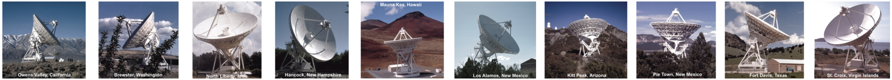
Correlators

Walter Brisken

NRAO

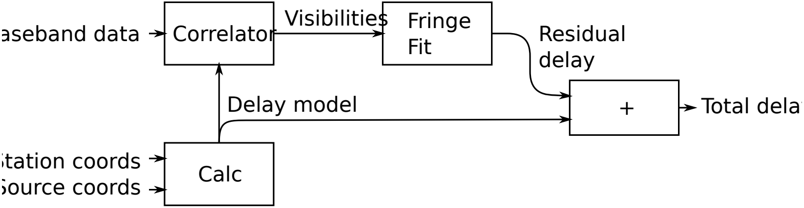
2019 Mar 15



Introduction

* I will explain how correlator products (visibilities) are formed
* I will explain the fringe fit process and creation of total delays
* I cover several Digital Signal Processing topics
* This is a very mathematical subject
  - Some calculation details are in the appendix
  - Several signal processing concepts are explained along the way
  - *Slow me down and ask questions as necessary!*

Overall context of this talk



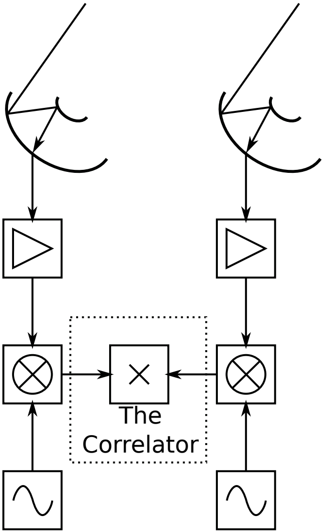
* Delay model calculation already covered
* Making use of total delay will be covered in later talks

Why learn about correlators?

* Understand interferometry data products
* Design interferometric experiments properly
* Implement or improve upon a correlator
* To operate a correlator
* To achieve an enhanced state of enlightenment

VLBI correlators

* Radio antennas/receivers measure electric field vectors
* These are handed to the correlator as voltage time series
* Here we are concerned with cross correlations of these
* 2 (or more) antennas and a correlator form a radio interferometer



Part 1: The real correlator

* Definition
* Correlation of functions
* Correlation of sampled data
* Noise and sensitivity
* The complex-valued visibility

What is a cross-correlator?

Formal definition

Any implementation the cross-correlation function,

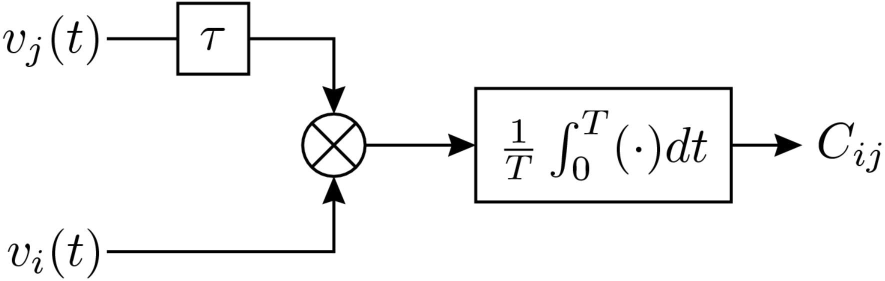
$$C_{ij}(\tau) = \mathrm{Corr}[v_i, v_j] = \langle v_i(t) v_j(t+\tau) \rangle$$

given two real-valued functions, $v_i(t)$ and $v_j(t)$.

Colloquial definition

The device that calculates the above for a VLBI (or other astronomical) observation across 2 or more antennas, each with 1 or 2 polarization components, 1 or more spectral windows with use of delay model functions $\tau_{ij}(t)$ appropriate for the source being studied. The VLBI correlator may also extract pulse cal tones and apply certain calibration to data.

Schematic



Some nuances

* The calculatd value, $C_{ij}(\tau)$, is a statistical quantity
  - Must average over many (independent) samples to be meaningful
  - For a bandwidth of $\Delta\nu$, one independent sample every $\Delta t = 1/2\Delta\nu$.
* Calculation is generally explicitly time-bounded
* Usually is computed on uniformly sampled data:

$$C_{ij}[k] = \frac{1}{N}\sum_{l=1}^{N} v_i[l]v_j[l+k]$$

  with integer k and l
* k or τ is called the *lag*

Example 1

- Use signals $v_1(t) = \sin 2\pi\nu t$ and $v_2(t) = \cos 2\pi\nu t$.
- Take limiting case as time range extends infinitely.

$$
\begin{aligned}
C_{12}(\tau) &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \cos 2\pi\nu t \, \sin 2\pi\nu(t+\tau)\, dt \\
&= -\frac{1}{2} \sin 2\pi\nu\tau
\end{aligned}
$$

- Narrow-band signals have large support over τ.
- Sums of pure tones (as here) have support even as $|\tau| \to \infty$.
- See appendix for detailed derivation.

Normalized correlation coefficient

* Often one is interested in a normalized value (independent of scale)

$$\Gamma_{ij}(\tau) \quad = \quad \frac{\langle v_i(t) v_j(t+\tau)\rangle}{\sqrt{\langle v_i(t)^2\rangle \langle v_j(t+\tau)^2\rangle}}$$

* The denominator is the geometric mean of the two signals' autocorrelations
* Γ_{ij} is a measure of how similar the two signals are
  - $\Gamma_{ij}(\tau) = \pm 1$ if and only if $v_i(t) \propto \pm v_j(t+\tau)$.
  - Otherwise $|\Gamma_{ij}(\tau)| < 1$
* For $v_1(t) = \sin 2\pi\nu t$ and $v_2(t) = \cos 2\pi\nu t$:

$$\Gamma_{ij}(\tau) = -\sin 2\pi\nu\tau$$

* Thus the cosine function is the same as the sine function with a $n - 1/4$ period shift.

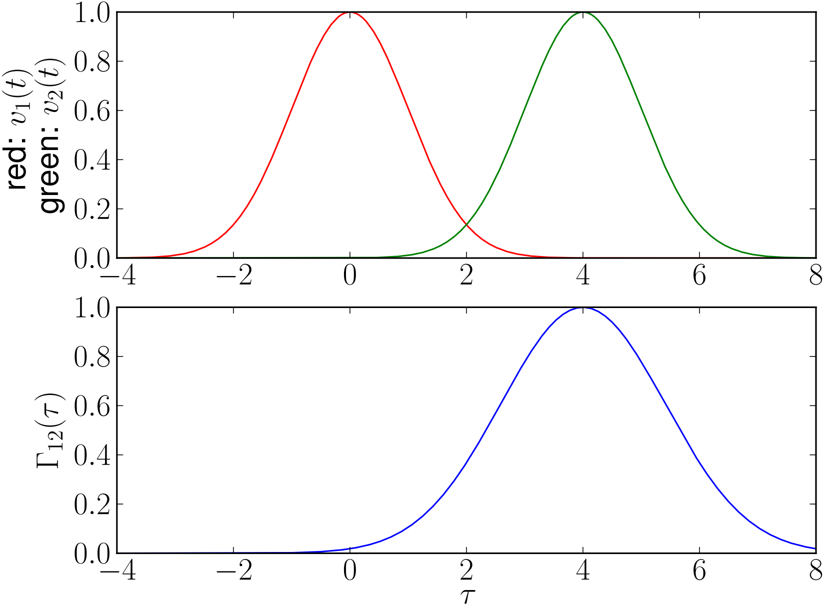
Example 2

* Cross-correlate $v_1(t) = e^{-t^2/2}$ against $v_2(t) = e^{-(t-t_0)^2/2}$
* For simplicity make use of $v_1(t) = v_2(t + t_0)$:

$$\begin{aligned} \Gamma_{12}(\tau) &= \frac{\int_{-\infty}^{\infty} v_1(t) v_2(t+\tau) dt}{\int_{-\infty}^{\infty} v_1(t)^2 dt} \\ &= e^{-(\tau - t_0)^2/4} \end{aligned}$$

* Result could be predicted without grungy math:
  - Correlation of time symmetric signals is equivalent to convolution
  - Convolution of two Gaussians is a wider Gaussian (sum in quadrature)
  - Signals are the same when $\tau = t_0$
* More complete derivation in appendix

Example 2 (continued)



Correlation of sampled data

- Sampled data can be treated in similar manner as a continuous function
- Replace integrals by sums
- Assume here that sampled data streams both be uniformly sampled at same interval, Δt
- Sampled signals *must* be band-limited with $\Delta\nu \leq 1/2\Delta t$ (Nyquist sampling theorem)
- Note: *sampled* does not imply *quantized*; ignore quantization here
- Given $v_i[l]$ and $v_j[l]$, the corresponding quantities are:

$$C_{ij}[k] = \frac{1}{N}\sum_{l=1}^{N} v_i[l]v_j[l+k]$$

$$\Gamma_{ij}[k] = \frac{\sum_{l=1}^{N} v_i[l]v_j[l+k]}{\sqrt{\sum_{l=1}^{N} v_i[l]^2}\sqrt{\sum_{l=1}^{N} v_j[l+k]^2}}$$

Example 3: Seismology

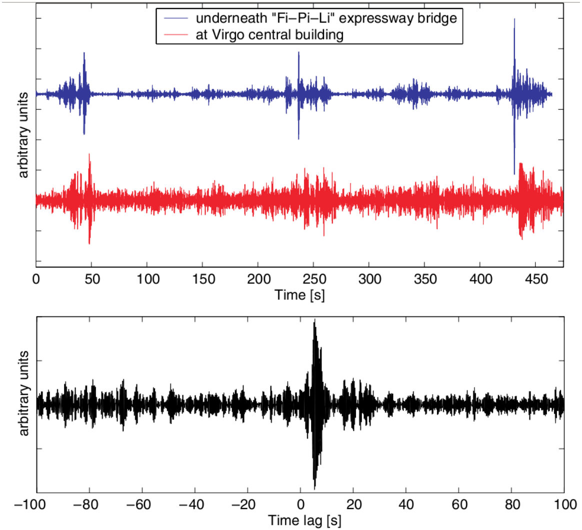
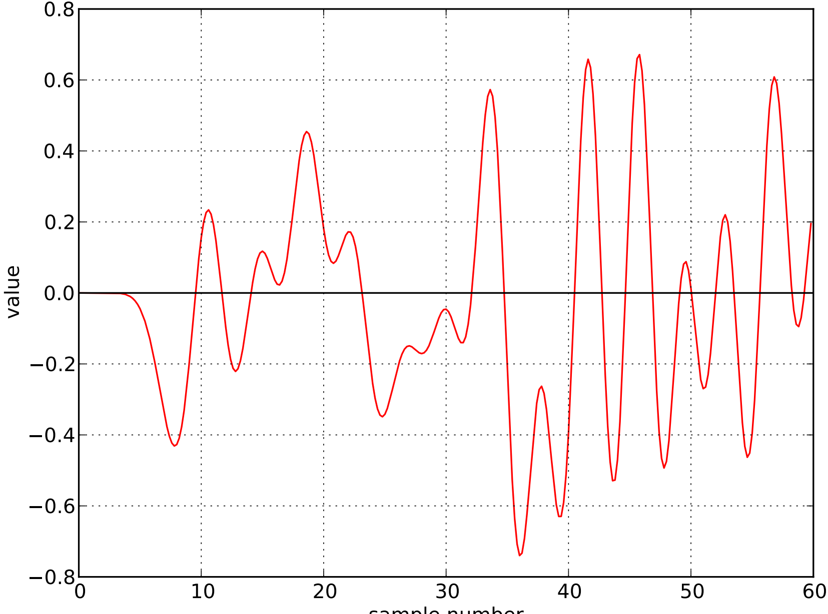
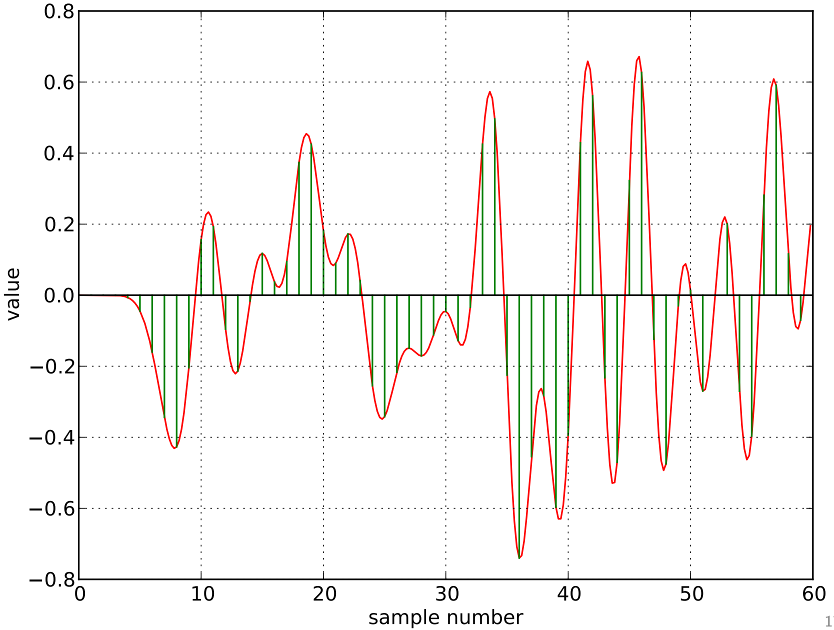


Image from Einstein Telescope design study document, 2011

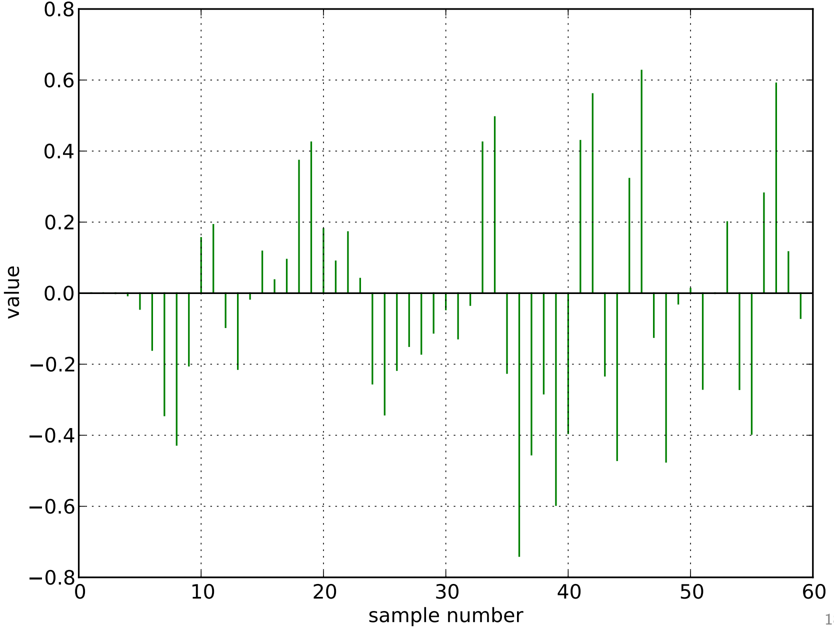
Sampling band-limited signal: original signal



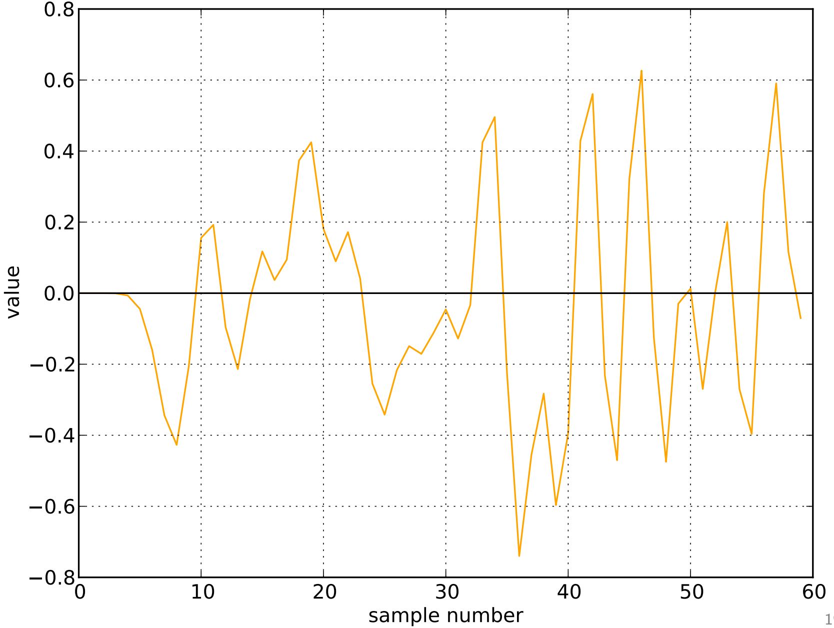
Capture signal every unit interval



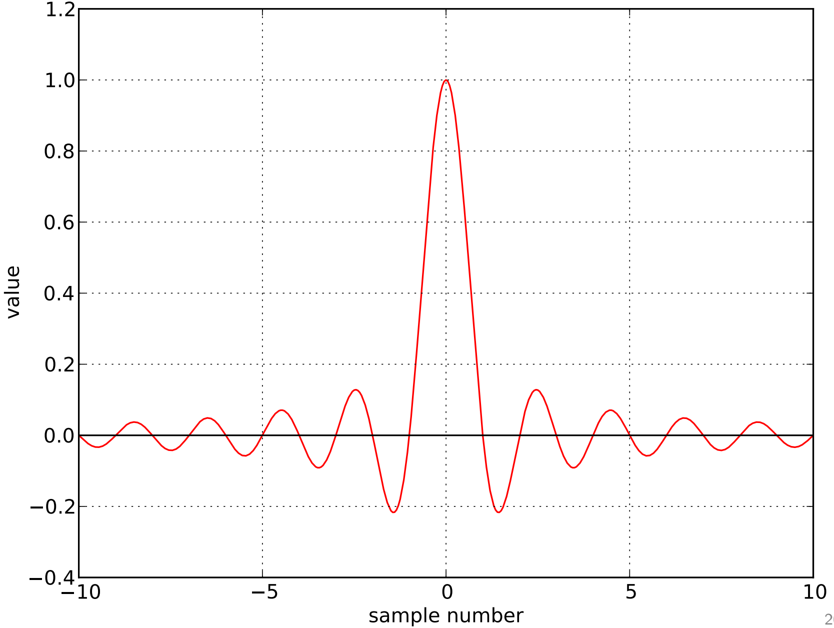
Retain only samples



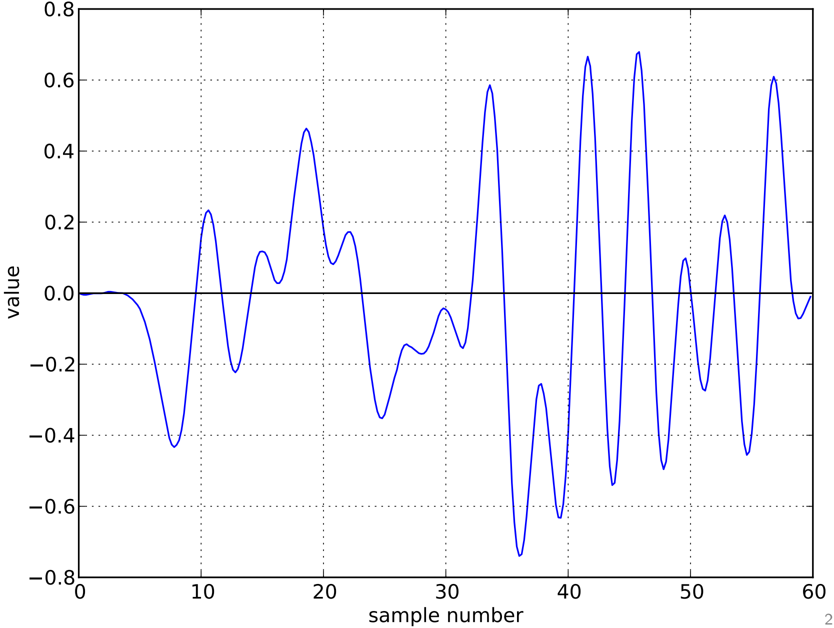
Naive signal reconstruction



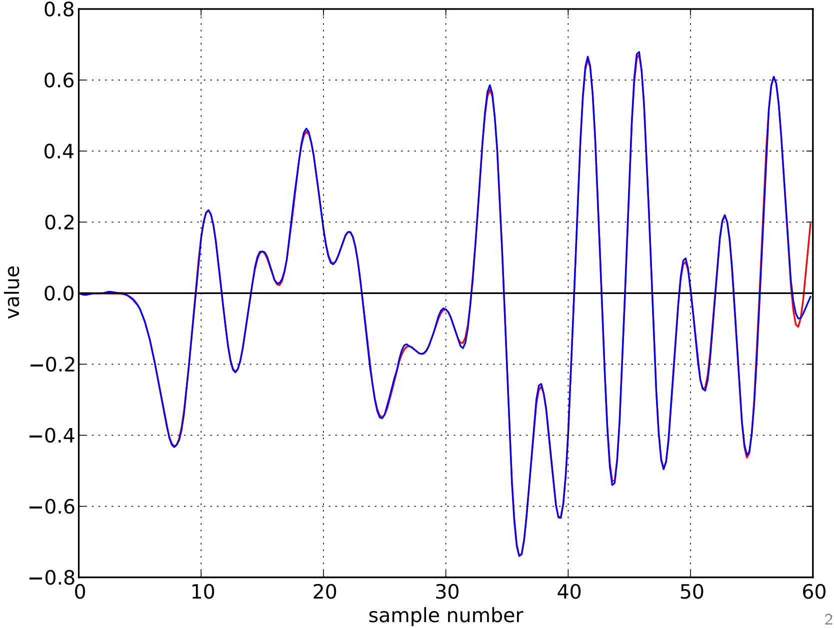
The interpolation function for 1^{st} Nyquist zone (sinc)



Properly interpolated function



Comparison of original and reconstructed signals



Correlation of Gaussian noise

* Take 2 sampled signals, $\mathfrak{g}_1[l]$ and $\mathfrak{g}_2[l]$, where
  * Each $\mathfrak{g}_i[k]$ is drawn from a zero mean, unit norm Normal distribution
  * $\langle \mathfrak{g}_i \rangle = 0$, $\left\langle \mathfrak{g}_i^2 \right\rangle = 1$ (which implies $C_{ij} = \Gamma_{ij}$)
  * $\langle \mathfrak{g}_j \mathfrak{g}_j \rangle = \delta_{ij}$ (defines uncorrelated noise)
* The expectation value of the correlation function vanishes

$$C_{12}[k] = \frac{1}{N} \sum_{i-1}^{N} \mathfrak{g}_1[l] \mathfrak{g}_2[l+k] = 0$$

* But its RMS does not

$$\sigma_{C_{12}[k]} = \frac{1}{\sqrt{N}}$$

* This is the basis for calculating interferometer sensitivity (see appendix)

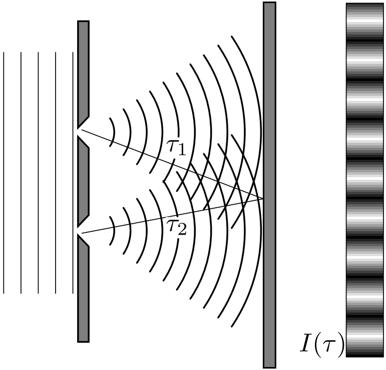
The (optical) double slit experiment

* The film at the image plane of a double slit is a correlator!
* $\tau = \tau_2 - \tau_1$ is the path length difference
* Monochrome signal hits mask $v(t) = \cos 2\pi\nu t$
* Signals at image:

$$\begin{aligned} v_1(t) &= \cos 2\pi\nu(t - \tau_1) \\ v_2(t) &= \cos 2\pi\nu(t - \tau_2) \end{aligned}$$

* Intensity at image:

$$\begin{aligned} I(\tau) &\propto \left\langle (v_1(t) + v_2(t))^2 \right\rangle \\ &= 1 + \cos 2\pi\nu\tau \end{aligned}$$



* This is an *additive correlator*
* The constant term is the total power
* The brightness ripples are *fringes*

Correlation of quasi-monochromatic signals

* As seen before cross correlation of two equal-frequency signals gives sinusoidal response with respect to τ.
* Sinusoids have two free parameters: amplitude and phase.
* Seems silly to need more than two measurements to completely characterize correlator response.
* Solution: measure two lags, separated by 90 degrees of phase!

$$C_{ij}(\tau) = C_{ij}(0)\cos 2\pi\nu\tau + C_{ij}(1/4\nu)\sin 2\pi\nu\tau$$

* For convenience, bundle into a single complex number

$$V_{ij} = C_{ij}(0) + iC_{ij}(1/4\nu)$$

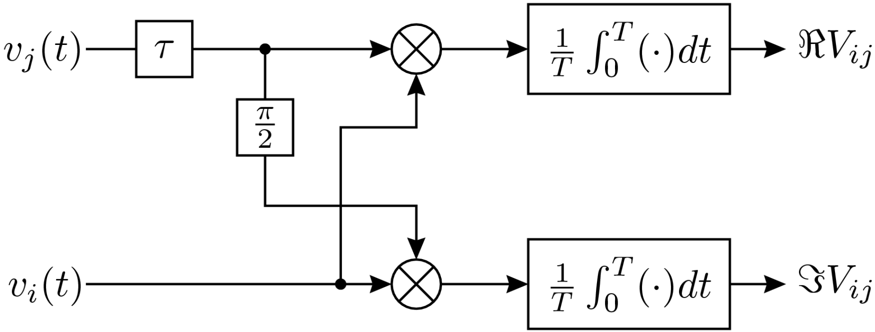
* This is proportional to the familiar *visibility*. And then

$$V_{ij}(\tau) = \mathrm{Re}\left(V_{ij}e^{-2\pi i\nu\tau}\right)$$

Part 2: The complex correlator

* The complex correlator
* The Hilbert transform
* Analytic signals
* Complex sampling

Schematic of complex correlator



$$V_{ij}(\tau) = \langle v_i(t)v_j(t+\tau)\rangle + i\,\langle v_i(t)\mathcal{H}[v_j](t+\tau)\rangle$$

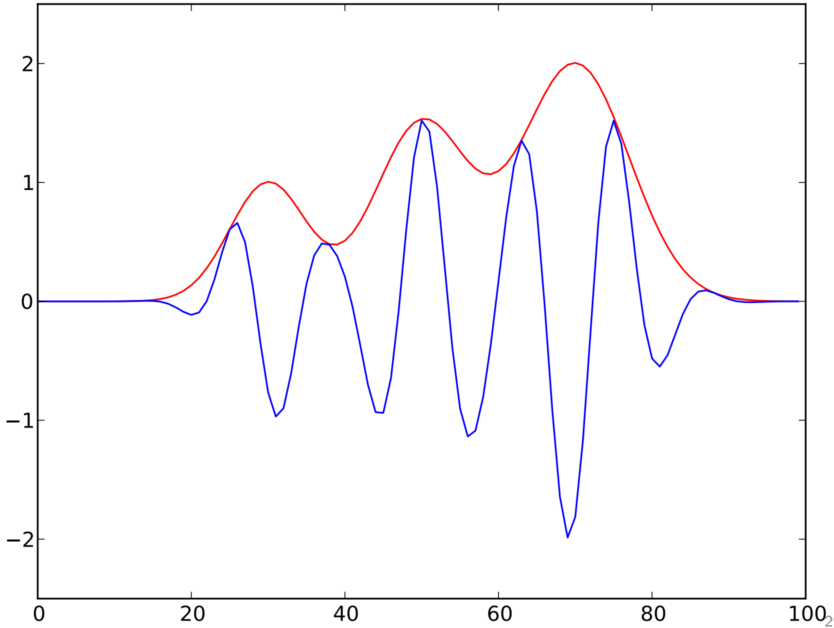
Analytic signals

* Given a real-valued signal $v(t)$, define analytic signal

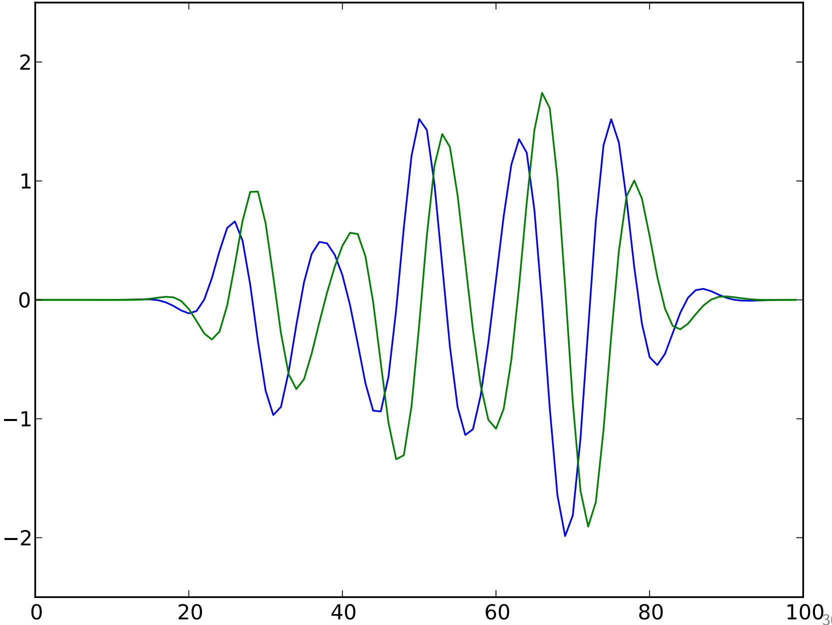
$$w(t) = v(t) + i\mathcal{H}[v(t)]$$

* Here $\mathcal{H}$ is the Hilbert transform
  - $\mathcal{H}[v(t)] = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{v(s)}{t-s} ds$
  - $\cos \rightarrow \sin$ and $\sin \rightarrow -\cos$
  - $\mathcal{H}[\mathcal{H}[v(t)]] = -v(t)$ (the operation is invertable)
* Analytic signals are mathematical tools
  - Allows complex multiplication
  - Simplifies Fourier transforms
  - Simplifies fringe rotation
* Remember: $\mathrm{Im}(w(t))$ is not physical
* See `https://en.wikipedia.org/wiki/Analytic_signal` for a good discussion

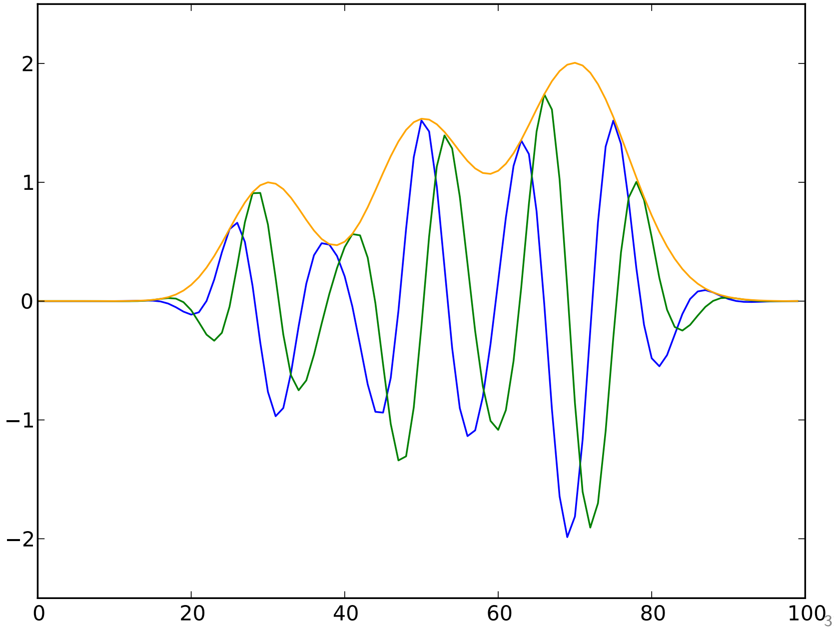
Hilbert transform example: amplitude modulated signal



Compute Hilbert transform (blue + i green is analytic)



Reconstruct envelope (sum blue & green in quadrature)



Analytic signal properties

* Energy content is double:

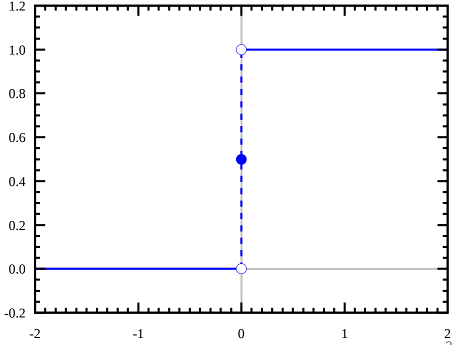
$$\int w(t)^* w(t) dt = 2 \int v(t)^2$$

* Fourier transform has no negative frequency components:

$$\mathcal{F}[w(t)](\nu) = 2H(\nu)\mathcal{F}[v(t)](\nu)$$

where the Heaviside step function is:

$$H(\nu) = \begin{cases} 0 & \nu < 0 \\ 1/2 & \nu = 0 \\ 1 & \nu > 0 \end{cases}$$



Making complex sampled data from real sampled data

* Start with a real sampled signal, $v[k]$
* Can compute the sampled equivalent of an analytic signal using discrete Hilbert transform

$$\mathcal{H}(v)[k] = \begin{cases} \frac{2}{\pi} \sum\limits_{n \text{ odd}} \dfrac{v[n]}{k-n} & k \text{ even} \\ \frac{2}{\pi} \sum\limits_{n \text{ even}} \dfrac{v[n]}{k-n} & k \text{ odd} \end{cases}$$

* Resultant signal, $v[k] + i\mathcal{H}(v)[k]$, carries duplicate information
* Can drop alternate samples to define:

$$w[k] = v[2k] + i\mathcal{H}(v)[2k]$$

* Note that sample rate simply is inverse bandwidth: $\Delta t = 1/\Delta\nu$
  * Clock rate of digital electronics can be halved!

Correlation of complex signals

* Make use of the power of complex numbers
* Note use of complex conjugation: *

$$
\begin{aligned}
\mathrm{Corr}[w_i, w_j] &\equiv \langle w_i^* w_j \rangle \\
&= \langle v_i v_j \rangle + i \langle v_i \mathcal{H}[v_j] \rangle - i \langle \mathcal{H}[v_i] v_j \rangle + \langle \mathcal{H}[v_i] \mathcal{H}[v_j] \rangle \\
&= 2 \langle v_i v_j \rangle + 2i \langle v_i \mathcal{H}[v_j] \rangle \\
&= 2V_{ij}
\end{aligned}
$$

* The above equation holds for continuous or sampled signals
* Equality of the second and third expressions can be shown through spectral analysis

Spectral decomposition of signals

* A band-limited signal can be expressed analytically as

$$w(t) = \int_0^{\Delta\nu} e^{2\pi i t\nu}\, \tilde{w}(\nu)\, d\nu$$

* Where $\tilde{w}(\nu)^*\, \tilde{w}(\nu)$ is proportional to the spectral power density of the signal at frequency ν.
* Nyquist sampling simply captures this each $\Delta t = 1/\Delta\nu$:

$$w[k] = \int_0^{\Delta\nu} e^{2\pi i k\Delta t\nu}\, \tilde{w}(\nu)\, d\nu$$

* The correlation function can be expressed as integral over frequency rather than over time:

$$V_{ij}(\tau) = \int_{\nu_1}^{\nu_2} e^{2\pi i\nu\tau}\, \tilde{w}_i(\nu)^*\, \tilde{w}_j(\nu)\, d\nu$$

Part 3: The lag (XF) correlator

* Concept
  - First cross-multiply and accumulate (X)
  - Then Fourier transform (F)
* Spectral response
* Realization of lag correlators in practice
* Examples of lag correlators

The complex lag correlator

- Fourier transforming a time series leads to its spectrum
- $V_{ij}(\tau)$ is a time series in τ
- What if we discrete Fourier transform it?
- Assume n lags, each spaced by the sample rate, Δt
- V_{ij} is complex-valued, so total bandwidth is $\Delta\nu = 1/\Delta t$

$$\begin{aligned}
\tilde{V}_{ij}[l] &\equiv \sum_{k=-\frac{n-1}{2}}^{\frac{n-1}{2}} e^{-2\pi ikl/n} V_{ij}[k] \\
&= \int_0^{\Delta\nu} A_n\left(\frac{l}{n} - \nu\Delta t\right) \tilde{w}_i(\nu)^* \, \tilde{w}_j(\nu) \, d\nu
\end{aligned}$$

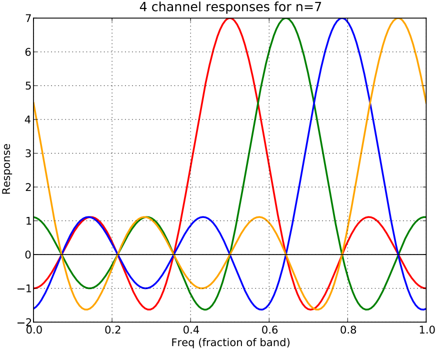
- Where did this come from? See appendix for details.
- What does it mean?

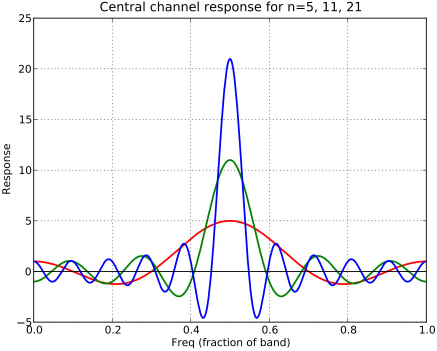
Interpretation

* Lag correlator response is equivalent to complex correlator with additional factor

$$A_n(x) = \sin n\pi x / \sin \pi x$$

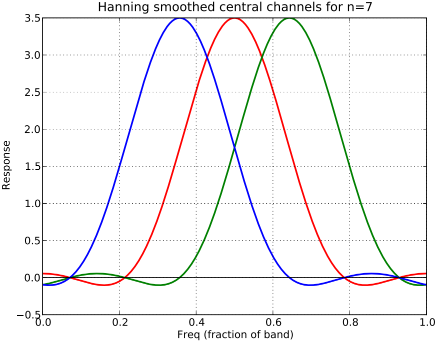
* The function A_n serves as a filter response
* Each output channel, l, has its own filter, shifted by $\Delta\nu/n$

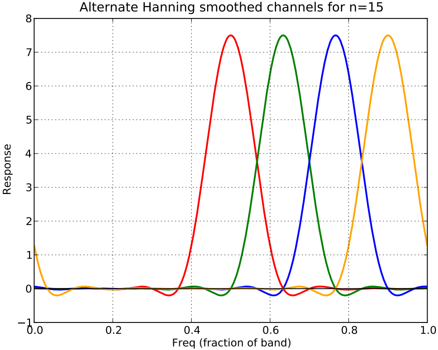




Hanning smoothing

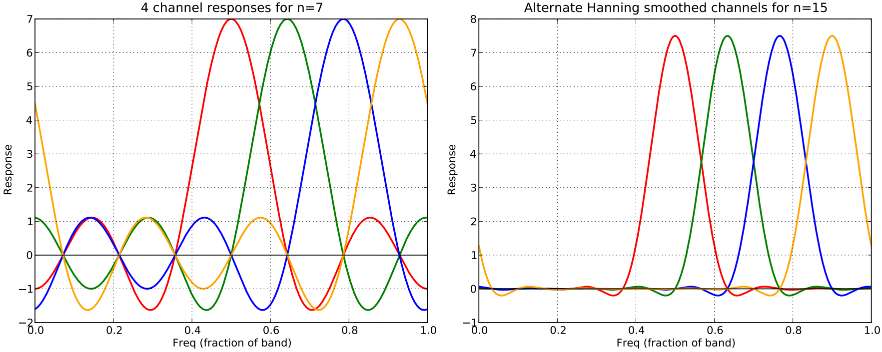
* Damp oscillitory spectra by smoothing with kernel $\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right)$
* Causes wider but much more contained spectral response
  * Can throw out every other channel without loss of information
* Effective in reducing impact of RFI





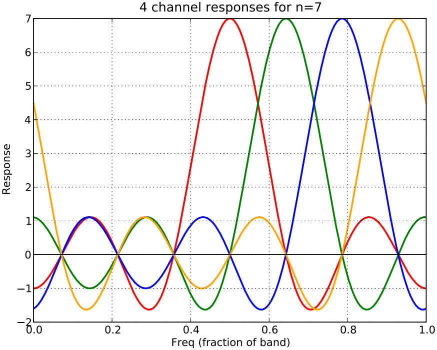
Comparison: with and without Hanning smoothing

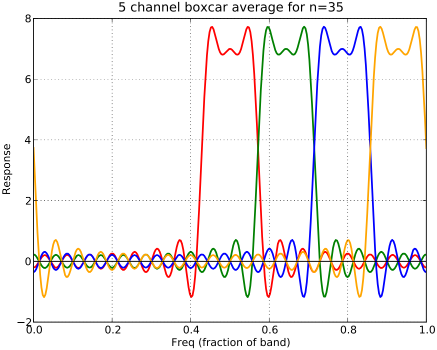
* Two spectra with same number of channels
* Second one improved but comes at higher computational cost
* Implications for RFI immunity?



Comparison: with and without channel averaging

* Two spectra with same number of channels
* Simple channel averaging does rather poor job (Gibb's effect)





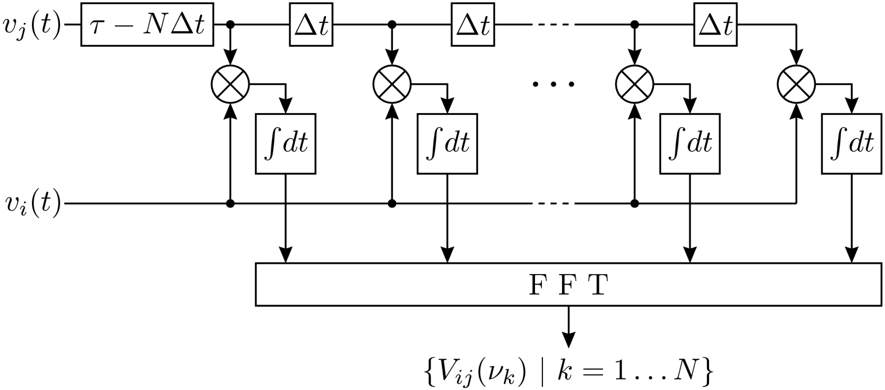
Why all the odd numbers?

- For symmetry sake want equal number of positive and negative lags
  - This is actually important when considering *closure quantities*
- We haven't yet discussed fractional sample correction
  - This allows calculation at $\tau \neq n\Delta t$
- Thus an odd number of lags is natural to consider
- All results generalize to even and odd numbers

Real lag correlators

* Conceptually same as complex lag correlators
* Need twice as many real lags for same response
  * Each lag is half as long (duration of a real sample rather than analytic complex sample)
* Half as many multipliers needed, but they run at twice the rate
* Use real-to-complex Fourier transform
* Spectral expansion of signals uses sines and cosines
* Both real and complex lag correlators used in practice

Schematic of (real) lag correlator

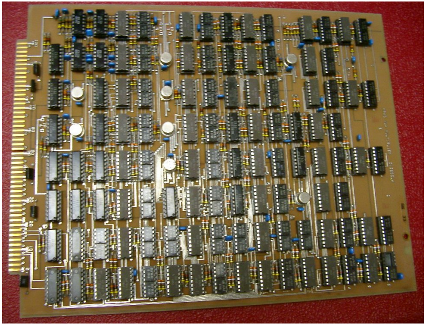


∗ Note: FFT usually performed in software even, on hardware correlators

Examples of lag correlators

* Mark4 (JIVE, Haystack, WACO, Bonn)
* 1997-present (mostly retired)

* Old VLA Correlator
* 1980-2008

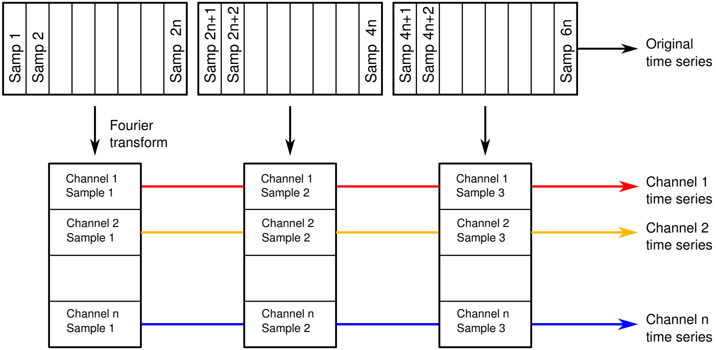


Part 4: The FX correlator

* Filter banks
* Concept
  * First Fourier transform (F)
  * Then cross-multiply and accumulate (X)
* Spectral response
* Realization of FX correlators in practice

FFT filter banks

- FFT incoming bandwidth $\Delta\nu$ real signal in blocks of $2n$
  - Shown below
- or FFT incoming bandwidth $\Delta\nu$ complex signal in blocks of n
- Produce n complex time series, each with bandwidth $\Delta\nu/n$



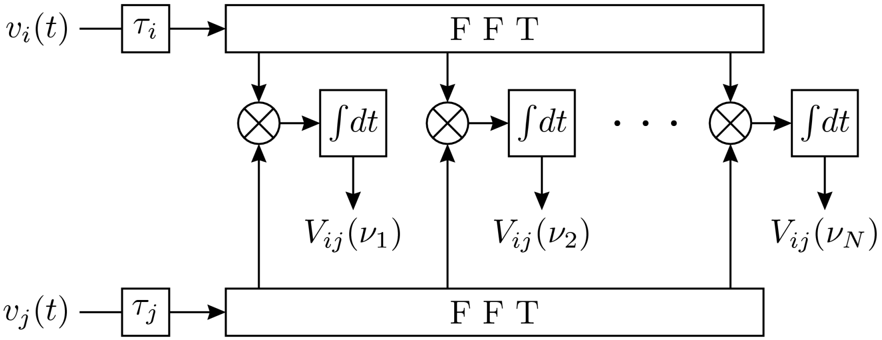
FFT filter bank frequency response

* Starting from a complex sampled signal, the filter bank output is:

$$
\begin{aligned}
\tilde{w}[l] &= \sum_{k=-\frac{n-1}{2}}^{\frac{n-1}{2}} e^{-2\pi ikl/n}\, w[k] \\
&= \int_0^{\Delta\nu} \sum_{k=-\frac{n-1}{2}}^{\frac{n-1}{2}} e^{2\pi ik(\Delta t\nu - l/n)}\, \tilde{w}(\nu)\, d\nu \\
&= \int_0^{\Delta\nu} A_n \left(\frac{l}{n} - \nu\Delta t\right) \tilde{w}(\nu)\, d\nu
\end{aligned}
$$

* Note symmetric summation; through universal relabeling of samples by 1/2 sample, an even number of samples can be accomodated.
  - Not possible in lag case because the parameter was the lag itself.
  - The process is equivalent to a *shifted FFT*

Schematic of FX correlator



FX correlator frequency response

* The visibility is computed as

$$\begin{aligned}\tilde{V}_{ij}[l] &= \langle \tilde{w}_i[l]^* \, \tilde{w}_j[l] \rangle \\ &= \int_0^{\Delta\nu} \left[A_n \left(\frac{l}{n} - \nu \Delta t \right) \right]^2 \tilde{w}_i(\nu)^* \, \tilde{w}_j(\nu) \, d\nu\end{aligned}$$

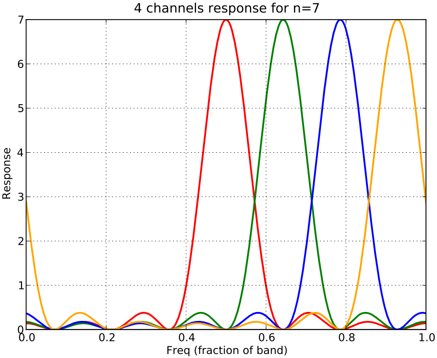
* Similar to lag correlator response but with extra factor of $A_n()$
  - Each filterbank contributes one factor

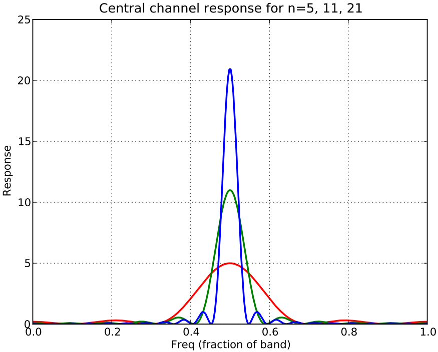
FX correlator frequency response

- Each channel's response is similar to that of the sinc^2 function:

$$A_n(x)^2 = (\sin n\pi x / \sin \pi x)^2$$

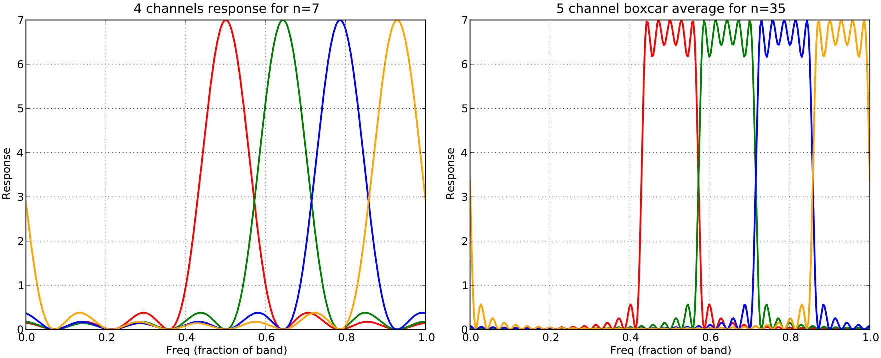
- Generally better than lag corretor output but worse than Hanning smoothed lag correlator output.





Comparison: with and without channel averaging

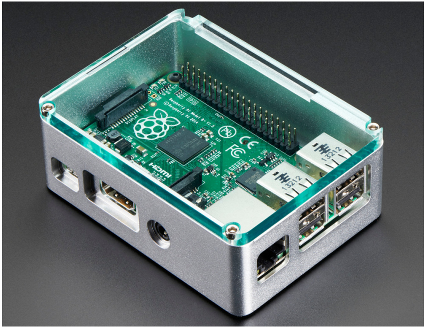
* Two spectra with same number of channels
* Simple channel averaging
* Neighboring channels fairly well isolated
* Peak sidelobes still rather high



Examples of FX correlators

* VLBA hardware correlator
* 1992-2009

* Most software correlators (e.g., DiFX and SFXC)
* Not tied to particular hardware
* DiFX can run on a Raspberry Pi!



What information is needed to correlate?

Time for you to brainstorm. . .

What information is needed to correlate?

* Start and stop times
* Frequencies of observation + bandwidth
* Location of the data
* Format of the data
* Location of antennas
* Coordinates of the source
* Clock offsets
* Correlator parameters: time and spectral resolution

Part 5: Fractional sample delay and fringe rotation

* Effect of delay error
* Fractional sample delay compensation
* Fringe rotation

Effect of a delay error

* Assume a broadband signal of uniform spectral density $|\tilde{w}(\nu)| = 1$
* Look at auto-correlation with a time lag of τ
* Consider one correlator channel with ideal spectral response between ν_1 and ν_2.

$$
\begin{aligned}
C_{ii}(\tau) &= \frac{1}{\nu_2 - \nu_1} \int_{\nu_1}^{\nu_2} d\nu \, e^{-2\pi i \tau \nu} \tilde{w}_i(\nu)^* \tilde{w}_i(\nu) \\
&= \frac{1}{\nu_2 - \nu_1} \int_{\nu_1}^{\nu_2} d\nu \, e^{-2\pi i \tau \nu} \\
&= e^{-2\pi i \tau \nu_0} \frac{\sin \pi \tau \Delta\nu}{\pi \tau \Delta\nu}
\end{aligned}
$$

* Where $\nu_0 = \frac{1}{2}(\nu_1 + \nu_2)$ is the channel center frequency
* And $\Delta\nu = \nu_2 - \nu_1$ is the channel bandwidth

Effect of a delay error

* There are two effects:
  * There is a phase shift of $2\pi\tau\nu_0$
  * There is an amplitude reduction (decorrelation) by amount

$$\frac{\sin \pi\tau\Delta\nu}{\pi\tau\Delta\nu}$$

* The phase error is correctable
* The amplitude can be restored
  * But decorrelation (loss of SNR) is permanent
  * This is devistating unless $\tau \ll 1/\Delta\nu$
* Remember these effects for when we discuss fringe fitting...

Fractional sample compensation

* As one observes an astronomical source, the correlator delay model, τ must change as the source moves across the sky.
* Source motion is smooth with time.
* Bulk delay is compensated by choosing which samples to correlate.
* Each incoming datastream can be offset from integer sample by as much as $\pm\frac{1}{2}$ of a sample.
* Compensation is handled differently on different correlator architectures.
* Spectral line (multi-channel) correlators simplify life: $\Delta t \ll 1/\Delta\nu$ in most cases so effective delay error is reduced.

Fringe rotation

* Essentially the time-dependent fractional sample compensation
* Various possible places to implement:
  - At end of each visibility spectrum calculation (as phase gradient)
  - During accumulation, after each FFT (as phase gradient; FX-only)
  - In time domain, directly on each sample (sample phase rotation)
* Magnitude depends on frequency, not bandwidth!
* Remember! Want to keep phase change well under 1 radian over any averaging period.

Post-integration fringe rotation

* The least costly (in terms of operations)
* Phase applied to visibility spectrum (part of fractional sample corr.)

$$V_{12}(\tau, \nu) = e^{-2\pi i \nu \Delta\tau} V_{12}([\tau], \nu)$$

* Where $[\tau]$ is the delay corresponding to the nearest integer number of samples, and
* $\Delta\tau = \tau - [\tau]$ is the fractional sample being compensated.
* This is valid when $T\dot{\tau}\nu \ll 1$
* Example: $b = 1$ km equatorial baseline at $\nu = 1$ GHz at zenith passage
    * Phase as function of time: $\phi(t) = 2\pi\nu b \sin(2\pi t/86400)/c$
    * The fringe rate, $\dot{\phi}(t) = 4\pi^2 \nu b \cos(2\pi t/86400)/(86400c)$, peaks at 1.5 rad/sec.
    * Thus post-integration fringe rotation is valid for $T \ll 0.6$ sec
* Often done on sub-integration basis.

Post-FFT fringe rotation (FX-only)

* With higher fringe rates, fringe rotation must be done on shorter timescales.
* FX correlators expose the spectrum after each FFT.
* Typical continuum correlator output has frequency resolution of 0.25 MHz, implying FFT timescales of 4μs.
* On a 8611 km baseline (longest VLBA), this is OK for $\nu \ll 20$ GHz.
* Use with care on continent-scale VLBI arrays!

Time-domain fringe rotation

- This is the most common form of fringe rotation used by VLBI
- Simply multiply each sample by $e^{2\pi i \nu_0 \Delta\tau}$ before correlating
- This makes for a complex-valued signal
  - But it is not an analytic signal!
- Note! This technique only works well for small fractional bandwidths
  - Same phase applied to all frequencies
  - Results in decorrelation near band edges by $\mathrm{sinc}\left(\pi \frac{\Delta\nu}{\nu_0}\right)$
  - Worst cast at VLBA: 128 MHz BW centered around 1.28 GHz
    - 1.6% decorrelation at band edge
    - 0.5% decorrelation averaged over band
    - This is still generally acceptable
  - Decorrelation grows as $\left(\frac{\Delta\nu}{\nu_0}\right)^2$

Part 6: The DiFX correlator

What DiFX (Distributed FX) does

* Decode incoming data
* Select data (coarse time delay)
* Fringe rotate
* Fourier transform
* Select sideband
* Apply fractional delay correction
* Cross-multiply
* Short-term accumulate
* Long-term accumulate
* Write visibility to disk
* *You will run DiFX at the demo tomorrow*

Part 7: Miscellaneous correlator topics

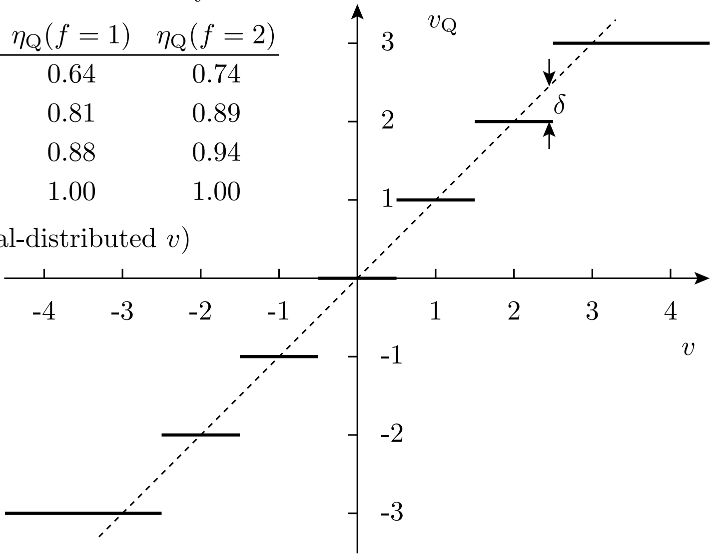
* Quantization
* Pulsar gating
* Other correlator functionality
* Design trade-offs

Quantization noise

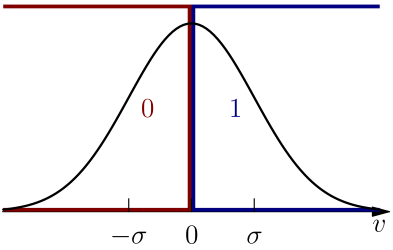
Quantization efficiency

levels	$\eta_Q(f=1)$	$\eta_Q(f=2)$
2	0.64	0.74
3	0.81	0.89
4	0.88	0.94
∞	1.00	1.00

(For normal-distributed v)



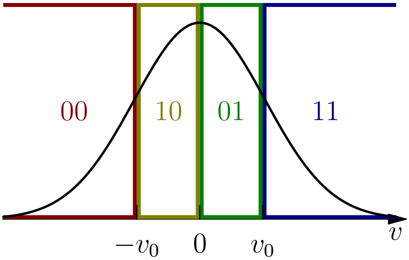
Real 2-state (1-bit) quantization



Code	Range	Value	Frac.
0	$-\infty$ to 0	$-\sqrt{2/\pi}\,\sigma$	50%
1	0 to ∞	$\sqrt{2/\pi}\,\sigma$	50%

* Values determined so as to minimize quantization noise
* Quantization efficiency $\eta_Q = 64\%$
* Effective number of bits, ENOB $= 1$

Real 4-state (2-bit) quantization



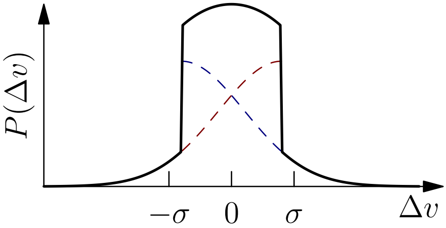
Code	Range	Value	Frac.
00	$-\infty$ to $-v_0$	$-\alpha R$	17%
10	$-v_0$ to 0	$-\alpha$	33%
01	0 to v_0	α	33%
11	v_0 to ∞	αR	17%

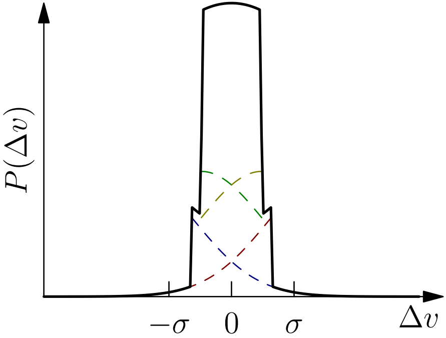
* Optimal values: $v_0 = 0.96\sigma$; $R = 3.3359 \longrightarrow \eta_Q = 88\%$
* ENOB $= 1.92$
* $\alpha = 0.4780\,\sigma$ determined so as to minimize quantization noise

Note: Different conventions for the codes exist (e.g., Mark5B, VDIF)

Quantization noise distribution

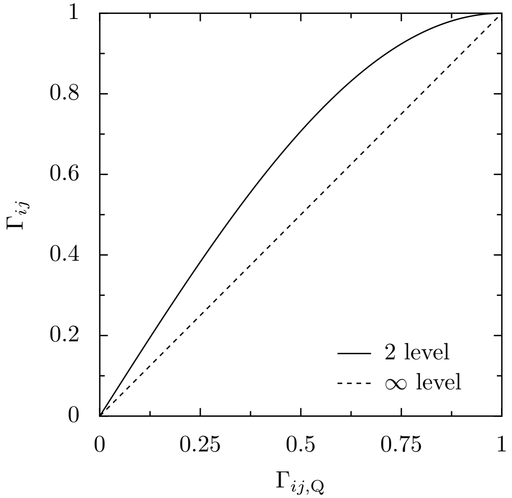
* Quantization noise is non-Gaussian
* Approaches uniform distribution
* Distributions for 1-bit and 2-bit sampling shown





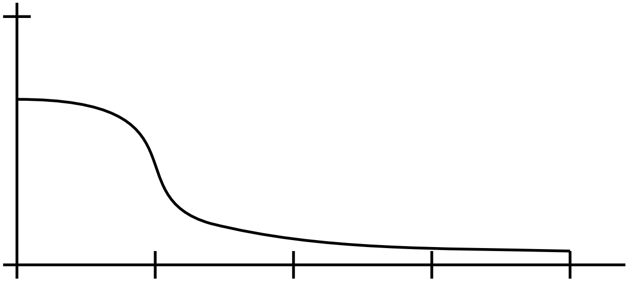
Quantization correction

* At low correlation, quantization decreases correlation
* Quantization causes predictable non-linearity at high correlation
* Linear correction is easy; full correction is complicated . . .



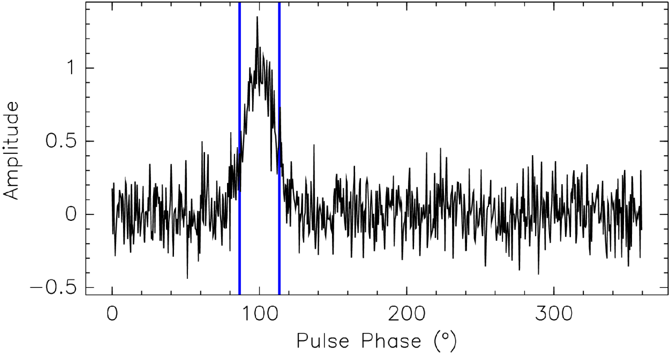
Quantization in the spectral domain

* 1-bit quantization is extreme case of harmonic distortion
* Power gets scattered into harmonics
* Oversampling allows partial discrimination of unwanted harmonics
  ◦ Increases signal to noise
  ◦ At a substantial data transmission cost
  ◦ Very quickly diminishing returns; better to use more bits



Pulsar gating

- Pulsars emit regular pulses with small duty cycle
- Period in range 1 ms to 8 s; usually $\Delta t \ll P_{\rm pulsar} < T$
- Blanking during off-pulse improves sensitivity
- Propagation delay is frequency dependent: best done on FX architecture



Other correlator functionality

* Pulse cal extraction
* Switched power extraction
* Data weights
* Multiple phase centers
* Spectral zooming
* Band matching
* Overlapped FFTs

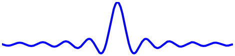
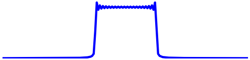
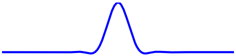
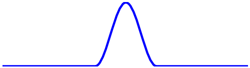
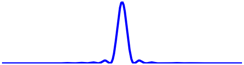
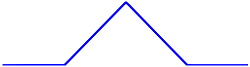
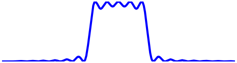
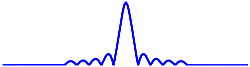
Trade-offs: hardware vs. software

* Hardware advantages
  - Can be 10-100× faster
  - Can be 10-100× more power efficient
  - Predictable operations once commissioned (usually)
  - Guaranteed real-time performance
* Software advantages
  - Short development timescales
  - COTS Hardware: cost effective
  - Generally more flexible
  - Extensible, even after deployed
* GPU-based correlators straddle the two
  - Higher compute density than CPUs
  - Less flexibility than CPUs
  - More difficult development than CPUs

Trade-offs: lag or FX architecture?

- Lag (XF) advantages
  - Can implement weights more precisely
  - Individual operations can be performed with small word sizes
  - Access to uncorrupted lag spectrum
    - Improved quantization correction
- FX advantages
  - Many fewer operations (increasingly so with larger spectra)
  - Improved native spectral response
  - Access to frequency domain on short timescales
    - Zoom bands and band-matching
    - More effective pulsar gating
    - Sub-integration RFI characterization

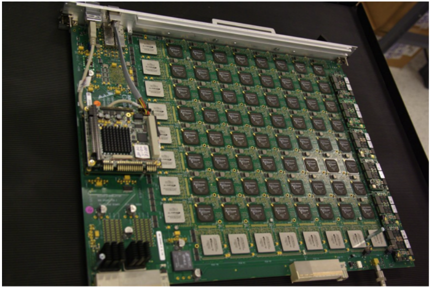
Spectral response and delay window duality

Processing	Spectral response	Delay window
lag		
lag w/Hanning		
FX		
FX w/boxcar		

* Related by Fourier transform
* Must take into consideration when calculating fringe SNR!

Hybrid correlators

* Example: Jansky VLA's WIDAR correlator
* 2008-present
* "Filter-bank XF" architecture
* Filterbank forms complex-valued sub-bands
* Each sub-band feeds a complex lag correlator



Left: WIDAR during construction   Right: WIDAR baseline board

Part 7: Practical considerations

* 1- or 2-bit quantization?
* What spectral resolution (or number of lags) is needed?
* What time resolution is needed?
* Do I need to generate all polarization products?

1- or 2-bit quantization?

* 1-bit sampling
  - Quantization efficiency $\eta_Q = 0.636$
  - Simplest to implement
* 2-bit sampling
  - $\eta_Q = 0.882$
  - $\eta_Q/\sqrt{2} = 0.624$
  - Slightly lower sensitivity at fixed bitrate
  - Quantization correction more linear
    - Improved performance in RFI environment
* Are there better quantization schemes? Yes…
* Why can 1- and 2-bit quantization work?
  - Absolute amplitude restored with total power measurements
  - The statistics of correlation are what matter

What spectral resolution is needed?

* Must obey frequency-time resolution product $\delta\nu\Delta t > 1$
  - Otherwise you are not correlating independent samples
  - Very rarely is this a limitation
* At low frequency RFI excision is better with more channels; can throw out affected data
  - Usually this means $\delta\nu \ll \Delta\nu$
* To accommodate typical clock uncertainties, open up the delay window to at least $\pm 2\mu$s
  - Implies spectral resolution of 0.25 MHz or better
* Number of (real) lags to accomplish is simply $2\delta\nu/\Delta\nu$

What time resolution is needed?

* Time resolution, $\Delta t =$ accumulation period
* Δt should be smaller than timechange of correlator statistics
  - Atmospheric / ionospheric pathlength changes
  - Delay model not accurate (antenna or source position error)
    - Residual rates usually measured in mHz or 10s of mHz
  - RFI environment
  - Source structure or spectrum change (e.g., pulsars)
* For most cm-wave VLBI, including most geodetic processing 1 or 2 seconds is fine
* For mm-wave VLBI and space VLBI, smaller number is usually needed

Should I correlate all polarization products?

* It depends
* Generally no harm in doing so, but increases computational load and output data size
* Many times not possible (e.g., only one polarization recorded)
* Are polarizations linear?
  - Probably; polarization basis rotates on sky differently at each antenna
  - Not required for array of equatorial mounted antennas
* Are there mixed linear and circular systems?
  - Yes, otherwise you will reduce your sensitivity on some baselines

Part 8: Fringe fitting

* The most primitive analysis step after correlation
* Data: time series of visibility spectra
  - For a single source
  - From within a single common spectral window
  - For a particular polarization product
  - Over a short enough time to prevent atmospheric decorrelation
  - Over a long enough time to achieve sensitivity requirements
* Remember: consequence of incorrect delay model:

$$\Delta\phi(\nu) = 2\pi\tau\nu$$

  where τ here is interpreted as the delay error
* Residual phase, $\Delta\phi$, is proportional to frequency
* Delay error cause: antenna/source positions, clock error, atmosphere

Delays

* Phase delay

$$D_\phi = -\frac{1}{2\pi}\frac{\phi}{\nu}$$

* Group delay

$$D_G = -\frac{1}{2\pi}\frac{d\phi}{d\nu}$$

* In non-dispersive media (e.g., vacuum) these are equal
* In dispersive media (e.g., ionosphere) they differ
* Group delay is the direct observable of VLBI observations
* Phase delay suffers from 2π ambiguities

Rates

* Phase rate

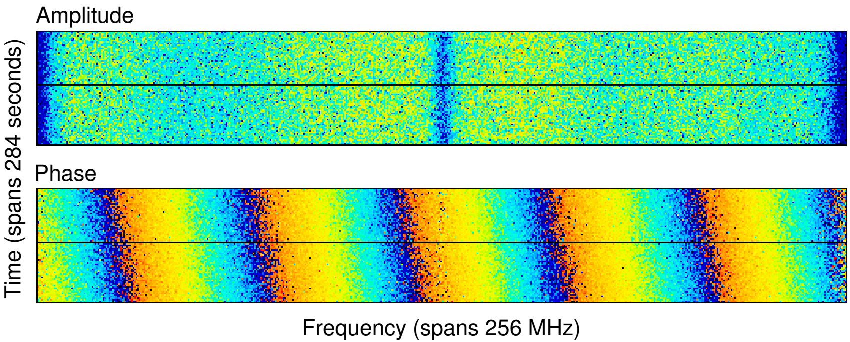
$$R_\phi = \frac{1}{2\pi}\frac{d\phi}{dt}$$

* Delay rate

$$R_D = \frac{1}{2\pi\nu}\frac{d\phi}{dt}$$

* The two quantities are easily and unambiguously convertable
* Fringe fitting naturally produces phase rate

Example data



$*$ Two adjacent 128MHz bands on HN-LA baseline at 4.8GHz

$*$ By eye: approx 5.5 turns of phase across 256MHz

$\longrightarrow$ $D_G \sim 5.5/256\text{MHz} = 21.5\text{ns}$

$*$ AIPS FRING result:

$\longrightarrow$ $D_G = 21.7\text{ns} \quad R_\phi = -0.6\text{mHz} \quad R_D = -1.3 \times 10^{-13}\text{sec/sec}$

Fringe fitting assumptions

* Observe sufficiently bright, sufficiently point-like source
* Antennas' bandpass response relatively phase flat
* During solution interval delay error evolves linearly in time
* Then phase can be expressed as:

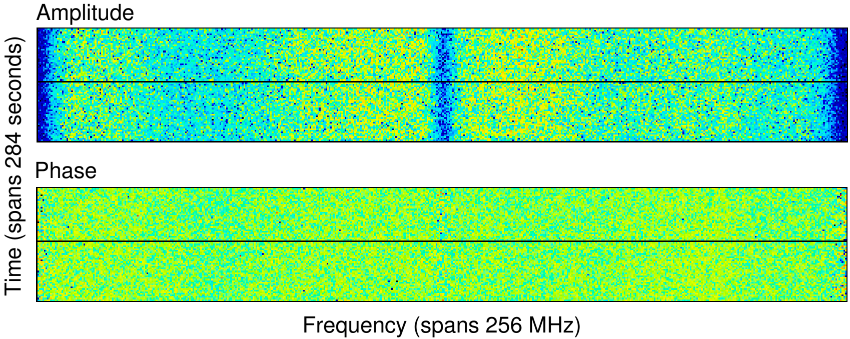
$$\phi(\nu, t) = \phi_0 - 2\pi D_G \nu + 2\pi R_\phi t$$

* Where ϕ_0 is the phase at the reference time and frequency
* Nice linear equation with 3 parameters, right?
* Not quite: phase only measured modulo 2π

General fringe fitting approach

* 2D FFT on $\phi(\nu, t)$
  - Maybe oversampled
* Identify peak valued pixel
* Centroid the peak $\longrightarrow (D_G, R_\phi, \phi_0)$ estimate
* Subtract $\phi_0 - 2\pi D_G \nu + 2\pi R_\phi t$ estimate from phases
  - Phases now all close to zero; 2π ambiguities less important
* Perform least-squares fit to these residuals
* Add fit values to estimates

Example data (corrected)



* Remaining phase ripple due to antenna bandpass
* Note increased phase noise near band edges

What could go wrong?

* No FFT peaks $\longrightarrow$ no solution
* Multiple FFT peaks $\longrightarrow$
* Aliased fringes

Appendices

* Trigonometric identities
* Symmetric power series sum
* Correlation of cosine and sine functions
* Correlation of Gaussian pulses
* Correlation of signals with noise
* Interferometer sensitivity

Trigonometric identities

$$
\begin{aligned}
\cos(x \pm y) &= \cos x \cos y \mp \sin x \sin y \\
\sin(x \pm y) &= \sin x \cos y \pm \cos x \sin y \\
\cos x \cos y &= \frac{1}{2}\left[\cos(x-y) + \cos(x+y)\right] \\
\cos x \sin y &= \frac{1}{2}\left[\sin(x-y) + \sin(x+y)\right] \\
\sin x \sin y &= \frac{1}{2}\left[\cos(x-y) - \cos(x-y)\right] \\
e^{ix} &= \cos x + i \sin x \\
\cos x &= \frac{e^{ix} + e^{-ix}}{2} \\
\sin x &= \frac{e^{ix} - e^{-ix}}{2i}
\end{aligned}
$$

Symmetric power series sum

$$
\begin{aligned}
A_{2m+1}(x) &= \sum_{k=-m}^{m} e^{-i2\pi xk} \\
&= \sum_{k=-m}^{\infty} e^{-i2\pi xk} - \sum_{k=m+1}^{\infty} e^{-i2\pi xk} \\
&= \left(e^{i(2m)\pi x} - e^{-i(2m-2)\pi x}\right) \sum_{k=0}^{\infty} e^{-i2\pi xk} \\
&= \left(e^{i(2m)\pi x} - e^{-i(2m-2)\pi x}\right) \frac{1}{1 - e^{-i2\pi xm}} \\
&= \frac{e^{i(2m+1)\pi x} - e^{-i(2m+1)\pi x}}{e^{i\pi xm} - e^{-i\pi xm}} \\
&= \frac{\sin(2m+1)\pi x}{\sin \pi x} \longrightarrow A_n(x) = \frac{\sin n\pi x}{\sin \pi x}
\end{aligned}
$$

Note: in the limit that $n \to \infty$, $A_n(x) \to \delta(x)$, $\frac{A_n(x/n)}{n} \to \operatorname{sinc} 2\pi \mathrm{x}$.

The lag correlator in detail

$$
\begin{aligned}
\tilde{V}_{12}[l] &\equiv \sum_{k=-\frac{n-1}{2}}^{\frac{n-1}{2}} e^{-2\pi i k l/n} V_{12}[k] \\
&= \sum_{k=-\frac{n-1}{2}}^{\frac{n-1}{2}} \frac{1}{N} \sum_{j=1}^{N} e^{-2\pi i k l/n}\, w_1[j]^*\, w_2[j+k] \\
&= \int_0^{\Delta\nu} d\nu_1 \int_0^{\Delta\nu} d\nu_2 \sum_{k=-\frac{n-1}{2}}^{\frac{n-1}{2}} \frac{1}{N} \sum_{j=1}^{N} e^{-2\pi i k l/n} \\
&\quad \times\, e^{2\pi i \Delta t j \nu_1}\, \tilde{w}_1(\nu_1)^*\, e^{2\pi i \Delta t (j+k) \nu_2}\, \tilde{w}_2(\nu_2) \\
&\quad \ldots
\end{aligned}
$$

The lag correlator in detail ...

$$
\begin{aligned}
\tilde{V}_{12}[l] &= \int_0^{\Delta\nu} d\nu_1 \int_0^{\Delta\nu} d\nu_2 \sum_{k=-\frac{n-1}{2}}^{\frac{n-1}{2}} e^{-2\pi i k\left(\frac{l}{n}-\nu_2\Delta t\right)} \\
&\quad \times \frac{1}{N}\sum_{j=1}^{N} e^{2\pi i \Delta t j(\nu_2-\nu_1)}\, \tilde{w}_1(\nu_1)^*\, \tilde{w}_2(\nu_2) \\
&\sim \int_0^{\Delta\nu} d\nu_1 \int_0^{\Delta\nu} d\nu_2\, A_n\left(\frac{l}{n}-\nu_2\Delta t\right) \delta(\nu_2-\nu_1)\, \tilde{w}_1(\nu_1)^*\, \tilde{w}_2(\nu_2) \\
&= \int A_n\left(\frac{l}{n}-\nu\Delta t\right)\, \tilde{w}_1(\nu)^*\, \tilde{w}_2(\nu)\, d\nu
\end{aligned}
$$

* Here $A_n(x) = \sin n\pi x / \sin \pi x$
* A_n is related to the sinc function
* See previous appendix for Derivation of function A_n

Correlation of cosine and sine functions w/ real correlator

$$
\begin{aligned}
C_{12}(\tau) &= \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} \cos 2\pi t \, \sin 2\pi(t+\tau)\, dt \\
&= \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} \frac{1}{2}\left[-\sin 2\pi\nu\tau + \sin 2\pi\nu(2t+\tau)\right] dt \\
&= \lim_{T\to\infty} \frac{1}{4T}\left[-t\sin 2\pi\nu\tau - \frac{1}{4\pi\nu}\cos 2\nu(2t+\tau)\right]_{-T}^{T} \\
&= \lim_{T\to\infty} -\frac{1}{2}\sin 2\pi\nu\tau + \mathcal{O}\left(\frac{1}{T}\right) \\
&= -\frac{1}{2}\sin 2\pi\nu\tau
\end{aligned}
$$

Correlation of Gaussian pulses

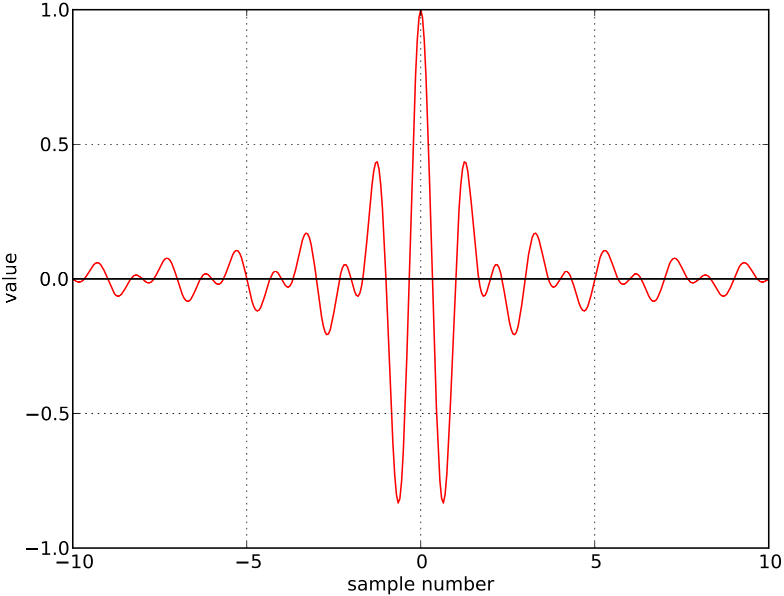
$$\begin{aligned}
\Gamma_{ij}(\tau) &= \frac{\int_{-\infty}^{\infty} e^{-t^2/2} e^{-(t-t_0+\tau)^2/2} dt}{\int_{-\infty}^{\infty} e^{-t^2} dt} \\
&= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-(t+t_0/2-\tau/2)^2/2} e^{-(t-t_0/2+\tau/2)^2/2} dt \\
&= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-t^2} e^{-(\tau-t_0)^2/2} dt \\
&= e^{-(\tau-t_0)^2/2}
\end{aligned}$$

Note use of Gaussian integral identity (twice):

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$$

The interpolation function for 2$^{\text{nd}}$ Nyquist zone

$$(\sin 2x - \sin x)/x$$



Correlation of signals with noise (at zero delay)

* $\mathrm{Corr}[v_i, v_j]$ is *bilinear* in its signal arguments:

$$\begin{aligned}\mathrm{Corr}[\alpha\, a + \beta\, b, \gamma\, c + \delta\, d] &= \alpha\,\gamma\,\mathrm{Corr}[a, c] \\ &+ \alpha\,\delta\,\mathrm{Corr}[a, d] \\ &+ \beta\,\gamma\,\mathrm{Corr}[b, c] \\ &+ \beta\,\delta\,\mathrm{Corr}[b, d]\end{aligned}$$

* A simplistic signal model for observation of a point source is

$$\begin{aligned}v_1[k] &= S[k] + N_1[k] = \sqrt{s}\,\mathfrak{g}_0[k] + \sqrt{n_1}\,\mathfrak{g}_1[k] \\ v_2[k] &= S[k] + N_2[k] = \sqrt{s}\,\mathfrak{g}_0[k] + \sqrt{n_2}\,\mathfrak{g}_2[k]\end{aligned}$$

* Where $S[k]$ and both $N_i[k]$ are all independent Gaussian noise streams.
* $\mathfrak{g}_i$ are unit norm zero mean Gaussian streams.
* For convenience, s and n_i are dimensioned as powers.

Correlation of signals with noise (at zero delay)

* Make use of bilinearity and previous relations:

$$\begin{aligned} C_{ij}[0] &= \langle SS \rangle + \langle N_1 S \rangle + \langle S N_2 \rangle + \langle N_1 N_2 \rangle \\ &= \frac{1}{N} \sum_{l=1}^{N} S[l]^2 \\ &= s \end{aligned}$$

* And normalized correlation coefficient:

$$\Gamma_{ij}[0] = \frac{s}{\sqrt{s+n_1}\sqrt{s+n_2}}$$

* In the low signal to noise limit ($s \ll \min n_1, n_2$)

$$\Gamma_{ij}[0] = \frac{s}{\sqrt{n_1 n_2}}$$

Correlation of signals with noise (at zero delay)

* Noise does not enter the expectation value of C_{ij}, but it does the uncertainty:

$$\sigma_{C_{ij}[0]} = \sqrt{\frac{2s^2 + n_1 s + s n_2 + n_1 n_2}{N}}$$

* Some messy statistics used, left as exercise to the astute reader!
* In the low signal to noise limit

$$C_{ij}[0] = s \pm \sqrt{\frac{n_1 n_2}{N}}$$

$$\Gamma_{ij}[0] = \frac{s}{\sqrt{n_1 n_2}} \pm \frac{1}{\sqrt{N}}$$

* Exercise to reader: consider the strong signal case.

Interferometer sensitivity

* Previous page allows one to write down the SNR for a measurement:

$$\mathrm{SNR} = \frac{s\sqrt{N}}{\sqrt{n_1 n_2}} = \frac{s\sqrt{N}}{\sqrt{\mathrm{SEFD}_1 \mathrm{SEFD}_2}}$$

* Usually instead the sensitivity of the baseline is expressed:

$$\Delta S = \frac{\sqrt{\mathrm{SEFD}_1 \mathrm{SEFD}_2}}{\sqrt{N}} = \frac{\sqrt{\mathrm{SEFD}_1 \mathrm{SEFD}_2}}{\sqrt{2\Delta\nu T}}$$

* SEFD is the *System Equivalent Flux Density*
  - SEFD $= T_{\mathrm{sys}}/g$ where g is antenna gain (units of K/Jy)
  - Equals the brightness (in Jy) of a source required to double antenna noise power (T_{sys})
  - VLBA antenna SEFD is typically 300 to 500 Jy.
* Additional efficiency factors may apply (e.g., quantization)

Finite energy signals

* Some signals are zero outside a finite time range
  ◦ Or diminish sufficiently fast such that $\lim_{T\to\infty}\int_{-T}^{T} v(t)^2 dt = C$
* Time averages of cross- and auto-correlations $\to 0$ as $T \to \infty$
* In such cases one can take the limit as follows:

$$\begin{aligned}
\Gamma_{ij}(\tau) &= \lim_{T\to\infty} \frac{\frac{1}{2T}\int_{-T}^{T} v_1(t)v_2(t+\tau)dt}{\sqrt{\frac{1}{2T}\int_{-T}^{T} v_1(t)^2 dt}\sqrt{\frac{1}{2T}\int_{-T}^{T} v_2(t+\tau)^2 dt}} \\
&= \lim_{T\to\infty} \frac{\int_{-T}^{T} v_1(t)v_2(t+\tau)dt}{\sqrt{\int_{-T}^{T} v_1(t)^2 dt}\sqrt{\int_{-T}^{T} v_2(t+\tau)^2 dt}}
\end{aligned}$$