FP7- Grant Agreement no. 283393 – *RadioNet3*

Project name: Advanced Radio Astronomy in Europe

Funding scheme: Combination of CP & CSA

Start date: 01 January 2012 Duration: 48 month

Deliverable 8.7

RFI mitigation

Due date of deliverable: 2014-01-31

Actual submission date: 2014-18-04
Deliverable Leading Partner: University of Orléans (France)

1 Document information

Document name:	UniBoard2 Firmware design document – RFI Mitigation
Type:	Report
WP:	8 (UniBoard²)
Participants:	Rodolphe Weber (UORL), Cédric Dumez-Viou (ObsParis), Nicolas Grespier (UORL)

1.1 Dissemination level

Dissemination Level		
PU	Public	**x**
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

1.2 Document history

Revision	Author	Date	Modification / Change
1.0	R.Weber	2014-02-04	Initial Version

1.3 Distribution list

ASTRON:	Andre Gunst, Eric Kooistra, Sjouke Zwier, Daniel van der Schuur, Harm Jan Pepping
JIVE:	Arpad Szomoru, Jonathan Hargreaves, Salvatore Pirruccio, Sergei Pogrebenko, Paul Boven, Harro Verkouter
UMAN:	Aziz Ahmedsaid, Ben Stappers
INAF:	Gianni Comoretto
BORD:	Benjamin Quertier, Alain Baudry, Stephane Gauffre
UORL:	Cédric Dumez-Viou, Rodolphe Weber, Nicolas Grespier
MPG:	Guenter Knittel, Reinhard Keller, Gundolf Wieching

1.4 Terminology

Beamlet	complex amplitude for a 200kHz given frequency band and 2 directions
Component	real amplitude for a given frequency band and 1 direction
EMBRACE	Electronic Multi-Beam Radio Astronomy Concept

DB	Data base
DPRAM	Dual Port RAM
FPGA	Field Programmable Gate Array
INR	Interference to noise ratio
IP	Intellectual property
FIFO	First in – First out memory management
ObsParis	Observatoire de Paris / Station de radioastronomie de Nançay
PFB	Polyphase Filter Bank
PLF	Polyphase Filter (part of PFB)
Powerlet	Data structure containing information about the astronomical data
FFT	Fast Fourier Transform
RAM	Random access memory
RFI	Radio Frequency Interference
RFIlet	data structure containing information about the detected RFI
SDM	Science data model
SKA	Square Kilometer Array
UORL	Université d'Orléans
VHDL	Very high-speed integrated circuit Hardware Description Language
Waveform	complex amplitude for a given frequency band and 1 direction

1.5 Reference

[1] A.-J. Boonstra and R. Weber, "RFI mitigation strategies for the SKA concepts which are based on phasedarrays," SKADS report DS4T3 Deliverable 6, Nov. 2009.

[2] G. W. Kant, P. D. Patel, S. J. Wijnholds, M. Ruiter, and E. Van der Wal, "EMBRACE: A Multi-Beam 20,000-Element Radio Astronomical Phased Array Antenna Demonstrator," *IEEE Trans. Antennas Propag.*, vol. 59, no. 6, pp. 1990–2003, Jun. 2011.

[3] A.-J. Boonstra and R. Weber, "RFI Mitigation Methods Inventory," SKADS report DS4T3 deliverable 1, Nov. 2009.

[4] F. Viallefond, "The Alma Science Data Model," vol. 351, p. 627, 2006.

[5] C. Dumez-Viou, "Restauration de sources radioastronomiques en milieu radioélectrique hostile : Implantation de détecteurs temps-réel sur des spectres dynamiques," University of Orleans, 2007.

[6] M. Ren, B. Gestner, and Carlos Maffrand, "Complex Gaussian Pseudo-random Number Generator :: Overview :: OpenCores." [Online]. Available: http://opencores.org/project,complex-gaussian-pseudo-random-number-generator. [Accessed: 18-Apr-2014].

1.6 Content

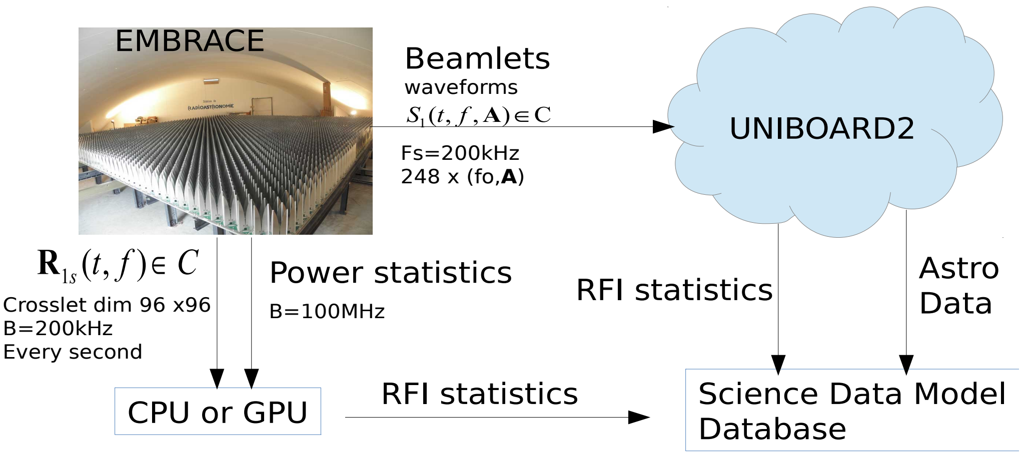

Figure 1: Architecture and main data flow overview. Fs is the sampling frequency, t is a time index, f is a frequency index, ***A*** *is a steering vector, B is the bandwidth,* $\mathbf{R}_{1s}$ *is a correlation matrix.* $S_1(t,f,\mathbf{A})$ *is a complex waveform. For a given instant t, the radio telescope EMBRACE generates 248 different waveforms.*

2 Introduction

We might consider that the RFI challenge in the SKA candidate sites in Australia and South Africa will not be so great. Under this assumption, one basic or recurrent scenario could be to carefully design the analogue parts, taking RFI threats into consideration, but to limit the digital measures to “flagging" [1]. In that case, the digital signal processing resources could be fully dedicated to regular signal processing tasks most of the time and could be partially re-used (scheduled) for observations facing specific RFI issues. However, it would be worthwhile to continuously monitor the quality of the data. Given the extreme sensitivity of the SKA telescope, this task has to be a by-product of the radio telescope it-self (i.e. an auxiliary antenna will not be sensitive enough). So, it would be interesting to implement some detection methods as regular signal processing tasks at station level and core level. The results could be linked to a kind of RFI statistics database or could be attached to the data for flagging.

The purpose of this project is to evaluate this concept. The EMBRACE radio telescope and the UNIBOARD² digital computing board will be used as an hardware framework for this evaluation:

- EMBRACE is a phased-array radio telescope [2]. Based on an LOFAR station firmware, it provides 200 kHz digital waveforms from 2 different directions. These waveforms will be used as data input for UNIBOARD². Besides, EMBRACE by it-self provides also some regular statistics (100 MHz power statistics and 200 kHz correlation statistics) on the processed data. These statistics will be added in the RFI monitoring as well but it is outside of the scope of this UNIBOARD JRA.
- UNIBOARD² will be configured for two main tasks. The first one will be the delivering of 64 bin power spectral estimators for all 200kHz waveforms. The second one will be the data quality monitoring of these waveforms at different time-frequency resolutions (200 kHz, 200/8 kHz and 200/64 kHz). At each frequency resolution, quality monitoring will be

based on several criteria (for example power, kurtosis, cyclostationary, [3]).

Different options are available for the flagged data:

- go-through option: data are just flagged, no modification of the data flow
- zero option: flagged data are replaced by zeros
- Gaussian option: flagged data are replaced by white Gaussian data.

Finally, the spectrum accumulation will be driven in real time by this data quality analysis. These approach is called oriented accumulation. Clean data and flagged data will not be accumulated in the same area. By this way, no data are lost and the end-user has the option to use or not the flagged data.

All these information will be stored in a dedicated database. To manage this amount of information, the database architecture will be set up through a science data model (SDM) structure (ref). This SDM has already been developed [4] for EMBRACE by Francois Viallefond (ObsParis) and it will be extended to include RFI information. Figure 1 shows the whole concept.

In the next section, the different parts implemented in Uniboard2 are explained in more details.

3 UNIBOARD² firmware description

UNIBOARD² receives the 248 complex waveforms S_1(t,fo,**A**). Each complex waveform is channelized by 8 and then by 8 again through 2 successive 8 bin polyphase filter banks (PFB). These PFB are maximally decimated (i.e. the output rate is the input rate divided by 8). At each stage, the waveforms

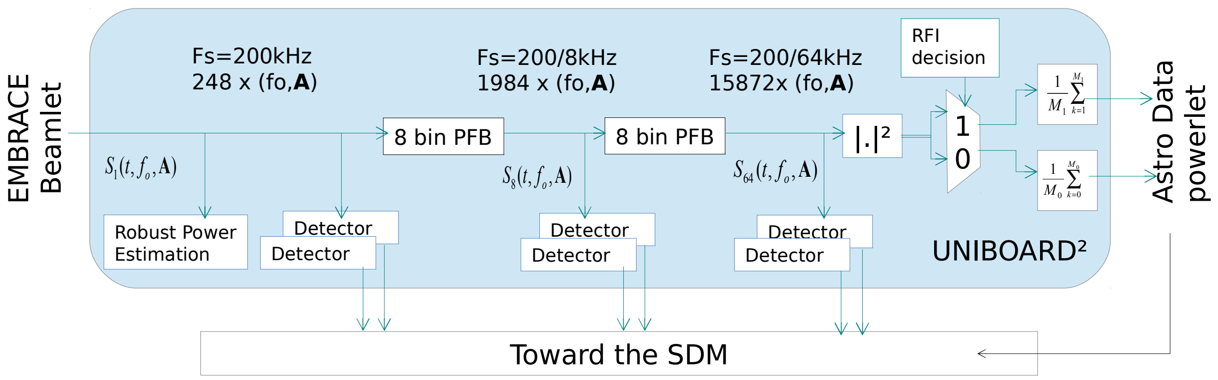

*Figure 2: Overview of the UNIBOARD² firmware. Fs is the sampling frequency at each stage, t is a time index, fo is a frequency index at the given resolution level Δf=1,8 or 64, **A** is a steering vector. Each polyphase filter bank (PFB) splits the input channel into 8 equispaced sub-channels. These sub-channels are decimated by 8. At each stage, detectors analyse the waveforms $S_{\Delta f}$(t,fo,**A**) and provide RFI information toward the SDM database and toward the RFI decision module. This last will control the final spectral accumulation. The accumulation number, M, is constant (M=M0+M1).*

$S_{\Delta f}$(t,fo,**A**) (with Δf=1,8 or 64) enter some RFI detectors. These detectors monitor the data and provide RFI information about data quality. These information are sent to the SDM database but they are also used internally for blanking tasks at each stage and for filtering task during the final spectral accumulation.
Assuming white noise and a narrow band interferer, the Interference to noise ratio (INR) will be 18 dB better in S_{64} than in S_1. The proposed multi-resolution approach is a way to improve the matching between detector time-frequency resolution and RFI characteristics. All these functions (see Figure 2) are now detailed.

3.1 The EMBRACE interface module

EMBRACE data stream is based on the beamlet concept : a beamlet represents the output radio telescope complex temporal amplitudes of two independent steering directions for a given 200kHz band. Thus, each beamlet contains 2 complex waveforms which can be decomposed into 4 components Xre, Xim, Yre, Yim, 16 bits each (see Figure 3). The suffix 're' and 'im' stand for real and imaginary respectively. The prefix 'X' and 'Y' represent two different targets in the sky for a given 200kHz band.

Let us defined S_1(t,fo,**A**) an individual 200kHz complex waveform (Xre+j.Xim or Yre+j.Yim) where 'fo' is the sub-band central frequency, '**A**' is the steering vector and 't' is the time. EMBRACE deliver 248 (124 beamlets per direction x 2 directions) of these complex waveforms. A second RF beam is generated by EMBRACE, but it will not be processed by our design.

The beamlets are provided through two 4Gbyte Ethernet links. The EMBRACE interface module (see Figure 4) will reorder the Ethernet data flow into a single waveform stream. Two FIFO memories (First In First out) of 3968 bits will be implemented through an Altera IP (see Table 1).

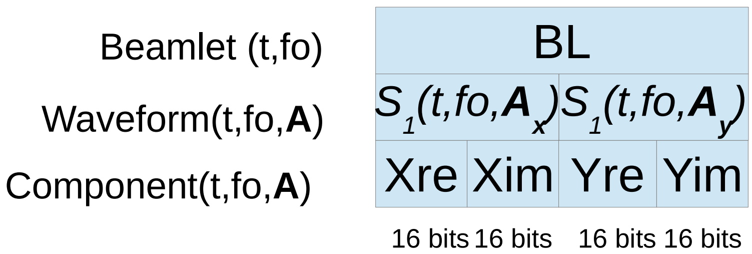

Figure 3: EMBRACE beamlet format and definition. EMBRACE delivers 124 beamlets, 248 waveforms and 496 components. The time index is t, the 200kHz frequency index is fo. These parameters are common to a given beamlet. Only the steering direction is different.

Parameters	**EMBRACE Interface module**
Number of beamlets per link	62
Number of links	2
Number of waveforms per beamlet	2
Number of components per beamlet	4
Amplitude dynamic	16 bits
Memory size (FIFO)	2x3968 bits

Table 1: Embrace interface module parameters and resource estimation.

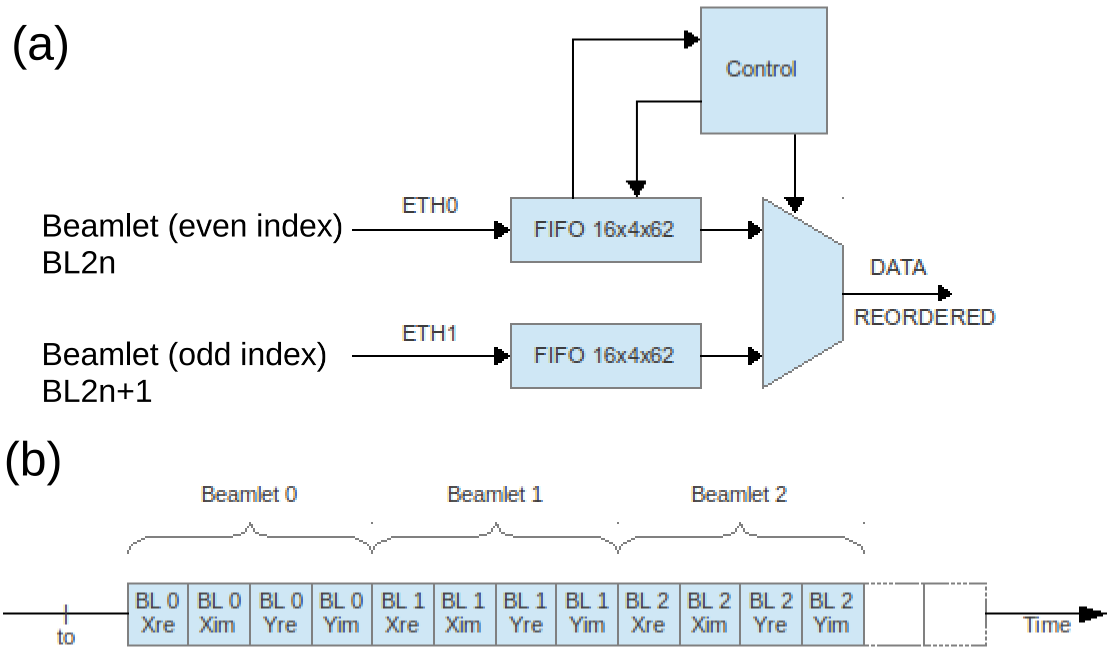

Figure 4: EMBRACE interface module. ***(a)*** *reordering module for 1 RF beam.* ***(b)*** *output of the reordering module. BLx stands for 'Beamlet number x'. (Xre+j.Xim) or (Yre+j.Yim) represent the complex amplitude of an individual 200kHz waveform for a given frequency and a given spatial direction.*

3.2 The polyphase filter bank module

Each PFB takes 8 input samples from a given waveform and provides 1 sample for 8 new sub-waveforms spanning the initial waveform bandwidth. In the following, these sub-waveforms, $S_{\Delta f}(t,fo,\mathbf{A})$, will be called waveforms as well. They will be identified by their frequency index and their resolution level (Δf=1, 8 or 64).

The PFB architecture is divided in two parts (see Figure 5). One is the polyphase filter (PLF) and the second is the Fourier Transform (FFT) applied to the PLF output. Data are resized and reordered (corner turner module) between this two modules.

In our application, the reference filter is 64 coefficient long. Figure 6 gives its impulse response and its filter bank representation in the frequency domain.

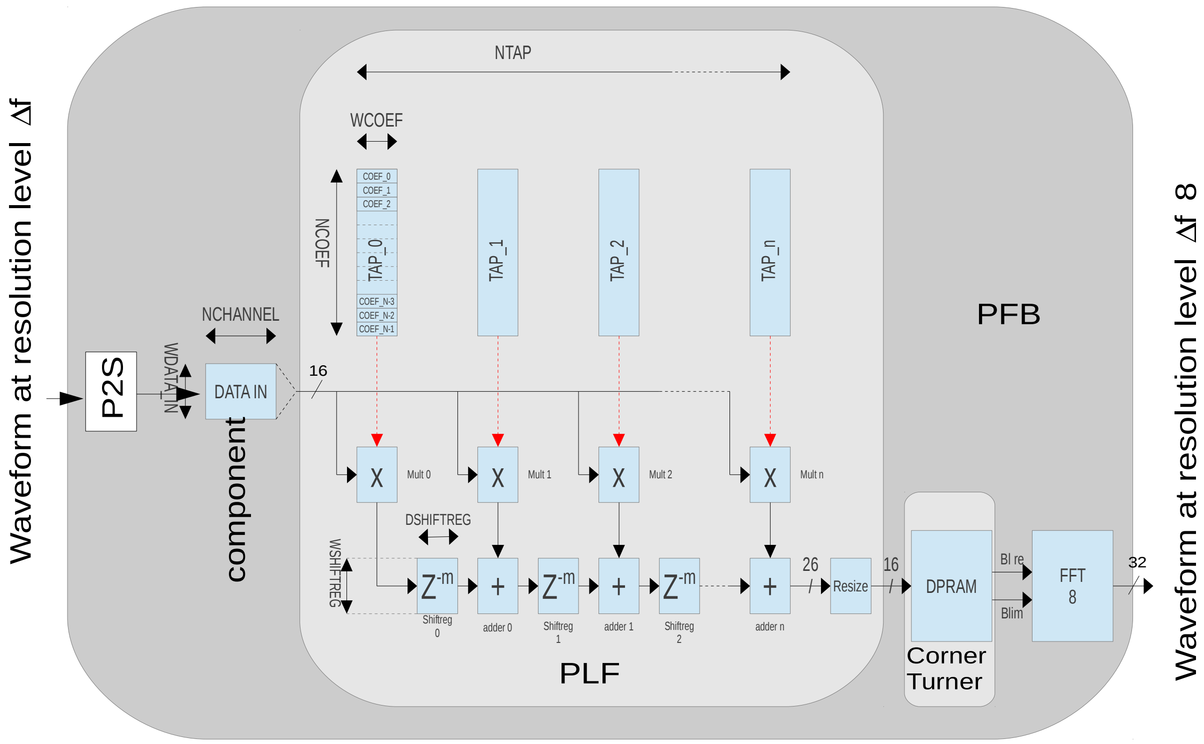

*Figure 5: The PFB architecture which includes the PLF block, the Corner Turner block and the FFT block. m is Nchannel *Ncoeff. The P2S module transforms a waveform flow into a serial component flow (i.e. a complex number flow into a "real then imaginary" flow.*

3.2.1 The PLF module

The PLF architecture is given in Figure 5. All the samples are complex. Table 2 gives the parameters defined for each PLF. For a given component, the data flow is based on the time order.

3.2.2 The FFT module

The FFT applied on each PLF has the same configuration (see Table 3). The FFT will be implemented through an Altera IP. The output of the FFT has a stream compliant with the PLF input (i.e. each FFT channel will be seen as a new waveform). So there is not need to reorganize the data flow after the FFT.

However, the FFT input is not compliant with a PLF stream. It is necessary to reorganize the data flow. Indeed, the FFT module needs a block of 8 continuous complex data corresponding to the filtering of 8 successive complex samples belonging to the same waveform, $S_{\Delta f}(t,fo,\mathbf{A})$. For example, after the first PFB, these samples are spaced by 248*8=1984 other samples from other waveforms.

3.2.3 The corner turner module

This module deinterleaves data before sending them to the FFT module (see Figure 7). This module uses 1 Altera DPRAM. This RAM will be split into 2 independent memory banks. Data reading and writing will swap between these 2 memory banks (see Table 4 for parameters and memory resources).

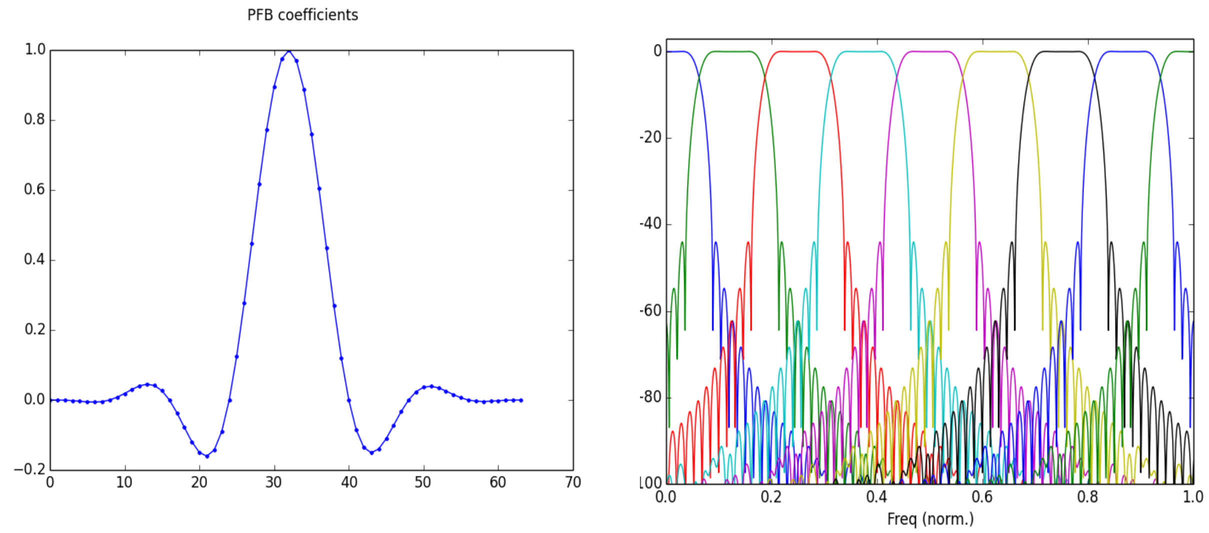

Figure 6: (right) Impulse response of the reference filter (left) Frequency representation of the PFB.

Parameters:	**PLF1**	**PLF2**
Number of taps (NTAP)	8	8
Coefficient number of bits (WCOEF)	18 bits	18 bits
Number of tap coefficients (NCOEF)	8	8
Number of components (NCHANNEL)	496	3968
Data number of bits (WDATA_IN)	16 bits	16 bits
Shift register Width (WSHIFTREG)	26 bits	26 bits
Shift register Depth (DSHIFTREG=m)	3968	31744
Estimated resources:		
DSP resources	7 adders, 8 multipliers	
memory	5777408 bits	

Table 2: PLF Parameters and resources estimation.

Parameters:	**ALTERA FFT**
size	8 bins
Input Data width	16 bits
Twiddle precision	18 bits
Output data width	20 bits
streaming	Variable (natural order)
Estimated resources:	
DSP resources	4 DSP blocks
memory	645 bits

Table 3: FFT Parameters and resources estimation.

Parameters:	**PFB 1**	**PFB 2**
Number of waveforms (N)	248	1984
Number of components (m)	496	3968
Number of tap coefficients	8	8
Component width	16 bits	16 bits
Estimated resources:		
Size of a writing or reading bank	3968 x 16 bits	31744 x 16 bits
memory	2 banks x 63488 bits	2 banks x 507904 bits

Table 4: Corner Turner module parameters and resources estimation

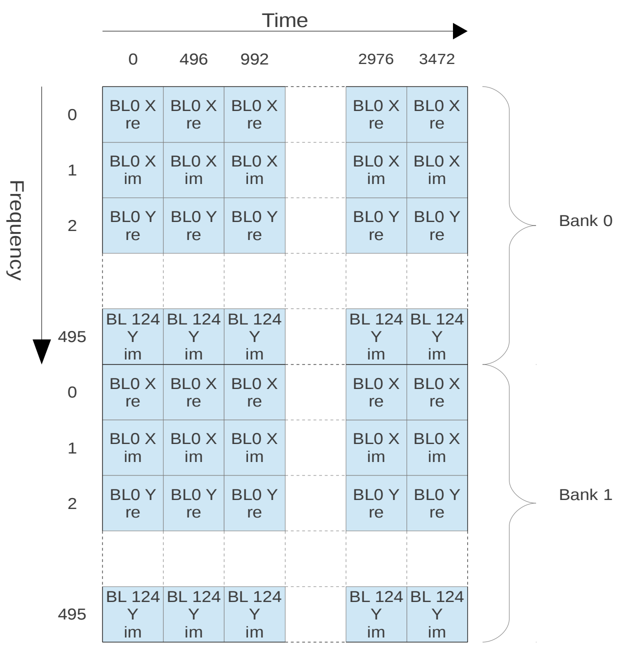

Figure 7: Corner Turner module. Data from the PLF (here PLF 1) are written row-wise (frequency axis) on one memory bank and read column-wise (time axis) on the other. BLx stands for 'Beamlet number x'. (Xre+j.Xim) or (Yre+j.Yim) represent the complex amplitudes of individual 200kHz waveforms for a given frequency and a given spatial direction.

3.3 Oriented power spectrum and Powerlet

After the last PFB, the power for each waveform is computed. According to a RFI flag generated by the RFI decision modules (see section 3.4), the power accumulation is done selectively (i.e. flagged and unflagged data are accumulated separately). In the current implementation (see Figure 8), the RFI decision is binary but multilevel decision could be easily implemented as well.

The Powerlet contains the following information, Powerlet=(t,fo,∆f,**A**, M, Power0,Power1, M0):

- configuration parameters stored in the database for the given acquisition:
  - ∆f is related to the waveform bandwidth
  - A is related to steering direction
  - M is related to the dump time
- dynamic parameters:
  - t is related to the time stamp of a M sample block
  - fo is related to the waveform central frequency (frequency index in the given resolution level ∆f)
  - Power0 is the accumulated clean power
  - Power1 is the accumulated power of RFI flagged samples.
  - M0 is the number of clean samples. M=M0+M1where M1 is the number of RFI flagged samples.

For efficiency purpose, all these information will be dispatched at different places into the Database architecture. Only the dynamic information generated by the accumulators will be put into the Uniboard Powerlet data stream. Configuration parameters will be stored once and for all.

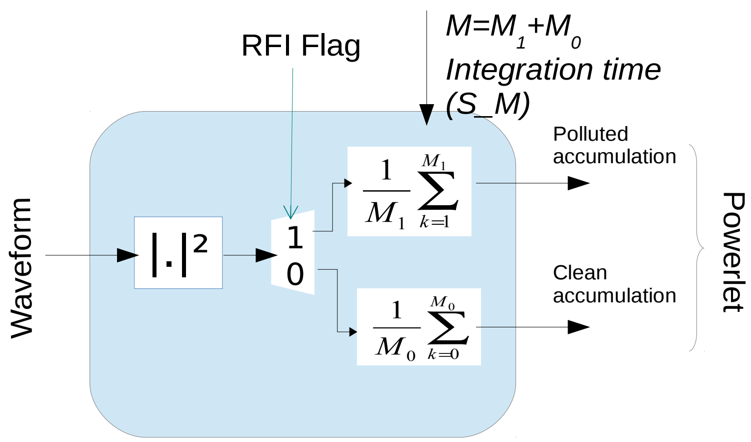

Figure 8: Oriented Power Accumulation Module. The RFI flag signal is generated by the RFI decision modules. M is the number of accumulated data. The dump time M is a static configuration parameter. The terms in brackets are the corresponding name used in the VHDL code.

3.4 RFI mitigation modules

Figure 9 gives an overview of an RFI mitigation module. It includes:

- a set of detectors: each detector analyses a set of *L* consecutive waveform time samples and provides a flag if the threshold is reached. The analysing window lengths, *L*, and the thresholds are static parameters.
- a synchronization part: Each detector has is own latency (mainly parametrized by *L*). Consequently, all RFI flags and waveforms must be synchronized with the longest latency. Simple one bit shift-registers can be used for RFI flags and a circular buffer can be used for the waveforms.
- A decision part: all individual RFI flags are merged to a single RFI decision flag.
- A blanking module: 3 modes are available (part of the static configuration):
    - *go-through mode*: waveform time samples are just flagged, no modification of the waveforms.
    - *zero option*: flagged waveform time samples are replaced by zeros
    - *Gaussian option*: flagged waveform time samples are replaced by synthetic complex Gaussian samples.

When RFI are detected, information about the detection context are sent to the RFIlet generator module.

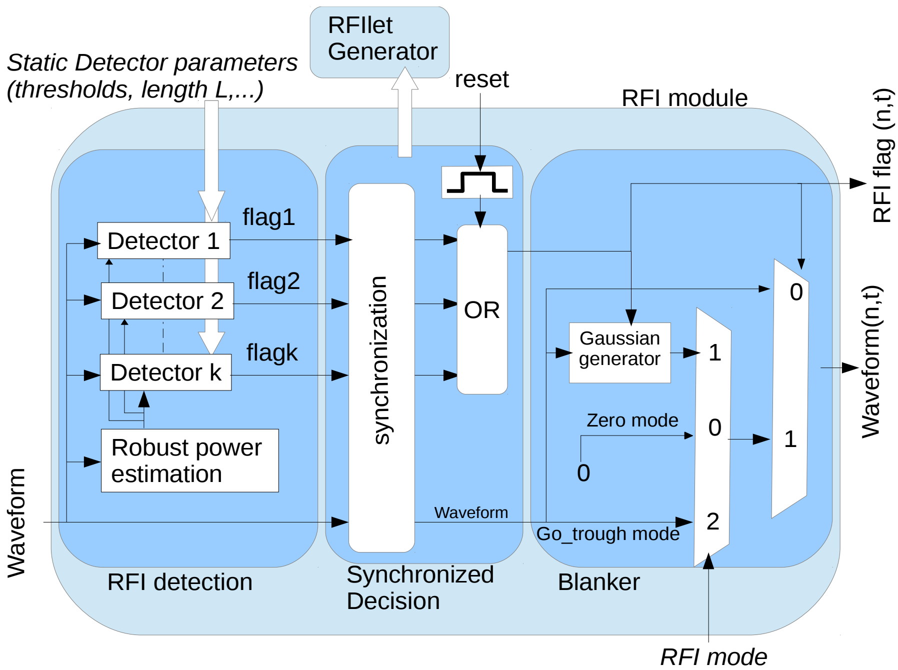

Figure 9: RFI module overview. Sets of waveform samples are send to the RFI detection module. Each detector evaluates the RFI contamination level. Then, the corresponding flags and the waveforms are sent to the synchronized decision module. The synchronisation removes the differential latencies generated by detector time responses. All detectors flags are merged to one global RFI decision. The next stage is the blanker module where several options are proposed to managed the polluted waveform samples.

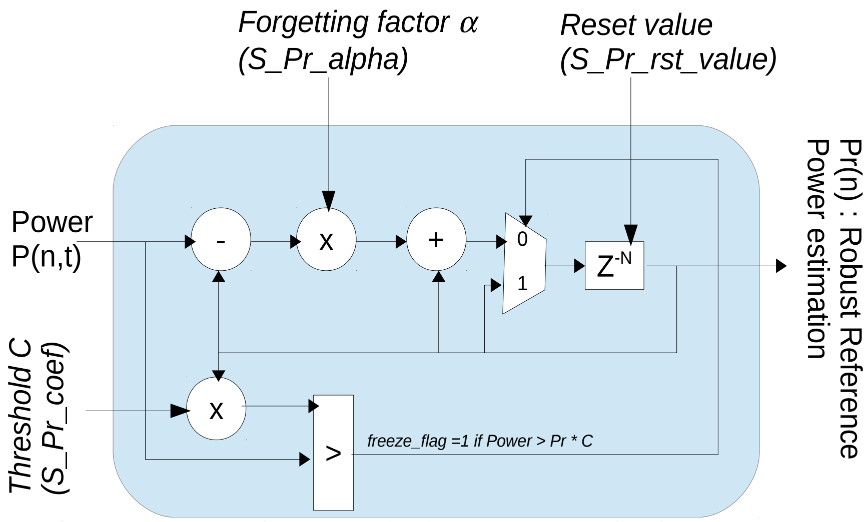

Figure 10: The robust power estimator. P(n,t) represents the instantaneous power at time t for waveform n. Pr is the robust power estimation. The forgetting factor α, the threshold C and the Reset value are static configuration parameters which control the detector behaviour. The terms in brackets are the corresponding name used in the VHDL code.

3.4.1 The detectors

Each detector receives the waveform stream, computes its criteria over *L* waveform samples (*L* can be different for each detector) and compares it with a given threshold. To normalize the criteria, a robust mean power (i.e. is estimated for each waveform.

3.4.1.1 Robust Power estimation

The approach is based on the χ2 distribution model assuming complex Gaussian noise waveforms when no RFI is present [5]. So, only the mean parameter is needed to fully define the signal statistics. The robust mean power value, *Pr*, is estimated recursively using a low-pass IIR filter with a threshold (see Figure 10):

$$Pr(n,t)=\alpha P(n,t)+(1-\alpha)Pr(n,t-1)$$

where *P(n,t)* represents the instantaneous power (i.e. X^2re+X^2im or Y^2re+Y^2im) of the current sample at time t for a given waveform n. The forgetting factor, α, is related to the number of samples used to compute the mean. It characterizes the memory effect of the IIR filter. The impulse response of this IIR filter is given by: $h(n)=\alpha\times(1-\alpha)^n$.

As shown in Figure 10, *Pr* is only updated when *P* is below a given threshold (i.e. outliers are not taken into account in the estimation). The threshold is calculated as *C x Pr* with *C* a coefficient defined by the user. *C* is related to the expected false alarm rate. An illustration of this estimator is given at Figure 15.

3.4.1.2 The strong detector

The strong detector (see Figure 11) is used to detect very short but strong RFI. The flagging occurs when *Ls* consecutive waveform samples exceed the power threshold. The power threshold is calculated as *Cs x Pr* with *Cs* a coefficient defined by the user. *Cs* is related to the expected false alarm rate.

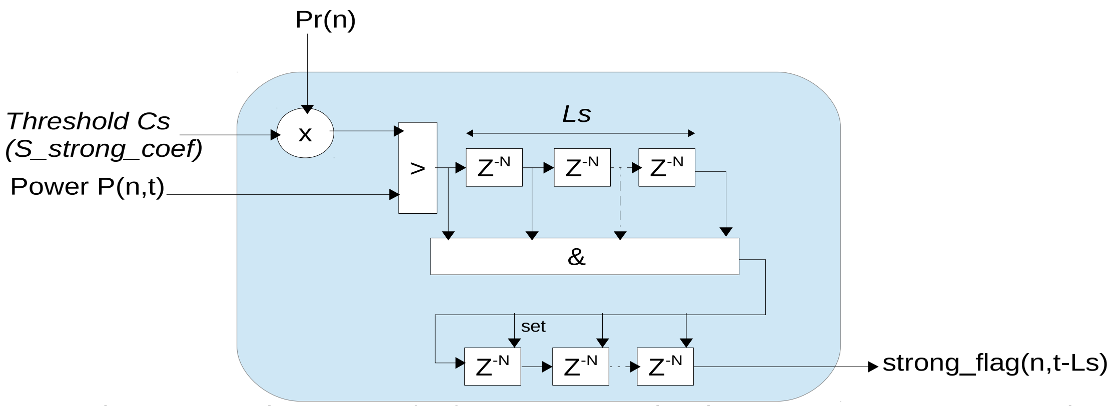

Figure 11: The strong detector. P(n,t) represents the instantaneous power at time t for waveform n. Pr is the robust power estimation. The threshold Cs and the number of time-lags Ls are static configuration parameters which control the detector behaviour. The terms in brackets are the corresponding name used in the VHDL code.

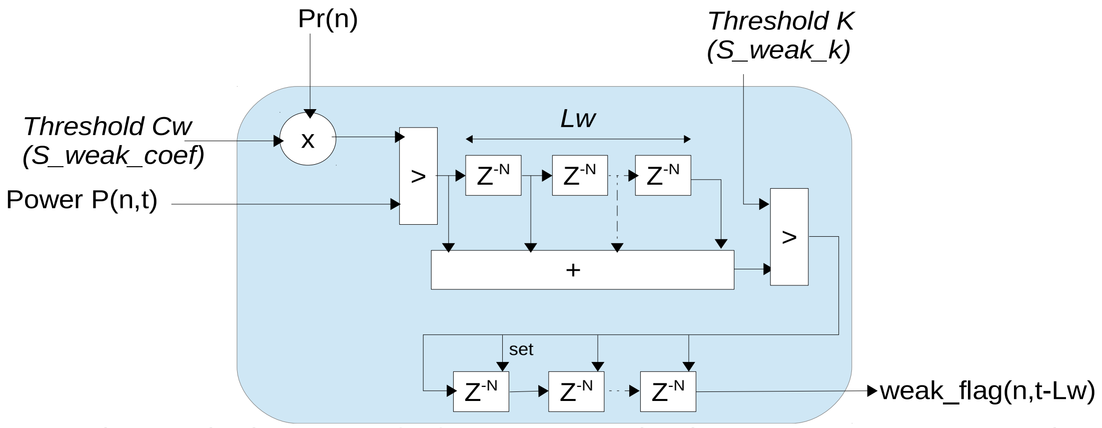

Figure 12: The weak detector. P(n,t) represents the instantaneous power at time t for waveform n. Pr is the robust power estimation. The threshold Cw and the number of time-lags Lw are static configuration parameters which control the detector behaviour. The terms in brackets are the corresponding name used in the VHDL code.

3.4.1.3 The weak detector

The weak detector (see Figure 12) is used to detect weak RFI but long RFI. The flagging occurs when at least *K* waveform samples among *Lw* consecutive waveform samples exceed the power threshold. The power threshold is calculated as *Cw x Pr* with *Cw* a coefficient defined by the user. *Cw* is related to the expected false alarm rate.

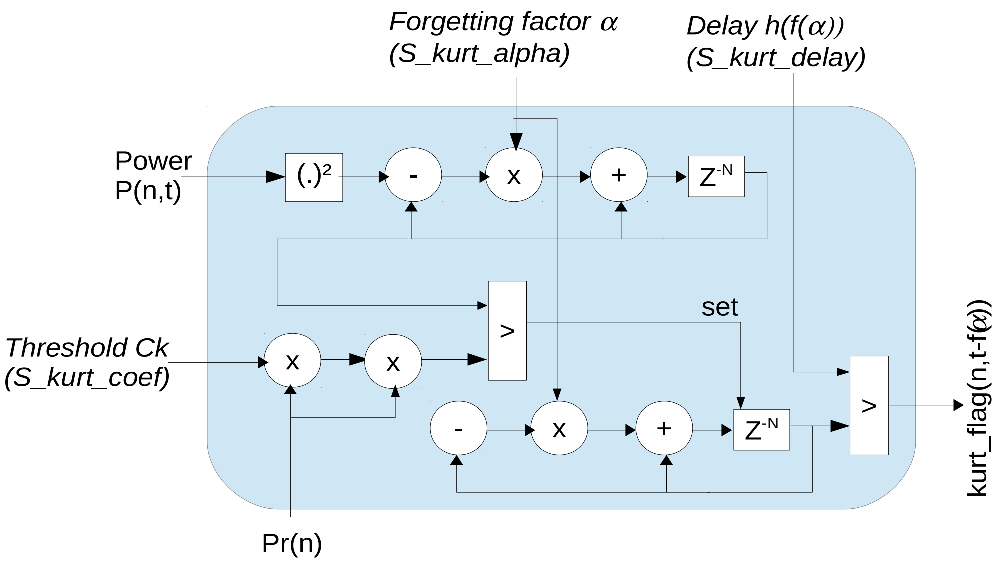

Figure 13: The kurtosis detector. P(n,t) represents the instantaneous power at time t for waveform n. Pr is the robust power estimation. The threshold Ck and the forgetting factor α, are static configuration parameters which control the detector behaviour. f(α) is the expected latency. The terms in brackets are the corresponding name used in the VHDL code.

3.4.1.4 The kurtosis detector

The kurtosis detector (see Figure 13) is based on the 4th order moment (i.e. mean of $|S(t,fo,A)|^4$). This moment is computed in a similar way as the robust power estimation, by using an IIR filter. The number of waveform samples involved in the averaging depends on the forgetting factor. The flagging occurs when the 4th moment exceeds the threshold. The threshold is calculated as $Ck \times Pr^2$ with Cw a coefficient defined by the user. *Ck* is related to the expected false alarm rate. The second IIR filter generates a flag which is compatible with the time response of the first IIR (remark : it could be also implemented by using a counter).

3.4.1.5 The Gaussian generator

The Gaussian generator (see Figure 14) will provide fake unpolluted waveform for replacing polluted waveform samples. The normalized Gaussian generator will be based on an Opencores module [6]. The waveform amplitude will be estimated in a similar way as the robust power estimation, by using an IIR filter both on real and imaginary parts.

3.4.2 The RFIlet generator

According to a RFI flag generating by the RFI decision module, the following information are send to the database:

- configuration parameters in the database for the given acquisition::
  - Δf is related to the waveform bandwidth
  - A is related to steering direction
  - all detector configuration parameters

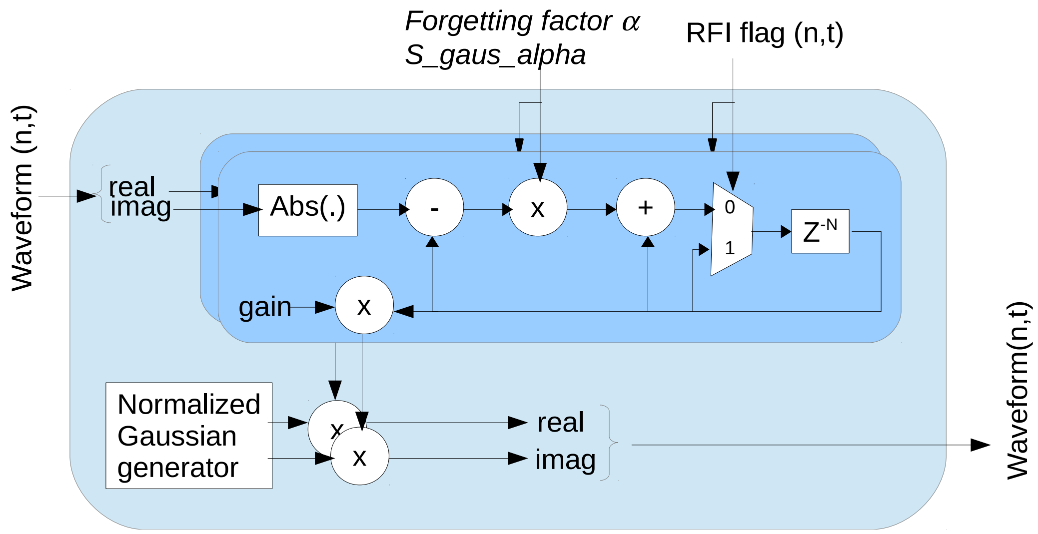

Figure 14: The Gaussian generator. The forgetting factor α, is a static configuration parameters which control the generator precision. The terms in brackets are the corresponding name used in the VHDL code

- dynamic parameters:

- t is related to the time stamp of the RFIlet
- ∆t is the duration of the RFIlet
- fo is related to the waveform central frequency (frequency index in the given resolution level)
- detector flags during the period [t t+∆t] for the considered waveform

For efficiency purpose, all these information will be dispatched at different places in the Database architecture. Only the dynamic information generated by the detectors will be put in the Uniboard RFIlet data stream. Configuration parameters will be stored once and for all.

4 Conclusion

All modules are based on the Altera Avalon Interface. All algorithms are based on very simple functions to preserve computational resources. However the memory resource is vital due to the multi signal approach and due to the synchronisation issue between the detectors and the data flow.

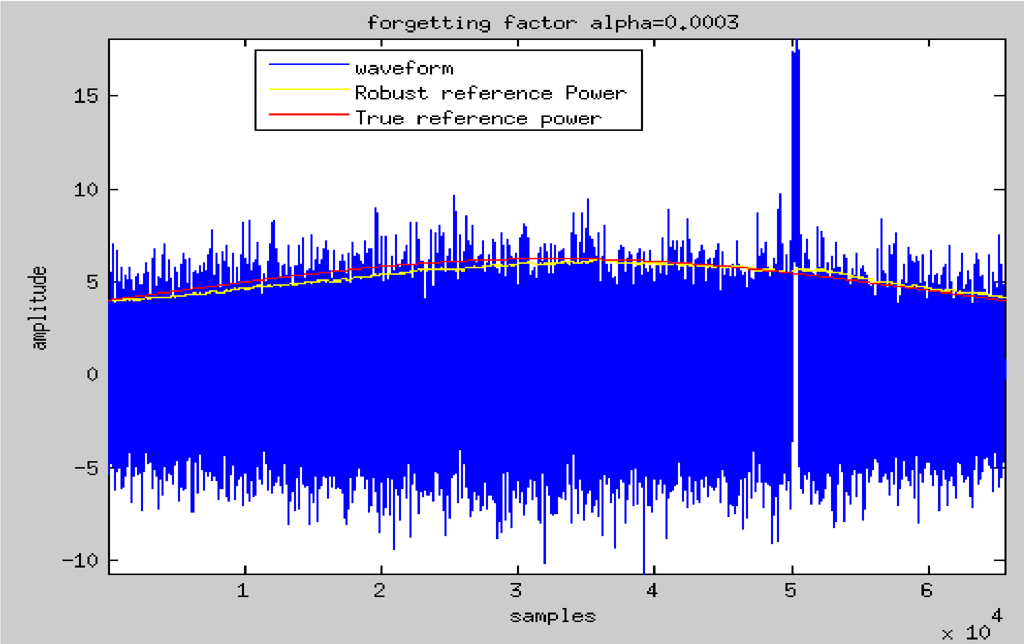

Figure 15: Robust reference power estimator example.

Copyright