

SDR equivalents to analog functions in a small-scale radio telescope

Marcus Leech, Science Radio Laboratories

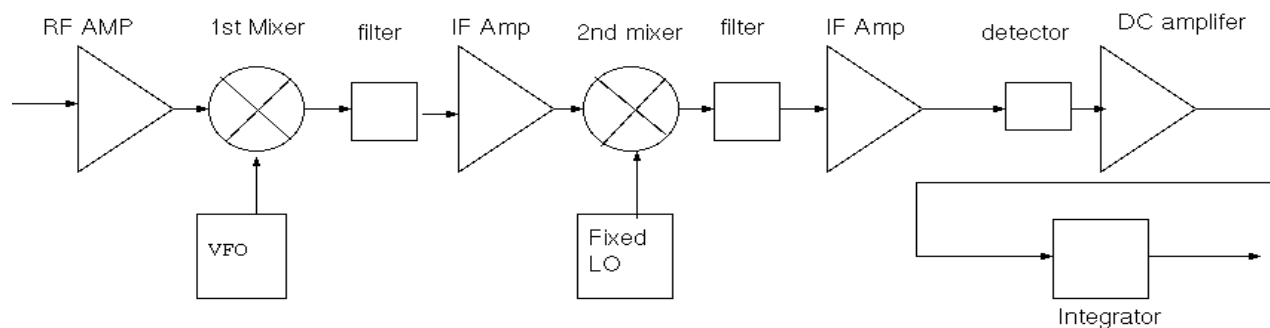
Introduction

Small-scale radio astronomy has typically been accomplished using simple analog circuitry, often constructed by the observer themselves from scrap-box parts. Observational goals are typically quite modest, and those modest goals are often reflected in the relatively-primitive analog receiver systems used.

With the advent of affordable *Software Defined Radio* technology comes an opportunity to re-examine the methods and techniques traditionally used by an amateur observer to construct a suitable radio telescope receiver.

Analog Total Power Receiver

In a typical small-scale observatory conducting total-power observations, the receiver is typically a single or dual-conversion superheterodyne design, with the final IF stage driving a hardware square-law detector, integrator, and DC amplifiers, as shown below:

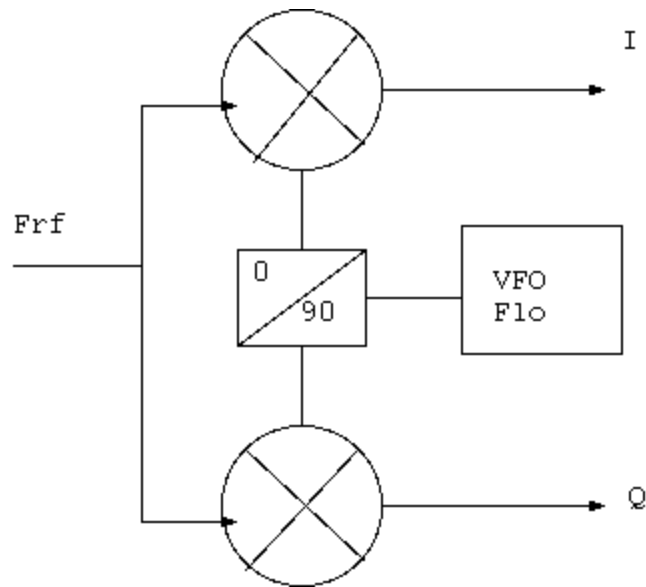


The signal arrives from the antenna, and is amplified by one or more stages of RF amplification (and sometimes bandpass filtering), and is then presented to the 1st mixer where the signal is converted to a fixed *intermediate frequency* by the mixing action between the incoming signal frequency and the frequency produced by the *Variable Frequency Oscillator*. The result is a fixed *intermediate frequency* that is filtered, and amplified by the 1st *IF amplifier*. The signal is then typically mixed **again** in the 2nd mixer stage, and converted to the final *intermediate frequency* with a fixed *Local Oscillator*. Once the final IF is produced, it is detected with a *square-law detector*, usually a detector diode. The detector produces an output voltage that is proportional to the incident AC power. This voltage is then amplified by the DC amplifier, and low-pass filtered with an integrator stage. The output of the integrator is then usually fed to a recording device of some sort.

Digital, SDR-based Total Power Receiver

In a digital receiver, the signal path is typically identical up to the output of the 1st IF amplifier, with some crucial differences. In an SDR-based receiver, the incoming signal is typically converted to a *quadrature-baseband* signal, producing the so-called ‘I’ (in-phase) and ‘Q’ (quadrature) signals. This is achieved using **two** mixers that are fed with the RF signal, and a local-oscillator that has both a *sine*

and *cosine* output. This is shown below:



Observe that there are two mixers, instead of the usual single mixer. There is also a so-called *phase-splitter* that produces both sine and cosine components from a single sine-component LO signal. It is also the case that the VFO is set so that F_{lo} is equal to F_{rf} , that is, the VFO operates at the same frequency as the desired reception center frequency. This is a so-called *direct-conversion* receiver. A direct-conversion receiver without so-called quadrature conversion suffers from a near-fatal flaw, in that both the *sum* and *difference* components in mixing perfectly overlap. In some types of radio-science, that would be acceptable, but in order for a direct-conversion design to “act” like a conventional receiver, a different signal representation is required. The I/Q signal is mathematically like a *complex* number with both a *real* and *imaginary* component. This complex (I/Q) representation allows us to disambiguate the sum/difference frequencies when we process the signal mathematically.

Consider an input signal a F , with a bandwidth of Bw , when that signal is processed by a direct-conversion quadrature mixer, the resulting *baseband* signal occupies a new spectrum that spans from $-Bw/2$ to $+Bw/2$. Now in the real world, there’s no such thing as a negative frequency, but the I/Q, or complex representation allows us to pretend that there is. It’s a mathematical abstraction that turns out to be incredibly useful in the real world.

Once we have our I and Q signals, they are typically low-pass filtered in hardware (analogous to the IF filters in a conventional receiver), amplified, and then sampled with a high-speed ADC.

The *low-pass* filters in a quadrature-digital receiver are also known as *anti-alias filters*. Recall the *Nyquist Sampling Theorem* that states that a signal of bandwidth B may be adequately sampled (and later reconstructed) using a sample rate of no more than $2B$. Put another way, a digital sampler, such as an ADC, operating at a sample rate of F_s can adequately sample a signal of bandwidth $F_s/2$. Any signal beyond $F_s/2$ becomes *aliased* into the input pass-band. Hence the need to have an anti-alias filter in front of the sampler in order to reduce frequencies above $F_s/2$. Many digital receivers use a fixed sampling rate, in order to simplify anti-alias filters, and it is usually the case that 5th or 7th order passive filters are used prior to sampling. In some variable-bandwidth designs, the anti-alias filters are based on a switch-capacitor filter technology that allows adjusting the low-pass corner frequency. This is how variable-rate sound cards for PCs perform anti-alias filtering.

Once the signal has been sampled at some rate that is convenient, it may be further processed in digital

hardware to reduce the sample rate, via a technique called *decimation*. In an agile digital receiver, it is typically the case that the signal is sampled at a high rate that is commensurate with available hardware budget, and then *decimated* according to the end application.

Once the signal has been decimated (which is also a form of digital low-pass filtering), it is presented to the final digital signal-processing chain for further processing and information extraction.

The digital processing chain

Once the signal has been digitized, and possibly decimated, it is presented to a general-purpose digital signal processing chain. We'll assume that chain "lives" inside a general-purpose computer.

We'll assume that our signal arrives in the I/Q format described above, but for *direct-sampled receivers*, the signal arrives as a single quantity. Such direct-sampled receivers are typically used only when the RF frequencies are relatively low (in the HF bands, typically).

Our signal arrives in the computer as I/Q samples that represent a swath of bandwidth (Bw) from $F_c - (Bw/2)$ to $F_c + (Bw/2)$, where F_c is the center frequency of interest. Keep in mind, though, the signal is in *baseband* form, with F_c at 0Hz.

Once we have the signal it can be further filtered, and even decimated further. It is sometimes the case that the sample rates offered by the front-end hardware don't precisely match the sample rates desired in the application, so further decimation is often performed within the computer.

Filters in a digital receiver are implemented typically with FIR or FFT topology filters, which give fairly-precise responses, with perfect symmetry, and flat passband responses. The design of such filters (or rather, the coefficients for such filters) is beyond the scope of this document, but tools exist for designing such filters.

Assuming that our signal is filtered appropriately, we need to detect it—that is, extract a number that is proportional to the incident power. In a digital receiver, we simply calculate:

$$X = (I*I) + (Q*Q)$$

Recall that the *power* flowing in a circuit is proportional to the square of the *voltage* across that circuit. Our digital signals are numerical samples of an RF *voltage*, and so *detection* is mathematically straightforward. Once the signal has been detected, it needs to be *integrated (low-pass filtered)*. In an analog receiver, the integrator would likely be constructed using an op-amp R-C low-pass filter. In a digital receiver, we use the *single-pole IIR filter* as a low-pass filter:

$$Y = (a * X) + ((1.0 - a) * Y)$$

As shown, Y represents the filter output, X the incoming signal from the detector, and 'a' is a parameter that determines the corner low-pass cutoff frequency (or, in more-familiar terms, the *integration time constant*).

Once the signal has been detected, and integrated (low-pass filtered), we can reduce its sample rate to a rate that is twice the *low-pass corner frequency* of the integrator. In a total-power radiometer, we typically use integration times from about 1 second up to several 10s of seconds. For consistency reasons, it's typical to reduce the post-integrator sample rate to some fixed value, like 2Hz.

Once the signal has been *integrated* and *decimated* down to a standard sample rate, we can scale it and offset it with simple arithmetic, before it is logged to an external file for later analysis and processing.

For example, if we wanted to convert the signal Y above (output of the integrator) to Jansky, we would

determine an offset value, and a coefficient value to apply to the signal prior to logging it:

$$Y1 = (Y + \text{offs}) * \text{Coeff}$$

Where ‘offs’ and ‘Coeff’ would need to be determined empirically for each situation. It is usual, however, to apply such scaling factors in a post-processing step, and that is also true for analog receivers.

An analog spectrometer

Recall the diagram of the analog total-power receiver given earlier. It has a VFO (Variable-Frequency Oscillator) that determines the center-frequency of operation, and two or more stages of Intermediate Frequency (IF) conversion and filtering.

If that receiver is modified so that there’s a very-narrow 3rd IF stage, with tightly-controlled analog filter and a precision detector, then by varying the VFO frequency in a swept-frequency fashion, and recording the outputs of the narrow-band 3rd IF detector, we can measure the frequency spectrum of a signal. Since the 3rd IF detector is very narrow, it allows us to estimate the power only over a tiny frequency portion of the incoming signal. By synchronizing the detector output with the so-called *sweep generator*, we can produce a spectral estimate.

A digital spectrometer: The *Fourier Transform*

Back in the 19th century a French mathematician and physicist, *Joseph Fourier*, invented a new mathematical technique for the analysis of complex vibrations and heat transfer. Today his keen mathematical insights into vibratory processes has revolutionized the way signals are analyzed and processed.

Various versions of the so-called *Fourier Transform* exist today, but for discrete-time-sampled digital signals, the technique most often applied is the *Fast Fourier Transform* (FFT).

Computing an FFT over a block of *discrete-time-samples* yields an estimate of the *frequency-spectrum* of that signal.

Consider our incoming I/Q signal from our ADC, that occupies a bandwidth of *Bw* Hz. If we calculate an FFT of length *Z* on the signal, we require *Z* samples of that signal to produce a spectral estimate over those samples:

$$O[0..Z] = \text{FFT}(S[0]..S[Z])$$

Where *S* is our incoming I/Q signal as a complex number. If our incoming signal has a bandwidth of *Bw*, then each “bin” of the output spectrum represents *Bw/Z* Hz of bandwidth. If our signal bandwidth is 1MHz, and our transform length is 1024, then each “bin” in the FFT output represents 976.25Hz of the incoming signal.

For a spectrally-stationary signal, frequency estimates become better as you integrate the output bins, and so it is typically the case that each FFT output bin is run through a *single-pole IIR filter*, just as in the total-power receiver:

$$E[i] = ((O[i] * a) + ((1.0 - a)*E[i]), i=0..Z$$

Where *E[x]* is the integrated estimate for bin ‘x’, *O[x]* is the FFT output for bin ‘x’.

Modern FFT implementations on modern desktop-CPU hardware can execute extremely quickly, which means that the resolution of the spectral estimates can be extremely high. SETI processing, for

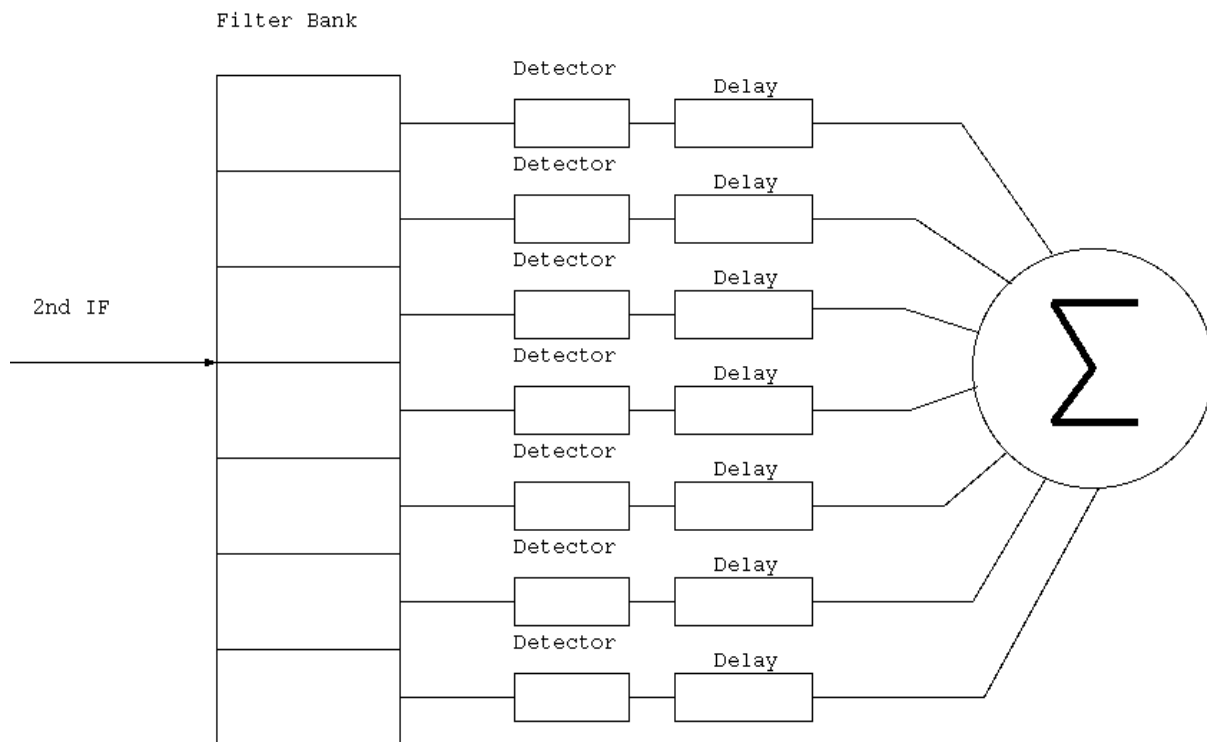
example, requires very high resolution, and thus very large FFTs. Such resolution would be difficult to obtain in an analog spectrometer. It is not unusual to find digital spectrometers in a *Software Defined Radio* to have FFTs with many 10s of thousands of FFT bins, giving very fine spectral resolution, capable of measuring very small doppler shifts, for example.

Analog Pulsar Processing: the Analog Filter Bank

Pulsars are rapidly-rotating neutron stars whose intense magnetic field causes a beam of electromagnetic radiation to emanate from each pole. When that beam intersects our antenna, we receive a brief pulse as the pulsar beam transits the beam of our antenna.

The peak power of these pulses, as received on earth is fairly weak, so they must be observed over significant bandwidths, in order to achieve adequate sensitivity. The problem is that the *interstellar medium* is *dispersive* in nature, and thus pulses at lower frequencies arrive **after** pulses at higher frequencies, which means that the pulses tend to become *smear*d, when observed over the entire bandwidth. The smearing is often so severe that the pulses get smeared **below** the noise floor of the detection apparatus.

In analog pulsar receivers a so-called *filter-bank* is applied to the 2nd IF to divide up the incoming bandwidth into small *sub-bands*, like so:



The resulting *sub-bands* are individually detected, and then *delayed* in a *variable analog delay line*, before being summed in an analog summing amplifier. In this way, all the pulses in the various sub-bands can be made to arrive simultaneously at the summing amplifier, and thus produce a non-smearred (or dramatically-less-smearred) pulse profile at the output of the summing amplifier.

Digital Pulsar Processing: Coherent De-dispersion

In 1975, *Hankins and Rickett*¹ determined the *transfer-function* of the interstellar medium. They were able to predict the so-called *dispersion measure* along any line of sight, knowing only the column depth of the medium (the distance to the observed object).

What this meant was that you could construct a *digital filter* to undo the effects of the interstellar medium, prior to detection of the signal. They published an algorithm (a series of mathematical steps) for determining the required coefficients of a complex FFT filter that would *de-disperse* the incoming signal, as long as you knew the *dispersion measure* (DM) along the line of sight, the observing bandwidth, and the observing center frequency.

It turns out that the required filters are quite large, when attempted over significant bandwidths, so it wasn't until fairly recently that de-dispersion **in real time** has been attempted by radio astronomers, since the necessary computing power is significant.

In a digital pulsar receiver, with real-time de-dispersion, the signal is processed like so:

$$Y = \text{FFT_FILTER}(S, DM, Fc, Bw)$$

Where Y is the de-dispersed signal, S is the incoming signal, DM is the *dispersion measure*, Fc is the center frequency, and Bw is the observing bandwidth. Once you have the de-dispersed signal, it can be processed with a detector, just as in a total-power receiver, although the output sample rates are typically much higher, on the order of several Khz, because pulsar pulses are quite narrow, although their repetition rates are usually quite long relative to their pulse widths.

Using this mathematical model, you can produce pulse profiles that are very accurate renderings of what the pulse shape actually looks like, without any appreciable smearing in time. Having high-fidelity pulse profiles allows astrophysicists to learn more about the nature of pulsars, neutron stars, and astrophysics in general.

Pulsar Folding: Digital Synchronous Detection

Recall that pulsar pulses are relatively weak, and exceedingly narrow. This means that the total energy captured by a radio telescope during a pulsar pulse is very small. We've already discussed how we use de-dispersion to reduce or eliminate *smearing* of a pulse, and thus improve the odds of detection.

Another technique that is used is so-called folding of the detected pulsar pulses. By knowing fairly precisely what the pulsar repetition rate is, we can add individual pulses together to produce a composite pulse profile with a higher SNR than a single pulse. Achieving this digitally is fairly straightforward, since you can manipulate samples in a fixed-length buffer that represents a single pulse period, and add these buffers together over time, to produce an integrated pulse profile.

The analog approach to this problem would be to use a series of analog delay lines and a summing amplifier with an appropriate integrator. But even in analog pulsar receivers, it is typically the case that the detector(s) output are sampled with a ADC with a sample rate of a few Khz, and so-called *folding* is achieved digitally, since the analog approach would be very cumbersome, and ADCs with

1 See: Hankins and Rickett, Pulsar signal Processing, Methods in Computational

Physics Vol 14, 1975.

Khz sample rates have been available for decades.

Transient Detection

Of some astrophysical interest are so-called cosmic transients—powerful pulses much like pulsar pulses, but random, sporadic, and usually extremely strong. Sometimes, these pulses are caused by extraordinary outbursts from existing and known pulsars, and sometimes their origins aren't clear.

Transient detection requires that you be able to sample the detector output at a fairly high rate, since transient pulses are quite narrow, similar to pulsar pulses, and that you be able to detect an event that qualifies as a transient, according to your criterion. The output of a detector in a radio telescope is very noisy both due to inherent noise in the electronics, and the noise of the cosmos itself. The task of transient detection lies in determining when a brief noise burst is statistically significant, and recording the transient and surrounding “context”.

With digital signal processing, applying such statistical tests is relatively straightforward, and well understood. You have to maintain a running estimate of the standard deviation of the incoming data, and “trigger” when a detected event exceeds the standard deviation by some configurable threshold. Quite straightforward, and cheap to compute, even at rates of several 10s of Khz on a modern computer.

A further refinement in transient detection is to use multiple receiver chains, which originate in different sky positions, and use the technique of *anti-correlation* to remove transients that are clearly of local origin (because they appear in one or more of the receiver outputs at the same time). Any transient event that is detected in only a single receiver is more likely to be of cosmic origin than one that is detected in more than one receiver.

It would be unlikely that you would construct a transient detector in a strictly-analog receiver, the required circuitry would be “finicky” and hard to characterize, but it is certainly theoretically possible.