Sub-channel stitching and truncation errors in the ALMA Tunable Filterbank

G. Comoretto

March 1, 2004

Abstract

In the ALMA Tunable Filterbank, the 2 GHz channel from one digitizer is divided into 32 sub-channels, that are separately correlated ad then stitched together. In this report, the accuracy of the sub-channel alignment is considered, in particular with regard to calibration errors due to truncation errors.

1 Introduction

If a digital filter is used in a correlator, the signal must be quantized several times. This causes calibration problems, especially if the quantization thresholds are not dynamically adjusted with high accuracy to the signal level.

In a hybrid correlator, these calibration errors cause alignment problems in the composite spectrum. Each sub-channel has a different total power level and this may cause both differences in calibration factors and offsets in the computed individual spectra. This, at the end, causes a degradation in the maximum dynamic range that can be obtained with such an instrument.

This problem has been raised for the ALMA enhanced filterboard. This board implements an hybrid correlator using the 32 time-multiplexed planes of the baseline correlator as 32 independently tunable sub-channels. This is a potential drawback for this design, as the original, time-multiplexed design does not suffer from this problem.

In this report we analyze this problem. In chapter 2 the problem is treated from an analytic standpoint, examining the effects of the quantization on the correlation function, the accuracy of the correction procedures, and their dependence on the signal level. This translates in requirements on total power measurements. In chapter 3 these concepts are validated by numeric simulations with realistic signals.

2 Analytic treatment

The analog signal from the IF processor is quantized several times before reaching the correlator:

- At the sampler, with 3 bit, 8 levels. The signal may vary by +/- 0.25 dB (+/- 3% amplitude), as the baseband processor output level is adjusted in 0.5 dB steps. Larger variations may occur, as the level is not dynamically adjusted.
- After the digital mixer, with 6 bit, 32 levels.
- After the first stage filter, with 8 bit (bit truncation)
- After the second stage filter, with 9 bit
- At the output stage, with 2 bit, 4 levels. Threshold levels are adjusted to match signal amplitude, but the adjustment is not dynamic: thresholds are adjusted at a given time, and do not track the signal.

When more quantization stages are done on the same samples, without intervening convolution operations, they can be treated together. For example, the last two quantizations are equivalent to a single 2 bit quantization, with the first one imposing only a finite step on the threshold setting.

Moreover, for the scope of this report, the first two quantizations modify the shape (amplitude, possibility of harmonic generation) of the whole 2 GHz channel, but do not affect the platforming process. Therefore we will focus on the effects of the 8 bit truncation after the 1st filter, and on the output 2 bit quantization. In the simulations, however, the whole digital filter will be considered.

2.1 Quantization effects

We have an analog signal, with RMS value σ, and whose correlation function is $R(\tau)$. The normalized correlation function is defined by $\rho = R/\sigma^2$.

In a digital system, the signal is quantized to a set of N discrete states, and is processed by a digital system, in our case a correlator. The digital correlator computes a (not normalized) correlation function

$$R_4 = f(\rho) = \sum w_{ij} P_{ij}(\rho) \tag{1}$$

where w_{ij} are the weights used in the correlation multiplication table, of size $N \times N$, and P_{ij} the joint probability of having two samples with states i and j at the multiplier inputs (see. e.g. Cooper, 1970 [3])[1]

We can expand the function $f(\rho)$ as $f(\rho) = a\rho + O(\rho^3)$, and define a *correlator gain* as

$$g = R_4/R = a/\sigma^2$$

Determining g is obviously of paramount importance to correctly platforming a spectrum computed by an hybrid correlator.

Gain, or in general the function $f(\rho)$, is a function of the weights w_{ij} and of the quantization thresholds $\{L_i\}$ (or, better, of the normalized threshold levels, $l_i = L_i/\sigma$). It must not be confused with the *quantization efficiency* ϵ (the degradation in signal to noise ratio), that can be computed by $\epsilon = a/\sigma_d^2$, with σ_d the RMS value of the digitized signal. Threshold levels and weights are chosen in order to maximize ϵ, with some technological constrains, but a is left unconstrained.

If thresholds are dynamically adjusted in order to keep L_i/σ at a constant, optimum level, then a (or $f(\rho)$) is also constant, and the sub-channels can be aligned measuring σ and multiplying each partial spectrum by σ^2. If this is not true, or if the adjustment is done with a finite precision, then the quantization correction depends on the total power, i.e. we have $f = f(\rho, \sigma)$.

In the following chapters we will analyze the two particular cases of large N quantizations, corresponding to the bit truncation at the first filter output, and the 2 bit case, corresponding to the output quantization.

2.2 Many bit case

When the number of quantization level is large, a linear approximation for $P_{ij}(\rho)$ can be used with good accuracy. Expanding the function $f(\rho)$ on powers of ρ (see appendix A), and with the assumption (used in both the correlator and the filter) that $w_{ij} = (2i+1)(2j+1)$, we obtain, at the first order:

$$g = \frac{2}{\pi\rho^2}\left(1 + 2\sum_{i>0}^{N/2} \exp\left(-\frac{1}{2}\frac{L_i^2}{\sigma^2}\right)\right)^2 \tag{2}$$

We then assume uniformly spaced thresholds, with a total quantization range expressed as the expected signal RMS level σ_e multiplied by a suitable factor α. σ_e has been introduced to allow the true signal RMS value, σ, to vary while thresholds are kept at a constant value. It is thus

$$L_i = i\,\alpha\sigma_e/N$$

[1] In this report we will assume that the quantization scheme is symmetric. We have N quantization levels, denoted by the index i ranging from $-N/2$ to $(N/2-1)$. Nonnegative i denote positive values for X. Level i is bonded by the thresholds L_i and L_{i+1}, with $L_0 = 0$ and $L_{\pm N/2} = \pm\infty$

If N is large enough, and $\alpha > 8$ (to prevent clipping), then the summation can be approximated with an integral, and $g = 4N^2\alpha^2/\sigma_e^2$, independently from σ. No quantization corrections are thus needed, as can be expected for an ideal many-bit analog-to-digital converter.

If α is lower, or σ/σ_e is significantly higher than 1, then the finite quantization range introduces truncation errors. A graph of g (normalized to its expected value) and of the quantization loss $(1-\epsilon)$ for 8 bit and 9 bit quantizers is shown in fig 1 as a function of $\sigma/\alpha\sigma_e$ The gain g (in blue) is the same for both cases, and is almost constant up to the point where truncation errors become significant, around $\alpha = 8$.

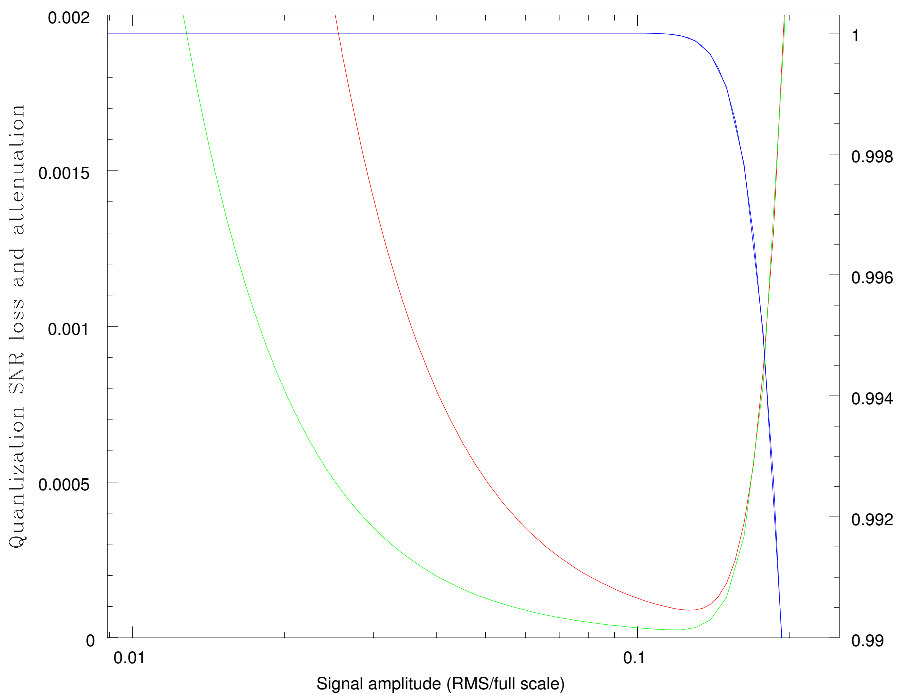

Figure 1: Gain (blue) and quantization loss for a 8 bit (red) and 9 bit (green) quantizers, as a function of the signal RMS level σ divided by the quantizer total range $\alpha\sigma_e$

The gain is unitary within 0.1% and 1% accuracy up to $\sigma/\alpha\sigma_e = 0.13$ and 0.2 respectively. The quantization loss is a few 0.01% in this range. To allow for variation in signal strength, a higher value for α must be adopted. With $\alpha = 16$ a dynamic range of 6 dB in the input signal level will not produce appreciable variations in g, while the maximum quantization loss is 0.04%. Under these assumptions, the 8 bit truncation of the first stage filter output do not produce appreciable ill effects.

2.3 2 bit case

The output quantizer, with $N = 2$, has a more severe effect on the correlation function. Expanding $f(\rho)$ to ρ^5 (as in appendix A) and expressing the expansion coefficients in term of $q = l_1 = L_1/\sigma$ gives:

$$
\begin{aligned}
R_4 &= a(q)\rho + b(q)\rho^3 + c(q)\rho^5 + O(\rho^7) \\
a(q) &= \frac{2}{\pi}\left(1 + 2\exp\left(-\frac{q^2}{2}\right)\right)^2 \\
b(q) &= \frac{1}{3\pi}\left(1 + 2(1-q^2)\exp\left(-\frac{q^2}{2}\right)\right)^2
\end{aligned}
$$

$$c(q) = \frac{1}{60\pi}\left(3 + 2(3 - 6q^2 + q^4)\exp\left(-\frac{q^2}{2}\right)\right)^2$$

If the thresholds (or the RMS signal amplitude) are different for the two channels, the coefficients are the geometric mean of the values computed for q_i and q_j.

To recover $R(\tau)$ from R_4, the above relation can be inverted to obtain ρ and the result multiplied by σ^2. Truncating the inverse relation to R_4^5 gives:

$$R = \sigma^2\rho = \sigma^2\left(a'(q)R_4 + b'(q)R_4^3 + c'(q)R_4^5 + O(R_4^7)\right) \quad (3)$$

$$a'(q) = \frac{s\sigma^2}{a} \quad (4)$$

$$b'(q) = -\sigma^2\frac{b}{a^4x)} \quad (5)$$

$$c'(q) = \sigma^2\left(\frac{3b^2}{a^8} - \frac{c}{a^7}\right) \quad (6)$$

In fig. 2 we plotted a' (red), b' (blue) and c' (cyan) as a function of $1/q = \sigma/L_1$ (the signal amplitude normalized to the threshold level used). The quantity $1/g = R/R_4$ is also plotted (green). The optimal value for 1/q is 1.01, and the plot covers the interval form 0.9 to 1.1 ($\pm$0.8 dB). In the Tunable Filter, threshold setting is done with 0.5% accuracy (0.05 dB step) in a 512×2bit lookup table, but since this adjustment is not performed dynamically, variations up to this level can be expected.

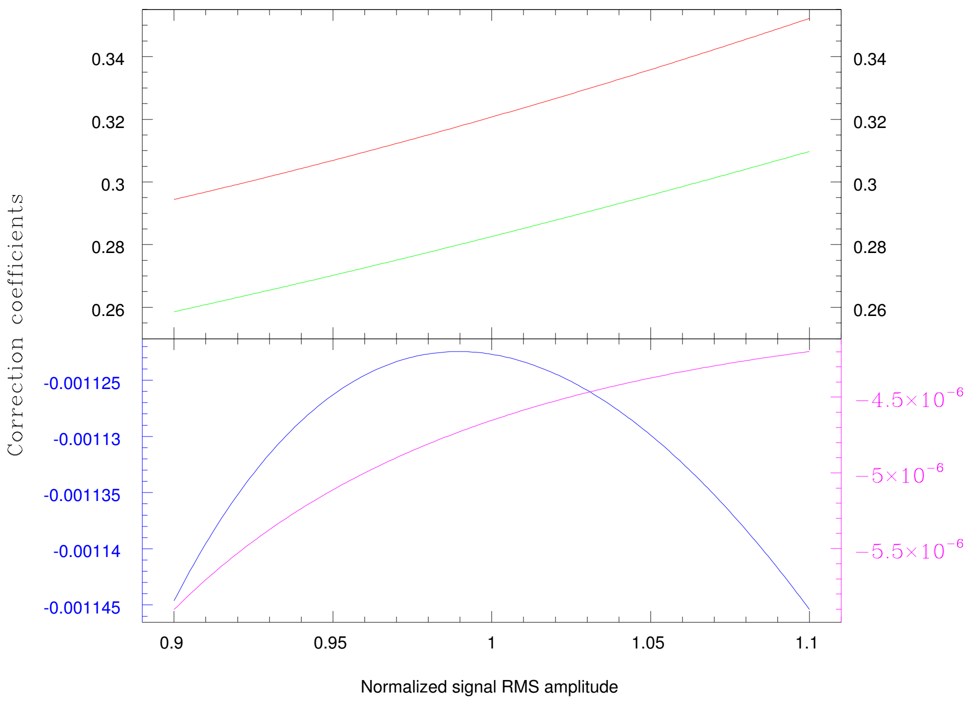

Figure 2: Polynomial coefficients of the correction function for 2 bit quantization. Horizontal scale is the signal RMS amplitude. Curves represent: red a', blue b', violet c', green R/R_4 for $\rho = 1$

This relation is accurate only for values of ρ not close to unity. The function $R_4(\rho)$ increases steeply near $\rho = 1$, and a more accurate relation (see e.g. Kogan [5] [6]) must be used. The relation 3 is however quite accurate up to $\rho = 0.3$, and much simpler to implement. In this report, we used this approximation to correct for quantization, and resorted to the more accurate integral relation of Hagen et al. [4] for the few data points with $\rho > 0.3$.

The main source of calibration error is the variability of a'. In the relevant range, the relative error in a' is about half the error in σ. To achieve a dynamic range of 40 dB, therefore, the signal level before the output digitizer must be measured with comparable accuracy. For a 62.5 MHz signal, this translates to an integration period of the order of 1 second. Total power measurements on the filter outputs must be done continuously during normal operation to achieve a calibration accuracy between sub-channels at 1/10 the spectral noise.

It should be noted however that total power can be derived from autocorrelation at zero lag. For fast observations, this should provide a reasonably good estimation of the total power level without stressing the communication network. For accurate measures, the total power meter for each sub-channel must be able to integrate for long periods, up to several seconds, and have to be sampled at this rate.

3 Simulations

To test the above concepts in a realistic situation, a functional equivalent of the whole filterbank has been simulated. The simulator is written in C++, and includes all the truncations and approximations used in the proposed filterbank. The mixer has been simulated using a lookup table, and filter coefficients are those computed for the filterboard. The signal level at the output of the second filter has been normalized to the optimum value for the 2-bit quantization stage within $\pm 2\%$, in order to simulate threshold adjustment with finite precision. The normalization factor has been recorded, and applied to the spectra after the clipping correction.

The test signal is composed of two partially correlated sequences. with a length of 64 Msample each (16 ms at 4 GHz). Both the uncorrelated and correlated parts are composed of shaped Gaussian noise, with the power spectrum shown in fig. 3, resp. in red and blue.

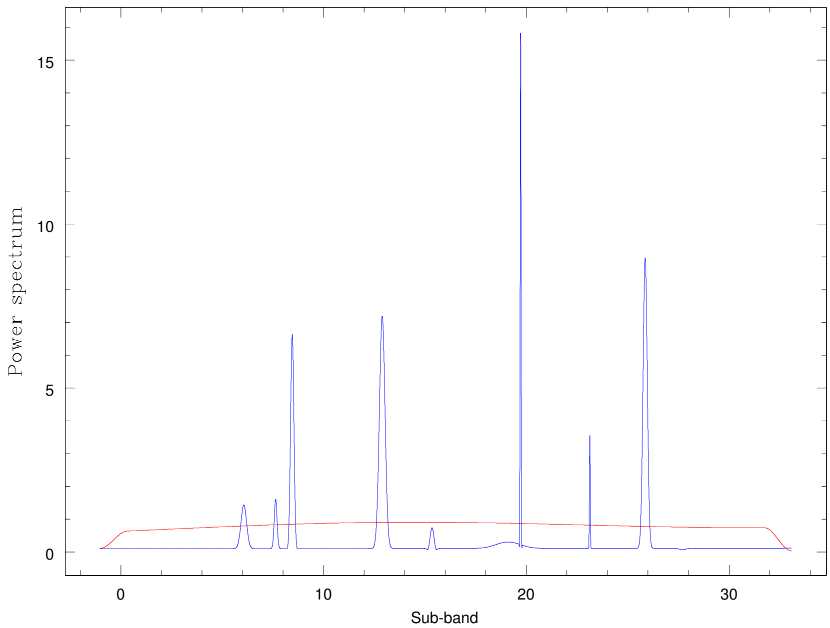

Figure 3: Test signal spectrum. Red is the uncorrelated part, and blue is the correlated one. Each spectral feature in the correlated part has a different phase. The correlated continuum level is 0.1. The spectrum is normalized to unity average spectral density. Horizontal scale is in channels (0 to 31).

The correlated part is composed both of a continuous component and several lines of various width and different phases. The relative total power in the correlated part is about 20% of the total, but in some sub-channels it is up to 3 times the uncorrelated part. In this way, a dynamic range of up to 6 dB among sub-channels is simulated. Peak correlated spectral density and correlated continuum level are resp. 16 times and 10% the average spectral density.

We analyzed this signal using several different procedures. For reference, a first reference cross spectrum has been computed using a simulated full-band analog crosscorrelator with the same spectral resolution and tapering function of the hybrid correlator.

The two sequences were processed by two identical software "filterbanks", and analyzed by a software digital correlator using the same multiplication table employed in the Alma baseline correlator. Each sub-channel was analyzed with 127 lags (63 negative, 63 positive, and zero), using Hanning tapering, and producing a 64 point cross spectrum. The 32 sub-channels overlap by 4 channels, and 2 channels at each edge were discarded.

The clipping correction used was that described in the previous chapter, for the 2-bit quantization only. At the moment, no correction for multiple quantizations was used. The quantity $q = L/\sigma$ was estimated by computing the variance of the second stage filter output. Individual sub-channels have been calibrated for filter shape, edge effects and tapering effects with the procedure given in [2].

For reference, we also analyzed the signals before the last 2-bit quantization stage, and the signals processed by a full accuracy hybrid correlator. In these cases, no quantization correction has been applied. The full precision hybrid correlator gives negligible errors, well below -40 dB, and thus we can reasonably assume that the errors seen in the remaining spectra are due to quantization.

The resulting hybrid spectrum has been divided by the nominal reference spectrum, giving the results in fig. 4. The left spectra refer to the full resolution data, while on the right these residuals have been smoothed with a 16 channels boxcar filter, to reduce the noise level. The black graphs refer to the 2-bit digitized data, while the red ones refer to the signals processed by the digital filterbank, but not re-quantized after the filtering. These spectra are dominated by quantization noise, that for a 16 ms data set analyzed at 2 MHz effective resolution and 2 bit quantization is about 0.3% (or 3% of the correlated continuum level). Smoothed spectra show residuals misalignment up to 1%, again dominated by the quantization noise. This corresponds to -30 dB with respect to the input RMS level, or -42 dB with respect to the peak correlated signal.

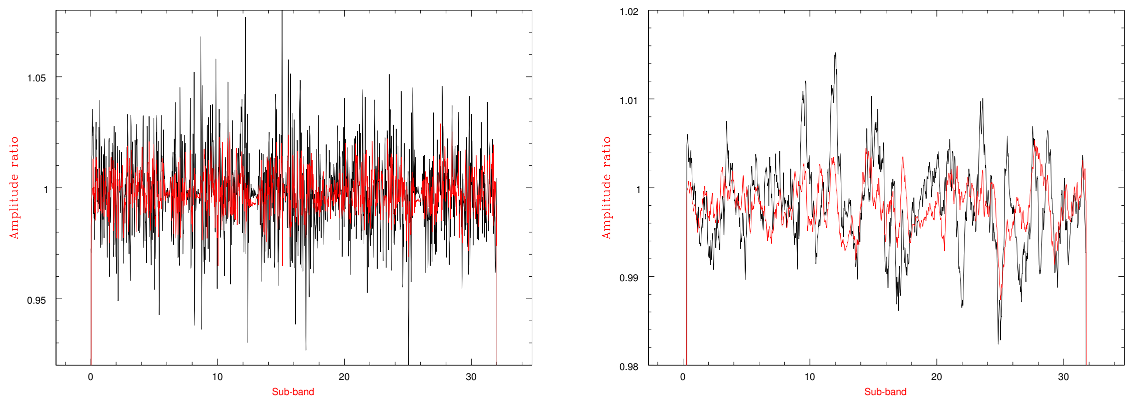

Figure 4: Ratio of hybrid cross spectrum to the original one. Horizontal scale is in channels (0 to 31). Left graph is for the full resolutions (64 channels/sub-channel), right is smoothed to 4 channels/sub-channel. Red graphs refers to signal processed by the filterbank but not quantized to 2 bit

In fig. 5 the difference between the reference and the hybrid spectra is shown. Black and red graphs refer to residuals for the 2-bit quantized and the unquantized digital outputs. Green graph (scale on the right) is the original cross spectrum. Differences may be up to a few parts in 10^{-4}, in correspondence of strong spectral features, giving rise to calibration errors of this order. These errors are apparent also for the unquantized digital output, and are thus due to the other quantizations (sampler, digital mixer) in the system. Sub-channel alignment is again much better than the noise. in this simulation (max. 1%, or 0.1% of the uncorrelated power), despite strong variations in sub-channel to sub-channel power level and the presence of strong spectral features.

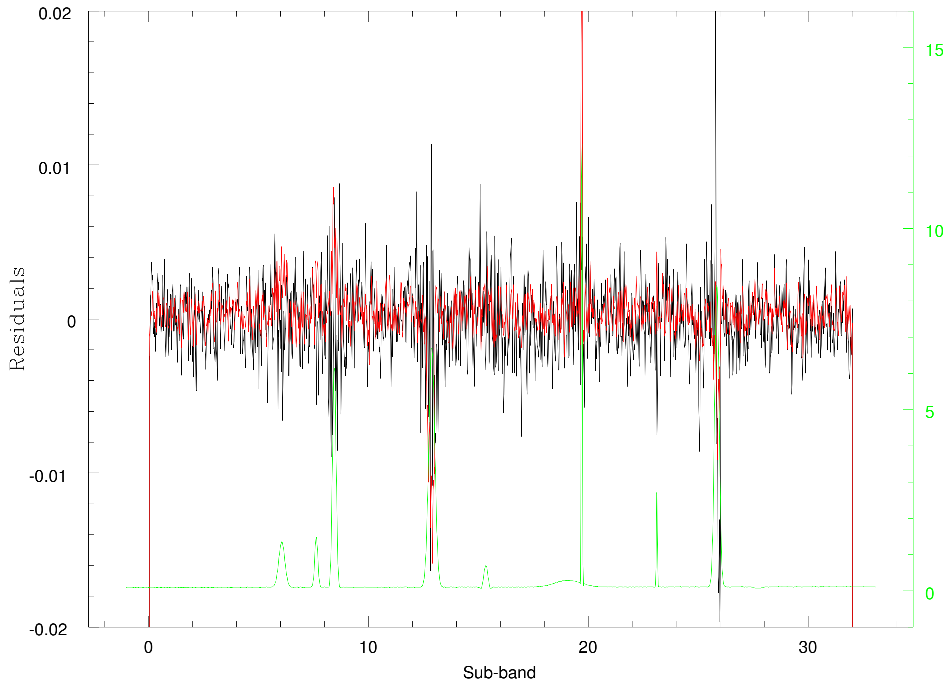

Figure 5: Residuals for the hybrid spectrum, both with (black) and without (red) final 2-bit requantization. Horizontal scale is in channels (0 to 31). Green line is the original cross spectrum, with scale on the right.

The same test signal, but with the correlated and uncorrelated parts respectively multiplied by 3 and 0.3, has been used to try to obtain a better signal to noise ratio despite the short integration time. The total power in the correlated part is thus ≈ 2 times that in the uncorrelated one. This causes however unrealistic variations in power level among sub-channels, up to more than 13 dB (see the amplitude measurements for sub-channels, represented as blue crosses in fig. 6), and is therefore an extreme sort of *stress test* for the system. The ratio of the hybrid cross spectrum to the reference one is shown in fig. 6, together with this latter (green line).

The sub-channel alignment is still within the noise (much better than 1%), except than in those sub-channels with a very strong spectral feature. Sub-channels 8, 12 and 25 have an in-channel power of more than 10 dB above other channels, and this reflects in clipping errors at the 8-bit quantization stage. A corresponding reduction in channel gain of up to 5% is visible in fig. 6. More work is in progress to see if better results can be obtained with some form of gain control after the 1st stage filter. Analyzing these sub-channel in detail one can see that the error is due to an incorrect calibration of the digitized autocorrelation function, producing errors of the order of 0.2% in the spectral shape. Since the total power in the sub-channel is ≈ 20 times that of other sub-channels, this error is magnified to the observed 5% level.

For comparison, the reference cross spectrum is also plotted, in green. The peak signal strength in correspondence with the larger errors is is more than 150 times the continuous signal, where the larger errors are visible. These errors correspond thus to a dynamic range of 3000 in this test case.

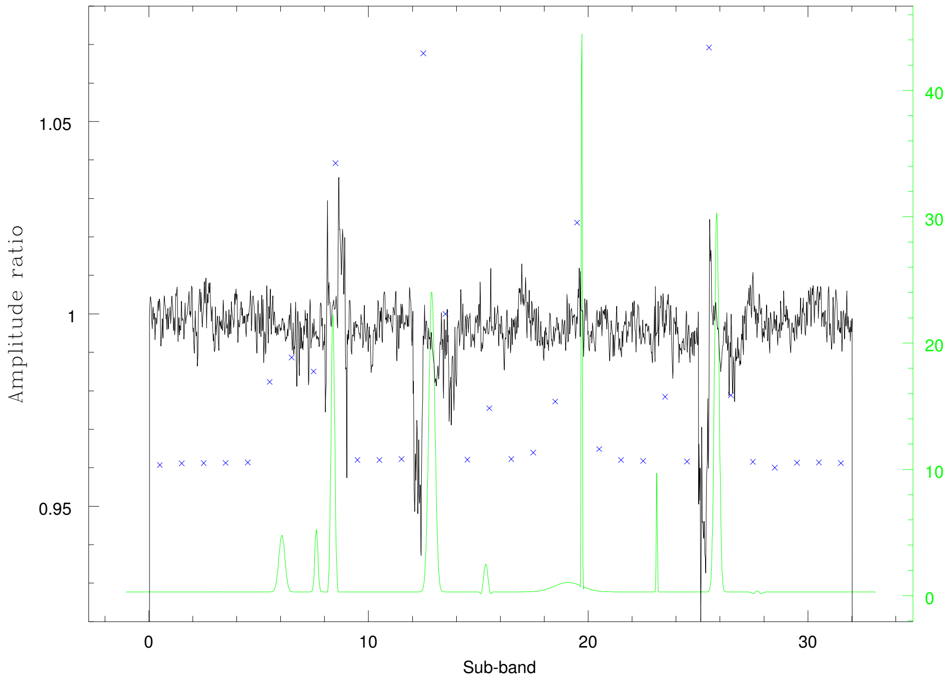

Figure 6: Ratio of hybrid cross spectrum to the original one. Horizontal scale is in channels (0 to 31). Green line is the original cross spectrum, with scale on the right. Blue crosses indicate RMS power in individual sub-channels, in arbitrary units (right scale).

4 Conclusions

These simulations show that the alignment procedure described allows sub-channel alignment better than the noise level for the data set used, i.e. $< 1\%$, or $< 0.1\%$ of the uncorrelated power, despite strong variations in sub-channel to sub-channel power level and the presence of strong spectral features.

Quantization errors in the sampler and digital mixer produce noticeable effects, of the order of 10^{-4} in the shape of strong spectral features under realistic conditions, but not contamination of other parts of the spectrum to better than $4\,10^{-5}$.

Pushing the simulation to very strong spectral signals, with total power in the spectral features several times the continuum or uncorrelated noise contributions, and with differences among sub-channels of more than 13 dB, we found that these quantization errors cause calibration errors of the order of a few percent, corresponding to a dynamic range of ≈ 3000. Thus a quantization correction procedure for these effects is desirable to achieve very high dynamic range.

A Correction for N bit quantization

Let $R(\tau)$ be the correlation function of two stochastic signals, X_1 and X_2, of null mean and standard deviation σ. If $x_{1,2} = X_{1,2}/\sigma$ are the normalized signals, let their joint distribution be bivariate, with correlation $\rho = R/\sigma^2$:

$$P(x_1, x_2; \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x_1^2 + x_2^2 - 2\rho\, x_1 x_2}{2(1-\rho^2)}\right) \tag{7}$$

In a digital correlator, the signal $X_{1,2}$ are quantized using a set of thresholds $\{L_i\}$, and the correlation multiplier uses a multiplication table with products w_{ij}. The multiplication table is assumed symmetric.We adopt the convention for the index i to have it ranging form $-(N/2)$ to $N/2-1$, with nonnegative index representing positive values. With this convention, for a full multiplier it is $w_{ij} = (2i+1)(2j+1)$, $L_0 = 0$, and $L_{\pm N/2} = \pm\infty$. It is also convenient to define $l_i = L_i/\sigma$ The digital correlation function $R_4(\tau)$ can be computed using:

$$
\begin{aligned}
R_4(\rho) &= \sum_{i,j=-N/2}^{N/2-1} w_{ij} P_{ij} \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}} \sum_{i,j=-N/2}^{N/2-1} w_{ij} \int_{l_i}^{l_{i+1}} \int_{l_j}^{l_{j+1}} \exp\left(-\frac{x_1^2 + x_2^2 - 2\rho\, x_1 x_2}{2(1-\rho^2)}\right)
\end{aligned}
$$

The integral is not easily computable. A possible solution, considering that usually the correlation coefficient is small, is to expand the above expression in powers of ρ, reducing the integral to a series of terms of the form

$$\rho^n \int x_i^p \exp(-x_i^2/2) dx_i \int x_j^q exp(-x_j^2/2) dx_j$$

. Terms with n even cancel due to symmetry in w_{ij}, while terms with n odd give equal contribution in all four quadrants. The resulting expression, to ρ^5, is then:

$$
\begin{aligned}
R_4(\rho) &= \frac{2}{\pi} \sum_{i,j=0}^{N/2-1} \sum_{k=1}^{5} \frac{\rho^k}{k!} w_{ij} \int_{l_i}^{l_{i+1}} Q_k(x_1) \exp\left(-\frac{x_1}{2}\right) dx_1 \int_{l_j}^{l_{j+1}} Q_k(x_2) \exp\left(-\frac{x_2}{2}\right) dx_2 \\
Q_1(x) &= x \\
Q_3(x) &= x(1-x^2) \\
Q_5(x) &= x(3-3x^2+x^4)
\end{aligned}
$$

If the weights w_{ij} are given by the above relation for a full multiplier, $w_{ij} = (2i+1)(2j+1)$, the integral gives:

$$
\begin{aligned}
R_4(\rho) &= \frac{2}{\pi} \sum_{k=1}^{5} \frac{\rho^k}{k!} (C_k(l_i))^2 \\
C_0 &= 1 + 2 \sum_{i=1}^{N/2-1} \exp(-l_i^2/2) \\
C_1 &= 1 + 2 \sum_{i=1}^{N/2-1} (1-l_i^2) \exp(-l_i^2/2) \\
C_2 &= 3 + 2 \sum_{i=1}^{N/2-1} (3-6l_i^2+l_i^4) \exp(-l_i^2/2)
\end{aligned}
$$

This relation is quite accurate for small ρ. For the 2-bit case, the error with respect to the more accurate integral relation given by Hagen [4] is 10^{-6} for $\rho = 0.3$, but quickly increases for larger ρ.

The signal amplitude before quantization can be recovered either using a dedicated total power meter, or using the autocorrelation measures available for all antennas. The autocorrelation at zero lag is in fact related to the threshold normalized amplitude q, for 4 bit quantization, by the relation $R_4(0) = 9 - 8\mathrm{erf}(q/\sqrt{2})$. This relation can be inverted numerically to obtain the value for the quantity q, i.e. the ratio of the actual threshold to the signal RMS amplitude.

B Quantization efficiency

For a quantizer with output values w_i, and full multiplication correlator, the quantization efficiency, i.e. the degradation of the signal to noise ratio, is given by the formula:

$$
\begin{aligned}
\epsilon &= 2 \left(\frac{\sum_i w_i A_i}{(\sum_i w_i^2 B_i)} \right)^2 \\
A_i &= \frac{1}{\sqrt{2\pi}} \int_{l_i}^{l_{i+1}} x \exp\left(-\frac{x^2}{2}\right) dx \\
&= \frac{1}{\sqrt{2\pi}} \left(\exp\left(\frac{-l_{i+1}}{2}\right) - \exp\left(\frac{-l_i}{2}\right) \right) \\
B_i &= \frac{1}{\sqrt{2\pi}} \int_{l_i}^{li+1} \exp\left(-\frac{x^2}{2}\right) dx \\
&= \frac{1}{2} \left(\mathrm{erf}\left(\frac{-l_{i+1}}{\sqrt{2}2}\right) - \mathrm{erf}\left(\frac{-l_i}{\sqrt{2}2}\right) \right)
\end{aligned}
$$

It is more convenient to compute $1-\epsilon$, that is the fractional increase in noise due to the quantization process.

In fig. 7 this quantity is plotted as a function of the input RMS amplitude, for 2 bit, 2×4 bit, and 4 bit quantization schemes. The amplitude is given in relation to the threshold of the second significant bit, i.e. the first after the sign bit.

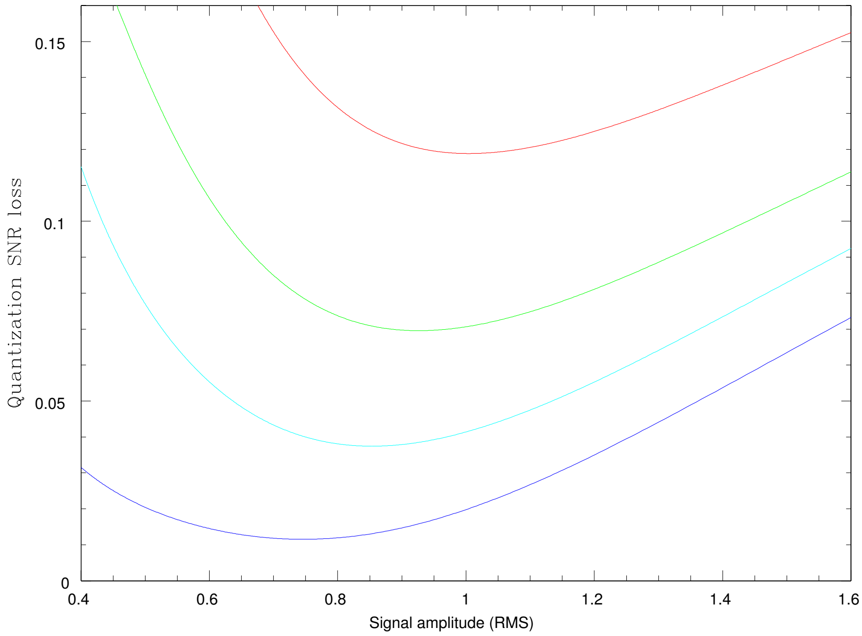

Figure 7: Quantization efficiency (left) for quantization schemes with: 2 bits, 4 level (red), 2×4 bits (green), 3 bits (cyan) and 4 bits (blue). Horizontal scale corresponds to an amplitude form 0.4 to 1.6 times (-8 to $+4$ dB) the central threshold in the quantization scheme.

References

[1] Comoretto, G.: *A digital BBC for ALMA*, ALMA memo 305 (2000).

[2] Comoretto, G.: *Phase effects in a hybrid correlator*, Arcetri Observatory Technical Memo 3/2003 (2003)

[3] Cooper, B.F.C.: *Correlation with Two-bit Quantization*, Ausl. J. Phys, **23**, 521 (1970).

[4] Hagen, J.B., Farley, D.T., Radio Science 8, 775 (1973)

[5] Kogan, EVLA Science Memo 5 (1993)

[6] Kogan, Radio Science 33(5), 1289 (1998)