VLBA Correlator Memo No. 75

(861120)

TWO-BIT CORRELATORS: MISCELLANEOUS RESULTS

FRED SCHWAB

November 19, 1986

Introduction. Two level (1-bit) and four level (2-bit) quantization schemes have been adopted in the design of the VLBA recording system. Four level quantization of the sampled signals is to be the primary mode of operation. For such coarsely quantized samples, if the true correlation is not close to zero then nonlinear corrections must be applied to the correlator outputs in order to obtain accurate estimates of the true correlation of the signals. For the four level case, accurate and economical correction formulae (e.g., polynomial or rational function approximations) are not available in the literature; that is the primary reason that I've written this memorandum.

In the traditional cross correlation spectrometer design, with post-correlation frequency analysis, quantization corrections are applied after correlation, but before applying the discrete Fourier transform operation (see Fig. 3–9 in [5]). In the design now under consideration for the VLBA (the 'FX' correlator), the Fourier transform operation is to precede cross multiplication. For this case, (after some data averaging) Fourier transformation back to the lag domain, followed by quantization correction, and followed by another Fourier transform, evidently are required if quantization corrections are to be applied.

In addition to presenting correction formulae—in the form of polynomial and rational functions—I'll also write down expressions for the correlator output for the case of non-identical quantization thresholds, since these formulae seem not to appear in the published literature. These results are neither very interesting nor illuminating, but, since I have them in hand, I want to record them for posterity.

Output of the Two-Bit Correlator. The operation of the idealized four level, or two-bit, quantizer is described by a step function q with values of ± 1 and $\pm n$ and steps at abscissae $a > 0$, $b < 0$, and 0:

$$q(x) \equiv \begin{cases} +n\,, & \text{for } x \geq a\,, \\ +1\,, & \text{for } 0 \leq x < a\,, \\ -1\,, & \text{for } b \leq x < 0\,, \\ -n\,, & \text{for } x < b\,. \end{cases}$$

For the case at hand, a two-input cross correlator, there are two quantization functions, q_1 and q_2—one for each data stream—with threshold abscissae a_i and b_i. For input signals $x(t)$ and $y(t)$, the correlator output is $\frac{1}{N}\sum_{k=1}^{N} q_1(x(t_k))q_2(y(t_k))$. For the case of zero-mean jointly stationary Gaussian input signals x and y, of unit variance and with cross correlation coefficient ρ, the expected value $r_u(\rho)$ of the correlator output is a weighted integral of the function

$$g(x,y,\rho) \equiv \frac{1}{2\pi\sqrt{1-\rho^2}}\, e^{-\frac{1}{2}\left(\frac{x^2-2\rho xy+y^2}{1-\rho^2}\right)}\,, \tag{1}$$

with piecewise constant weights as shown in Figure 1. This expectation,

$$r_u(\rho) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} q_1(x)q_2(y)g(x,y,\rho)\,dx\,dy\,, \tag{2}$$

can be expressed in terms of the bivariate normal integral (see [1, p. 936 ff.] for properties of the bivariate normal integral),

$$L(h,k,\rho) \equiv \int_h^{\infty}\int_k^{\infty} g(x,y,\rho)\,dx\,dy\,. \tag{3}$$

Explicitly, referring to Equations 2 and 3, one has, by inspection, $r_u(\rho) = n^2[L(b_1,b_2,\rho) + L(b_1,a_2,\rho) - L(b_1,-\infty,\rho) + L(a_1,b_2,\rho) + L(a_1,a_2,\rho) - L(a_1,-\infty,\rho) - L(-\infty,b_2,\rho) - L(-\infty,a_2,\rho) +$

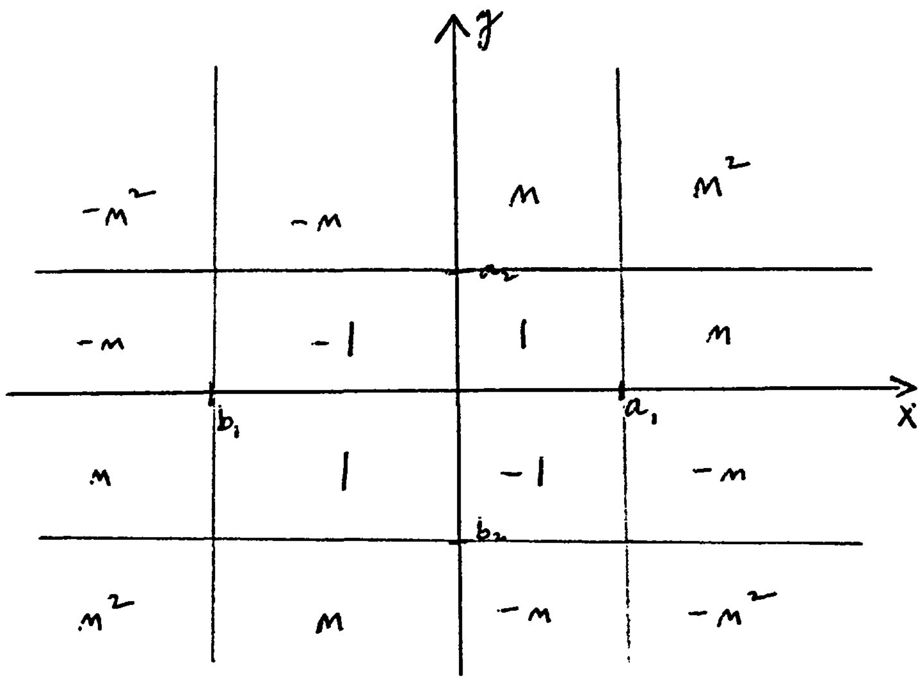

Figure 1. The expected value of the correlator output is a weighted integral of

$$g(x,y,\rho)=\frac{1}{2\pi\sqrt{1-\rho^2}}e^{-\frac{1}{2}\left(\frac{x^2-2\rho xy+y^2}{1-\rho^2}\right)}.$$

The piecewise constant weights are illustrated here.

$L(-\infty,-\infty,\rho)] + n[-2L(b_1,b_2,\rho) - 2L(b_1,a_2,\rho) + L(b_1,-\infty,\rho) + 2L(b_1,0,\rho) - 2L(a_1,b_2,\rho) - 2L(a_1,a_2,\rho)+L(a_1,-\infty,\rho)+2L(a_1,0,\rho)+L(-\infty,b_2,\rho)+L(-\infty,a_2,\rho)-2L(-\infty,0,\rho)+2L(0,b_2,\rho)+2L(0,a_2,\rho) - 2L(0,-\infty,\rho)] + L(b_1,b_2,\rho) + L(b_1,a_2,\rho) - 2L(b_1,0,\rho) + L(a_1,b_2,\rho) + L(a_1,a_2,\rho) - 2L(a_1,0,\rho) - 2L(0,b_2,\rho) - 2L(0,a_2,\rho) + 4L(0,0,\rho)$. This simplifies to

$$\begin{aligned} r_u(\rho) =&(n-1)^2[L(a_1,a_2,\rho)+L(a_1,b_2,\rho)+L(b_1,a_2,\rho)+L(b_1,b_2,\rho)+1] \\ &+2(n-1)[L(a_1,0,\rho)+L(b_1,0,\rho)+L(a_2,0,\rho)+L(b_2,0,\rho)] \\ &-n(n-1)[Q(a_1)+Q(b_1)+Q(a_2)+Q(b_2)]+\frac{2}{\pi}\arcsin\rho\,, \end{aligned} \tag{4}$$

where $Q(x) \equiv \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-t^2/2}\,dt = \frac{1}{2}\left(1-\operatorname{erf}\frac{x}{\sqrt{2}}\right) = \frac{1}{2}\operatorname{erfc}\frac{x}{\sqrt{2}}$.

For the case of identical quantization functions, $q_1 \equiv q_2$, and symmetrically placed quantizer threshold abscissae, $a = -b \equiv v_0$, Equation 4 simplifies to

$$\begin{aligned} r_u(\rho) =&(n-1)^2[L(v_0,v_0,\rho)+2L(v_0,-v_0,\rho)+L(-v_0,-v_0,\rho)] \\ &+4(n-1)[L(v_0,0,\rho)+L(-v_0,0,\rho)]-n^2+1+\frac{2}{\pi}\arcsin\rho\,. \end{aligned} \tag{5}$$

This is equivalent to the expression given by Hagen and Farley in [6],

$$r_u(\rho) = \frac{1}{\pi}\int_0^\rho \left\{(n-1)^2\left(e^{-v_0^2/(1+x)}+e^{-v_0^2/(1-x)}\right)+4(n-1)e^{-v_0^2/(2(1-x^2))}+2\right\}\frac{dx}{\sqrt{1-x^2}}. \tag{5'}$$

Now going back to the case of non-identical quantization thresholds, a formula equivalent to Equation 4 is

$$\begin{aligned} r_u(\rho) = r_u(0) + \int_0^\rho \Big\{&(n-1)^2(g(a_1,a_2,x)+g(b_1,b_2,x)+g(a_1,-b_2,x)+g(a_2,-b_1,x)) \\ &+2(n-1)(g(a_1,0,x)+g(b_1,0,x)+g(a_2,0,x)+g(b_2,0,x))+\frac{2}{\pi\sqrt{1-x^2}}\Big\}\,dx\,, \end{aligned} \tag{4'}$$

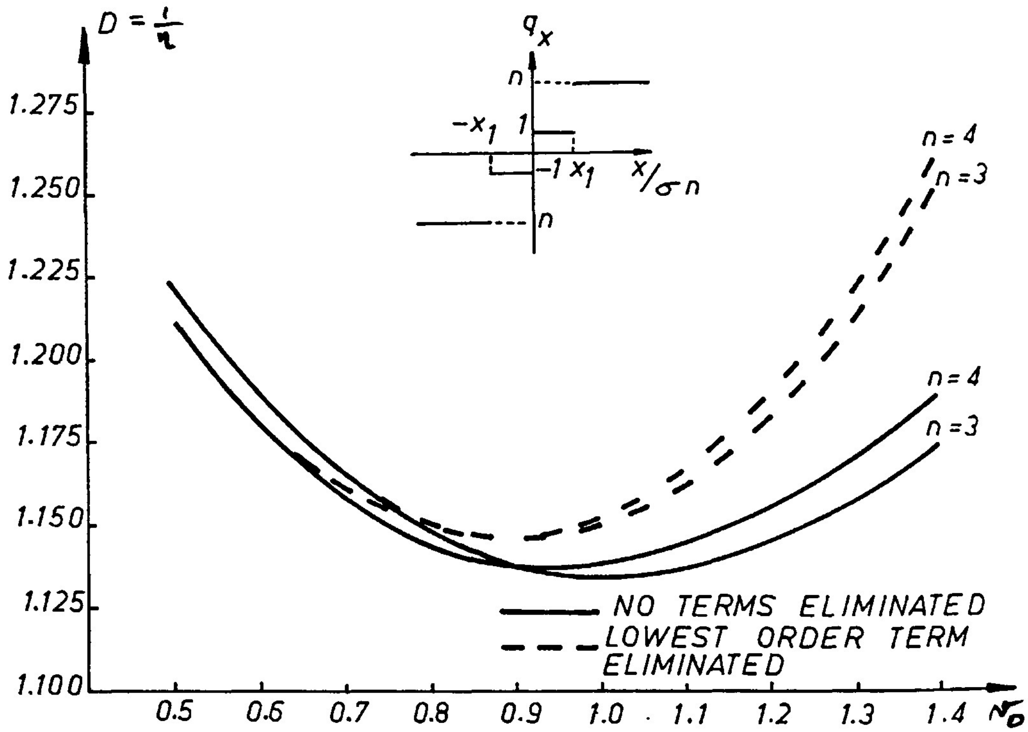

Figure 2. Plots of the reciprocal $D = 1/\eta$ of the correlator efficiency, for $n = 3$ and $n = 4$, as a function of the quantizer threshold setting v_0. (Adapted from Bowers and Klingler, [4].) (The dotted line cases are not considered in this memorandum.)

where, for conciseness, I haven't written out the exponentials in full. This integral representation follows from a theorem due to Price [7]: the mth derivative of the expected correlator output (for input signals of unit variance) is given by $\frac{d^m r_u}{d\rho^m} = \left\langle \frac{\partial^m q_1}{\partial x^m} \frac{\partial^m q_2}{\partial x^m} \right\rangle$. This method of derivation also is used in [6].

Note that for non-identical quantization thresholds $r_u(0) \neq 0$ and $r_u(-1) \neq -r_u(1)$, in general.[1]

For identical thresholds, the *normalized correlator output* is obtained by dividing the correlator output by $r_u(1)$, the expected value of the correlator output for $\rho = 1$. For $\rho = 1$ and identical quantization thresholds, Equations 5 and 5′ reduce to $r_u(1) = n^2 - (n^2 - 1)\,\mathrm{erf}\,\frac{v_0}{\sqrt{2}}$. We'll name this normalized output r: $r(\rho) = r_u(\rho)/r_u(1)$ is the *normalized, expected correlator output.*

Correlator Efficiency. For $|\rho| \ll 1$ and identical thresholds, the signal-to-noise ratio relative to that of an ideal analog correlator—i.e., the so-called *correlator efficiency* η—assuming band-limited noise equi-distributed over a rectangular band, and assuming sampling at the Nyquist rate, is given by

$$\eta = \left.\frac{dr}{d\rho}\right|_{\rho=0} = \frac{2}{\pi r_u(1)} \left((n-1)e^{-v_0^2/2} + 1\right)^2 . \tag{6}$$

The maximum value of η, namely $\eta \approx 0.8825$, occurs at $n \approx 3.336$, $v_0 \approx 0.9816$. As shown in Table 1 and in Figure 2, this is a rather broad maximum. For n held fixed at $n = 3$ or $n = 4$, and the threshold level v_0 chosen optimally, η stays approximately equal to 0.88. For n an integer, the optimal efficiency occurs at $n = 3$.

[1] One has $r_u(0) = (n-1)^2[1 + Q(a_1)Q(a_2) + Q(a_1)Q(b_2) + Q(b_1)Q(a_2) + Q(b_1)Q(b_2) - Q(a_1) - Q(b_1) - Q(a_2) - Q(b_2)]$, $r_u(1) = (n-1)^2[Q(\max(a_1, a_2)) + Q(a_1) + Q(a_2) + Q(\max(b_1, b_2))] - (n-1)[(n-2)(Q(a_1) + Q(a_2)) + n(Q(b_1) + Q(b_2))] + n^2 + 1$, and $r_u(-1) = (n-1)^2 \left[\begin{Bmatrix} 0, & \text{if } a_1 \geq -b_2 \\ 1 - Q(a_1) - Q(b_2), & \text{otherwise} \end{Bmatrix} + \begin{Bmatrix} 0, & \text{if } a_2 \geq -b_1 \\ 1 - Q(a_2) - Q(b_1), & \text{otherwise} \end{Bmatrix} \right] - (n-1)[n(Q(a_1) + Q(a_2)) + (n+2)(Q(b_1) + Q(b_2))] + n^2 - 2.$

TABLE 1. Optimal Quantization Thresholds		
n	v_0	η
3	0.99568668	0.8811539496
3.3358750	0.98159883	0.8825181522
4	0.94232840	0.8795104597

Best Straight-Line Approximations. The relative error curve corresponding to the best linear minimax[1] approximation $\tilde{\rho}$ to $\rho = r^{-1}$ over the range $0 \le r \le r(\rho_{\max})$ has its extrema at $r = 0$, $r = r(\rho_{\max})$, and it has a single zero at some intermediate point in that interval. Approximations of this type are shown in Table 2 (for $n = 4$ and v_0 chosen optimally, as in line 3 of Table 1).

TABLE 2. Best Linear Minimax Approximations		
$\rho_{\max}$	$\tilde{\rho}(r)$	$\max\limits_{0 \le r \le r(\rho_{\max})} \left\lvert \dfrac{\rho - \tilde{\rho}(r(\rho))}{\rho} \right\rvert$
0.1	$1.1368324r$	1.44×10^{-4}
0.2	$1.1363416r$	5.76×10^{-4}
0.3	$1.1355232r$	1.30×10^{-3}
0.4	$1.1343733r$	2.31×10^{-3}
0.5	$1.1328754r$	3.62×10^{-3}
0.6	$1.1309772r$	5.29×10^{-3}
0.7	$1.1285298r$	7.45×10^{-3}
0.8	$1.1251166r$	1.04×10^{-2}
0.9	$1.1191837r$	1.57×10^{-2}
1.0	$1.0641069r$	6.41×10^{-2}

These straight-line approximations are relevant to the FX spectral processor design if one chooses not to Fourier transform from the frequency domain to the lag domain to apply a quantization correction. This is because the minimax errors tabulated above indicate, as a function of the maximum correlation one wishes to consider, the degree to which it is valid to assume that there exists a linear relation between the values of the true cross correlation function and those of the Fourier transform of the cross spectrum obtained at the output of the FX processor.

Best Rational Minimax Approximations. I haven't constructed an exhaustive table of rational approximations, but for $n = 4$ and for v_0 chosen optimally, the following two—one an eleventh degree polynomial approximation in r, but the other just the quotient of a third degree polynomial and a first degree polynomial in r^2, multiplied by r—are examples of interesting best rational minimax approximations, valid for $|\rho| \le 0.95$:

$$\begin{aligned}\tilde{\rho}(r) =& 1.1371289r - 5.261628 \times 10^{-2} r^3 + 1.32608 \times 10^{-1} r^5 \\ & - 5.664224 \times 10^{-1} r^7 + 9.861185 \times 10^{-1} r^9 - 6.5297935 \times 10^{-1} r^{11} ,\end{aligned} \tag{7}$$

and

$$\tilde{\rho}(r) = \frac{1.1369813 - 1.2487891r^2 + 4.5380174 \times 10^{-2} r^4 - 9.1448344 \times 10^{-3} r^6}{1 - 1.0617975r^2} r . \tag{8}$$

With (7) the maximum relative error is 1.17×10^{-4}, but with (8) it is 2.77×10^{-5}—about four times smaller.

The degree of a rational function $R = p/q$ is usually defined to be the degree l of the numerator, plus the degree m of the denominator. $l + m + 1$ is the number of coefficients essential in order to define R (the number is not $l + m + 2$, because of cancellation). If we factor out an r in (7), to get

$$\begin{aligned}\tilde{\rho}(r) =& (1.1371289 - 5.261628 \times 10^{-2} r^2 + 1.32608 \times 10^{-1} r^4 \\ & - 5.664224 \times 10^{-1} r^6 + 9.861185 \times 10^{-1} r^8 - 6.5297935 \times 10^{-1} r^{10}) r ,\end{aligned} \tag{7'}$$

[1] A so-called *minimax* approximation is one which minimizes the maximum error, in absolute value, over the interval of approximation. And here I'm considering relative error, rather than absolute error.

then we have a polynomial of degree 5 in r^2, which is also a rational function of degree 5. But in (8) we have a superior approximation which is based on a rational function of degree 4 in r^2. This illustrates a common feature of rational approximations: that frequently a rational approximation is superior to a polynomial approximation of the same, or even higher, degree. (8) is the best rational approximation of degree 4, over the range $|\rho| \leq 0.95$.

Remarks.

- In VLBA Correlator Memorandum No. 74, John Benson mentions that the FX simulator program uses a default quantizer threshold setting $v_0 = .674\sigma$. The quantizer output levels are not given in the memo, but in talking with John I learned that the effective relative weights are $\pm 1, \pm 2$; this corresponds to $n = 2$, in the terminology of the present memo. Thus, from Equation 6 above, the efficiency η of the simulator's (default) quantization scheme is $\approx .82183$. The ratio of that number to the efficiency corresponding to optimal quantization ($n \approx 3.336$, $v_0 \approx .9816$) is 93.1%. I'd recommend that the simulator program be modified to use a higher efficiency quantization scheme since, in the final design, one certainly wouldn't want to do it as it's done now.
- With the FX design, a further loss in efficiency results from data segmentation. The amount of loss depends on the shape of the data window: more highly concentrated data windows (i.e., heavier tapering) lead to greater loss, but most of this loss can be recovered by employing sufficient overlap. For typical data windows, almost all of the lost efficiency can be recovered with 50% to 65% overlap of the data segments; further, in our application we should never need very heavy tapering—so, although there would be significant losses with no overlap, 50% overlap would suffice to recover almost all of the loss. I'll cover this in another memo.
- After a look at an earlier draft of this memo, Jon Romney pointed out that I hadn't considered the effect of moving the quantizer transition from -1 to $+1$ away from the abscissa $x = 0$. It would be straightforward to include such a shift in Equation 4, but I haven't done so. Since it's unlikely that the VLBA design will employ phase shifting to alleviate the effects of asymmetric quantizer thresholds, we may have to consider the point that Jon raised and we may have to do some quantitative work on asymmetries, to gain peace of mind.
- After a look at an earlier draft of this memo, John Granlund questioned why, in the section on rational approximations to the two-bit analog of the Van Vleck correction, I hadn't given approximations for the cases $n = 3$ and $n \approx 3.336$, since these values of n lead to higher efficiency than the case $n = 4$. For $n = 3$, and v_0 chosen optimally as in line 1 of Table 1, the approximation

$$\tilde{\rho}(r) = \frac{1.1347043 - 3.0971312r^2 + 2.9163894r^4 - .89047693r^6}{1 - 2.6892104r^2 + 2.4736683r^4 - .72098190r^6} r, \tag{9}$$

yields a maximum relative error of 1.51×10^{-4}. For $n \approx 3.336$ and v_0 chosen optimally as in line 2 of Table 1, the approximation

$$\tilde{\rho}(r) = \frac{1.1329552 - 3.1056902r^2 + 2.9296994r^4 - .90122460r^6}{1 - 2.7056559r^2 + 2.5012473r^4 - .73985978r^6} r, \tag{10}$$

yields a maximum relative error of 1.46×10^{-4}. And for $n = 4$, and v_0 chosen optimally as in line 3 of Table 1, the approximation

$$\tilde{\rho}(r) = \frac{1.1368256 - 3.0533973r^2 + 2.8171512r^4 - .85148929r^6}{1 - 2.6529114r^2 + 2.4027335r^4 - .70073934r^6} r, \tag{11}$$

yields a maximum relative error of 1.50×10^{-4}. These three approximations are valid for all $|r| \leq 1$. To obtain higher accuracy I would need to split the range of approximation into two or more parts (I can easily do so, on request, if anyone needs better approximations). With the FX design, I don't see any need to restrict the choice to integer weights (though for a lag-domain digital correlator non-integer weights are pretty much ruled out, and $n = 4$ might be easier to implement in hardware than $n = 3$).

- For any of the standard quantization schemes—i.e., two-level, three-level, multi-level, hybrid cases, or whatever—one can neatly express the correlator ouput in terms of the bivariate normal integral, as I've done for the four-level case in Equations 4 and 5.

References

1. Abramowitz, M. and Stegun, I. A. (Eds.), *Handbook of Mathematical Functions*, Second Edition, Nat. Bur. Stand. Appl. Math. Ser. 55, U.S. Govt. Printing Office, Washington, D.C., 1970.
2. Bowers, F. K. and Klingler, R. J., "Quantization noise of correlation spectrometers", *Astron. Astrophys. Suppl.*, 15 (1974) 373–380.
3. Cooper, B. F. C., "Correlators with two-bit quantization", *Aust. J. Phys.*, 23 (1970) 521–527.
4. D'Addario, L. R., Thompson, A. R., Schwab, F. R., and Granlund, J., "Complex cross correlators with three level quantization: design tolerances", *Radio Science*, 19 (1984) 931–945.
5. D'Addario, L. R., "Cross correlators", Lecture No. 3 in *Synthesis Imaging: Course Notes from an NRAO Summer School Held in Socorro, New Mexico, August 5–9, 1985*, Edited by R. A. Perley, F. R. Schwab, and A. H. Bridle, Workshop No. 13, NRAO, Green Bank, W. Va., 1986, pp. 31–48.
6. Hagen, J. B. and Farley, D. T., "Digital-correlation techniques in radio science", *Radio Science*, 8 (1973) 775–784.
7. Price, R., "A useful theorem for nonlinear devices having Gaussian inputs", *IRE Trans. Inf. Th.*, IT-4 (1958) 69–72.