Synthesis Imaging in Radio Astronomy II
ASP Conference Series, Vol. 180, 1999
G. B. Taylor, C. L. Carilli, and R. A. Perley (eds.)

4. Cross Correlators

J. D. Romney
National Radio Astronomy Observatory, Socorro, NM 87801, U.S.A.

Abstract. This lecture presents the general concepts of correlation and spatial coherence functions, as they apply to correlator implementations for synthesis-imaging arrays. A uniform treatment, based in the spectral domain, is applied to the both the conventional lag and the spectral or 'FX' correlator architectures. Parallel presentations of both architectures are given, including their strengths and weaknesses, although somewhat more emphasis is placed on the FX correlator because it is less familiar.

1. Introduction

The correlator of a synthesis-imaging array is the subsystem in which the interferometers are actually formed. Although a conceptual interference pattern may be considered to exist just from the act of receiving a common wavefront at the antennas of the array, it is in the correlator that interference fringes are formed and the complex visibility or spatial coherence function — introduced in the first lecture as the fundamental observable of synthesis imaging — is measured. Some have referred to the correlator as the "lens" of a synthesis array for this reason.

This lecture will address the concepts and theory of correlation in general. The theoretical treatment, based more in the spectral domain than has been customary, is designed to lead naturally to both the conventional lag correlator, and the spectral-domain or 'FX' architecture. Both connected-element synthesis arrays like the VLA, and VLBI arrays like the VLBA, are treated uniformly.

Following the spectral-domain basis of the theoretical treatment, and departing from previous treatments, I also will assume throughout that the desired output of the correlator is spectroscopic — i.e., a set of visibility measurements on a closely-spaced grid in radio-frequency space. This is clearly the requisite measurement for investigations of spectral emission or absorption lines; subsequent lectures will show that it is also the most appropriate form for constructing images of continuum sources, or for making high-precision astrometric or geodetic measurements.

2. Interfaces

The fundamental inputs to the correlator are digital samples from each antenna of the observing array. Lectures such as this sometimes cover details of the sampling process, but since the sampling occurs in other parts of the system, these details seem mainly to distract from consideration of the correlator itself, and in any case cannot be accommodated in the time available. Thus, I will take the position that the correlator's job is to measure the *sampled* signal.

The samples are obtained either in real time in connected-element arrays, or reproduced some time later from data tapes recorded in VLBI arrays. In the latter case, the correlator reproduces and resynchronizes the recorded samples, and thus recreates in one room the situation that existed across the world as the

observations occurred. The delay between observation and correlation currently amounts to about two weeks for the VLBA correlator, although the range of turn-around times is limited in practice only by the performance of package-delivery services at the minimum, and the lifetime of a particular VLBI data recording system at the maximum.

The correlator's basic output is a set of measurements of the complex visibility function, spanning a total of four dimensions: the spatial-frequency vector (u, v, w) and the observing frequency ν. In the conventional ordering imposed by the correlator implementation, these results consist of spectra (over ν), for each member of the set of baselines, at a sequence of observing times. The baselines comprise all pairwise combinations of the observing stations, including "zero-baseline", single-dish measurements; observing time advances at intervals on the order of a second. Each baseline and time corresponds to a different (u, v, w) vector.

3. Background

At this point it's necessary to introduce some fundamental concepts and the formalism to be used throughout the lecture. The following subsections will have to summarize, in a very superficial fashion, some areas which are the subject of entire graduate-level courses, but I have given several references for further reading. And since I will be using mathematics principally as an heuristic tool, I will not pursue rigorous derivations.

3.1. Correlation and Coherence

The subjects of correlation and coherence can be approached from the point of view of optics (for which Born & Wolf 1980 is the definitive general reference), of communications engineering (where Blackman & Tukey 1958) is the classic work, although concerned only with what radio astronomers would call the "zero baseline" case), or of stochastic processes (see, *e.g.,* Parzen 1962). The terminology common in the latter field is closer to that of radio astronomy, and will be used in the following treatment.

We start with a quasi-monochromatic electromagnetic plane wave with an electric field component given by the real part of

$$E(t) = A(t)e^{+i2\pi\nu_0 t} \tag{4–1}$$

where $A(t)$ is a complex, band-limited, covariance-stationary, ergodic stochastic process. $A(t)$ can be considered to modulate a carrier at frequency ν_0. The wave propagates *opposite* to the unit vector pointing at the source, i.e., along $\hat{\mathbf{k}} = -\hat{\mathbf{s}}$.

Although $E(t)$ is an "analytic signal" as used in the usual complex-representation formalism, I will focus on $A(t)$, which is complex. Although the analytic-signal formalism thus will not be particularly useful, I will nevertheless work with complex functions because they permit a clearer presentation. Generally, physical signal processes will be the real parts of these functions, and I will point out the few cases where this somewhat careless usage requires correction.

We will be interested principally in the spectrum of $A(t)$, and so we introduce at this point its *spectral representation*

$$A(t) = \int_{-\infty}^{+\infty} s(\nu) e^{+i2\pi\nu t}\, d\nu. \tag{4–2}$$

It must be cautioned that $s(\nu)$ is strictly a mathematical device, and in fact has no physical meaning and is not measurable. Mathematicians are reluctant even to give it a name. Only when we reach Eqs. (4–6)–(4–7) below do we obtain what is usually thought of as a "spectrum". Nevertheless, we can still assert that $s(\nu)$ is zero outside a limited band $|\nu| \le B$.

The interferometer measures what is called variously the cross-covariance, cross-correlation, or visibility function, defined as

$$\Gamma_{ij} = \langle E_i(t+\tau_{ij}) E_j^*(t) \rangle, \tag{4–3}$$

where $\tau_{ij} = (\mathbf{r}_i - \mathbf{r}_j) \cdot \hat{\mathbf{k}}/c$ is the light-travel time between observation of the same plane wave at the two spatial positions labeled i and j. The superscript * indicates complex conjugation.

In terms of the band-limited quasi-monochromatic signal $A(t)$,

$$\Gamma_{ij} = \langle A(t+\tau_{ij}) A^*(t) \rangle e^{+i2\pi\nu_0\tau_{ij}} = \gamma_{ij}(\tau) e^{+i2\pi\nu_0\tau_{ij}}, \tag{4–4}$$

where

$$\gamma_{ij}(\tau) = \langle A(t+\tau_{ij}) A^*(t) \rangle \tag{4–5}$$

is the covariance function of the stochastic process $A(t)$. The angle brackets in each of Eqs. (4–3)–(4–5) denote an expectation or statistical average. Since $A(t)$ has been declared to be ergodic, these can be approximated adequately by averages with respect to t.

Since τ_{ij} is realized by receiving the wave at different locations on the Earth, it will change as diurnal rotation changes the angle between the baseline vector and source unit vector; this gives rise to the "fringe" signal of interferometry which we must measure. Here $\gamma_{ij}(\tau)$ represents the complex amplitude and phase which modulate the fringe signal given by the phasor in τ_{ij}.

Using Eq. (4–2) we can develop a spectral representation for $\gamma_{ij}(\tau)$, as

$$\begin{aligned} \gamma_{ij}(\tau) &= \left\langle \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} s_i(\nu) s_j^*(\nu') e^{+i2\pi[(\nu-\nu')t + \nu\tau_{ij}]}\, d\nu\, d\nu' \right\rangle \\ &= \int_{-\infty}^{+\infty} S_{ij}(\nu) e^{+i2\pi\nu\tau_{ij}}\, d\nu, \end{aligned} \tag{4–6}$$

where

$$S_{ij}(\nu) = s_i(\nu) s_j^*(\nu) = \int_{-\infty}^{+\infty} \gamma_{ij}(\tau) e^{-i2\pi\nu\tau}\, d\tau \tag{4–7}$$

can now properly be called the "cross-power spectrum" of the process $A(t)$. A great deal of elegant mathematics has been devised to establish Eqs. (4–6)–(4–7) on a rigorous basis, even though the integrals in the Fourier transform relationship formally do not exist.

3.2. Observations

In this section I'll outline the major steps that occur at each observing station to produce the samples processed by the correlator. The quasi-monochromatic plane wave described by Equation (1) is received at station i, at position $\mathbf{R}_i(t)$, with a delay

$$\tau_i = \mathbf{R}_i(t + \tau_i) \cdot \hat{\mathbf{k}}/c \tag{4–8}$$

relative to the time the wave passes the coordinate origin. Note that τ_i must be defined recursively, in terms of the position of the station at an earlier time when the wave was received there. The coordinate origin relative to which $\mathbf{R}_i$ is measured is arbitrary, but is usually taken as some point near the center of a connected-element array, or the center of the Earth for a VLBI array. The time dependence of τ_i is not shown explicitly.

The received field

$$E_i(t) = E(t + \tau_i) = A(t + \tau_i)e^{+i2\pi\nu_0(t+\tau_i)} \tag{4–9}$$

is converted to baseband by a single-sideband mixer with LO frequency ν_0:

$$V_i(t) = E_i(t)e^{-i2\pi\nu_0 t} = A(t + \tau_i)e^{+i2\pi\nu_0\tau_i}. \tag{4–10}$$

This equation is not physically correct as written, since the LO signal is actually real, and the sum-of-frequencies component is removed by bandpass filtering. Nevertheless, the result is valid. A complete treatment also must account for frequency and phase offsets among the LOs at different stations, which I must gloss over here.

The baseband signal is sampled at an interval Δt, through a non-linear function $f[\cdot]$:

$$R_i(n) = f[V_i(t = n\Delta t)]. \tag{4–11}$$

Sampling will impose both a loss of sensitivity and a modification of the original spectrum. Techniques for correcting both effects in post-correlation analysis are well known. As I remarked at the outset, however, I consider here that the correlator's job is only to measure the sampled signal.

All subsequent manipulations of the signal operate on discrete, digital samples. However, most people find continuous functions and integrals more intuitive than series and summations, and I will proceed using the continuous forms whenever possible, which in fact will turn out to cover most of the material of this lecture. When necessary, I will resort to the discrete notation.

In a connected-element synthesis instrument, the digital samples can then be routed directly to the correlator. In a VLBI array, they are recorded on magnetic tape, in a format which includes embedded time tags to allow the time of each individual sample to be re-established. Offsets among the atomic time standards at the various stations (which can be kept reasonably small with modern technology) would have to be included in a more complete treatment.

4. The Correlator Frontend

At the input to a VLBI correlator system, the recorded samples are reproduced from the tapes, and the sample times re-established. It's sometimes possible

to play back the tapes at a speed several times that used in recording, so that correlation can proceed faster than real time.

I have assigned to the correlator's "frontend" two other, essential functions, which compensate for the effects of the array geometry on the signals received. While it's not necessary (nor, is some cases, possible) to perform this compensation at the frontend of the correlator, it is convenient to introduce them conceptually at this point, because they follow naturally from the preceding section. At this point it's appropriate to discuss these functions in a completely station-based manner, although often they are thought of in a baseline sense. Subsequent sections will go into greater detail on their application in the two alternative correlator architectures, and will consider the baseline-based approach where necessary.

Implementation of these functions is an area where significant differences arise among specific correlator designs. Although a hierarchy of linear and/or polynomial approximations is almost always used, the number of coefficients and their update intervals vary widely.

4.1. Delay Compensation

Under the simplifying assumptions adopted earlier, compensation for the interferometer delay is extremely straightforward. Each reproduced signal $V_i(t)$ is delayed by δ_i:

$$P_i(t) = V_i(t - \delta_i) = A(t + \tau_i - \delta_i)e^{+i2\pi\nu_0\tau_i}. \qquad (4\text{–}12)$$

(Remember, the actual signal is a set of discrete samples of $V_i(t)$, $P_i(t)$, *etc.*) Ideally, one would choose

$$\delta_i = \tau_i \qquad (4\text{–}13)$$

to obtain the "white light fringe", as will be shown below. Departures from this ideal in different correlator architectures are discussed in subsequent sections.

The regular sampling implicit in Eq. (4–11) limits delay tracking at the correlator input to integral units of the sample interval. Techniques which can implement or approximate fractional-sample delay tracking will be discussed in subsequent sections. Connected-element interferometer systems, like the VLA, may implement a continuously-varying delay tracking by applying a timing offset to the sample clock, eliminating or minimizing the requirement for delay tracking in the correlator. This is feasible in a connected-element system where the correlator's operation can be embedded in that of the overall interferometer, but is substantially more difficult, although not impossible, in a VLBI correlator. Such a scheme was considered in the design of the VLBA, but was rejected as too difficult, and as an impediment to global VLBI compatibility.

In a VLBI instrument, it is possible to implement arbitrarily large delays simply by offsetting the tapes during playback. In practice, a digital buffer is necessary to accommodate minor variations in tape speed, so that at least a part of the delay can be, and inevitably is, tracked by manipulating the buffer pointers.

4.2. Phase Compensation

A phase rotation must also be applied to the signal, in order to stop the fringe signal as described earlier. We denote the phase compensation function as $\theta_i(t)$.

This phase rotation is almost always applied as a *complex* mixer, for reasons explained later. The *real* output of this rotator is then

$$X_i(t) = P_i(t)e^{-i\theta_i(t)} = A(t + \tau_i - \delta_i)e^{+i2\pi\nu_0\tau_i - i\theta_i(t)}. \qquad (4\text{–}14)$$

while the imaginary part is phase-shifted by $\pi/2$. Again the ideal phase rotation will be shown later to be correct:

$$\theta_i = 2\pi\nu_0\tau_i. \qquad (4\text{–}15)$$

Again, connected-element interferometer systems may eliminate the need for phase compensation in the correlator, by applying small offsets to the local oscillator frequencies in the baseband conversion step, similarly to the offset sampling described in the preceding subsection.

5. The Lag Correlator

With this background, we can consider the conventional or lag correlator architecture quite straightforwardly. Generalize Eq. (4–13) to provide a range of "lags", additional delay increments whose spacing, for the observing bandwidth B limiting the quasi-monochromatic process $A(t)$, is given by $\Delta\tau = 1/2B$:

$$\delta_i(n) = \tau_i + n\Delta\tau \qquad n \in [-\tfrac{N}{2}, \tfrac{N}{2}-1], \qquad (4\text{–}16)$$

and, using Eqs. (4–16) and (4–15) in Eq. (4–14), define a set of lagged signals

$$L_i(t,n) = A(t - n\Delta\tau). \qquad (4\text{–}17)$$

Then the correlation obtained by the lag correlator is

$$\begin{aligned} C_{ij}(n) &= \langle L_i(t,0)L_j^*(t,n)\rangle = \langle A(t)A^*(t - n\Delta\tau)\rangle \\ &= \langle A(t + n\Delta\tau)A^*(t)\rangle = \gamma_{ij}(n\Delta\tau). \end{aligned} \qquad (4\text{–}18)$$

Comparison with Eq. (4–5) shows that we have thus obtained the complex amplitude and phase which modulate the fringe signal in Eq. (4–4), and confirms that Eqs. (4–13) and (4–15) are correct. The N values of the cross-correlation function $C_{ij}(n)$ are now stationary, and can be integrated as long as desired, limited only by the range of residual fringe frequencies to be preserved.

Finally, to obtain the desired cross-power spectrum, it is then only necessary to apply the Fourier transform of Eq. (4–7). (This is a bit too glib; Eq. (4–7) implies $\gamma_{ij}(\tau)$ is sampled over an infinite range of τ, when in fact only N lags are obtained; the effect of this truncation is explored below.)

A M-station lag correlator, of which the VLA correlator is a large-scale example with $M = 27$, consists of $M(M-1)/2$ baseline units, each performing the complex multiplications and accumulations implied by Eq. (4–18). With only one primary unit type, arranged to form all possible station pairs, an overall block diagram is hardly necessary. A detailed view of the processing within each such unit appears in Fig. (4–1). The boxes labeled δ_i and θ_j represent the frontend delay and phase compensation, each applied to only one of the

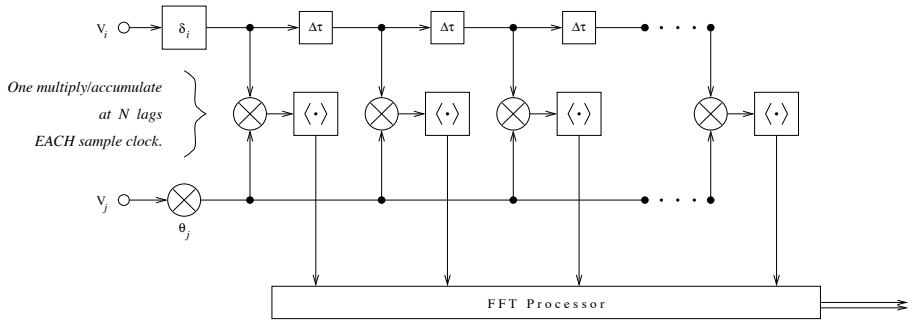

Figure 4–1. Lag correlator baseline processing.

station input streams v_i and v_j for reasons explained below. The individual one-lag delays marked $\delta\tau$ generate the lagged signals I have denoted as $L_j(t,n)$ in Eq. (4–17). These are multiplied by an unlagged signal corresponding to $L_i(t,0)$ in the mixer symbols, and the boxes denoted $\langle\cdot\rangle$ represent the statistical average to complete the implementation of Eq. (4–18). The final box is the Fourier transform corresponding to Eq. (4–7), although this is actually done off-line in many lag systems, including the VLA correlator. As indicated in the note at the left, each lag cell must be multiplied and accumulated on each sample clock, which will turn out to be a major contrast to the FX architecture.

5.1. Lag Correlator: Fractional-Sample Delay

I pointed out previously that delay compensation at the correlator input necessarily is restricted to integral samples. We are now in a position to see the effects of this, and to consider some solutions. Instead of Eq. (4–13), write

$$\delta_i = \tau_i + \varepsilon_i \tag{4–19}$$

where ε_i represents a fractional-sample error due to truncation or rounding of δ_i. Then in Eq. (4–17) we will have

$$L_i(t,n) = A(t - n\Delta\tau - \varepsilon_i), \tag{4–20}$$

and in Eq. (4–18),

$$C_{ij}(n) = \gamma_{ij}(n\Delta\tau - (\varepsilon_i - \varepsilon_j)). \tag{4–21}$$

On applying Eq. (4–7), the result of this lag shift will be to change $S(\nu)$ to

$$S'(\nu) = e^{-i2\pi\nu(\varepsilon_i - \varepsilon_j)} S(\nu). \tag{4–22}$$

Thus any fractional-sample error in delay compensation induces a *phase slope* across the observed band. Expressing the error in units of samples, f,

$$\varepsilon_i - \varepsilon_j = \frac{f}{2B}, \tag{4–23}$$

where B is the observing bandwidth,

$$S'(\nu) = e^{-i\pi f \frac{\nu}{B}} S(\nu). \tag{4–24}$$

To simplify the further analysis of this effect, we will assume, as is often the case, that the required delay compensation drifts through many sample intervals during the correlator's on-line integration time. Thus we regard f as uniformly distributed between some symmetric bounds, $f \in [-f_0, +f_0]$, and compute the expectation of the degraded correlator output spectrum:

$$S'_{f_0}(\nu) = \frac{1}{2f_0}\int_{-f_0}^{+f_0} S'(\nu)\,df = \frac{S(\nu)}{2f_0}\int_{-f_0}^{+f_0} e^{-i\pi f\frac{\nu}{B}}\,df = S(\nu)\cdot\frac{\sin\pi f_0\frac{\nu}{B}}{\pi f_0\frac{\nu}{B}} \quad (4\text{–}25)$$

The degradation of a continuum signal, $S(\nu) = S_0$, is obtained by integrating again, this time with respect to ν:

$$S'_0 = \frac{1}{B}\int_0^B S'_{f_0}(\nu)\,d\nu = \frac{S_0}{B}\int_0^B \frac{\sin\pi f_0\frac{\nu}{B}}{\pi f_0\frac{\nu}{B}}\,d\nu = S_0\cdot\frac{\mathrm{Si}(\pi f_0)}{\pi f_0} \quad (4\text{–}26)$$

where $\mathrm{Si}(\cdot)$ is the "sine integral" function.

We consider now some specific practical cases of f_0. Even if the error for each station is kept within $f \in [-\frac{1}{2}, +\frac{1}{2}]$ by careful rounding, independent station-based delay compensation can allow combined errors extending to $f \in [-1, +1]$ on a baseline basis. The effect of this is seen in Fig. (4–2a), where the two heavy lines show the phase shift, according to Eq. (4–24), across the observing band for the extreme values $f = \pm 1$, and the hatched area between them is the entire region covered for this case of $f_0 = 1$. From Eq. (4–25), we see that a spectral component at the upper band edge, $\nu = B$, is completely washed out, while the amplitude of one at the band center is reduced by $\sin(\frac{\pi}{2})/\frac{\pi}{2} = 0.637$. Across the entire band, Eq. (4–26) shows the amplitude of a continuum signal is reduced by $\mathrm{Si}(\pi)/\pi = 0.589$. These are all quite unsatisfactory.

Thus, lag correlators always require a fundamentally baseline-oriented delay compensation, where the rounding can ensure that $f \in [-\frac{1}{2}, +\frac{1}{2}]$. One common way to accomplish this is to use station-based delays with an extra, baseline-dependent *vernier delay* of $(0, \pm 1)$ bit. However implemented, this scheme achieves $f_0 = \frac{1}{2}$, as illustrated in Fig. (4–2b), with loss factors of 0.637 at the upper band-edge, 0.900 at the center, and 0.873 across a continuum band.

A further improvement requires coordination of the delay and phase compensation, inserting an extra phase shift of $\frac{\pi}{2}$ into the phase rotator whenever a one-sample delay shift occurs. Instead of Eq. (4–15) we have

$$\theta_i = 2\pi\nu_0\tau_i + \frac{\pi}{2}\frac{\delta_i}{1/2B} = 2\pi(\nu_0\tau_i + \frac{B}{2}\delta_i), \quad (4\text{–}27)$$

and Eqs. (4–20)–(4–22) become

$$L_i(t,n) = A(t - n\Delta\tau - \varepsilon_i)e^{+i\pi B\delta_i}, \quad (4\text{–}28)$$

$$C_{ij}(n) = \gamma_{ij}(n\Delta\tau - (\varepsilon_i - \varepsilon_j))e^{+i\pi B(\delta_i - \delta_j)}, \quad (4\text{–}29)$$

$$S''(\nu) = e^{-i2\pi[\nu(\varepsilon_i-\varepsilon_j)+\frac{B}{2}(\delta_i-\delta_j)]}S(\nu) = e^{-i2\pi[(\nu-\frac{B}{2})(\varepsilon_i-\varepsilon_j)+\frac{B}{2}(\tau_i-\tau_j)]}S(\nu). \quad (4\text{–}30)$$

In Eq. (4–27), since τ_i and δ_i are nearly equal, we see that the effect of the extra phase steps is to shift the phase tracking from the edge to the center of the

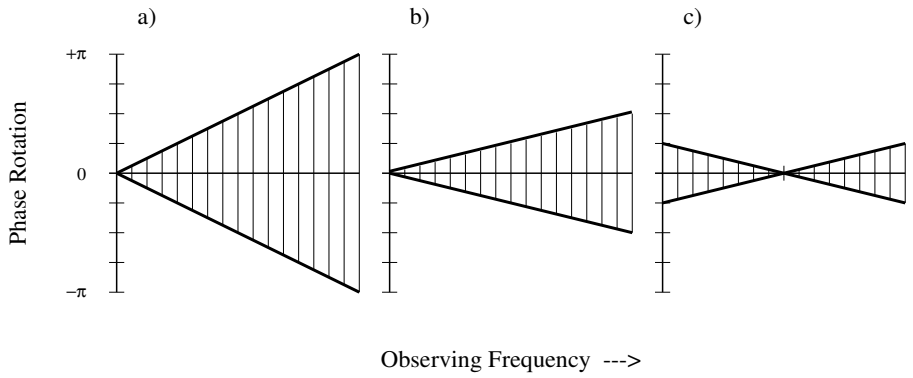

Figure 4–2. Phase shifts induced across the observing band for various configurations of delay and phase compensation.

observed band. This is reflected in the second term in the phasor of Eq. (4–30); in the first term, we obtain the desired change where the phase slope now pivots about $\nu = B/2$.

For this configuration, Eqs. (4–25) and (4–26) can be modified straightforwardly to

$$S''_{f_0}(\nu) = S(\nu) \cdot \frac{\sin \pi f_0 (\frac{\nu}{B} - \frac{1}{2})}{\pi f_0 (\frac{\nu}{B} - \frac{1}{2})} \tag{4–31}$$

and

$$S''_0 = S_0 \cdot \frac{\mathrm{Si}(\pi f_0/2)}{\pi f_0/2} \tag{4–32}$$

Assuming the baseline-based delay tracking introduced above, with $f_0 = \frac{1}{2}$, the phase shift variation across the band is as shown in Fig. (4–2c). Spectral components at the band center are unaffected, while those at *both* band edges are reduced by $\sin(\frac{\pi}{4})/\frac{\pi}{4} = 0.900$. The amplitude of a continuum signal is reduced by $\mathrm{Si}(\frac{\pi}{4})/\frac{\pi}{4} = 0.966$.

No matter which of these variants is used, it is necessary to apply some sort of correction in post-correlation software for the effects of integral-sample delay compensation in the lag correlator. Also, those cases where the assumption of rapidly changing delay fails can and must be corrected after correlation. These aspects of calibration are beyond the scope of this lecture.

5.2. Lag Correlator: Phase Rotator Implementation

A lag correlator must form a large number of lag products, since this must be done for every baseline, and so the efficiency of each lag is a critical design factor. This has limited lag correlator designs to at most four-level by four-level products until very recently. Thus with two- or four-level signal inputs, only primitive phase-rotation functions can be used. One very popular choice has been the three-level rotator (Clark, Weimer, & Weinreb 1972), shown in Fig. (4–3) along with the sinusoidal function which is being approximated.

Since the function is nearly a square wave, it is not surprising that a significant fraction of its power appears at frequencies other than the desired fun-

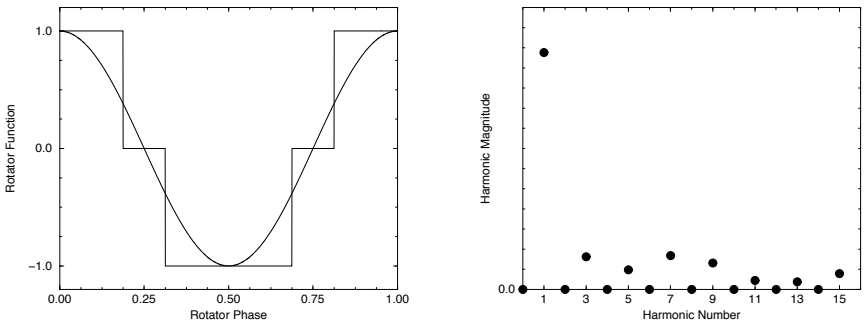

Figure 4–3. The three-level fringe rotator function, and its spectrum.

damental. Harmonic components will shift the input signals by the wrong frequency; these shifted components will not be stationary during an integration and this fraction of the input signal power will be lost. Consequently, lag correlators using this rotator function reduce the amplitude of the output fringe signal by a factor 0.96, which must be corrected in the calibration process.

Use of such a fringe rotation function has a further consequence for lag correlators. It is not possible to do the straightforward station-based fringe rotation implied by Eq. (4–15), for chance coincidences of different harmonics of two different fringe rotation functions will produce unexpected spurious correlations. Thus, as was true of the delay, phase rotation must be implemented on a baseline basis in a lag correlator.

5.3. Lag Correlator: Spectral Response

The limited range of lags, N in Eq. (4–16), has an important consequence in the profile of spectral lines observed with the synthesis instrument. The effect is fairly obvious, but should be developed here for comparison with the FX correlator architecture, where the corresponding situation is less well known.

We write $N\Delta\tau = T$, and represent the truncation of $\gamma_{ij}(\tau)$ as

$$\gamma'_{ij}(\tau) = \gamma_{ij}(\tau) \cdot \sqcap(\frac{\tau}{T}). \tag{4–33}$$

The "unit rectangle function" $\sqcap(x)$, an extremely useful device introduced in Bracewell's (1965) classic work on Fourier transform theory, is unity for $x \in [-\frac{1}{2}, \frac{1}{2}]$ and zero elsewhere. Using $\gamma'_{ij}(\tau)$ in Eq. (4–7), we can then apply the Convolution Theorem of the Fourier transform, and write immediately

$$S'(\nu) = S(\nu) * T\ \mathrm{sinc}(T\nu), \tag{4–34}$$

where the symbol $*$ indicates the convolution operation and the 'sinc' function $\mathrm{sinc}(s) = \sin(\pi s)/(\pi s)$, shown in Fig. (4–4), is the Fourier transform of $\sqcap(x)$.

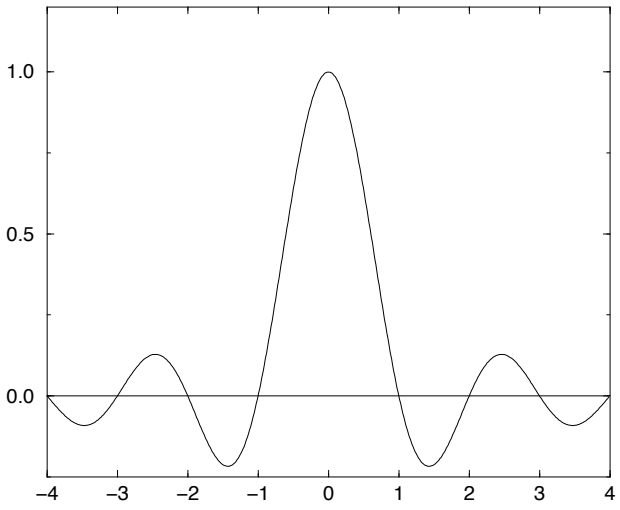

Figure 4–4. The sinc(·) function.

6. The FX Correlator

At the simple level adopted at the beginning of the preceding section for the lag correlator, the derivation of the spectral-domain or 'FX' correlator is almost too trivial to write down. If only as a framework for the description of the hardware stages, however, I will do so anyway. Starting again with the compensated input signal given by Eq. (4–14), for delay and phase compensation specified by Eqs. (4–13) and (4–15), we have directly

$$X_i(t) = A(t). \tag{4–35}$$

The FX correlator immediately Fourier transforms this signal — the 'F' of FX — to give

$$s_i(\nu) = \int_{-\infty}^{+\infty} X_i(t) e^{-i2\pi\nu t}\, dt \tag{4–36}$$

The FFT algorithm makes the entire FX concept practical, but is not a fundamental requirement. Next, these "station spectra" are *cross-multiplied* — the 'X' — yielding

$$S_{ij}(\nu) = s_i(\nu) s_j^*(\nu) \tag{4–37}$$

The cross-power spectrum in Eq. (4–37) can then be integrated. Again, the integration time is limited only by the desired range of residual fringe frequencies.

The name 'FX' was originated by Chikada et al. (1987), who also built the first such correlator, to indicate this reversal of the order of operations compared to the conventional lag correlator. The group at NRAO that designed the VLBA correlator considered one design of each type, and adopted the complementary term 'XF' which is sometimes used for the latter architecture.

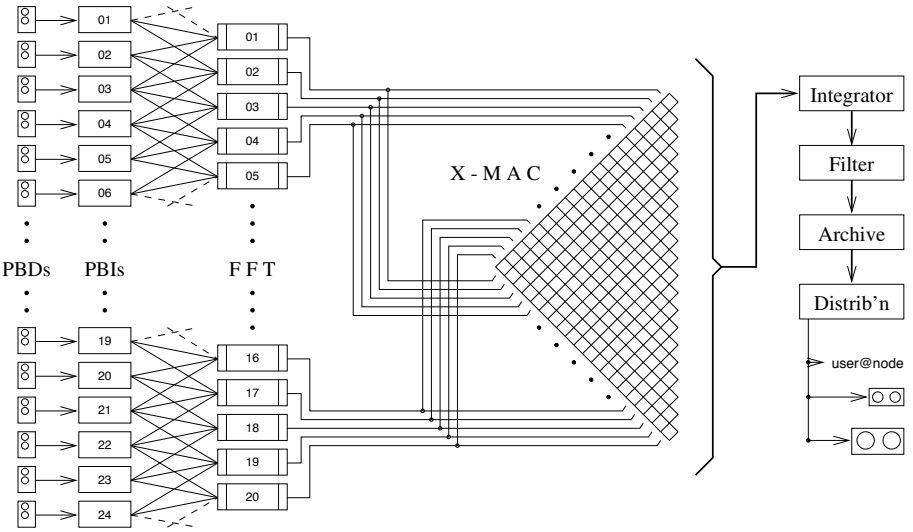

Figure 4–5. Overview of the VLBA Correlator, as representative of the FX architecture.

The VLBA correlator is presented here as representative of the FX architecture. Since this scheme is less familiar, and involves two basic unit types — rather than only one as in the lag correlator — an overview is shown in Fig. (4–5). The 24 VLBI tape playback drives (PBDs) and the 20 "playback interfaces" (PBIs) are only of interest here as implementing the integral-sample delay compensation. Phase compensation is performed in the unlabeled small boxes at the inputs to the 20 FFT processors, which themselves perform the 'F' operation of the correlator. The small boxes at the FFT outputs will be discussed in the next subsection. In the cross-multiplier/accumulator (X-MAC) a single square corresponds to the 'X' operation of the one-baseline correlator described here, and represents both Eq. (4–37) and a subsequent partial integration, which in turn is extended to longer times in the Integrator. Other components of the figure are not germane to the present discussion.

A detailed view of an FX correlator's baseline processing appears in Fig. (4–6). Here the frontend delay (δ) and phase (θ) compensation are shown applied separately to each station's input stream (v), since — as will be discussed below — this is feasible in the FX architecture. The "FFT processors" load segments of N points in a time series, each of which is Fourier transformed to produce "station spectra" which are then cross-multiplied, point by point, in the mixer symbol and accumulated. The accumulators yield the desired spectral-domain output directly after each integration cycle. As indicated in the note at the left, each spectral point must be cross-multiplied and accumulated only once for each N sample clocks — a factor of N lower rate than in the lag architecture.

Since FX correlators remain fairly unusual, I thought it would be useful to mention some, other than the VLBA correlator, that have been built. Following the original system built by Chikada et al. (1987) at Nobeyama Radio Observatory, Japanese scientists have built several other FX correlators, at Nobeyama and at the National Astronomical Observatory in Mitaka. Most notable among

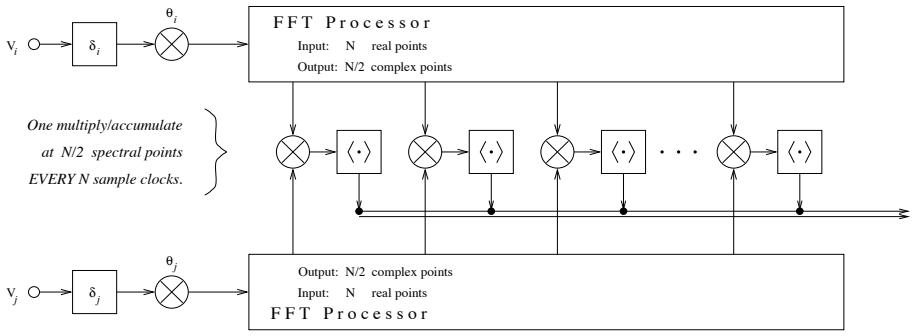

Figure 4–6. FX correlator baseline processing.

these, for the interests of local NRAO personnel, is the correlator built as part of the VSOP Space VLBI mission (Horiuchi et al. 1998). In the United States, a single-dish "Spectral Processor" was completed in 1990 at the NRAO Green Bank site. Some particular applications of that system are mentioned in subsequent sections where the relevant FX capability is discussed.

6.1. FX Correlator: Fractional-Sample Delay

The effect of integral-sample delay compensation at the correlator's front end is fundamentally the same for FX and lag correlators, but significantly different in the practical application. Again we use Eq. (4–19) to represent the integral-sample compensation with fractional-sample errors, and replace Eq. (4–35) with

$$X_i(t) = A(t - \varepsilon_i). \tag{4–38}$$

And similarly to the lag case, the FX correlator output is seen to be modified to

$$S'_{ij}(\nu) = e^{-i2\pi\nu(\varepsilon_i - \varepsilon_j)} S(\nu). \tag{4–39}$$

Although the same phase slope arises in both correlator architectures, in the FX scheme we can correct the slope at the Fourier transform output, before cross-multiplication and before integration. This operation is indicated in Fig. (4–5) by the small unlabeled boxes at the FFT outputs. Applying the correction at this point is equivalent to interpolating the fractional-sample delay at the correlator input. It allows the FX correlator to operate in all cases with *no fractional-sample loss at all.*

6.2. FX Correlator: Phase Rotator Implementation

Although the first stage of an FX correlator's Fourier transform could be optimized to accept the low precision (typically one or two bits) usually used in sampling, practical designs make all stages as similar as possible. Thus, the input typically can be presented at the same moderate precision as is needed at the Fourier transform's output. This allows the phase rotation to be done at that point using a multi-level rotator function, at essentially no extra cost.

Signals passed through such a rotator lose no significant power to harmonics. Thus, beyond just retaining the fringe power lost in a three-level rotator, it

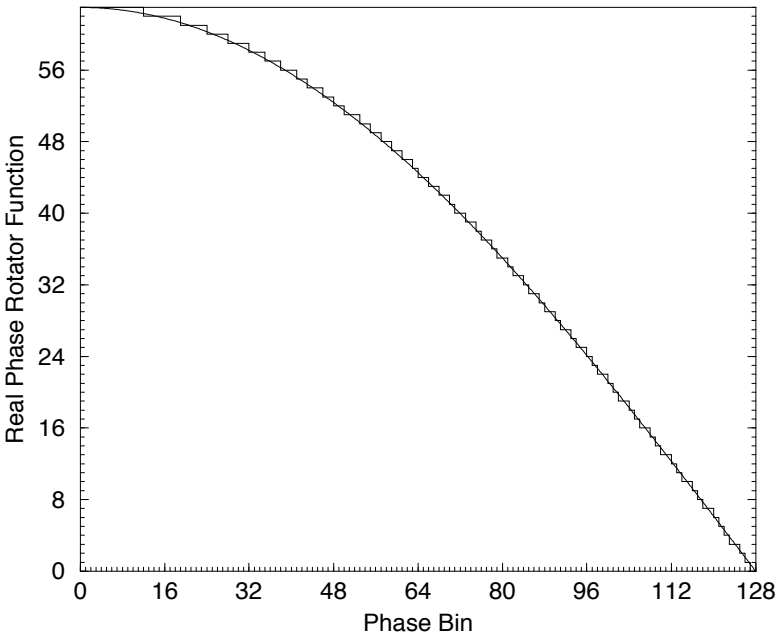

Figure 4–7. The VLBA correlator's multi-level phase rotator function.

becomes practical to implement station-based phase rotation without spurious correlations — and without the cost of a large number of baseline-based rotation elements.

Fig. (4–7) shows the multi-level rotator function used in the VLBA correlator — or actually just the real part of the first quadrant; all the rest is determined by symmetry. This function uses a 9-bit input phase word (spanning all four quadrants), and provides a 7-bit signed output value. This approximates the desired sinusoidal function so closely that only 5.3×10^{-5} of the input power goes into harmonics; no individual harmonic is more than about 25 dB below the fundamental.

6.3. FX Correlator: Spectral Response

Reversing the order of F and X operations has an interesting effect on the spectral profile. A practical FX correlator transforms a limited number, N, of samples in its F component — just as the lag correlator can form only a limited number of lags.

Proceeding as for the lag correlator, we write $N\Delta t = T$ (which is now an extent in time, rather than lag), and in place of Eq. (4–33),

$$X_i'(t) = X_i(t) \cdot \sqcap(\frac{t}{T}) \qquad (4\text{–}40)$$

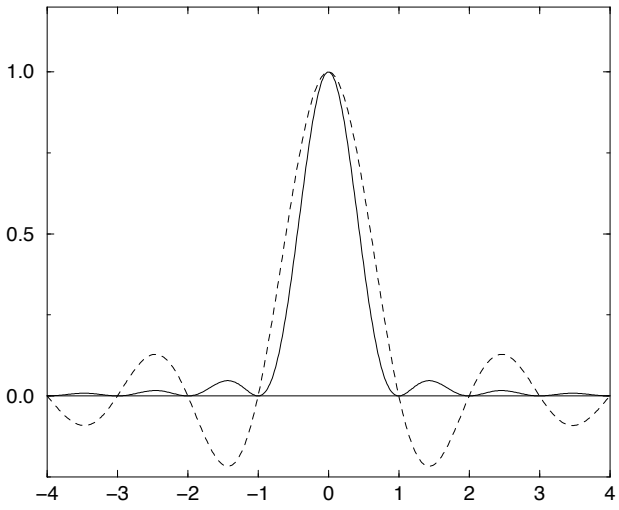

Figure 4–8. The $\mathrm{sinc}^2(\cdot)$ function, with $\mathrm{sinc}(\cdot)$ [dashed] for comparison.

The corresponding convolution in the spectral domain,

$$s'_i(\nu) = s_i(\nu) * T\ \mathrm{sinc}(T\nu), \tag{4–41}$$

is applied to *each* of the station spectra, *before* cross-multiplication. Consequently, the cross-power spectrum is convolved twice by the $\mathrm{sinc}(\cdot)$ function:

$$S'_{ij}(\nu) = S(\nu) * T^2\,\mathrm{sinc}^2(T\nu). \tag{4–42}$$

The $\mathrm{sinc}^2(\cdot)$ spectral response of the FX correlator, shown in Fig. (4–8), has both advantages and disadvantages compared to the $\mathrm{sinc}(\cdot)$ response of the lag correlator (shown by a dashed curve in Fig. (4–8) for comparison). It is narrower, and has lower sidelobes, which generally are preferable characteristics for spectroscopy. Indeed, the FX-type Green Bank Spectral Processor has been exploited specifically to suppress interference signals. On the other hand, the narrow profile causes a greater "ripple" as a narrow line is scanned across the discrete spectrum. These characteristics are compared in the following table.

6.4. FX Correlator: Segmentation

In addition to the narrower spectral response, truncation of the input time series has a further effect on the performance of the FX correlator. This can be made evident by applying an inverse Fourier transform to Eq. (4–42) to derive the effective cross-correlation function measured by the correlator:

$$\gamma'_{ij}(\tau) = \gamma_{ij}(\tau) \cdot T \wedge (\tau/T). \tag{4–43}$$

The "unit triangle function" $\wedge(x) = \sqcap(x) * \sqcap(x)$ is the self-convolution of $\sqcap(x)$, having unit amplitude at $x = 0$ and tapering at unit positive and negative slope

Table 4–1. Spectral Response of Correlators

Correlator: Response:	Lag $\mathrm{sinc}(\cdot)$	FX $\mathrm{sinc}^2(\cdot)$
Full Width at 50% [spectral channels]	1.21	0.88
Full Width at 10% [spectral channels]	1.82	1.48
Height of First Sidelobe	−0.22	+0.05
Half-Channel Ripple	0.64	0.41

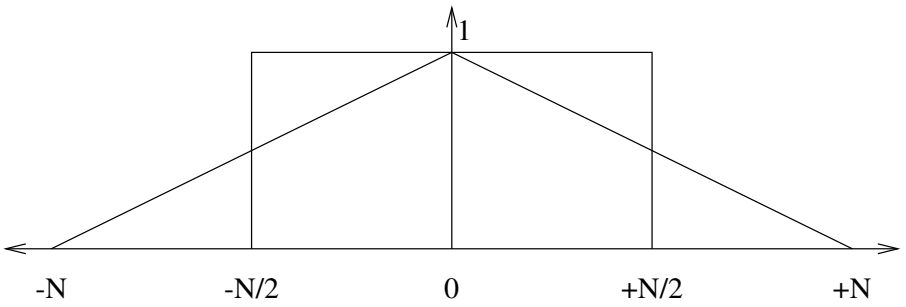

Figure 4–9. Relative density of lag measurements for lag and FX correlators.

to zero at $x = (-1, 1)$. Note that the width of $\wedge(x)$ is twice that of $\sqcap(x)$, while their integrals are equal.

Eq. (4–43) reveals two interesting characteristics of the FX correlator's response in the lag domain. First, transforming data samples in segments of length N yields a range of $2N$ lags covering $\tau \in (-N\Delta t, +N\Delta t)$. More importantly, however, these lags are heavily tapered by the triangle function; there are fewer pairs of samples within the set of N which can form large lags than small lags. Put another way, an FX correlator makes the same number of correlation measurements as a lag correlator with N lags, but they are distributed differently. The FX correlator reaches twice the lag range, but makes few measurements at the extremes of the range. This effect is shown schematically in Fig. (4–9).

The tapered distribution of independent cross-correlation measurements implies an oppositely-shaped distribution of noise in the measurements. The effect of this noise on a particular observation depends on the spectral characteristics of the source observed. A continuum spectrum has little power anyway beyond the first few lags away from zero, so the triangle tapering has little effect, while a narrow spectral line has significant power even at large lags. Before treating this effect of segmentation more quantitatively, we should consider an enhancement which can be applied when sufficient processing capacity is available: overlapping segments.

Operating with overlap factor f, an FX correlator advances its input data stream by only N/f samples for each N-point transform — requiring f times the processing capacity, of course. Some typical cases are shown in Fig. (4–10).

Overlapping does not change the triangle taper on the effective cross-correlation, nor the spectral response. It merely generates f times the number

Figure 4–10. Cases of overlapped segmentation for $N = 32$ and $f = \{1, 2, 4\}$.

of correlation measurements. Some of these measurements are new: at large lags, $|\tau| \geq \frac{f-1}{f}N$, overlapping segments contain pairs of samples not otherwise combined. For smaller $|\tau|$, however, previously measured correlations are just repeated in the overlapping segments.

Thus we define a measurement weight

$$W(\tau) = \begin{cases} 1, & |\tau| \leq \frac{f-1}{f}T \\ f(1 - \frac{|\tau|}{T}), & \frac{f-1}{f}T \leq |\tau| \leq T \\ 0, & |\tau| \geq T \end{cases} \tag{4–44}$$

with $T = N\Delta t$ as previously. This function, a trapezoid with ever steeper ends as f increases (and degenerating to a triangle for $f = 1$), gives the fraction of all possible correlations which are measured at each lag τ. Fig. (4–11) shows $W(\tau)$ for some practical cases of f.

Then, for cross-correlation amplitude given by $C(\tau)$, we can adopt this relative measure of signal-to-noise ratio:

$$R = \frac{\int_{-T}^{+T} W(\tau)C(\tau)\,d\tau}{\int_{-T}^{+T} C(\tau)\,d\tau}. \tag{4–45}$$

R is normalized to unity for the non-overlapped case $f = 1$ which yields the same number of correlation measurements as a lag correlator.

Finally, we can evaluate Eq. (4–45) in two limiting cases. For a continuum source, $C(0) = C_0$ and vanishes elsewhere, and

$$R_{\text{continuum}} = 1, \qquad \text{for any } f. \tag{4–46}$$

And for a monochromatic line, $C(\tau) = C_0$ for all τ;

$$R_{\text{line}} = \int_{-T}^{+T} W(\tau)\,d\tau = \frac{2f-1}{f}. \tag{4–47}$$

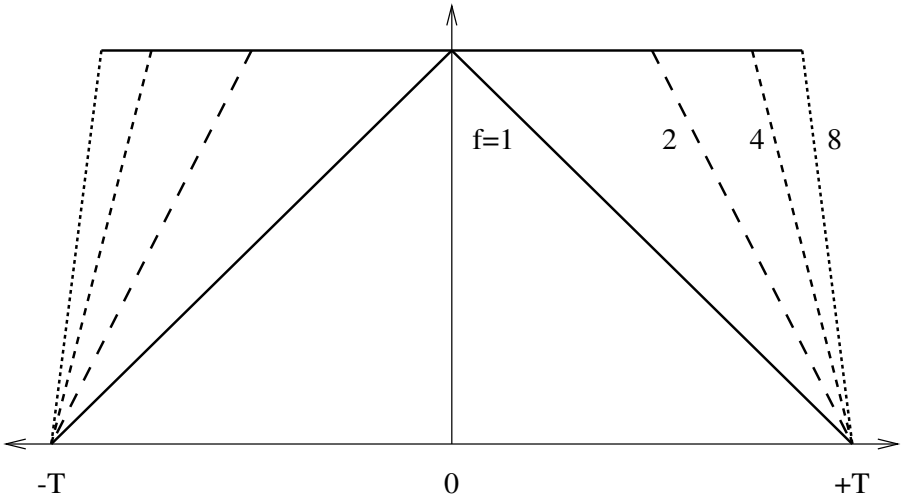

Figure 4–11. The measurement-weight function $W(\cdot)$ for some practical cases of f.

6.5. Miscellaneous FX Topics

Concluding my discussion of the FX correlator, I'd like to mention a few other aspects in which it differs significantly from the lag correlator. I'll start with a negative feature: it doesn't cope well with episodically invalid data. In a lag correlator, it's possible, at least in principle, to have a data-validity signal accompany each sample, and then inhibit correlation, independently at each lag, when either sample is invalid. In practice, this requires enough additional signal paths and normalization counters that it is seldom done — but in an FX correlator, it can't be done at all. Any practical FX design must use the FFT algorithm, which requires equi-spaced data samples. There is no way to omit an invalid sample. Generally this is not a problem at all for connected-element synthesis arrays. In VLBI arrays, the expedient of counting the errors for each FFT data segment, and invalidating the entire FFT output if the count exceeds some threshold, is usually satisfactory for typical recording and reproduction errors.

Next, I'll turn to two features which can only be implemented in an FX architecture. An input time-domain window can be applied at the same point as the phase rotation. Smoothing in the spectral domain, as commonly done in post-processing, typically applies only functions with a few points close to the peak, to minimize the computational load. With the input time-domain window, however, any real function can be used.

A pulsar gate can be applied at the FFT output, in the spectral domain. At that point, all delay and phase compensation has been performed, the signal had been resolved into narrow spectral bins, and an array-wide pulsar gate function — depending only on frequency and pulse phase — can be applied. The gate thus performs a form of de-dispersion impossible in a lag correlator, where the time-domain samples span an entire baseband channel. The Green Bank Spectral Processor uses this approach for single-dish pulsar de-dispersion. Such a gate was included in the design of the VLBA correlator, and in the original

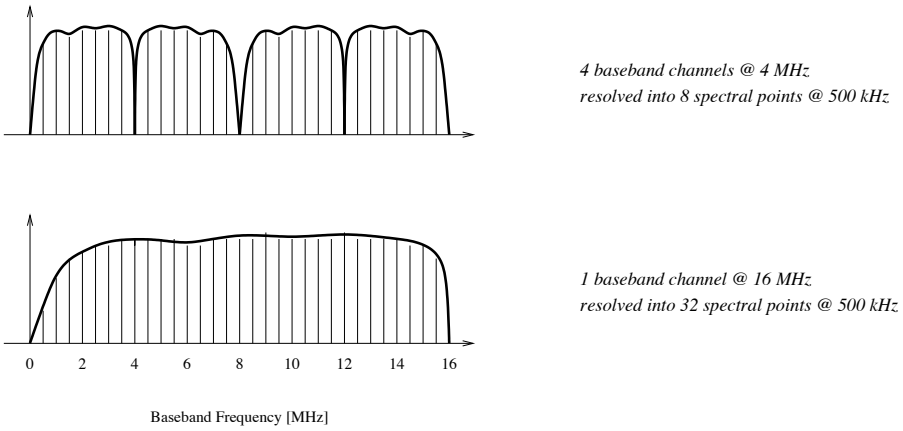

Figure 4–12. Example of an FX "hybrid mode".

hardware fabrication, but only became available recently with the completion of the necessary software.

Finally, we come to the dangerous "hybrid modes". It's possible to configure an FX correlator so as to transform, say, four contiguous baseband channels at relatively low spectral resolution, and cross-multiply the spectra against that from a single channel of four times the bandwidth, resolved into four times as many spectral bins. The individual bins from both sides thus match up, with the same frequency spacing. Such modes may be useful in VLBI arrays, to deal with incompatibilities among the disparate signal-processing and recording systems. Fig. (4–12) shows an example, which I must emphasize is purely conceptual, and has never been implemented. I've labeled these hybrid modes "dangerous" because, once people understand the concept, they expect any conceivable version to be achievable. In practice, each such case must be planned as part of the hardware design effort.

7. FX and Lag Correlator Intercomparison

To consolidate the material presented earlier, this final section will compare the two alternative correlator architectures from several quite different points of view. I'll begin with some very general considerations, then turn to a very particular basis for comparison — cost, and conclude by summarizing the other advantages and disadvantages of each.

Fig. (4–13) illustrates the four basic data types involved in interferometric correlation, and presents the lag and FX architectures graphically as two different routes from the common input, station sample streams, to the common result, baseline cross-power spectra. Two general conclusions should be apparent from this figure, as well as from the parallel derivations of the two correlators I presented earlier. First, the lag and FX correlators represent opposite sides of the Convolution Theorem of the Fourier transform: the transform of the convolution (correlation) equals the product of the transforms. And second, the FX architecture minimizes the number of operations required to perform the task,

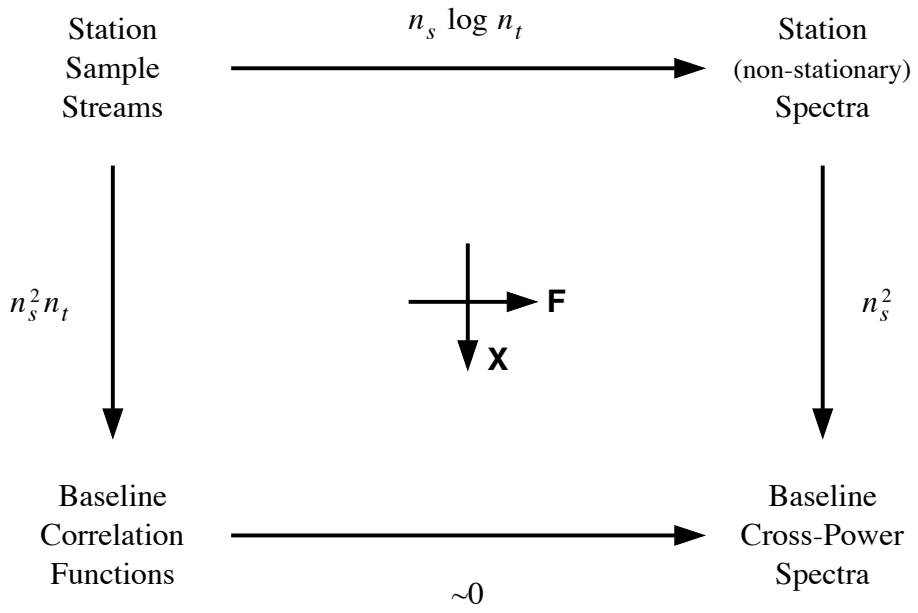

Figure 4–13. Comparative data processing paths for lag and FX architectures.

by exploiting the efficient FFT algorithm, and by organizing as much processing as possible on a station basis.

The figure shows, in terms of the system dimensions — n_s stations, and n_t samples transformed or lags formed — the operations needed along each of the four paths, and leads naturally into a more detailed discussion of the relative costs of the two alternatives.

7.1. Cost

Costs of the lag and FX architectures were compared during the design of the VLBA correlator. Two different approaches were followed, one based on simple operation counts (Romney 1986), and one on estimates of the required number of logic gates (Romney 1987). Both results are summarized here, for a correlator supporting n_s station inputs, each with n_c baseband channels (of sample rate r_0), and with n_t samples per segment transformed or lag correlations accumulated. The computations are based on $n_s^2/2$ baselines, including the real, single-dish "self-spectra".

The simplest approach to counting operations is just to count multiply operations. It can be shown that to a very good approximation, one addition accompanies each multiplication in both cases.

The lag correlator forms n_t lags per baseline and channel, each at rate r_0. As described earlier, phase rotation can only be performed on one input signal, so that each lag has one complex and one real input, requiring 2 multiplies. Thus the aggregate multiplier rate for cross-correlation is

$$r_L = r_0 n_s^2 n_c n_t. \tag{4–48}$$

The contribution of the subsequent FFT operation can be neglected, since these lags can be accumulated for some time beforehand, so r_L is then the total multiply rate for the lag correlator system

For simplicity, assume the FX correlator's FFTs are implemented using a straightforward radix-2 Cooley-Tukey algorithm. Then the transform consists of $\frac{1}{2}n_t log_2 n_t$ "butterfly" stages, of 4 multiplies each, per station and channel. Transforms are executed at a rate r_0/n_t, so the aggregate multiplier rate in the FX correlator is

$$r_F = 2r_0 n_s n_c log_2 n_t. \tag{4–49}$$

The cross-multiplier must process $n_t/2$ points of each station/channel spectrum (because only one side of the spectrum contains signal power), at the same rate r_0/n_t. Since the spectra are complex, 4 multiplies are required per point. The aggregate cross-multiplication rate is then

$$r_X = r_0 n_s^2 n_c. \tag{4–50}$$

The ratio of multiplies for the two architectures,

$$R_{\frac{Lag}{FX}} = \frac{r_L}{r_F + r_X} = \frac{n_s n_t}{2log_2 n_t + n_s}, \tag{4–51}$$

shows that the FX architecture requires significantly fewer operations when *either* of n_s or n_t is large.

The gate-count approach was developed to circumvent one very serious oversimplification of the operation counts: the multiplies are of different complexities, with the lag correlator requiring only a few bits. Some serious preliminary work is required before a gate-count comparison can be attempted. Fairly thorough designs under both architectures, *using the same generation of microelectronic technology* (preferably as current as possible), must be developed.

From these one can extract the basic gate-count units: g_L gates per lag per baseline, g_F gates per FFT stage, and g_X gates per baseline cross-multiplied. It's important to ensure that the operations included in these basic gate-count are as functionally comparable as possible.

Then the lag correlator's total gate count (using the same system parameters as previously) is

$$G_L = n_c \frac{n_s^2}{2} n_t \cdot g_L, \tag{4–52}$$

and the corresponding total gate count for the FX correlator

$$G_{FX} = n_c n_s \lceil \log_4 n_t \rceil \cdot g_F + n_c \frac{n_s^2}{2} \cdot g_X. \tag{4–53}$$

Note that a radix-4-plus-2 FFT implementation is now assumed; the logarithm is to base 4, and the integer ceiling operation $\lceil \cdot \rceil$ accounts for the case where a radix-2 stage is required for n_t an odd power of 2.

Obviously, the results of such an analysis will depend sensitively on the gate-count factors g_L, g_F, and g_X, which can change drastically with time. Numerical results from the VLBA correlator study, done more than a decade ago, thus are no longer of any relevance. They did, however, show that the FX

architecture would require fewer total gates, for any number of stations, for even quite modest spectral resolution — for the VLBA correlator's dimensions.

As developments in microelectronic technology make individual gates less expensive, however, other considerations may become more important, at least for very large-scale systems. The FFT stage of an FX system performs a data-*expansion* operation, and for the Millimeter Array currently under development it appears that the task of routing multi-bit signals to the cross-multiplier stage for an FX correlator would be prohibitive.

7.2. Summary of Other Advantages

The FX architecture allows a number of desirable features to be implemented which are not possible in a lag correlator. Some, such as truly station-based phase rotation and exact fractional-sample delay compensation, eliminate sensitivity losses as well as permitting more efficient implementation of essential functions. Others make unique, specialized observations possible (the spectral pulsar gate), or reduce the need for post-correlation processing (the time-domain window, and indeed the station-based phase rotation belongs in this category too). Its $\mathrm{sinc}^2(\cdot)$ spectral response may be superior, although it must be said there is as yet no experience with this. Although subject to segmentation loss, overlapped processing allows this to be restored, and to better sensitivity than an equivalent lag correlator.

The lag correlator is more tolerant of noisy transmission or recording of its input signals. And it actually obtains the cross-correlation function, which may be useful for certain cases of the correction for the non-linear sampling imposed on the observed signals.

References

Blackman, R. B., & Tukey, J. W. 1958, *The Measurement of Power Spectra* (New York: Dover).

Born, M., & Wolf, E. 1980, *Principles of Optics*, Sixth Ed. (Oxford: Pergamon).

Bracewell, R. 1965, *The Fourier Transform and Its Applications* (New York: McGraw-Hill).

Chikada, Y., et al. 1987, Proc. I. E. E. E., 75, 1203.

Clark, B. G., Weimer, R., & Weinreb, S. 1972, *The Mark II VLB System*, NRAO Electronics Division Internal Report 118.

Horiuchi, S., Kameno, S., Nan, R., Shibata, K., Inoue, M., Kobayashi, H., Murata, Y., Fomalont, E., and VSOG team 1998, *Advances in Space Research*, in press.

Parzen, E. 1962, *Stochastic Processes* (San Francisco: Holden-Day).

Romney, J. D. 1986, *Introduction to the Spectral-Domain ("FX") Correlator*, VLBA Correlator Memo No. 60.

Romney, J. D. 1987, *Spectral- vs. Lag-Domain Correlation: Comparison of Fundamental Hardware Requirements*, VLBA Correlator Memo No. 80.