

Design of a FIR filter using a FPGA

G. Comoretto¹

¹Osservatorio Astrofisico di Arcetri

Arcetri Technical Report N° 5/2002
revision 2.1
Firenze, November 2002

Abstract

In the hybrid correlator proposed for ALMA, a large fraction of the total complexity and cost is represented by the digital filterbank. In this report, an alternative design for the filter unit is presented. This includes a digital baseband converter, to select an arbitrary portion of the input band, and a two-stage filter. The baseband converter allows the selection of a 62.5 MHz portion of the input band arbitrarily positioned in the 2 GHz IF bandwidth. In this way, it is possible to cover it with partially overlapped sub bands, allowing a better control of edge effects. The filter section is composed by a 128-tap coarse FIR filter, used to allow a 1:32 decimation of the input data, and a 64 tap sharp filter, equivalent to a 2048 tap filter operating at the original data rate. The whole filter can be fit in a Xilinx XC2V1500 FPGA (or a corresponding Altera Stratix), with a reduction of a factor 4 in total gate usage.

1 Introduction

In the hybrid correlator proposed for ALMA, a large fraction of the total logic and correlator cost is represented by the digital filter bank. In this architecture [1], denoted in this report as *basic FIR design*, 32 filters, with a number of taps of the order of a few 1024 each, are required to split the input bandwidth into the same number of parallel sub-bands. In each filter, the desired band is chosen using an appropriate filter shape, and then aliased to baseband by resampling the filtered data at a reduced clock rate. For flexibility, it should be possible to select any sub-band in any filter. Reduced bandwidth can be achieved by cascading filters together, at the expense of a reduced number of sub-bands.

No provision has been made for tuning the sub-bands, i.e. each sub-band must fall inside one of 32 equal slices of the input bandwidth. At reduced bandwidth, a similar restriction applies, i.e. the input bandwidth is always divided into N fixed sub-bands (with N a power of 2), and one individual sub-band is chosen among them. No provision is made for fine control of the individual sub-band phase, to implement a "fractional bit shift". Fractional bit shift is provided by adjusting the phase of the sampling clock, using a PLL to generate a small frequency offset in this clock.

In this report a slightly different approach is suggested, with a full digital SSB converter (in VLBA terminology, *Baseband Converter*, or BBC) to convert each sub-band to the required correlator input bandwidth. The circuit is composed of a digital local oscillator (LO), with enough resolution to provide *fractional bit shift* phase correction, followed by a two-stage FIR filter with (almost) fixed coefficients. This approach has the following advantages:

- fractional bit shift in sub-bands is easier to control than in the sampler, where resynchronization problems of the digital stream may arise with a variable-phase clock¹.
- Sub-band stitching can be done by using two extra correlator slices and packing the sub-bands closer to each other. This relaxes the requirements on the filter sharpness (wider transition band), and improves the amplitude and phase response near the sub-band edge.
- Sub-bands can be positioned without any restriction in the input band, giving more observing flexibility. For example, it is possible to perform high resolution observations on several lines arbitrarily placed in the IF band.
- The possibility of placing the sub-band arbitrarily within the main IF band allows for more efficient filter design. In particular, the 2-stage FIR filter, described in this report, becomes possible, with considerable saving in total filter size and cost.

The digital BBC has some disadvantages with respect to the basic FIR filter design. These consist mainly in a worse SNR, due to the extra quantization in the mixer, and in possible intermodulation effects, because of the nonlinearity in the quantization process. However these effects can be controlled [6], and the overall signal degradation can be negligible (1.5% degradation in SNR). With respect to a WIDAR-type correlator[2], a digital LO placed before the bandpass filter has the disadvantage of producing intermodulations and ghost images in presence of strong input lines. These effects can be reduced to an acceptable level by using a more sophisticated, 4-bit representation for the LO sine wave, compared to the 2-bit representation that is sufficient in the WIDAR system. The resulting spur free dynamic range is about 38 dB. that may be inadequate for ALMA. A spur free dynamic range of 50 dB is achievable using a more accurate 6-bit sinewave representation, at the expense of a 50% increase in the first filter size. This has the further advantage of reducing quantization noise in the mixer to a negligible 0.6%.

The design presented in this report is an evolution of the design presented in [8], with corrections and integrations from the hybrid correlator design group. A low-pass design for the filter has been adopted instead of a band-pass. The spur free dynamic range has been increased. Implementation considerations are based on a more realistic clock frequency of 125 MHz, instead of 250 MHz.

The structure of the proposed filterbank is described in chapter 2. The architecture of the BBC is described in detail in chapter 2.2.

¹If fractional bit shift in the sampler is already implemented for the first generation correlator, this advantage is irrelevant.

2 Architecture

The system proposed here is composed of a bank of identical BBCs. Each BBC is composed of a digital LO, and a fixed filter. Filter shape is the same for all sub-bands, and need not to be changed to adjust bandwidth².

The other parts of the correlator (correlator unit and the antenna unit before the filter) is identical to that assumed in the basic hybrid correlator design. The basic structure for the filterbank processing unit is shown in fig. 1

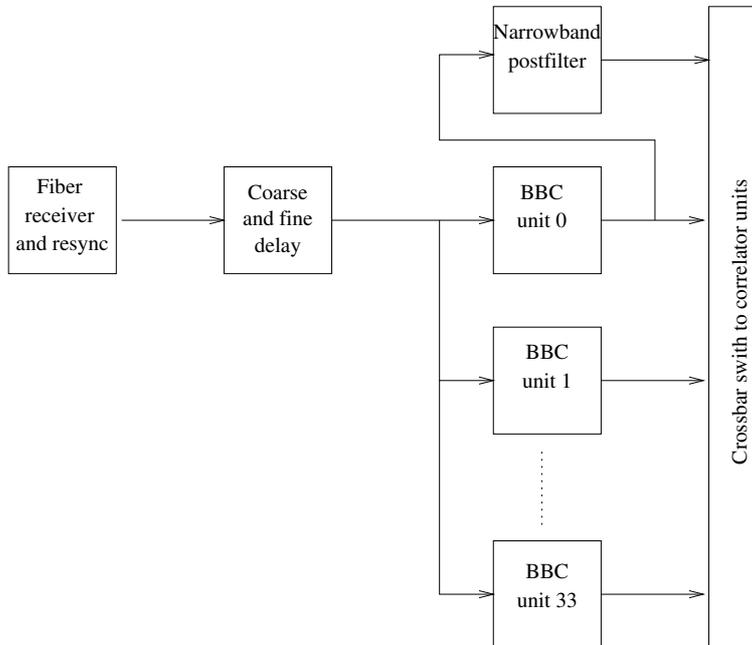


Figure 1: Structure of one filterbank unit of the hybrid correlator. Signal received from the fiber link is compensated for geometric delay, and split into 34 sub-bands. Each sub-band can be freely positioned within the IF. Signal is then transmitted to the correlator units. An extra narrow band filter can be used to further narrow one of the sub-bands.

At full bandwidth, 34 BBCs are used to implement 34 slightly overlapping sub-bands, covering the whole 2 GHz IF bandwidth.

2.1 Reduced bandwidth configurations

Several options are available to implement bandwidths narrower than the full IF. The subject is thus quite complex, and must be treated in the framework of the general architecture of the hybrid correlator. This will be the subject of a next report. In this chapter, we will only make general considerations, to show what the possibilities are.

The simplest way to observe a reduced bandwidth is to use less sub-bands, and cascade together the correlator slices. For example, to observe half the total band, one can use alternate sub-bands and cascade in pairs the correlator units. In this way, the total number of spectral channels is constant, and the resolution increases linearly with the inverse bandwidth B^{-1} .

If the correlator implements some form of channel recirculation, it is possible to reduce the bandwidth of each slice, increasing the corresponding number of spectral channels per slice. The total number of spectral channels increase with B^{-1} , and the resolution with B^{-2} . The second stage filter may use tap

²If the correlator allows channel recirculation, increased filter sharpness is required. This can be accommodated in several ways, cascading the second stage filters or using tap recirculation. This topic will be treated in a separate report

recirculation to increase the total number of taps, thus obtaining sharper band edges. The amount of overlap remains the same, so it is sufficient to increase the number of taps as B^{-1} . No cascading of different filters is required.

Assuming 1024 channels per IF (32 channels per sub-band), one can obtain, *without recirculation*, a maximum resolution of 62.5 kHz (1024 channels over a 62.5 MHz bandwidth) ³.

With a maximum recirculation factor of 8 in the correlator, the number of channels increase with the inverse bandwidth up to a total bandwidth of 250 MHz, for a spectral resolution of $250 \text{ MHz}/8192 = 30\text{kHz}$. Further reduction in the bandwidth results in a linear increase in resolution, up to a resolution of 7.8 kHz for a 62.5 MHz bandwidth. Correlator units always operate at full rate. Increase in correlator resolution is obtained by correlator channel recirculation, and/or by cascading correlator lags.

In principle, one can increase the filter sharpness by:

- cascade the first stage filters from different slices
- cascade the second stage filters from different slices
- recirculate taps in the second stage filter

From the discussion above, what is really needed is only the possibility to recirculate filter taps. Anyway all these solutions are technically possible, producing a filter sharpness (transition region) proportional to the square bandwidth. Their effective usefulness and implementation problems will be discussed in a separate report.

2.1.1 Narrowband post-filter

To implement bands narrower than a single slice (total bandwidth $< 62.5\text{MHz}$), a single narrowband post-filter can be used. If recirculation is used in the correlator, the minimum bandwidth is 7.8 MHz with a resolution of 1 kHz, and the need for a narrowband filter is much reduced.

A symmetric filter with coefficient recirculation can implement in principle arbitrary decimation factors. The correlator shift register is then clocked at a corresponding reduced rate, with all correlation resources cascaded together as a single 1024 channels correlator⁴.

The number of taps is inversely proportional to the output bandwidth, giving a constant number of multiplications per unit time. The design for such a filter will be subject of a further report. In this way, without channel recirculation, spectral resolution increases linearly with decreasing bandwidth up to 1 MHz. Apart from the different implementation, these performances are identical to those of the basic hybrid correlator.

2.2 Filter architecture

To reduce filter size and cost, a two stage filter has been used. The first filter is used to reduce the sampling rate to 1/32 of the input rate (125 MHz), with a passband of 62.5 MHz (the required final passband) and a transition band sufficient to prevent aliasing. Filter output is complex, thus representing a total effective bandwidth of 125 MHz (passband + guard bands). In this way it is possible to have a flat passband response and a high stop band rejection with a very limited number of taps (128 in the proposed design). The second filter operates at a much reduced sample rate, and can thus obtain a given performance with a number of taps approximately reduced by a factor of 32. The 64-tap filter proposed here is thus equivalent to a single pass 2048 tap filter. In this design, stop band rejection is determined by the first filter, while bandpass shape is determined by the second filter.

³The proposed hybrid correlator has 32 channels per sub-band, in full polarization mode (all 4 Stokes parameters computed). For dual polarization mode, the number of channels double, and quadruple for a single polarization mode. This corresponds resp. to 8192 or 16384 channels per polarization over the 8 GHz bandwidth. In this report, we will always assume the full polarization mode, and implicitly consider the cases with less polarizations and more channels.

⁴Using a Xilinx blockRAM to implement delays and to store tap coefficients, decimation factors of up to 512K can be implemented. This corresponds to a bandwidth of a hundred Hz with a resolution of a fraction of a Hz, well below any conceivable application. For decimation factors greater than 32-64 (1-2 MHz BW), however, noise aliasing begins to affect performances.

Tap coefficients can be determined initially by separate optimization of the two stages, using standard algorithms. Then the second filter is modified to compensate for the small roll-off in the passband due to the first filter.

To avoid increasing quantization losses, the second filter uses a many-bit representation for the signal. In the proposed design, 10 bits are used for the signal, and 11 for tap coefficients.

The second filter rejects half of its input bandwidth by a large factor. In this way about half of the folded noise present after the first filter is strongly rejected, and the total folded noise is decreased by ≈ 3 dB. For reduced bandwidth operations, the second filter rejects a higher fraction of its input bandwidth. The total noise contribution comes from the aliased images of the second filter passband, and is thus constant, independently of the final bandwidth and decimation.

2.3 Specifications

In this chapter, the main specifications for a digital BBC are given. Specifications derive mainly from hybrid correlator specifications, as detailed in the Phase 2 proposal for Work Element 6.325.2570. A summary of the proposed specifications is listed in tab. 1.

Each BBC must satisfy the basic specifications for the ALMA hybrid correlator. Some further specifications deal with the capability of frequency tuning. The bandpass specifications derive from the capability of overlapping the sub-bands.

Input data is given as a 32 time multiplexed stream, with an input frequency of 125 MHz, for a total data rate of 4 GS/s (2 GHz bandwidth). Data is represented with 3 bit, using any convenient code. An input format using 16x multiplexed data at 250 MHz is also possible, implementing a 1:2 demultiplexing stage internally to the BBC.

Output data is given as a non-multiplexed data stream, at a clock of 125 MHz, using a resolution of 3 or 4 bit, for a bandwidth of 62.5 MHz. Data is rescaled, using uniform spaced thresholds, to guarantee maximum efficiency in the correlator.

The output data represents an arbitrary sub-band of 62.5 MHz, SSB converted to baseband, with no frequency folding. The actual bandpass is 30/1024 of the input bandwidth, with guard bands folded around slice edge. For a input bandwidth of 2 GHz, this means that the total bandwidth is 58.594 MHz, (1.953-60.547 MHz), with a guard band extending from -1.95 to +1.95 MHz of each slice boundary. With a 32-channel correlator per sub-band, this corresponds to deleting the first and last channel of each sub-band, keeping the remaining 30 channels. With 34 sub-bands, one obtains 1020 usable channels over 2 GHz, i.e. the first and last 4 MHz of each IF band are not usable.

Reduced bandwidth must be available, to allow for data recirculation in the correlator. A recirculation factor to a factor of 8 must be supported, with a corresponding reduced bandwidth of down to 7.8 MHz. The band shape should scale with the bandwidth, i.e. the guard band is always a fixed fraction of the total bandwidth.

Ripple in bandpass should not exceed ± 0.17 dB. A out-of-band attenuation of more than 45 dB is required for high dynamic range operation. To reduce to an acceptable value the noise folded back in the passband, a stop band rejection higher than 40 dB on average is required. This is implied in the above specification of 45 dB for dynamic range. For a decimation factor of 32, and a stop band rejection of 45 dB, the folded noise level is about 30 dB below the in-band noise, contributing to a 0.1% of the total noise. For reduced bandwidth, the *average* stop band rejection must be higher, around 50 dB. As noted at the end of chapter 2.2, this requirement is automatically satisfied by the two stage architecture, as the second filter rejects most of the folded noise present at its input.

The hybrid architecture can be used to implement a fractional delay correction. The fractional delay is currently implemented using a variable phase clock, with a resolution of 15 ps. If it is possible to set the LO frequency of each band with an accuracy of 3.6 millihertz, the maximum phase error due to the resolution step corresponds to 15 ps over 125 MHz after 1 second. It is necessary to specify both initial phase and frequency with this accuracy, and to update these quantities on the fly, in order to track the model delay.

If fractional bit delay correction is not required, the LO setting accuracy must allow the correct positioning of the overlap in the sub-bands. For a recycling factor of 8, the required overlap corresponds to 1/8192 of the total 2 GHz band, i.e. 0.24 MHz. 14 bit of resolution in the LO are thus sufficient. It is

Bandwidth per baseband	2 GHz
Number of sub bands	34
Bandwidth per sub band	62.5 MHz
Overlap between adjacent sub bands	9%
Recirculation	2, 4, 8x
Bandwidth at max. recirculation	7.8 MHz
Sub band tuning range	± 2 GHz
Bandpass ripple	< 0.2 dB
Out of band attenuation	> 45 dB
Spurious free dynamic range	> 45 dB
Phase switching modes in filter	90° step
Programming time	0.1 s

Table 1: Specifications for the hybrid correlator relevant to the filter unit

not necessary to specify initial phase, it is sufficient to have it reset to a known value (e.g. zero) at the beginning of the integration.

Additional 90° and 180° phase switching should be included in the LO.

3 Implementation

The internal structure of a digital BBC is shown in fig. 2.

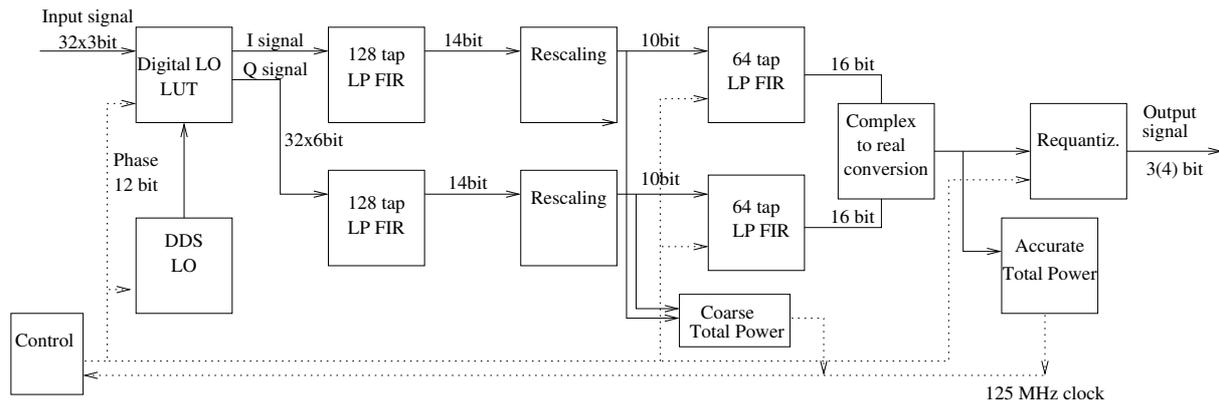


Figure 2: Structure of a digital BBC. The signal is mixed with a quadrature LO, filtered by a first broad filter re-quantized to 10 bit, filtered by a second sharp filter, converted to real representation, rescaled and re-quantized to a final resolution of 3 or 4 bits. Total power meters are used to monitor signal level.

The BBC is composed of a digital oscillator (DDS), a digital quadrature mixer, a first broad band filter and a second sharp band filter. Both these filters are complex, and operate as low pass. The output from the second filter is converted to real by shifting the central part of the bandwidth by $1/4$ of the sampling frequency. The real value output is rescaled and re-quantized, and the in-band total power is measured.

The mixer selects a region of the input band that is subsequently filtered through the broad band filter. This filter has a bandwidth of 62.5 MHz, equal to the final required bandwidth, but is just sharp enough to allow a decimation by a factor of 32. The output of the filter is thus a slice 125 MHz wide, of which 62.5 MHz represents the desired data. Guard bands extend by 62.5 MHz on each side of the passband, and are folded back in the two upper 31.2 MHz of the complex output signal.

The second filter operates at the decimated frequency. It is a half bandwidth low pass filter, and selects the desired central portion of the band, rejecting the wide transition regions of the first filter. It

also compensates for the roll off in the first filter shape, obtaining a final passband ripple of better than 0.2 dB peak to peak.

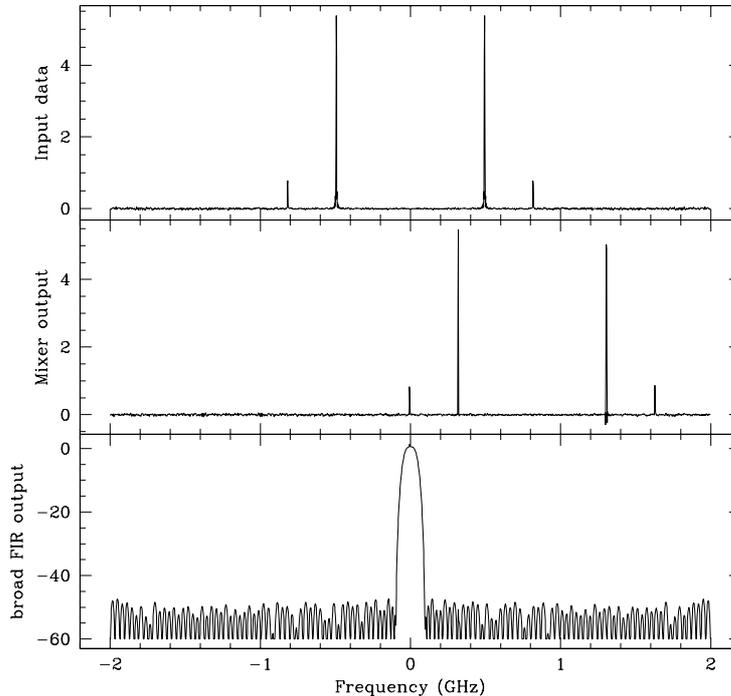


Figure 3: Spectral processing of a simulated signal. From top: (a) Input real signal; (b) Mixer output; (c) Undecimated broad filter output. Graphs have a logarithmic (dB) scale.

Signal processing on a simulated input spectrum is shown in fig. 3 and 4. The complex spectrum of the (real, analog) input signal is shown in fig. 3a. The signal is composed of white noise, a strong out-of-band tone (-20dB), and a weaker (-30dB) in-band tone. The simulated signal is 2.5 ms long. After digitization and conversion using a 6-bit digital mixer, the signal spectrum is shown in fig. 3b. The signal is then filtered by the broad band filter, resulting in the spectrum shown in fig. 3c. This latter shows the rejection of the strong unwanted tones, approximately 45 dB relative to the input level.

The spectrum in fig. 4a shows the signal after decimation. The guard band is heavily aliased, but the central region, containing the signal of interest, is unaffected by aliasing. The sharp filter output is shown in fig. 4b, that is then translated in frequency by 1/4 band and converted to real, obtaining the final real spectrum shown in fig. 4c.

3.1 Local oscillator and mixer

The local oscillator is a time multiplexed DDS. Each sample has an associated time $t = t_i + t_j$, where $t_i = i\tau_m$ is a multiple of the demultiplexed clock period $\tau_m = 1/125$ MHz, and $t_j = j\tau_s$, where j identifies a time multiplexed branch, is a multiple of the sampler clock period $\tau_s = 1/4$ GHz. The clock phase is thus also composed of two parts, $\phi_j + \phi_i$. ϕ_i is generated by a standard DDS register operated at a clock frequency of 125 MHz, set to the desired final frequency ν modulus 125 MHz. $\phi_j = j\tau_s\nu$ is a constant phase offset, different in each time multiplexed branch. The local oscillator is thus composed of two parts. A 36 bit DDS, operating at 125 MHz, generates a common phase for all branches. Each branch has a phase offset register and an independent sine/cosine lookup table to generate the quadrature signals.

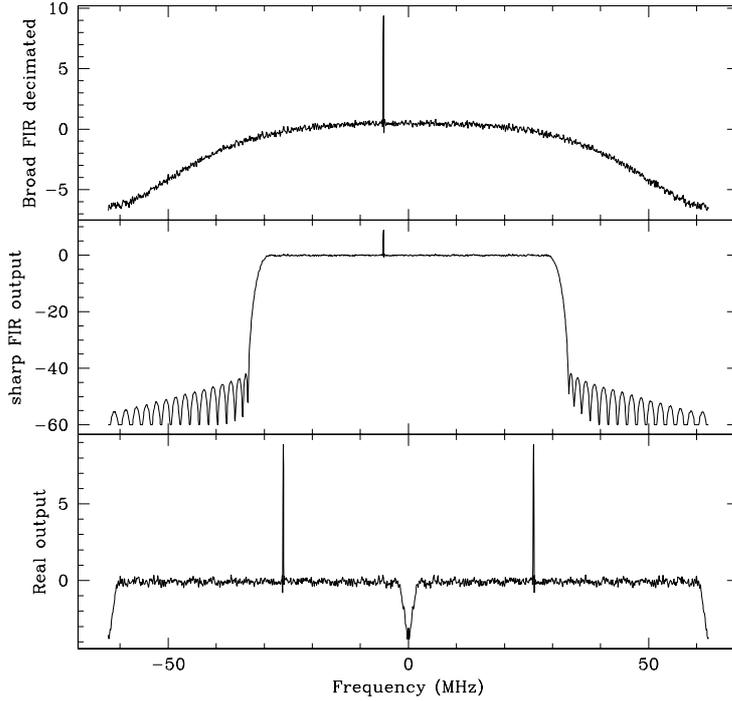


Figure 4: Spectral processing of a simulated signal. From top: (a) Decimated broad filter output; (b) Sharp filter output; (c) Real signal sent to the correlator. All plots are on a logarithmic vertical scale.

The total frequency resolution is equal to the resolution of the DDS, and the frequency span is equal to the DDS total span multiplied by the time multiplexing factor. The particular window of 125 MHz is selected by appropriately choose the phase offsets ϕ_j .

The DDS is implemented as a 3-stage pipelined adder of 12 bit per stage. This implies that any frequency change takes effect with a delay of 3 cycles of the 125 MHz clock. The LO resolution is $125/(2^{12})\text{MHz}= 1.8 \text{ mHz}$. If no fractional bit delay is needed, a single stage DDS, with 12 bit resolution, is sufficient, giving a frequency step of 30 kHz. The saved complexity is however marginal, and the extra resolution may be useful for other purposes.

The mixing operation is done in a lookup table, where the input signal magnitude selects one of four 6-bit sine waves, of corresponding amplitude. Each sine is represented using 512 points (for half turn), for a total phase resolution of $1/1024$ turn. The result sign is computed XOR-ing the phase MSB and the input signal sign. It would be possible to further reduce the LUT size using a one quadrant sine representation, an reducing the phase to the first quadrant. For Xilinx FPGA's, where the RAM size is 2048 bytes, this is not necessary, but could be useful in FPGA's with smaller RAM blocks.

The spur free dynamic range (SFDR) due to LO harmonics is determined by two factors, the number of bits in the result and the phase quantization step. With the values adopted in this design, the resulting sinewave, weighted for the input signal statistics (see [6]), gives a SFDR around 50 dB. The harmonic content of the reproduced sinewave is shown in fig. 5. The spectrum shows spurs corresponding to the phase quantization step (harmonics $1024N$), but these are below the noise floor due to the 6 bit quantization. For comparison, the SFDR for a 4 bit mixer is around 32 dB, and a phase quantization of 256 points would be adequate. A 5 bit mixer gives a SFDR slightly less than 40 dB.

The mixer slice conceptual schematic is shown in fig. 6. Each parallel sample is split in sign and magnitude, and the 2 bit magnitude is combined with 9 bit of phase to address a $2048*5$ LUT RAM.

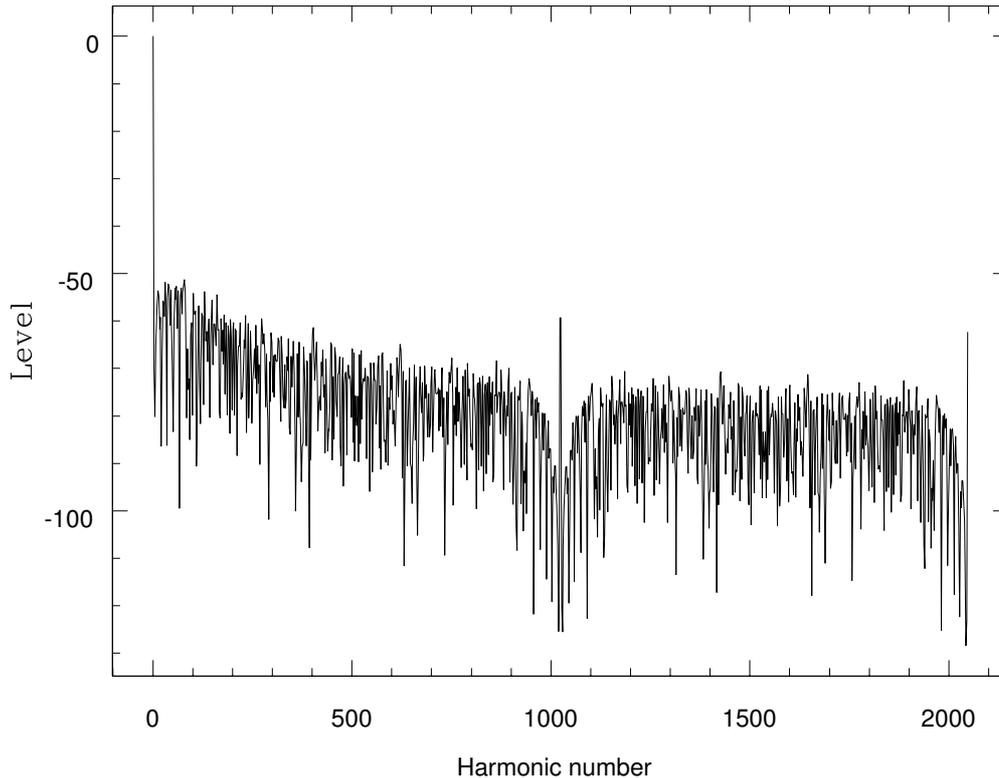


Figure 5: LO harmonic content. LO signal harmonics are around 50 dB. Spurs at harmonic number 1024 and 2048 are due to the phase quantization step.

The RAM output represents the magnitude of the converted value. The sign is independently processed in a separate LUT. Using a dual-port RAM, both the I and Q values for a given sample can be obtained, with a resolution of 1/1024 turn and 6 bits of accuracy⁵.

The phase of each slice is offset by a programmable value, to take into account the relative delay with respect to the 125 MHz clock. This also resolves the ambiguity due to having the local oscillator operating at 125 MHz. In this way, the total tuning range for the DDS is extended by 32 times, to ± 2 GHz (a total of 40 bit resolution over the usable 2 GHz band). The DDS and the phase offsets in each slice must be reprogrammed together every time the frequency is changed.

RAM values are loaded at startup. Since they represent a table with a fixed representation of sine and cosine values, common to all the 32 RAM, and need not to be changed when the filter is reprogrammed, they can be loaded from an external ROM, or included in the FPGA configuration.

3.2 Broad band filter

The I and Q mixer outputs are processed by two identical FIR filters, with 128 taps each. The filters reduce the bandwidth of the complex signal and resample it. The filter output represents a complex signal with both a sampling rate and an effective bandwidth of 125 MHz. Filter conceptual schematic is shown in fig 7. Each tap in the filter is implemented as two lookup tables, one for the top 4 bits, and one for the remaining two bits. To save space, this latter is grouped in such a way that a 4 bit LUT implements two taps (see Xilinx technical note in [12]). The result for the two branches is added together, taking into account the necessary 4 bit shift.

⁵Up to 10 bits are possible giving the available RAM sizes, but this increases the complexity of the FIR filter.

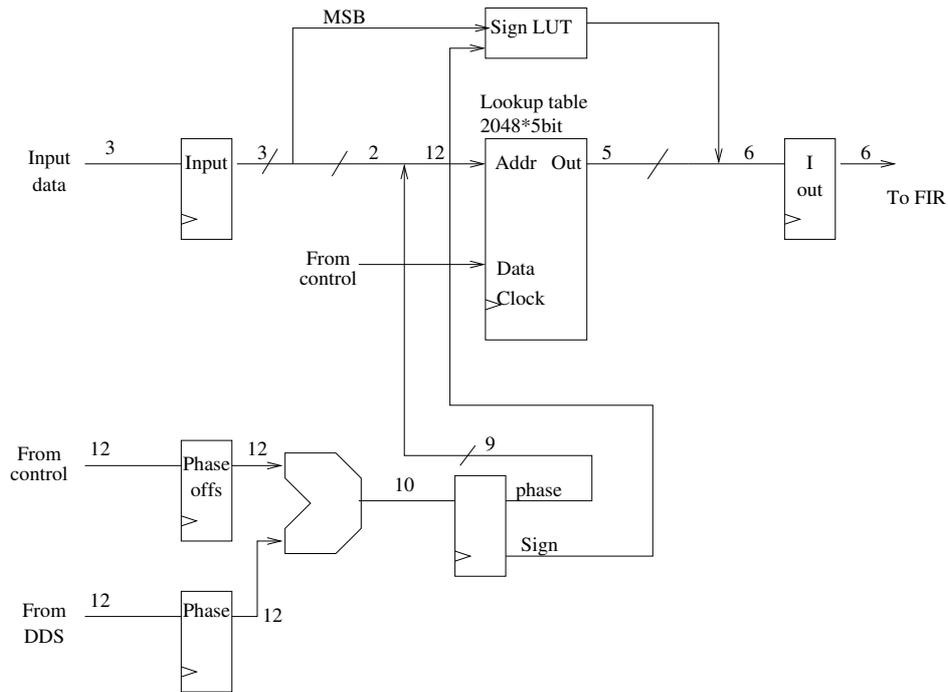


Figure 6: Design of a mixer slice. Each parallel sample is processed in a similar way, with an appropriate value for the phase offset. Multiplication and sine/cosine generation is performed in a LUT memory, loaded at startup.

The filter type is low pass, with a passband of $1/32$ of the input band and guard bands of $1/32$ (folded) on each side. The filter is fixed (the fixed coefficients loaded with the design), since tuning is performed using the digital LO. Avoiding the necessity of dynamically changing tap coefficients, the circuit complexity can be significantly reduced.

The filter has been designed using the Remez algorithm. This produces an equiripple design, but with different ripple in the passband and stop band. We choose to have a better stop band rejection at the expense of a worse passband equalization. The relatively large roll-off (1 dB) is compensated in the second filter. Larger stop band attenuation can be achieved at the expense of a larger roll-off. A roll-off of more than 1 dB might however introduce spurious truncation effects.

Using infinite accuracy representation for the tap coefficients, the filter would have a rejection of about 50 dB. Truncating the coefficients to 9 bits (8 bits plus sign) or 8 bits degrades the rejection to something better of 48 and 47 dB, resp. A 8 bit representation appears a good compromise between Filter response for ideal (infinite resolution) and 8-bit coefficients is shown in fig. 8

The filter output is represented as a stream of 10 bit samples (a truncation and justification mechanism is provided to select the 10 more significant bits in any circumstance), with a complex data rate of 125 MHz, for a total bandwidth of 125 MHz (but with only 62.5 MHz of useful bandwidth).

3.3 Sharp filter

The signal is then filtered using a 64-tap complex FIR filter, implemented as two real filters with 64 taps each. Coefficient width is 11 bit, and each multiplier produces a 22 bit output. This accuracy has been chosen mainly because 11 bit multipliers are readily available in Xilinx Virtex FPGA, but lower resolution is acceptable. Performance degradation is small for coefficient truncation up to 8 bit.

Filter is even, and data from positive and negative lags are summed together to save multipliers. Since only alternate outputs need to be computed (see chap. 3.4), each multiplier is run at 125 MHz, and computes odd and even taps on alternate clock cycles (with coefficient recirculation) to give one output

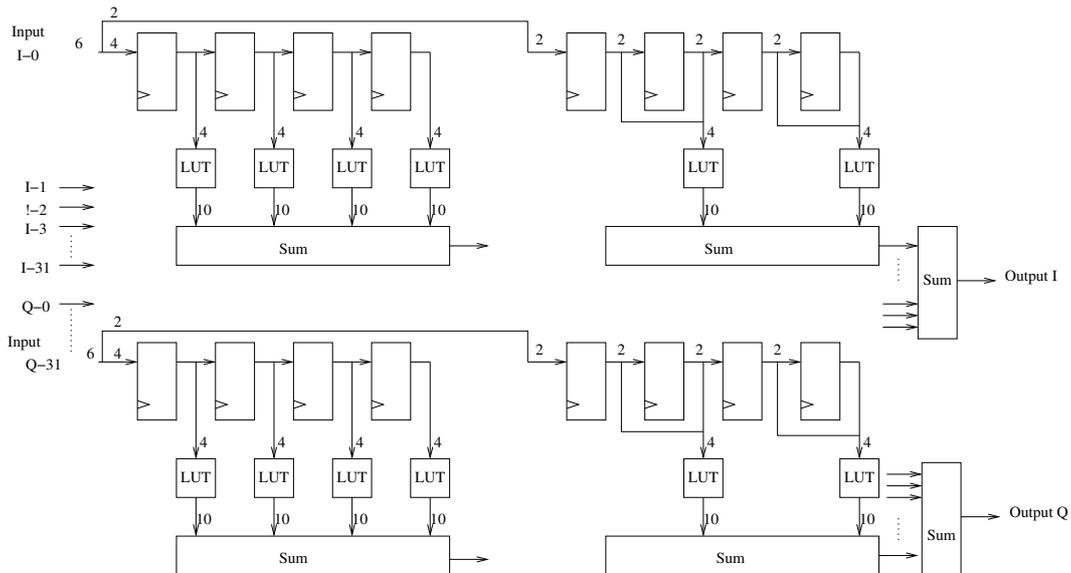


Figure 7: Coarse FIR schematic. Signals from I and Q mixers are multiplied by coefficient taps in LUT tables. Input is from 32 time multiplexed streams, output is to 2 (I and Q) streams.

sample at 62.5 MHz data rate. Filters produce an output on alternate 125 MHz clock cycles, i.e. real samples on even cycles and imaginary samples on odd cycles.

In this way, only 16 multipliers are used for each filter, 32 in total. The filter output is 10 bit, at a data rate of 62.5 MHz, with alternate I and Q samples. The block schematic for one filter is shown in fig. 10.

Different tradeoffs are possible for the filter shape. It is possible to obtain an equiripple design with ± 0.2 dB of in-band ripple and an uniform rejection of 45 dB on the unwanted section of the sub-band. With a less equiripple design, it is possible to obtain a better in-band flatness (± 0.1 dB), at the expense of a lower rejection (-42 dB) on the first sidelobe adjacent to the passband, affecting the first useful channel.

Using 8 bit coefficients degrade slightly the performances. The in-band ripple increases by 0.02 dB and is less regular, but is always less than ± 0.1 db. The first 4 sidelobes after the transition band have an amplitude of -44 dB, instead of -46 dB.

Total band shape (of both coarse and sharp FIRs), for 11 bit coefficients obtained with least square minimization, is shown in fig. 9 left. Sharp FIR provides additional rejection over 1/2 of the total bandwidth, further reducing aliased noise contribution. Shape around the passband is shown enlarged in fig. 9 right, with nominal, guard and passband indicated by ticks above the plot. The horizontal scale in this plot has been expressed in channels, assuming 32 spectral channels per sub band, to better evaluate the transition region effects.

For reduced bandwidth, the output data rate is reduced by the fractional bandwidth, and the number of taps is increased by the same factor. The total number of multiplications is therefore constant. The filter architecture has been designed for a decimation factor up to 16, corresponding to a bandwidth of 1/8 the nominal one. The delay line is implemented using the 16-bit distributed RAM resources as FIFO's. A 16 bit RAM holds the coefficients for each tap. A last-in-first-out (LIFO) buffer must be used at the folding point of the delay chain, to fold back samples before inserting them in the reverse delay branch.

Coefficients must be reloaded every time bandwidth is changed. If the band shape scales exactly with the bandwidth, it would be possible to use a fixed coefficient table, using one coefficient every N for larger bandwidths. Unfortunately this is not our case, because the filter shape must compensate for the first filter roll-over, and thus does not scale with the bandwidth.

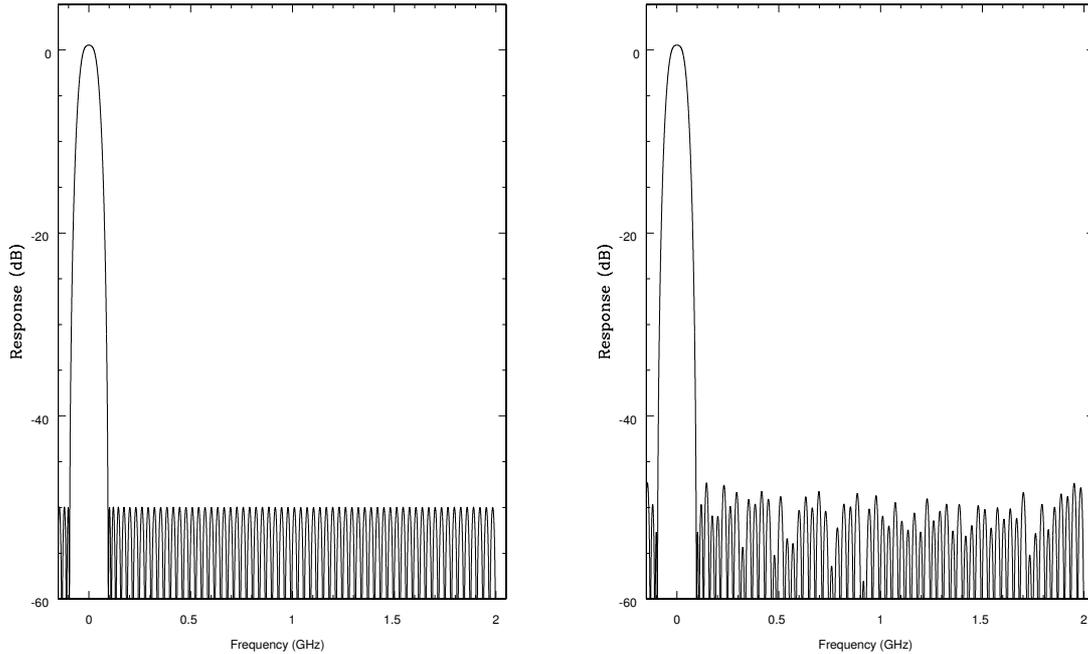


Figure 8: Coarse FIR passband. Left plot is for infinite resolution tap coefficients, right plot for coefficients rounded to 8 bits

3.4 Output section

The complex output from the filter must be converted to a real signal, and requantized to the final 3-bit value used in the correlator.

Complex to real transformation is performed translating the output of the sharp filter in frequency by $1/4$ of the sample frequency. If C_n is the filter output at sample n , the real signal S_n is given by $S_n = \text{Re}(C_n \exp(2\pi i n/4))$. The exponential assumes cyclically the values $(1, -i, -1, i)$, i.e. has only integer real or imaginary components, and can be simply implemented by the circuit shown in fig.11.

The circuit selects alternatively the I and Q output, and thus only alternate values need to be computed by each filter branch. This is exploited in the filter implementation, saving half the multipliers needed.

Filter output statistics is collected by a digital total power meter, implemented using a hardware multiplier. DC offset is also monitored, and optionally removed.

The filter output is then multiplied by a programmable rescaling factor, and truncated to 3 or 4 bits. In this way, uniform quantization levels with arbitrary spacing can be generated. The rescaling factor is programmed by external logic using the informations gathered by the total power module.

3.5 Considerations on FPGA resource usage

The filter described in this report can be implemented in one of the last generation FPGA's available through Altera (Stratix family) or Xilinx (Virtex 2 family). In this report I will analyze a possible implementation on a Xilinx XC2V1000 and XCV1500 Virtex chips (cost around 200\$ and 300\$ resp. for moderate volume). There are no technical stopovers for using an Altera FPGA, and no attempt has been made in comparing the two implementations.

The Virtex2 family is rated for operation up to 350 MHz, but most internal resources cannot work above 250 MHz. At this frequency it is difficult to implement adders longer than 12-14 bit, and block multipliers can handle signals of a maximum of 10-11 bits. Even at this frequency, however, routing delays may prove the design impossible to operate. Power dissipation and thermal problems are also

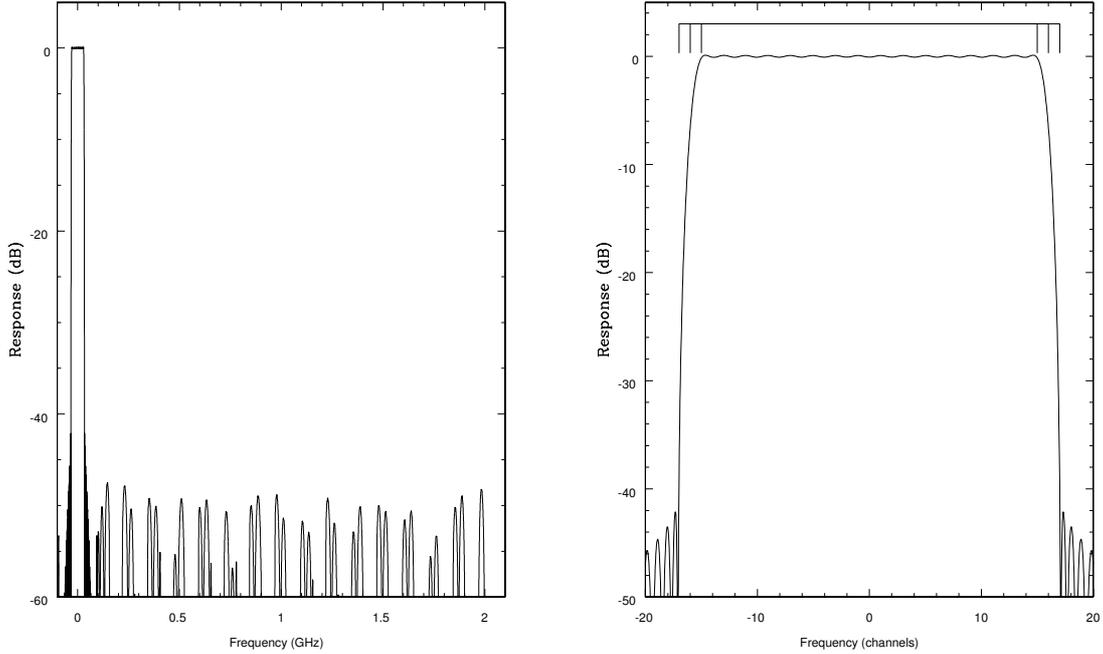


Figure 9: Global filter passband (coarse and sharp FIRs. Left plot is for whole band (frequency scale in GHz), right plot is a zoom around the bandpass (frequency scale in correlation channels, 32 channels per sub band). Ticks above the band indicate the guard band (wider), nominal (1/32 of IF, intermediate), and pass bands (30/32 of the nominal band)

more severe at a higher clock frequency. A more conservative clock frequency of 125 MHz has thus been adopted.

Extended simulation, both using software tools and an hardware demonstrator, may show the feasibility of running some sections at 250 MHz, This would further reduce FPGA usage, but has not been considered here.

The most resource intensive part of the circuit is the first FIR. FIR taps are implemented using lookup tables (LUTs), and 15 LUTs are needed for each tap. With 2 FIRs of 128 taps each, this requires 3840 LUTs. Approximately the same number of LUTs are required for the adder chain, totaling 7680 LUTs.

The digital mixer requires one RAMBLOCK and ≈ 45 LUTs for each input stream, and the DDS requires about 150 LUTs. Therefore the LO/mixer requires 1600 LUTs.

The second FIR filter is based on block multipliers. A total of 32 multipliers are required (of 40 available), while other 3 multipliers are used by the rescaler and total power circuit. Each section of the second FIR filter requires a total of about 100 LUTs, to implement 4 taps. A total of 32 sections are required (128 taps, 64 times 2), for a total of 3200 LUTs.

Control logic, total power and re-quantization circuitry probably does not require more than 200 LUTs.

The grand total is therefore 12700 LUTs, more than the total number of 10240 available LUTs in the chip.

If the filter taps need not to be changed (bandwidth selection can be obtained by tuning the LO, narrower bandwidth can be implemented reducing the number of parallel slices processed, and for very narrow bandwidth an extra filter operating at reduced clock rate can be considered), it is possible to reduce the number of LUTs used in the first filter. Half the taps can be represented using only 4 bits, instead of the 7 used in the central section of the FIR kernel. This translates in a saving of approximately 3000 LUTs. Also in the second filter about half the taps can be represented with 4 bits less than in the central core, but in this case the total saving is less than 500 LUTs, and filter re programmability is lost.

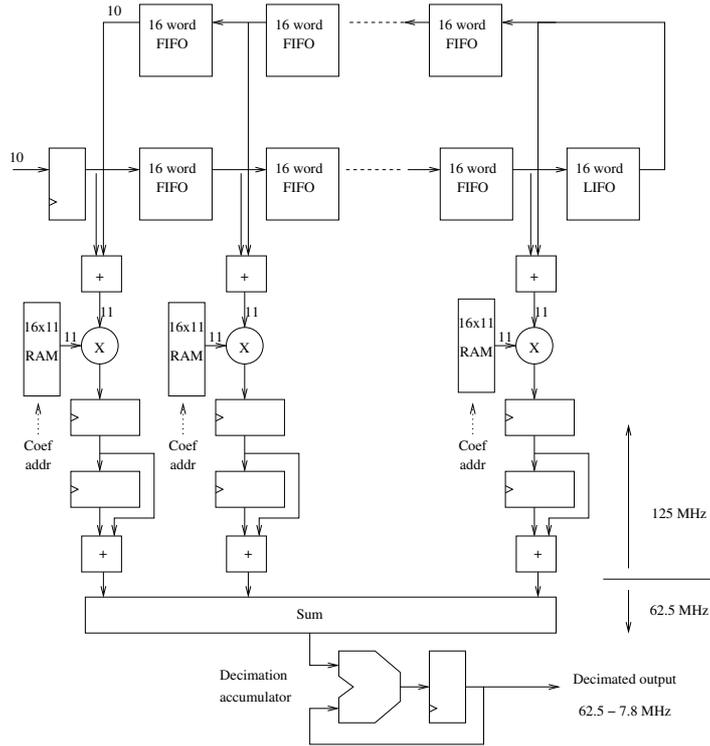


Figure 10: Schematic of the second (sharp) FIR. Only one time multiplexed slice is shown. Delay line is folded back to exploit filter symmetry. O/E is a signal to distinguish between odd and even clock cycles.

With these modifications, one obtains an unrealistic final occupation close to 100%. It is still possible to fit the whole design in a single FPGA if a 4 bit digital mixer is used, with reduced (32 dB) spurious free dynamic range. The design fits reasonably in the larger XC2V1500.

Another possible approach would be to split the design in two parts, one for the mixer and the broad filter, and another for the second filter. The first FPGA would have an occupation of 75%, that guarantees a good routing. The second FPGA would have a very low occupation, due to the heavy use of the multiplier resources. Probably multiplier use is more efficient for Altera FPGA's, and one can think at a mixed implementation, with Xilinx for the first stage and Altera (implementing many filter channels in a single FPGA) for the second stage. This would also make easier to cascade these filters in reduced bandwidth configurations.

Using larger chips, more than one channel can be implemented in a single FPGA. The larger chips currently available are the XC2V8000 (cost around 7500\$), that would host 5 channels (at 50% occupation, limited by number of multipliers). At the current rate of growth of FPGA size, it appears feasible to host all 34 SSB in a single unit within a few years. With the current available devices, 7 FPGA's would be required for the whole filterbank, but the more economic solution would be to use 17-34 smaller devices.

4 Conclusions

A two-stage digital filter with an equivalent number of 2048 taps can be implement in a single XCV1500 field programmable gate array. A digital LO and SSB converter can also fit in the same chip. A 32-channel digital filter can be reasonably fit into 7 of the larger Xilinx chip available today, even if this is not the more economic solution. Such a design would allow for a simplification and cost reduction of the second generation correlator filter board. A digital LO would allow for band overlapping, and for fractional delay compensation.

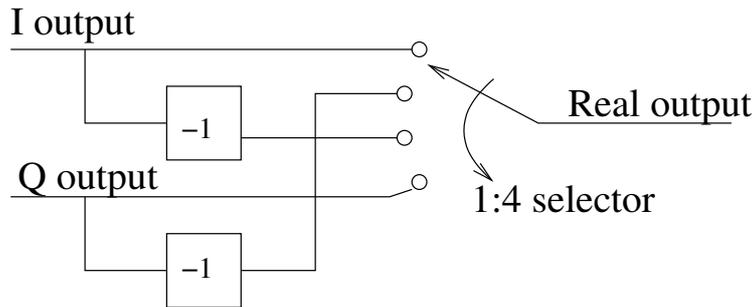


Figure 11: Complex to real converter. I and Q samples from the complex filter are fed with both signs to a 1:4 selector. The input signal must have a total bandwidth of half the useful range (from $-f_s/4$ to $+f_s/4$). Output has the same data rate and total bandwidth, translated to positive frequencies only (from 0 to $+f_s/2$).

References

- [1] A. Baudry, A.W. Gunst: "ALMA Filter Bank Specifications and Delay Tracking", ASTRON report (in preparation) (2001)
- [2] B.R. Carlson, P.E. Dewdney: "Efficient wideband digital correlation", Electronics Letter, IEEE, **36-11**, 987 (2000)
- [3] B.R. Carlson: "A Closer Look at 2-Stage Digital Filtering in the Proposed WIDAR Correlator for the EVLA", NRC-EVLA Memo 03 (2000)
- [4] B.R. Carlson: "Refined EVLA WIDAR Correlator Architecture", NRC-EVLA Memo 14 (2001).
- [5] B.R. Carlson: "WIDAR Correlator Sensitivity Losses", NRC-EVLA memo 26 (2001)
- [6] G. Comoretto: "A digital BBC for the Alma interferometer", Alma report n. 305 (1999)
- [7] G. Comoretto: "Possible designs for a hybrid correlator", Arcetri Internal Report n. 8/2000
- [8] G. Comoretto: "Design of a FIR filter using a Xilinx FPGA", Arcetri Internal Report n. 4/2002
- [9] R. Escoffier and J. Pisano: "Test Report of the Baseline ALMA Correlator Digital Filter" Alma Report n. 409
- [10] B. Quertier: "Proposal for a Future Correlator Filter Board", ASTRON report (in preparation) (2001)
- [11] Harris HSP43168 FIR filter data sheet
- [12] V. Pashram, A. Miller, K. Chapman: "Transposed form FIR Filters", Xilinx application note 219 Filters

Contents

1 Introduction

1

2	Architecture	2
2.1	Reduced bandwidth configurations	2
2.1.1	Narrowband post-filter	3
2.2	Filter architecture	3
2.3	Specifications	4
3	Implementation	5
3.1	Local oscillator and mixer	6
3.2	Broad band filter	8
3.3	Sharp filter	9
3.4	Output section	10
3.5	Considerations on FPGA resource usage	11
4	Conclusions	13