



Article

FPGA-Based Architectures for Acoustic Beamforming with Microphone Arrays: Trends, Challenges and Research Opportunities

Bruno da Silva ^{1,2,3,*}, An Braeken ¹ and Abdellah Touhafi ^{1,2}

¹ Department of Industrial Sciences (INDI), Vrije Universiteit Brussel (VUB), 1050 Brussels, Belgium; an.braeken@gmail.com (A.B.); abdellah.touhafi@vub.be (A.T.)

² Department of Electronics and Informatics (ETRO), Vrije Universiteit Brussel (VUB), 1050 Brussels, Belgium

³ Department of Electronics and Information Systems (ELIS), Ghent University (UGent), 9052 Ghent, Belgium

* Correspondence: bruno.da.silva@vub.be

Received: 7 June 2018; Accepted: 1 August 2018; Published: 3 August 2018



Abstract: Over the past decades, many systems composed of arrays of microphones have been developed to satisfy the quality demanded by acoustic applications. Such microphone arrays are sound acquisition systems composed of multiple microphones used to sample the sound field with spatial diversity. The relatively recent adoption of Field-Programmable Gate Arrays (FPGAs) to manage the audio data samples and to perform the signal processing operations such as filtering or beamforming has lead to customizable architectures able to satisfy the most demanding computational, power or performance acoustic applications. The presented work provides an overview of the current FPGA-based architectures and how FPGAs are exploited for different acoustic applications. Current trends on the use of this technology, pending challenges and open research opportunities on the use of FPGAs for acoustic applications using microphone arrays are presented and discussed.

Keywords: FPGA; architectures; acoustics; microphone array; beamforming

1. Introduction

Sensor arrays have been useful tools for radar-related applications, radio astronomy and other applications including indoor localization systems or environmental monitoring. Over the past decades, many systems composed of microphone arrays have been developed and evaluated for sound source location, separation and amplification in difficult acoustic environments. Large microphone arrays like [1–3] have been built in the early 2000s to evaluate different algorithms for speech enhancement targeting conference rooms. Small microphone arrays are nowadays ubiquitous in many consumer devices such as laptops, webcams or smartphones as an aid for speech recognition [4] due to the improvement of the recognition when compared to a single omnidirectional microphone [5]. Multimodal computing applications such as person tracking systems, hearing aid systems or robot-related applications benefit from the use of microphone arrays. The computational demand of such applications is not always satisfied when targeting real-time or large microphone arrays.

Field-Programmable Gate Array (FPGA) is a semiconductor device providing thousands of arrays of programmable logic blocks and specific-operation blocks. FPGAs are reprogrammed to enable the hardware description of the desired application or functionality requirements. Customized designs built with this technology achieve low latency, performance acceleration and a high power efficiency, which made them very suitable as hardware accelerators for applications using a microphone array. FPGAs are well-suited for high-speed data acquisition, parallel processing of sample data streams and to accelerate audio streaming applications using microphone arrays to achieve real-time.

Moreover, the multiple I/Os that FPGAs offer, facilitates the interface of microphone arrays with a relatively large number of microphones. As a result, FPGAs have displaced Digital Signal Processors (DSPs) for acoustic applications involving microphone arrays in the recent past.

FPGA technology has disruptive characteristics which are changing the way microphone arrays are used for certain acoustic applications. The latest FPGA-based architectures are capable to fully embed all the computational tasks demanded by highly constraint acoustic imaging applications. Nevertheless, many FPGA-based architectures must solve similar challenges when interfacing and processing multiple data streams from large microphone arrays. Moreover, different FPGA-based architectures have been proposed for sound source localization. Here, a review of how FPGAs are used for applications based on microphone arrays is presented. This survey intends to present the most relevant uses of FPGAs combined with microphone arrays. The challenges, trends and potential use that FPGA's technology has for acoustic-related microphone arrays are also presented and discussed.

An introduction of microphone arrays, detailing the most common types of microphones and explaining the type of processing done with microphone arrays is presented in Section 2, followed by a brief explanation of the FPGA technology in Section 3. A categorization of the state-of-the art based on the role of the FPGA when processing the signal data coming from the microphone arrays is proposed in Section 4. Such an overview provides a perspective on how FPGAs have been adopted for the different types of applications involving microphone arrays. An analysis of advanced FPGA-based architectures of acoustic beamformers is presented in Section 5. This analysis focused on how existing beamforming techniques are implemented on FPGAs, the level of integration on the FPGA and how the selection of the microphone affects the architecture. Moreover, the reasons of the incremental adoption of FPGAs is further discussed in Section 6, providing an overview of the recent trends. The current challenges and research opportunities are discussed in Section 7. Finally, the conclusions are drawn in Section 8.

2. Microphone Arrays

Microphone arrays have advanced together with recent developments in microphone technology. Large microphone arrays like [2], composed of hundreds of Electret-Condenser microphones (ECMs), have evolved to compact microphone arrays composed of Micro-Electro-Mechanical Systems (MEMS) microphones to be integrated in smartphones, tablets or voice assistants, such as Amazon's Alexa [6]. Microphone arrays have been used in hearing aids [7], for echo cancellation [8], in ambient assisted living [9], in automotive [10,11], for biometrical systems [12,13], for acoustic environmental monitoring [14], for detection and tracking of Unmanned Aerial Vehicles (UAVs) [15,16], or for using UAVs for rescue activities [17], for speech enhancement [5,18] and in many other applications [19]. The miniaturization of the packaging while keeping the quality of the microphones' response has lead to compact microphone arrays and created opportunities for new acoustic applications.

2.1. Type of Microphones

The most popular type of microphones to compose microphone arrays is briefly described here. Microphones can be grouped based on their transducer principle, that is, how the microphone converts the acoustic vibration into an electrical signal. There are several types of transducers such as condenser or capacitor microphones, dynamic microphones, piezoelectric microphones,... Over all the variety of the available microphones, two main categories have predominated when building microphone arrays: ECMs and MEMS microphones (Figure 1).

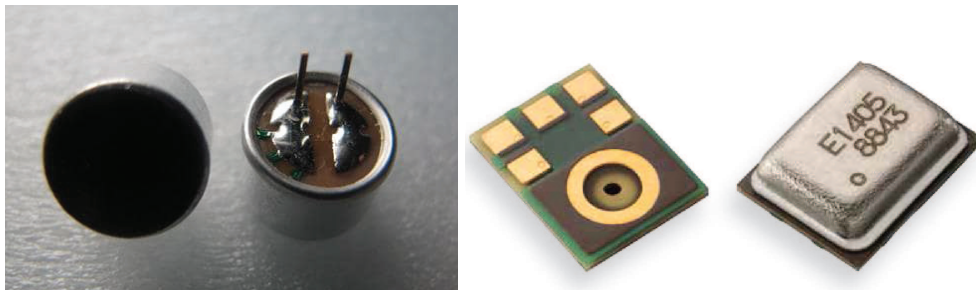


Figure 1. Example of ECMs (left) and MEMS microphones (right).

2.1.1. ECMs

ECMs are a type of condenser microphone composed of conductive electrode members on different plates, one of which is a moveable diaphragm. One of the plates is a stable dielectric material called electret with a permanent electric charge, which eliminates the need for a polarizing power supply. ECMs only require certain power supply to power an integrated preamplifier. The capacitance of the parallel plate capacitor changes when the distance between the two plates varies when the sound wave strikes the surface of the moveable diaphragm. ECMs have a whole acoustic frequency response and a low distortion in the signal transmission since the capacitance effect varies due to an electromechanical mechanism. The relatively small package of ECMs made them the preferred choice to build the first large microphone arrays like in [1–3]. The output format of ECMs composing microphone arrays has been traditionally analog. The output impedance rounds from a few hundred to several thousand ohms [20], which must be considered when selecting the codec. This impedance is determined in case of the ECMS by the value of the load resistance with a corresponding change in sensitivity [21].

2.1.2. MEMS Microphones

A MEMS microphone is a miniature microphone, usually in the form of a surface mount device, that uses a miniature pressure-sensitive diaphragm to sense sound waves. Similarly to ECMs, the variations of the diaphragm directly determine the capacitance. This diaphragm is produced by surface micromachining of polysilicon on a silicon substrate or etched on a semiconductor using standard Complementary Metal Oxide Semiconductor (CMOS) processes [22]. MEMS microphones include significant amounts of integrated circuits for signal conditioning and other functions within the same package of the sensor since it shares the same fabrication technologies used to make integrated circuits.

MEMS microphones are categorized based on the type of output, which can be analog or digital [23]. Analog MEMS microphones present an output impedance of a few hundred ohms and an offset DC voltage between ground and the supply voltage. Despite this offset avoids the clipping of the peaks of the highest amplitude output signals, it also leads to a high-pass filter effect which might attenuate low frequencies of interest. Regarding the high impedance, a possible solution to avoid attenuations at the output side is the use of programmable gain amplifiers at the codec side.

Because MEMS microphones are produced on a silicon substrate, a clear benefit of digital MEMS microphones is the easy integration of the transducer element together with an amplifier and an Analog-to-Digital Converter (ADC) in the same die or in the package of the microphone. As a result, digital MEMS microphones drastically reduce all the required circuitry to interface the digital signal processor unit. Due to integrating all the ADC circuitry into the microphone's package, digital MEMS microphones provide advantages in the design phase. Each new design iteration requires adaptations in the signal conditioning circuitry when using analog microphones.

The encoded output format of digital MEMS microphones are Pulse Density Modulation (PDM) or Inter-IC Sound (I^2S) output interface. PDM represents an oversampled 1-bit audio signal and

brings low noise. The data from two microphones share the same data line at different shared clock edge, guaranteeing their synchronization. Same principle can be applied to microphone arrays, where multiple digital MEMS microphones can be synchronized by using the same clock. The synchronization of the microphones is crucial in microphone arrays, which might determine what type of MEMS microphone to use since arrays composed of analog MEMS microphones must be synchronized at the ADC. I^2S MEMS microphones present the same properties as the PDM MEMS microphones, but integrate in the silicon all the circuitry required for the PDM demodulation and multi-bit Pulse Code Modulation (PCM) conversion. Thus, I^2S MEMS microphones output is filtered and decimated at baseband audio sample rate.

2.1.3. Considerations

The selection of the type of microphones when building a microphone array is determined by the different features that each microphone's technology provides. The type of microphones and the output data format determine the overall output response and the digital signal processing requirements.

Despite ECMs and MEMS microphones operate as condenser microphones, MEMS microphones benefit from the enormous advances made in silicon technology over the past decades and present several advantages [24,25] that make them more suitable for many acoustic applications.

- MEMS microphones have less sensitivity to temperature variations than ECMs.
- MEMS microphones' footprint is around 10 times smaller than ECMs.
- MEMS microphones have a lower sensitivity to vibrations or mechanical shocks than ECMs.
- ECMs have a higher device-to-device variation in their frequency response than MEMS microphones.
- ECMs need a specific soldering process and are unable to be undertaken re-flow soldering, while MEMS can.
- MEMS microphones have a better power supply rejection compared to ECMs, facilitating the reduction of the components' count of the audio circuit design.

The advantages of the MEMS technology explains why MEMS microphones have slowly replaced ECMs as the default choice for microphones arrays since their introduction by Knowles Acoustics in 2006.

The output format is another relevant factor to be considered since it directly affects to the requirements of the digital signal processing system. Analog microphones demand certain considerations when selecting the codec due to the high impedance and the voltage offset at the microphone's output. Codecs, such as digital pre-amplifiers [26], convert the analog signals from analog microphones, in particular ECMs. This pre-amplifier drives an over-sampled sigma delta ADC to PDM output data. This type of integrated circuits facilitates the interface of analog microphones with digital processing systems by providing a compatible digital data format like PDM or I^2S audio bus [27]. The use of digital MEMS microphones, however, reduces the complexity of the hardware since they do not require external amplifiers. This fact makes digital MEMS microphones immune to Radio Frequency (RF) noise and less sensitive to electromagnetic interference compared to analog versions [28].

At the digital signal processing system side, the PDM data format produced at a high-sample rate needs to be demodulated to an analogue form before being heard as audio or converted to PCM format if it needs to be digitally analysed [29]. The operations required to demodulate the oversampled PDM signals consists of a multi-filter stage for the PDM demodulation and PCM conversion [30]. The integration of the PDM demodulation in the silicon reduces the I^2S MEMS microphones' flexibility since they present a fixed demodulator architecture [31]. The PDM demodulation circuitry integrated on the chip is a fixed decimator by a factor of 64 followed by a low-pass filter to remove the remaining high frequency components in the signals. The microphone operates as an I^2S slave, transferring the PCM data word length of 24 bits in 2's complement, as depicted in Figure 2. Due to the fixed decimation

factor, the digital signal processing system must wait several clock cycles before to receive the PCM signal from each microphone. This solution might satisfy some acoustic applications requirements, but it certainly reduces the opportunities of exploring alternative demodulation architectures based on the target application demands. For instance, different design strategies related to the architecture of the PDM demodulation are proposed in [32] to accelerate a particular type of acoustic application.

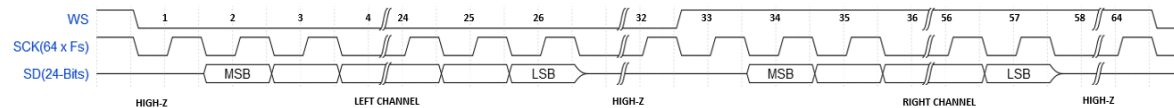


Figure 2. Example of two microphones with I^2S output format sharing the interface to a digital signal processing system [31]. SCK, WS and SD are the serial data clock, serial data-word select and the serial data output for the I^2S interface respectively.

2.2. Microphone Array Processing

Microphone arrays exploit the processing of signals captured by multiple spatially-separated microphones. The distance between microphones results in a difference in the path length between the sound sources and the microphones. This difference results in a constructive interference when the path length is equal for both microphones, obtaining an amplification of the signal by a factor of the number of microphones. The difference is dependent on the angle of incidence of the acoustic wave and the distance between the microphones. Therefore, microphone arrays are able to reinforce the sound coming from a particular direction while attenuating the sound arriving from different directions. The microphone arrays' frequency response depends on [33]:

- the number of microphones
- the spacing between microphones
- the sound source spectral frequency
- the angle of incidence

A high number of microphones improves the frequency response by increasing the Signal-to-Noise Ratio (SNR) [34] and by spatially filtering more precisely the sound field. Regarding the microphone's spacing, a large distance between microphones improves the array's response for lower frequencies, a short spacing prevents spatial aliasing [35]. The array geometry, referring to the position of the microphones in the array, is a wide research field [36] because the geometry aims to enhance acoustic signals and separate them from noise based on the acoustic application [37]. Figure 3 depicts some examples of array geometries.

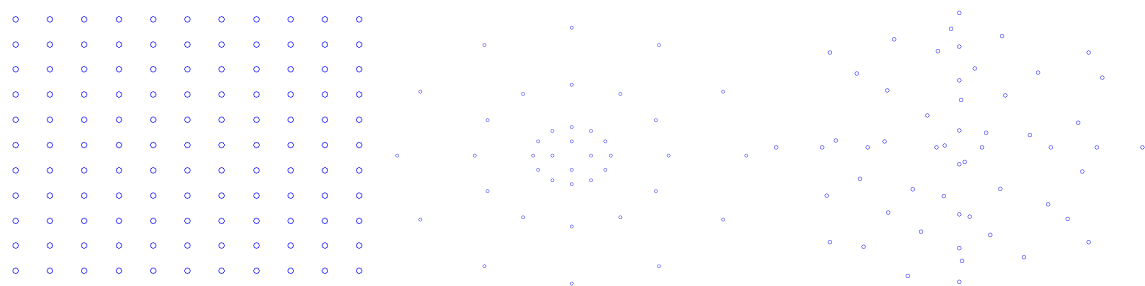


Figure 3. Examples of microphone arrays geometries.

The angle of incidence can be modified, performing a spatially filtering of the sound field, by adapting the path lengths of the input data of the microphones. The concept of steering the microphone's response in a desired direction is called beamforming [38,39]. The beamforming methods can be applied in the time domain or in the frequency domain. Time-domain beamformers apply

different time delays to each microphone to compensate for the path length differences from the sound source to the microphone arrays. The basic time-domain beamformer is the well-known Delay-and-Sum. The time delays can be also integrated in Finite Impulse Response (FIR) filters, like one per microphone, performing the Filter-and-Sum beamformer. Both beamformers can be also applied in the frequency domain. In that case, the signal received from each microphone is separated into narrow-band frequency bins through discrete Fourier transformation, before applying phase shift corrections to compensate the difference in path lengths. The beamforming operations present a high-level of parallelism and demand a very low latency when targeting real-time applications. Both are well-known features that FPGA present nowadays.

3. FPGA Technology

FPGAs are semiconductor devices composed of logic blocks interconnected via programmable connections. The fundamental units of an FPGA are Configurable Logic Blocks (CLBs) consisting of Look-Up Tables (LUTs) constructed over simple memories, SRAM or Flash, that store Boolean functions. Each LUT has a fixed number of inputs and is coupled with a multiplexer and a Flip-Flop (FF) register in order to support sequential circuits. Likewise, several CLBs can be combined for implementing complex functions by configuring the connection switches in the programmable routing fabric. The flexibility of FPGAs enables the possibility of embedding application specific architectures, which can be tuned to target performance, power efficiency or low latency.

Figure 4 depicts the FPGA's design flow. FPGAs are programmed with Hardware Description Languages (HDL), which describe the desired functionality to be mapped onto the reconfigurable hardware. The hardware description elaborated by the designer is used by the vendor's synthesizer in order to find an optimized arrangement of the FPGA's resources implementing the described functionality. During the synthesis, the design is translated to Register Transfer Logic (RTL) netlists. This feature distinguishes FPGAs from Application-Specific Integrated Circuits (ASICs), which are custom manufactured for specific design tasks. The application's functionality can also be described through digital logic operators represented as schematics diagrams. The netlists generated at the synthesis stage are used during the implementation stage to perform several steps: translation, mapping, place and routing. The translation merges the incoming netlists and constraints into the vendors' technology of the target FPGA. The mapping fits the design into the available resources, such as CLBs and I/Os. The place and route step places the design and routes the components to satisfy the timing constraints. Finally, a bitstream, which is used to program the FPGA with the design, is generated and downloaded to the device.

FPGA's design flow demands the design verification at the implementation stage, which is done through logic or timing simulations. Moreover, the synthesis and implementation stages require many compute-intensive operations, demanding minutes to hours to be completed based on the used amount of FPGA's resources. As a result, the overall design flow becomes a high-time and effort demanding task. In the recent past, several High-Level Synthesis (HLS) tools have been developed to alleviate the hardware description by using high-level descriptive languages, such as C/C++ [40,41] or OpenCL [42]. This high-level approach allows the increment of the reusability of the hardware descriptive code and facilitates the debugging and verification process.

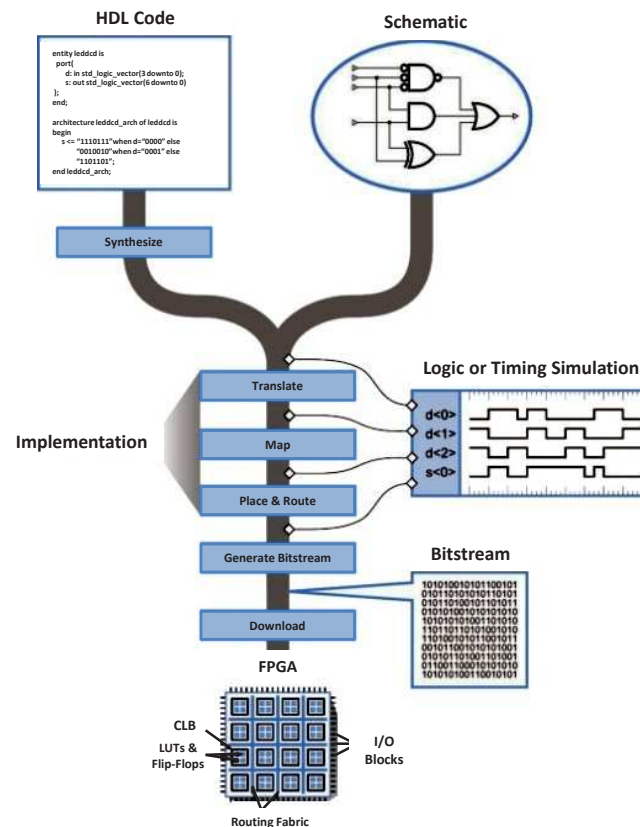


Figure 4. FPGA's Design Flow.

FPGA's resources have increased in the latest years following the improvements in the RTL technology. This increment in the available resources enables the embedding of general-purpose soft-core processors using the reconfigurable logic. Most of these customizable processors are 32-bits processors with a Reduced Instruction Set Computer (RISC) architecture. Existing open source soft-core processors, such as the OpenRISC [43], and especially the recent RISC-V [44] architecture, have been proposed in recent years as alternative to the Xilinx's Micro/PicoBlaze [45] or the Intel/Altera's Nios-II [46] soft-core processors. The use of these soft-core processors are extended in control-related applications or for the management of communication processes. The designer's effort to program such general-purpose processors is reduced due to the use of high-level languages. Although this type of embedded processors allows a fine tune customization at instruction levels and can be easily modified, their performance is not very high as they operate in a range from 50 MHz until 250 MHz. In recent years, there has been a move towards System-on-Chip (SoC) and FPGAs have been combined with hard-core processors, which are processors implemented with a fixed architecture in the silicon. Hard-core processors together with FPGA fabric provide a larger interconnection bandwidth between both technologies and achieve faster processing speed since they are not limited by the reconfigurable logic speed. Figure 5 depicts a Xilinx Zynq SoC FPGA serie [47], composed of a Processing System (PS), which is a dual-core ARM Cortex-A, and a Programmable Logic (PL) based on Artix-7 or Kintex7 FPGA fabric. Such SoC FPGAs demand, however, a hardware/software co-design to be fully exploited.

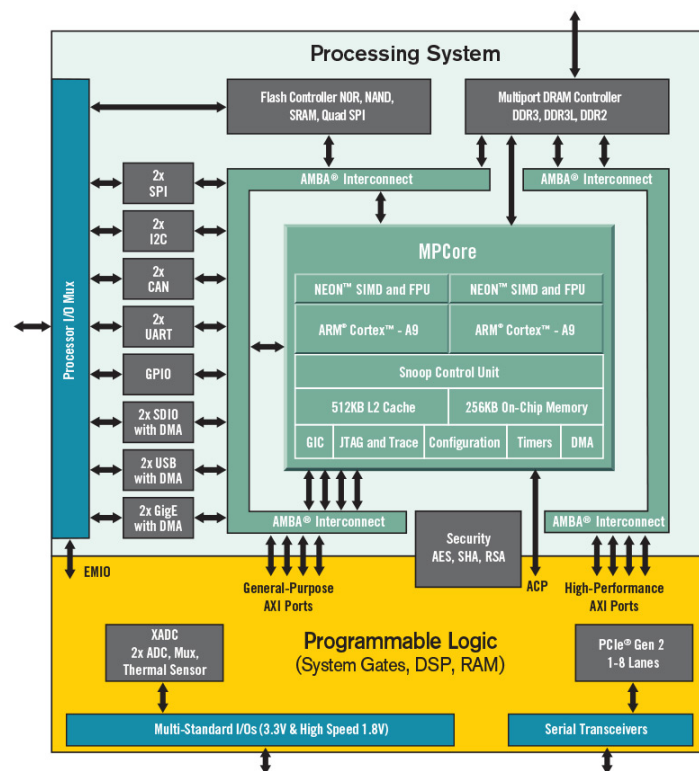


Figure 5. The SoC FPGAs Zynq serie [47] is composed of a hard-core processor (PS) and the FPGA reconfigurable logic (PL).

4. Categorization of FPGA-Based Designs for Microphone Arrays

There are many examples of the use of FPGAs for microphone arrays' applications. Beyond the type of acoustic applications, the FPGA's designs can be grouped into three main categories based on the embedded functionality:

- FPGAs satisfy the low latency and the deterministic timing required for the management of multiple data streams coming from multiple microphones. In several acoustic applications, FPGAs are used for the audio signal treatment by grouping the multiple data streams in an appropriated format before being processed. A common example is the serialization of the parallel incoming signals from the microphone array.
- Microphone arrays can be used to locate sound sources. Several FPGA-based designs embed not only the acquisition, demodulation and filtering of the data stream from the microphones, but also the required algorithms to locate sound sources. Further classification can be done based on the level of complexity of such algorithms, and the consequent computational demand.
- Highly constraint acoustic imaging applications have been developed on FPGAs in order to satisfy real-time demands and high computational requirements. The real-time computation of tens of microphones used for acoustic imaging applications demands a highly efficient performance architecture to properly exploit and achieve the performance that FPGAs offer nowadays.

The state-of-the-art for each of these categories is described more into detail in the following sections.

4.1. FPGA-Based Audio Acquisition Systems

An example of an FPGA-based acquisition system is described in [48]. The authors present a general-purpose acoustic array composed of 52 analogue MEMS microphones. Their acquisition

system, based on [49], includes an FPGA to provide real-time processing capabilities. The microphone array is composed of analog MEMS microphones, each demanding a circuitry composed of a two-stage 20 dB amplifiers and a decoupling capacitor. Additionally, to connect to the FPGA, each channel requires an ADC with 12 bits of resolution and a maximum sampling rate of 5 Mbps. One application of this acquisition system is the sound source separation using the Independent Component Analysis (ICA) technique [48]. The same authors propose in [50] the use of beamforming, matched filtering, spectral processing, and ICA for imaging, tracking and identification of objects. In both applications, the main tasks of the FPGA are the formatting and transmission of the recovered audio streams to a desktop Personal Computer (PC) through an Ethernet link.

There are more examples. Table 1 summarizes the main characteristics of the most relevant designs using FPGAs to rearrange the data format of the data streams from microphone arrays. The operations on the FPGA, however, are not significantly performance demanding since the embedded operations basically involve decimation, filtering and data formatting.

Table 1. Summary of related work targeting audio signal arrangements on FPGAs.

Reference	Application	Year	Type of MIC	Model of MIC	MICs Per Array	FPGA	Operations
[3]	Acoustic Data Acquisition System	2004	Analog ECM	Panasonic WM-54BT	1020	Xilinx 3000E	Data formatting and transmission
[48,50]	Acoustic Data Acquisition System	2010	Analog MEMS	Knowles Acoustics SPM0208	52	Xilinx Virtex-4 XC4VFX12	Sampling, formatting and transmission
[51]	Sound-Source Location	2011	Analog MEMS	Knowles Acoustics SPM0208	52	Xilinx Virtex-4 XC4VFX12	Sampling, formatting and transmission
[52]	Calibration for Acoustic Imaging	2014	Digital MEMS	Knowles Acoustics SPM0405HD4	64	FPGA array PXI-7854R	Data Acquisition
[53]	Evaluation of Stent Effectiveness	2016	Digital MEMS	Analog Devices ADMP521	4	Unspecified FPGA	Data acquisition and transmission

4.2. FPGA-Based Sound Locators

Nowadays many applications need to determine the sound source locations with a different degree of accuracy, timing demand and power efficiency. Several sound source location algorithms are used to perform this task. These algorithms determine the Direction-of-Arrival (DoA) of the sound wave and can be classified according to the method used [19,54]:

1. Time-Difference of Arrival (TDOA)
2. Steered Response Power (SRP)
3. High-Resolution Spectral Estimation (HRSE)

TDOA-based sound locators derive their source location from the calculation of the time-delay estimation relative to pairs of spatially separated microphones, and the knowledge of the fixed position of the microphones in the array. The second class refers to those sound locators whose estimation is based on maximizing the SRP of a beamformer. The last category includes those methods relying on an application of a signal correlation matrix.

4.2.1. FPGA-Based Designs of TDOA-Based Sound Locators

Relevant FPGA-based designs using TDOA for sound source localization are summarized in Table 2. The authors in [55] are among the first ones to fully embed the operations required to locate sound sources on an FPGA-based architecture. Their solution uses a general cross-correlation (GCC)-based TDOA to locate the sound sources and it reaches up to 12 microphones processed in parallel by placing multiple TDOA estimation modules. Moreover, the authors estimate that a larger version including up to 100 microphones can be supported when operating at 100 MHz, due to further readjustments and optimizations on their design. To overcome the high resource consumption, the authors in [56] propose a less-resource demanding TDOA algorithm to locate the

sound sources. Instead of using GCC, which requires extensive multiplications, their architecture implements a variation of TDOA based on the Average Magnitude Difference Function (AMDF), which only demands basic operations such as accumulations and subtractions. The authors also compare their architecture against the state-of-the-art, such as [55], obtaining similar performance with a much less complex and hardware demanding algorithm.

Table 2. Summary of FPGA-based microphone arrays related work using the TDOA estimation to perform the sound source location.

Reference	Application	Year	Type of MIC	Model of MIC	MICs Per Array	FPGA	Source Location Technique
[55]	Sound-Source Localization	2003	Not Specified	Not Specified	2	Xilinx Virtex-II 2000	GCC-TDOA
[57,58]	Countersniper System	2005	Analog ECM	Not Specified	3	Xilinx XC2S100 FPGA or ADSP-218x DSP	Shockwave and Muzzle Blast Detectors
[56]	Sound-Source Localization	2009	Analog	Not Specified	2	Altera DE2-70 Cyclone-II	AMDF-based TDOA
[59]	Sound-Source Localization	2010	Analog ECM	Not Specified	8	Xilinx Spartan-3 XC3S200	MCALD
[60]	Sound-Source Localization	2015	Analog	Not Specified	8	Xilinx Spartan-3E XC3S400	SNN and TDOA
[61]	Speech Enhancement	2017	-	MS Kinect microphones	2	Xilinx Spartan-6 LX45	TDOA
[62]	Sound-Sources Localization	2017	Analog ECM	Not Specified	4	Xilinx Zynq 7020	GCC-TDOA

The sound source detection using distributed microphone arrays is usually demanded by military applications such as sniper detection and localization, or man-wearable passive acoustic arrays to detect gunshots [63]. Several counter-sniper systems have been proposed in [57,58,64] using FPGAs to manipulate the incoming data from microphone arrays. The real-time source location is possible due to fully embedding their TDOA system on the FPGA, consisting of a shockwave and a muzzle blast detector [64]. Moreover, the Bluetooth communication between the distributed acoustic sensing nodes is managed by a picoBlaze soft-core processor on the node's FPGA. This distribution of the computational tasks between the FPGA logic and the soft or hard-core processors reduces the overall effort required to develop the HDL description of certain processes, such as the communication management. The combination of FPGAs with hard-core processors can also be used for more computational intensive tasks. For instance, the authors in [62] propose the use of a SoC FPGA, which embed a hard-core processor, together with a microphone array composed of 4 elements to locate sound sources. While the FPGA part computes the cross-correlations between pairs of microphones to the GCC-TDOA, the hard-core processor estimates the 3D location of sound sources. This system only needs 28 ms to compute the six TDOA, required to cross-correlate the 4 microphones.

Although simple TDOA-based approaches use FPGAs as part of the sound locator systems [59,61], innovative designs like [60] consider the use of neural networks for acoustic applications. Their architecture embeds on an FPGA a Spike Neural Network (SNN) whose first layer computes the TDOA for each pair of microphones while the second layer evaluates the response for the beamed orientations.

4.2.2. FPGA-Based Designs of SRP-Based Sound Locators

Sound source locators based on SRP use spatial filtering techniques to perform their estimations. Spatial filtering techniques, more-known as beamforming [38], enable the steering of the microphone array response to beam particular directions. As previously discussed, some signals experience constructive interferences while others experience destructive interference depending on the focused direction. Nonetheless, the complexity of steered-beamformers leads to a higher computational

demand when compared to TDOA-based locators [39]. Table 3 summarizes the most relevant features of the FPGA-based designs using SRP-based beamforming for sound source localization.

Conventional beamforming techniques are largely used in speech-array applications such as speech enhancement in conference rooms, allowing the audio enhancement of the speaker's speech while suppressing the surrounding noise. A real-time acoustic beamforming system, composed of 12 MEMS microphones, is proposed in [65] to improve speech intelligibility. The system uses an FPGA to implement a real-time, high-throughput and modular Delay-and-Sum beamformer. The selection of analogue MEMS microphones in their array demands the use of an ADC per microphone and a Serial Peripheral Interface (SPI) to communicate with the FPGA. Digital MEMS microphones, instead, do not only encapsulate the ADC circuitry but also have the potential to offer similar performance as high-quality analogue microphones for some applications. A comparison between digital MEMS microphones and analogue microphones is presented in [66]. This paper describes the design and implementation of an eight-element digital MEMS microphone array for distant speech recognition, which is compared to an analogue equivalent composed of eight high-quality analogue microphones. While the analogue system records at 48 KHz, the digital array uses an FPGA to demodulate the acquired audio signal. The beamforming stage consists of a Wiener filter, to remove the stationary noise of the channels, followed by a Filter-and-Sum beamformer. Different recognition models are applied to the recorded speeches in order to compare the response of both arrays. The results show that the absolute difference in word-error-rate (WER) between both arrays is around 14% worse for the digital array when a none recognition technique is applied. Despite recognition techniques decrease the WER of both types of arrays, the digital array shows a better response to such techniques, reducing to 4.5% the absolute difference in WER when compared to the analogue array.

Large microphone arrays especially benefit from using FPGAs to perform the beamforming operations. The same authors of [67] present in [68] a large microphone array composed of 300 digital MEMS microphones for indoor real-time speech acquisition. The array is decomposed in sub-arrays where parallel Delay-and-Sum beamforming operations are performed in the time domain. The quality of the captured speech meters cannot be, however, as good as with a hand-held microphone despite the large number of microphones composing the array. The use of Delay-and-Sum beamforming is also proposed in [34] for sound source location. The authors propose a FPGA-based implementation with a different strategy to process the data stream from the 52 microphones of their array. Instead of implementing individual filters for each microphone, the authors propose the execution of the Delay-and-Sum beamforming algorithm directly over the digital output signals from the microphones. This strategy has the potential benefit of drastically saving area and power consumption due to the significant reduction of the number of filters needed.

Table 3. Summary of FPGA-based microphone arrays related work using SRP-based sound locators.

Reference	Application	Year	Type of MIC	Model of MIC	MICs Per Array	FPGA	Source Location Technique
[66]	Distant Speech Recognition	2010	Digital MEMS	Knowles Acoustics SPM0205HD4	8	Xilinx Spartan-3A	Adaptive Filter-and-Sum
[68]	Speech Acquisition	2012	Digital MEMS	Not Specified	300	Multiple unspecified FPGAs	Time Domain Delay-and-Sum
[65]	Sound-Source Localization	2013	Analog MEMS	Not Specified	12	Xilinx Spartan-3E 1200	Time Domain Delay-and-Sum
[34]	Sound-Source Localization	2014	Digital MEMS	Analog Devices ADMP521	52	MicroSemi Igloo	Time Domain Delay-and-Sum
[69,70]	Sound-Source Localization	2015	Digital MEMS	Analog Devices ADMP621	33	Xilinx Spartan-6 LX25	Time Domain Delay-and-Sum
[71]	Deforestation Detection	2015	Digital MEMS	ST Microelectronics MP34DT01	8	MicroSemi Igloo 2	Time Domain Delay-and-Sum
[72]	Deforestation Detection	2016	Digital MEMS	ST Microelectronics MP32DB01	4, 8 or 16	Xilinx Spartan 6 FPGA	Time Domain Delay-and-Sum
[73]	Enhancement of Audio Signals	2016	Digital MEMS	AKUSTIC AKU242	7	Xilinx Zynq 7020	MVDR
[74,75]	Sound-Sources Localization	2016	Analog MEMS	InvenSense INMP504	48	Intel/Altera's DE1-SoC board	Adaptive Filter-and-Sum
[76]	Sound-Source Localization	2017	Digital MEMS	Analog Devices ADMP521	4, 8, 16 or 52	Xilinx Zynq 7020	Time Filter-Domain Delay-and-Sum
[77]	Hearing Aid System	2017	Analog MEMS	Analog Devices ADMP401	2	Xilinx Artix-7 A100	Adaptive Null-forming
[32]	Sound-Source Localization	2017	Digital MEMS	Analog Devices ADMP521	4, 8, 16 or 52	Xilinx Zynq 7020	Time Domain Filter-Delay-and-Sum
[78]	Hearing Aid System	2018	Digital MEMS	Knowles Acoustics SPM0405HD4H	48	Intel/Altera EP4CE15F17C8N	Time Domain Delay-and-Sum
[79]	Sound-Sources Localization	2018	Digital MEMS	InvenSense ICS-41350	4, 8, 16 or 52	Microsemi SmartFusion2 M2S050	Time Domain Delay-and-Sum

An example of a fully embedded beamforming-based acoustic system for localization of the dominant sound source is presented in [69,70]. Their FPGA-based system consists of a microphone array composed of up to 33 MEMS microphones. The tasks embedded on the FPGA involve the audio signal demodulation, the filtering, a Delay-and-Sum beamformer and a Root Mean Square (RMS) detector. The FPGA implementation is done using the Xilinx System Generator tool, which incorporates several libraries for the Matlab/Simulink tool enabling a high-level prototyping of FPGA designs.

The authors in [71] also fully embed a beamforming-based acoustic system composed of digital MEMS microphone arrays acting as node of a Wi-Fi-based Wireless Sensor Network (WSN) for deforestation detection. The power consumption, however, is a critical parameter for WSN-related applications. Their architecture uses an extremely low-power Flash-based FPGA, which allows to only consume 21.8 mW per node in the network, to compute their 8-elements microphone array. A larger version of this microphone array, composed of 16 microphones, is proposed by the same authors in [72]. Because of the additional computational operations, their architecture migrates to a Xilinx Spartan6 FPGA, leading to 61.71 mW of power consumption. Low-power architectures for WSN nodes to perform sound source localization are, however, not an exception. The authors in [79] propose a multi-mode architecture implemented on a extremely low power Flash-based FPGA, achieving a power consumption as low as 34 mW for a 52-elements microphone array. The proposed multi-mode architecture has been also exploited by the same authors in [76], where their architecture dynamically adapts the quality of the sound source location by adjusting the angular resolution of the beamed orientations. Their architecture also decomposes the microphone array in subarrays in order to modify in runtime the number of active microphones. This architecture is further accelerated in [32], where a performance analysis is discussed to perform sound localization in real-time.

The use of SoC FPGAs, like the Xilinx Zynq serie [47], for microphone arrays provides additional capabilities and facilitates the embedding of certain signal processing operations which otherwise would demand a higher effort to be embedded on a standalone FPGA. For instance, the authors in [73]

use a microphone array for real-time speech enhancement by embedding on a SoC FPGA all the operations to reduce the noise and interference. While the FPGA part manages the demodulation and filtering operations needed to retrieve the original audio signal, the hard-core processor performs the post-processing operations composed of the sound-source localization using SRP-PHAT and the speech enhancements using the Minimum Variance Distortionless Response (MVDR) beamformer to guarantee that the desired sound remains undistorted. Similarly, the work presented in [74,75] uses SoC FPGA architecture to locate sound sources by using a microphone array composed of 48 MEMS microphones. The system performs a Filter-and-Sum beamforming technique at the FPGA side while the hard-core processor is only used for communication.

More applications embed beamforming operations on FPGAs to perform sound localization. For instance, an FPGA-based hearing aid system using a microphone array is presented in [77]. The authors propose an adaptive beamforming algorithm [80] to introduce nulls in particular directions to cancel noise. Similarly, an FPGA-based hat-type hearing aid system composed of 48 MEMS microphones is presented in [78]. Their system provides a flexible sound directivity by allowing the user to emphasize the audio signals coming from several directions.

4.2.3. FPGA-Based Designs of HRSE-Based Sound Locators

The HRSE-based type of sound locators (Table 4) include adapted beamforming methods from the field of high-resolution spectral analysis such as Minimum Variance Spectral Estimation (MVSE) or the popular Multiple Signal Classification (MUSIC) algorithm based on eigenanalysis techniques. Despite this type of sound locators provide high-resolution results, typically consisting of sharp peaks, they also present limitations under certain acoustic conditions such as reverberations, which limits their adoption for speech-source location applications.

The authors in [67] evaluated the potential use of an acoustic sensor composed of a microphone array and an FPGA as part of a runway incursion avoidance system. The HRSE Capon beamformer, also known as MVSE, is preferred instead of the mostly-adopted Delay-and-Sum beamformer to obtain a higher resolution as result due to dynamically adapting the weights of the Capon beamformer to best fit the acoustic environment. The FPGA-based system samples, filters and transmits the data acquired by the microphone array composed of 105 MEMS microphones to a computer, where the beamforming operations and the additional tracking operations are performed.

Combinations of sound localization techniques are also possible when targeting distributed networks. The authors in [81–83] propose a distributed network of FPGA-based microphone array nodes for sound source localization and separation. The architecture includes a voice activity detector (VAD), the MUSIC algorithm for the sound-source location and the Delay-and-Sum beamforming technique for the sound sources separations at the network level. Each node is composed of a 16-element microphone array and a SoC FPGA, which uses the hard-core processor for the network communication. Moreover, the node's power consumption is reduced due to the VAD, which wakes up the system only when human voice is detected.

Table 4. Summary of FPGA-based microphone arrays related work using HRSE-based sound locators.

Reference	Application	Year	Type of MIC	Model of MIC	MICs Per Array	FPGA	Source Location Technique
[81–83]	Sound-Source Localization and Separation	2010	Analog ECM	Sony ECM-C10	16	Xilinx Virtex-4 FX (SZ410 Suzaku board)	MUSIC and Delay-and-Sum
[67]	Detection and Tracking of Aircrafts	2010	Digital MEMS	Not Specified	105	Unspecified FPGA	Capon

4.3. FPGA-Based Acoustic Imaging

Acoustic cameras are devices able to visually represent the sound waves due to the combination of multiple acquisition channels and beamforming techniques. The high computational needs of

acoustic cameras, where the multiple incoming audio signals from the microphone array are processed for a high number of beamed directions need an independent computation, making FPGAs the most suitable technology to build real-time acoustic imaging systems. The implementation of FPGA-based acoustic cameras have only be possible in recent years due to the increment of the computational power of current FPGAs.

A relatively high-resolution FPGA-based acoustic camera is proposed in [84], where all the operations needed to generate acoustic heatmaps are embedded in a Xilinx Spartan 3E FPGA. Despite their architecture achieves up to 10 frames-per-second (FPS) for acoustic image resolutions of 320×240 pixels, their architecture includes no filter beyond the inner filtering during the ADC conversion of the incoming data from their analogue ECMs. Furthermore, the acoustic images include ultrasound acoustic information since the frequency response reaches up to 42 kHz due to a missed high-pass filtering stage.

Another example is presented in [85], where the authors use an FPGA to implement a real-time acoustic camera. The authors justify the use of digital MEMS microphones in order to eliminate the analogue front-end and the use of ADCs. Moreover, their architecture uses a customized filter together with Delay-and-Sum beamforming operations in the FPGA. Nevertheless, the authors do not provide further information about the power consumption, the timing or FPS neither the output resolution, which is assumed to be 128×96 as is mentioned in [86].

The performance achieved with FPGA-based acoustic imaging systems has resulted in commercial products, as detailed in [87]. The authors describe a beamforming-based device (SM Instruments' Model SeeSV-S200 and SeeSV-S205) to detect squeak and rattle sources. The proposed devices are mainly composed of digital MEMS microphones arrays and an FPGA. Moreover, the authors affirm that, due to offering a good performance in high frequency range, beamforming techniques perform well detecting squeak and rattle noise since this kind of noise mainly consists of high frequency components. The FPGA implements the beamforming stage, supporting up to 96 microphones and generating sound representations up to 25 FPS with an unspecified resolution.

Digital MEMS microphones are combined with FPGAs for robot-based applications in [88]. The authors firstly propose an automated microphone array shape calibration in order to accurately estimate the array elements facing the noisy and reverberant environments of the real-world robotic operations. Such calibration is based on TDOA of moving noise sources. Despite up to 128 digital MEMS microphones can be managed due to the high I/O available in the chosen FPGA, only 44 microphones are used for their calibration example. The authors extend their audio acquisition system to perform acoustic imaging by performing a generalized inverse beamforming [89] on a standard laptop. Their system reaches up to 60 FPS with an unspecified resolution, thanks to the multi-thread computation of the Fast-Fourier Transforms (FFTs) required for the beamforming operations.

Instead of beamforming in the frequency domain, the authors in [90] present a time-domain Linearly Constrained Minimum Variance (LCMV) beamformer [91] embedded on an FPGA. The aim of the system is to visually track auto vehicles and to characterize the acoustic environment in real-time. The microphone array is composed of 80 MEMS microphones, whose positions are determined by a multi-objective genetic algorithm [92]. The filter's coefficients required for their beamformer are generated and stored in the host-PC and loaded to the FPGA when needed. The FPGA is in charge of the filtering and the Filter-and-Sum beamforming operations. It is interesting to notice that the authors decided to use a HLS tool called Xilinx Vivado HLS [40] as part of the tool flow to implement the FPGA design.

A heterogeneous system is proposed in [93]. This system combines an FPGA, an embedded hard-processor, a Graphics Processing Unit (GPU) and a computer desktop to generate acoustic images using a planar MEMS microphone array composed of 64 digital MEMS microphones. Their modular approach allows to distribute the computational operations between the different devices by using LabVIEW from National Instruments [94]. In the full embedded mode, the Xilinx Zynq 7010 performs

the signal demodulation and filtering on the FPGA part while computing the beamforming operations on the embedded hard-processor. This acoustic imaging system is used to estimate the real position of the fan inside a fan matrix [95] and to create virtual microphone arrays for higher resolution acoustic images in [96].

FPGAs can be also combined with desktop PCs to perform 3D impulsive sound-source localization method, as in [97,98]. The proposed system computes the Delay-and-Sum beamforming operation on the PC while the FPGA filters the acquired audio signals and displays through VGA the acoustic heatmap generated on the PC.

Table 5 summarizes the most relevant features of the FPGA-based acoustic cameras. Parameters like the FPS or the acoustic image resolution, which determine the number of beamed directions, reflect the performance and the image quality of the FPGA-based architectures respectively. Most of the architectures summarized in Table 5 do not only use an FPGA for the acoustic imaging operations. They usually combine FPGAs with other hardware accelerators like GPUs [93] or with multi-core processors [98] to compute the beamforming operations, the filter's coefficients or to generate the visualization. Nevertheless, there is not a clear answer why recent FPGA-based acoustic cameras are not fully embedded, as in [84].

Table 5. Summary of FPGA-based microphone arrays related work targeting acoustic imaging applications.

Reference	Application	Year	Type of MIC	Model of MIC	MICs Per Array	Device	Beamforming Algorithm	Resolution	Real-Time	Power
[84]	Acoustic Imaging	2010	Analog ECM	Ekulit EMY-63M/P	32	Xilinx Spartan-3E XC3S500E	Time-Domain Delay-and-Sum	320×240	10 FPS	Not Specified
[88]	Robotic Applications	2012	Digital MEMS	Not Specified	44	Xilinx Spartan-6 LX45	Frequency-Domain Generalized Inverse	Not Specified	60 FPS	Not Specified
[85]	Acoustic Imaging	2014	Digital MEMS	Not Specified	32	Xilinx Spartan-6 XC6SLX16	Time-Domain Delay-and-Sum	128×96	Not Specified	Not Specified
[87]	Detection squeak and rattle sources	2014	Digital MEMS	Analog Devices ADMP 441	30 or 96	National Instruments sbRIO or FlexRIO (Xilinx Zynq 7020)	Time-Domain Unspecified Beamforming	Not Specified	25 FPS	Not Specified
[90]	Acoustic Imaging	2015	Analog MEMS	InvenSense ICS 40720	80	Xilinx Virtex-7 VC707	Linearly Constrained Minimum Variance	61×61	31 FPS	75 W
[93,95,96]	Acoustic Imaging	2016	Digital MEMS	ST Microelectronics MP34DT01	64	National Instruments myRIO (Xilinx Zynq 7010)	Frequency-Domain Wideband	40×40	33.4 ms to 257.3 ms	Not Specified
[97,98]	Acoustic Imaging	2017	Analog MEMS	ST Microelectronics MP33AB01	25	Xilinx Artix-7 XC7A100T	Time-Domain Delay-and-Sum	Not Specified	Not Specified	Not Specified
[99]	Acoustic Imaging	2018	Digital MEMS	Knowles Acoustics SPH0641LU4H	16	Xilinx Zynq 7020	Time-Domain Delay-and-Sum	160×120 up to 640×480	32.5 FPS	Not Specified

5. FPGA-Based Architectures for Acoustic Beamforming

Applications related to sound source location or acoustic imaging are constrained and high-performance demanding. Such applications are capable to fully exploit many of the FPGA's features. An analysis of the existing FPGA-based architectures not only provides an idea about the most demanding features but also provides an inspiration on how FPGAs can be further exploited to satisfy future sound-related applications. Here, a detailed overview of the most performance demanding FPGA-based architectures which integrate acoustic beamforming in the embedded tasks is presented.

5.1. FPGA-Based Audio Signal Demodulators for Acoustic Beamforming

The architectures summarized in Tables 1–5 exemplify how FPGAs can perform many different operations required for sound-related applications using microphone arrays. These tasks change based on the number of microphones of the array, the type of microphone, the beamforming technique and the application. For instance, many architectures present similar audio signal demodulation strategies due to the selection of analog or digital microphones to compose their sensor array. The demodulation of the acquired acoustic signal is one type of operation where FPGAs have been replacing DSPs for signal processing applications in the recent past. The acquired data from the microphones demands certain signal processing operations to retrieve the original audio signal, which are ideally embedded on FPGAs. Figure 6 depicts the operations performed on the FPGAs when using analog microphones. Such type of microphones requires a signal conditioning circuitry and ADCs before interfacing the FPGA. Unlike digital microphones, the input data is already audio signal and does not need any demodulation. This fact facilitates the embedding of the beamforming operations since FPGA's resources are not consumed for implementing these demodulation operations. Filtering is required in order to remove noise from the audio signal. One may notice how the implementation of the filtering is introduced after the beamforming operation in order to reduce the noise and the signals at undesired frequencies. Due to the relatively low resource consumption of Cascaded Integrator-Comb (CIC) filters [100] and serial FIR filters, they are both embedded together with the beamformer. Thus, the architectures (a), (b) and (c) in Figure 6, correspond to the designs presented in [65,84,90] respectively, are able to fully embed the filtering and the beamforming operations on the FPGA. For instance, the filter operations in [65] are done in a two-stage filtering process composed of one CIC and one low-pass FIR filter. One of the advantages of filtering after the Delay-and-Sum beamforming algorithm is the reduction of the consumed area. The overall number of filters is no longer associated to the number of microphones of the array, because the incoming audio signals are delayed and summed before the filtering operations. As a consequence, the area consumption is significantly reduced. The available resources of the target FPGA and the unnecessary audio demodulation thanks to using analog microphones determine the embedding of the operations further than the type of application. An exception is the architecture (d) in Figure 6, presented in [98], where the authors prefer to compute the beamforming operations in a PC.

5.2. Partially Embedded FPGA-Based Acoustic Beamformers

The PDM demodulation starts with a PDM to PCM conversion by using a CIC filter (also known as SINC filters) [100] at the first stage of decimation to reduce the sampling frequency. This component is followed by a couple of half-band low-pass decimation filters and a low-pass FIR filter to further reduce the sampling frequency and to remove the high-frequency noise introduced by the sigma-delta converter which is integrated in the digital MEMS microphones. The described PDM demodulation is applied in [66,88] as shown in architectures (a) and (b) in Figure 7 respectively.

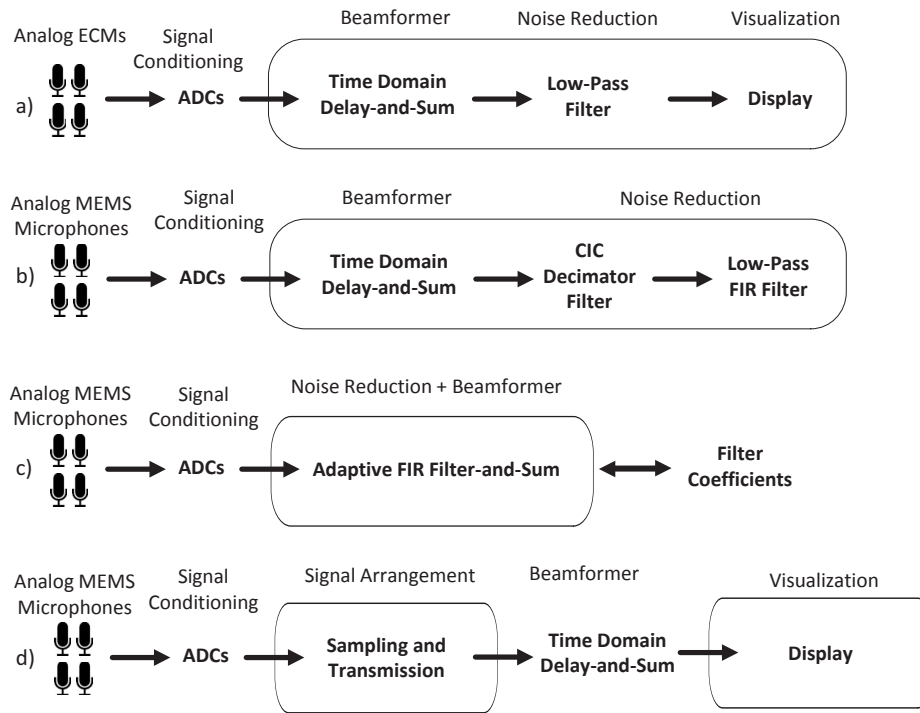


Figure 6. FPGA-Based architectures for analog microphone arrays. The framed operations are performed on the FPGA. The architectures depicted in (a–d) are proposed in [65,84,90,98] respectively.

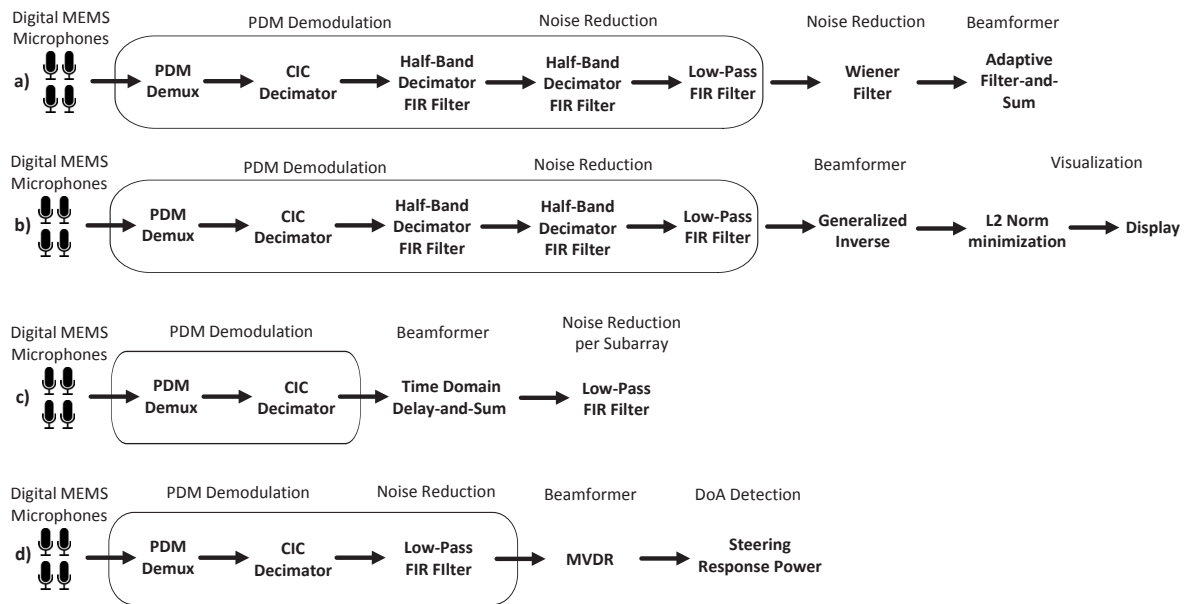


Figure 7. FPGA-Based architectures for acoustic beamforming in time domain using digital MEMS microphones. The framed operations are performed on the FPGA. The architectures depicted in (a–d) are proposed in [66,68,73,88] respectively.

The cost of embedding the PDM demodulation on the FPGA is directly determined by the number of microphones in the array since the PDM demodulation proposed in [100] demands several cascaded filters per digital MEMS microphone. Optionally, the number of cascaded filters can be reduced when the input signals are Time-Division Multiplexed (TDM). The cost, however, are additional clock cycles to reset the filters, to operate at a different clock domain or to use extra internal memory to preserve

the filter's intermediate operations per input signal. The architectures (c) and (d), proposed in [68,73] respectively, present simplified demodulations where the cascaded half-band filters are replaced by a higher-order low-pass FIR filter or simply removed. For instance, the architecture in [68] drastically reduces the number of filters by removing the half-band low-pass filters and reducing the number of low-pass FIR filters to correspond to the number of subarrays. Although the resource consumption is significantly reduced thanks to simplifying the PDM demodulation, the beamforming operations are, in contrast, not embedded on the FPGA. This fact is independent of the time or frequency domain of the beamforming technique, as shown in Figure 8 where several architectures performing beamforming techniques in the frequency domain are depicted. Furthermore, the architectures (a) and (b) in Figure 8, corresponding to the architectures presented in [67,93] respectively, do not even embed the FFT operations due to their high resource consumption, even considering that FPGAs support high-performance FFT implementations [101].

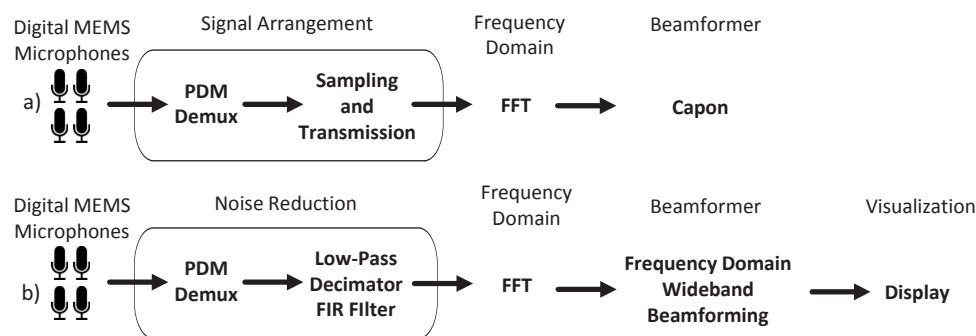


Figure 8. FPGA-Based architectures for acoustic beamforming in frequency domain using digital MEMS microphones. The framed operations are performed on the FPGA. The architectures depicted in (a) and (b) are proposed in [67,93] respectively.

5.3. Embedded FPGA-Based Acoustic Beamformers

Fully-embedded architectures are, unexpectedly, rare. Figure 9 depicts a few examples of FPGA-based architectures where the PDM demodulation and the beamforming operations are embedded. The available resources and the achievable performance that current FPGAs provide, facilitate the signal processing operations demanded by the PDM demodulation and the beamforming techniques. The authors in [69] propose the architecture (a) depicted in Figure 9, which fully embeds the PDM demodulation detailed in [30] together with a Delay-and-Sum beamformer. The sound-source localization is performed through the RMS calculation. The architecture (b) depicted in Figure 9 includes a novel PDM demodulation based on Cascaded Recursive-Running Sum (CRRS) filters to build their acoustic camera in [85]. Different authors in [34,72] use the architecture (c) depicted in Figure 9. Instead of implementing individual PDM demodulators for each microphone, the authors propose the execution of the Delay-and-Sum beamforming algorithm over the PDM signals. The output of the Delay-and-Sum, which is no longer a 1-bit PDM signal, is filtered by windowing and processed at the frequency domain. This strategy has the potential benefit of saving area and power consumption due to the drastic reduction of the number of filters needed as shown in [79]. The architecture (e) in Figure 9 also performs the Delay-and-Sum beamforming algorithm over the PDM signals but does not calculate the SRP in the frequency domain. Similarly, the architecture (d) in Figure 9 also calculates the beamforming algorithm and the SRP in the time domain. The architecture (d) depicted in Figure 9 is a generalization of the different versions of a Filter-Delay-and-Sum beamformer initially presented in [76], accelerated in [32] and improved in [99]. Nevertheless, they all have in common a filtering stage composed of several cascaded filters before performing a Delay-and-Sum beamformer. While architectures like (e) present a lower resource consumption, they are not as fast as architectures of the form (d), which due to their specific characteristics, can be further accelerated [32].

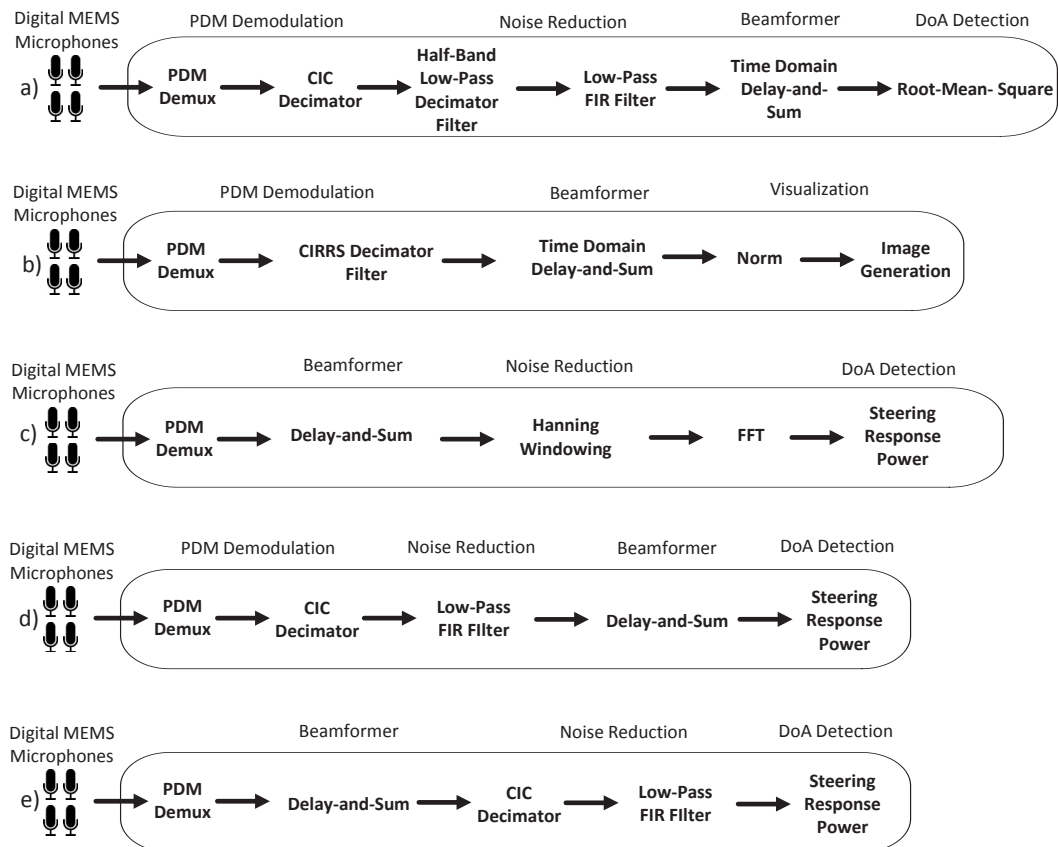


Figure 9. FPGA-Based architectures fully embedding all the operations to satisfy real-time application demands. The framed operations are performed on the FPGA. The architectures depicted in (a–d) are proposed in [30,32,34,79,85] respectively.

The type of microphones determines what operations must be embedded on the FPGA. For instance, analog microphones demand external circuitry for the ADC conversion before interfacing the FPGA while digital PDM microphones require cascaded filters for the signal demodulation. Nowadays FPGAs provide enough resources to perform in real time complex beamforming algorithms involving tens of microphones. Nevertheless, the choice of the architecture is strongly linked to the characteristics and constraints of the target application.

6. Trends

The high level of parallelism achievable on FPGAs well-suits not only for multiple customized data path processes, such as audio signal demodulation (Table 1) but also to perform complex audio beamforming operations (Sections 4.2.2, 4.2.3 and 4.3). The embedding of such computational-demanding signal processing operations on FPGAs has only been possible in the recent past due to several factors. Figure 10 depicts the categorization of the presented related work. The evolution of the number of microphones in the array over the last years reflects some interesting facts of how FPGAs are used. Notice, however, that some designs like [68] are not included since their FPGA-based system is composed of multiple FPGAs. In the early 2000s, the first uses of FPGAs to compute microphone arrays signals were to mainly embed simple applications for sound-source location applying TDOA [55,57]. Such applications require a minimum number of microphones since the traditional GCC used for TDOA grows exponential with the number of microphones. FPGAs started to be seriously considered in the following decade, being involved in a broader type of applications such as in [72,78,84,98]. Several factors might justify the increasing adoption of the FPGAs' technology:

1. Cheaper, smaller and fully integrated microphones, like digital MEMS microphones [102], facilitate the construction of larger arrays, increasing the computational demands beyond of what microprocessors or DSPs can deliver.
2. FPGAs have also benefited from the Moore's law [103], and due to a higher transistor integration in the same die, FPGAs offer larger reconfigurable resources.
3. Advances in the FPGAs' design tool chain, like the HLS tools [40,104], have reduced the overall effort to develop and to accelerate new and existing applications on FPGAs.

Cheaper and smaller microphones facilitate the construction of larger arrays. The MEMS technology to build microphones has been available since the early 2000s [22], but only introduced in commercial devices in 2006 after Apple introduced a 3-microphone array in their iPhone 6. The replacement of ECMs by MEMS microphones to build microphone arrays only started around 2010, when Knowles Acoustics lost the monopoly of the MEMS microphones commercialization [105]. As a result, several acoustic applications, mostly related to sound source location, have been implemented on FPGA (Figure 10).

In the early 2000s, FPGAs were only considered for audio signal demodulation of the incoming data stream from microphone arrays composed of tens to hundreds of microphones [3] or as sound source locators for 2-microphone arrays [55]. Over the last years, applications demanding a relatively large number of microphones have also used FPGAs to embed the most computational demanding operations. The trend is to embed more complex applications on the FPGA. Despite some FPGA-based architectures still allocate on the FPGA's resources the audio demodulation operations, over the last years FPGAs are no longer used exclusively for audio demodulation but also to embed complex applications. Current FPGAs provide a larger amount of resources, including DSPs and internal blocks of memory (BRAM), allowing the implementation of more complex architectures targeting real-time signal processing applications. For instance, SoC FPGAs such as Xilinx Virtex-II FPGAs used in [55] have been replaced by larger SoC FPGA such as Xilinx Zynq SoC FPGA-based board for acoustic imaging, as in [93].



Figure 10. Categorization of the FPGA-based microphone arrays publications.

New HLS tools such as Xilinx Vivado HLS [40], LabVIEW from National Instruments [94], Xilinx System Generator [104] or HDL Coder for Matlab/Simulink [106], reduce significantly the overall development cost. For instance, Xilinx System Generator [104] incorporates several libraries in Matlab/Simulink tool which allows a high-level prototyping of FPGA designs. HLS tools have been

helpful to develop complex designs, as in [66,69,90,93]. The reduction of the overall effort to develop complex designs on FPGAs is one of the advantages of such tools. Moreover, the distribution of the computational tasks in heterogeneous systems, such as SoC FPGAs or including GPUs, as in [93], is simplified due to design at a higher level.

Although the spatial resolution increases with a large number of microphones per array [107], the additional benefit of incrementing the number of microphones decreases when considering the increment of the computational demand. It comes from the fact that the added value of increasing the number of microphones starts to decrease after certain amount. For instance, the number of microphones per FPGA not only did not increase over the last two years, but even decreased. Moreover, the integration of a large number of microphones in a planar array becomes extremely challenging without increasing the microphone spacing, leading to microphone arrays of several meters long [87].

The decrement of the number of microphones per array also occurs for applications related to acoustic imaging. Acoustic cameras are extremely performance demanding, specially when targeting real-time. The fact is that the computation in parallel of tens to hundreds of incoming signals from microphones can simply consume the FPGA's available resources. Moreover, acoustic cameras need to steer to hundreds of thousands orientations in order to provide acceptable image resolutions [85]. Therefore, the trend of FPGA-based acoustic cameras is to converge to a balance between the FPGA's resource consumption, the target performance and the desired acoustic image resolution.

A trend of FPGA-based architectures for microphone arrays is to embed more complex acoustic applications while reducing the number of microphones of the array. FPGAs are no longer only considered for audio signal demodulation but also as a platform on which computational demanding acoustic applications such as acoustic imaging applications can be embedded. Constraints, such as a real-time response or power efficiency, become more relevant when targeting new acoustic applications like acoustic imaging applications or WSNs-based applications. Modern FPGAs not only provide a higher number of resources where to embed complex applications, but also integrate CPUs and even GPUs in the same die [108] or become extremely power efficient when considering the Flash-based FPGAs [109].

The FPGA technology, however, is far of being exploited. Nowadays FPGAs present interesting features which have not been already explored such as dynamic partial reconfiguration. Therefore, it is expected that the incoming FPGA-based acoustic applications not only fully embed their operations on the FPGAs but also exploit some of the unique features that this technology offers.

7. Challenges and Research Opportunities

The current state-of-the-art of FPGA-based acoustic applications have been summarized in Tables 1–5. Although the characteristics of these FPGA-based designs have been discussed in the previous sections, important features, such as achievable performance or the power efficiency have not been analyzed. Their relevance for the nowadays acoustic applications is, however, critical when choosing technology.

Performance

FPGA's technology offers unique features which could satisfy the most performance demanding acoustic applications. The low latency usually required by acoustic applications such as speech enhancement is achievable on FPGAs when fully embedding the signal processing operations. Most of the current FPGA-based designs still perform the computational-hungry operations on general purpose processors, demanding high-bandwidth I/O connection in order to satisfy the low latency required for real-time applications. Besides such constraints, only a few designs fully embed all computations on the FPGA [34,65,69,72,84]. Furthermore, several architectures, such as those found in [62,73,74,93], have already considered SoC FPGAs to distribute the tasks between the different technologies. These heterogeneous platforms present new opportunities to combine acoustic applications with other types

of applications. For instance, the use of different types of sensors (e.g. Infrared cameras) can be combined with acoustic microphone arrays for smart surveillance. Acoustic cameras already combine traditional RGB cameras by overlapping images. Many applications could exploit such combination by using SoC FPGAs to process in real-time the sensing information from each device.

Dynamism

Despite real-life environments present dynamic acoustic responses, most of the FPGA-based architectures cannot adapt their response to different acoustic contexts. Despite certain solutions like [66,74,77] consider adaptive beamforming techniques, the overall response of the system varies in a short range. Many applications need to change their behaviors based on the acoustic context [110], such as applications targeting specific sound sources [111], where a simple adaptation of the filters is not enough and a different feature extraction is needed. Adjustments on the number of active microphones, the acquisition time or the target sound source demand the implementation of complex context-switch controllers. FPGAs provide a unique feature which allows to partially reconfigure parts of the embedded functionality in runtime. FPGA's dynamic partial reconfiguration [112,113] provides the context-switch capability which is not present in other technologies. An example of the potential benefit of using partial reconfiguration is shown in [76], where the proposed architecture uses a low-level reconfiguration to dynamically adjust the angular resolution of their sound locator. For instance, the authors in [114] present a SoC FPGA implementation of a Frost's beamformer. Despite the authors do not target microphone arrays, the architecture seems to be compatible and the principles of their approach are applicable for different types of sensor arrays. Their architecture presents two interesting approaches. Firstly, the distribution of the computations between the ARM Cortex-A9 hard-core Processor System (PS) and the Programmable Logic (PL), and secondly, the use of partial reconfiguration to adjust the Frost's beamformer.

Power Efficiency

The dynamism required by advanced acoustic applications also leads to power efficiency. FPGAs are well-known by their power efficiency, offering a higher Operations-per-Watt than general purpose processors [115] and different hardware accelerators, such as GPUs [116,117]. WSN-related applications are very sensitive to power efficiency since the network is often built on battery-based nodes. The use of microphone arrays as sensing nodes of WSN applications demand power efficient solutions. FPGA-based solutions have already shown how the power efficiency of microphone arrays can be increased. For instance, the power consumption per microphone has decreased from 400 mW using DSP in [2] to only 77 mW per microphone in [55], and more recently, to only 27.14 mW for the overall system in [72].

One can conclude that FPGA's characteristics have not been fully exploited and many acoustic-related applications can benefit from this technology. Recent heterogeneous FPGA SoCs provide enough resources to not only embed acoustic applications but also extend functionalities by combining with different types of applications while satisfying the performance and the power efficiency demands. Acoustic applications involving machine learning, such as acoustic scene recognition [118,119] or learning situations in home environments [120], can directly benefit from FPGA's features. Further than FPGA SoCs, FPGAs standalone offer a unique feature, such as partial reconfiguration, which can certainly provide the flexibility that many multi-modal applications such as human-robot interaction [121] or multimodal acoustic imaging demand [122]. FPGAs still have much to offer to acoustic applications using microphone arrays.

8. Conclusions

In this paper, we have not only shown how FPGAs are used for many different acoustic applications, with a clear trend in the complexity of the embedded operations, but also the pending challenges that FPGA-based designs must face in the incoming years. Nevertheless, FPGAs have

demonstrated to be an ideal platform where to embed the most demanding acoustic applications. Despite the variety of demands that multiple embedded algorithms used for similar acoustic applications present, FPGAs have proven to be flexible platforms able to satisfy performance and power demands. Moreover, the detailed FPGA-based architectures have shown that acoustic beamforming is no longer used only for sound localization but it has been extended for acoustic imaging applications. Dynamism, performance and power efficiency are, however, still challenges to be faced.

Author Contributions: Conceptualization, Methodology, Investigation and Writing-Original Draft Preparation: B.d.S. Supervision, Writing-Review & Editing: A.B. Supervision and Funding Acquisition: A.T.

Funding: This work was supported by the European Regional Development Fund (ERDF) and the Brussels-Capital Region-Innoviris within the framework of the Operational Programme 2014-2020 through the ERDF-2020 Project ICITYRDLBRU. This work was also partially supported by the CORNET project "DynamIA: Dynamic Hardware Reconfiguration in Industrial Applications" [123] which was funded by IWT Flanders with reference number 140389.

Acknowledgments: The authors would like to thank Xilinx for the provided software and hardware under the University Program Donation.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Silverman, H.F.; Patterson, W.; Flanagan, J.L.; Rabinkin, D. A digital processing system for source location and sound capture by large microphone arrays. In Proceedings of the 1997 IEEE International Conference on IEEE 1997 ICASSP-97 Acoustics, Speech, and Signal Processing, Munich, Germany, 21–24 April 1997; Volume 1, pp. 251–254.
2. Silverman, H.F.; Patterson, W.R.; Flanagan, J.L. The huge microphone array. *IEEE Concurr.* **1998**, *6*, 36–46.
3. Weinstein, E.; Steele, K.; Agarwal, A.; Glass, J. *Loud: A 1020-Node Modular Microphone Array and Beamformer for Intelligent Computing Spaces*; Technical Report; Massachusetts Institute of Technology: Cambridge, MA, USA, 2004.
4. Tiete, J.; Domínguez, F.; da Silva, B.; Touhafi, A.; Steenhaut, K. MEMS microphones for wireless applications. In *Wireless MEMS Networks and Applications*; Elsevier: New York, NY, USA, 2017; pp. 177–195.
5. Moore, D.C.; McCowan, I.A. Microphone array speech recognition: Experiments on overlapping speech in meetings. In Proceedings of the 2003 IEEE International Conference on (ICASSP'03) IEEE Acoustics, Speech, and Signal Processing, Hong Kong, China, 6–10 April 2003; Volume 5, p. 497.
6. Microsemi, Alexa Voice Service Development Kit (ZLK38AVS2). Available online: <https://www.microsemi.com/product-directory/connected-home/4628-zlk38avs> (accessed on 15 July 2018).
7. Widrow, B.; Luo, F.L. Microphone arrays for hearing aids: An overview. *Speech Commun.* **2003**, *39*, 139–146.
8. Kellermann, W. Strategies for combining acoustic echo cancellation and adaptive beamforming microphone arrays. In Proceedings of the IEEE International Conference on ICASSP-97 1997, Acoustics, Speech, and Signal Processing, Munich, Germany, 21–24 April 1997; Volume 1, pp. 219–222.
9. Villacorta, J.J.; Jiménez, M.I.; Val, L.D.; Izquierdo, A. A configurable sensor network applied to ambient assisted living. *Sensors* **2011**, *11*, 10724–10737.
10. Grenier, Y. *A Microphone Array for Car Environments*; IEEE: Piscataway, NJ, USA, 1992; pp. 305–308.
11. Fuchs, M.; Haulick, T.; Schmidt, G. Noise suppression for automotive applications based on directional information. In Proceedings of the (ICASSP'04) IEEE International Conference on IEEE Acoustics, Speech, and Signal Processing, Montreal, QC, Canada, 17–21 May 2004; Volume 1, pp. I–237.
12. Izquierdo-Fuente, A.; Del Val, L.; Jiménez, M.I.; Villacorta, J.J. Performance evaluation of a biometric system based on acoustic images. *Sensors* **2011**, *11*, 9499–9519.
13. Del Val, L.; Izquierdo-Fuente, A.; Villacorta, J.J.; Raboso, M. Acoustic biometric system based on preprocessing techniques and linear support vector machines. *Sensors* **2015**, *15*, 14241–14260.
14. Blumstein, D.T.; Mennill, D.J.; Clemins, P.; Girod, L.; Yao, K.; Patricelli, G.; Deppe, J.L.; Krakauer, A.H.; Clark, C.; Cortopassi, K.A.; et al. Acoustic monitoring in terrestrial environments using microphone arrays: Applications, technological considerations and prospectus. *J. Appl. Ecol.* **2011**, *48*, 758–767.

15. Nakadai, K.; Kumon, M.; Okuno, H.G.; Hoshiba, K.; Wakabayashi, M.; Washizaki, K.; Ishiki, T.; Gabriel, D.; Bando, Y.; Morito, T.; et al. Development of microphone-array-embedded UAV for search and rescue task. In Proceedings of the 2017 IEEE/RSJ International Conference on IEEE Intelligent Robots and Systems (IROS), Vancouver, BC, Canada, 24–28 September 2017; pp. 5985–5990.
16. Case, E.E.; Zelnio, A.M.; Rigling, B.D. Low-cost acoustic array for small UAV detection and tracking. In Proceedings of the IEEE National, NAECON 2008 Aerospace and Electronics Conference, Dayton, OH, USA, 16–18 July 2008; pp. 110–113.
17. Hoshiba, K.; Washizaki, K.; Wakabayashi, M.; Ishiki, T.; Kumon, M.; Bando, Y.; Gabriel, D.; Nakadai, K.; Okuno, H.G. Design of UAV-embedded microphone array system for sound source localization in outdoor environments. *Sensors* **2017**, *17*, 2535.
18. Doclo, S.; Moonen, M. GSVD-based optimal filtering for single and multimicrophone speech enhancement. *IEEE Trans. Signal Process.* **2002**, *50*, 2230–2244.
19. Brandstein, M.; Ward, D. *Microphone Arrays: Signal Processing Techniques and Applications*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013.
20. Stetron, Stetron Electret Condenser Microphones (ECM) Catalog. Available online: <https://www.stetron.com/category/microphones/> (accessed on 15 July 2018).
21. Hosiden, Guide for Electret Condenser Microphones. <http://www.es.co.th/Schemetic/PDF/KUC.PDF> (accessed on 15 July 2018).
22. Neumann, J.; Gabriel, K. A fully-integrated CMOS-MEMS audio microphone. In Proceedings of the 12th International Conference on, 2003 IEEE TRANSDUCERS, Solid-State Sensors, Actuators and Microsystems, Boston, MA, USA, 8–12 June 2003; Volume 1, pp. 230–233.
23. Lewis, J. Analog and Digital MEMS Microphone Design Considerations. Available online: <http://www.analog.com/media/en/technical-documentation/technical-articles/Analog-and-Digital-MEMS-Microphone-Design-Considerations-MS-2472.pdf> (accessed on 15 July 2018).
24. Lewis, J.; Moss, B. MEMS microphone: The future for hearing aids. *Analog Dialog.* **2013**, *47*, 3–5.
25. Song, Y. *Design, Analysis and Characterization of Silicon Microphones*; State University of New York: Binghamton, NY, USA, 2008.
26. Fairchild, FAN3850A Microphone Pre-Amplifier with Digital Output. Available online: <http://www.onsemi.com/PowerSolutions/product.do?id=FAN3850A> (accessed on 15 July 2018).
27. Philips Semiconductors, I2S Bus Specification. Available online: <https://www.sparkfun.com/datasheets/BreakoutBoards/I2SBUS.pdf> (accessed on 15 July 2018).
28. MEMS Analog and Digital Microphones, ST Microelectronics. Available online: http://www.st.com/content/ccc/resource/sales_and_marketing/promotional_material/flyer/83/e5/0d/ba/f0/40/4b/a3/flmemsmic.pdf/files/flmemsmic.pdf/jcr:content/translations/en.flmemsmic.pdf (accessed on 1 June 2018).
29. Lewis, J. Understanding Microphone Sensitivity. Available online: <http://www.analog.com/en/analog-dialogue/articles/understanding-microphone-sensitivity.html> (accessed on 31 July 2018).
30. Hegde, N. Seamlessly Interfacing MEMS Microphones with Blackfin Processors. EE-350 Engineer-to-Engineer Note. 2010. Available online: <http://www.analog.com/media/en/technical-documentation/application-notes/EE-350rev1.pdf> (accessed on 31 July 2018).
31. TDK Invensense, ICS-43434 Datasheet. Available online: <https://www.invensense.com/wp-content/uploads/2016/02/DS-000069-ICS-43434-v1.2.pdf> (accessed on 15 July 2018).
32. da Silva, B.; Braeken, A.; Steenhaut, K.; Touhafi, A. Design Considerations When Accelerating an FPGA-Based Digital Microphone Array for Sound-Source Localization. *J. Sens.* **2017**, *2017*, 6782176:1–6782176:20.
33. Seltzer, M.L. Microphone Array Processing for Robust Speech Recognition. Ph.D. Thesis, CMU, Pittsburgh, PA, USA, 2003.
34. Tiete, J.; Domínguez, F.; da Silva, B.; Segers, L.; Steenhaut, K.; Touhafi, A. SoundCompass: A Distributed MEMS Microphone Array-Based Sensor for Sound Source Localization. *Sensors* **2014**, *14*, 1918–1949.
35. Petersen, D.; Howard, C. Simulation and design of a microphone array for beamforming on a moving acoustic source. In Proceedings of the Acoustics 2013, Victor Harbor, Australia, 9–11 November 2013.
36. Sarradj, E. A generic approach to synthesize optimal array microphone arrangements. In Proceedings of the 6th Berlin Beamforming Conference, BeBeC-2016-S4, Senftenberg, Germany, 5–6 March 2016.

37. Krim, H.; Viberg, M. Two decades of array signal processing research: the parametric approach. *IEEE Signal Process. Mag.* **1996**, *13*, 67–94.
38. Van Veen, B.D.; Buckley, K.M. Beamforming: A versatile approach to spatial filtering. *IEEE ASSP Mag.* **1988**, *5*, 4–24.
39. Mucci, R. A comparison of efficient beamforming algorithms. *IEEE Trans. Acoust. Speech Signal Process.* **1984**, *32*, 548–558.
40. Xilinx Vivado High-Level Synthesis. Available online: <https://www.xilinx.com/products/design-tools/vivado/integration/esl-design.html>. (accessed on 1 June 2018).
41. Intel/Altera HLS Compiler. Available online: <https://www.altera.com/products/design-software/high-level-design/intel-hls-compiler/overview.html> (accessed on 15 July 2018).
42. Intel/Altera FPGA SDK for OpenCL. Available online: <https://www.altera.com/products/design-software/embedded-software-developers/opencl/overview.html> (accessed on 15 July 2018).
43. OpenRISC Project: Open processor for Embedded Systems. Available online: <https://openrisc.io/> (accessed on 30 July 2018).
44. RISC-V Foundation. Available online: <https://riscv.org/> (accessed on 30 July 2018).
45. Xilinx MicroBlaze Soft Processor Core. Available online: <https://www.xilinx.com/products/design-tools/microblaze.html> (accessed on 30 July 2018).
46. Intel/Altera Nios-II Soft Processor. Available online: <https://www.intel.la/content/www/xl/es/products/programmable/processor/nios-ii.html> (accessed on 30 July 2018).
47. Xilinx Zynq-7000 SoC. Available online: <https://www.xilinx.com/products/silicon-devices/soc/zynq-7000.html> (accessed on 1 June 2018).
48. Turqueti, M.; Saniie, J.; Oruklu, E. MEMS acoustic array embedded in an FPGA based data acquisition and signal processing system. In Proceedings of the 2010 53rd IEEE International Midwest Symposium on IEEE Circuits and Systems (MWSCAS), Seattle, WA, USA, 1–4 August 2010; pp. 1161–1164.
49. Turqueti, M.; Rivera, R.A.; Prosser, A.; Andresen, J.; Chramowicz, J. CAPTAN: A hardware architecture for integrated data acquisition, control, and analysis for detector development. In Proceedings of the 2008 NSS'08 IEEE Nuclear Science Symposium Conference Record, IEEE, Piscataway, NJ, USA, 25–28 October 2008; pp. 3546–3552.
50. Turqueti, M.; Kunin, V.; Cardoso, B.; Saniie, J.; Oruklu, E. Acoustic sensor array for sonic imaging in air. In Proceedings of the 2010 IEEE Ultrasonics Symposium (IUS), San Diego, CA, USA, 11–14 October 2010; pp. 1833–1836.
51. Kunin, V.; Turqueti, M.; Saniie, J.; Oruklu, E. Direction of arrival estimation and localization using acoustic sensor arrays. *J. Sens. Technol.* **2011**, *1*, 71–80.
52. Havránek, Z.; Beneš, P.; Klusáček, S. Free-field calibration of MEMS microphone array used for acoustic holography. In Proceedings of the 21st International Congress on Sound and Vibration, Beijing, China, 13–17 July 2014.
53. Akay, M.; Dragomir, A.; Akay, Y.M.; Chen, F.; Post, A.; Jneid, H.; Paniagua, D.; Denktas, A.; Bozkurt, B. The Assessment of Stent Effectiveness Using a Wearable Beamforming MEMS Microphone Array System. *IEEE J. Transl. Eng. Health Med.* **2016**, *4*, 1–10.
54. DiBiase, J.H.; Silverman, H.F.; Brandstein, M.S. Robust localization in reverberant rooms. In *Microphone Arrays*; Springer: Berlin, Germany, 2001; pp. 157–180.
55. Nguyen, D.; Aarabi, P.; Sheikholeslami, A. Real-time sound localization using field-programmable gate arrays. In Proceedings of the 2003 (ICASSP'03) IEEE International Conference on Acoustics, Speech, and Signal Processing, Hong Kong, China, 6–10 April 2003; Volume 2, p. 573.
56. Wang, J.F.; Jiang, Y.C.; Sun, Z.W. FPGA implementation of a novel far-field sound localization system. In Proceedings of the TENCON 2009 IEEE Region 10 Conference, Singapore, 23–26 November 2009; pp. 1–4.
57. Simon, G.; Maróti, M.; Lédeczi, Á.; Balogh, G.; Kusy, B.; Nádas, A.; Pap, G.; Sallai, J.; Frampton, K. Sensor network-based countersniper system. In Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems, Baltimore, MD, USA, 3–5 November 2004; pp. 1–12.
58. Lédeczi, Á.; Nádas, A.; Völgyesi, P.; Balogh, G.; Kusy, B.; Sallai, J.; Pap, G.; Dóra, S.; Molnár, K.; Maróti, M.; others. Countersniper system for urban warfare. *ACM Trans. Sens. Netw.* **2005**, *1*, 153–177.
59. Aleks, I.; Hocenski, Ž.; Horvat, P. Acoustic Localization based on FPGA. In Proceedings of the MIPRO, 2010 33rd International Convention, IEEE, Opatija, Croatia, 24–28 May 2010; pp. 656–658.

60. Faraji, M.M.; Shouraki, S.B.; Iranmehr, E. Spiking neural network for sound localization using microphone array. In Proceedings of the 2015 23rd Iranian Conference on. Electrical Engineering (ICEE), IEEE, Tehran, Iran, 10–14 May 2015; pp. 1260–1265.
61. Biswas, T.; Mandal, S.B.; Saha, D.; Chakrabarti, A. Coherence based dual microphone speech enhancement technique using FPGA. *Microprocess. Microsyst.* **2017**, *55*, 111–118.
62. Sledevič, T.; Laptik, R. An evaluation of hardware-software design for sound source localization based on SoC. In Proceedings of the 2017 Open Conference of IEEE, Electrical, Electronic and Information Sciences (eStream), Vilnius, Lithuania, 27 April 2017; pp. 1–4.
63. Đurišić, M.P.; Tafa, Z.; Dimić, G.; Milutinović, V. A survey of military applications of wireless sensor networks. In Proceedings of the 2012 Mediterranean Conference on, IEEE, Embedded Computing (MECO), Bar, Montenegro, 19–21 June 2012; pp. 196–199.
64. Sallai, J.; Hedgecock, W.; Volgyesi, P.; Nadas, A.; Balogh, G.; Ledecz, A. Weapon classification and shooter localization using distributed multichannel acoustic sensors. *J. Syst. Archit.* **2011**, *57*, 869–885.
65. Abdeen, A.; Ray, L. Design and performance of a real-time acoustic beamforming system. In Proceedings of the 2013 SENSORS, Baltimore, MD, USA, 3–6 November 2013; pp. 1–4.
66. Zwysig, E.; Lincoln, M.; Renals, S. A digital microphone array for distant speech recognition. In Proceedings of the 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), Dallas, TX, USA, 14–19 March 2010; pp. 5106–5109.
67. Hafizovic, I.; Nilsen, C.I.C.; Kjølervbakken, M. Acoustic tracking of aircraft using a circular microphone array sensor. In Proceedings of the 2010 IEEE International Symposium on Phased Array Systems and Technology (ARRAY), Waltham, MA, USA, 12–15 October 2010; pp. 1025–1032.
68. Hafizovic, I.; Nilsen, C.I.C.; Kjølervbakken, M.; Jahr, V. Design and implementation of a MEMS microphone array system for real-time speech acquisition. *Appl. Acoust.* **2012**, *73*, 132–143.
69. Salom, I.; Celebic, V.; Milanovic, M.; Todorovic, D.; Prezelj, J. An implementation of beamforming algorithm on FPGA platform with digital microphone array. In Proceedings of the Audio Engineering Society Convention 138, Audio Engineering Society, Warsaw, Poland, 7–10 May 2015.
70. Todorović, D.; Salom, I.; Čelebić, V.; Prezelj, J. Implementation and Application of FPGA Platform with Digital MEMS Microphone Array. In Proceedings of the Proceedings of 4th International Conference on Electrical, Electronics and Computing Engineering, Kladovo, Serbia, 5–8 June 2017; pp. 1–6.
71. Petrica, L.; Stefan, G. Energy-Efficient WSN Architecture for Illegal Deforestation Detection. *Int. J. Sens. Sens. Netw.* **2015**, *3*, 24–30.
72. Petrica, L. An evaluation of low-power microphone array sound source localization for deforestation detection. *Appl. Acoust.* **2016**, *113*, 162–169.
73. Kowalczyk, K.; Wozniak, S.; Chyrowicz, T.; Rumian, R. Embedded system for acquisition and enhancement of audio signals. In Proceedings of the 2016 IEEE, Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), Poznan, Poland, 21–23 September 2016; pp. 68–71.
74. Azcarreta Ortiz, J. Pyramic Array: An FPGA Based Platform for Many-Channel Audio Acquisition. Master's Thesis, Universitat Politècnica de Catalunya, Barcelona, Spain, 2016.
75. Bezzam, E.; Scheibler, R.; Azcarreta, J.; Pan, H.; Simeoni, M.; Beuchat, R.; Hurley, P.; Bruneau, B.; Ferry, C.; Kashani, S. Hardware and software for reproducible research in audio array signal processing. In Proceedings of the 2017 IEEE International Conference on IEEE, Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017; pp. 6591–6592.
76. da Silva, B.; Segers, L.; Braeken, A.; Touhafi, A. Runtime reconfigurable beamforming architecture for real-time sound-source localization. In Proceedings of the 26th International Conference on Field Programmable Logic and Applications, FPL 2016, Lausanne, Switzerland, 29 August–2 September 2016; pp. 1–4.
77. Samtani, K.; Thomas, J.; Varma, G.A.; Deepu, S.; others. FPGA implementation of adaptive beamforming in hearing aids. In Proceedings of the 2017 39th Annual International Conference of the IEEE, Engineering in Medicine and Biology Society (EMBC), Seogwipo, Korea, 11–15 July 2017; pp. 2239–2242.
78. Inoue, T.; Ikeda, Y.; Oikawa, Y. Hat-type hearing aid system with flexible sound directivity pattern. *Acoust. Sci. Technol.* **2018**, *39*, 22–29.

79. da Silva, B.; Segers, L.; Braeken, A.; Steenhaut, K.; Touhafi, A. A Low-Power FPGA-Based Architecture for Microphone Arrays in Wireless Sensor Networks. In Proceedings of the 14th International Symposium, Applied Reconfigurable Computing. Architectures, Tools, and Applications 2018, Santorini, Greece, 2–4 May 2018; pp. 281–293.
80. Luo, F.L.; Yang, J.; Pavlovic, C.; Nehorai, A. Adaptive null-forming scheme in digital hearing aids. *IEEE Trans. Signal Process.* **2002**, *50*, 1583–1590.
81. Takagi, T.; Noguchi, H.; Kugata, K.; Yoshimoto, M.; Kawaguchi, H. Microphone array network for ubiquitous sound acquisition. In Proceedings of the 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), Dallas, TX, USA, 14–19 March 2010; pp. 1474–1477.
82. Kugata, K.; Takagi, T.; Noguchi, H.; Yoshimoto, M.; Kawaguchi, H. Intelligent ubiquitous sensor network for sound acquisition. In Proceedings of the 2010 IEEE International Symposium on IEEE, 2010 Circuits and Systems (ISCAS), Paris, France, 30 May–2 June 2010; pp. 1414–1417.
83. Izumi, S.; Noguchi, H.; Takagi, T.; Kugata, K.; Soda, S.; Yoshimoto, M.; Kawaguchi, H. Data aggregation protocol for multiple sound sources acquisition with microphone array network. In Proceedings of the 2011 Proceedings of 20th International Conference on IEEE Computer Communications and Networks (ICCCN), Lahaina, HI, USA, 31 July–4 August 2011; pp. 1–6.
84. Zimmermann, B.; Studer, C. FPGA-based real-time acoustic camera prototype. In Proceedings of the 2010 IEEE International Symposium on IEEE, Circuits and Systems (ISCAS), Paris, France, 30 May–2 June 2010; pp. 1419–1419.
85. Sanchez-Hevia, H.; Gil-Pita, R.; Rosa-Zurera, M.; others. FPGA-based real-time acoustic camera using pdm mems microphones with a custom demodulation filter. In Proceedings of the 2014 IEEE 8th IEEE, Sensor Array and Multichannel Signal Processing Workshop (SAM), A Coruna, Spain, 22–25 June 2014; pp. 181–184.
86. Sánchez-Hevia, H.A.; Mohino-Herranz, I.; Gil-Pita, R.; Rosa-Zurera, M. *Memory Requirements Reduction Technique for Delay Storage in Real Time Acoustic Cameras*; Audio Engineering Society Convention 136; Audio Engineering Society: New York, NY, USA, 2014.
87. Kim, Y.; Kang, J.; Lee, M. Developing beam-forming devices to detect squeak and rattle sources by using FPGA. In Proceedings of the INTER-NOISE and NOISE-CON Congress and Conference Proceedings. Institute of Noise Control Engineering, Melbourne, Australia, 16–19 November 2014; Volume 249, pp. 4582–4587.
88. Perrodin, F.; Nikolic, J.; Busset, J.; Siegwart, R. Design and calibration of large microphone arrays for robotic applications. In Proceedings of the 2012 IEEE/RSJ International Conference on IEEE, Intelligent Robots and Systems (IROS), Vilamoura, Portugal, 7–12 October 2012; pp. 4596–4601.
89. Suzuki, T. L1 generalized inverse beam-forming algorithm resolving coherent/incoherent, distributed and multipole sources. *J. Sound Vib.* **2011**, *330*, 5835–5851.
90. Netti, A.; Diodati, G.; Camastra, F.; Quaranta, V. FPGA implementation of a real-time filter and sum beamformer for acoustic antenna. In Proceedings of the INTER-NOISE and NOISE-CON Congress and Conference Proceedings. Institute of Noise Control Engineering, San Francisco, CA, USA, 9–12 August 2015; Volume 250, pp. 3458–3469.
91. Bourgeois, J.; Minker, W., Eds., Linearly Constrained Minimum Variance Beamforming. In *Time-Domain Beamforming and Blind Source Separation: Speech Input in the Car Environment*; Bourgeois, J., Minker, W., Eds.; Springer US: Boston, MA, USA, 2009; pp. 27–38.
92. Diodati, G.; Quaranta, V. Acoustic Sensors Array for Pass-by-Noise Measurements: Antenna Design. In Proceedings of the International Congress on Sound and Vibration ICSV22, Athens, Greece, 10–14 July 2015.
93. Izquierdo, A.; Villacorta, J.J.; del Val Puente, L.; Suárez, L. Design and evaluation of a scalable and reconfigurable multi-platform system for acoustic imaging. *Sensors* **2016**, *16*, 1671.
94. LabVIEW FPGA Module. Available online: <http://www.ni.com/labview/fpga/> (accessed on 1 June 2018).
95. del Val, L.; Izquierdo, A.; Villacorta, J.J.; Suárez, L. Using a Planar Array of MEMS Microphones to Obtain Acoustic Images of a Fan Matrix. *J. Sens.* **2017**, *2017*.
96. Izquierdo, A.; Villacorta, J.J.; del Val, L.; Suárez, L.; Suárez, D. Implementation of a Virtual Microphone Array to Obtain High Resolution Acoustic Images. *Sensors* **2017**, *18*, 25.

97. Seo, S.W.; Kim, M. Estimation of 3D ball motion using an infrared and acoustic vector sensor. In Proceedings of the 2017 International Conference on IEEE, Information and Communication Technology Convergence (ICTC), Jeju, Korea, 18–20 October 2017; pp. 1047–1049.
98. Seo, S.W.; Kim, M. 3D Impulsive Sound-Source Localization Method through a 2D MEMS Microphone Array using Delay-and-Sum Beamforming. In Proceedings of the 9th International Conference on Signal Processing Systems, Auckland, New Zealand, 27–30 November 2017; pp. 170–174.
99. da Silva, B.; Segers, L.; Rasschaert, Y.; Quevy, Q.; Braeken, A.; Touhafi, A. A Multimode SoC FPGA-Based Acoustic Camera for Wireless Sensor Networks. In Proceedings of 13th International Symposium on Reconfigurable Communication-centric Systems-on-Chip, ReCoSoC 2018, Lille, France, 9–11 July 2018; pp. 1–8.
100. Hogenauer, E. An economical class of digital filters for decimation and interpolation. *IEEE Trans. Acoust. Speech Signal Process.* **1981**, *29*, 155–162.
101. Garrido, M.; Acevedo, M.; Ehliar, A.; Gustafsson, O. Challenging the limits of FFT performance on FPGAs. In Proceedings of the 2014 14th International Symposium on Integrated Circuits (ISIC), Saint-Malo, France, 13–16 September 2014; pp. 172–175.
102. Bogue, R. Recent developments in MEMS sensors: A review of applications, markets and technologies. *Sens. Rev.* **2013**, *33*, 300–304.
103. Underwood, K. FPGAs vs. CPUs: trends in peak floating-point performance. In Proceedings of the 2004 ACM/SIGDA 12th International Symposium on Field Programmable Gate Arrays, Monterey, CA, USA, 22–24 February 2004; pp. 171–180.
104. Xilinx System Generator for DSP. Available online: <https://www.xilinx.com/products/design-tools/vivado/integration/sysgen.html> (accessed on 1 June 2018).
105. Dusan, S.V.; Lindahl, A.; Andersen, E.B. System and Method of Mixing Accelerometer and Microphone Signals to Improve Voice Quality in a Mobile Device. U.S. Patent 9,363,596, issued 7 June, 2016.
106. Mathworks HDL Coder Toolbox. Available online: <https://www.mathworks.com/products/hdl-coder.html> (accessed on 1 June 2018).
107. Nordholm, S.; Claesson, I.; Dahl, M. Adaptive microphone array employing calibration signals: an analytical evaluation. *IEEE Trans. Speech Audio Process.* **1999**, *7*, 241–252.
108. Zynq Ultrascale+ SoC. Available online: <https://www.xilinx.com/products/silicon-devices/soc/zynq-ultrascale-mpsoc.html> (accessed on 1 June 2018).
109. Greene, J.; Kaptanoglu, S.; Feng, W.; Hecht, V.; Landry, J.; Li, F.; Krouglyanskiy, A.; Morosan, M.; Pevzner, V. A 65nm flash-based FPGA fabric optimized for low cost and power. In Proceedings of the 19th ACM/SIGDA iNternational Symposium on Field Programmable Gate Arrays, Monterey, California, USA, 27 February–1 March 2011; pp. 87–96.
110. Cowling, M.; Sitte, R. Comparison of techniques for environmental sound recognition. *Pattern Recognit. Lett.* **2003**, *24*, 2895–2907.
111. Heittola, T.; Mesaros, A.; Virtanen, T.; Eronen, A. Sound Event Detection in Multisource Environments Using Source Separation. Machine Listening in Multisource Environments. 2011. Available online: https://www.isca-speech.org/archive/chime_2011/cm11_036.html (accessed on 31 July 2018).
112. Kao, C. Benefits of partial reconfiguration. *Xcell J.* **2005**, *55*, 65–67.
113. Lie, W.; Feng-Yan, W. Dynamic partial reconfiguration in FPGAs. In Proceedings of the Intelligent Information Technology Application, 2009, IITA Third International Symposium on IEEE, Nanchang, China, 21–22 November 2009; Volume 2, pp. 445–448.
114. Llamocca, D.; Aloï, D.N. Self-reconfigurable implementation for a switched beam smart antenna. *Microprocess. Microsyst.* **2018**, *60*, 1–14.
115. Degnan, B.; Marr, B.; Hasler, J. Assessing trends in performance per Watt for signal processing applications. *IEEE Trans. Very Large Scale Integr. Syst.* **2016**, *24*, 58–66.
116. Betkaoui, B.; Thomas, D.B.; Luk, W. Comparing performance and energy efficiency of FPGAs and GPUs for high productivity computing. In Proceedings of the 2010 International Conference on Field-Programmable Technology, Beijing, China, 8–10 December 2010; pp. 94–101.
117. GPU vs. FPGA Performance Comparison, White Paper. Available online: http://www.bertendsp.com/pdf/whitepaper/BWP001_GPU_vs_FPGA_Performance_Comparison_v1.0.pdf (accessed on 1 June 2018).

118. Dai, W. Acoustic Scene Recognition with Deep Learning. 2016. Available online: https://www.ml.cmu.edu/research/dap-papers/DAP_Dai_Wei.pdf (accessed on 31 July 2018).
119. Parascandolo, G.; Heittola, T.; Huttunen, H.; Virtanen, T.; others. Convolutional recurrent neural networks for polyphonic sound event detection. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2017**, *25*, 1291–1303.
120. Brdiczka, O.; Crowley, J.L.; Reignier, P. Learning situation models in a smart home. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **2009**, *39*, 56–63.
121. Budkov, V.Y.; Prischepa, M.; Ronzhin, A.; Karpov, A. Multimodal human-robot interaction. In Proceedings of the 2010 International Congress on IEEE Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), Moscow, Russia, 18–20 October 2010; pp. 485–488.
122. Zunino, A.; Crocco, M.; Martelli, S.; Trucco, A.; Del Bue, A.; Murino, V. Seeing the Sound: A New Multimodal Imaging Device for Computer Vision. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Santiago, Chile, 7–13 December 2015; pp. 6–14.
123. Mentens, N.; Vandompe, J.; Vliegen, J.; Braeken, A.; da Silva, B.; Touhafi, A.; Kern, A.; Knappmann, S.; Rettowski, J.; Al Kadi, M.S.; Göhringer, D. DynamIA: Dynamic hardware reconfiguration in industrial applications. In Proceedings of the International Symposium on Applied Reconfigurable Computing (ARC) 2015, Bochum, Germany, 13–17 April, 2015; pp. 513–518.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).