Ether
ES
Sound

Overview - An introduction to the technology

Rev. 2.0c

October 20, 2004

Table of Contents

1 Overview

EtherSound enhances established technologies to provide easy-to-implement, high-quality audio networks. The patented EtherSound protocol provides deterministic, very low-latency, synchronous transmission of audio over a standard Ethernet infrastructure.

EtherSound maintains a fully digital path between networked audio devices. Using a 100 Mbits/s network, up to 64 channels of 48 kHz 24-bit bi-directional PCM digital audio, control data, and bi-directional status data may be transported among a virtually infinite number of connected devices.

Off-the-shelf Ethernet components (i.e. switches) can be used to extend the number of audio devices, as well as the distance between the devices on the network (i.e. using optical fiber). A list of tested Ethernet components is available on the Digigram web site.

Configuration and set-up are simpler than analog systems. EtherSound drastically reduces installation costs of systems for public address, installed sound, residential installations, etc…

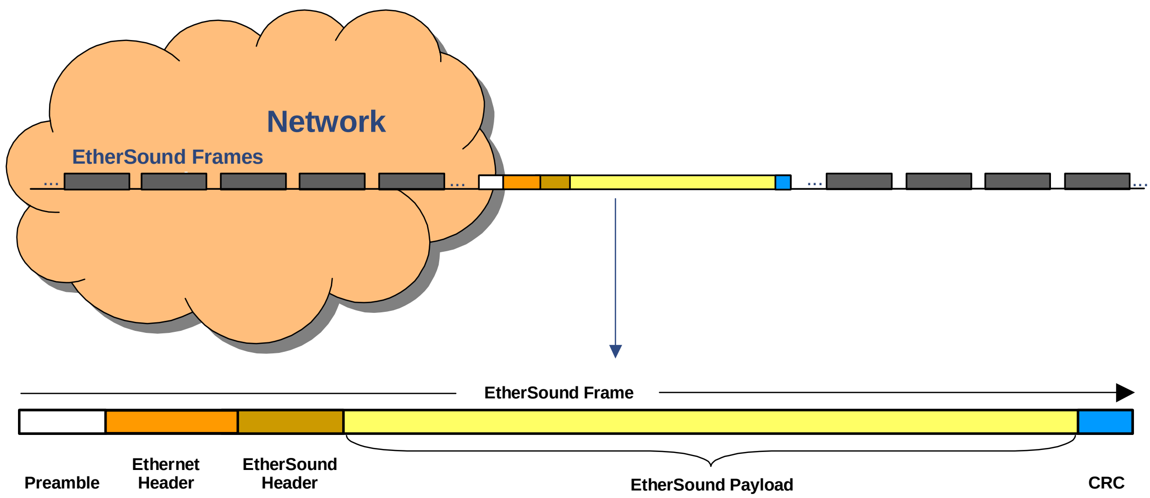
2 The EtherSound Protocol

The EtherSound protocol developed by Digigram is based on true Ethernet frames and provides a means of communication between EtherSound- capable devices.

2.1 EtherSound Frames

EtherSound frames are wrapped up in standard Ethernet frames and consist of two main parts:

- the **EtherSound Header**, that includes all important information relevant to the protocol,
- the **EtherSound Payload**, the actual data transmitted via the network.



The EtherSound Payload, just like for standard Ethernet frames, is divided in a set of packets. Each packet is made of:

- the Packet Header, where the packet type and subtype are defined,
- the Packet Data, the actual data.

2.2 Audio clocks

The EtherSound network is fully synchronous and carries its own clock. The Primary Master generates the network audio clock, each downstream ES device's audio clock is synchronized with the Primary Master audio clock:

- Derived from the EtherSound frames
- Synchronized on a distributed Word Clock whenever phase consistency is required

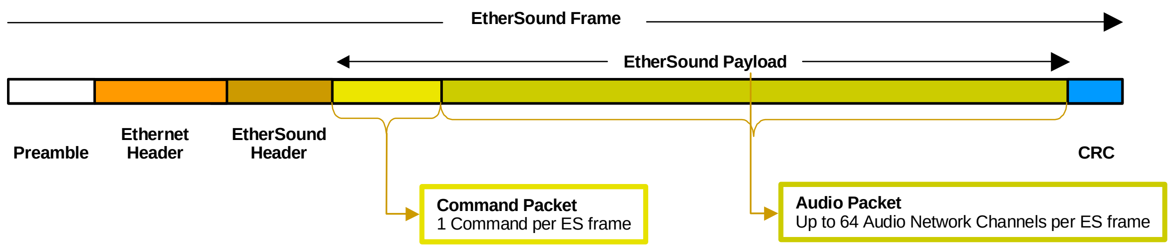
Clock distribution and timings aspects are discussed in chapter 6.

2.3 EtherSound Protocol V1

The current implementation of the EtherSound protocol (V1) is based on 100 Mbps networks and features a payload of two data packets:

- a Command Packet that transmits one command. This command can be a control command (write) or a status request command (read).
- an Audio Packet for transmitting up to 64 24-bit audio streams at 44.1 kHz or 48 kHz.

Therefore, the EtherSound frame frequency equals the sampling frequency.



The EtherSound V1 frame has been designed to fully comply with 802.3 in order to allow use of standard Ethernet switches.

2.3.1 Dealing with high sampling rates

The flexibility of audio data stored in EtherSound frames allows for using several network channels to transport samples of a given audio stream at a sampling rate of an exact multiple of the network frequency.

In other words, a 48 kHz EtherSound network is capable of carrying audio streams at 96 kHz. The network frequency remains unchanged, but the audio data of the 96 kHz audio streams is stored using two EtherSound network channels instead of one. Any combination is possible, among which:

- 64 audio streams at 48 kHz
- 62 audio streams at 48 kHz and one audio stream at 96 kHz
- 48 audio streams at 48 kHz and 8 audio streams at 96 kHz
- 32 audio streams at 48 kHz and 16 audio streams at 96 kHz
- 32 audio streams at 96 kHz

and so on…

2.4 Control data

Control data is embedded in each EtherSound frame to control a specific device by writing or reading the device's internal registers. A complete description of control capabilities is provided in chapter 7.

The maximum bit rates are 768 kbits/s for write commands and 384 kbits/s for read commands.

3 Types of EtherSound devices

There are four types of EtherSound hardware devices differing in the way they act in an EtherSound network:

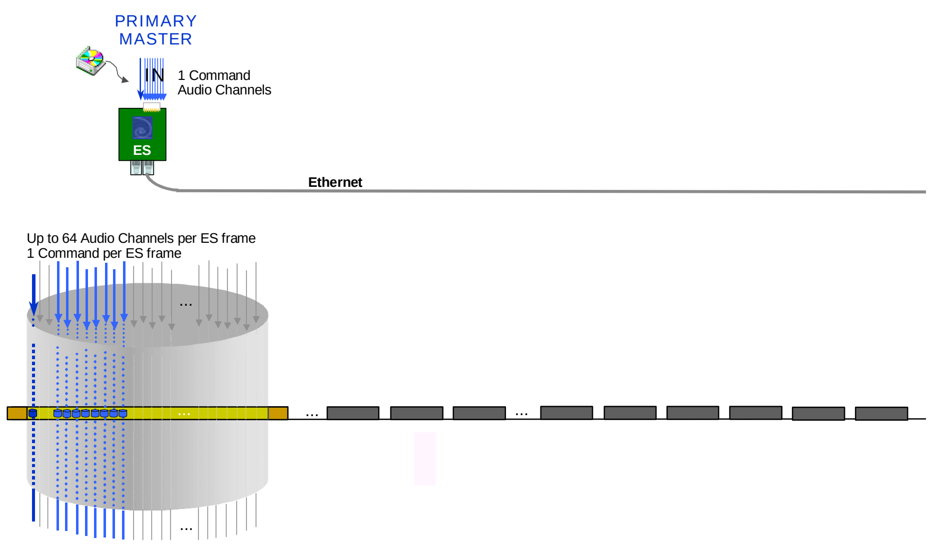
- Primary Master
- Master
- Slave
- Master/Slave

3.1 Primary Master

The first EtherSound device in the network is called the **Primary Master**. As well as being a source of audio for the network, the Primary Master provides the commands and audio clock.

In other words, it is the Primary Master that initializes and builds the EtherSound frames and the first to fuel these frames with audio data. By connecting its "Ethernet IN" port to a host computer (PC) you can remotely control and manage the entire EtherSound network via software.

Should a failure of the Primary Master occur, the next master device in the network automatically becomes a Primary Master. If the problem disappears, the system automatically returns to its initial state.



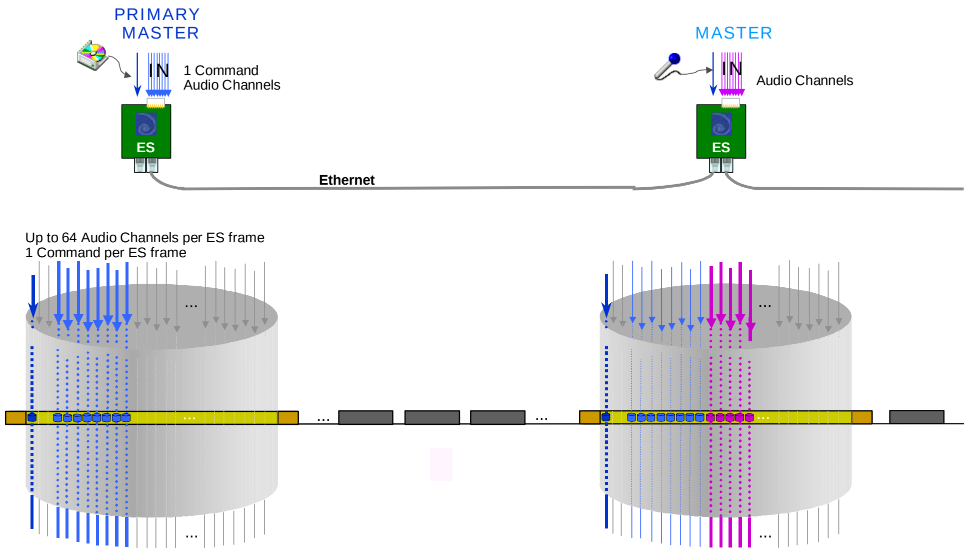
Two data flows are defined in the EtherSound naming convention:

The broadcast data flow generated by a Primary Master is called **downstream** and is available to any device of a given EtherSound network.

The unicast data flow sent back to a Primary Master is called **upstream**.

3.2 Master

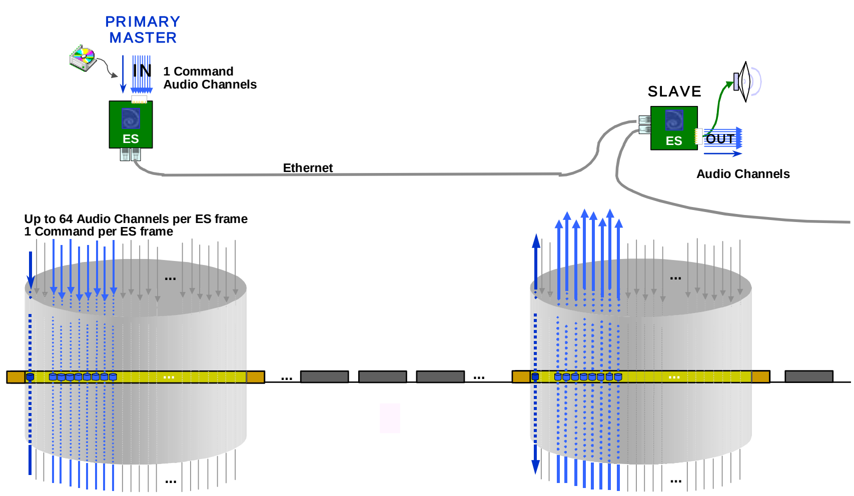
A device in the network located downstream from the Primary Master and contributing additional audio channels into the EtherSound stream is called **Master**. A Master device answers to the status requests and commands of the Primary Master. A Master device is always located downstream of a Primary Master device.



When used within a bi-directional loop, a setting defines whether the audio data is to be written into the downstream or the upstream.

3.3 Slave

An EtherSound device that receives the EtherSound stream and restores standard audio is called a **Slave** device. A Slave device answers to the status requests and commands of the Primary Master.



When used within a bi-directional loop, a setting defines whether the audio data is to be read from the downstream or the upstream.

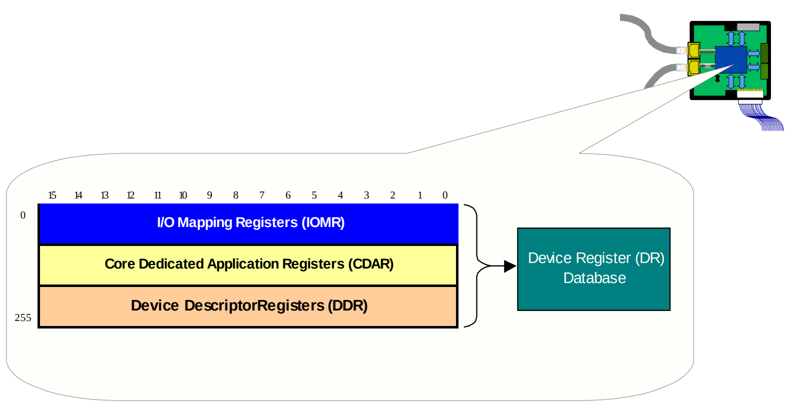
3.4 Master/Slave

A Master/Slave device combines both functionalities. It can read audio data from the EtherSound network and write audio data onto the EtherSound network. A Master/Slave device answers to the status requests and commands of the Primary Master.

4 EtherSound registers

Each EtherSound FPGA firmware has an internal database of 256 16-bit-Device Registers (DR). These Device Registers, also called the DR database, provide configuration information status or manage the behavior of the EtherSound device. These Device Registers are grouped in three separate categories:

- DDR: Device Descriptor Registers (48 registers) describe the status and the configuration of the EtherSound Kernel of the device. All EtherSound devices include these registers.
- IOMR: I/O Mapping Registers (a maximum of 128 registers) describe the EtherSound network channel assignments. The number of IOMR depends on the audio capacity of the EtherSound device.
- CDAR: Core Dedicated Application Parameters (a maximum of 80 registers) The Core Dedicated Application Registers (CDAR) are reserved for specific optional functions that are not included by default in every EtherSound device, such as RS232, GPIOs, analog gains, parameters for µcontroller or DSP software.



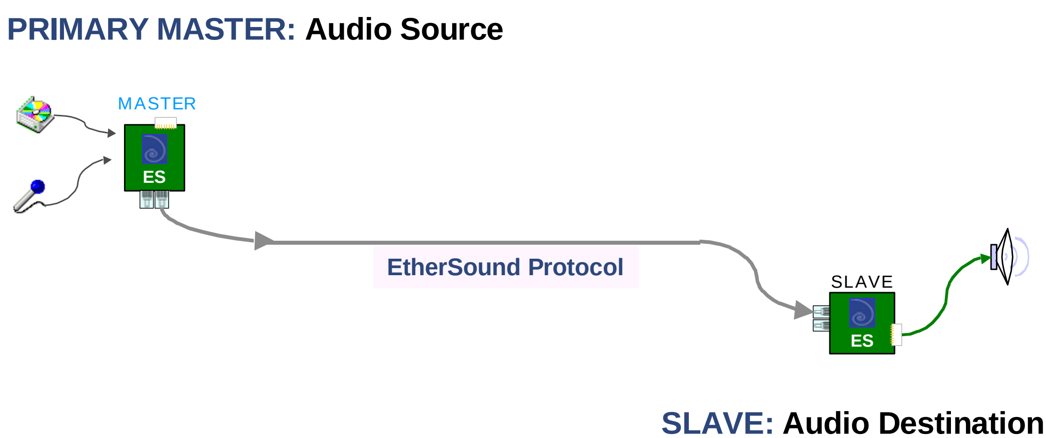
5 The EtherSound network and audio transport

5.1 Ethernet / IEEE 802.3 compatibility

The EtherSound technology is compatible with IEEE 802.3x standards and operates on full duplex switched, Fast Ethernet networks. Data is typically transported via dedicated Local Area Networks (LANs) with a minimum bandwidth of 100 Mbps/s (100BaseTx) full duplex.

5.2 The simplest configuration

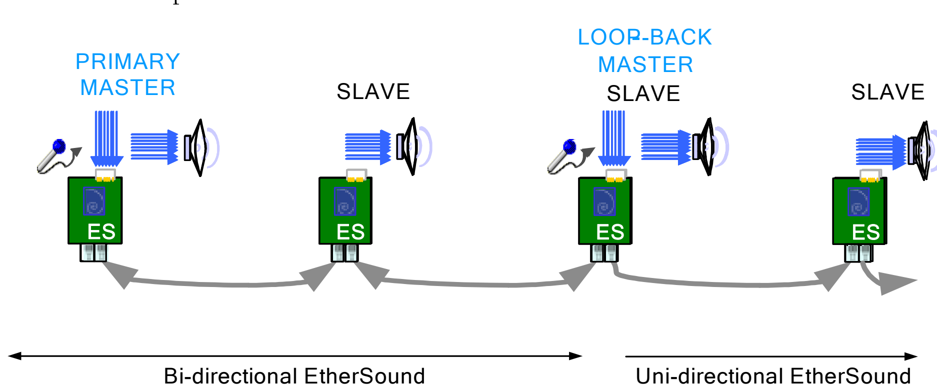
The simplest EtherSound configuration consists of one Primary Master and one Slave device in a uni-directional configuration:



In this case, audio is only transported in the downstream and therefore is uni-directional.

5.3 Bi-directional EtherSound

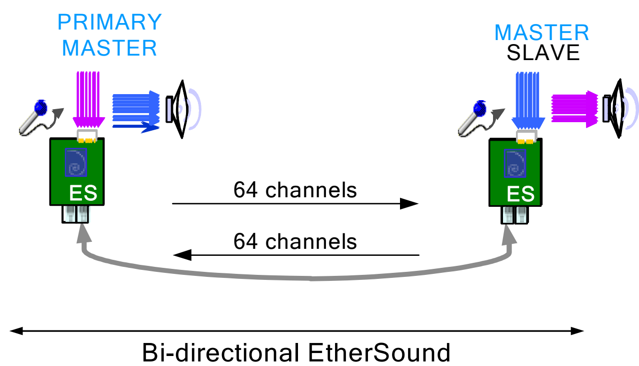
Bi-directional EtherSound stands for an EtherSound implementation and set-up featuring ubiquitous audio. In other words, audio can be inserted anywhere and is available anywhere. This feature is achieved by defining a loop-back device within a daisy chain that will send the EtherSound frame back to the Primary Master using the full-duplex feature of Fast Ethernet. Thus, any network channel inserted into the downward stream is available to any device upstream of the loop-back device.



Devices located downstream of the loop-back device are only provided with uni-directional audio service. Therefore, added audio is only available to downstream devices.

4.3.1 A bi-directional high speed, high capacity bridge using EtherSound

One specific application of bi-directional EtherSound consists in building a high speed, high capacity bridge between two audio devices. Going into details, the Primary Master writes audio data into the downstream which is read by the Master/slave device which in turn writes audio data in the upstream. In this configuration, 64 channels of 24-bit 48 kHz audio are available in both directions, for a combined link capacity of 128 channels.



Running at higher sampling rates, this bridge could accommodate thirty-two 96-kHz audio streams in each direction.

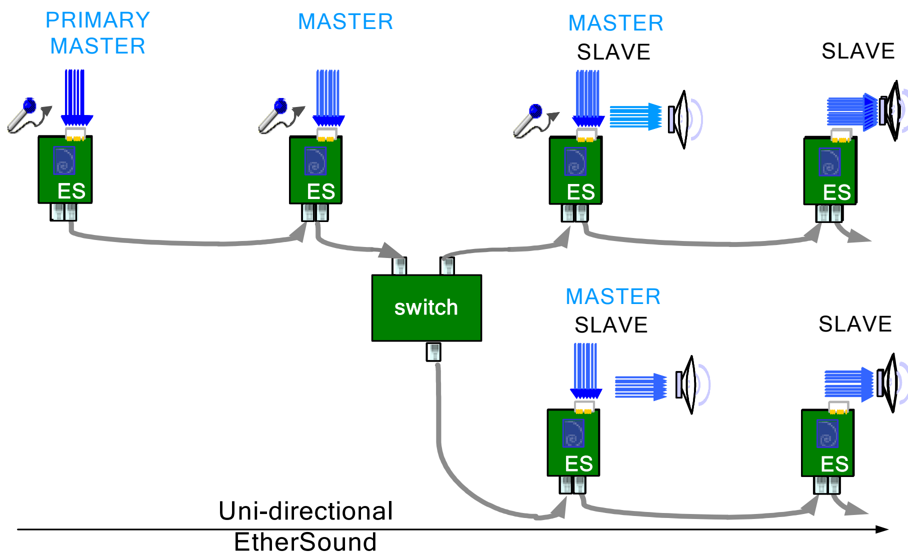
5.4 Uni-directional EtherSound

On the other hand, a EtherSound network can also be a purely uni-directional network.

Uni-directional EtherSound is compatible with various network topologies and may be Daisy Chain, Star or a combination of both.

Audio transmission is in only one direction. When an audio channel is "*extracted*" by a device from the EtherSound stream, that channel continues to remain available for all further downstream slave devices, until a master device inserts audio in this channel.

In this example, the Master/Slave device extracts the audio channels from the network, and is configured so that it applies some DSP processing and audio mixing before contributing the result down to the network as a new audio channel.



Audio data inserted by the Master/Slave of each daisy chain is only available to Slave devices located downstream.

5.5 Network wiring

A virtually unlimited number of devices may be daisy-chained, with no additional hardware, using the Ethernet-standard Unshielded Twisted Pair (UTP) Category 5 cabling. Devices may be up to 100 meters apart with no limitation on the distance between the first and last devices in the chain. Longer distances may be covered using media converters and fiber optics.

Common Ethernet switches may also be used to distribute uni-directional EtherSound transmissions to networked audio devices, facilitating more complex network architectures and extending the distance between devices.

A document dedicated to network architecture issues as well as a list of tested components is available for download on the Digigram web site.

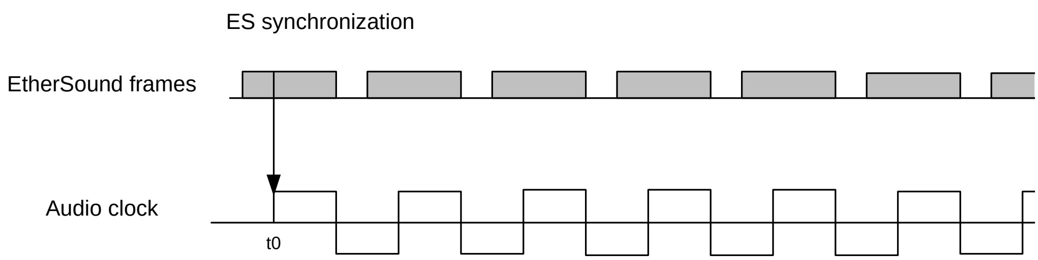
6 Clock distribution and timing aspects

6.1 Audio clocks

The audio clock of an EtherSound device can be derived from the network or be embedded. The network clock is generated by the first device connected on an EtherSound network, the Primary Master (see next section for an in depth description).

6.1.2 Network derived audio clock

The audio clock is derived from the incoming EtherSound frame:



An embedded PLL takes care of lowering the audio clock jitter.

All devices are synchronous to the Primary Master. Phase varies according to propagation delay.

6.1.3 External clock

An external synchronization feature is available when phase accurate synchronization across inputs and outputs of several devices is required. Depending on the network architecture and the sampling frequency, up to 8 consequent devices can be synchronized and in phase. If more devices are connected, the subsequent devices will have latency increase steps of one sample.

For obvious reasons, the external synchronization clock needs to be in sync with the Primary Master.

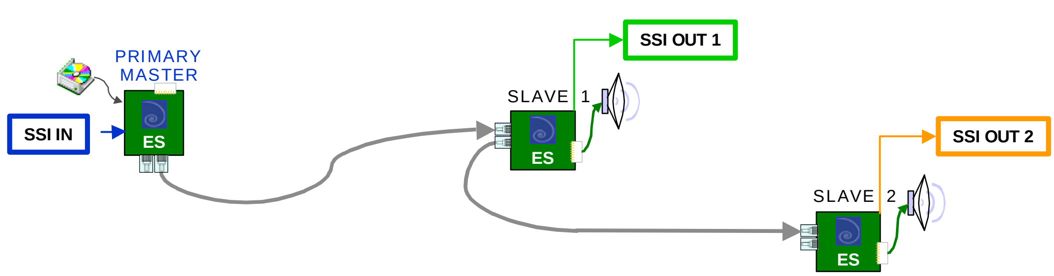
6.2 Timing and Latency of the Digigram Reference Designs

Some facts:

- Typical end-to-end, EtherSound transmission time (SSI In to SSI Out) is 6 samples.
- Less than 1.6 µs of latency is added per EtherSound device passing audio downstream in case of a daisy chained connection due to network output and input buffering.

6.2.1 A case study using a standard daisy chain

The case studies mentioned here do not include timing due to A/D and D/A conversions.



Assuming a daisy chained network running at 48 kHz and a distance of 100 meters (330 feet) between the devices:

- SSIOUT1 will have a latency of 125 µs (6 samples @ 48 kHz) + 1.5 µs (± 0.1 µs) + 0.48 µs (100 meters of cable) = 127 µs (± 0.1 µs),
- SSIOUT2 will have a latency of 125 µs (6 samples @ 48 kHz) + 2 * 1.5 µs (± 0.1 µs) + 2 * 0.48 µs = 129 µs (± 0.2 µs),
- Time difference between SSI OUT 1 and SSI OUT 2 is 2 µs (± 0.1 µs), about 1/10 sample @ 48 kHz, meaning that about ten EtherSound units can be daisy chained before getting a one sample time difference.

Furthermore, crossing switches may introducing important latency which might turn into undeterministic phase shift.

6.2.2 A case study using Word Clock synced devices in a daisy chain

If audio inputs and outputs need to be phase aligned, Word Clock can be distributed to EtherSound devices.

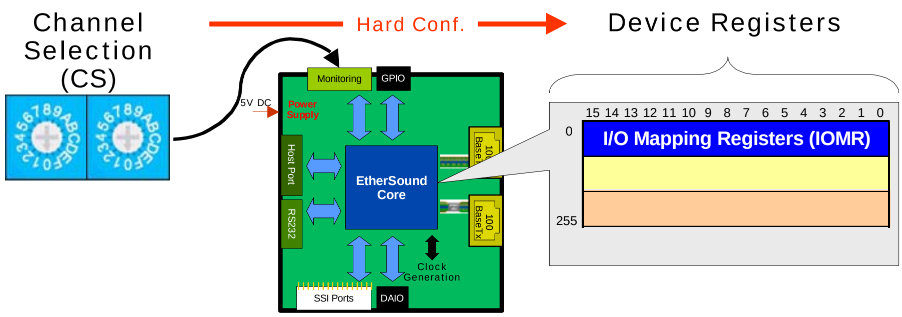
Assuming, same as above, a daisy chained network of Word Clock synchronized devices running at 48 kHz and a distance of 100 meters (330 feet) between the devices:

- SSIOUT1 will have a latency of 125 µs (6 samples @ 48 kHz),
- SSIOUT2 will have a latency of 125 µs (6 samples @ 48 kHz),
- SSIOUT10 will have a latency of 146 µs (7 samples @ 48 kHz).

7 Controlling EtherSound devices

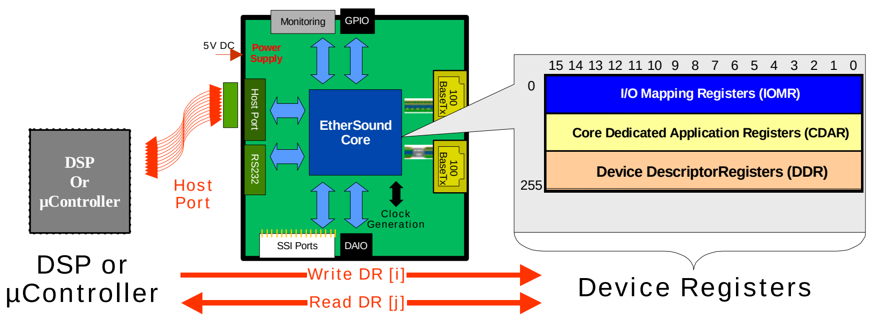
7.1 Local hardware management

By using rotary wheel, Dip Switches, or external logic directly connected to the specific monitoring port signals. This solution allows to control directly the I/O Mapping Registers (IOMR) (i.e. the local EtherSound network channel assignments) without any software as well as the loop back functions for bidirectional devices. It is available with Master and Slave devices.

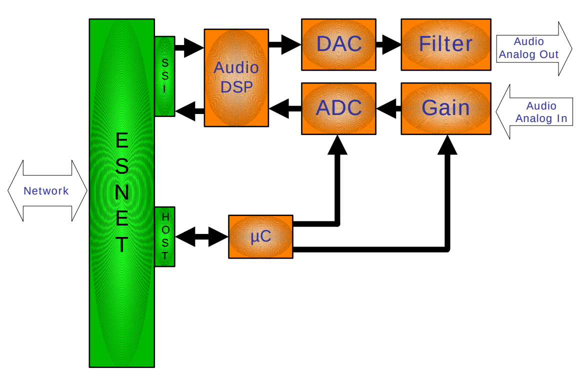


7.2 Local software management

By software, using the Host Port from an external µController or a DSP. This solution allows reading and writing in any local Device Register.



The onboard microcontroller can take care of many non EtherSound tasks such as A/D and D/A management and local HUI.



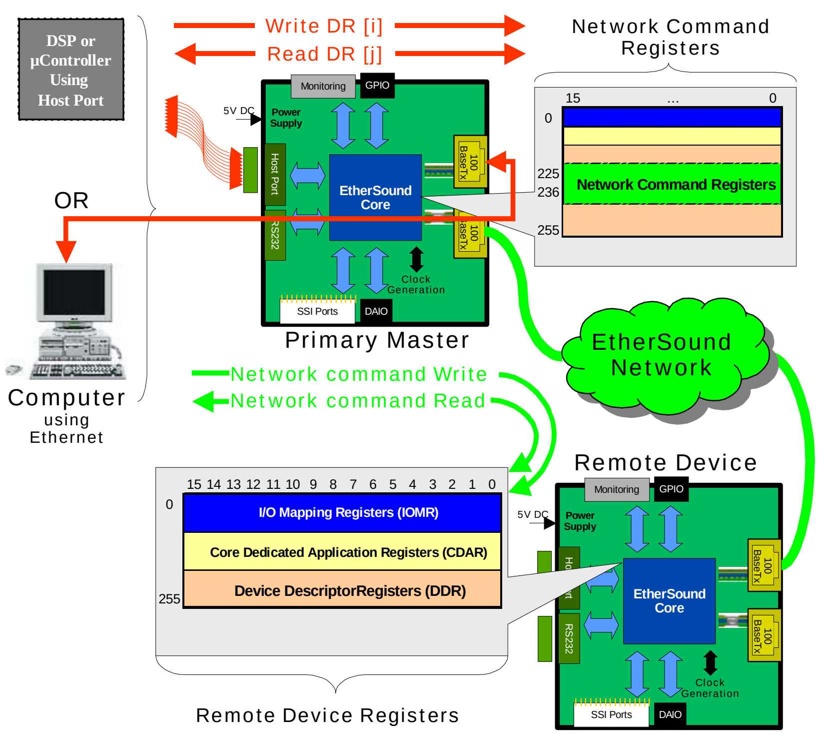
7.3 Remote management

Among the functions taken care of by the Primary Master are the control of the connected devices. As mentioned when describing the EtherSound frame, each frame includes a Command Packet. The purpose of this packet is to transport control command from the Primary Master to a designated connected device and to carry return values back to the Primary Master. The control command is inserted in the downstream frame by the Primary Master and read by the target device which in turn writes its return value in the upstream. This specific feature, available only on the Primary Master, allows to read and write any Device Register of any EtherSound device on the network. Whatever the audio transport, status information is always bi-directional so that it may be provided by any device on the network and read via the Primary Master.

The control commands allow management of the EtherSound network channels within each device as well as GPIO, serial port or any remote control of device specific functions of any kind.

Control of the Primary Master is available through:

- A local data port called the host port.
- A PC based API and its associated software.



7.3.1 Remote management using the host port

The Host port of the Primary master allows to access any register of any device connected on the network.

7.3.2 Remote software management using the ESM API

Software management is provided using the "Ethernet IN" interface connected to an external computer and thanks to a specific Ethernet protocol. This solution allows reading and writing in any local Device Register. It is only available on a Primary Master, but allows addressing any register of any device.

Digigram provides a PC/Windows Application Programming Interface (ESM API) to control any register on any EtherSound equipment of an EtherSound network through the connected Primary Master.

No knowledge of the Ethernet protocol is needed as this API provides a strong level of abstraction of the underlying network or networks such as:

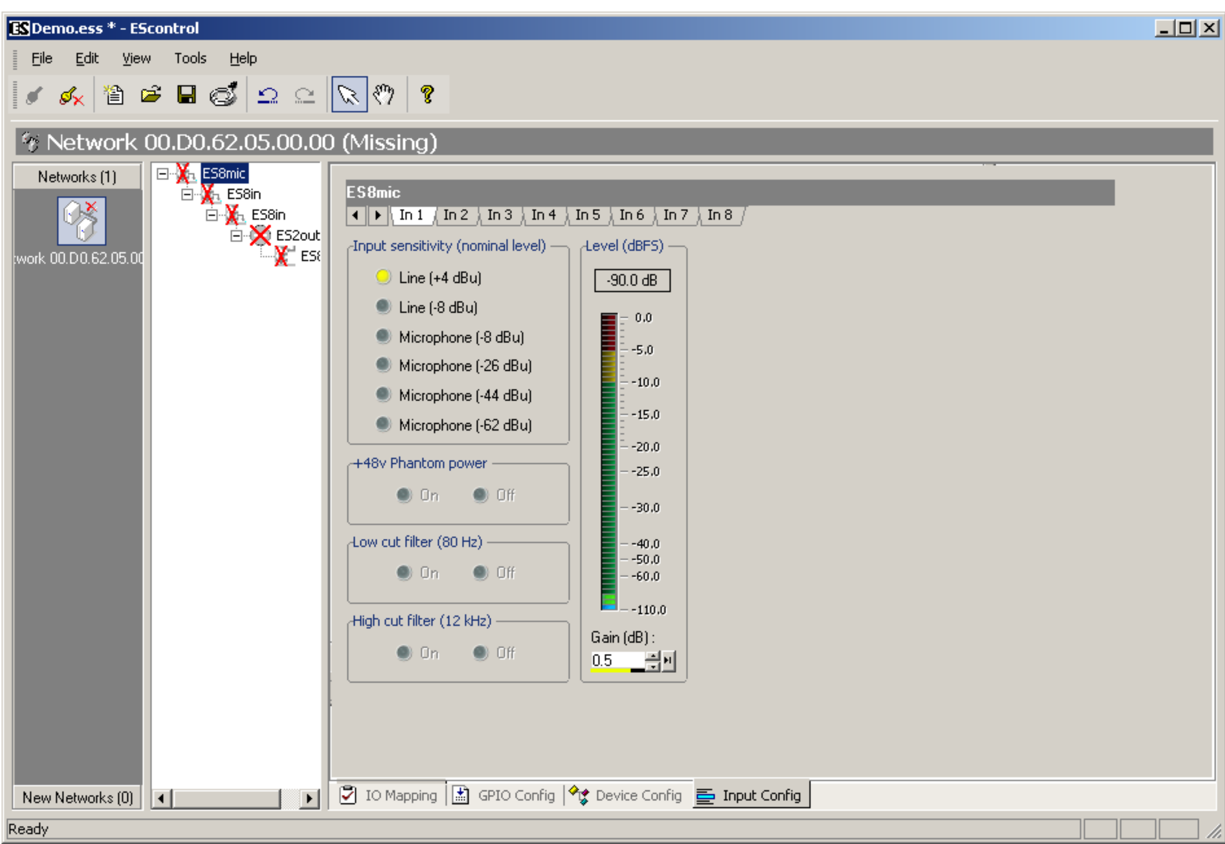
- Network enumeration
- Device enumeration
- Management of multi-vendor-devices networks
- Device identification
- Routing control
- GPIO, RS232 management
- Access to specific features

7.3.3 6.3.3 Remote software management using EScontrol

Digigram developed an application using this API. The Microsoft Windows based graphical interface of EScontrol provides easy control of an EtherSound network.

EScontrol provides access and control of:

- EtherSound routing
- EtherSound standard features management (GPIO, network errors)
- EtherSound product specific features if required extensions have been installed.

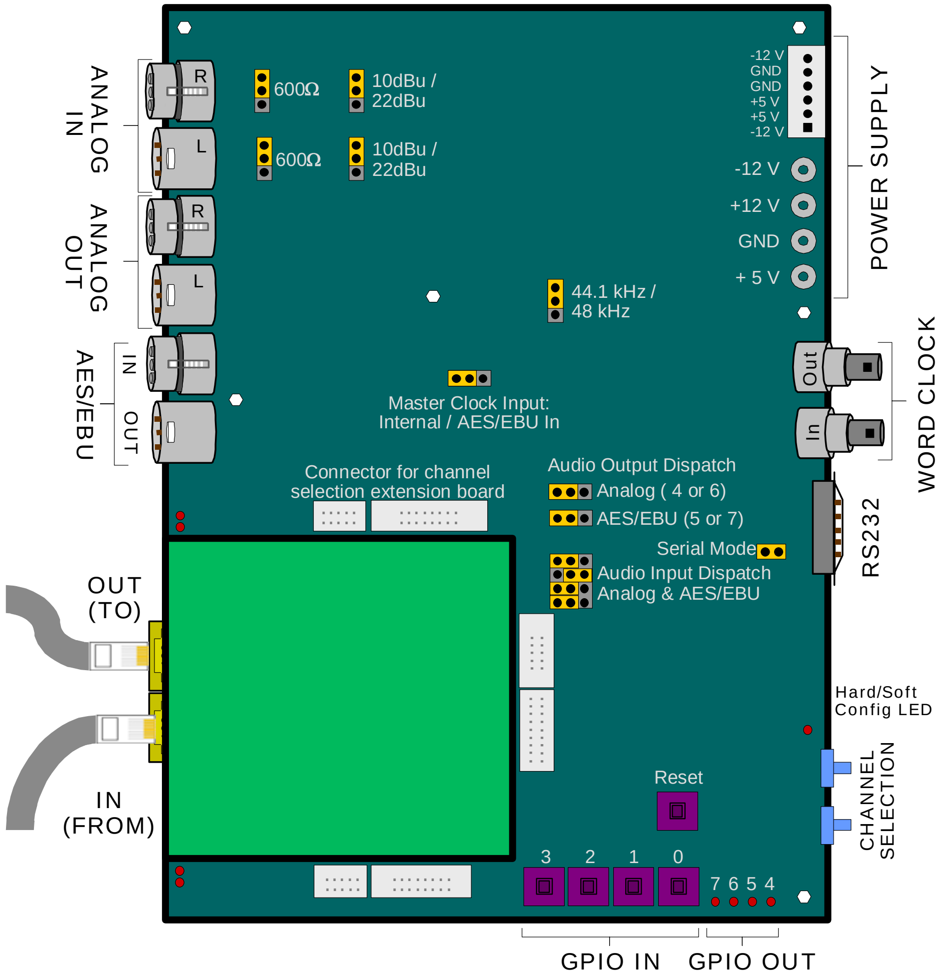


8 Evaluating the EtherSound technology

The ESnet Evaluation Board is a ready-to-use platform for testing and evaluating the EtherSound technology. EtherSound hardware implementations may be designed to operate in master mode, slave mode, or both. Devices may contribute audio channels to the EtherSound stream, playback audio channels, or both.

Each Digigram ESnet Evaluation Kit comes complete with two ESnet Evaluation Boards, software and documentation to help you set-up and test a minimum EtherSound environment.

The Evaluation Boards include a large variety of connections in order to provide a comprehensive range of test and evaluation capabilities.


ANALOG IN
R
L
ANALOG OUT
AES/EBU
IN
OUT
600Ω
10dBu / 22dBu
44.1 kHz / 48 kHz
Master Clock Input: Internal / AES/EBU In
Connector for channel selection extension board
Audio Output Dispatch
Analog (4 or 6)
AES/EBU (5 or 7)
Serial Mode
Audio Input Dispatch
Analog & AES/EBU
-12 V
GND
GND
+5 V
+5 V
-12 V
+12 V
+ 5 V
POWER SUPPLY
Out
In
WORD CLOCK
RS232
OUT (TO)
IN (FROM)
Hard/Soft Config LED
CHANNEL SELECTION
Reset
3 2 1 0
7 6 5 4
GPIO IN
GPIO OUT

•

9 Appendix A: Glossary

BROADCAST

A configuration where an equipment transmits simultaneously towards all the other equipments.

CDAR

CDAR are the Core Dedicated Application Registers. These EtherSound registers are reserved for Digigram or client-specific peripheral functions, such as RS232, GPIO, analog gain….

DAIO

DAIO stands for Dedicated Application Inputs Outputs. It is a set of 10 pins available in the Digigram Reference Designs. They require a specific logic programming of the FPGA.

DAISY CHAIN

The **Daisy Chain** is a network topology where all devices are “serially” linked one to the other.

DDR

DDR are the Device Descriptor Registers. They describe the status and the configuration of the EtherSound Kernel of the device.

ETHERNET

Most used Local Area Network (LAN), originally developed by Digital Equipment Corporation, Intel Corporation and Xerox Corporation (DIX), standardised by the IEEE Committee, ISO normalized. Standard off-the-shelf components widely available at reasonable cost.

FPGA

A Field Programmable Gate Array is a programmable electronic device and allows for flexible designs along with on the field software updates.

FRAME

A **Frame** is a set of characters that are transmitted as an entity according to a defined format. The frame follows a coding procedure at the physical level before emission. The EtherSound Frame is fully compliant to Ethernet 802.3 standard.

FULL DUPLEX

A transmission is said to be **Full Duplex** when data can be transmitted and received simultaneously.

GPIO

GPIO stands for General Purpose Inputs Outputs. The ESnet MS 88 Eeprom 300K Reference Design and the ESnet MSx 88 Eeprom 200K Reference Design have 8 GPIO pins that can be fully configured to either retrieve or output an Low Voltage Differential Signaling (LVDS) level to remotely control external devices such as relays and switches. GPIOs can be used for example with a microphone equipped with a two-position switch that will be connected in that case to the GPIO inputs.

HOST PORT

The Host Port is an interface provided on the ESnet MS 88 Eeprom 300K Reference Design and the ESnet MSx 88 Eeprom 200K Reference Design for communication with a µController or a DSP on an Application Board.

The communication will be performed via the EtherSound Device Registers.

IEEE 802

Committee created in 1980 that has established standards for informatics equipment connection. IEEE 802.3-4-5 describes the Physical and Link Layers (MAC) from the ISO OSI reference model. Those different physical layers can interface with the IEEE 802.2 norm that describes the upper part of the Link Layer (LLC).

LATENCY

The latency of a device measures the insertion delay it will add in a system.

MASTER

A device in the network that is downstream from the Primary Master that contributes additional audio channels into the EtherSound stream is called **Master.**

A Master answers to the status requests and commands of the Primary Master. Also see *Master/Slave* and *Primary Master*.

MASTER/SLAVE

A device may be configured to both contribute and extract audio channels. This device is called Master/Slave.

A Master/Slave answers to the status requests and commands of the Primary Master.

MULTICAST

A configuration where an equipment transmits simultaneously towards a group or list of other equipment.

NETWORK CHANNEL

Elementary slot of the audio packet of an EtherSound frame. In EtherSound protocol V1 an audio packet contains sixty-four 24-bit network channels.

PRIMARY MASTER

The first EtherSound device in the network is called the Primary Master. As well as being a source of audio for the network, the Primary Master provides the commands and audio clock.

SLAVE

An EtherSound device that receives the EtherSound stream and restores standard audio is called a Slave.

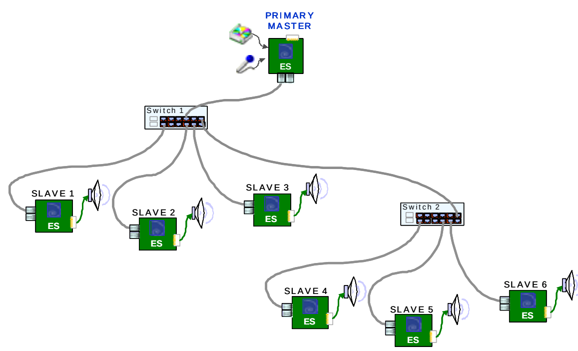
A slave answers to the status requests and commands of the Primary Master.

SSI

The Synchronous Serial Interface is the most common way to send/receive data to/from a Standard Audio DAC or ADC. This interface is provided on the ESnet MS 88 Eeprom 300K Reference Design and the ESnet MSx 88 Eeprom 200K Reference Design in form of 8 data wires (4 IN allowing 8 Channel Upload, 4 OUT allowing 8 Channel Download) and Clock control.

STAR

Star is a network topology where all devices are connected to a same unit (a switch in the following picture) that is handling all the communications.



UNICAST

A configuration where one equipment transmits towards a single equipment.

Digigram S.A.
Parc de Pré Milliet
38330 Montbonnot
FRANCE
Tel: +33 (0)4 76 52 55 01
Fax: +33 (0) 4 76 52 53 07
E-mail: info@digigram.com

Digigram Inc.
2101 Wilson Boulevard, Suite 1004, Arlington, VA 22201
USA
Tel: +1 703 875 9100
Fax: +1 703 875 9161
E-mail: input@digigram.com

Digigram Asia Pte Ltd.
350 Orchard Road - #19-07 Shaw House
Singapore 238868
SINGAPORE
Tel: +65 6291 2234
Fax: +65 6291 3433
E-mail: info_asia@digigram.com