclined) which is a classic non-linear operation. Due to the analytical difficulty presented in calculating the probability of detection of a string of ones through the Bayesian, or likelihood ratio, approach, it would be better to obtain bounds on performances. For the present case, this problem becomes fantastically simple.

The Chebyshev inequality is introduced now. Because of its great generality, this limit is often too low to be more than a rough guide in determining probabilities in particular cases, but such inequalities are invaluable in proving limiting properties.

$$P[|X-\mu| \geq \lambda\sigma] \leq \frac{VarX}{[\lambda\sigma]^2} \tag{7.13}$$

which gives a loose upper bound on the probability of a random variable overshooting the $\lambda\sigma$ limits on either side for a two-sided distribution. Now one has to invoke the ergodic theorem to effectively deal with time averages rather than statistical averages. This is true because the law of large numbers can be safely invoked where observations consisting of a few million samples are considered. Armed with these invocations we reinterpret 7.13 now. The upper bound on the probability of finding a noise excursion beyond the $\lambda\sigma$ limit on either side at most is given by

$$\frac{VarX}{(\lambda\sigma)^2} \tag{7.14}$$

or

$$\frac{VarX}{2\,(\lambda\sigma)^2} \tag{7.15}$$

at most for excursions above the positive $\lambda\sigma$ limit. In other words, when we observe above the positive $\lambda\sigma$ level, only at most $\left(\frac{VarX}{2(\lambda\sigma)^2} \times 100\right)$ % of the time do we expect the excursions to be due to noise. This means that $\left(1 - \frac{VarX}{2(\lambda\sigma)^2}\right)$ portion of the time we must observe no signal. Now consider the situation where a pulsar signal is riding on the noise, and its peak is well above the positive $\lambda\sigma$

limit. Now, except for the noise excursions above $\lambda\sigma$ there is possibly nothing other than the pulsar signal. In summary, $\left(1 - \frac{VarX}{2(\lambda\sigma)^2}\right)$ of the time, any excursion statistically corresponds to the pulsar signal. Interpreting probability as a 'relative frequency of occurence'(which is allowed under the present circumstances as already discussed), the probability of those signal samples exceeding the $\lambda\sigma$ threshold being pulsar signal is given by

$$P_D = 1 - \frac{VarX}{2\left(\lambda\sigma\right)^2} \tag{7.16}$$

For the one-bitting operation, this translates into the following. The proportion of the occurence of continuous string of '1's, the count of which exceeds a predetermined value, is given by 7.16 and the proportion of the occurences of the irregular sequence of '1's and '0's above the positive $\lambda\sigma$ threshold is given by 7.15. Thus with sufficiently high probability we must be able to detect pulsar signals that exceed even as low a threshold as 2σ. For $\lambda = 2$, the probability of correct detection of such a string of '1's is at least 87.5%. Table 7.1 gives the values for other threshold values.

Table 7.1: Probability of detection of pulsar signals for various thresholds

λ	P_D
2	87.500%
3	94.444%
4	96.875%
5	98.000%
6	98.611%

The table, though self-explanatory, is better supplemented with a graphical plot. Notice the asymptotical behaviour of probability, which is quite expected. At this point, it is worth reminding that when we speak about the probability of detection of pulsar, we are actually referring to the probablitiy of *correct* detection of pulsar signals. The probability of finding pulsar signals above the λ threshold level increases, as it can be explained in the following two ways:

1. The higher the threshold level λ, the lower do we expect the noise excursions to reach that level

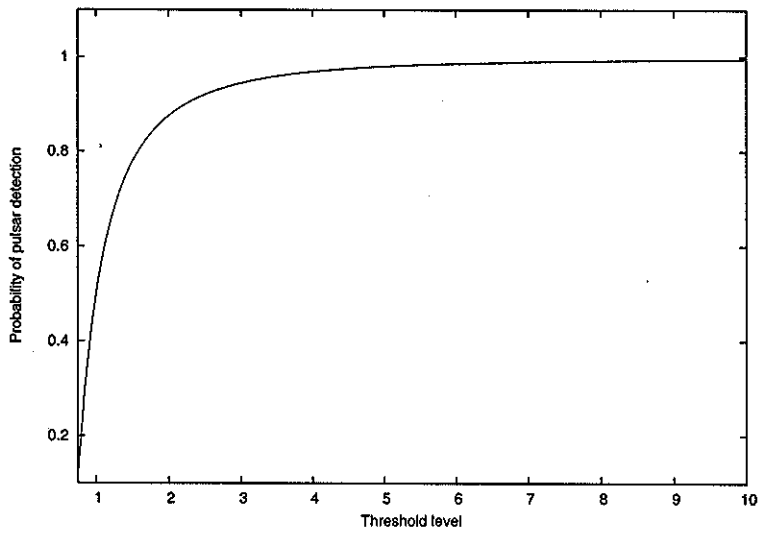


Figure 7.2: Probability of detection of pulsars as a function of λ

2. The higher threshold level also means that if the pulsar is strong enough for us to fix such a high threshold, the better is the probability of detecting it

However for weak pulsars, a lower value for λ is preferred, and even a value of $\lambda = 2$ is satisfactory. While it must be remembered that Chebyshev inequality gives only a loose lower bound, values obtained from actual observations should not be really discouraging as 87.5% is quite a promising value for probability of correct detection. The graph also provides for a pictorial guide to choosing the optimum threshold level, depending on a case-to-case basis. A value for λ between 2 and 3 should be satisfactory for most cases.

7.1.3 Preservation of SNR by one-bitting

One-bitting reduces storage space when doing search observations. But how much of the SNR do we lose because of a non-linear operation such as one-bitting? It is now shown that if the SNR of the input signal is itself small, then the loss in SNR is not severe and the degradation is asymptotic. However for the high SNR signals, the detection probability improves with signal strength inspite of a deterioration in SNR, as has been shown previously.

Again, reminding that the noise is zero-mean Gaussian, the signal plus noise model has a slightly biased mean because of the signal pulses. The bias will be

stronger towards the positive for stronger pulsars. The bias also depends on the duty cycle of the pulse. Due to the bias in the mean, the one-bitting will result in more '1's than '0's. In the case of noise only, statistically we expect an equal number of '1's and '0's. Let us call the proportion of '1's p and the proportion of '0's q, of course it being understood that $p + q = 1$. Now the contribution to the '1's by the signal is actually $p - 0.5$ because the contribution from the noise has to be decoupled. When the signal alone is present and there is no noise the proportion of '0's will be

$$1 - (p - 0.5) = 1 - p + 0.5 = q + 0.5$$

Let us now calculate the SNRs before and after the one-bitting.

Call $\langle P \rangle$ the signal power and σ^2 the noise power.

$$\mathrm{SNR_i} = \frac{\langle P \rangle}{\sigma^2} \tag{7.17}$$

After one-bitting, it will be

$$\mathrm{SNR_o} = \lim_{N \to \infty} \frac{\frac{1}{N} \sum_{i=1}^{N(p-0.5)} (1)^2}{\frac{1}{N} \sum_{i=1}^{N(0.5)} (1)^2} \tag{7.18}$$

which simplifies to

$$\mathrm{SNR_o} = \frac{p - 0.5}{0.5} \tag{7.19}$$

As a matter of fact, p is always greater than 0.5. This formula is consistent with common-sense because when there is no signal, $p = 0.5$ and the SNR=0. The maximum achievable output SNR is now only 1. Therefore even if the input SNR becomes indefinitely high, the output SNR saturates at the value of 1. However for low input SNR values for which p falls in the range (0.5,1], there is a one-to-one correspondence between the input and output SNRs as given by the straight-line portion with a non-zero slope in the plot. We conclude that for low SNR signals, there is no substantial degradation in the output SNR due to the one-bitting operation.

7.1.4 Effect of powerline harmonics

Powerline interference can affect the probability of correct detection of a string of '1's in the follwing manner: If the noise is riding on the 50 Hz sinusoid, one-

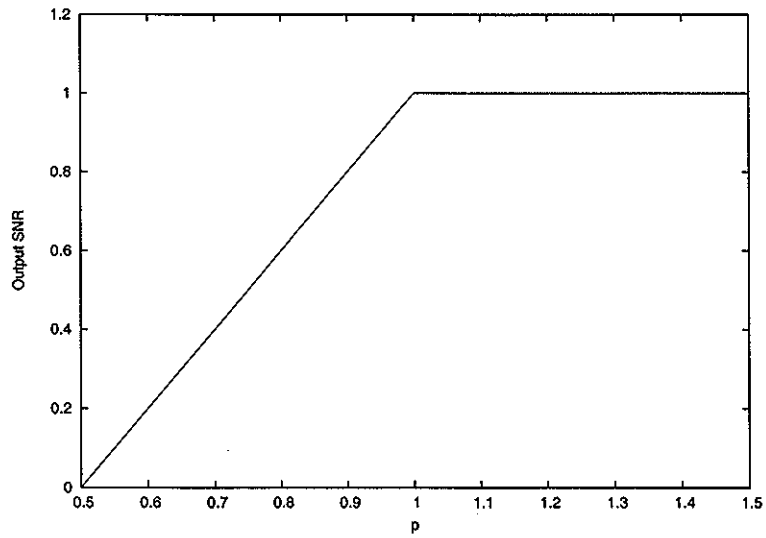


Figure 7.3: Output SNR as a function of p

bitting can play havoc with the data. If the 50 hz is extremely strong as we have found in some observational data, all one can get after one-bitting is a regular string of '1's and '0's alternating 50 times a second.

Therefore the above analysis holds good without the 50 Hz presence because it has been derived altogether ignoring the powerline interference. Hence it is a very lucrative idea to filter out the 50 Hz interference before the running mean subtraction and one-bitting is done. This will pay rich dividends in terms of probability of detecting pulsars.

7.2 Performance in Real-Time

The real-time version of the filter is tested and its performance is evaluated through various parameters like the time it takes for various values of integrations and FFT lengths. The different modes in process_psr are also explained.

7.2.1 Modes in process_psr

The program process_psr is a very flexible and feature-packed program. It can run in the following modes:

- **Through mode**: In this mode the filtering routine is bypassed. The data is acquired and various checks such as marker checks and GPS time-stamping

proceed as usual.

- **Filter mode**: In this mode, the data passes through the filtering subroutine. There are again two submodes within the filtering subroutine. It can either return *float* type data or *short int* type data. In the former, there is no extra computational effort. In the latter, however, a type conversion from *float* to *short int* is required. This introduces a substantial computational overload.

- **Bit-packing mode**: In this mode the data is buffered and the running mean is subtracted. Thereafter, the data is encoded into one, two, four, eight or sixteen bits, which is user-programmable.

Each of these modes can be invoked independently of all other modes, meaning there can be any combination of these modes. The timing estimation proceeds when all the modes are enabled, except the bit-packing which is by-passed.

7.2.2 Real-Time Processing Cost and Timing Estimation

When filtering is done in real-time, there are a lot of issues to be taken care of. The GMRT pulsar data flow chain is very flexible allowing the observer to set the sampling interval in multiples of $16\mu s$, which is given as the number of accumulations or integrations. This number is usually an integral power of two, ranging in principle from 2 to 64 or 128. The default for most pulsars is 32 integrations. However if higherresolution data is required, one may opt for fewer integrations. In this case, the real-time filtering program should be flexible enough to accommodate the changes in the number of integrations.

The filtering program for the real-time data chain is the same as the one for offline filtering. The number of blocks to filter is not calculated apriori because the filter has to run for as long as the pulsar is observed. However, the number of samples to accumulate in one block for Fouriér transforming is user-programmable. This also defines the time duration for which the powerline sinusoid is assumed stationary. The upper bound on this measure was estimated earlier through preliminary experiments as 16 seconds. For smaller lengths of FFT, the time duration also goes down and the powerline frequency is more stationary. But, the compromise for stationarity of powerline frequency is the coarse bin resolution of the FFT.

The various time estimates for the real-time data acquisition program into which the filtering algorithm has been seamlessly merged, have been measured and should serve as a rough guide to the observer for choosing the proper parameters.

The real-time data acquisition program[1] incorporates within itself as a small subroutine the real-time filtering program. But it was found that the filtering subroutine takes the lion's share of processing power. Therefore, with the filtering subroutine included, it is necessary to check whether the data-acquisition program is able to keep up with the data rates served by the radio telescope signal processing engine. To this end, a few tests were performed. Before explaining them, a brief sketch of the output formats available at the filtering subroutine follows. The filtering subroutine takes its input data as blocks of length equal to the FFT length, for all the 256 channels. The format of this data is *float*. This is absolutely essential because the FFT rountine requires *float* data type to work with sines and cosines of angles, whose values are always real-valued between 0 and 1. At the end of filtering, the filtering routine passes the filtered raw data to the process_psr program from where it was called. Now, the filtering program has the option of returning the result in either *float* format or returning it as a *short int* by typecasting it. Typecasting from *float* to *short int* is a very expensive operation as we found out, and hence the timing estimates are available for these two options separately.

When using the normal 32 or 16 integrations, the observer can safely choose a default FFT length of 32K or 64K. Since this corresponds to 16 seconds of data, there is no process lag, besides these FFT lengths also ensure sufficient bin resolution. However for smaller FFT lengths, to achieve the same bin resolution, the FFT length has to go up the same factor by which the number of integrations goes down. The load on the system becomes too heavy in this case for it to successfully keep up with the data rates. However, the bin resolution can be compromised to some extent because we are now dealing with smaller durations during which the powerline sinusoids are much more stationary than when it is for 16 seconds.

For *short int* data output from the filter

- 4 integrations - maximum FFT length = 8K

[1]Written by R.M.Dabade

- 8 integrations - maximum FFT length = 16K

For *float* type output from the filter

- 4 integrations - maximum FFT length = 16K
- 8 integrations - maximum FFT length = 64K

It is very obvious that the typecasting from *float* to *short int* takes heavy processing power; consequently the program is not able to keep up with the high data rates at fewer integrations.

7.3 Optimized FFT for accelerated performance

The volume of literature available on FFT is immense. Early work on FFT consisted of optimizing the mathematical operations, namely additions and multiplications. Many efficient algorithms for FFT like the Winograd DFT and Cook-Toom algorithm found ready applications in FFT hardware. These were available as software routines for programming on PCs as well[11]. However, with the exponential improvement of processor speeds, better algorithms have been devised that exploit the inherent architecture of the processor itself. The best known and perhaps the friendliest of such a family of FFT routines is the FFTW library of Fast Fouriér Transforms[5]. The *realft* routine takes 4 ms to compute an FFT of length 32 K, whereas the FFTW took just 1 ms. Since it is faster than the realft routine of the same length by a factor of four, it is highly desirable in a real-time environment where the system requirements are dictated by the data rate. In order to give the data acquisition program enough time to keep up with the data rate, we decided to exploit the adaptive capabilities of the FFTW routines.

Chapter 8

Results and Discussion

Filtering was performed on a few data files which had notoriously high strengths of powerline harmonics and good results have been obtained. Numerically, the performance of the filtering routine can be quantified if we have measures of SNR before and after filtering. We expect low values for SNR before filtering because of the contribution to the noise power from the 50 Hz and its harmonics. The SNR should show reasonable improvement if the interference has indeed been filtered.

The SNR before and after filtering for each of the following cases was observed and the results are summarized below in a table.

- De-dispersed data before and after filtering
- Data de-dispersed before filtering and after filtering with true DM
- Data de-dispersed before filtering and after filtering with zero DM

Table 8.1: SNR values for average pulse profile of pulsar b0943+10 before and after filtering

Mode	**Peak SNR**	**Best SNR**
Data de-dispersed with true DM	33.877491	73.5323563
Filtering after dedispersion with true DM	44.113739	149.072739
Dedispersion with true DM after filtering	46.0323296	162.669907
De-dispersion with zero DM before filtering	4.9392066	9.74692345
De-dispersion with zero DM after filtering	5.25243902	25.0268936

A discussion of the above table now proceeds.

- When data is de-dispersed with the true DM value and not filtered, it shows a poor SNR due to the presence of the 50 Hz in the average pulse profile.

- When the same de-dispersed data is now filtered for the powerline harmonics, there is a considerable improvement in the SNR. The peak and best SNRs have both increased dramatically.

- When the raw data itself is filtered prior to de-dispersion, the improvement over the first case is quite along expected lines. But there is a marginal improvement in both the peak and best SNRs over the second mode. This can readily be explained as follows: When the data is de-dispersed before filtering, the noise is not uncorrelated across channels due to the presence of 50 Hz. Whereas when the data is first filtered prior to de-dispersion, the removal of 50 Hz actually guarantees, to some extent, that the noise across channels is now uncorrelated. When uncorrelated random variables are added, the central limit theorem can be applied with good validity.

- When data is de-dispersed with zero DM, the delay spread is not corrected for and pulse smearing results. However the powerline harmonics may add up almost coherently if the channel-to-channel phase difference is too small. This severely degrades the SNR as the result shows.

- However, if the data is filtered and then de-dispersed, there is a marginal improvement in the SNRs. Though the powerline harmonics have been filtered out, the noise due to electronics still dominates; the pulse energy across a narrow band of 62.5 KHz is not sufficient to make drastic changes in the SNR. However,with channel collapse with proper DM, the improvement is substantial as case 3 clearly shows.

Chapter 9

Conclusion and further work

Three methods were devised and the tracking filter is found the most satisfactory solution. It is therefore chosen as the standard algorithm. It is used to filter both de-dispersed data as well as raw data.

A solution for filtering the powerline harmonics in real-time from pulsar data now exists in the GMRT data chain. The performance is found to be satisfactory when run on a few data files in the simulation mode. This is a software solution running on the PC in the data chain. In future, one can think of integrating the filtering algorithm in the hardware itself so that there is little or no powerline interference in the observation data. This also does away with the need of overheads on the data acquisition PCs as it is found that the forward and inverse FFTs load the processor intensively.

The outcome of the project, other than the filtering routine itself, is the handle on the mechanism of interference entering into the data. The statistics of the movement and strengths of the powerline harmonics collected from the raw data file give us meaningful plots of the powerline trajectories in the data. Comparing with the grand bandshape, one can also identify the RFI channels that are modulated by the powerline sinusoid and those RFI channels that are not.

The probability of pulsar detection improves because of filtering as shown semi-qualitatively. This is a major step towards all-sky pulsar search observations using the GMRT.

Looking into the future, the following remedial meaasures are suggested.

- We find that 50 Hz sometimes leaks in through strong RFI channels, as in

channels 66 and 182 of the USB of 225 MHz, for example, and sometimes through low power RFI channels strongly modulated by 50 Hz, as in channels 41 and 42 of b0943+10 data, also at 225 MHz. Since GMRT now has the facility to locate RFI sources, these potential interferors can be located and remedial action be taken.

- Achieve the targeted bin resolution of 0.06 Hz/bin at 4 integrations.
- Identify antennas picking up 50 Hz and turning them off before every observation run
- Have a facility to automatically turn-off in real-time those channels picking up 50 Hz, i.e., setting their gains to zero.

Bibliography

[1] Basu, Baidyanath. *An Introduction to Astrophysics*, Prentice-Hall of India Ltd.,

[2] Chengalur, Jayaram N., Gupta, Yashwant & Dwarakanath, K.S. *Introduction to Low-frequecny Radio Astronomy.*

[3] Dahimiwal, Kiran H. & Gupta, Yashwant. *GPS Time Stamp for Pulsar Receiver*, NCRA Tech. Rep. R00199.

[4] Gupta, Yashwant. *Refractive Scintillation Effects on Pulsar Observations*, Ph.D Thesis.

[5] Frigo, Matteo & Johnson, Steven G. *FFTW: An Adaptive Software Architecture for the FFT*, ICASSP 1998, vol. 3, pp. 1381-1384.

[6] Joshi, B.C. *Pulse emission properties of radio pulsars*, Ph.D Thesis.

[7] Kraus, J.D. *Radio Astronomy*, Cygnus Quasar Books.

[8] Lamport, Leslie. LaTeX A Document Preparation System.

[9] McDonnough,Robert & Whalen, Anthony, *Detection of Signals in Noise*, Academic Press.

[10] Miller, Diane Fisher. *Basics of Radio Astronomy for the GAVRT*, Jet Propulsion Laboratory, CalTech. (http://www.jpl.nasa.gov/radioastronomy)

[11] Press, William H. et al. *Numerical Recipes in C*, Cambridge University Press.

[12] Smith, Francis G. *Pulsars*, Cambridge University Press.

[13] VanTrees, Harry L. *Optimum Array Processing*, John Wiley & sons.

[14] Venkataraman, G. *Chandrasekhar and His Limit*, Universities Press.