



**HAL**  
open science

## **micarray - A System for Multichannel Audio Streaming over Ethernet**

Christoph Reitbauer, Harald Rainer, Markus Noisternig, Bernhard  
Rettenbacher, Franz Graf

► **To cite this version:**

Christoph Reitbauer, Harald Rainer, Markus Noisternig, Bernhard Rettenbacher, Franz Graf. micarray - A System for Multichannel Audio Streaming over Ethernet. 5th Congress of the Alps Adria Acoustics Association, Sep 2012, Petrcane, Zadar, Croatia. pp.1-1. hal-01161015

**HAL Id: hal-01161015**

**<https://hal.science/hal-01161015>**

Submitted on 16 Jul 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## 5th Congress of Alps-Adria Acoustics Association

12-14 September 2012, Petrčane, Croatia

Hrvatsko akustičko društvo  
**HAD**  
Acoustical Society of Croatia

### MICARRAY – A SYSTEM FOR MULTICHANNEL AUDIO STREAMING OVER ETHERNET

Christoph Reitbauer<sup>a</sup>, Harald Rainer<sup>a</sup>, Markus Noisternig<sup>b</sup>, Bernhard Rettenbacher<sup>a</sup>, Franz Graf<sup>a</sup>

<sup>a</sup> Space and Acoustics, DIGITAL Institute for Information and Communication Technologies,  
JOANNEUM RESEARCH, Leonhardstrasse 59, 8010 Graz, Austria (harald.rainer@joanneum.at)

<sup>b</sup> Acoustic and Cognitive Spaces Research Group, UMR STMS IRCAM-CNRS-UPMC, 1 place Igor-  
Stravinsky, 75004 Paris, France (markus.noisternig@ircam.fr)

**Abstract:** Nowadays several commercial off-the-shelf products for the use of multichannel audio data transmission over Ethernet are available. For certain applications these products have either technical limitations and/or they are too expensive. Therefore a highly flexible and cost efficient FPGA based data acquisition system has been developed. With this system it is possible to transmit 64 channels of sensor data at a resolution of 24 bits and an adjustable sampling frequency up to 192 kHz. To fulfill all of the technical requirements, the analog frontend was designed to support the IEPE standard. In this manner it is possible to drive a wide range of different sensor types, e.g. accelerometers, pressure sensors, force sensors, and microphones. Furthermore a small-sized microphone pre-amplifier for the operation of an omnidirectional back electret condenser microphone was designed and crafted. With the aid of these miniature microphone modules different array geometries can be realized.

Key words: streaming audio, fpga, udp, ethernet, iepe, piezoelectric, transducer arrays  
PACS number: 43.38.Kb, 43.38.Lc, 43.38.Hz, 43.38.Fx, 43.58.Fm

#### 1. INTRODUCTION

In recent years research and industries have shown growing interest in directional microphone arrays for speech communication systems (e.g. acoustic source localization and separation, speech enhancement and de-reverberation) as well as for the measurement and analysis of three-dimensional sound fields (e.g. planar microphone arrays for near-field acoustic holography and spherical microphone arrays for room acoustic sensing). Typical array configurations consist of a few dozen to perhaps hundreds of microphones. Unfortunately, the per-channel cost and complexity of these systems (i.e. microphones, pre-amplifiers and signal conditioners, data acquisition, multichannel audio streaming) limits their use in many applications.

In addition, the high channel count and the often-required long cable runs limit the use of traditional analog audio transmission technologies. Therefore, practical and cost-effective digital audio distribution and management solutions over Ethernet are required, which allow for the transmission of synchronous uncompressed multichannel

audio data with lowest latency. To provide the highest possible flexibility, a typical solution should distribute audio over standard CAT5 UTP LAN cables and 1000Base-T switching hubs while remaining fully IEEE 802.3ab compliant. There are multiple Audio over Ethernet technologies available on the market today, e.g. CobraNet (Cirrus Logic), Dante (Audinate), Ethersound (Digigram), SuperMAC / HyperMAC (Klark Teknik), which are often used for audio and broadcast engineering applications.

In many practical array applications the small spacing between microphones requires a sufficiently high sampling rate to prevent time-scaling artifacts; research projects related to this article typically require audio sampling rates up to 192 kHz. To the best of the authors' knowledge only Dante and SuperMAC / HyperMAC provide sampling rates higher than 96 kHz but with a relatively small channel count. Brüel & Kjaer's LAN-XI system is a distributed network data acquisition hardware solution for measurement applications with a maximum sampling rate of 262 kHz and a high channel count. However, the high hardware and software purchase costs

of the above mentioned systems required the development of a solution that fits the requirements of the envisaged research projects but benefits from significant cost savings. In the following, this project will be referred to as “micarray”.

The “micarray” project aims at developing a modular and cost efficient data acquisition system that supports the IEPE (Integrated Electronics Piezo Electric) [1] standard for piezoelectric transducers. Many manufacturers offer a wide range of IEPE compatible sensors, such as microphones, accelerometers, stress sensors, and transducers for specific applications. The choice of the IEPE standard thus one allows to implement multimodal sensor systems.

Pollow et al. have recently presented a comparable system [2] for the streaming of audio data over Ethernet. The main difference to micarray is that this system doesn't support the IEPE standard. The NIST Mark-III microphone array [3] is an open-hardware platform for speech signal processing, which only provides relatively low sampling rates; thus it was not suitable for our applications.

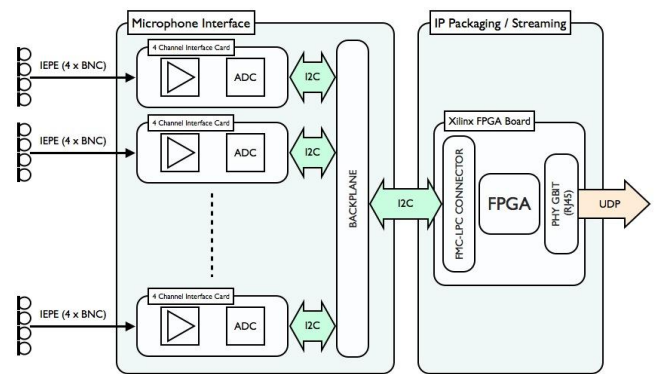
## 2. HARDWARE DEVELOPMENTS

Different microphone arrays for the recording and analysis of sound fields have been recently released to the market. However, none of the available arrays could fulfill all of the technical requirements of our projects. For this reason, IRCAM and JOANNEUM RESEARCH co-developed a digital hardware platform for capturing the data of a large number of microphones with excellent audio quality. Audio is captured by high quality analog-to-digital converters (ADCs), processed by a FPGA unit and transmitted over Gigabit Ethernet. The FPGA unit and transmission protocol was developed by our industrial research partner xFace.

Specifications of the developed platform:

- 64 input channels
- 24-Bit PCM audio signal resolution
- Adjustable sampling frequency up to 192 kHz
- Latency from ADC to Ethernet < 1ms
- Xilinx Spartan-6 based FPGA
- I<sup>2</sup>S interface and power supply via FFC cables to FPGA board
- Network interface, 1000Base-T, with RJ45 connector
- UDP based data streaming protocol (xFaceStream®)
- UDP and ARP protocols for configuration and status requests

Figure 1 depicts the main functional units of the spherical microphone array. In order to reduce the complexity of the design and to provide a maximum flexibility at the same time the main functional units are separated on different printed circuit boards (PCBs) and are fully encapsulated and self contained.

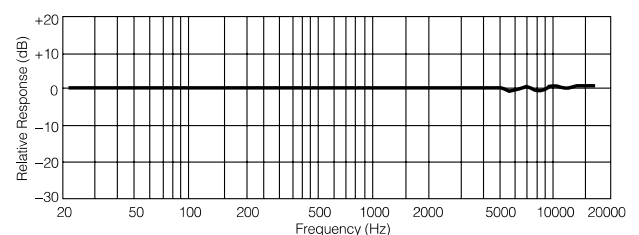


**Fig.1.** Microphone signal conditioning and general processing structure for streaming the microphone array audio signals over IP

One of the main objectives of the circuit design is to reduce the total harmonic distortions (THD) and to improve the signal-to-noise ratio (SNR). To reduce electromagnetic interferences PCBs were made in four layers and only surface-mounted devices (SMD) were used. In four-layer boards the ground plane is routed underneath the power plane and both are close to the signal planes, thus reducing radio-frequency noise. Microphones are in general high impedance and thus cause strong electromagnetic inductions in combination with long conductor runs. Surface-mount technology allows for closer component placement due to the reduced physical dimensions of SMDs; the small traces help to eliminate electromagnetic interferences.

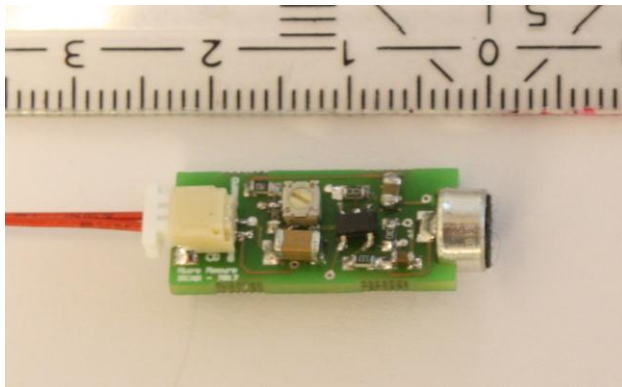
### 2.1 Microphone pre-amplification and signal conditioning

The microphone pre-amplification stage is designed for, but not limited to, the use with Panasonic WM-61 microphones. These small general purpose back electret condenser microphones provide a high SNR, high sensitivity, and a high resistance to vibrations. The typical frequency response curve of WM-61 microphones is depicted in Figure 2.



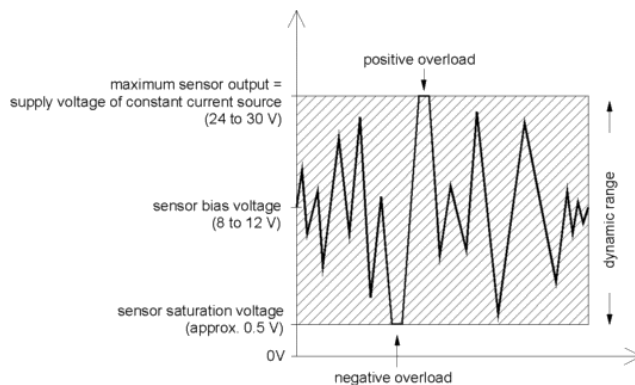
**Fig.2.** Typical frequency response curve of the Panasonic WM-61 electret microphone (cf. data sheet provided by the manufacturer).

The microphone pre-amplifier board implements the well-established IEPE standard. It supports the microphone capsule with a standard operation voltage of 2 V and transforms the high-impedance output of the microphone into a low-impedance voltage signal that can be transmitted over long distances. The circuit is powered by a constant current, which is provided by the subsequent ADC board. With reference to the IEPE standard the constant current should be in the range of 2 mA to 20 mA. The lower the constant current the higher the output impedance and sensitivity to electromagnetic interferences; within this project the constant current was set to be adjustable in a range of 4 mA to 6 mA. The microphone pre-amplifier prototype with implemented IEPE standard is depicted in Figure 3.



**Fig.3.** Microphone pre-amplifier prototype with Panasonic WM-61 back electret condenser microphone.

The microphone pre-amplifier was designed to support a dynamic range of 30 dB (SPL) to 130 dB (SPL), which corresponds to a sound pressure dynamic range of 0.63 mPa to 63.25 Pa. With a sensitivity of 100 mV/Pa the peak output voltage calculates to a range from approx. 90  $\mu$ V to 9 V and is thus compatible with the IEPE standard for a supply voltage of 24 V. Figure 4 illustrates the dynamic range of IEPE compatible sensors.



**Fig.4.** Dynamic range of IEPE compatible sensors [1].

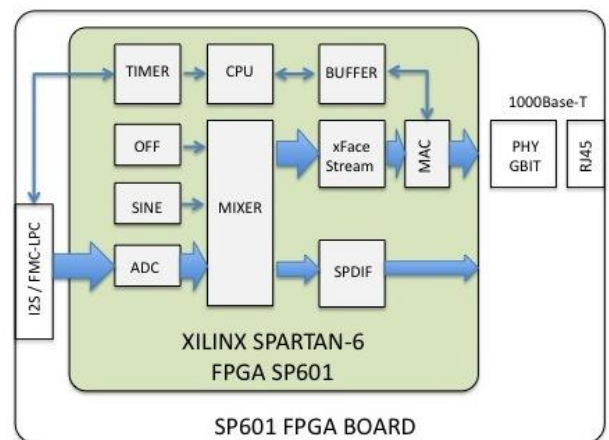
## 2.2 Data acquisition and digitalization board

The data acquisition and digitalization board provides four constant current powered IEPE inputs, an input signal amplifier for each channel, and a low noise analog-to-digital conversion (ADC) unit. All data acquisition boards are connected via a backplane to the Xilinx Spartan-6 FPGA board, which also provides the power supply and clock signals. I<sup>2</sup>S (Integrated Interchip Sound) interface is used for connecting the digital audio devices together. I<sup>2</sup>S is an electrical serial bus interface standard that separates clock and data signals, thus resulting in very low jitter connections. In the I<sup>2</sup>S standard any device can act as the system clock master by providing the necessary clock signals; a slave will usually derive its internal clock signal from an external clock input.

The analog-to-digital converter board uses a Texas Instruments PCM4202 high performance Delta-Sigma audio ADC that supports 24-bit linear PCM output data with sampling frequencies up to 216 kHz. This integrated chip supports the I<sup>2</sup>S interface and provides an overall THD+N > -105 dB. The PCM4202 requires an external system clock from which the modulator oversampling and digital sub-system clocks are derived. As mentioned above, the system clock is provided by the FPGA via the I<sup>2</sup>S interface.

## 2.3 FPGA board and multichannel audio streaming over IP

The Spartan-6 FPGA board collects all digital audio signals from the data acquisition and digitalization boards and converts them to the proprietary xFaceStream<sup>®</sup> streaming protocol. Audio data is buffered and converted into UDP packets that are transmitted to the client computer via Gigabit Ethernet. xFaceStream<sup>®</sup> allows for a high throughput and low latency; each packet contains a 64-bit high precision timestamp for data synchronization. Figure 5 illustrates the FPGA board processing structure.



**Fig.5.** FPGA board architecture.

UDP sends data, also referred to as datagrams, to other hosts on an IP network using a simple transmission protocol without implicit handshaking, i.e. without requiring any prior communication to set up the transmission channels and data paths. It is well suited to real-time streaming applications but provides an unreliable service as datagrams may for instance arrive out of order, appear duplicated or get lost. The timestamps provided by the xFaceStream® protocol allow a client-side identification of out-of-order and duplicate datagrams without loosing the real-time capability.

Table 1 summarizes the required bandwidths for different sample rates and numbers of audio channels. Using a standard 1000Base-T GBit Ethernet connection it is possible to send 64 audio channels with 24-bit PCM at 192 kHz sampling rate.

**Table 1.** Required bandwidth for different numbers of audio channels and sampling rates.

Sampling Frequency [kHz]	Number of audio channels	Required Bandwidth [MBit/s]
48	32	51.6
96   48	32   64	103.2
192   96	32   64	206.4
192	64	412.8

### 3. CONCLUSION

In this paper a system for multichannel audio streaming over Ethernet was presented. It fits the requirements of the different research projects but benefits from significant cost savings compared to commercial network audio systems.

Implementing the IEPE standard to the analog frontend offers the possibility to drive a wide range of different sensor types such as microphones, accelerometers, stress sensors, and transducers for specific applications, thus increasing the flexibility of the presented data acquisition system.

In the future, the system will be applied to different applications with a variety of different hardware configurations, such as e.g. spherical microphone arrays for near-field holography and room acoustic analysis, multimodal sensor systems for inspection and testing, and distributed sensor networks for traffic noise analysis.

### 4. ACKNOWLEDGEMENTS

The authors would like to express their gratitude to Johannes Wolkerstorfer (xFace) for the excellent work on the development of the xFaceStream® Ethernet streaming protocol. Appreciation is extended to Emmanuel Flety and Gérard Bertrand for numerous informative discussions and their assistance on all aspects of implementation and electronic circuit design. Last, but not least, the authors wish to thank Thibaut Carpentier for his help with software and driver development. This project was supported in part (second author) by the French National Research Agency project ANR-CONTINT “Sample Orchestrator 2 (SOR2)”.

### REFERENCES

- [1] M. Weber: **IEPE Standard**, retrieved from [http://www.mmf.de/iepe\\_standard.htm](http://www.mmf.de/iepe_standard.htm) (accessed July 12th, 2012)
- [2] M. Pollow, P. Dietrich, B. Krechel, M. Vorländer: **Unidirektionale mehrkanalige Audioübertragung über Ethernet**, *Fortschritte der Akustik: DAGA 2011*, Düsseldorf, Germany, 2011
- [3] V. Stanford, and C. Rochet: **The NIST Mk-III Microphone Array, and Applications of Adaptive Beam Forming to Speech**, *5<sup>th</sup> International Workshop on Microphone Array Systems – Theory and Practice*, Pommersfelden, Germany, 2003
- [4] Ch. Reitbauer: **Entwicklung eines echtzeitfähigen Systems zur akustischen Detektion der Fahrtrichtung**, *Master Thesis at the Graz University of Technology*, Graz, Austria, 2011