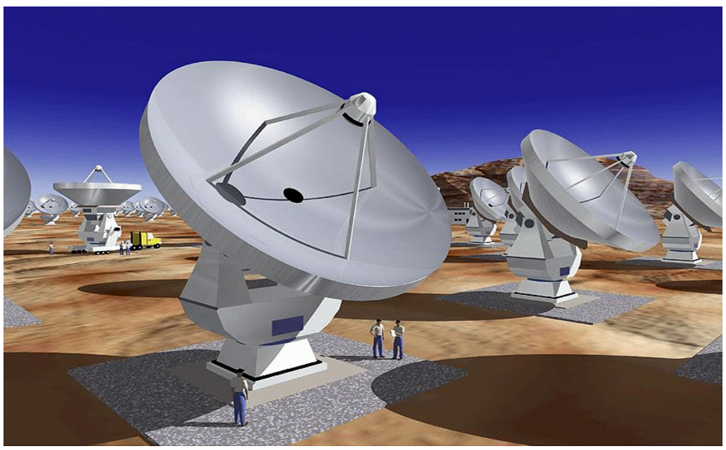
Author: A.W. Gunst Verified by: C.M. de Vos	Date of issue: June 16, 2001 Kind of issue: Public	Scope: Development Doc.nr.: Report-005 ASTRON-28000-R1
Responsible: C.M. de Vos Approved by: C.M. de Vos	Status: Preliminary Revision nr.: 0.2	File: G:\data\tekkamer\28000-r1\475repor.pdf
The ALMA FC study carried out at ASTRON is a contribution to the European ALMA Backend Electronics Team.		

Simulation Results of Low Bit FIR Filters



Distributionlist:

Group:	For Information:
S. Alliot	A. van Ardenne
A. Bos	A.J. Boonstra
J.D. Bregman	C.A. Leering
G.N.J. van Diepen	
A. Doorduin	
A.W. Gunst	
C.M. Loose	
K. van der Schaaf	
G.W. Schoonderbeek	
C.M. de Vos	

Documentrevision:

Revision	Date	Section	Page	Modification
0.1	April 27, 2001	-		Creation
0.2	June 16, 2001	-		Continued

Contents

Introduction 4

1 Simulation Setup 4

2 Effect of Input Signal 9

3 Effect of Number of Time Samples 14

4 Effect of Coefficient Quantization 17

5 Effect of In Between and Output Quantization 17

6 Skipping Most Significant Bits 20

Conclusion 21

References 22

Introduction

In the European approach of the Atacama Large Millimeter Array a number of digital FIR filters will be used to process the same number of subbands separately. The complexity of these filters is determined by the specifications and the number of bits representing the input, in between and output signals. The effect of the specification on the FIR complexity is discussed in [1], while the effect of code word representation is already discussed in [2]. However [2] is not fully applicable for ALMA because the number of input and output bits is much more than used for ALMA. Another issue which distinguishes that document from this [1] is the input signal. In [2] the filter characteristics are obtained using an impulse function at the input, while Gaussian input signals are assumed throughout this document. The differences with impulse response functions is outlined in Section 2. To show the effects of quantizers in the FIR filters it is important to involve the correlator in the system as well. The model is extensively discussed in Section 1. This document treats the effect of the quantizers on the filter characteristic, given a certain filter. Once, the effects are known a trade off can be made between the number of bits at each stage and the number of taps given the specifications. The design space of a certain implementation contains the number of bits and the number of taps. For each choice of the parameters there is not necessary a unique solution. Choosing the optimum solution given a specification is not topic of this document.

In this document equidistant quantization levels are assumed. This was done to structure the simulations. A next step can be, to sacrify the equidistance between the quantization steps. For a signal to noise ratio point of view, this can be more profitable. Further the direct form FIR filter structure is assumed.

The number of time samples used during simulations is motivated in Section 3, while the effect of coefficient quantization is discussed in Section 4. The in between and output quantization is outlined in Section 5. In the last section most significant bit reduction is discussed, while in the rest of the document only least significant bit reduction is assumed.

1 Simulation Setup

The Matlab [3] simulation setup is shown in Fig. 1. Within this model two pseudo random noise sources $y_1(t)$ and $y_2(t)$ are considered. Both are correlated with normalized correlation ρ. This is accomplished by adding two uncorrelated noise sources of power 1 by a correlated noise source of power 1 weighted with b, while keeping the total power at a constant level [4]. This is done for

[1] This document can be viewed best in full color

each point on the correlation function. Representing the pseudo random time signals $y_1(t)$ and $y_2(t)$ as stochastic signals results in the following relations:

$$Y_1 = \frac{1}{\sqrt{1+b(\tau)^2}}\Big(N_1 + \mid b(\tau) \mid N_0\Big) \quad (1)$$

$$Y_2 = \frac{1}{\sqrt{1+b(\tau)^2}}\Big(N_2 + b(\tau) N_0\Big) \quad (2)$$

$$b(\tau) = \sqrt{\frac{|\rho(\tau)|}{1-|\rho(\tau)|}} \cdot \mathrm{sign}(\rho(\tau)), \quad (3)$$

with $\rho(\tau)$ the normalized correlation function. The factor $\sqrt{1+b(\tau)^2}$ is used for holding the total output noise power at 1.

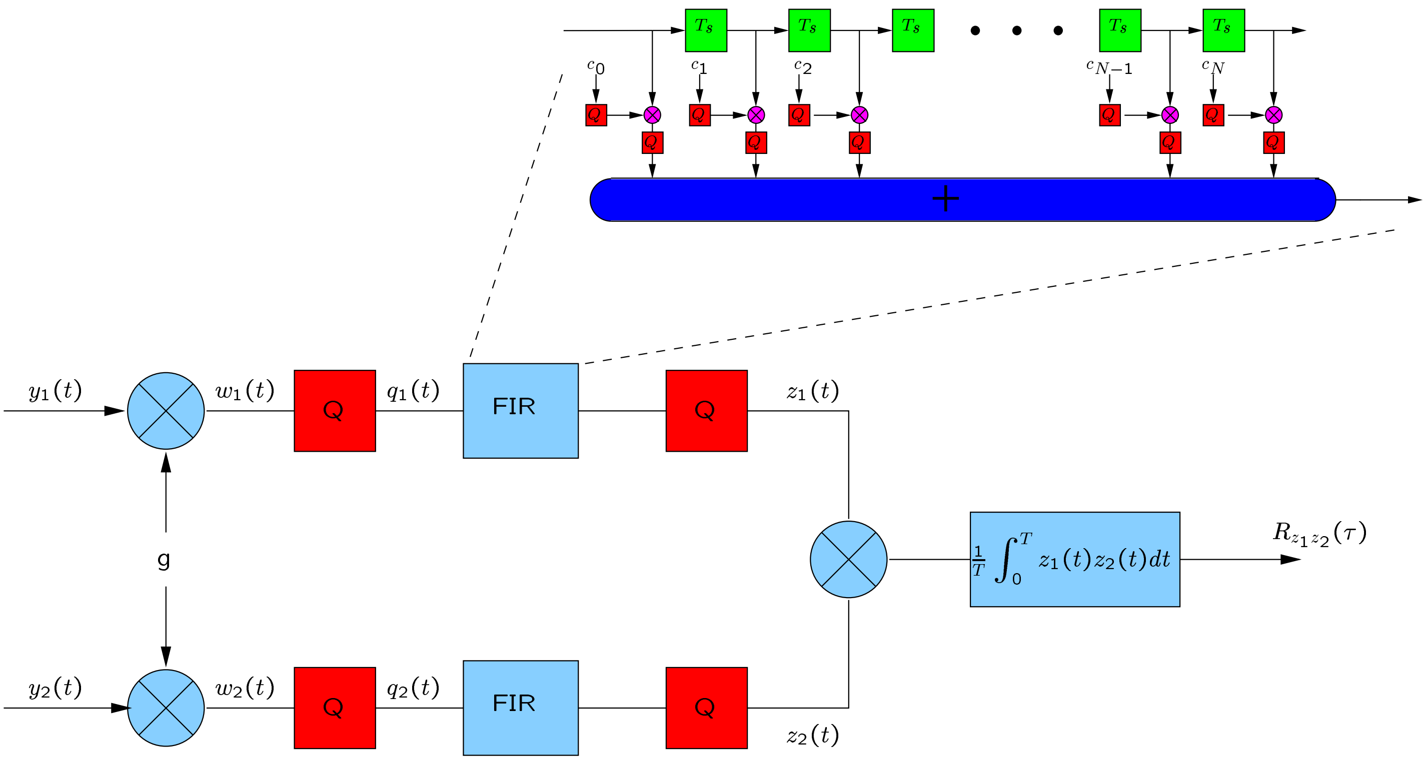


Figure 1: Simulation model used for determining re-quantization effects in FIR filters.

The complete signal is amplified with a factor g. Assuming a fixed midriser quantizer with unity stepsize, quantity g determines the number of reference levels crossed. The RMS noise value of 1 is scaled to g LSB.

Simulations are done using a lowpass filter with a passband at approximately a quarter of the sample frequency f_s. A relative broad filter is used, because a limited number of points can be used throughout the simulation. The simulation time increases, when the amount of taps is increased. Further the current specifications of the ALMA FIR filters are used, i.e. a ripple of 1 dB, an attenuation of 40 dB and a transistion region of 10 percent [1]. Using an equiripple design, resulted in a 58 taps filter. The ideal frequency response is depicted in Fig. 2. The lowpass filter is implemented in a straigthforward manner using the transversal structure.

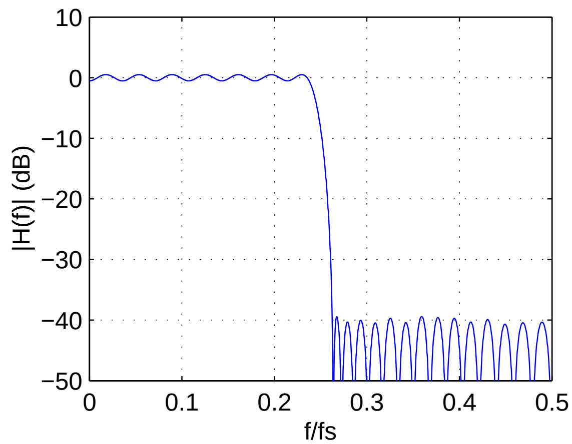


Figure 2: Amplitude response of a 58 taps continuous FIR filter.

Each quantizer in the FIR filter has a riser at the origin and unit step size. The input signal is coded with n_i bits. The number of bit used to represent the coefficients is set to n_c. The result after multiplication is quantized to n_b bit. After adding all products the sum is re-quantized to n_e bit. In Fig. 3 a re-mapping of code words is given. The number of extra bits due to the addition equals n_a (can be determined when the coefficients are known [2]. From the figure can be verified that a total gain is introduced during re-quantization. When no gain is introduced during input quantization the total gain equals

$$G = 2^{n_b-1-(n_i+n_c-2)} \cdot 2^{n_e-1-(n_b+n_a-1)} = 2^{n_e-n_i-n_c-n_a+1} \tag{4}$$

This gain is compensated in the simulation program.

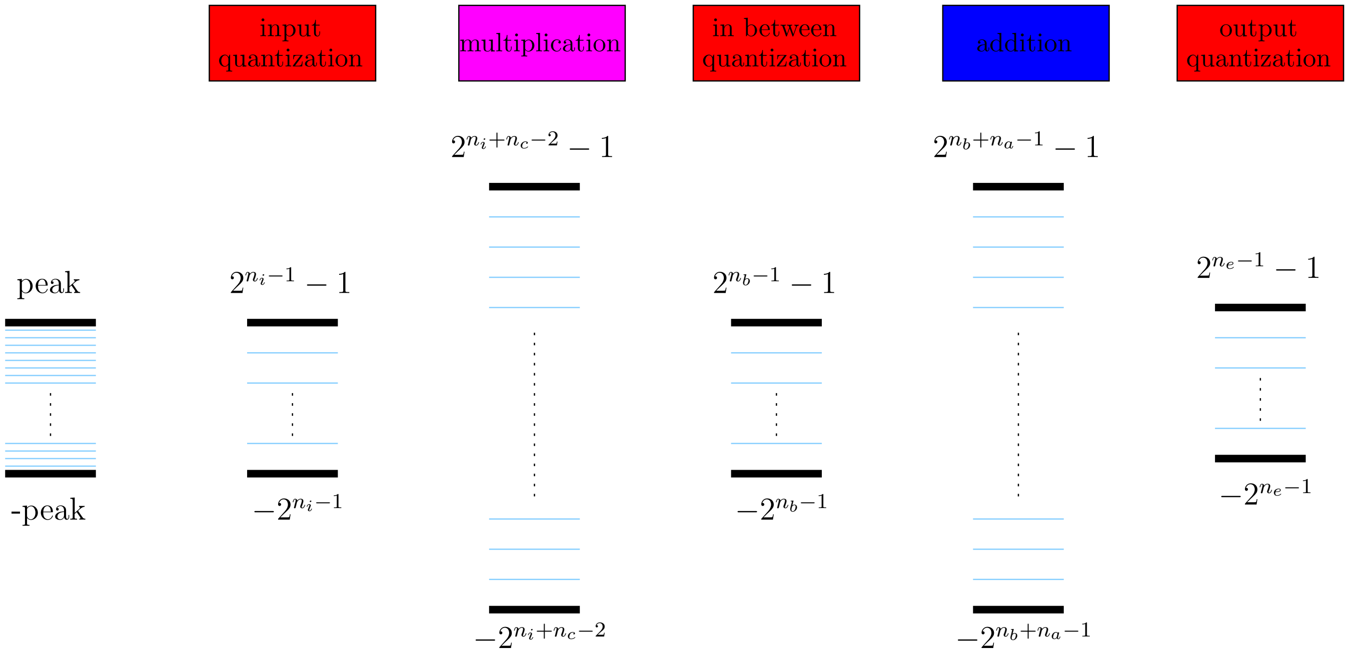


Figure 3: Code word progress in FIR filter due to (re-)quantization.

Another parameter which must be set is the number of time samples and the number of lags used in the correlator. Increasing the number of time samples while fixing the lags results in more accurate simulations (the variance on the correlation function decreases with the number of time samples). On the other hand increasing the number of lags while fixing the number of time samples results in a more accurate frequency resolution, while sacrifying the variance on the correlation function. For ALMA 32 lags are used for each subbands. Suppose 32 subbands are applied, than the total number of lags equals 1024. Because the lowpass filter used in this report has a passband till approximately a quarter of the sample frequency, the number of lags which can be used in that subband can be $\frac{1024}{2} = 512$. Therefore the number of points on the correlation function is chosen to be 512, 256 negative lags and 256 positive lags. Before discussing the results first quantities are defined to quantify the filter characteristic. Fig. 4 defines four important quantities which determines the filter characteristic. For the filter under consideration the passband frequency F_p is set to $0.2375 f_s$, while the stopband frequency is set at $0.2625 f_s$. The actual passband and stopband frequencies are determined by the resolution. In all simulation 512 correlation points are assumed and therefore the frequency range goes in steps of $\frac{1}{512}$. Given the passband and stopband frequencies the ripple R_p and stopband attenuation R_s are determined. In many cases the passband frequency is larger than for the continuous filter. When determining the stopband attenuation for this case a smaller value is obtained. Therefore another stopband attenuation definition is used, further refered as R_{s2}. The first valley of the stopband frequency range is determined. Thereafter the worst attenuation within this frequency range is determined, which is defined as R_{s2}. Accordingly the first frequency from $0.25f_s$ having the worst attenuation is determined. This frequency point is defined as F_{s2}. When the ripple is determined, afterwards a frequency point F_{p2} is determined where the filter characteristic is still within ripple. This value F_{p2} can be larger than F_p. So, R_p and R_s are determined given F_p and F_s, while the other quantities are determined from R_p and the worst attenuation in the stopband. The new quantities are defined in Fig. 5.

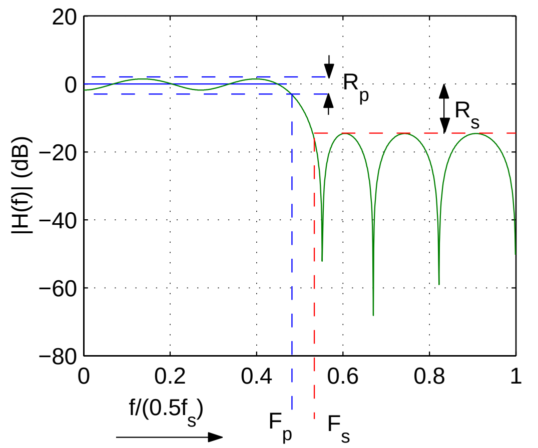


Figure 4: Definition of FIR filter quantities, with f_s the sample frequency.

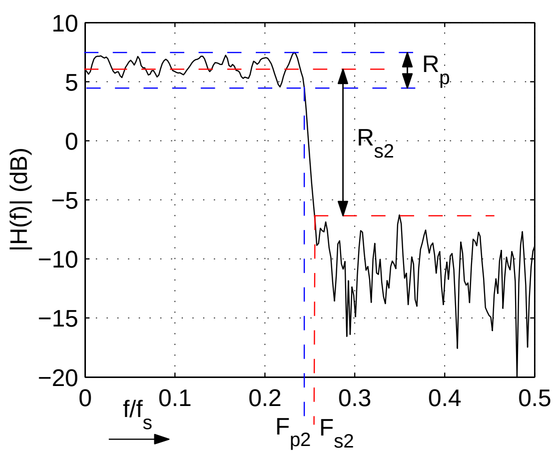
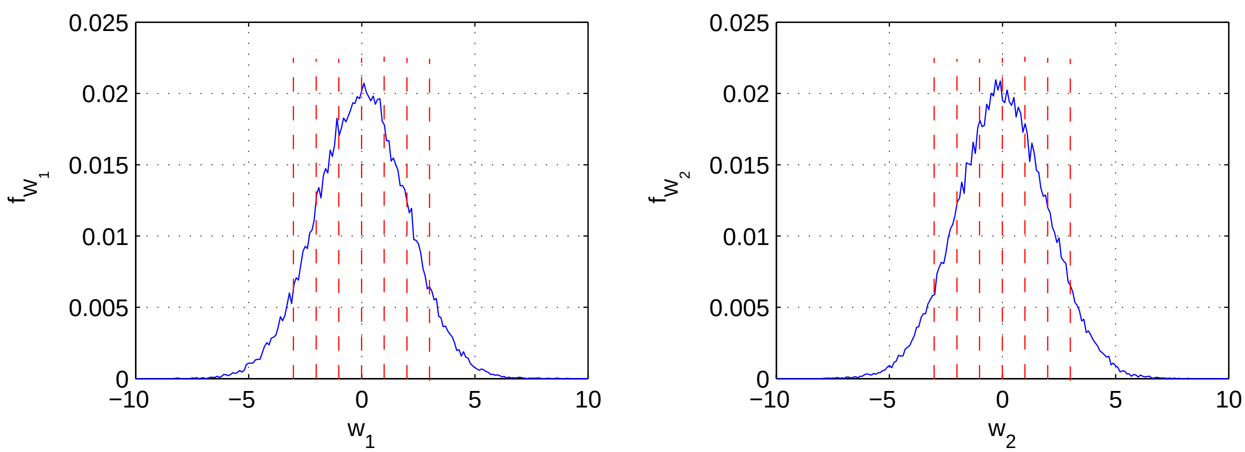


Figure 5: Additional finite length FIR filter definitions.

2 Effect of Input Signal

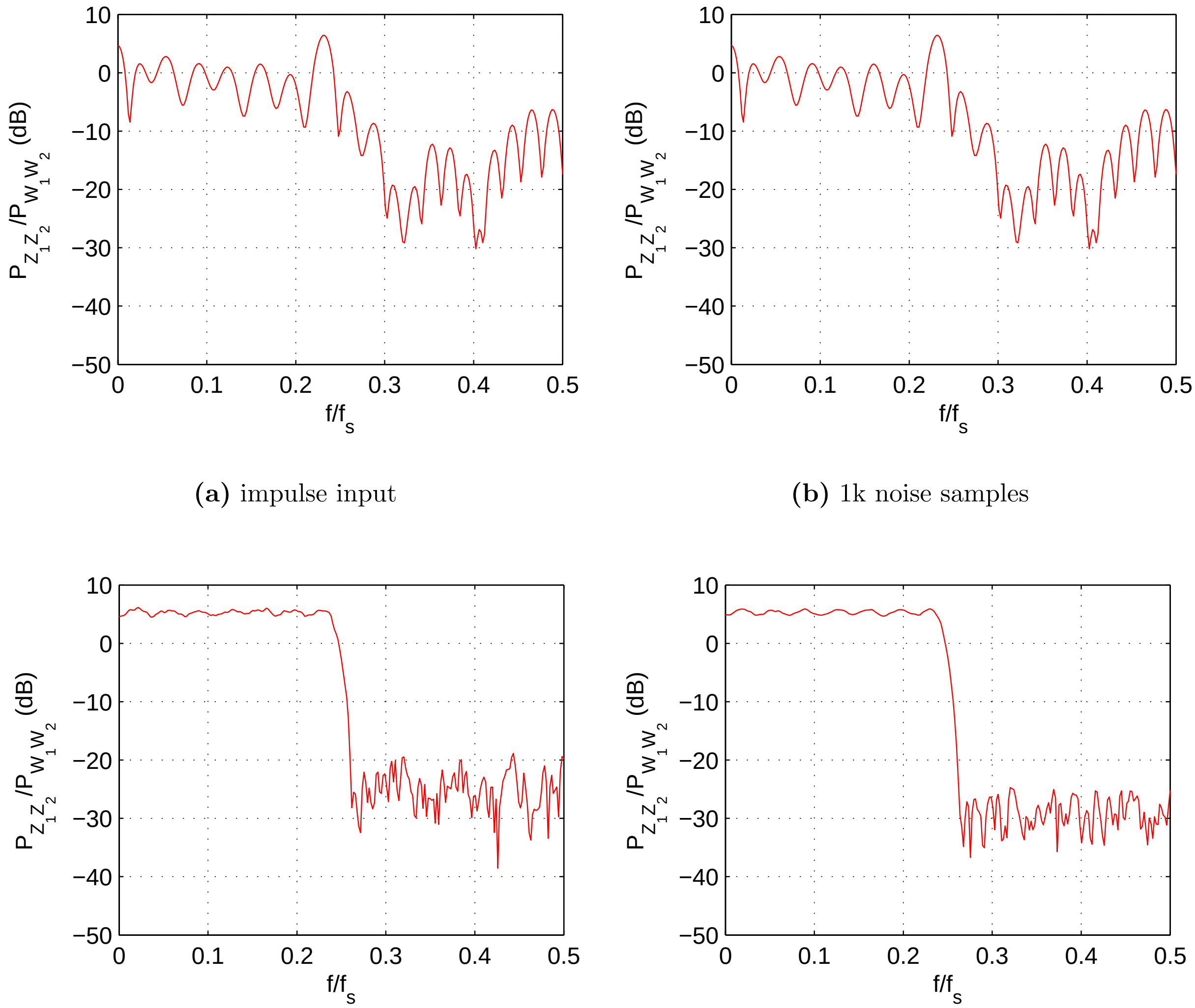
In [2] the effect of finite length FIR filters is discussed for an impulse function at the input. The filter characteristic obtained with an impulse function at the input is independent on the number of time samples. For an partly uncorrelated noise signal the filter characteristic becomes better when more time samples are taken (this is discussed extensively in Section 3). In this section both are compared. Normally the filter response caused by an impulse at the input is enough to describe the filter for all classes of input signals. However, when (re-)quantization is applied this statement is not valid anymore, i.e. the filter characteristic depends on the input signal. For ALMA noise inputs are of interest. Therefore throughout this report noise signals are considered. The used noise signals are generated in Matlab with a mean of zero and a variance of 1. It is already mentioned that after adding the correlated and uncorrelated noise part the total variance is still 1. Thereafter the signal is scaled with g. This determines the total variance. In ALMA a three bit A/D converter is applied. According to [5] for three bit the optimal noise power at the input assuming a unity stepsize A/D converter equals 2.1 for small correlation coefficients. For this report $g = 2$ is chosen. The probability density functions of both signals in the model depicted in Fig. 1 after amplifying it with g are given in Fig. 6.



(a) impulse input **(b)** 64k noise samples

Figure 6: Probability density function of signal W_1 (a) and W_2 (b), using 64k time samples (levels of A/D converter are indicated with the dashed lines).

To show the difference between the impulse response function of the filter under consideration and the filter characteristic when using noise signals, both are depicted in Fig. 7 (for all spectra a kaiser window is assumed with parameter 10). In fact for the impulse input function the autocorrelation is calculated. The amplification in the passband is caused by the first quantizer. The transfer characteristics for an impulse function and 64k noise samples without input quantizers is depicted in Fig. 8. From Fig. 7(b), (c) and (d) can be concluded that increasing the number of time samples enhances the filter transfer characteristic. This is because the noise at the input is partly correlated. By increasing the number of samples, the amount of uncorrelated noise reduces. Therefore the filter characteristic obtained with Gaussian noise, depends on the correlation coefficient. Fig. 9 depicts the filter characteristicis obtained with n_e=6 for 1k, 8k and 64k time samples.



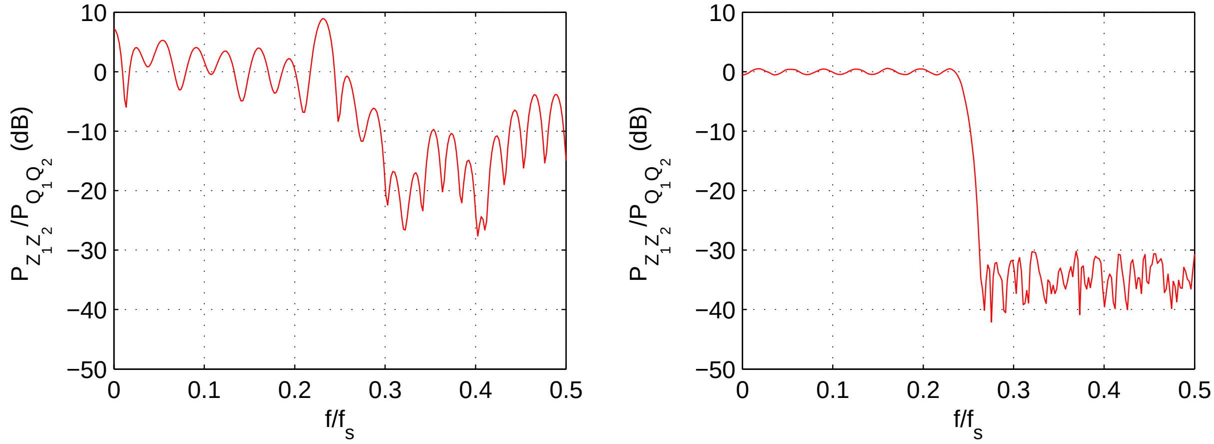
(a) impulse input

(b) 1k noise samples

(c) 8k noise samples

(d) 64k noise samples

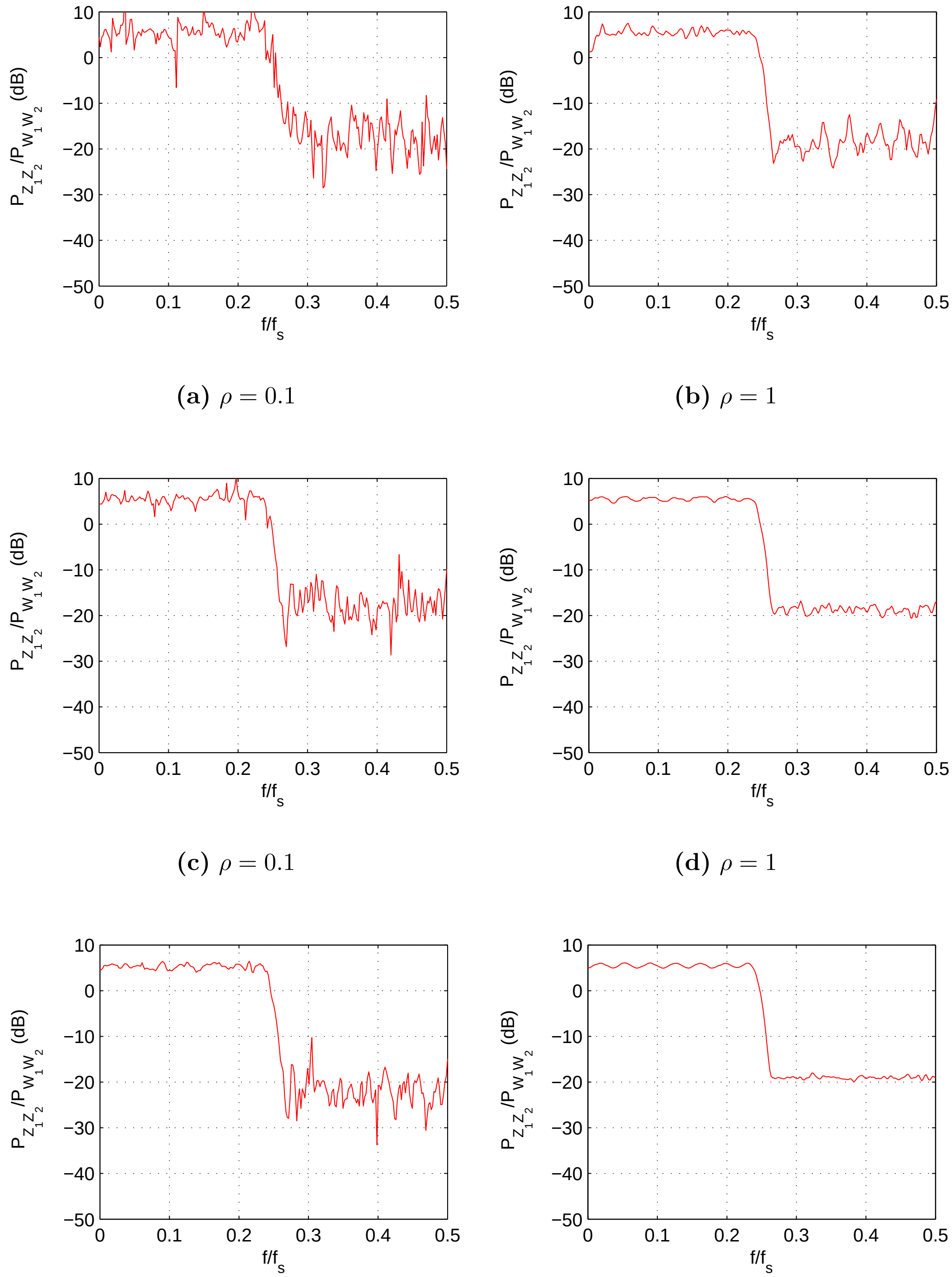
Figure 7: Filter characteristics obtained with (a) an impulse and partly (ρ=0.5) correlated Gaussian noise of 1k time samples (b) 8k time samples (c) and 64k time samples (d) for a 58 taps FIR filter with 3 input bits and 6 output bits. For the coefficients 16 bits are used. For both signals the same model is used.



(a) impulse input

(b) 64k noise samples

Figure 8: Filter transfer characteristic excluding the input quantizers for an impulse input (a) and partly correlated Gaussian noise of 64k time samples (b) at the input.



(a) $\rho = 0.1$ **(b)** $\rho = 1$

(c) $\rho = 0.1$ **(d)** $\rho = 1$

(e) $\rho = 0.1$ **(f)** $\rho = 1$

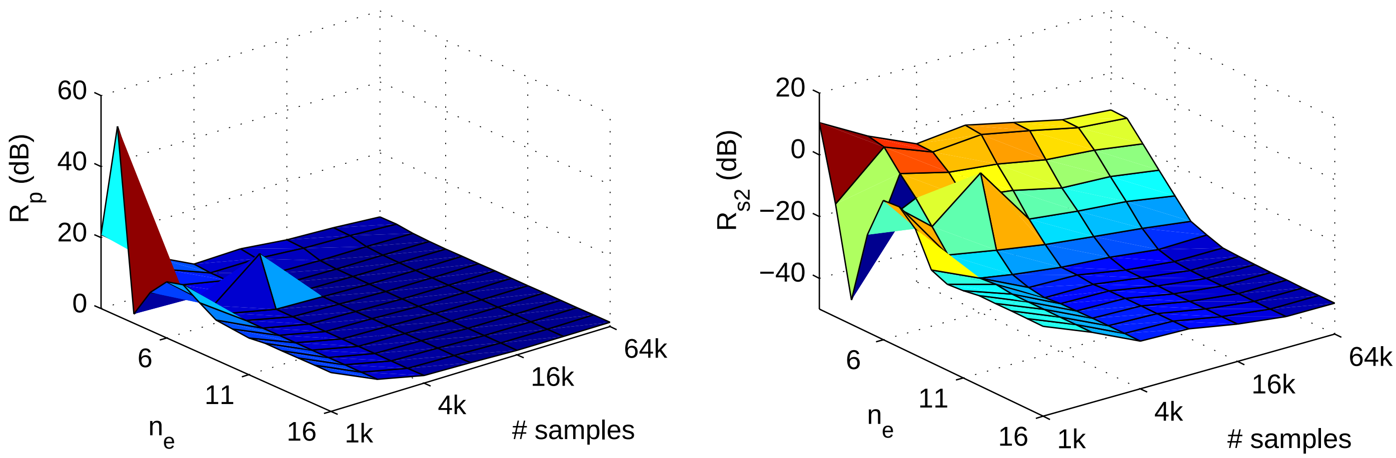
Figure 9: Filter characteristics obtained with (a) and (b) 1k time samples, (c) and (d) 8k time samples and (e) and (f) 64k time samples for noise with a correlation coefficient of 0.1 and 1 (n_e=6).

In the remainder the filter characteristics obtained with the model of Fig. 1 are considered for Gaussian input signals.

3 Effect of Number of Time Samples

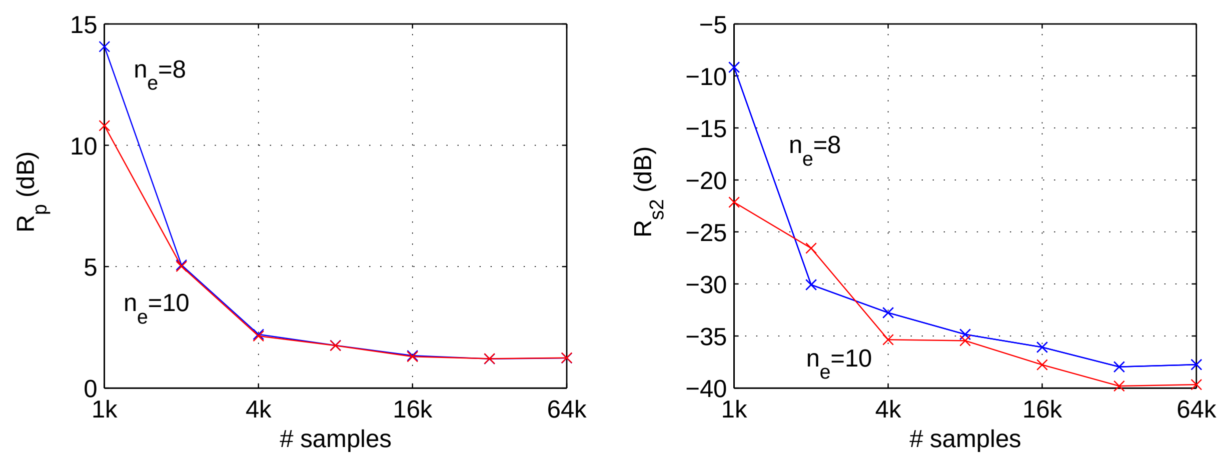
The variance of the correlation function decreases linearly with the number of time samples, given a fixed number of lags. A doubling in time samples results in a decrease of the variance with a factor 2. The number of time samples used is limited by the simulation time. The dependency of the filter characteristic on the number of time samples is only the case when partly uncorrelated noise is applied. If the noise signal at the input has a small correlation coefficient, than for an infinite number of time samples the theoretical filter characteristic should be obtained if only the ouput bits are quantized. Quantization of the coefficients results immediately in a degradation of the filter response, which cannot be cancelled due to integration. For ALMA a partly correlated noise source is present at the input. Therefore the number of coefficient bits must be chosen not too low.

In Fig. 7 the effect of increasing the number of points was already shown. Doubling the number of time samples results in 3 dB less variance on the power spectral density (this is verified without the filters). So, the reduction in variance going from Fig. 7(b) to Fig. 7(c) equals 9 dB. These figures also show the large passband ripple for a less amount of time samples. Therefore more time samples are necessary in order to determine the ripple accurately. In Fig. 10(a) the ripple is shown as a function of n_e and the number of time samples, while Fig. 10(b) depicts the stopband attenuation. From both figures can be seen that at least 8k samples are necessary to obtain an accurate estimation of the ripple given n_e. Fig. 11 enhances the visualisation of the convergence of the ripple when increasing the number of samples. The ripple converges quicker than the stopband attenuation does. Because of the definition of R_{s2} the size of the stopband differs also (because this section deals only with the influence on the number of time samples this is not discussed here). Trading off the convergence and an acceptable simulation time resulted in the use of 64k time samples. For an infinite amount of time samples the filter transfer characteristic would approach the continuous one (assuming no coefficient and in between quantization).



(a) **(b)**

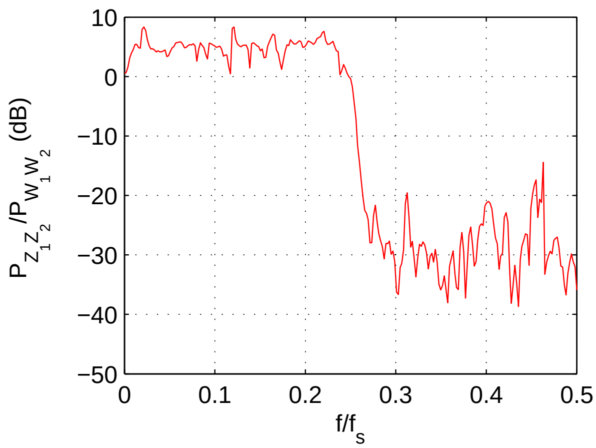
Figure 10: Ripple R_p (a) and stopband attenuation R_{s2} (b) as a function of n_e and the number of time samples. The correlation coefficient of the input signal is 0.5.



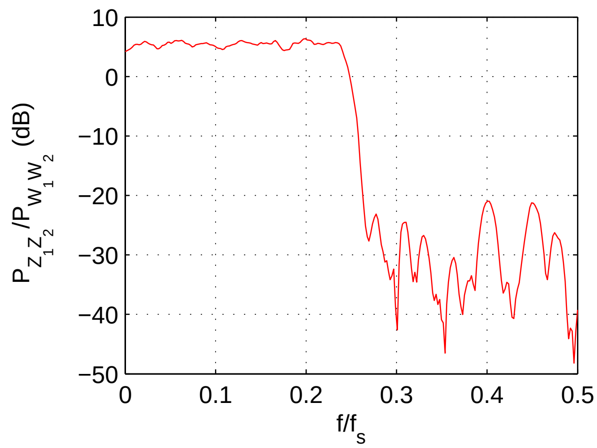
(a) **(b)**

Figure 11: Ripple R_p (a) and stopband attenuation R_{s2} (b) as a function of the number of time samples for n_e is 8 and 10 bit. The correlation coefficient of the input signal is 0.5.

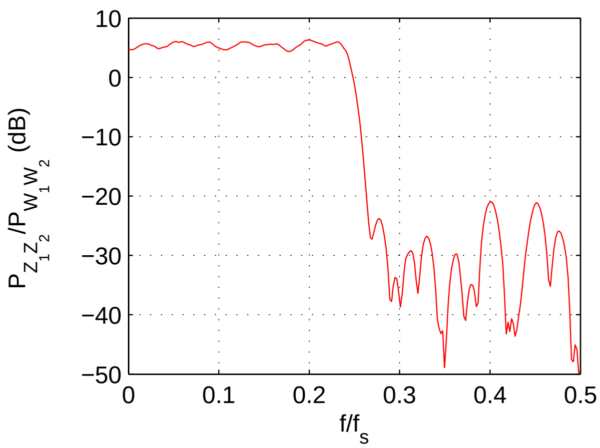
In the previous discussions only the number of end bits are reduced. It looks that for enough time samples the theoretical filter frequency response can be obtained. However this is not the case when the number of coefficient bits is reduced as shown in Fig. 12. Reducing the number of coeffient bits results directly in a degradation in filter performance, which can not be undo with using partly correlated noise at the input. This seems logical because the coefficients determine the theoretical filter characteristic. When the number of coefficient bits are reduced the theoretical filter characteristic is changed.



(a) 1k samples



(b) 8k samples

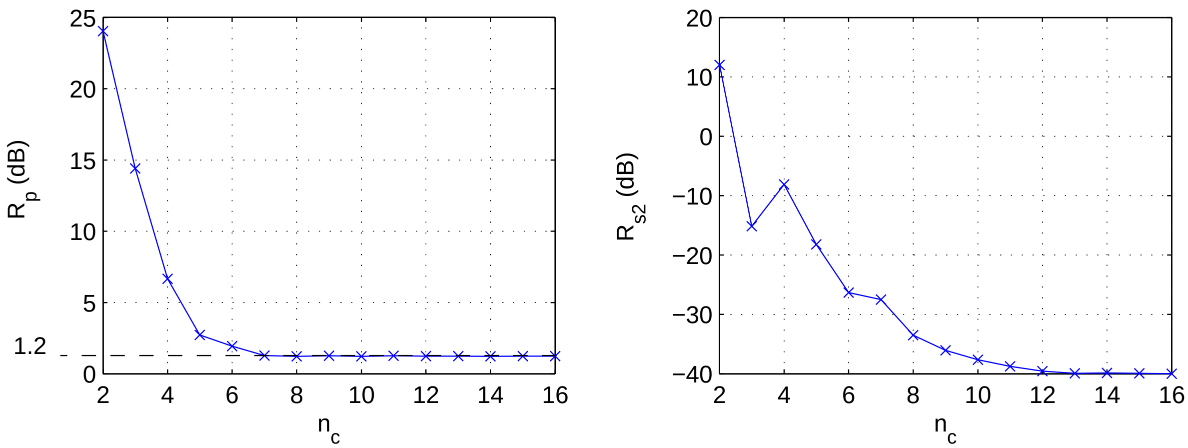


(c) 64k samples

Figure 12: Filter characteristic with n_c=6 for (a) 1k, (b) 8k and (c) 64k time samples.

4 Effect of Coefficient Quantization

In Fig. 13 the ripple and stopband attenuation as function of the number of coefficient bits are shown. The number of output bits and in between bits is set to 16. To sacrefice only a little bit of the specification 12 bits are chosen. This results in 1.2 dB ripple in the passband, with F_{p2}=0.2383 and a stopband suppression of 39.5 dB with F_{s2}=0.2656.



(a) **(b)**

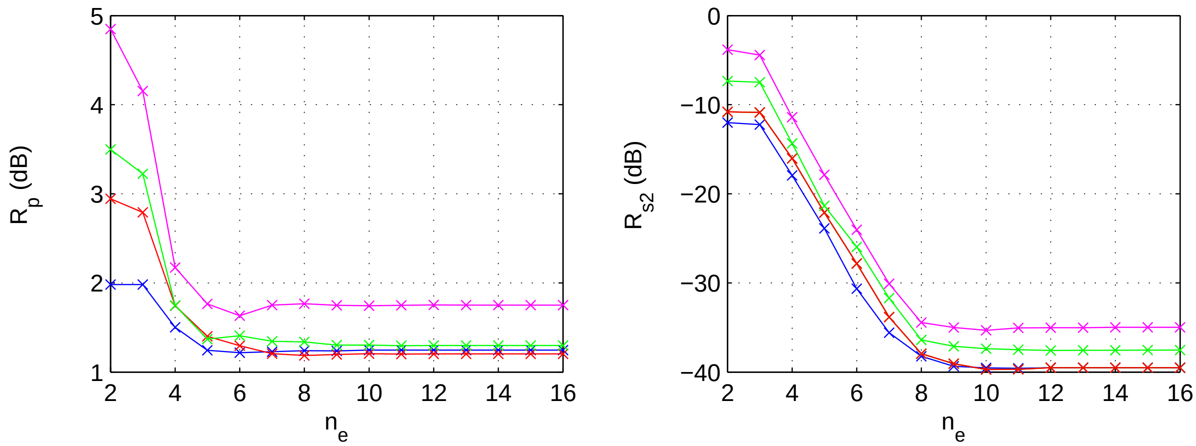
Figure 13: Ripple and stopband attenuation as a function of coefficient bits (64k time samples are used).

5 Effect of In Between and Output Quantization

Now the input bits and coefficient bits are fixed, the number of bits to represent the in between results must be chosen. The re-quantization is done after multiplying the coefficient with the input signal. Since the number of output bits is constant (i.e. the number of correlator input bits in ALMA is specified), the output re-quantizer is determined by setting the number of in between bits. Two extreme re-quantization schemes can be chosen:

- no in between quantization, because working at full precision is the best strategy;
- full in between quantization, because the quantization will add uncorrelated noise which will be reduced by correlation.

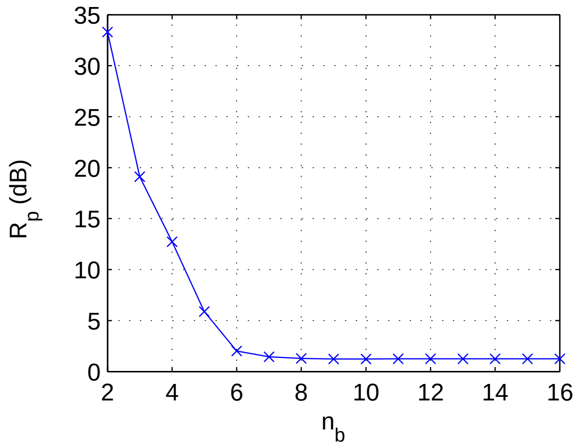
First Fig. 14 shows the ripple and attenuation as function of the number of end bits when the number of in between bits equals 16 (the first extreme case). From this is seen that the performance degrades significantly when the number of bits is reduced. Especially less than 8 output bits, given the number of time samples affects the filter charactertistic. When more time samples are used it is expected that the theoretical filter characteristic is approximated. In the ALMA case two input bits are used for the correlator. For the moment three bits are taken (a redundant most significant bit is assumed). Then definitely more time samples are necessary to show a better filter characterisitic.



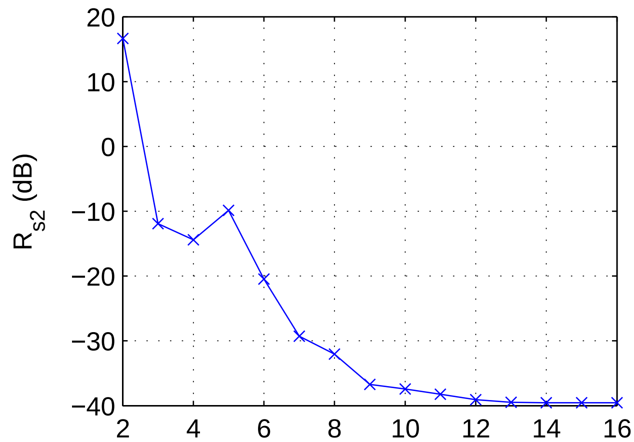
(a) (b)

Figure 14: Ripple (a) and stopband attenuation (b) as a function of output bits (8k, 16k, 32k and 64k time samples are used). The number of input bit is three, the number of coefficient bit 12 and the number of in between bit is 16.

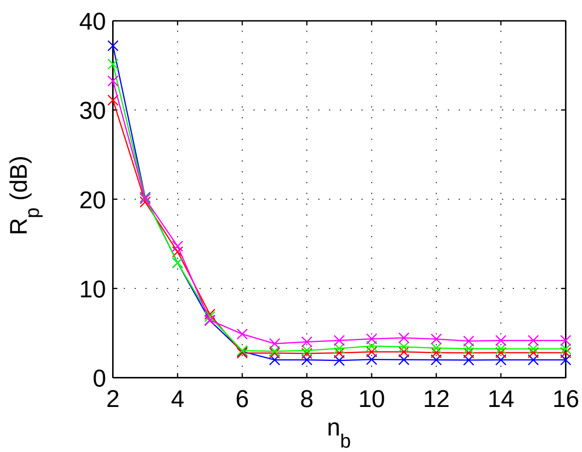
The effect of the number of in between bits when the output is coded with 16 bit is shown in Fig. 15(a) and (b). The results for three output bits is shown in Fig. 15(c) and (d). From this is seen that it is of no use to code the in between bits in more than four bits when the suppression is considered. Due to the coefficients used three extra bits are generated by the addition. So, the result after multiplication and in between quantization is increased by maximally three bits. Using three output bits results in maximally 6 in between bits. The reason that the number of in between bits can be less is because the in between quantization adds partly uncorrelated noise. Increasing the number of time samples beyond 64k gave not the expected improvements. This may be caused by the Matlab simulation (pseudo random noise was used).



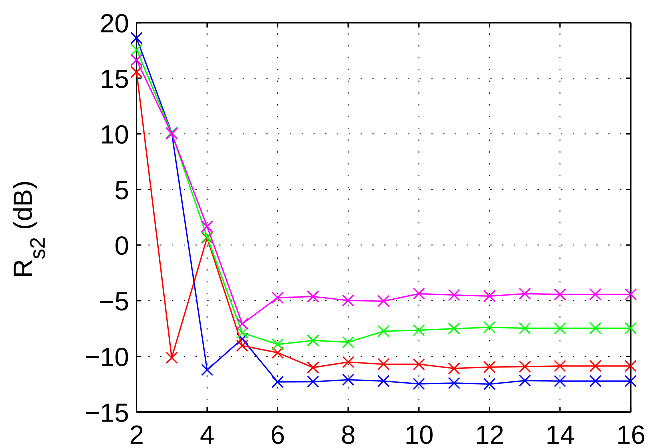
(a) n_e=16



(b) n_e=16



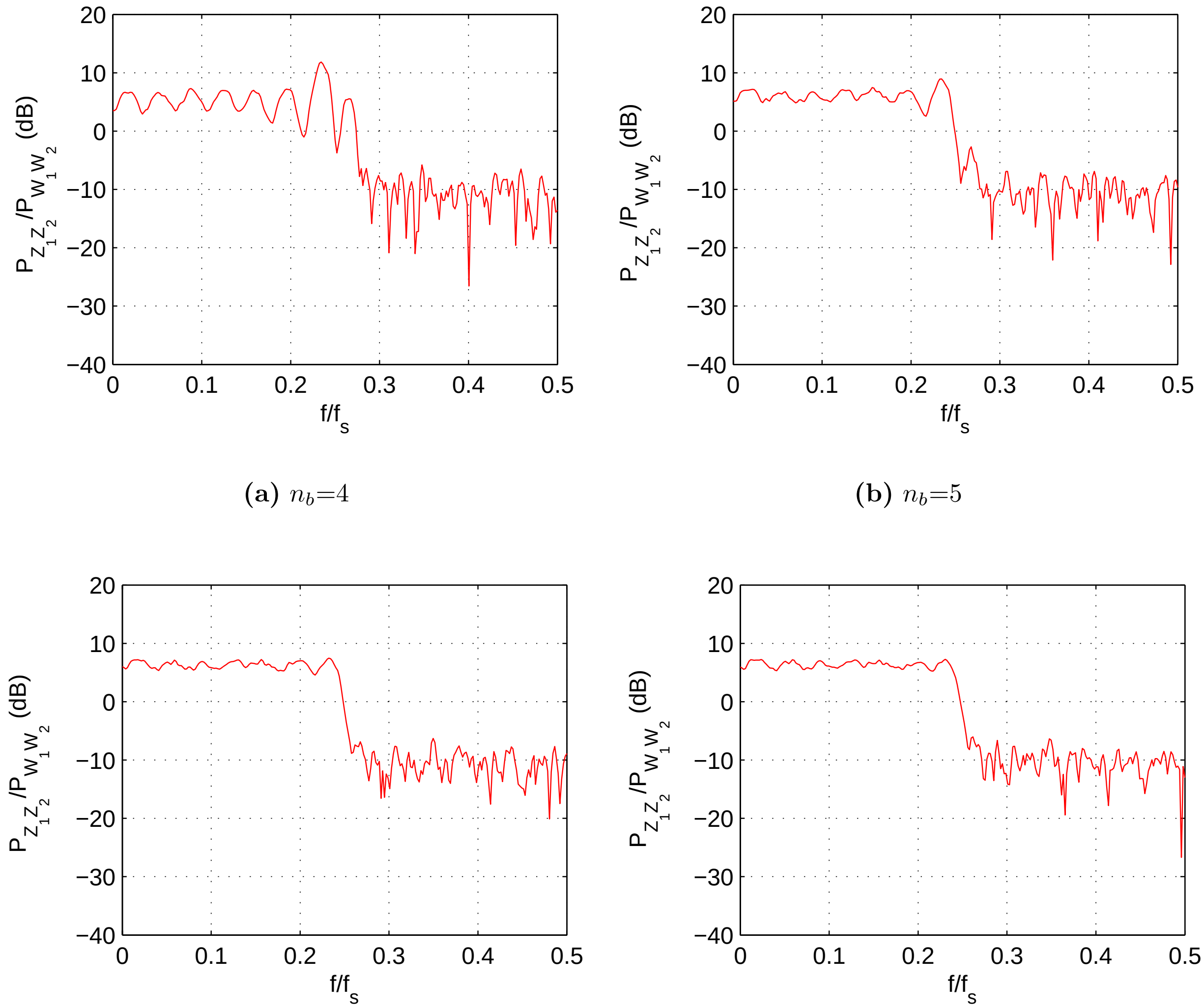
(c) n_e=3



(d) n_e=3

Figure 15: Ripple (a), (c) and stopband attenuation (b), (d) as a function of in between bits. The number of input bit is 3, the number of coefficient bit 12 and the number of output bit is 16 (a),(b) and 3 (c), (d) (8k, 16k, 32k and 64k time samples used).

The filter charactertistiscs for n_b equals 4, 5, 6 and 7 is shown in Fig. 16.



(a) n_b=4

(b) n_b=5

(c) n_b=6

(d) n_b=7

Figure 16: Filter responses for n_b is 4 (a), 5 (b), 6 (c) and 7 (d). The number of input bit is three, the number of coefficient bit 12 and the number of output bit is 3.

6 Skipping Most Significant Bits

The range of the finite length FIR filter output is determined by the coefficients and the input signal. It is conservative to use the full range of the output, when the probability for crossing the most significant bit is very low. The following relation holds when using random input signals [6]:

$$\sigma_e^2 \leq \sigma_i^2 \, ||F||_2^2 \, , \tag{5}$$

with the two norm defined as

$$||f||_2^2 = \sum_{k=0}^{\infty} f^2(k) \tag{6}$$

Filling in the coefficients of 12 bits and taking n_i=3 and 1 bit for the addition results in n_e=15 bit. The 2 norm equals 3203. Multiplying it with the input variance results in 12812. To represent this, 14 bit is required. Therefore one MSB can be skipped, when using random noise signals. This appeared also from the output signal. From the three output bits applied in this report, only two are used.

The range of the finite length FIR filter output is determined by the coefficients and the input signal. It is conservative to use the full range of the output, when the probability for crossing the most significant bit is very low. When using white inputs with $S_{uu}(f) = 1$ then the output v is bound to:

$$\sqrt{\mathrm{E}\left[v^2(k)\right]} = ||f||_2 \tag{7}$$

The following relation holds when using random input signals [7]:

$$\sigma_e^2 \leq \sigma_i^2 \, ||F||_2^2 \, , \tag{8}$$

with the two norm defined as

$$||f||_2^2 = \sum_{k=0}^{\infty} f^2(k) \tag{9}$$

Filling in the coefficients of 12 bits and taking n_i=3 and 1 bit for the addition results in n_e=15 bit. The 2 norm equals 3203. Multiplying it with the input variance results in 12812. To represent this, 14 bit is required. Therefore one MSB can be skipped, when using random noise signals. This appeared also from the output signal. From the three output bits applied in this report, only two are used.

Conclusions

To solve the re-quantization issues for ALMA, Gaussian noise input signals are used. The results presented in this report are obtained with a model including the correlator. By using more time samples the theoretical filter characteristic can be obtained. However the number of coefficient bits limits the filter characteristic. For the filter used in this report 12 coefficient bits seems not to harm the filter characteristic significantly. Because the number of output bits was limited to three, it is of no need to use a lot of in between bits. Four bits appeared enough.

When using random input signals one bit can be skipped from the most significant side. xxx seems reasonable. PERFORMANCE VRAAG: is minder bits input beter of slechter? minder bits weg te halen. anders implementeren steeds 1 optelling en weer quantizeren, dat is misschien veel beter. CWtje erin doen.

References

[1] A.W. Gunst. FIR filter complexity in ALMA. *Internal ASTRON Report: Report-004 ASTRON-28000-R1*, April 2001.

[2] A.W. Gunst. Simulation results of finite length FIR filters. *Internal ASTRON Report: Report-002 ASTRON-28000-R1*, November 2000.

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

[4] M. Kendall and A. Stuart. *The Advanced Theory of Statistics*. Charles Griffin & Company, fourth edition, 1977.

[5] 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.

[6] L.R. Rabiner and Bernard Gold. *Theory and Application of Digital Signal Processing*. Prentice Hall, Englewood Cliffs, 1975.

[7] R. A. Roberts and C . T. Mullis. *Digital Signal Processing*. Addison-Wesley, Massachusetts, 1987.