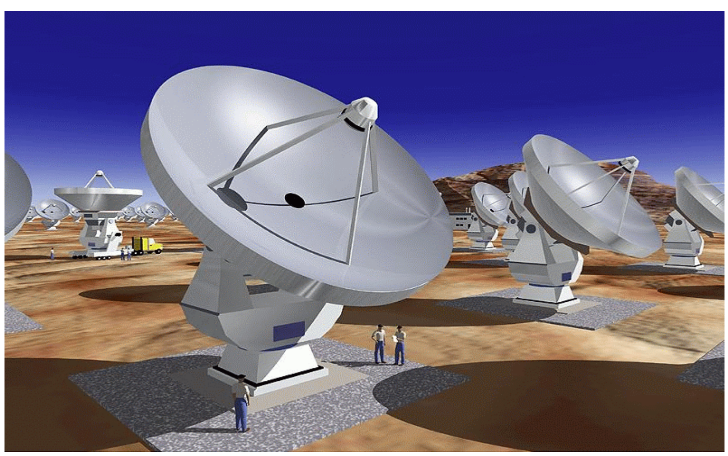
Author: A.W. Gunst Verified by: J.D. Bregman	Date of issue: March 28, 2002 Kind of issue: Public	Scope: Development Doc.nr.: Report-014 ASTRON-28000-R1
Responsible: C.M. de Vos Approved by: C.M. de Vos	Status: Final Revision nr.: 1.1	File: G:\data\tekkamer\28000\ 54Brepor.pdf
The ALMA FC study carried out at ASTRON is a contribution to the European ALMA Backend Electronics Team.		

Modeling the Effects of Re-quantization for ALMA-FC:

Cascading of Degradation Factors



Document history:

Revision	Date	Section	Page	Modification
0.1	November 14, 2001	-		Creation
0.2	January 28, 2002	-		Continued
0.3	February 22, 2002	Section 10		Corrections and section added
1.0	February 26, 2002	-		Released
1.1	March 28, 2002	-		Revised

Abstract

In the European hybride XF approach for the Atacama Large Millimeter Array correlator, the signals are split up in subbands before correlation. To split the signal in subbands filters are used. The filters generate additional bits. Because the size of the correlator increases with the square of the number of bits, a bit reduction before correlation is necessary. In this document the degradation of the system is determined when quantization before filtering and re-quantization after filtering is applied. From this is concluded that the total degradation of the system can be represented by the cascading of the degradation factors due to the quantizer and re-quantizer.

Because the input quantizer adds quantization noise, a scaling must be introduced before the re-quantizer. Also the scaling of the filtering stage must be incorporated. When doing so, the degradation of the quantizer approximates the degradation of the re-quantizer when the same amount of bits are used.

Furthermore it is concluded that the best strategy is, to use the same amount of quantization bits as re-quantization bits. Then, the performance increase is significant. Increasing the number of quantization bits and re-quantization from 2 to 3, results in a performance increase of 21 percent. From 3 to 4 bits results in an additional performance increase of 7 percent.

Rationale

This report describes the progress made in the determination of the total degradation factor of the system under consideration. To emphasize the progress made during the study this section was added.

The filters introduced in the European hybride XF approach extends the number of bits. Because the size of the processing system after filtering grows with the squared number of bits an additional re-quantization stage is required. Due to the re-quantization stage the degradation factor of the total system is increased.

To quantify the degradation factor of the total system first a theoretical first order estimation is performed for a quantizer. For the first order estimation the quantization noise is assumed as an added noise component. The estimation is extended for the degradation after filtering and re-quantization for wideband input noise and narrow band input noise signals. From both was concluded that the total degradation factor can be described by a cascading of the degradation factors of the separate quantizers.

To determine an estimate of the degradation factor from simulations a narrowband correlated noise signal plus a wideband uncorrelated noise signal was considered. When the correlation coefficient after quantization is plotted as function of the input correlation coefficient, the degradation factor can be determined from the slope of this curve. A poly fit function is used to estimate the slope in the zero point.

Using narrowband correlated noise in the passband results in a lower degradation factor after filtering, when the input signal is quantized. This is due to the fact that the added quantization noise is wideband and filtered in the stopband. Hence, the correlation coefficient after filtering increases and the degradation factor decreases. The theoretical improvement in degradation due to the filtering is calculated and is depedent on the filter characteristics.

The procedure of the determination of the degradation factor is dependent on the correlation coefficient. Therefore the degradation factors are determined for a number of correlation coefficients. From this was concluded that a correlation coefficient of 0.1 is appropriate for the rest of the simulations, because in that region the degradation factor remains nearly constant.

For the simulations the degradation factor at several stages was obtained when a correlation coefficient of 0.1 was used. When the amount of input bits was varied the degradation factor of the re-quantizer shows dependencies with the input quantizer. This was beyond the expectations.

To gain a better understanding of this effect the degradation was simulated as function of the number of re-quantization bits. When the insights in the quantization and re-quantization processes progressed it was found that the amount of degradation due to the re-quantizer was distorted because the power of the signal was increased due to the input quantizer. Taken this power into account reduces the dependency between the quantizer and re-quantizer.

According to the overall degradation it was found that the degradation was approximately the cascading of both. The optimal situation is to choose the number of input bits the same as the number of output bits.

The overall degradation factor is enhanced, when the quantizer and re-quantizer levels are chosen properly. For two input bits and two re-quantizer bits the degradation factor is reduced from 1.76 to 1.26 when more optimal levels are chosen. The error made in the degradation factor can be reduced by choosing a smaller band of correlated noise or by using more time samples. Furthermore a manual (0,0) point can be inserted to reduce any offset errors.

Finally the correction for variance increase due to input quantization is made in finite word length precision (8 bit). The simulation results approximates the results obtained without the finite precision compensation.

Contents

Introduction

In the European hybride XF approach for the Atacama Large Millimeter Array correlator a number of digital FIR filters will be used to split the wide band input signal in subbands prior to digital cross correlation [1]. The FIR filters increase the number of bits in the system. To cope with the correlator complexity the number of bits is reduced before the cross correlator. This will degrade the signal. It is important to determine the amount of degradation due to the re-quantization.

The degradation effects of digital cross correlation have been studied already for Gaussian signals clipped to 1 bit by [2] and two bit by [3, 4, 5]. However, the output of a digital filter operated in the few bit region is non Gaussian. A computer simulation of re-quantization after a FIR filter was already performed for the MMA [6]. In [6] a number of auto correlation spectra is presented to show the effects of re-quantization. However the cross correlation spectra and degradation factor due to re-quantization was not discussed.

The goal of the documents in the "Modeling the Effects of Re-quantization for ALMA-FC" series is to determine the spectra and degradation factor after correlation due to re-quantization. For this a simulation model is necessary which is discussed in [7]. In the model noise and Continuous Wave (CW) signals can be used. The effects in the spectral estimation after re-quantization is covered in [8]. In this document the degradation factor due to the re-quantization is discussed.

In the first section a first order estimation of the degradation factor after quantization is derived. This is extended to re-quantization in Section 2.1 and Section 2.2 for wideband noise and narrowband noise respectively. The degradation factor in the simulation model is defined at several stages. How the degradation factors are determined is covered in Section 3. The next section deals with the simulated degradation factor after filtering. The simulated degradation as function of input correlation coefficient, input quantization and re-quantization is discussed in respectively Section 5, Section 6 and Section 7. The overall degradation is covered in Section 8. The optimization of the quantizer and re-quantizer levels is outlined in Section 9. Finally in Section 10 the results with finite length computations are given.

1 Degradation After Quantization

In this section a first order estimation of the degradation factor after quantization is given. For this, the quantization noise is assumed to be uniformly distributed. This is approximated by setting the maximal range default from -4σ to 4σ. Furthermore the quantization noise between both channels of the correlator is assumed to be uncorrelated. This is valid only when the input correlation coefficient is also small. This applies to the signals used in radio astronomy where the system noise is normally much larger than the astronomical noise. The correlation coefficient measured by a continuous correlator is given by:

$$\rho_c = \frac{\sigma_{Ci}^2}{\sigma_{Ci}^2 + \sigma_{Ui}^2}, \tag{1}$$

where σ_{Ci}^2 is the correlated input noise power and σ_{Ui}^2 the uncorrelated input noise power.

When the noise signal is quantized, an uncorrelated quantization noise power of σ_{Qi}^2 is introduced (a correlated quantization noise part can be added in the numerator of the equation, but complicates the approximation). Hence, the correlation coefficient after quantization can be calculated using

$$\rho_d = \frac{\sigma_{Ci}^2}{\sigma_{Ci}^2 + \sigma_{Ui}^2 + \sigma_{Qi}^2}, \tag{2}$$

The quantization power for a n_i bit signed quantizer is defined as

$$\sigma_{Qi}^2 = \frac{q^2}{12}, \tag{3}$$

with a level increment

$$q = \frac{k \cdot g}{2^{n_i - 1}} \tag{4}$$

when a maximal range of $-k \cdot g$ to $k \cdot g$ is considered. The input noise power is g^2 [7] and can be written as

$$g^2 = \sigma_{Ci}^2 + \sigma_{Ui}^2 \tag{5}$$

Substituting Eq. (4) in Eq. (3) yields

$$\sigma_{Qi}^2 = \frac{k^2 \cdot g^2}{12 \cdot 2^{2n_i - 2}} = \frac{k^2 \cdot g^2}{3 \cdot 2^{2n_i}} \tag{6}$$

Substituting Eq. (6) in Eq. (2) yields

$$\rho_d = \frac{\sigma_{Ci}^2}{(\sigma_{Ci}^2 + \sigma_{Ui}^2) \cdot (1 + \frac{k^2}{3 \cdot 2^{2n_i}})} = \frac{\rho_c}{(1 + \frac{k^2}{3 \cdot 2^{2n_i}})} \tag{7}$$

n_i	$\frac{q}{g}$	g	k
2	0.995	1.005	1.99
3	0.585	1.709	2.34
4	0.335	2.985	2.680
5	0.184	5.435	2.944
6	0.104	9.615	3.328
7	0.056	17.86	3.584
8	0.030	33.33	3.840

Table 1: *Optimal power relation for 2 till 8 bit correlators.*

Definining the degradation factor as

$$D = \frac{\rho_c}{\rho_d} \tag{8}$$

results in an estimated degradation of

$$\hat{D} = 1 + \frac{k^2}{3 \cdot 2^{2n_i}} \tag{9}$$

using a quantizer range of $-k \cdot g$ to $k \cdot g$. Since, uniform quantization noise is assumed, the equations are not valid when the noise is clipping extremely. When k is chosen small, i.e. the maximal range of the quantizer is small, the quantizer clipps and the derived approximations are not valid anymore. However, when k is too large the levels of the quantizers are not at an optimum and therefore the degradation factor becomes larger. In [9] the optimum equi-distant levels by simulation are determined for 2 to 8 bit, when also clipping is included. From those results g can be calculated, when $q = 1$ is assumed. The value of k can be calculated using Eq. (4). The results as a function of the number of input bits are listed in Table 1.

In Table 2 the degradation factor obtained in [9] and using the first order estimation of Eq. (9) are derived when the optimum levels are used. The first order estimation of the degradation factor approximates the simulated one, obtained in [9]. However, it can be seen that the estimation is optimistic (clipping is not incorporated). Furthermore the results for $k = 1, 2, 3$ and 4 are also listed. The results found in [9] are approximated for $k = 3$. The estimated degradation factor as function of the number of input bits is depicted in Fig. 1 for k is 1, 2 and 3.

n_i	D [9]	$\hat{D}(k\ opt.)$	$\hat{D}(k=1)$	$\hat{D}(k=2)$	$\hat{D}(k=3)$	$\hat{D}(k=4)$
2	1.135	1.083	1.021	1.083	1.188	1.333
3	1.039	1.025	1.005	1.021	1.047	1.083
4	1.012	1.009	1.002	1.005	1.012	1.021
5	1.004	1.003	1.0003	1.001	1.003	1.005
6	1.001	1.001	1.0001	1.0003	1.0007	1.001

Table 2: *Estimation of degradation factor for a 2 to 6 bit correlator (for one bit the quantization noise cannot be assumed to be uniform).*

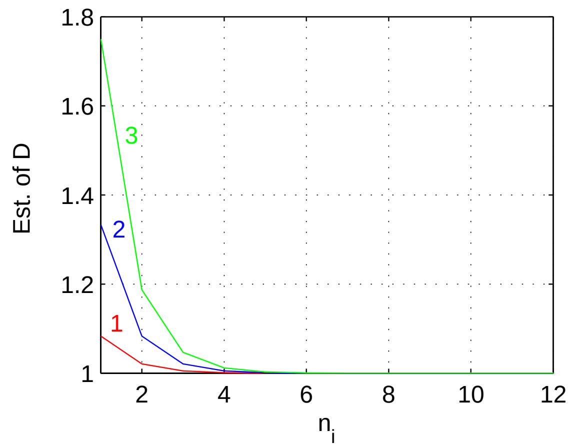


Figure 1: Estimated degradation factor for k is 1, 2 and 3.

2 Estimation of Degradation After Re-quantization

In the previous section the degradation factor is estimated when only quantizers are considered. In this section the degradation factor is estimated when the quantized signal is filtered and re-quantized. When the input noise signals are wideband, the filter will not change the degradation factor. This is discussed in Section 2.1. However, when the correlated noise part is within the passband of the filter only (narrowband noise), the degradation factor after filtering will decrease. This is caused by the quantization noise which is wideband and filtered in the stopband. Hence, the total amount of quantization noise will decrease after filtering and therefore the performance is improved. This is topic of Section 2.2.

2.1 Wideband Noise Signals

In this section wideband signals are assumed. According to [7] the noise power after filtering is the noise power before filtering multiplied with $\|\mathbf{c}\|_2$, with $\mathbf{c}$ the coefficient vector of the filter. The coefficients can be quantized as well. The correlated noise part after filtering σ^2_{Cf} can be written as

$$\sigma^2_{Cf} = \sigma^2_{Ci} \cdot \left(\|\mathbf{c}\|_2\right)^2 \tag{10}$$

The total noise power after filtering can be written as the sum of the correlated noise part σ^2_{Cf}, uncorrelated noise part σ^2_{Uf} and quantization noise part σ^2_{Qf}, all after filtering.

$$\sigma^2_{Cf} + \sigma^2_{Uf} + \sigma^2_{Qf} = \left(\sigma^2_{Ci} + \sigma^2_{Ui} + \sigma^2_{Qi}\right) \cdot \left(\|\mathbf{c}\|_2\right)^2 = \left(\sigma^2_{Ci} + \sigma^2_{Ui}\right) \cdot \left(1 + \frac{k^2}{3 \cdot 2^{2n_i}}\right) \cdot \left(\|\mathbf{c}\|_2\right)^2 \tag{11}$$

Hence the correlation coefficient after filtering is

$$\rho_f = \frac{\sigma^2_{Cf}}{\sigma^2_{Cf} + \sigma^2_{Uf} + \sigma^2_{Qf}} = \frac{\sigma^2_{Ci} \cdot \left(\|\mathbf{c}\|_2\right)^2}{\left(\sigma^2_{Ci} + \sigma^2_{Ui}\right) \cdot \left(1 + \frac{k^2}{3 \cdot 2^{2n_i}}\right) \cdot \left(\|\mathbf{c}\|_2\right)^2} = \rho_d \tag{12}$$

From this equation is seen that the degradation factor is not changed after filtering, as expected.

Just before the re-quantizer the noise is re-scaled with a scaler s_2 [7] which is defined as

$$s_2 = \frac{g_o}{2^{n_i + n_c + n_a - 2}}, \tag{13}$$

with n_i the number of input bits, n_c the number of coefficient bits and n_a the number of added bits due to the filtering. The scaler will not affect the correlation coefficient.

Re-quantizing the filtered result, adds again uncorrelated quantization noise. The correlated noise component after re-quantization σ^2_{Cr} will not be affected by the re-quantization stage (quantization noise is assumed uncorrelated)

$$\sigma^2_{Cr} = s_2^2 \cdot \sigma^2_{Cf} = s_2^2 \cdot \sigma^2_{Ci} \cdot \left(\|\mathbf{c}\|_2\right)^2 \tag{14}$$

The total noise power just before the re-quantization is

$$\sigma^2_t = s_2^2 \left(\sigma^2_{Cf} + \sigma^2_{Uf} + \sigma^2_{Qf}\right) \tag{15}$$

Since quantization adds noise, the total noise power after the re-quantization is

$$\sigma^2_t + \sigma^2_{Qr} = \left(\sigma^2_{Ci} + \sigma^2_{Ui}\right) \cdot \left(1 + \frac{k^2}{3 \cdot 2^{2n_i}}\right) \cdot s_2^2 \cdot \left(\|\mathbf{c}\|_2\right)^2 + \sigma^2_{Qr} \tag{16}$$

with σ_{Qr}^2 the quantization noise added by the re-quantizer. The quantization step q equals

$$q = \frac{k_2 \cdot \sigma_t}{2^{n_e - 1}} \tag{17}$$

with an output quantizer range of $-k_2 \cdot \sigma_t$ to $k_2 \cdot \sigma_t$. Hence the quantization noise is

$$\sigma_{Qr}^2 = \frac{q^2}{12} = \frac{k_2^2 \cdot \sigma_t^2}{3 \cdot 2^{2n_e}} \tag{18}$$

The correlation coefficient at the output of the re-quantizer is

$$\rho_r = \frac{\sigma_{Cr}^2}{\sigma_t^2 + \sigma_{Qr}^2} = \frac{\sigma_{Cr}^2}{\sigma_t^2 \cdot (1 + \frac{k_2^2}{3 \cdot 2^{2n_e}})} \tag{19}$$

Using Eq. (15), Eq. (13), Eq. (14) and Eq. (11) results in a final expression of

$$\begin{aligned}
\rho_r &= \frac{s_2^2 \cdot \sigma_{Ci}^2 \cdot (\|\mathbf{c}\|_2)^2}{s_2^2 \cdot \left((\sigma_{Ci}^2 + \sigma_{Ui}^2) \cdot (1 + \frac{k^2}{3 \cdot 2^{2n_i}}) \cdot (\|\mathbf{c}\|_2)^2\right) \cdot (1 + \frac{k_2^2}{3 \cdot 2^{2n_e}})} \\
&= \frac{\sigma_{Ci}^2}{(\sigma_{Ci}^2 + \sigma_{Ui}^2) \cdot (1 + \frac{k^2}{3 \cdot 2^{2n_i}}) \cdot (1 + \frac{k_2^2}{3 \cdot 2^{2n_e}})} \\
&= \frac{\rho_c}{(1 + \frac{k^2}{3 \cdot 2^{2n_i}}) \cdot (1 + \frac{k_2^2}{3 \cdot 2^{2n_e}})} \\
&= \frac{\rho_d}{(1 + \frac{k_2^2}{3 \cdot 2^{2n_e}})}
\end{aligned} \tag{20}$$

The total estimation of the degradation factor equals

$$\hat{D}_t = \frac{\rho_c}{\rho_r} = (1 + \frac{k^2}{3 \cdot 2^{2n_i}}) \cdot (1 + \frac{k_2^2}{3 \cdot 2^{2n_e}}) \tag{21}$$

Defining the degradation factor due to the first quantizer as $\hat{D}_1$ and due to the re-quantizer as $\hat{D}_2$ results in a total degradation factor of:

$$\hat{D}_t = \hat{D}_1 \cdot \hat{D}_2 \tag{22}$$

So, the total degradation factor is the cascading of the degradation factor due to the input quantizer and due to the re-quantizer.

2.2 Narrowband Noise Signals

In this section the noise is assumed to consist out of a narrowband correlated part, which is in the passband of the filter and a wideband uncorrelated part. Again, small correlation coefficients are assumed because the assumption of uncorrelated quantization noise has to hold. Using no quantization at all the correlation coefficient before filtering is

$$\rho_c = \frac{\sigma_{Ci}^2}{\sigma_{Ci}^2 + \sigma_{Ui}^2} \tag{23}$$

and after filtering

$$\rho_{cf} = \frac{\sigma_{Ci}^2}{\sigma_{Ci}^2 + \sigma_{Ui}^2 \cdot G} \tag{24}$$

with G the gain of the filter. Since the correlated noise part is within the passband, this part of the noise is not affected by the filtering operation (passband gain is assumed 1). When the first quantization stage before filtering is inserted the correlation coefficient becomes

$$\rho_d = \frac{\sigma_{Ci}^2}{(\sigma_{Ci}^2 + \sigma_{Ui}^2) \cdot (1+K)}, \tag{25}$$

with K a constant which is dependent on the scale of the quantizer (see Eq. (7)). Filtering this signal results in a correlation coefficient of

$$\rho_f = \frac{\sigma_{Ci}^2}{\sigma_{Ci}^2 \cdot (1+K) + \sigma_{Ui}^2 \cdot (1+K) \cdot G} \tag{26}$$

Adding a re-quantization stage results in a correlation coefficient of

$$\rho_f = \frac{\sigma_{Ci}^2}{(\sigma_{Ci}^2 \cdot (1+K) + \sigma_{Ui}^2 \cdot (1+K) \cdot G) \cdot (1+K_2)} = \frac{\sigma_{Ci}^2}{(\sigma_{Ci}^2 + \sigma_{Ui}^2 \cdot G) \cdot (1+K) \cdot (1+K_2)} \tag{27}$$

This can be rewritten as

$$\rho_f = \frac{\rho_{cf}}{(1+K) \cdot (1+K_2)} \tag{28}$$

and hence the total degradation factor is the cascading of the degradation factor of the quantizer stage and the re-quantizer stage again.

3 Determination of the Degradation Factor in the Simulation Model

In the previous sections, a first order estimation of the degradation factor was derived. In this section the degradation factor is determined from the simulation model [7]. In order to do so, the degradation factor is determined from the correlation function. In Fig. 2(a) the input correlation function is depicted. Fig. 2(b) shows the correlation function after quantization with one bit. The peak of the correlation function at lag zero shows the initial input correlation coefficient of 1 before quantization and after one bit quantization. The correlation coefficient is for most of the lags smaller than 0.1. To determine the degradation, for each τ point the ideal correlation coefficient is plotted against the correlation coefficient after quantization. This results in 513 points, which are plotted in Fig. 2(c). The points can be connected to each other with a line. From the correlation function Fig. 2(a) is seen that the correlation coefficient becomes not smaller than -0.25. Furthermore the sudden crack of ρ_c (region from approximately 0.9 to 1), is caused by the small peak in the correlation function. There are no intermediate points. When the main lobe was broader, more large values for the correlation coefficient are found. This is the reason that a lowpass filtered correlated noise band is used. The smaller the filter is, the broader the main lobe will be.

Since, the degradation for 1 bit quantization is well known, the simulated digital correlation coefficient can be compared with the correlation coefficient found from theory. For 1 bit the relation between the ideal correlation function $\rho_c(\tau)$ and digital correlation function $\rho_d(\tau)$ obeys the following relation [2]

$$\rho_d(\tau) = \frac{2}{\pi} \arcsin(\rho_c(\tau)) \tag{29}$$

The theoretical relation for 1 bit is depicted in Fig. 2(d). This figure approximates the points obtained from the simulation (Fig. 2(c)).

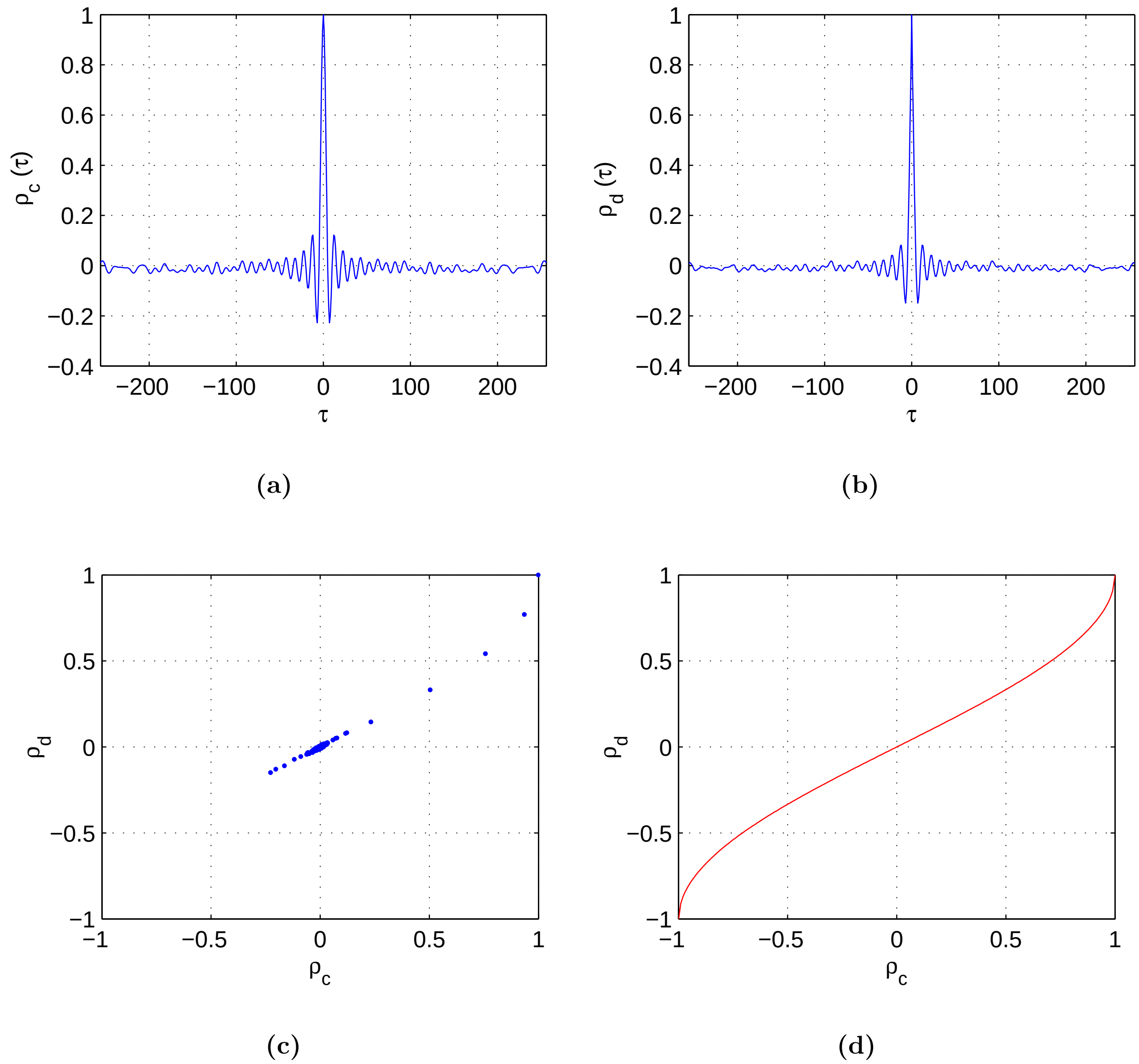


Figure 2: Input (a) and digital (b) correlation function as a function of τ for one bit and digital correlation coefficient as a function of input correlation coefficient for simulation (c) and from the Van Vleck function (d) ($g = 0.25$, ρ=0.5 and 64k time samples).

The degradation factor is determined from the slope of the graph wherein the degraded correlation coefficient is plotted against the ideal correlation coefficient, like Fig. 2(c). To determine the slope the first correlation points in the correlation function (Fig. 2(a)), from $\tau = 0$ to the first zero crossing point are used. In this way a single line is obtained with a limited number of points. The number of points taken depends on the bandwidth of the correlated noise part (determines first zero crossing) and the resolution in delay steps. Therefore a broad main lobe is desired, i.e. a small band limited correlated noise signal. A correlated noise band of $0.1f_s$ is chosen. The number of lags is set at a constant (512). A *polyfit* function of Matlab [10] is used to find a polynomial fit of the data obtained. The *polyfit* function finds the coefficients of a polynomial of degree n that fits the data in a least squares sense [11]. A degree n of 1 is chosen. Hence, the data is fitted to a straight line $y(x) = a + b \cdot x$. The goal is to find b with the *polyfit* function. Since, it is important for the polynomial fit to have data points already on a straight line, a small input correlation coefficient is used. However, for small correlation coefficients the simulated values are influenced relatively more by the variance. Therefore the correlation coefficient must not be too small. The input correlation coefficient is chosen in Section 5.

According to Eq. (55) in Appendix A the errors variance made in b, considering a normal distributed error of the data points with variance $\sigma^2_{\rho_{xy}}$, equals [11]

$$\sigma_b^2 = \frac{N_p \cdot \sigma^2_{\rho_{xy}}}{N_p \sum_{i=1}^{N_p} x_i^2 - \left(\sum_{i=1}^{N_p} x_i\right)^2} \tag{30}$$

with N_p the number of data points defined and

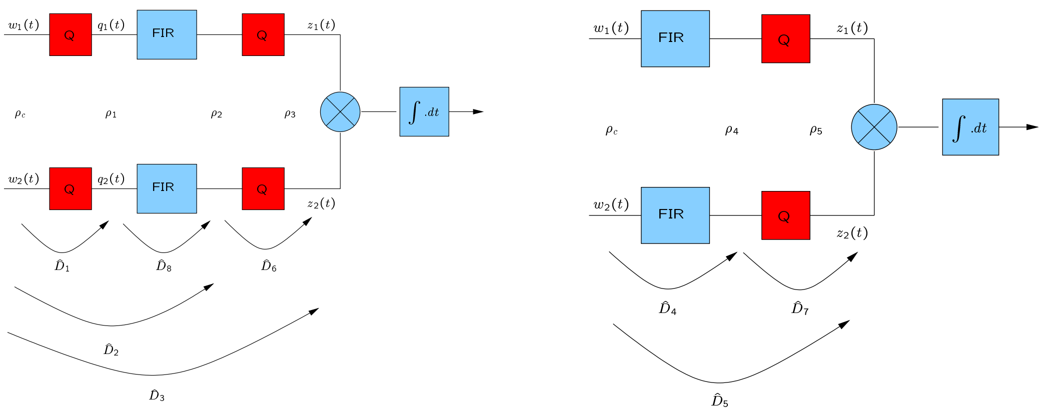
$$\sigma^2_{\rho_{xy}} = \frac{1+\rho_c^2}{N} \tag{31}$$

Hence, the variance of b can be rewritten as

$$\sigma_b^2 = \frac{N_p \cdot (1+\rho_c^2)}{N\left(\cdot N_p \sum_{i=1}^{N_p} x_i^2 - \left(\sum_{i=1}^{N_p} x_i\right)^2\right)}, \tag{32}$$

with N the number of time samples. For N = 64k, N_p=5, ρ_c = 0.5 and the datapoints $\{0.50, 0.47, 0.38, 0.25, 0.12\}$ the variance of coefficient b and hence the error in the degradation factor is approximately $3 \cdot 10^{-5}$. The standard deviation equals approximately $6 \cdot 10^{-3}$. Considering a maximal error of $4 \cdot \sigma$ results in a maximum error of 0.02 (probability of errors larger is in the order of 10^{-6}). When a correlation coefficient of 0.1 is used (with datapoints $\{0.095, 0.093, 0.076, 0.048, 0.026\}$) the variance in the degradation factor is approximately $9 \cdot 10^{-4}$, while the maximum error for a $4 \cdot \sigma$ boundary is 0.12.

The correlation function is obtained at each stage of the system, to understand the correlation coefficient degradation throughout the system. This is done with and without input quantization. The correlation coefficients are defined in Fig. 3.



(a) **(b)**

Figure 3: Definition of several correlation coefficients at different stages of the simulation model with input quantizer (a) and without input quantizer (b).

From the correlation coefficients a number of degradation factors can be defined:

$$\begin{aligned}
D_1 &= \frac{\rho_c}{\rho_1} \\
D_2 &= \frac{\rho_c}{\rho_2} \\
D_3 &= \frac{\rho_c}{\rho_3} \\
D_4 &= \frac{\rho_c}{\rho_4} \\
D_5 &= \frac{\rho_c}{\rho_5} \\
D_6 &= \frac{\rho_2}{\rho_3} \\
D_7 &= \frac{\rho_4}{\rho_5} \\
D_8 &= \frac{\rho_1}{\rho_2}
\end{aligned} \tag{33}$$

The end to end degradation factor of the system is D_3 and according to Section 2.2 this can be calculated with

$$D_3 = D_1 \cdot D_8 \cdot D_6 \tag{34}$$

4 Estimation of Degradation After Filtering

Depending on the spectral distribution of correlated and uncorrelated noise, spectral filtering does change the degradation factor observed in the output signal of the filter. Normally the observed noise is wideband and for that case the spectral filtering does not change the degradation factor. However in this document band limited noise was assumed. The degradation factor after filtering is derived in this section.

Recalling the equation for calculating the correlation coefficient

$$\rho_c = \frac{\sigma_{Ci}^2}{\sigma_{Ci}^2 + \sigma_{Ui}^2}, \tag{35}$$

with σ_{Ci}^2 the correlated noise power and σ_{Ui}^2 the uncorrelated noise power and re-arranging this, yields for the uncorrelated part

$$\sigma_{Ui}^2 = \frac{\sigma_{Ci}^2 \cdot (1 - \rho_c)}{\rho_c} \tag{36}$$

When the correlated noise part is in the passband, the correlation coefficient after filtering becomes

$$\rho_f = \frac{\sigma_{Ci}^2}{\sigma_{Ci}^2 + G \cdot \sigma_{Ui}^2}, \tag{37}$$

with G a filter dependent gain of wideband noise signals, which can be written as

$$G = \left(1 - \frac{f_c}{0.5f_s}\right) \cdot \delta_2 + \frac{f_c}{0.5f_s} \cdot \delta_1 \tag{38}$$

with f_s the sample frequency, f_c the cut off frequency, δ_1 the gain in the passband (assumed 1) and δ_2 the gain in the stopband. Substituting Eq. (36) in Eq. (37) yields

$$\begin{aligned} \rho_f &= \sigma_{Ci}^2 \cdot \frac{1}{1 + G \cdot \frac{1-\rho_c}{\rho_c}} \\ &= \frac{\rho_c}{\rho_c + G \cdot (1 - \rho_c)} \end{aligned} \tag{39}$$

Using a low pass filter with $f_c = 0.25f_s$ results in a G of $\frac{\delta_2}{2} + \frac{1}{2}$. Hence the correlation after filtering is

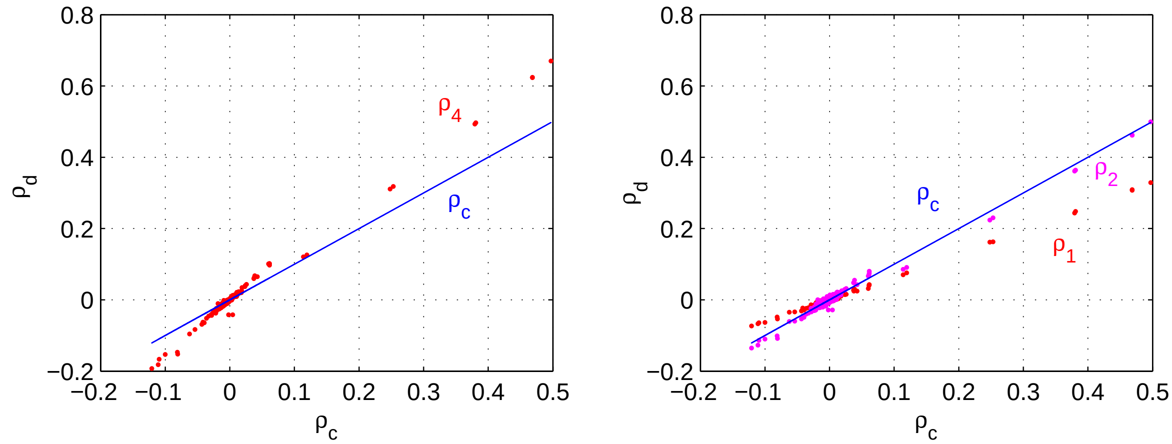
$$\rho_f = \frac{\rho_c}{\rho_c + \left(\frac{\delta_2}{2} + \frac{1}{2}\right) \cdot (1 - \rho_c)} \tag{40}$$

Assuming a correlation coefficient of 0.5, a low pass filter with a suppression of 20 dB results in a correlation coefficient after filtering of approximately 0.66. The simulation gave a correlation coefficient of 0.67 after filtering. The filtered correlation coefficient ρ_4 as function of the continuous correlation coefficient is drawn in Fig. 4(a). A couple of deteriorated small points can be notified. Not all points are concentrated in the origin, when the input correlation coefficient is zero. However, most of the points are located at the origin, when the input correlation coefficient is zero. The reason for the deteriorated points is the filter. Due to the filter the correlation function is changed and also the zero crossings are affected. Therefore for an input correlation coefficient of zero, the output correlation coefficient is not zero. To compare the result of a quantized signal after filtering, the reference signal must be also after filtering. From this two additional degradation factors can be defined:

$$
\begin{aligned}
D_9 &= \frac{\rho_4}{\rho_2} \\
D_{10} &= \frac{\rho_4}{\rho_3}
\end{aligned}
\tag{41}
$$

It is expected that D_9 is approximately D_1 and D_{10} is approximately $D_1 \cdot D_6$.

The correlation coefficient after quantizing the signal with 1 bit is depicted in Fig. 4(b). This line approximates the van Vleck curve (indicated as ρ_1, the dots are not connected). For a maximum input correlation coefficient of 0.5, a digital correlation coefficient of 0.33 is obtained. Filtering the quantized signal increases the correlation coefficient as shown by the ρ_2 dots. After filtering the correlation line approximates the input correlation. Again this is caused by the correlated noise, which is in the passband. In practice the correlated noise will be wide band as well and the correlation coefficient is not influenced by the filter operation. For clarity also the ideal correlation coefficient ρ_c is drawn.



(a) **(b)**

Figure 4: The filtered correlation coefficient ρ_4 (a) and quantized (filtered) correlation coefficients (b) as a function of the input correlation coefficient (n_i=1, n_e=1, $\rho = 0.5$, $g = 1$ and the number of time samples is 64k).

5 Degradation as Function of Correlation Coefficient

In Section 3 was mentioned that the input correlation coefficient must not be too large and not too small. Therefore in this section the degradation factor dependency on the correlation coefficient is determined for one input bit and three output bits. In this way a correlation coefficient region can be determined, where the degradation factor is more or less constant. The degradation factors are determined for a number of input correlation coefficients. The results are listed in Table 3. Three regions can be distinguished. In the upper region ($\rho_c > 0.2$) the correlation coefficient line is not linear. Degradation factor D_1 is decreasing because the derivative of the line is decreasing for that region. For the lower region the variance of the correlation coefficient measurements is relative large compared with the measured correlation coefficients. Here, degradation factor D_1 and D_6 are far from constant. In the center region both degradation factors are constant, because the relation is almost linear. Therefore, the results in the remainder of this document are discussed for a correlation coefficient of 0.1. The degradation factors D_9 and D_{10} are determined by a correlation coefficient ρ_f which is larger than the input correlation coefficient. Here less constant values can be observed. In Section 2.1 was derived that the total degradation factor can be approximated as a cascading of the degradation factors of each operation. So, the following equations must hold:

$$D_3 = D_1 \cdot D_8 \cdot D_6 \tag{42}$$

ρ_c	ρ_f	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}
0.025	0.048	1.66	0.94	1.08	0.52	0.76	1.15	1.39	0.54	1.78	2.03
0.05	0.09	1.57	0.79	0.84	0.48	0.63	1.06	1.29	0.51	1.64	1.75
0.1	0.18	1.60	0.79	0.82	0.50	0.60	1.03	1.18	0.50	1.57	1.62
0.2	0.33	1.60	0.83	0.86	0.55	0.62	1.03	1.12	0.52	1.50	1.55
0.3	0.46	1.58	0.87	0.91	0.60	0.66	1.04	1.10	0.55	1.45	1.51
0.4	0.57	1.54	0.90	0.93	0.65	0.71	1.03	1.10	0.59	1.39	1.43
0.5	0.66	1.50	0.93	0.96	0.70	0.76	1.03	1.09	0.62	1.34	1.37

Table 3: *Degradation factors as function of correlation coefficient for $n_i = 1$ and $n_e = 3$.*

and

$$D_5 = D_4 \cdot D_7 \tag{43}$$

Concentrating on the results for a correlation coefficient of 0.1 results in

$$D_1 \cdot D_8 \cdot D_6 = 1.60 \cdot 0.50 \cdot 1.03 = 0.82 \sim D_3 \tag{44}$$

and

$$D_4 \cdot D_7 = 0.5 \cdot 1.18 = 0.59 \sim D_5 \tag{45}$$

Furthermore D_4 approximates D_8, while the input correlation coefficient for both situations differ. D_6 does not approximate D_7 for reasons that will become clear in Section 7. For an input correlation coefficient of 0.1, a ρ_2 of 0.12 and a ρ_4 of 0.18 was measured. For larger correlation coefficients the slope is larger and hence the degradation factor is smaller. This effect is also shown in Table 3 when D_1 is compared for all correlation coefficients. Increasing the input correlation coefficient results in a decrement of the degradation factor.

From Section 4 was expected that D_9 is approximately D_1 and D_{10} is approximately $D_1 \cdot D_6$. For a correlation coefficient of 0.1, D_1 equals 1.60 and D_9 equals 1.57. Finally D_{10} is 1.62 and $D_1 \cdot D_6$=1.65.

6 Degradation as Function of Input Quantization

In this section an overview is given of the simulation results when the output re-quantizer is set to a constant of 3 bits and the number of input bits is varied. The results for the degradation factor D_1 can be compared with the results found from the estimated degradation factors in Section 1,

n_i	$\hat{D}_1(k\ opt.)$	$D_1(k\ opt.)$	$\hat{D}_1(k=4)$	$D_1(k=4)$
2	1.083	1.132	1.333	1.306
3	1.025	1.063	1.083	1.081
4	1.009	1.033	1.021	1.013
5	1.003	0.998	1.005	1.002
6	1.0009	1.002	1.001	0.999

Table 4: *Estimated and simulated degradation factor after first quantization stage for a 2 to 6 bit correlator, when k is chosen optimal and more conservative (4).*

considering quantization only. For this a conservative value for k of 4 is chosen. Furthermore a more optimal value for k is used. The estimated results together with the simulation results are listed in Table 4 for 2 to 6 input bits. The table shows that the estimation is more valid for a small number of bits. For these cases the degradation factor is relative large and robust for small deviations.

Table 5 shows the simulated degradation factors as a function of n_i. In this table is shown that the degradation factor D_8 is almost constant when the number of input bits is varied (variation < 0.01). The small changes are caused by the variation in correlation coefficient (see Table 3). The correlation coefficient after the first quantization stage is dependent on the number of input bits. The correlation coefficient after 1 bit quantization is smaller due to the large uncorrelated quantization noise added.

Interesting to note is that for some reason the degradation factor D_6 is influenced by the number of input bits. This was not expected from the first order estimations derived in Section 1. When no input quantizer is present, the described effect is not notified (D_7 remains constant). So, the combination of input quantization and filtering causes a not predicted increase in the degradation factor of the re-quantizer to a certain limit. Now, the question arises how much the variation is when a different number of output bits are chosen. This is addressed in the next section.

Furthermore Table 3 shows that $D_1 \approx D_9$ and $D_1 \cdot D_6 \approx D_{10}$.

7 Degradation as Function of Re-quantization

In the previous section was notified that the degradation factor of the re-quantizer is influenced by the input quantizer. Therefore the results of D_6 are measured when n_i and n_e are varied from 1 to 8 bits. In Fig. 5(a) the degradation factor of the re-quantization stage D_6 is plotted against the number of input bits (from 1 to 12). For a small number of input bits the degradation

n_i	D_1	D_2	D_3	D_6	D_8	D_9	D_{10}	$D_1 \cdot D_6$
1	1.596	0.790	0.816	1.032	0.501	1.57	1.62	1.65
2	1.306	0.638	0.685	1.072	0.494	1.27	1.36	1.40
3	1.081	0.533	0.592	1.109	0.497	1.06	1.18	1.20
4	1.013	0.511	0.587	1.150	0.504	1.01	1.17	1.16
5	1.002	0.504	0.580	1.149	0.503	1.00	1.15	1.15
6	0.999	0.502	0.590	1.173	0.503	1.00	1.17	1.17
7	1.000	0.503	0.593	1.178	0.503	1.00	1.18	1.18
8	1.001	0.504	0.598	1.187	0.503	1.00	1.19	1.19

Table 5: *Simulated degradation factors for a 1 to 8 bit correlator when k=4 and* $n_e = 3$ *bit.*

factor increases significantly when the number of input bits is increased. The value converges to a constant level, i.e. the level when no input quantization is used (D_7). From the figure is seen that the degradation factor D_6 is worse than D_1 given the same amount of bits.

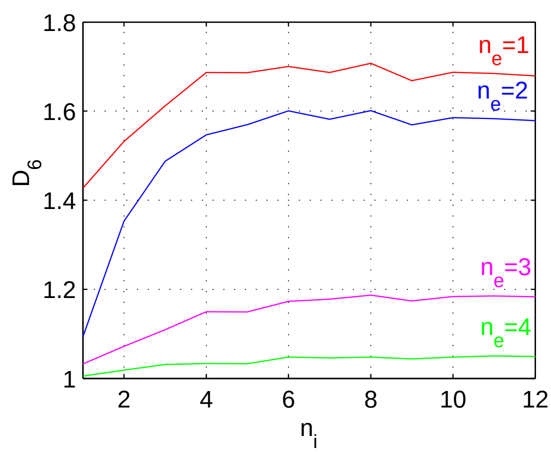


Figure 5: Degradation factor D_6 as function of n_i for n_e is 1, 2, 3 and 4 (64k time samples, 512 lags used and an input correlation coefficient of 0.1).

The degradation factor D_6 is larger than expected for two reasons. First the gain of the filter is compensated with the number of coefficient bits and number of generated bits due to the addition in the FIR filter. The real gain of the filter is determined by $\|\mathbf{c}\|_2$ and is less (added bits are determined with a ceiling operation, see [7]). Compensating with the number of bits leads to more suppression and therefore the levels of the re-quantizers are not put on the same positions as the input quantizer. The re-quantizer levels are more conservative in this way, therefore the degradation factor will be larger. However, g_o in the second scaler s_2 (Eq.(13)) can be set in such a way that the result is compensated for the real gain in the filter [1]. Additionaly g_o can also set the levels of the re-quantizer to a different setting than the input quantizer. For clarity a new s_2 is defined:

$$s_{2n} = \frac{g_o}{\|\mathbf{c}\|_2 \cdot g} \tag{46}$$

When the re-quantizer is set to the same level as the input quantizer and the number of bits of both is the same, the g terms are cancelled. Taking quantizer ranges of -4σ to 4σ results for noise only in a g of 2^{n_i-3} and a g_o of 2^{n_e-3}.

Using a better compensation for the filter gain, results in more optimal degradation factors. However, the dependency of the input quantizer is still not cancelled completely. This is due to the extra quantization noise which is added in the input quantization stage. In this way the variance is increased and the optimal scaling has to be modified, dependent on the number of input bits. For this purpose g_o can be used again. Doing so, results in a re-quantization which is more or less independent of the input quantizer. For this case g_o must be made dependent on the variance of the noise after filtering [2].

The differences in D_6 when traditional scaling, optimal filter scaling and additionally input variance compensation scaling is used is listed in respectively Table 6, Table 7 and Table 8. From the tables can be seen that the results are not changed for 1 bit (the transistion level is at zero and not dependent on scaling). Table 6 shows a larger degradation factor as Table 7 for a small number of input bits, because the levels are more conservative. Correction for the added variance due to the input quantization deteriorates the degradation factor for a small number of input bits. This is due to the non optimal scaling which is used. A rather conservative range of -4σ to 4σ was chosen. When, there is no correction for the extra variance the scaling is larger and hence the range is taken smaller, i.e. there is more clipping. This results in a more optimal degradation factor. In Section 9 the optimal levels are exploited. In Table 8 a more constant D_6 is shown, when the number of input bits vary (except for 1 bit). When for the last situation n_e is varied from 1 to 4 and n_i is set to 3 a degradation factor D_6 of respectively 1.61, 1.37, 1.09 and 1.01 is measured. If now n_i is varied from 1 to 4 and n_e is set to 3 a degradation factor D_1 of respectively 1.60, 1.31, 1.08 and 1.01 is measured. Both are approximating each other and therefore the quantization of the input quantizer and re-quantizer can be treated separately and are independent of each other in the first order. The behavior of the re-quantizer is the same as for the quantizer.

[1] This is not possible when using finite precision computations. See Section 10

[2] This is not possible when using finite precision computations. See Section 10

	n_e			
n_i	1	2	3	4
1	1.43	1.09	1.03	1.01
2	1.53	1.35	1.07	1.02
3	1.61	1.49	1.11	1.03
4	1.69	1.55	1.15	1.03

Table 6: *Degradation factor D_6 when the traditional scaling (Eq.(13)) is used as function of n_i and n_e.*

	n_e			
n_i	1	2	3	4
1	1.43	1.06	1.00	1.00
2	1.53	1.28	1.04	1.01
3	1.61	1.32	1.08	1.02
4	1.69	1.38	1.08	1.03

Table 7: *Degradation factor D_6 when the real scaling of the filter is incorporated (Eq.(46)) as function of n_i and n_e.*

	n_e			
n_i	1	2	3	4
1	1.43	1.42	1.10	1.03
2	1.53	1.35	1.09	1.03
3	1.61	1.37	1.09	1.01
4	1.69	1.41	1.10	1.04

Table 8: *Degradation factor D_6 when the real scaling of the filter is used and the variance is compensated for the added input quantization noise as function of n_i and n_e.*

	n_e							
n_i	1	2	3	4	5	6	7	8
1	2.28	2.26	1.75	1.64	1.61	1.61	1.60	1.60
2	2.00	1.76	1.43	1.35	1.32	1.31	1.31	1.31
3	1.74	1.48	1.18	1.10	1.09	1.08	1.08	1.08
4	1.71	1.43	1.11	1.05	1.02	1.01	1.01	1.01
5	1.69	1.44	1.10	1.03	1.01	1.01	1.00	1.00
6	1.70	1.43	1.10	1.03	1.01	1.00	1.00	1.00
7	1.69	1.43	1.10	1.03	1.01	1.00	1.00	1.00
8	1.71	1.44	1.10	1.03	1.01	1.01	1.00	1.00

Table 9: *Total degradation factor $D_1 \cdot D_6$ ($\approx D_{10}$) as a function of the number of input bits and output bits.*

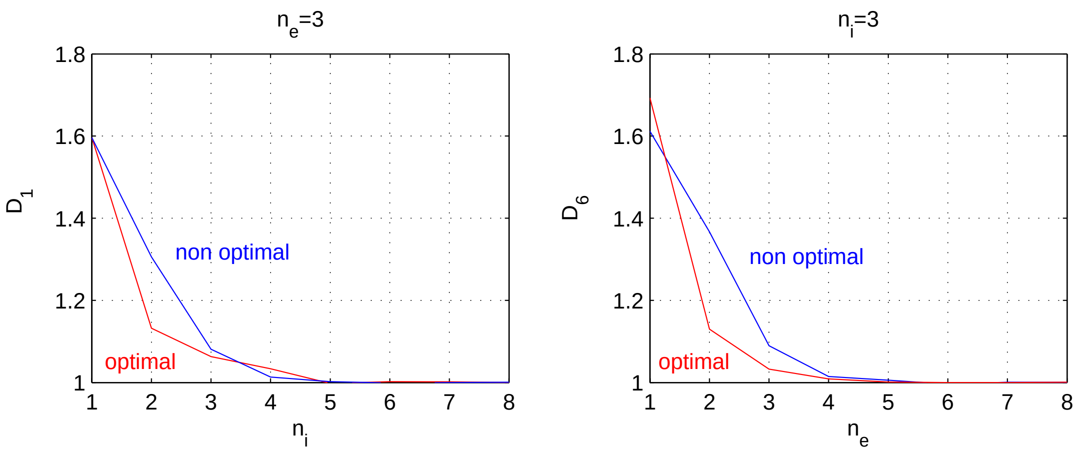
8 Overall Degradation

In this section the overall degradation factor $D_1 \cdot D_6$ is discussed when the number of input bits and number of output bits are varied from 1 to 8. For this the variance after filtering is measured and the signal is set for both quantizers to the -4σ to 4σ range. The results of the total degradation of the system $D_1 \cdot D_6$ are listed as function of n_i and n_e in Table 9. Because the effects of quantization and re-quantization are more or less the same, a symmetric table was expected. This is approximately true. However, the amount of re-quantizer bits tends to have more influence than the amount of input bits, i.e. it is better to have more re-quantizer bits than input bits (if $n_i > 1$ and $n_e > 1$). The differences are minimal, but never the other way around. This can be explained by taking the filter into consideration. Half of the input quantization noise is filtered by the filter. Because the correlated noise signal is in the passband, nothing of the desired correlated noise is filtered and therefore the correlation coefficient is increased due to the filtering. This leads to an decrease in degradation factor. For wideband signals this does not apply. According to Table 9 the best strategy is to choose the number of input bits the same as the number of re-quantization bits. The degradation factor is still quite high. Therefore Section 9 determines the degradation factor for more optimal equi-distant levels.

9 Optimal Quantization and Re-quantization Levels

In the previous sections a conservative quantizer range was chosen. In this section the quantizer range is chosen more optimal. The levels however, remain equi-distant. The scaling is based on the results found in [12]. The value of g and k is listed in Table 1 as a function of n_i. The input quantizer and output quantizer (with quantities g_0 and k_2) are both scaled according to Table 1. Furthermore the scaling of the re-quantizer is compensated for the variance increase due to the input quantizer, after filtering. Eq. (46) is used for the filter scaling.

The degradation factor D_1 and D_6 as a function of the number of input bits and output bits for the non-optimal and optimal case are shown in Fig. 6(a) and Fig. 6(b) respectively. The optimal degradation factor found after input quantization approximates the one found in the literature.



(a) **(b)**

Figure 6: The optimal and non-optimal degradation factor D_1 as function of the number of input bits (a) and the optimal and non-optimal degradation factor D_6 as function of the number of output bits (b) ($g = 1$, $n_c = 16$, 512 lags and 64k time samples).

In Table 10 degradation factor D_1 is listed as function of the number of input bits and output bits. From the table can be seen that degradation factor D_1 is independent of the number of output bits. From Table 11 can be seen that the degradation factor of the re-quantizer is dependent on the number of bits of the quantizer. This is probably caused by the correlation coefficient which is changed dependent on the number of input bits. Another effect can be the number of time samples and the procedure to determine the degradation factor.

The error can be reduced by taking more time samples, a smaller band of correlated noise (results in more data points for the data fitting) or by inserting the zero-zero point manually in the correlation coefficient graph. The last option can be done to force a line through the origin. This reduces the variance of the error slightly (number of data points is increased from 5 to 6). Doing so, results in a more predictable value for D_6 when n_e is 1 bit and n_i is varied. The results for degradation factor D_6 are listed in Table 12 for n_i and n_e ranging from 1 to 4 bit. From now on

	n_e							
n_i	1	2	3	4	5	6	7	8
1	1.60	1.60	1.60	1.60	1.60	1.60	1.60	1.60
2	1.13	1.13	1.13	1.13	1.13	1.13	1.13	1.13
3	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06
4	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03
5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
7	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
8	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 10: *Degradation factor D_1 as a function of the number of input bits and output bits when optimal equi-distant levels are used.*

	n_e							
n_i	1	2	3	4	5	6	7	8
1	1.43	1.09	1.04	1.01	1.01	1.00	1.00	1.00
2	1.50	1.09	1.03	1.01	1.00	1.00	1.00	1.00
3	1.69	1.13	1.03	1.01	1.00	1.00	1.00	1.00
4	1.67	1.13	1.04	1.01	1.00	1.00	1.00	1.00
5	1.60	1.12	1.04	1.01	1.00	1.00	1.00	1.00
6	1.69	1.15	1.04	1.01	1.00	1.00	1.00	1.00
7	1.68	1.14	1.04	1.02	1.00	1.00	1.00	1.00
8	1.67	1.14	1.03	1.02	1.00	1.00	1.00	1.00

Table 11: *Degradation factor D_6 as a function of the number of input bits and output bits when optimal equi-distant levels are used.*

	n_e			
n_i	1	2	3	4
1	1.48	1.07	1.03	1.01
2	1.50	1.11	1.03	1.01
3	1.63	1.13	1.04	1.01
4	1.60	1.13	1.04	1.01
5	1.57	1.12	1.04	1.01
6	1.61	1.14	1.04	1.01
7	1.61	1.14	1.04	1.01
8	1.61	1.14	1.04	1.02

Table 12: *Degradation factor D_6 as a function of the number of input bits and output bits when optimal equi-distant levels are used (64k time samples used). Now the (0,0) point is inserted in the correlation coefficient graph, to be used for the polynomial fit.*

	n_e			
n_i	1	2	3	4
1	1.50	1.07	1.03	1.01
2	1.52	1.11	1.03	1.01
3	1.54	1.13	1.04	1.01
4	1.55	1.13	1.04	1.01
5	1.56	1.13	1.04	1.01
6	1.56	1.13	1.04	1.01
7	1.56	1.13	1.04	1.01
8	1.55	1.13	1.04	1.01

Table 13: *Degradation factor D_6 as a function of the number of input bits and output bits when optimal equi-distant levels are used (1M time samples used).*

all the results listed are included with the (0,0) data point.

To reduce the variance of the error, in Table 13 the degradation factor D_6 is listed when 1M time samples are used. From this table can be seen that the results of D_6 is more independent of n_i. The degradation factor D_6 has a larger variance in the error for a low amount of input bits than for a larger amount of input bits, because the correlation coefficient after quantizing is smaller for a low amount of bits. This can explain the variation for a low amount of bits. Assuming a correlation coefficient of 0.1 before filtering and 0.18 after filtering, results for 64k samples, 6 data points in an error of 0.11 and 0.06 respectively (assuming a $4 \cdot \sigma$ bound). So, when the correlation coefficient is smaller the error is larger. Using 1M samples results in errors of maximal 0.03 and 0.01.

	n_e							
n_i	1	2	3	4	5	6	7	8
1	2.32	1.66	1.60	1.56	1.56	1.55	1.55	1.55
2	1.72	1.26	1.17	1.14	1.13	1.13	1.13	1.13
3	1.60	1.17	1.08	1.05	1.04	1.04	1.04	1.04
4	1.57	1.15	1.05	1.02	1.01	1.01	1.01	1.01
5	1.56	1.14	1.04	1.01	1.01	1.00	1.00	1.00
6	1.56	1.13	1.04	1.01	1.00	1.00	1.00	1.00
7	1.56	1.13	1.04	1.01	1.00	1.00	1.00	1.00
8	1.55	1.13	1.04	1.01	1.00	1.00	1.00	1.00

Table 14: *Degradation factor $D_1 \cdot D_6$ as a function of the number of input bits and output bits when optimal equi-distant levels are used (1M time samples used).*

	n_e							
n_i	1	2	3	4	5	6	7	8
1	2.24	1.60	1.54	1.51	1.50	1.50	1.50	1.50
2	1.70	1.24	1.15	1.12	1.12	1.12	1.12	1.12
3	1.59	1.17	1.07	1.05	1.04	1.04	1.04	1.04
4	1.56	1.15	1.05	1.02	1.01	1.01	1.01	1.01
5	1.56	1.14	1.04	1.01	1.01	1.00	1.00	1.00
6	1.56	1.13	1.04	1.01	1.00	1.00	1.00	1.00
7	1.56	1.13	1.04	1.01	1.00	1.00	1.00	1.00
8	1.55	1.13	1.04	1.01	1.00	1.00	1.00	1.00

Table 15: *Degradation factor D_{10} as a function of the number of input bits and output bits when optimal equi-distant levels are used (1M time samples used).*

For the last situation Table 14 shows the total degradation factor $D_1 \cdot D_6$ and Table 15 shows the total degradation factor D_{10}. Degradation factor D_{10} is smaller because the correlation coefficient after filtering is larger (0.18 in stead of 0.1). This statement was verified by a simulation with an input correlation coefficient of 0.18.

From the results in Table 15 can be seen that the performance increase of a 2 bit quantizer and re-quantizer to 3 bit is more significant (1.24-1.07=0.17) than from 3 to 4 (1.07-1.02=0.05). This can be translated in a collective area reduction of 17 respectively 5 percent. Furthermore it can be seen that it is more profitable to use more re-quantization bits than quantization bits, than the other way around. This is caused by the filtering and the assumption of narrowband correlated noise. The best strategy is to use an equal amount of quantizer and re-quantizer bits.

More optimal degradation factors can be found when the levels are also chosen not equi-distant [4, 5]. This is not covered in this document. From the results obtained can be concluded that the total degradation factor using non-equidistant levels can be calculated from the results of the single quantizers with non-equidistant levels.

10 Finite Length Computations

In Section 7 two additional modifications to the model were made in order to make the re-quantizer degradation independent of the quantizer degradation:

1. correction for the real filter scaling;
2. correction for the variance increase due to the input quantization noise.

Both corrections cannot be implemented straightforward when finite length computations are used. The first correction cannot be used at all (the filter scaling is already corrected with s_2). For the second correction the noise power after filtering is measured. The power is related to the expected power when no input quantizer is present. This relation is multiplied with the signal. For finite precision the relation must be quantized as well. Also the power measurement must be implemented. Another strategy is to use an estimation, based on the assumption of uniform noise. Considering a quantization step q of 1 the correction becomes:

$$\frac{g}{g + \frac{1}{\sqrt{12}}} \tag{47}$$

Quantizing this relation with 8 bit and using no additional filter correction results in approximately the same degradation factors as obtained before. For the sake of completeness the results of the total degradation factor $D_1 \cdot D_6$ for 1M samples is given in Table 16.

	n_e							
n_i	1	2	3	4	5	6	7	8
1	2.32	1.81	1.63	1.58	1.56	1.55	1.55	1.55
2	1.72	1.30	1.19	1.15	1.14	1.13	1.13	1.13
3	1.60	1.19	1.09	1.06	1.05	1.04	1.04	1.04
4	1.57	1.15	1.06	1.02	1.02	1.01	1.01	1.01
5	1.56	1.15	1.05	1.02	1.01	1.00	1.00	1.00
6	1.56	1.15	1.05	1.01	1.01	1.00	1.00	1.00
7	1.56	1.15	1.05	1.01	1.01	1.00	1.00	1.00
8	1.55	1.15	1.05	1.01	1.00	1.00	1.00	1.00

Table 16: *Degradation factor $D_1 \cdot D_6$ as a function of the number of input bits and output bits when optimal equi-distant levels are used and finite length computations are considered (1M time samples used).*

Conclusions and Recommendations

In this document a first order estimation of the total degradation factor of a system consisting out of a quantizer, filter and a re-quantizer is given. From this is concluded that the degradation of the total system can be approximated by the cascading of the degradation factor of the quantizer and degradation factor of the re-quantizer. Because the input quantizer adds quantization noise, a scaling must be introduced before the re-quantizer. Also the scaling of the filtering stage must be incorporated. When doing so, the same optimal levels of the re-quantizer can be set. In this way it was shown that the operation of the re-quantizer was not dependent on the quantizer in first order. Furthermore a first order estimation was done, representing the quantization noise as an additive noise source. The results of the estimation approximates the simulation.

From the results appeared that the best strategy is to use the same amount as input bits as re-quantization bits. Just like a single quantizer the performance increase for a low amount of bits is much higher than for a large amount of bits. For a single quantizer, the performance increase from 2 to 3 bit is about 10 percent. Going from a 2 bit quantizer and re-quantizer to a 3 bit quantizer and re-quantizer in the simulation model results in a performance increase of 21 percent. And from 3 to 4 bit in a performance increase of 7 percent. These numbers can be translated directly in an increase of collective area.

For further study it is recommended to address the effect of optimal quantization within a FIR filter implementation. This will complicate the simulation model even more and increase the simulation time required. Furthermore non equi-distant quantizer and re-quantizer levels can be incorporated. Finally a correction can be estimated, which reduces the effects of quantization and re-quantization.

References

[1] *Proceedings of the Second Meeting of the European ALMA Backend Electronics Team*, February 2001.

[2] J. H. Van Vleck and D. Middleton. The spectrum of clipped noise. *Proceedings of IEEE*, 54(1):2–19, 1966.

[3] J. B. Hagen and D. T. Farley. Digital-correlation techniques in radio science. *Radio Science*, 8(8, 9):775–784, 1973.

[4] B. F. C. Cooper. Correlators with two bit quantization. *Australian Journal of Physics*, 23:521–527, 1970.

[5] F .K. Bowers and R. J. Klingler. Quantization noise of correlation spectrometers. *Astron. Astrophys. Suppl.*, 15:373–380, 1974.

[6] R. Escoffier, L.R. D'Addario, C. Broadwell, J. Webber, and F. Schwab. Computer simulation of a FIR filter for the MMA. *MMA Memo No. 248*, February 1999.

[7] A.W. Gunst. Modeling the effects of re-quantization for ALMA-FC: The simulation setup. Technical Report Report-005 ASTRON-28000-R1, ASTRON, Dwingeloo, the Netherlands, February 2002.

[8] A.W. Gunst. Modeling the effects of re-quantization for ALMA-FC: Effects in spectral estimation. Technical Report Report-013 ASTRON-28000-R1, ASTRON, Dwingeloo, the Netherlands, February 2002.

[9] A. W. Gunst. The use of multi-bit A/D converters in radio astronomy. Master's thesis, University of Twente, Enschede, The Netherlands, june 1999.

[10] *Matlab: High Performance Numeric Computation and Visulation Software*. The Math Works Inc., 2000. Reference: `www.mathworks.com`.

[11] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C*. Cambridge University Press, 1995.

[12] A. W. Gunst, A. B. J. Kokkeler, and G. W. Kant. A/D converter research for SKA. In *Proceedings of the International Conference on Technologies for Large Antenna Arrays*, Dwingeloo, The Netherlands, 12-14 april 1999. NFRA.

A Error in Straight Line Data Fitting

Consider N_p data points (x_i,y_i), each with variance σ_i^2 (normal distributed), which have to be fitted into a straight line $y(x) = a + b \cdot x$ (degree $n = 1$). According to [11] (Section 15.2, Eq. (15.2.9)) the variance of a and b can be calculated with

$$\begin{aligned} \sigma_a^2 &= \frac{S_{xx}}{\Delta} \\ \sigma_b^2 &= \frac{S}{\Delta} \end{aligned} \tag{48}$$

with S defined as

$$S = \sum_{i=1}^{N_p} \frac{1}{\sigma_i^2} \tag{49}$$

and Δ as

$$\Delta = S \cdot S_{xx} - (S_x)^2 \tag{50}$$

Furthermore S_{xx} is defined as

$$S_{xx} = \sum_{i=1}^{N_p} \frac{x_i^2}{\sigma_i^2} \tag{51}$$

and S_x as

$$S_x = \sum_{i=1}^{N_p} \frac{x_i}{\sigma_i^2}, \tag{52}$$

with N_p the number of data points. The data points are dependent on various parameters, therefore an upper bound for variance σ_b^2 is defined (the variance of a is not of interest for this document). Furthermore the variance on each data point is assumed equal (is justified in [8] because $N >> M$ and the data points are near lag zero). Than,

$$S = \frac{N_p}{\sigma_{\rho_{xy}}^2}, \tag{53}$$

with $\sigma^2_{\rho_{xy}}$ the variance on each data point. Substituting Eq. (51), Eq. (52) and Eq. (49) in Eq. (50) and assuming the variance on each data point to be equal, results in

$$\begin{aligned}\Delta &= \sum_{i=1}^{N_p} \frac{1}{\sigma_i^2} \cdot \sum_{i=1}^{N_p} \frac{x_i^2}{\sigma_i^2} - \left(\sum_{i=1}^{N_p} \frac{x_i}{\sigma_i^2} \right)^2 \\ &= \frac{N_p}{\sigma^2_{\rho_{xy}}} \cdot \frac{1}{\sigma^2_{\rho_{xy}}} \sum_{i=1}^{N_p} x_i^2 - \frac{1}{\sigma^4_{\rho_{xy}}} \left(\sum_{i=1}^{N_p} x_i \right)^2 \\ &= \frac{1}{\sigma^4_{\rho_{xy}}} \left[N_p \sum_{i=1}^{N_p} x_i^2 - \left(\sum_{i=1}^{N_p} x_i \right)^2 \right]\end{aligned} \tag{54}$$

Hence, the variance of b yields,

$$\sigma_b^2 = \frac{N_p \cdot \sigma^2_{\rho_{xy}}}{N_p \sum_{i=1}^{N_p} x_i^2 - \left(\sum_{i=1}^{N_p} x_i \right)^2} \tag{55}$$