

Arecibo Upgrade Notes
COMPARISON OF DATA ACQUISITION METHODS: DATA RATES

Jim Cordes
 12 February 1993

Four methods of data acquisition for pulsar science may be considered, one involving predetection recording and three that are post-detection:

- (1) *Direct baseband sampling:* Usually a complex baseband signal of bandwidth $B_b = B/2$ is sampled at Nyquist intervals $\delta t = (2B_b)^{-1} = B^{-1}$, where B is the IF bandwidth. The time-bandwidth product $\equiv 1$. Full polarization analysis (four Stokes parameters) may be done by recording two polarized baseband signals (e.g. LHCP & RHCP). However, the same amount of data must be recorded even if one is interested in analyzing only the sum of the RHCP and LHCP powers, as in pulsar searching. With direct recording, any conceivable kind of analysis may be performed but at the expense of larger data rates. The data rate to magnetic tape for N_{pol} polarization channels is (in bits s^{-1})

$$R_{tape} = 2mN_{pol}B \quad (1)$$

with m bits per number (2m bits per complex sample).

- (2) *Filter bank sampling:* Detected outputs of an analog filter bank are sampled at intervals $\delta t = 2\Delta\nu_{ch}$ per channel, where $\Delta\nu_{ch}$ is the *voltage* bandwidth of each IF filter. Prior to recording, RHCP and LHCP may be summed. Full polarization means recording 4 numbers per frequency channel (auto and cross products of predetection outputs between two polarization channels). The time bandwidth product ≥ 1 , depending on time averaging of $\tau/\delta t$ individual samples. The data rate to tape for $N = B/\Delta\nu_{ch}$ channels is

$$R_{tape} = \frac{2mN_{pol}B}{(\tau/\delta t)N_{\Sigma}}, \quad (2)$$

where we have assumed that channel spacings are equal to the voltage bandwidth (See notes in Appendix). We have defined N_{Σ} as the number of polarization channels that have been summed after detection but before recording to tape.

- (3) *DFT Filter Bank:* The data rate into an FFT device is the same as the direct, baseband sampling rate. The complex data out of an FFT device that handles N_{pol} polarizations are at a rate

$$R_{FFT} = 2mN_{pol}B. \quad (3)$$

For searching, the squared magnitudes of two polarization channels may be summed, reducing the rate by a factor of 4. However, optimal S/N requires overlapping of data blocks in the time domain by 50% so that aliasing of noise is reduced significantly. Antialiasing is probably more important in searching, where detection levels are minimized, than in timing and other monitoring studies. Overlap increases the data rate by 2. The net sample interval per DFT channel is $\Delta t_{FFT} = N_{FFT}/2B$, with overlap, and the net rate to tape is

$$R_{tape} = \frac{2mN_{pol}B}{(\tau/\Delta t_{FFT})N_{\Sigma}}, \quad (4)$$

where a factor $\tau/\Delta t_{FFT}$ allows for summing over multiple computation times of the FFT.

- (4) *Correlator sampling:* Correlation functions with N_{lags} unique values (positive and zero lags) are produced at time intervals in units of ‘fundamental integration times’ $\Delta t_{corr} = 2N_{lags}(2B)^{-1} = N_{lags}/B$. This is the absolute minimum time needed to compute the N_{lags} correlation values with the stipulation that the same number of lagged products contributes to each lag. As with the DFT filter bank, minimization of aliasing requires that the *sample interval* be half the integration time. Being a post-detection device, polarizations may be summed for pulsar searching where total power is of interest. Polarization work requires computation and recording of 2 autocorrelation and 2 cross correlation functions. The time bandwidth product ≥ 1 , depending on time averaging prior to recording in excess of the fundamental sample interval. The data rate out of the correlator is $R_{corr} = 2mN_{pol}B$ for correlator dumps at intervals $\Delta t_{dump} = N_{ch}/2B$. With summing over multiple correlator dumps, the data rate to tape is, in general,

$$R_{tape} = \frac{2mN_{pol}B}{(\tau/\Delta t_{dump})N_{\Sigma}}. \quad (5)$$

Table 1 compares the four methods by showing data rates and sample intervals *prior* to any real-time accumulation before recording as well as after any such accumulation. Table 2 gives values of parameters for different observing goals. The choice of data acquisition clearly depends on the intended analysis. For some analyses, there are clear-cut advantages.

Searching: For searching, the post-detection schemes have the advantage that polarization summing may be performed, yielding a factor of two smaller data rate to tape (for equal quantization m). Taking as an example, $B = 10$ MHz, $N_{pol} = 2$ (eg. using the 430 MHz line feed at Arecibo) we have $R_{tape,direct} = 40m$ Mbits s^{-1} while $R_{tape,FFT} = 20m$ Mbits s^{-1} . For an FFT length $N_{FFT} = 256$ over $B/4 = 2.5$ MHz and using 4 FFT chips per polarization, ($\times 2$ for overlap to imply 8 chips) a sample interval of $51.2\mu s$ is achieved. Thus, with 16 chips, a search device can be implemented that has 1024 total channels over 10 MHz each sampled at $51.2\mu s$.

As wider total bandwidths B are considered, more channels are needed in post-detection devices to yield the same dispersion smearing. With searching, a time resolution less than $\sim 50\mu s$ is unnecessary, so the time-bandwidth product (per channel) can become very large, thereby decreasing the recorded data rate to far below that needed for direct baseband recording.

Timing & Polarization: Timing and polarization studies require greater time resolution than does searching. This is especially true for millisecond pulsars with pulse widths as small as $40\mu s$, for which sample intervals in the range of 1 to $10\mu s$ are desirable. However, pulse waveforms may be averaged over long times (\sim minutes), thus reducing the data rate to magnetic tape or disk to much less than what is needed for searching.

If signal averaging is done in real time (with dedispersion subsequent to Faraday unwrapping in the case where full Stokes parameters are wanted), then pre-and-post-detection methods are equally efficacious for many pulsars. Pre-detection methods are advantageous, however, for observations of pulsars with large DM or at low frequencies. The achievable time resolution when dispersion is removed by predetection methods is $\Delta t_{pre} = B^{-1}$ while for post-detection methods the best time resolution is $\Delta t_{post} = (8.3DM\nu_{GHz}^{-3})^{1/2}\mu s$. If we ignore pulse broadening due to scattering, which also gets larger for large DM and low frequencies, then predetection methods are superior

for $\Delta t_{pre} < \Delta t_{post}$, or

$$B > 0.35 \text{ MHz} \left(\frac{\nu_{GHz}^3}{DM} \right)^{1/2}. \quad (6)$$

At 430 MHz, total bandwidths greater than $\sim 0.1(DM)^{-1/2}$ MHz yield better time resolution than any post-detection method, but only up to dispersion measures of ~ 200 for which scattering limits the time resolution.

APPENDIX

Notes on Analog Filter Banks:

For a voltage response at IF of $H(\nu - \nu_{IF})$ with bandwidth $\Delta\nu_{ch}$, the power spectral response to a sinusoid of frequency ν is $|H(\nu - \nu_{IF})|^2$ while the Fourier transform of detected noise is $H(\nu) * H^*(\nu)$. The bandwidth of detected noise is larger than that of the IF signal by 2 for an ideal rectangular passband and by $\sqrt{2}$ for a gaussian passband. Sampling at intervals $\delta t = (2\Delta\nu_{ch})^{-1}$ yields no aliasing for the ideal passband. Sampling at this rate for a gaussian causes frequency components beyond the 1/e point of the detected signal's spectrum to be aliased. If IF filter channels are spaced such that they overlap at their half-power points (i.e. of $|H|^2$), the spacing is $\Delta\nu_{ch}$ for an ideal passband and $\sqrt{2\ell n 2} \Delta\nu_{ch} = 1.18 \Delta\nu_{ch}$ for gaussian filters.

Notes on DFT Filter Banks:

An N-point DFT produces channel center frequencies $\nu_k = (N\delta t)^{-1}$, where δt is the sample interval and the response to a complex sinusoid of frequency ν is (to within a constant phase factor)

$$H_k = N^{-1} \frac{\sin N\pi(\nu - \nu_k)\delta t}{\sin \pi(\nu - \nu_k)\delta t}. \quad (A1)$$

The peaks of channels fall at the first nulls of contiguous channels and the power responses, $|H_k|^2$, of adjacent channels cross at the levels $(2/\pi)^2 \approx 0.41$ for $N \gg 1$. DFT's calculated from contiguous, non-overlapping blocks of data of length $T = N\delta t$ yield a time series for each frequency channel that is undersampled. For this case, the folding frequency (for the transform of one of these time series) is $1/2T$, so voltage amplitudes as large as $2/\pi = 0.64$ at $\nu - \nu_k = 1/2T$ are aliased. By computing DFT's of data blocks that overlap by 50%, the folding frequency is pushed out to the first null of H_k . These considerations are for the predetection case. The squared magnitude of each DFT output will have a spectrum that is identical to that of the magnitude because the self convolution of a sinc function is the same sinc function.

TABLE 1
COMPARISON OF SAMPLING METHODS

Sampling Method	tB product	Sample Interval (per channel) Δt	Data Rate to Tape R_{tape}
baseband, complex (pre-detection)	1	B^{-1}	$2mN_{pol}B$
filter bank, DFT, or correlator (post-detection)	≥ 1	$N_{ch}/2B$	$\frac{2mN_{pol}B}{(\tau/\Delta t)N_{\Sigma}}$

m = number of bits per independent sample (e.g. complex samples = 2m bits)
 B = total bandwidth analyzed
 N_{pol} = number of polarization channels sampled
 N_{Σ} = number of polarization channels summed before recording.
 $\tau/\Delta t$ = post-detection number of samples summed before recording.
 Post-detection cases: overlap of DFT's or correlation functions by 50% is assumed.

TABLE 2
POST DETECTION SIGNAL PARAMETERS

Observation	N_{pol}	N_{Σ}	N_{beams}
Searching	2	2	≥ 1
Timing	4	1	1
ISS	2	1,2	1

N_{beams} = number of independent beams in a multiple feed system