Quantization Effects in Digital Upconversion and Digital Beam Forming for TIGER Radar System

T. Salim, J. Devlin, J. Whittington
Department of Electronic Engineering
La Trobe University Bundoora, VIC 3086 Australia

1 INTRODUCTION

Quantization is a representation of data samples with a certain number of bits per sample after rounding to a suitable level of precision. Quantization errors in a Digital Signal Processing (DSP) system can be introduced from three sources; one source is input quantization, a second is coefficient quantization and the third is the finite precision in the arithmetic operations [1]. The quantization error in the arithmetic operations can be controlled by carefully selecting the size of buffer registers according to the input word length. Quantization errors from input and filter samples are considered in this article. The effects of quantization errors and the tradeoffs required between precision and hardware resources are discussed in relation to the implementation of the DSP in Field Programmable Gate Array (FPGA).

This article is divided into two main sections; quantization effects for upconversion and quantization effects for digital beamforming (DBF). Fixed length samples cause reduction in the filter dynamic range and gain resolution. We discuss finite precision effects on filter zeros and its frequency spectrum in section 2.1. Quantization error increases with filter order and this issue is described in section 2.2. Higher precision levels cause reduced quantization offsets. This is presented in section 2.3. Quantization produces similar effects in phased signals, however, its effects are different when considering multiple beam generation. Effects of precision levels on the beam patterns are described in section 3.1. In section 3.2, quantization effects are discussed for the first and second sidelobes.

2 QUANTIZATION EFFECTS ON UPCONVERSION

In multirate systems, upconvcersion can be achieved with oversampling and filtering techniques [2]. For the proposed digital TIGER system, input Gaussian pulses are upsampled to produce higher order Nyquist zones. A high pass FIR filter is employed to acquire a spectral zone at the expanded band edge. In this case, higher efficiency is possible by exploiting filter symmetry. For a higher throughput rate, polyphase implementation of the FIR filters can be employed [3]. Since signal amplification is performed in the analog domain, a high speed 14 bit DAC is used for digital to analog conversion.

Finite precision causes similar effects in the input data samples and filter coefficients. Fixed word length effects on filter coefficients, filter length and dynamic range are described in the following sections.

2.1 Sensitivity of Filter Coefficients to Quantization

Finite precision plays a significant role in the dynamic range of filter gain and DC offset. A large number of quantization levels will decrease the quantization error; on the other hand it requires larger silicon space to implement the design. The quantization affects the input Gaussian pulse and the filter coefficients. The pole and zero maps show perturbations in Figure 1 when samples are restricted to finite word length. The filter coefficients in the lower parts are constrained to 14 bit quantized samples and the length of the filter is 100 taps. This constraint arises from the fast DAC of 14 bit width used for converting a digital signal into the analog domain.

Since the dynamic range of the quantizer is less than that of the filter coefficients, the quantized coefficients are disturbed from the unit circle. The gain of the quantized filter response is displayed in Figure 1 which is distinctly less than that for the infinite precision filter. For these simulations infinite precision representation is regarded as floating point, which provides significantly better precision than the quantization levels discussed here. The zeros around $Z = -1$ are responsible for passband attenuation and are less displaced. As the dynamic range of the quantizer is increased to match the filter coefficients, the signal to quantization noise ratio (SNR) improves, but at the cost of increased hardware resources. Similar results can be obtained for the input Gaussian pulse when quantized to specified fourteen bit word lengths.

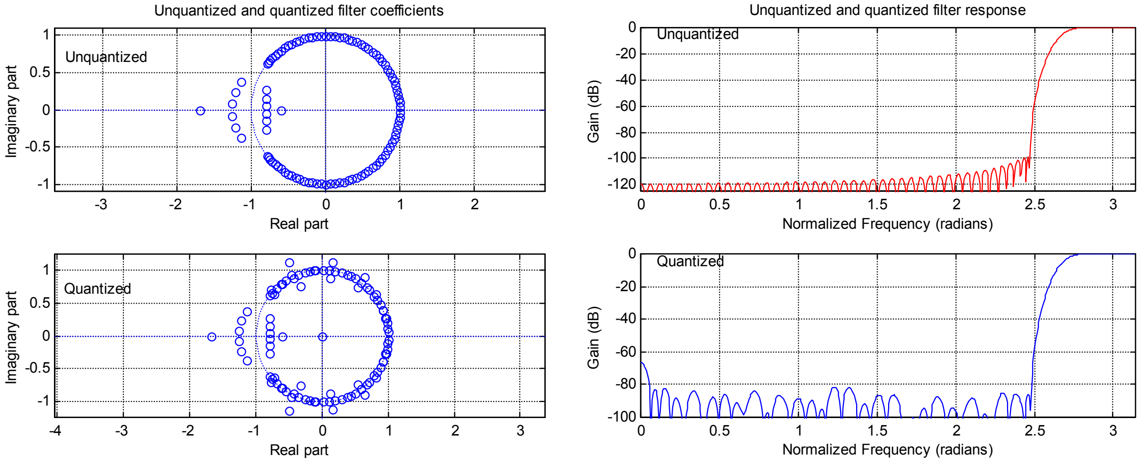

Figure 1 Quantization effects on (a) filter coefficients (b) frequency response

Finite precision is hardware efficient since the system data width is less than the infinite precision (or floating point) case. Quantization reduces a few out of 100 coefficients to zero, which will further ameliorate the memory cell and arithmetic processing requirement. Quantization also reduces the filter gain compared to infinite precision samples; however this reduction is acceptable as long as it remains within an attenuation limit. The fourteen bit quantizer provides more than 80dB attenuation which is better than the standard of 60dB used by many communication systems.

2.2 Quantization Effects on Filter Order

For direct conversion transmission, a cascaded design performs better than a single stage. This is because quantization errors are reduced with a lower filter order. Secondly a lower order design requires less logic resources. Quantization errors vary with the length of a filter and we now study the effects of the filter order on the quantization error.

A simulated result is shown in Figure 2, where quantization error is plotted against variable filter order. The quantization is performed by rounding the infinite precision samples to the closest fixed point value. The quantization error increases with increased filter order, since the highest power index in the filter polynomial is the most affected by the rounding. When the quantizer is increased with one more bit in the precision, the error is reduced by approximately 6dB as would be expected.

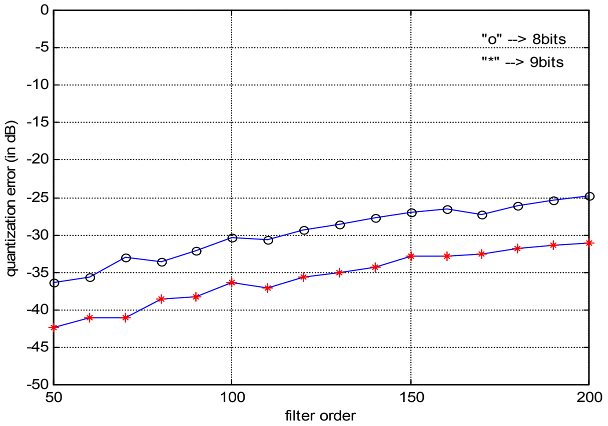

Figure 2 Effect of quantization at different filter order

The lower order filter provides better dynamic range than the higher order for eight and nine bit quantizers. This fact is also evident in Figure 2. At lower filter order of fifty, accumulative quantization error is around -43dB and at higher order of 200, it is -31dB. The 12dB difference is equivalent to two additional bits in quantization.

Non-linear effects of the quantization can be reduced using a smaller filter order in the modulator. Since the cascaded design comprises a filter of lower order, compared with the single model, it introduces less quantization error than the single stage.

2.3 Quantization and Word length

The dynamic range of the scaled filter depends on the number of bits assigned to the quantizer. For maximum signal power, the quantizer range should be equal to the signal magnitude. An FIR filter with filter variance σ_f^2 and quantization noise variance σ_n^2 has a signal to noise ratio of

$$SNR = 10\log\left(\frac{\sigma_f^2}{\sigma_n^2}\right) \tag{1}$$

This expression can be used to estimate the appropriate word length for the FPGA implementation. A comparison of SNR versus word precision using the above expression has been calculated and is shown in Figure 3. From this graph it is evident that for each bit added to the word length, there is approximately a six decibel improvement in the SNR. For a higher precision level, a system can still be implemented, but at the cost of increased FPGA logic resources.

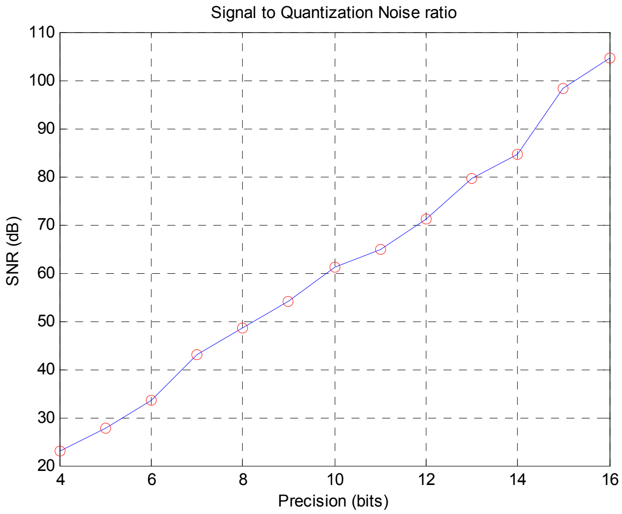

Figure 3 SNR and precision level.

3 QUANTIZATION EFFECTS ON DIGITAL BEAMFORMING

The quantization of infinite precision samples into fixed word length degrades the phased signals. As was discussed in the previous section, the use of more levels for higher precision decreases the quantization error at the expense of larger hardware resources. For a reduced precision level, quantization error is spread to the main beams and to the grating lobes as well. In this section we present effects of quantization on beam resolution and associated grating lobes.

3.1 Quantization effects on Beam Pattern

Phased signals have similar quantized effects on main beam resolution as the filter samples. However non-linearity arises in the sidelobes since the quantizer is not of adequate resolution to represent small changes that affect the sidelobe levels.

In order to investigate the quantization effects, an example is presented with fixed word length delay samples. The coefficients of the time vector are quantized into four and ten bits; the increased number of bits will reduce the quantization effect. For an actual design the fixed bit width will depend on available hardware resources. The quantized beam in Figure 4 shows that a four bit fixed number does not adequately represent the beam pattern and thus introduces quantization noise. The ten bit numbers will also introduce quantization error, but at a lower level as shown in Figure 4(b).

As can be seen from this simple example, the four bit quantization compromises the sidelobes at the -20dB level, while the ten bit quantization provides a reasonably faithful reconstruction of the theoretical sidelobes at this level. Therefore we conclude that for the 14 bit DAC of the proposed system, the sidelobe level will be essentially unaffected by the quantization at the -20dB level.

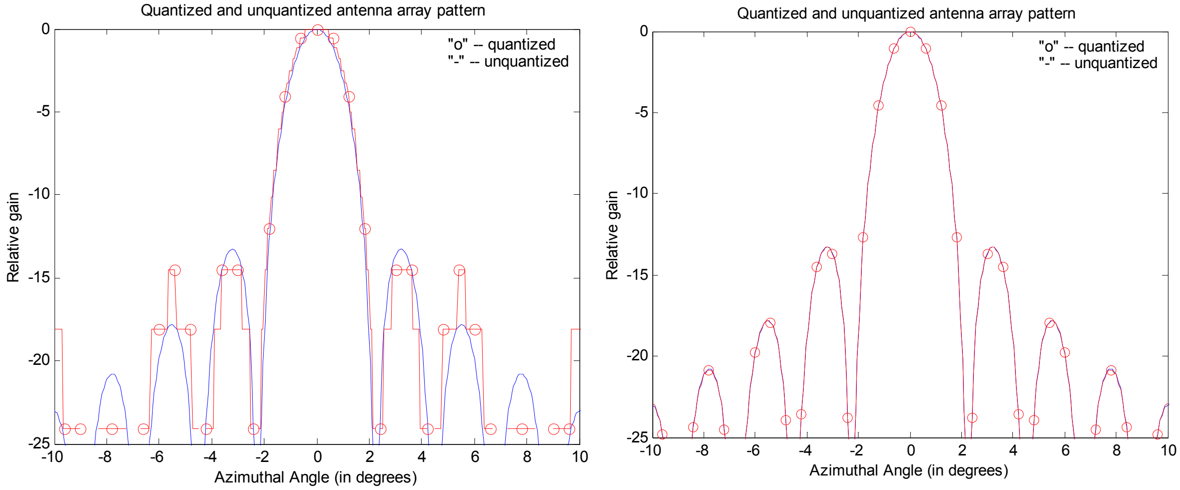

Figure 4 Quantization effects on beam pattern when samples are rounded to (a) four bit precision and (b) ten bits.

3.2 Sensitivity of Sidelobe Levels to Quantization

Quantization causes gain errors in sidelobe levels. Higher resolution in quantization introduces lower quantization error. The graph in Figure 4 shows that the four bit samples result in a quantization error which reduces the first sidelobe gain while producing a gain error in the second sidelobe. The quantization error changes the dynamic range of the grating lobes and degrades the adjacent beam resolution for multiple beam systems.

A simulated graph is displayed in Figure 5 to demonstrate non-linear behavior of the quantizer in the sidelobe resolution. For a lower order quantizer, the quantization step is not perfectly matched with the sidelobe levels. For the first sidelobe, the quantized resolution is less than the infinite precision case, although it approaches the floating point value with increasing quantized levels. Figure 5(a) shows that for a three bit quantizer, the first sidelobe resolution is at -18dB, while at ten bits it approaches the infinite precision value of -13.5dB. Unlike the first sidelobe, the second sidelobe exhibits higher resolution error at a lower precision level, since the quantizer can not represent the dynamic range adequately. Again, quantization error reduces with an increase in the number of bits.

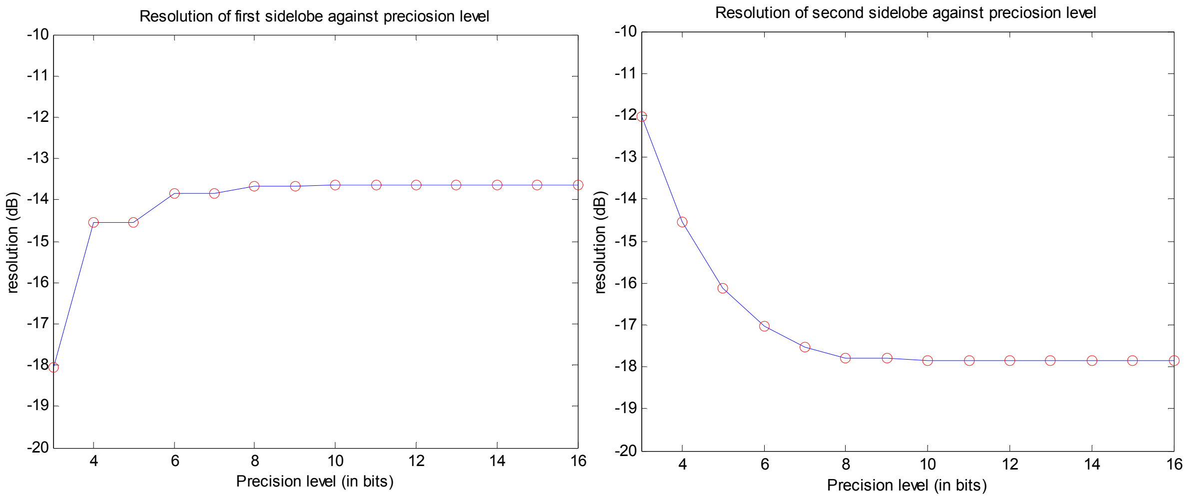

Figure 5 Quantization effects on sidelobe levels, (a) first sidelobe (b) second sidelobe.

4 DISCUSSIONS

In this paper, effect of fixed word lengths on signal upconversion and phasing for the digital TIGER system have been discussed. For the digital upconversion process, the quantization error can be described using pole/zero filter and frequency response plots. Filter resolution and stop band attenuation are degraded when quantization is introduced. For an increase in filter order, the quantization error increases as the highest order in filter polynomial is effected the most. To overcome this limitation, the number of precision levels of a quantizer can be increased, however this will require increased logic resources for FPGA implementation. Quantization effects in phasing are more complex than in the filter quantization since finite precision degrades the sidelobe resolution. For lower precision levels, the quantization error exhibits non-linear behavior in the second sidelobe. The quantization error is higher for lower precision levels. In order to overcome these non-linear effects, a precision level of more than eight bits is required. Performance of the proposed digital system will be effectively unaffected by the fixed word length limitations since a system data bus of at least 14 bits is suggested.

5 REFERENCES

[1] A. B. Sripad and D. L. Snyder, 'A Necessary and Sufficient Condition for Quantization Errors to be Uniform and White', *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. ASSP-25, No. 5, October 1977, pp. 442-448.

[2] T. Salim., J. Devlin., J. Whittington., 'Investigation of multirate techniques for digital generation of Transmitter signals for TIGER Radar', *IEEE INMIC Conference,* Lahore Pakistan, December 2001.

[3] T. Salim, J. Devlin, J. Whittington, G. Tobin 'Design and Realization of HF Up Converter using Multistage Polyphase Filter', *Fourth International Conference on Modelling and Simulation (MS02),* Melbourne, Australia, November 2002.