

An Ultra Low-Power WOLA Filterbank Implementation in Deep Submicron Technology

R. Brennan, T. Schneider
Dspfactory Ltd
611 Kumpf Drive, Unit 200
Waterloo, Ontario, Canada N2V 1K8

P. Balsiger, Ch. Calame, A. Drollinger,
L. Grisoni, A. Heubi,
F. Pellandini, D. Sun, Ch. Waelchli,
Institute of Microtechnology
University of Neuchâtel
Rue A.-L. Breguet 2, 2000 Neuchâtel,
Switzerland

Abstract

The availability of deep Submicron technology opens the door to advanced algorithms specifically targeted for ultra low-power portable applications like hearing aids. These applications are extremely constrained by small physical size and extremely low power consumption requirements. To achieve these strict requirements, every component must be justified. The Weighted Overlap-Add (WOLA) filterbank discussed here is an important component that meets these difficult requirements.

Introduction

Many signal processing algorithms can be cast into a filtering (frequency-domain) framework. These include dynamic range compression, noise reduction, sub-band coding and directional processing, voice activity detection and echo cancellation. For these types of real-time audio signal processing applications, the filtering requirements are strict: i) low group delay, ii) high degree of adjustability, iii) high fidelity. A frequency domain approach is an efficient method of meeting these constraints while delivering low power and flexibility. This paper describes a WOLA filterbank that utilizes a novel architecture and processing scheme. A key figure of merit in these applications is the energy required for a given amount of processing. Typically this is ex-

pressed in μJ for a Fast Fourier Transform (FFT) operation.

The WOLA filterbank is tiny. On a $0.35\mu\text{m}$ process, memory (2 kwords), occupies 4.4 mm^2 and logic is 2.2 mm^2 . Less than 30 kbytes are used for the entire coprocessor. Typical power consumption at 1 Volt is $300\text{ }\mu\text{W}$ for a WOLA coprocessor in a typical 16-channel configuration. The chip is typically clocked at 1.28 MHz.

Basic Concept

The WOLA design provides a highly flexible time-frequency representation amenable to sub-band adaptive, sub-coding and other similar applications [1], [2], [4]. The coprocessor easily interfaces with any standard DSP processor, AD- and DA-converters.

The co-processor has two main sub-blocks (Figure 1): the WOLA and the Input/Output processor (IOP). Input samples are stored in a circular input FIFO. Every R (input block size) samples a WOLA analysis transformation is performed on L samples ($L \gg R$). A general purpose DSP (the "control DSP") is used to analyze the spectrum and to apply, via the shared RAM, gains for each frequency band. Then, the WOLA coprocessor performs a WOLA synthesis transformation and stores the results in the output FIFO. The IOP is responsible for interpolating (decimating) the outgoing (incoming) samples.

The WOLA can also operate in stereo mode. In this mode, the WOLA processes two simultaneous data streams. Following analysis, the control DSP performs a final butterfly separation step, applies gains to the separated stereo channels and then mixes both channels. The mixed frequency-domain signals are then returned to the time domain via a WOLA synthesis transformation. Stereo supports the implementation of phase-dependent algorithms including direction of arrival estimation, echo cancellation and sub-band beamforming.

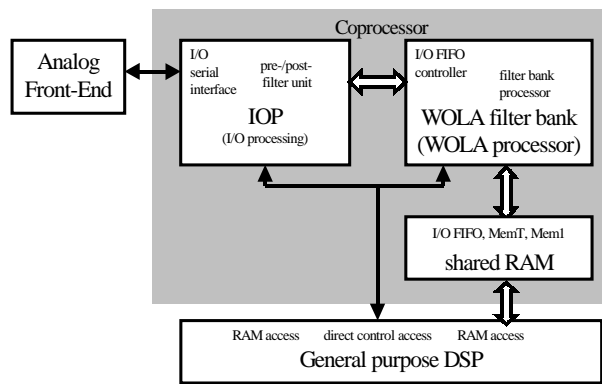


Figure 1. Overview of the co-processor's environment

Theoretical aspect of the WOLA filterbank

Over the last two decades, multi-rate digital signal processing techniques have been considerably developed and widely practiced in various engineering disciplines. The conditions to obtain perfect reconstruction (PR) maximally decimated (or critically sampled) filter banks have been extensively investigated and well-documented [5]. PR systems impose severe constraints that are not suitable in some applications. For applications requiring significant adjustment in the frequency bands, other structures are preferable [2].

The WOLA structure can meet these design constraints [3]. Furthermore, as will be shown below, when implemented using

Block-Floating-Point (BFP) arithmetic, excellent performance is obtained.

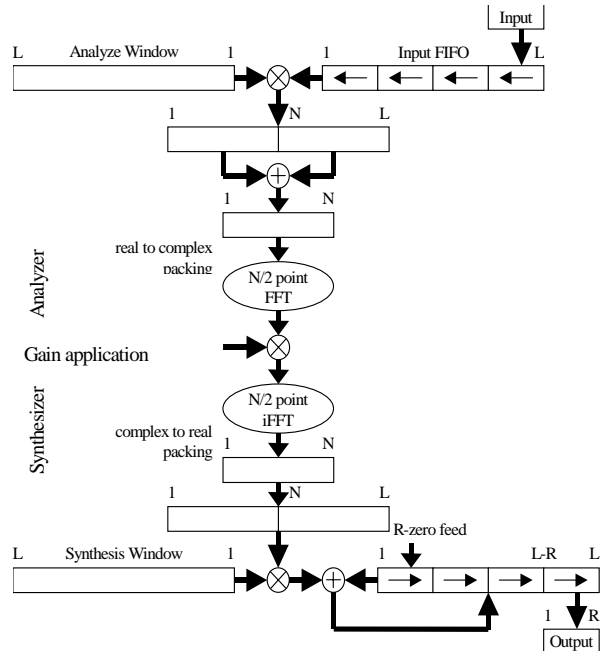


Figure 2. Simplified block diagram of a WOLA filter bank

WOLA structure

Figure 2 shows a simplified block diagram of an oversampled WOLA filter bank [2], [3]. The input step size (R) is the FFT size (N) divided by the oversampling ratio (OS). The use of oversampling provides two benefits (i) the gain of the filterbank bands can be adjusted over a wide range without the introduction of audible aliasing and (ii) a group delay versus power consumption trade-off can be made.

In operation, the input FIFO is shifted and R new samples are stored. The input FIFO is then windowed with a prototype low pass filter of length L . The resulting vector is added modulo N (i.e., "folded") and the FFT of the resulting windowed time segment is computed. Because an FFT is used, the outputs from the analysis filterbank provide

both magnitude and phase information (i.e., they are complex).

To generate a modified time-domain signal, the channel gains are applied to the $N/2$ FFT outputs (channel signals) and an inverse FFT is computed. The resulting time-domain “slice” is then windowed with a synthesis window and accumulated into the output FIFO. This generates R samples that are shifted out of the output FIFO. Finally, R zeros are shifted into the output FIFO and the entire process repeats for the next block of R input samples.

Equally spaced bands can be generated by using an Odd FFT and a square-wave modulator on the time-domain input and output signals. For stereo processing, two mono signals are interleaved and processed as a complex signal. After analysis the complex result is separated to get the individual mono signal spectra.

Block floating point arithmetic

Block-floating-point (BFP) computation units are used to increase the dynamic range and reduce the quantization error in order to improve the SNR of the WOLA filterbank. The BFP strategy decreases the quantization error without increasing the computation complexity. This is achieved by dividing data into non-overlapped groups (passes) and formatting the data at each node in data flow path with common exponent.

Implementation of the WOLA

The implementation of the WOLA coprocessor was guided by three primary constraints: (a) minimal size (both gate count and silicon area) (b) minimal power consumption (c) flexibility; we required a WOLA filterbank design that supported programmable configurations.

The Input/Output Processor (IOP)

The input and output FIFO's are realized as circular buffers. The input FIFO has two data domains: the WOLA-processing domain that is used by the WOLA processor and the write-in domain that stores the new samples. The output FIFO has also two data domains: the WOLA-processing domain and the read-out domain (which contains the data that are ready to send out).

The IOP contains an interface module and pre-/post- processing unit for interpolating and decimation filtering. A pole-zero pair is used as a DC removal filter. To save area and power, the number of arithmetic units is limited to one MAC unit, one adder, one shifter and one rounding unit. These resources are shared between the DC-removal, the decimation and the interpolation filters.

The WOLA Control System

A simple, yet very flexible controller commands the WOLA processor. WOLA processing is divided into different passes. The characteristics of a pass are i) all operations are the same (radix2, radix4, etc) and ii) every read and write address for the data and coefficients can be generated by a bit-reverse addressing unit. Thus, each pass can employ a fixed configuration.

Data-path

The data-path consists of a multiplier array and an ALU bank (Figure 3).

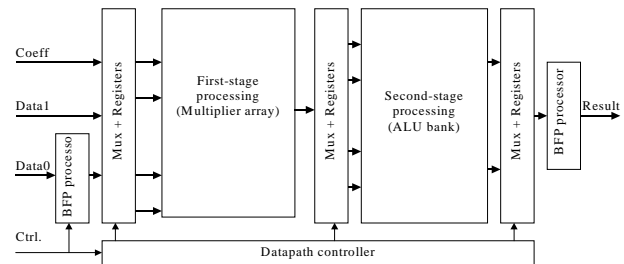


Figure 3. Block diagram of the data-path.

This structure achieves an efficient data-flow. N-point FFT processing is achieved by computing a specific number of radix4 or radix2 transforms.

Results

The noise floor and THD+N measurements of the coprocessor's data-path units show that the noise floor is approximately -115 dB. A highly selective filterbank configuration (16-channels with 14 ms group delay) provides about -65 dB THD+N. Note that the realized THD+N is dependent on the selected WOLA parameters (FFT size, oversample factor and window length) as well as the window coefficients. A large number of combinations that trade-off fidelity (THD+N) versus power consumption versus group delay is possible.

The time and cycles used to perform WOLA analysis and synthesis is strongly configuration dependent. Table 1 shows the cycles required for analysis and synthesis and also gives the ratio between the used time and the time that is available if a system clock of 1.28 MHz and a sampling frequency of 16 kHz are used.

N	L	OF	#cycles (analysis + synthesis)	$t_{used}/t_{available}$ (analysis + synthesis)
16	128	2	192	38%
32	256	2	258	26%
32	128	2	328	33%
32	128	2	289	29%
128	128	4	788	39%

N : FFT size (complex)
L : prototype filter length (window size)
OF : oversampling factor ($R=N/OF$)

Table 1. Computation cycle and time for different configurations

A first implementation was done on a 0.35 μ m CMOS five-metal layer technology. The

total die size (including pads) is 7.86 mm². A second implementation has been done on a 0.18 μ m CMOS four-layer metal technology. The measured power consumption is 0.4 mW @ 1.7 volt supply voltage and less than 250 μ W @ 1 volt for a typical algorithm.

Figure 4 compares the co-processor performance (in terms of execution time in microseconds at an equivalent chip clock frequency of 10 MHz) to standard DSP processors. Note that the BDSP9124 and the DSP-24 are dedicated FFT DSPs, which use significantly more resources than the WOLA coprocessor described here does.

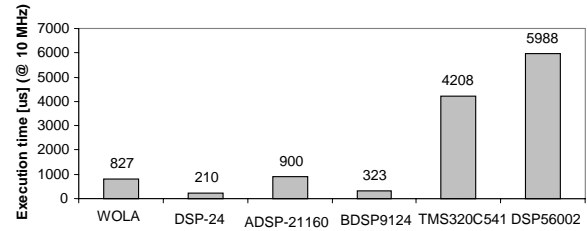


Figure 4. Execution time in μ sec for 1024 complex point FFT sizes @ 10 MHz. ADSP-21160 (Analog Devices), BDSP9124 (Butterfly DSP), WOLA (IMT, Dspfactory), DSP-24 (DSP Architectures) and C6x and C50 (Texas Instruments)

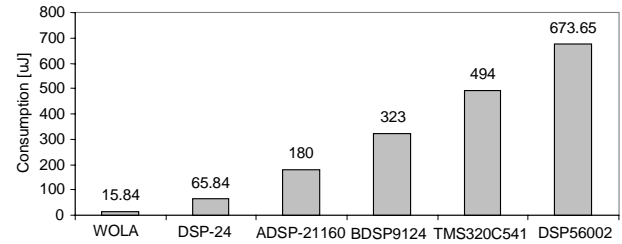


Figure 5. Effective power consumption for one 1024 complex point FFT. Power supply: WOLA 1.7V, DSP-24 3.3V, ADSP-21160 5V, BDSP9124 5V, TMS320C541 5V, DSP56002 5V

Because the maximal FFT size of the WOLA co-processor is limited to 128 complex points (256 real points) its 1024 point complex-FFT execution time is an interpolated value.

The most important value for lowest power applications is the effective power consumption used to perform an FFT. Figure 5 shows the effective power consumption for different processors. The processors from Analog Devices, Butterfly DSP and DSP Architectures are focussed on high throughput rate and not on low-power implementations. Figure 5 clearly shows that the WOLA offers roughly 4.5 times the performance when power consumption is considered.

Conclusions

This paper presents an efficient design and the results for a novel real-time WOLA filterbank. The co-processor has been implemented in two deep sub-micron technologies. The entire co-processor requires approximately 30 k gates.

The WOLA filterbank is ideal for use in a wide range of frequency domain processing applications as mentioned in the introduction. Other applications include personal listening devices requiring head-related transforms and other similar algorithms, which can be implemented in the frequency domain. The WOLA is especially suited to applications that require low delay (< 10 ms).

In wireless applications the WOLA can be used to perform simultaneous high-fidelity equalization, noise reduction and AGC (dynamic range compression). Using the stereo processing mode, the WOLA can be used to implement two-microphone, adaptive frequency domain noise cancellation for use in handsets. The stereo processing mode is also well suited for frequency domain echo cancellation.

The large amount of signal separation between the WOLA output channels offers a high-degree of orthogonality between all channels. This greatly speeds the conver-

gence of adaptive algorithms and makes for very efficient subband coders and decoders. The low delay is also attractive for many subband-coding applications.

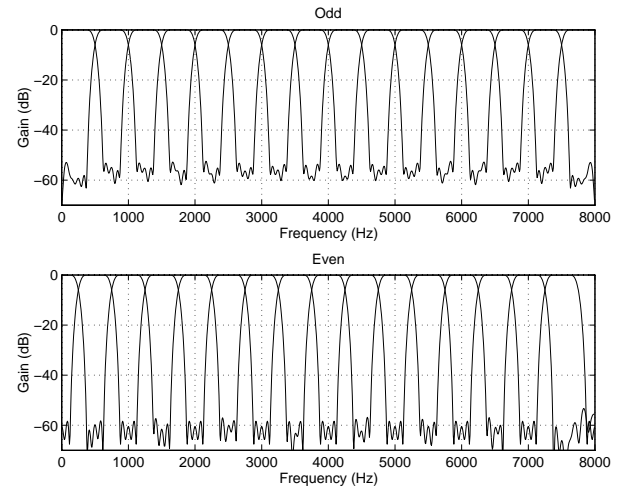


Figure 6. Channel frequency responses for 16-channel filterbank in even and odd stacking with 14-ms group delay

The combination of a general-purpose (control) DSP and the WOLA provides an ideal flexibility versus power consumption trade-off.

References

- [1] R. Brennan & T. Schneider, "Filterbank Structure and Method for Filtering and Separating an Information Signal into Different Bands, Particularly for Audio Signals in Hearing Aids", PCT Patent Publication WO9847313A210, October 22, 1998.
- [2] R. Brennan & T. Schneider "A Flexible Filterbank Structure for Extensive Signal Manipulations in Digital Hearing Aids," Proc. ISCAS-98, Monterey, CA.
- [3] R. E. Crochiere & L. R. Rabiner, "Multirate Digital Signal Processing", Prentice-Hall, 1983.
- [4] T. Schneider & R. Brennan "A Multichannel Compression Scheme for a Digital Hearing Aid," Proc. ICASSP-97, Munich, Germany.
- [5] P. P. Vaidyanathan, "Multirate Systems And Filter Banks", Prentice-Hall, 1993.