VLBA Correlator Memo No. 80
(870220)

Spectral- *vs.* Lag-Domain Correlation: Comparison of Fundamental Hardware Requirements

JONATHAN D. ROMNEY
National Radio Astronomy Observatory
Charlottesville, Virginia

1987 February 13

Economy of implementation was "the principal advantage" cited in VLBA Correlator Memo 60 leading us to incline toward the spectral-domain (FX) correlator scheme. That memorandum (C60; this notation is used hereafter for all memoranda in this series) illustrated the arithmetical dialectic involved by a simple heuristic comparison of aggregate multiplier rates for the two correlator concepts. It was shown that the many-station high-resolution correlator specified for the VLBA requires a lag (XF) scheme to support over *two-and-a-half orders of magnitude more* multiply operations per unit time than an equivalent FX system. But it was also acknowledged that some multiplies are more equal than others, and that lag multiplies are significantly simpler than the multi-bit operations required in the FX algorithm. Nevertheless, the margin of advantage is sufficient that this calculation remains a useful, straightforward comparison at a level independent of technological considerations and architectural details.

At the opposite extreme in detail is a full cost comparison of complete system designs. One major disadvantage of pursuing such a comparison is the cost and effort involved in designing two systems when only one will be built. In the case of the VLBA, however, we already have in hand the lag correlator architecture developed by Caltech and NRAO personnel during 1984–85 and described in Memo C41; this remains the most thoroughly developed design for a VLBA correlator, although it is now several years out of date. A second serious difficulty with a such a "bottom line" comparison is that the results are likely to reflect primarily the set of specifications and modes supported, and the cost estimates for control and support equipment, rather than the costs to implement the FX and XF correlator algorithms.

To address the latter question, it thus seems essential to develop a comparison at an intermediate level of detail, in terms of the "fundamental hardware" requirements for the two correlator schemes. By this I mean only that hardware actually supporting the primary data paths through the correlator, thereby ignoring everything required for control and support, accumulation, and post-correlation processing. Chikada *et al.* (*Indirect Imaging*, ed. Roberts, 1983, p. 387) performed a similar comparison, introducing "complexity factors" in their version of the C60 calculation. Since practical correlator architectures are available implementing both approaches using commercially available gate arrays, it is both convenient and realistic to express the fundamental hardware requirements in units of gates. And since gates have a fairly well-established (but time-variable) dollar value, this comparison provides good cost estimates for actually doing the computations required in both algorithms.

It may be worth observing at the outset, though, that comparing "apples with apples" is fundamentally impossible, *for any realistic, multi-purpose correlator system.* The two architectures scale so differently with configuration that it is completely unrealistic to apply the same set of specifications to designs of both types.

The next two sections develop the gate requirements for FX and XF correlators, respectively, for configurations specified (in the notation of C60) by the numbers of (IF) channels n_c, stations n_s, and points transformed or lags calculated n_t (generally assumed to be a power of 2). Conclusions are drawn from these results, at increasing levels of specificity, in several subsequent sections, followed in turn by a discussion of several factors affecting the accuracy and significance of the conclusions.

FX Fundamental Hardware

The FX architecture and gate array design described by Escoffier in Memo C71, and by Escoffier and Greenberg in C72, form the basis for this calculation. In particular, I assume a radix-4-plus-2 implementation, using the multi-purpose gate array — Figure 5 of C72 — to support both the F and X processes. Different sections of the gate array are used in the two cases. The *complex multiplier*, containing 2128 gates, is common to both; this is followed by the 2986-gate *radix-4 combiner* to implement an FFT stage, or by the *accumulator adder* of 1920 gates for use in the baseline cross-multiply matrix. (These gate counts reflect recent further development of Greenberg's array design.) Thus the gate requirement for the FX correlator is given by

$$\left.\begin{aligned} G_{\text{FX}} &= n_c n_s \lceil \log_4 n_t \rceil \cdot g_{\text{F}} + n_c \frac{n_s^2}{2} \cdot g_{\text{X}} \\ g_{\text{F}} &= 5114 \quad [\text{ gates / channel / station / stage }] \\ g_{\text{X}} &= 4048 \quad [\text{ gates / channel / baseline }] \end{aligned}\right\} \quad (1)$$

Note that the logarithm is to base 4! The integer-ceiling operation $\lceil\cdot\rceil$ is required to account correctly for the final stage when it functions in radix-2 mode. The expression for the number of baselines provides for $n_s(n_s - 1)/2$ complex baseline cross-power spectra plus n_s real station spectra.

For later reference, we should review at this point just what elements of the signal path are supported by these gates. The Fourier transform and baseline cross-multiplication, of course, are fundamental to the FX algorithm. In accomplishing these, data are passed between stages in a complex floating-point representation (described in C71) comprising two 5-bit mantissa fields with a common 4-bit exponent. Within the gate array, however, only the complex multiplier exploits this format; the radix-4 combiner and accumulator adder are 16- and 32-bit fixed-point processors, respectively. In addition to the basic operations, the gates counted in (1) also support multi-level fringe rotation (*i.e.*, multiplication by a complex phasor, but *not* fringe-phase computation) and a dynamic, exact correction for the fractional sample delay error (known as "fractional bit shift" in one-bit correlators), both as described in C72. Finally, since the output spectra are produced only every n_t samples, the fundamental hardware accomplishes as well a short-term integration, amounting to 64 μs at the maximum 32 MHz sample rate.

XF FUNDAMENTAL HARDWARE

The 1984–85 XF correlator study summarized in C41 proposed developing a gate array to implement most economically the $\sim 10^5$ lags required. Almost simultanously with that study, planning began at the Netherlands Foundation for Radio Astronomy (NFRA) for a large European VLBI correlator, and to support this and other projects a gate array of remarkably similar specifications was designed by Bos (NFRA Internal Technical Report 176, NFRA Notes 488 & 490). Despite some differences in details, the similarity is sufficient — and, of course, far from fortuitous — that it would be accurate to say the NFRA array *is* the VLSI development proposed in C41. Accordingly, this calculation is based on the particulars of that array, which supports 16 real lags, equivalent to 8 complex lags, in an 8000-gate array, with (Bos reports privately) a 65% utilization factor. The XF correlator gate requirement is therefore

$$\left.\begin{aligned} G_{\mathrm{XF}} &= n_c \frac{n_s^2}{2} n_t \cdot g_{\mathrm{XF}} \\ g_{\mathrm{XF}} &= 650 \quad [\text{ gates / channel / baseline / complex lag }] \end{aligned}\right\} \quad (2)$$

As in C60, it is assumed here that the correlation functions can be averaged long enough that the F process is negligible. The dependence on n_s is identical to that in equation (1), but not entirely accurate for the lag case. Since the baseline correlation functions require $n_s(n_s - 1)/2$ complex lags, while the real, symmetric station autocorrelations require $\frac{1}{4}$ as many, the correct expression is $\frac{n_s^2}{2} - \frac{n_s}{4}$, but for convenience I have ignored the small second term. Quantizing the factor n_t is unnecessary since in any practical case it will be a multiple of the 16-lag quantum anyway.

The gates counted in (2) include three-level fringe rotation (but not fringe-phase computation, which is omitted from the NFRA gate array), and medium-term accumulation, up to 33 ms at 32 MHz sample rate. No proper fractional sample delay error correction is possible; with the VLBA's 32 MHz sample rate, this would require Fourier transforming the correlation functions at intervals as short as 2 ms, making this process no longer negligible.

COMPARISON — GENERAL REMARKS

Before discussing specific configurations and specifications, it will be instructive — and for some readers, perhaps, sufficient — to consider some general consequences of (1) and (2) above. The ratio $G_{\mathrm{FX}}/G_{\mathrm{XF}}$ is a monotonically decreasing function of both n_s and n_t for all realistic values of those variables; *i.e.*, not surprisingly, the fundamental hardware requirements favor the FX over the XF correlator for sufficiently many stations, *or* sufficiently high spectral resolution. The table below shows the locus of the crossover $G_{\mathrm{FX}} = G_{\mathrm{XF}}$, where temporarily n_s and n_t are treated as continuous, and the quantization of FFT stages in (1) is ignored.

$G_{\mathrm{FX}} = G_{\mathrm{XF}}$															
n_s :	0.8	1	1.5	2	3	3.2	4	5	7	10	13.3	15	20	28.6	∞
n_t :	64	50.8	32	24.3	16.9	16	13.6	11.8	10.0	8.7	8	7.8	7.4	7	6.2

One has to wonder, after inspecting the preceding table, why anyone would want to build a large XF correlator — at least, if economy in fundamental hardware is a basic concern. For even very modest spectral resolution, an FX correlator is clearly superior for any number of stations. The table suggests that a "moderate" size 10- or 15-station XF system, *for continuum only*, might be sensible if $n_t = 8$ is acceptable. Note, however, that with so few lags the XF correlator suffers an additional 2.5% sensitivity loss due to imperfect fringe-sideband rejection (Thompson, Moran, and Swenson, *Interferometry and Synthesis in Radio Astronomy*, 1986, p. 301). A practical limit is probably $n_t \geq 16$ — for which the FX algorithm is favored, even for a very small 3-station continuum-only system; note in particular that use of the NFRA gate array would require this limit, since it actually implements 16 *real* lags in series.

While suggestive, comparison in the present form would be decisive only for purists interested in gate economy as a metaphysical ideal. The next section considers actual magnitudes and costs of the hardware requirements for configurations of interest.

COMPARISON — SPECIFIC CONFIGURATIONS

The tables in this section present the gate requirements, calculated from equations (1) and (2) above, in several basic configurations relevant to the VLBA. Both G_{FX} and G_{XF} are given in units of *Megagates*. For gate arrays of the type and scale considered in this memorandum, an active gate costs about 1¢, so that each Megagate unit represents about $10,000 in cost of fundamental hardware.

Here and in the following section, I consider both 10- and 20-station configurations as representing realistic extremes for n_s. I think most would agree that $n_s = 10$ is inadequate as the basic correlator array, but it occurs as a natural tradeoff in some XF configurations. And $n_s = 20$ appears in part for historical reasons, again related to the XF arithmetic of equation (2) — but also because it provides dual 10-station capability, in either correlator scheme, in addition to accommodating a respectable extended array.

Continuum configurations are characterized by $n_c = 8$ and small n_t. The former is only half the channel capacity of the VLBA recording system, but this ratio has long been part of the correlator specifications. Several low-resolution cases are tabulated below, the lower limit being set by the practical restriction $n_t \geq 16$ proposed above.

Continuum Configurations ($n_c = 8$)

	$n_s = 10$		$n_s = 20$	
n_t	G_{FX}	G_{XF}	G_{FX}	G_{XF}
16	2.4	4.2	8.1	16.6
32	2.8	8.3	8.9	33.3
64	2.8	16.6	8.9	66.6

The differences between the two algorithms are unambiguous but not pronounced. Even though the continuum case is the least favorable to the FX correlator (for fixed array size), it still requires only half as many gates as an XF for a 20-station, minimum-resolution system. But the difference is "only" about 80 K$.

High-resolution spectroscopy represents the opposite extreme, where $n_c = 1$ and n_t is as large as possible. The cases tabulated range from an unacceptable $n_t = 256$ — which is the best resolution achievable in NRAO's Mark 2 correlator — to $n_t = 2048$, an octave beyond the VLBA specifications. (Recall that only $n_t/2$ useful spectral points are obtained.)

Spectroscopic Configurations $(n_c = 1)$

	$n_s = 10$		$n_s = 20$	
n_t	G_{FX}	G_{XF}	G_{FX}	G_{XF}
256	0.41	8.3	1.22	33.3
512	0.46	16.6	1.32	66.6
1024	0.46	33.3	1.32	133.1
2048	0.51	66.6	1.42	266.2

Here the advantage of the FX algorithm is apparent. It achieves the VLBA resolution specification, $n_t = 1024$, with ***only 1% of the gates*** required for a 20-station XF system, at a cost savings of 1.3 M$; in a 10-station system, the cost difference is 330 K$. Furthermore, another octave of resolution is achievable with only an 8% increase in gate count. As described in C72, this extra hardware also supports a proper correction for fractional-sample delay error, a sensitivity enhancement of particular importance for spectroscopy.

Comparison — Practical Specifications

The VLBA has been designed as a multi-purpose instrument, and the correlator therefore must support effectively a range of configurations. In assembling configurations from the preceding menus into a coherent system specification, the realities of the correlator algorithm inevitably will dominate the combination of configurations chosen. Put the other way around, the specification will be tailored to the algorithm chosen. This is particularly so for a lag correlator, where the required hardware is more extensive and must be used efficiently.

This section, then, compares two rather different practical systems: the FX correlator described in C72, and the XF architecture of C41. ***These systems are not directly comparable***, but each represents the best compromise among configurations subject to the limits imposed by equations (1) and (2), respectively. The tables for each system show both G_{FX} and G_{XF}, however, to make clear how well each algorithm performs in a foreign as well as a native environment.

In the FX scheme, the weak dependence on G_{FX} on n_t precludes any effective tradeoff between array size and resolution. The relatively small gate requirements in any configuration, documented in the preceding section, make the small inefficiency of a fixed 20-station array tolerable; balancing this inefficiency is the simplification achieved by eliminating a whole dimension of mode switching. Thus, two basic configurations — continuum and (high-resolution) spectroscopic — suffice to show the range supported. The "continuum" case, $n_t = 256$, is a natural consequence of the tradeoff between array size and channels; reduction of this resolution in the correlator backend represents a negligible cost. C72 also proposed another inefficiency: fixed 2048-point transforms with spectral averaging at the accumulator inputs to reach smaller effective n_t values. To reflect this effect, the third entry in the table below shows the actual fundamental hardware required for the C72 correlator.

FX Specifications

n_s	n_c	n_t	G_{FX}	(G_{XF})
20	1	2048	1.42	(266.2)
20	8	256	9.75	(266.2)
20	8	2048	11.39	(2129.9)

The organization of the XF correlator is more straightforward (or perhaps just more familiar). Spectroscopic modes at low, medium, and high resolution are shown to illustrate the tradeoff between resolution and array size, and the wideband continuum mode is an extreme case of the orthogonal tradeoff between resolution and channels. All these modes use (almost) exactly the same fundamental hardware, with switching required to support the reconfiguration.

XF Specifications

n_s	n_c	n_t	G_{XF}	(G_{FX})
10	1	1024	33.3	(0.46)
14	1	512	32.6	(0.75)
20	1	256	33.3	(1.22)
20	8	32	33.3	(8.93)

Comparison at this level, again, unanimously favors the FX correlator. It enjoys a 3:1 advantage in gate count, costs 220 K$ less, and represents a far more capacious spectroscopic system. An FX implementation of the XF-tailored specifications could still be realized for little more than $\frac{1}{4}$ of the cost. Conversely, an XF implementation of the FX-tailored specifications (including a slight inefficiency incorporated into the FX at a cost of 16 K$) would be a 3-Megalag machine costing over 21 M$!

Significance

Are any of these distinctions significant? The cost differential of 220 K$, equivalent to about three Congressional salaries, certainly must be considered seriously. However, in a complete correlator system the cost of fundamental hardware will be heavily diluted by "auxiliary" equipment including directly associated support and control hardware, post-correlation processing facilities, and computers for system-wide control. In several quite different and independent VLBA correlator cost estimates in recent years, fundamental hardware has accounted consistently for $\frac{1}{6}$ to $\frac{1}{4}$ of the total hardware cost, so it becomes conceivable that "noise" in the auxiliary cost might exceed the FX correlator's favorable differential.

A review of the possible differences, though, suggests that auxiliary costs will, if anything, also be lower for the FX system. Most of the higher-level control functions should be equivalent in FX and XF correlators. In direct support for the fundamental processing, the XF system requires more hardware to support fringe-phase computation, which it must do on a baseline basis; balancing this, the FX needs additional accumulation capacity to match the NFRA gate array. Neither of these factors is very significant, however.

The primary discrepancy appears to lie in post-correlation processing, where the XF has still to justify the F in its name. Typically, 45 1024-lag baseline correlation functions must be transformed, at intervals ranging (see C41) from 0.5 to 10 seconds, requiring computing rates between 0.2 and 4 Mflops. The upper end of this range is required to support wide-field mapping programs, and imposes an additional cost of perhaps 100 K$ on the XF correlator.

Summary

⋆ Analysis of the "fundamental hardware" requirements for primary data paths in the FX and XF algorithms, using gate-array designs based on similar levels of microelectronic technology, shows that a many-station and/or high-resolution correlator can be implemented more economically using the FX algorithm.

⋆ Only for a very small, 3-station continuum-only system is the XF correlator superior; a 10-station continuum XF correlator costs at least twice as much (in fundamental hardware) as an FX implementation, but the magnitude of the difference is small, ~ 20 K$.

⋆ An FX correlator for many-station high-resolution spectroscopy can be realized with only 1% of the gates required for an equivalent XF system.

⋆ Comparison of practical, multi-mode correlator designs for the VLBA demonstrates that the FX requires only $\frac{1}{3}$ of the gates, and costs 200 K$ less, compared to an XF system of decidedly lower performance.