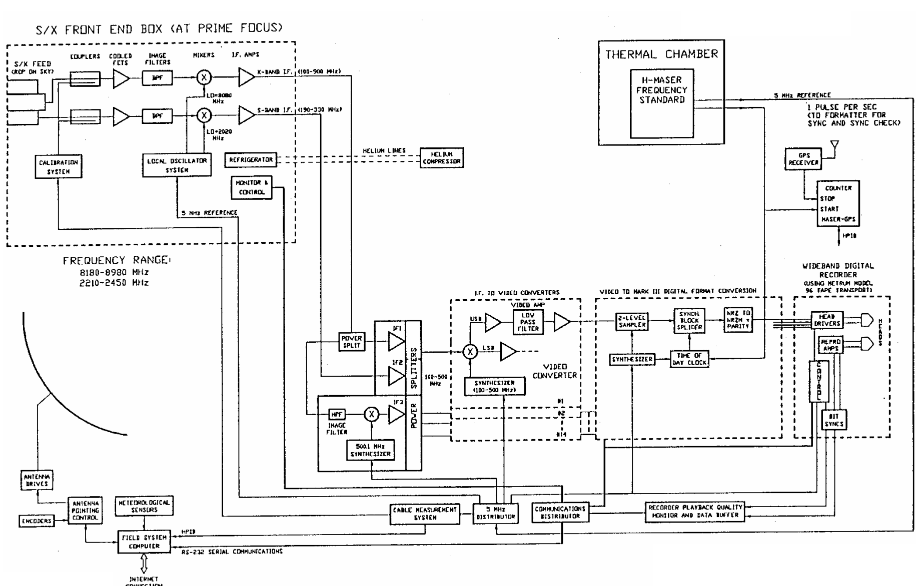
VLBI Electronics

A **system**-level (not component-level) **introduction** to (primarily) **analog** electronics for cm-wavelength VLBI

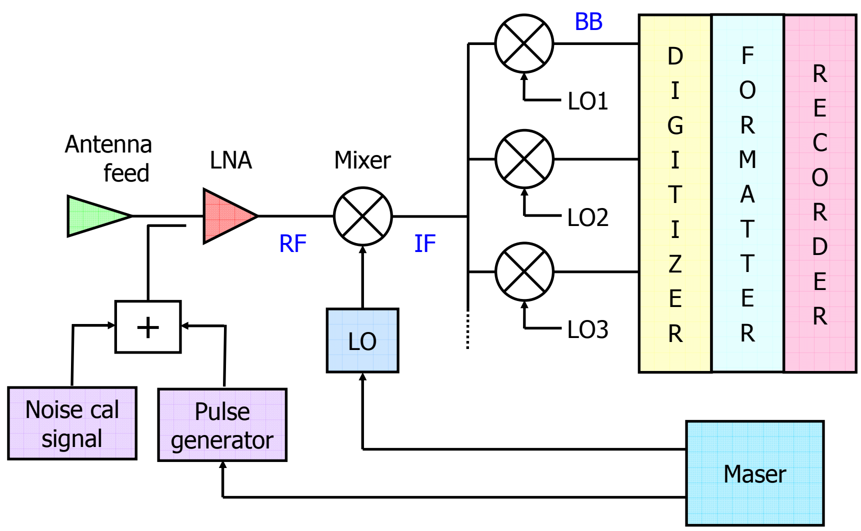
Major topics –

Thermal noise	Mixers
Feed polarization	Images & sidebands
Noise calibration	Phase-locked oscillators
Phase calibration	2-bit sampling
Low-noise amplifier	Aliases & Nyquist zones
Gain compression	Digital backends

Block diagram of S/X data acquisition system



Simplified block diagram of S- or X-band electronics



Radio telescopes

- Emission from radio sources is extremely weak.
  - For example, if the **largest** single radio telescope on Earth (Arecibo 305-meter) were to observe the **strongest** celestial radio source (supernova remnant Cas A) at S-band over a bandwidth of **1 GHz** for **50 years**, the total energy collected would light a 10-W light bulb for 20 milliseconds.
- In a typical geodetic observation, noise level is 100 to 10000 times stronger than signal from radio source.
- In order to detect signal at correlator, antennas must stay on source for tens of seconds to minutes to get enough SNR.
- Could do better science/geodesy with more scans per hour or more SNR per scan.
- How?
  - Build bigger telescopes to raise signal level from radio source.
  - Build telescopes that slew faster.
  - Decrease noise level into receiver.

Thermal noise

- Any opaque object above absolute zero temperature emits "thermal" electromagnetic radiation.
  - Any object opaque at microwave frequencies emits microwaves.
- At microwave frequencies, power P_{ant} received by an antenna surrounded by an enclosure at absolute temperature T is

$$P_{ant} = k_B \cdot T \cdot \text{bandwidth}$$

  where k_B = Boltzmann's constant = 1.38×10^{-23} W K^{-1} Hz^{-1}.
- For a 50%-efficient 20-m antenna with 50-K system noise observing a 1-Jy source over 1 GHz:

  P_{ant} from noise = 7×10^{-13} W

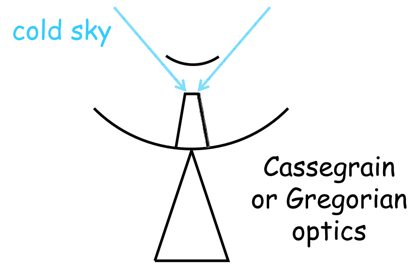
  P_{ant} from source = 8×10^{-16} W → noise/source ~ 1000
- A transmission line (e.g., waveguide or coax) decreases the SNR of the signal it carries in two ways:
  - Attenuates the signal (reduces "S") by factor (1 – loss)
  - Adds noise (increases "N") by amount $k_B \cdot T \cdot$ bandwidth $\cdot$ loss

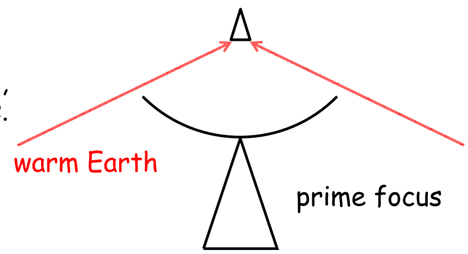
System noise budget

Source of noise	*Typical antenna temperature*	*Major dependencies*
Radio source	0-1 K	
Cosmic microwave backgrnd	3 K	
Milky Way Galaxy	0-1 K	frequency, direction
Ionosphere	0-1 K	time, frequency, elevation
Troposphere	3-30 K	elevation, weather
Antenna radome	0-10 K	
Antenna	0-5 K	
Ground spillover	0-30 K	elevation
Feed	5-30 K	
Cryogenic LNA	5-20 K	
Total	16-130 K	

Options for minimizing extraneous noise in telescopes

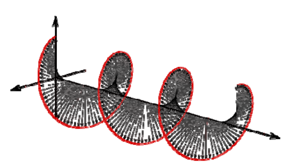
- Use telescope optics that minimizes pickup of 300-K ground radiation.
  - In dual-reflector (Cassegrain or Gregorian) optics, spillover from feed intercepts sky, not ground, at most elevations. → lower noise than with prime-focus feed
  - Dual reflectors allow "shaped" optics to reduce feed spillover. → reduced noise at low elevation
- Waveguide feeds have lower loss, and hence noise, than coax feeds.
- Cool non-waveguide feeds.
- Use lowest-noise first-stage amplifers.





Feed polarization

- Electromagnetic waves may be viewed as being a combination of either
  - Vertical and horizontal components of a linearly polarized (LP) wave or
  - Right and left circular components of a circularly polarized (CP) wave.



- Feeds are designed to be most sensitive to either LP or CP waves.
- At microwave frequencies, LP feeds are generally easier to construct and have lower noise than CP.
- However, because
  - feed orientation relative to the source changes on an alt-az telescope as the Earth rotates,
  - orientation is different for widely separated telescopes, and
  - vertical and horizontal LP signals from a source are uncorrelated,

  signal correlation between two LP telescopes can vary, and even drop to zero, as the Earth rotates.
- CP feeds avoid this problem.

Noise calibration

- Noise calibration system measures changes in the power sensitivity.
  - It tells nothing about phase instabilities.
- A signal of known strength is injected ahead of, in, or just after the feed, and the fractional change in system power is measured.
- The system temperature T_{sys} (equivalent to the system power $k_B T_{sys}$ per Hertz) is then calculated from the known cal signal strength as

$$T_{sys} = T_{cal} / (P_{on}/P_{off} - 1)$$

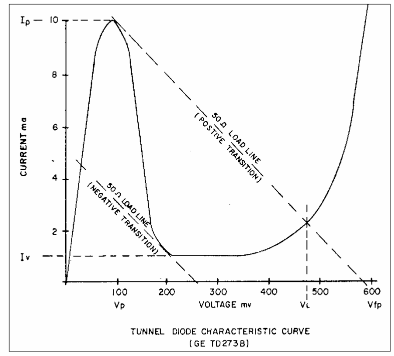
- Even if the true value of T_{cal} is unknown, the calculated T_{sys} values are useful as measures of relative changes in sensitivity.
- If T_{cal} is small ($\lesssim$5% of T_{sys}), continuous T_{sys} measurements can be made by firing the cal signal periodically (VLBA uses an 80 Hz rep rate) and synchronously detecting the level changes in the backend.
- For reliable T_{sys} measurements
  - T_{cal} must be stable (may require physical temperature control), and
  - the system gain ahead of the injection point must be stable.
- Field System does T_{sys} calculations and reports raw data and results.

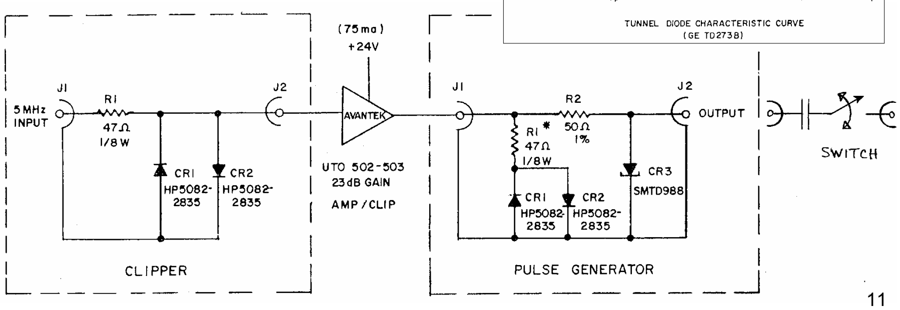
Phase calibration

- Phase calibration system measures changes in system phase/delay.
  - Amplitude of phase cal signal recovered at baseband can be used to infer variations in T_{sys}, but that is a secondary feature.
- A train of narrow pulses ("clock ticks") is injected typically at the same location as the noise cal signal.
- Pulses of width t_{pulse} with a repetition rate of N MHz correspond to a series of frequency tones spaced N MHz apart from DC up to a frequency of ~ $1/t_{pulse.}$
  - E.g., pulses of width ~50 ps yield tones up to ~20 GHz.
  - Typical pulse rate is 1 MHz, but it will be 5 or 10 MHz in VLBI2010.
- Older pulse generators used tunnel or step recovery diodes; newer generators use high-speed digital logic devices.
- Precision of instrumental phase/delay measurements can be no better than stability of phase cal electronics.
  - Temperature sensitivity of "digital" phase cal is < 1 ps/°C.
  - If temperature or mechanical stability of cable carrying reference signal up antenna is inadequate, need to measure cable length.

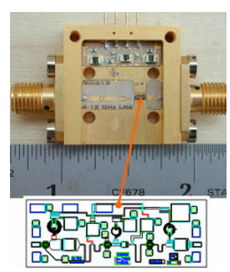
Tunnel diode pulse generator

- 1970s-era circuit below illustrates how a 5 MHz sinewave is converted to a 1 MHz pulse train.
  - Tunnel diode creates a 5 MHz square wave with fast rise/fall.
  - Capacitor “differentiates” to make positive & negative pulses.
  - Switch passes every 5th positive pulse. → 1 MHz pulse train





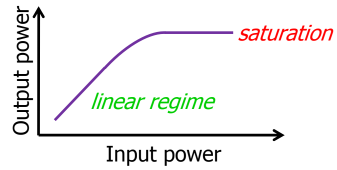
Low-noise amplifier

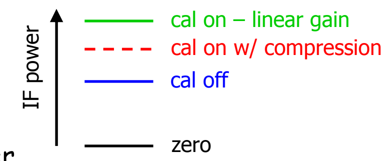


- LNA sets electronic contribution to noise level.
- Typical characteristics:
  - 5-20 K noise temperature
  - ~35 dB gain
  - cryogenic operation
- Gain should be high enough that noise contribution from downstream electronics is not significant.
- Power output capability (1-dB compression point) should be high enough that amplifier is not driven into saturation by RFI or phase cal pulses.
  - Equally true for any amplifier or other device ahead of sampler.
- A filter may be placed after LNA to remove image or strong RFI, but a filter placed ahead of the LNA is generally avoided because of the penalty in system noise.

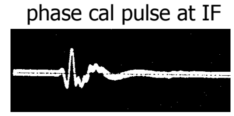
Gain compression

- Gain of analog electronics ahead of sampler should always be linear.
  - If T_{sys} doubles, noise at sampler input should double.
- At high power, gain becomes nonlinear.
- Operation near saturation causes erroneous T_{sys} measurements.
- Nonlinear gain causes intermodulation products in presence of RFI. → Reduced sensitivity at frequencies besides RFI.
- Phase cal pulse may be >10 times stronger than system noise.
  - An amplifier operating <10 dB below saturation will saturate during pulse.
    - Pulse amplitude is reduced, and spurious signals are generated.
- Wideband RF amps should be >20 dB below saturation when phase cal is off.



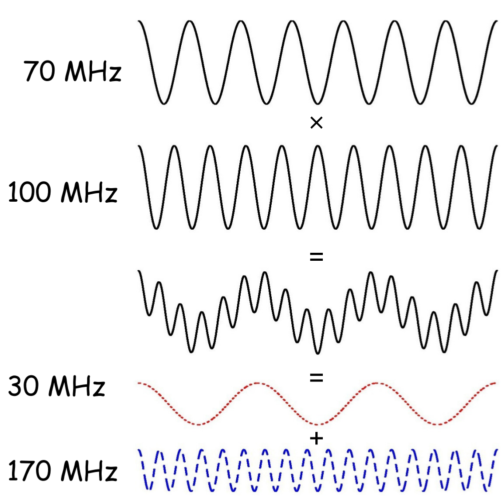


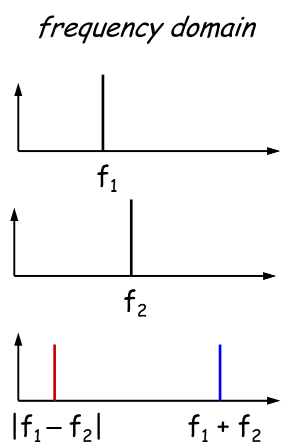
Oscilloscope trace of phase cal pulse at IF



Mixer action

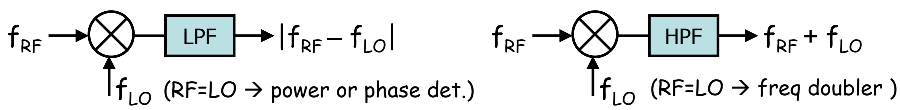
A mixer multiplies two signals together.



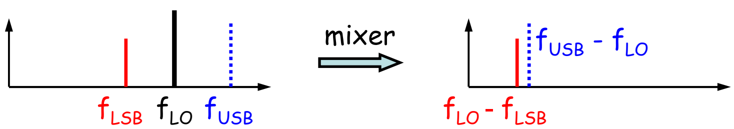


Sidebands and images

- From two input tones (sinewaves), a mixer creates output tones at the sum and difference frequencies.
- By filtering the output with a lowpass or highpass filter, the circuit becomes a downconverter or an upconverter:



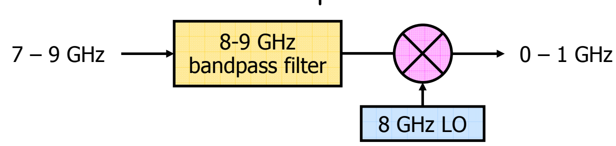
- In a downconverter, $f_{RF} > f_{LO} \Rightarrow$ upper sideband (USB) conversion
  $f_{RF} < f_{LO} \Rightarrow$ lower sideband (LSB) conversion



- For RF frequencies at $f_{LSB} = f_{LO} - f_{IF}$ and $f_{USB} = f_{LO} + f_{IF}$, the mixer output is at the same frequency f_{IF}.
- The image is the RF signal from the undesired sideband that has the same IF frequency as the desired signal.

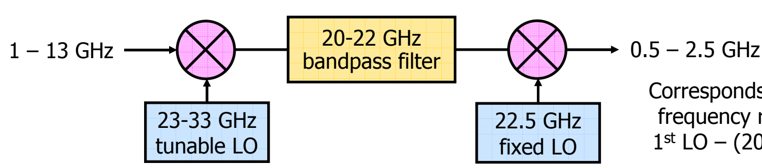
Image rejection

- Unless precautions are taken when downconverting, SNR will be lost if image signals are allowed to fall on top of desired signals in the IF.
- There are two common methods for rejecting the image.
  - Pre-mixer filter. Example:



Output corresponds to 8-9 GHz at input – 7-8 GHz "image" is attenuated in IF output

  - Updown converter – an upconverter with adjustable LO followed by a downconverter with fixed LO. Example:

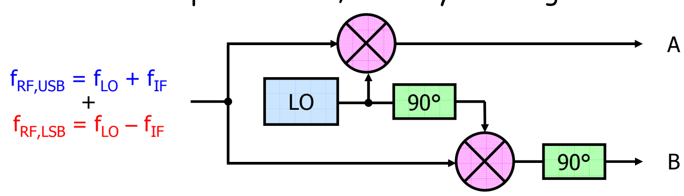


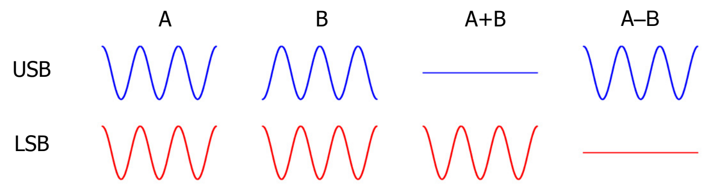
Corresponds to input frequency range of 1st LO – (20-22 GHz)

- An updown converter is more complex but allows the input frequency "window" to be varied over a wide range by varying the LO.

Image separation

- A sideband-separating (or image-reject) mixer puts downconverted sidebands out on separate lines, thereby avoiding SNR loss.





⇨ A+B gives downconverted LSB, A–B gives USB.

- VCs/BBCs use image-reject mixers to downconvert IF to baseband.
- Typical image rejection of 20 dB is not sufficient when LO frequency is integer MHz and phase cal is present.

Real mixers

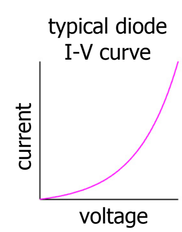
- Analog mixers rely on nonlinear current and/or voltage behavior of a device, such as a diode, to multiply two waveforms together.
- I-V curve can be represented by polynomial:

$$I(V) = a + bV + cV^2 + dV^3 + \cdots$$

- Over a small region of I and V, first 3 terms dominate.
- For V = sum of voltages R and L, quadratic term yields

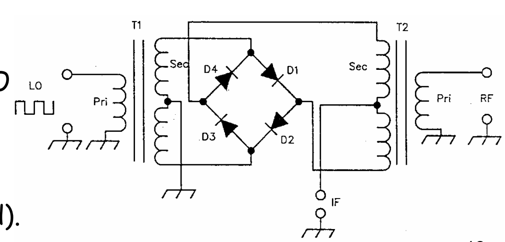
$$c\,(R + L)^2 = c\,R^2 + c\,L^2 + 2c\,RL \rightarrow \text{It's a mixer!}$$

- But a diode also makes 1st and 2nd harmonics of R and L.



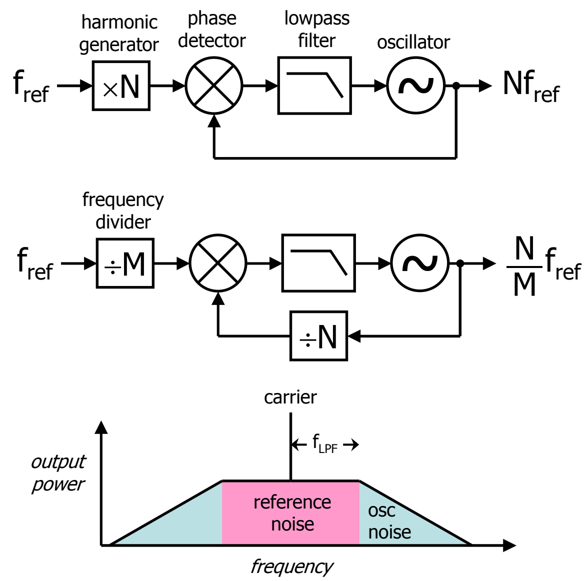
- Symmetry of 4-diode double-balanced mixer eliminates many RF/LO harmonics & products.
- Most mixers use square wave LO to improve performance.
- Even ideal DB mixers generate intermodulation products at frequencies mRF ± nLO (m,n odd).

Double-balanced mixer



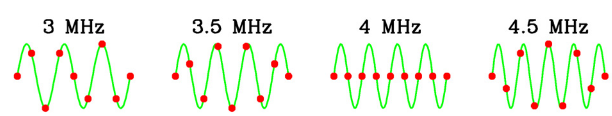
Phase-locked oscillators

- PLOs use a reference frequency to steer an oscillator on long time scales (> μs to ms).
- PLO phase noise outside the loop BW (= lowpass cutoff) is lower than that of ref signal multiplied to output frequency.
- 1st circuit is common in receiver LOs.
- VC/BBC PLO uses 2nd circuit with 5 MHz f_{ref} and M=500. → Output can be varied in steps of 10 kHz.

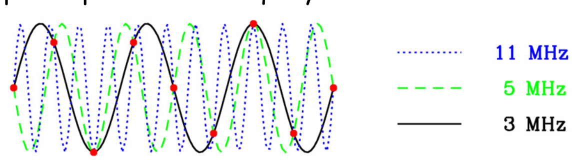


Sample rate, aliases, and Nyquist zones

- When data are sampled at a uniform rate, problems occur when the data include frequencies both above and below half the sample rate.
- Example of 4 sinewaves sampled at 8 MSps over 1 μs (samples in red):

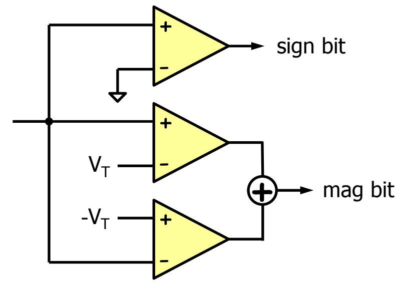


- Nyquist zone = frequency range from N x (f_{samp}/2) to (N+1) x (f_{samp}/2).
  - E.g., for f_{samp} = 8 MSps, NZs are 0-4 MHz, 4-8 MHz, 8-12 MHz, etc.
- Frequencies in different NZs that yield the same frequency in the sampled data are "aliases" of each other (e.g., 3.5 & 4.5 MHz above).
- If the data include frequencies from only one NZ, the waveform at the sampler input can be uniquely reconstructed from the samples.



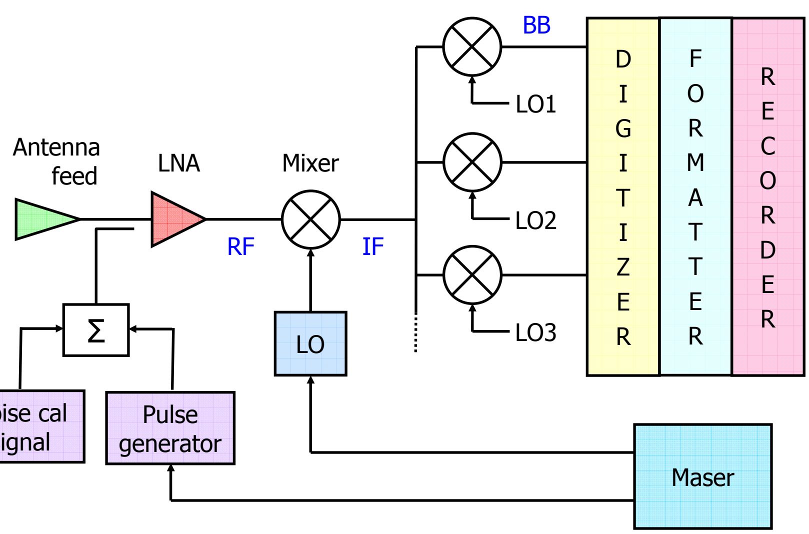
1- and 2-bit sampling

- For a **fixed** data rate (bits per second) in **continuum** observations:
  - 1-bit sampling (signal sign) gives maximum SNR.
    - Spread the samples over as much BW as possible.
  - 2-bit sampling (sign and magnitude) is almost as good as 1-bit.
    - Use half the BW as 1-bit sampling.
  - 3 or more bits give much worse SNR.
- In 1-bit sampling, rms input voltage simply must be >> sampler DC offset.
- In 2-bit sampling, SNR is maximized if 36% of samples are in high-mag state.
- Field System can monitor statistics via Mark 4 decoder or Mark 5B recorder.

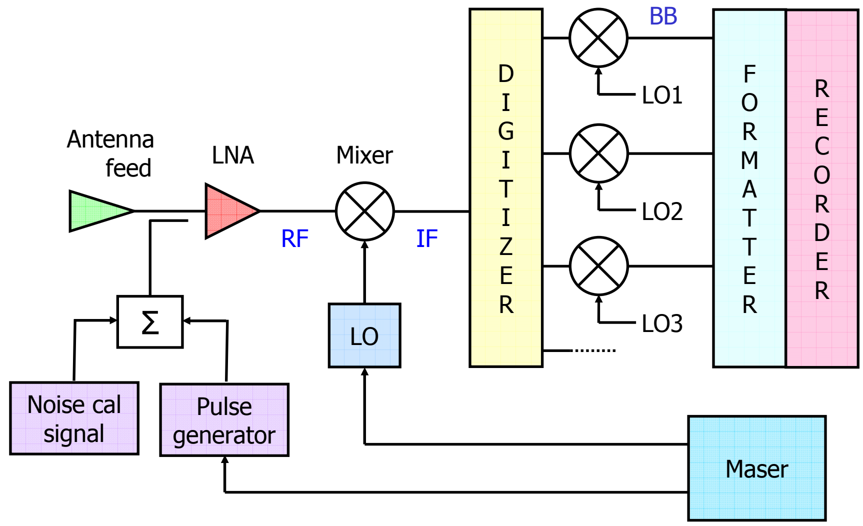


- 2-bit samples become in effect 1-bit if
  - input level is too low and mag bit is always '0', or
  - input level is too high and mag bit is always '1'.
- Automatic gain control in VC/BBC maintains sampler input at correct level for 2 bits.

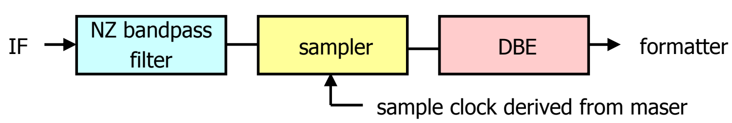
How does a system with DBE differ from one with ABE?



Here's how!

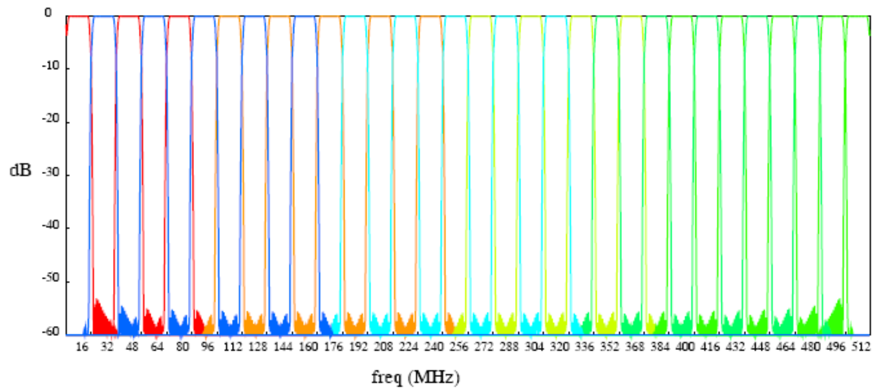


Digital backend



- Each Nyquist zone is typically 512 or 1024 MHz wide.
- Samples are typically 8 bits, to accommodate RFI and spectral slope.
- DBE splits NZ into multiple narrower channels, resampled at 2 bits.
- Method of channelization depends on firmware flavor:
  - Digital downconverter (DDC or DBBC)
    - Functions the same as analog BBC, with a tunable LO to downconvert an arbitrary slice of the input frequency range.
    - Multiple DDCs needed to provide multiple baseband channels for recording. (One FPGA can accommodate multiple DDCs.)
  - Polyphase filter bank (PFB)
    - Splits the input into 2^N channels, which are downconverted.
    - Limited flexibility in choosing input-to-output frequency mapping.

Example of polyphase filter bank channelization



FAQ (and InFAQ)

- Why are VC/BBC frequencies usually set to xxx.99 MHz (or xxx.49 MHz for astronomy)?
  - Short answer: To avoid contamination of phase cal at baseband.
  - Image and aliased phase cal tones may fall at the same frequency as the desired tone.
    - For example, for VC frequency xxx.50 MHz, image, alias, and desired tones all appear in baseband at nn.50 MHz.
  - The early Mark I and Mark III correlators supported only a very limited set of phase cal frequencies, of which 10 kHz was the least susceptible to corruption by images and aliases, hence xxx.99 MHz.
  - More modern correlators (e.g., VLBA, Mark 4, DiFX) support a far wider range of tone frequencies.

FAQ (and InFAQ) – cont'd

- Why does geodesy observe in two frequency bands, S and X?
  - To remove the effect of the ionosphere from the observed delays.
  - Because the ionospheric delay depends strongly on frequency (delay ~ frequency^{-2}), its value can be estimated accurately from data taken at two widely separated frequencies like S and X.
  - Other major errors in the delay are either independent of frequency (e.g., troposphere) or are measured by other means (e.g., instrumentation).
- Why does the Mark 4 IF distributor split each IF into separate low (100-220 MHz) and high (220-500 MHz) frequency bands?
  - Because the VC mixers are imperfect devices.
  - The strongest mRF ± nLO DB mixer intermodulation product is m=1 and n=3, which is ~13 dB below the m=n=1 fundamental.
  - If the IF input spectrum is flat, ~5% more power will appear at BB from 3x the desired IF frequency, with a corresponding SNR loss.
  - A low-band bandpass filter removes those 3x frequencies.

Some lessons learned the hard way

- When there is a problem, 90% of the time…
  - if it's an instrument, you forgot to plug it in;
  - otherwise, the problem is a bad connector.
- Use a spectrum analzyer to look for strange things in every signal you can get at. For example…
  - RFI, ripple, oscillation, bandpass rolloff, etc. in broadband RF/IF
  - Spurious signals and phase noise sidebands in phase cal
  - Phase noise and modulation on LOs and LO reference signals
- Use an oscilloscope to look for strange things at IF or baseband.
  - Check over a wide range of time spans.
  - Trigger internally, on maser-derived timing signal, and on line sync.
- Wiggle/shake things and see what happens (especially connectors!).
- Attenuators are a quick fix for many problems, e.g., amplitude ripple and amplifier oscillation caused by reflections.
- Use 50Ω power dividers to split RF/IF signals, not tees.
- The fewer connectors in your system, the happier your life will be.