Music & the Internet
MUMT301

Gabriel Vigliensoni
Schulich School of Music
McGill University

Plan

- Review of last class and assignment #3
- Internet technologies (i.e., Ethernet, TCP/IP, IP Headers, TCP Header, DNS, Ports, DHCP, FTP, SSH, HTTP)
- Sound file formats
- Audio compression
- Introduction to JavaScript
- Assignment #4

Assignment 3

- Andrew
- Amelya
- Carlos
- Emmanuel
- Jina
- Matthew
- Olivier
- Patrick
- Ryan
- Samuel
- Truman
- Willis

Sound file formats

- **Exponential growth** of audio material on the Internet since 1995
- “**mp3**” most searched term in 1999
- Big impact of coded audio: mostly everybody is using **MPEG-1/2 Layer-3**
  - also known as MP3
- Broadly speaking, **sound content** is delivered in **two format categories**:
  - As **recorded sound**:
    - often called *waveform* sound
  - As **structured audio**:
    - Sounds are generated in a dynamic manner at runtime
    - MIDI, MODs (e.g., trackers)

Structured Audio

- Structured audio formats provide **data to support dynamic construction of sound** through hardware and software
- *Sequencers* and *trackers* control
  - the **timing of sounds**, i.e., when individual sound elements start or stop
  - **sound attributes** such as volume, pitch, and other features
- *Sound elements* can be
  - short sections of **sound samples or loops** or
  - **data elements that characterize a sound** so that a *synthesizer* can produce the actual sound

Structured Audio

- Structured audio **does not convey audio**, the sounds are generated in a dynamic manner at *runtime.* Most common formats are:
- **MIDI** (1983)
  - **technical standard** to allow the **communication between electronic music instruments** (and computers)
  - Organized by the MIDI association
  - MIDI 1.0 Detailed specification, MIDI message example
  - MIDI classical archives
  - MIDI world
- **Module files or MOD**s (late 80s) are used in music *tracker* software
  - Arrangement of discrete musical notes positioned at discrete chronological positions on a timeline
  - E.g., OpenMPT, 16bitshock, Soundbox, FruityLoops, Renoise
  - http://modarchive.org/

Recorded sound

- Audio file data is stored in a **binary representation**
  - it is necessary to have a **format specification** in order to know how to read a given format's data
- <u>Sustainability of Digital Sound Formats. Library of Congress Collections</u>
- <u>Comprehensive list of audio file formats</u>
- The most common audio file formats are/were:
  - SND or AU (NeXT, Sun)
  - AIFF (Apple, SGI)
  - WAV (Microsoft)
  - MP3 (MPEG)
  - FLAC (OpenSource)
- Audio data can be stored in compressed or uncompressed formats

Uncompressed formats

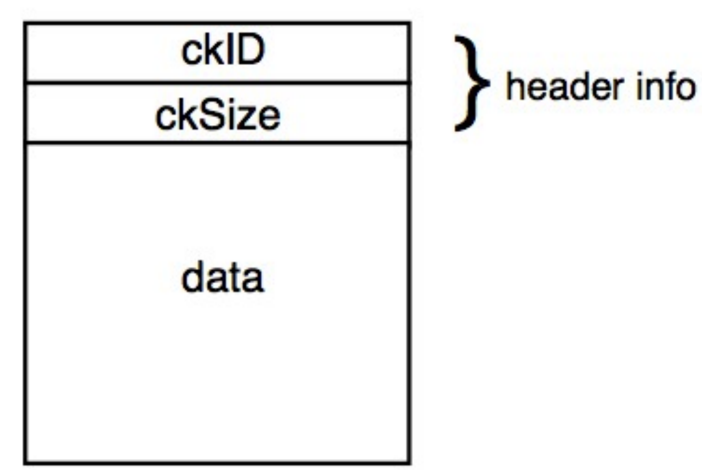
- LPCM (Linear Pulse Code Modulation Audio):
  - **digital representation of analog signal**
  - **magnitude** of the signal is **sampled regularly** at uniform intervals
  - **quantized** to a series of values in binary code
  - defining characteristics:
    - **sampling rate** (44.1kHz, 48kHz, 88.2kHz, 96kHz, 176.4kHz, 192kHz)
    - **bit depth** (8, 16, 20, 24, 32 bits per sample)
  - used for uncompressed encoding of audio data in the **Compact disc Red Book Standard** (i.e., Audio CD, 1982)

Compact Disc

- **Optical** storage of digital audio (data)
- Introduced in 1982 by Sony and Phillips (together)
- **Originally for sound** only, but later for all types of data
- 74–80 minutes of audio at 44.1KHz/16bit/stereo
- How much data?
    - 88,200 bytes/sec per channel
    - 5,292,000 bytes/minute per channel

Uncompressed formats RIFF File formats

- RIFF (Resource Interchange File Format) is a **tagged file structure for multimedia resource files**
- Developed by Electronic Arts (the gaming company!)
- Aimed to **facilitate data transfer between different software and companies**
- Generic container file format (AIFF, WAVE, RIFF, SMF)
- RIFF is **not a file format**, but a **file structure** that defines a class of more specific file formats
- Based on **headers** pointing to "**chunks**" of data

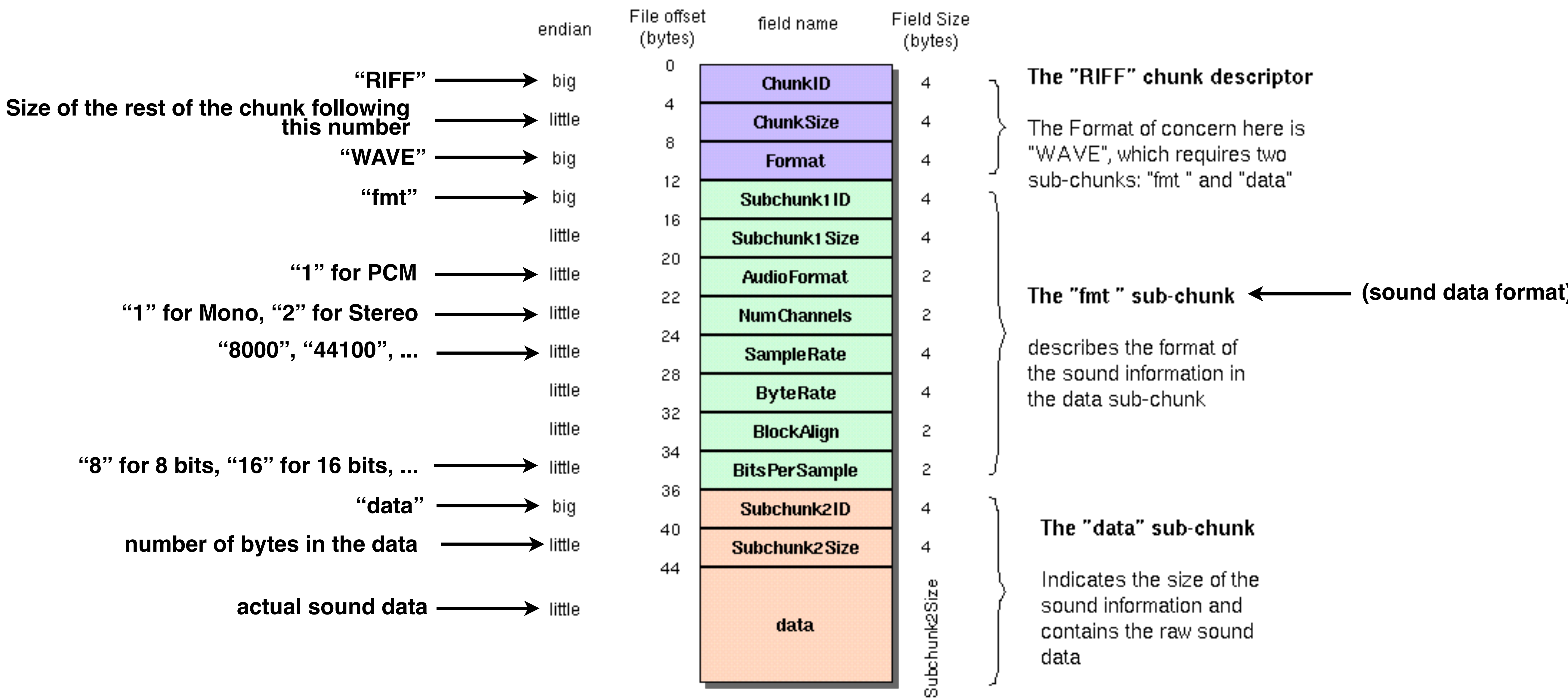


A chunk.

Uncompressed formats
RIFF File formats

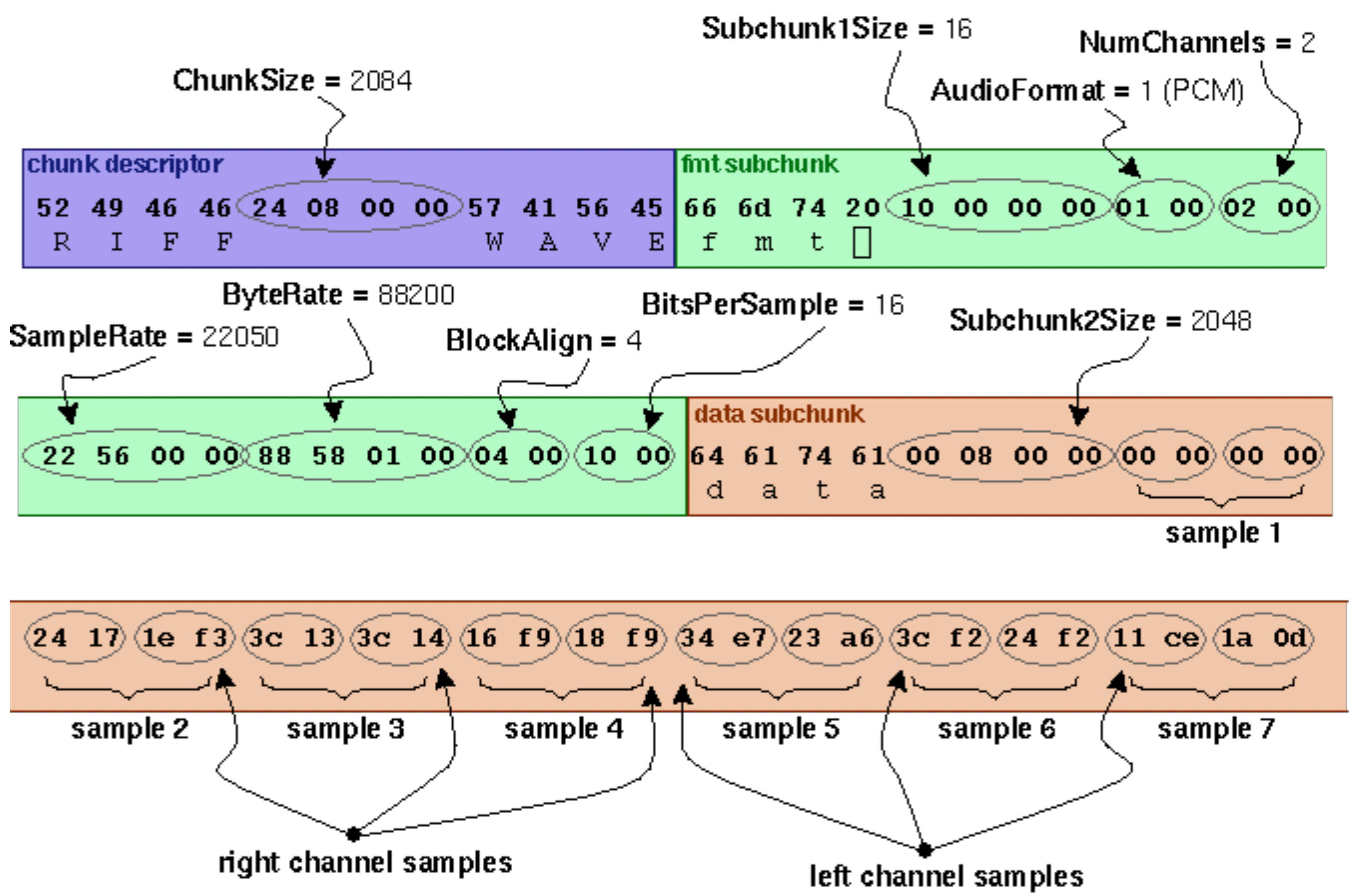
- AIFF (Audio Interchange File Format, 1989)
- WAVE (Waveform Audio File Format, 1991)

Uncompressed formats
The Canonical WAVE file format



Taken from http://soundfile.sapp.org/doc/WaveFormat/

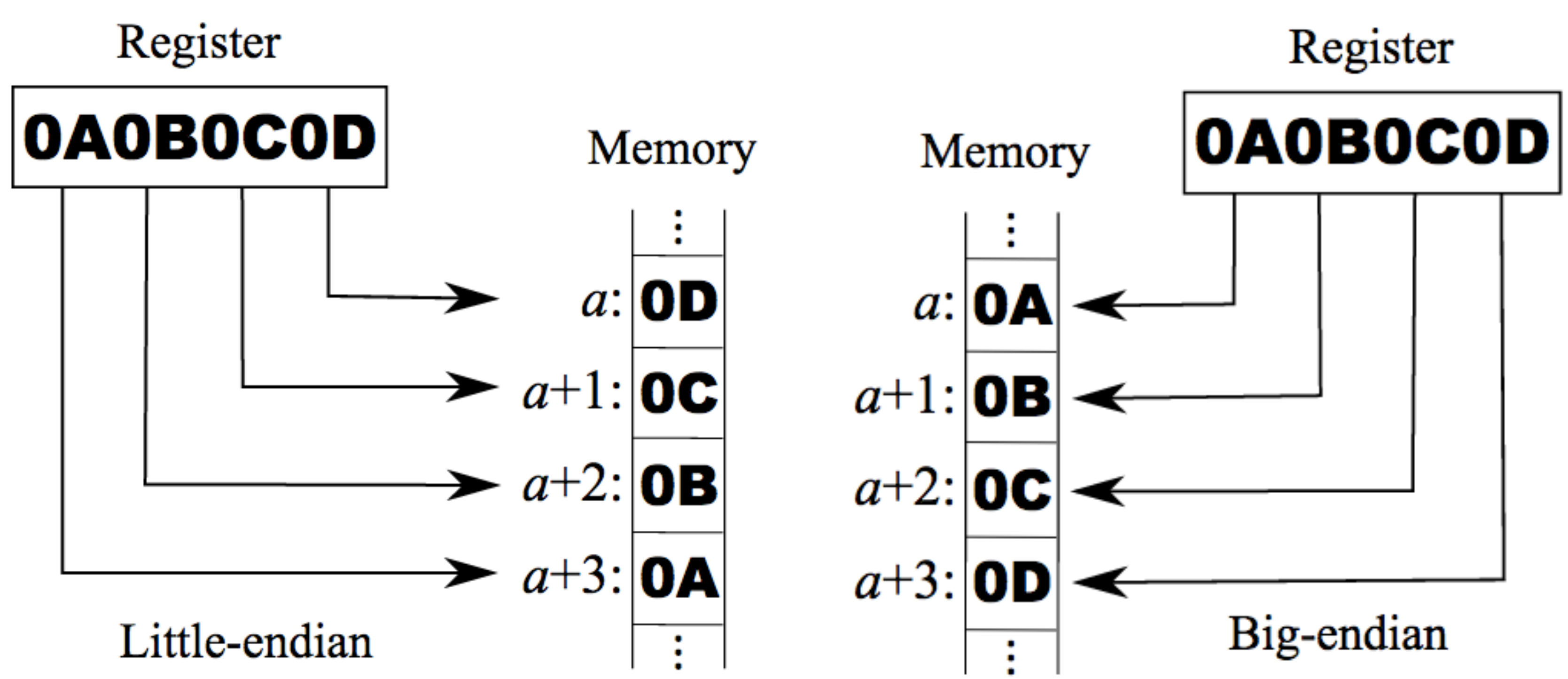
Uncompressed formats
The Canonical WAVE file format



Taken from http://soundfile.sapp.org/doc/WaveFormat/

Endianness

- In WAVE files, data bytes are “little endian” ordered, When looking at multiple bytes, the first byte is smallest.
- In AIFF files, data bytes are “big endian” ordered. When looking at multiple bytes, the first byte (lowest address) is the biggest.



Taken from http://en.wikipedia.org/wiki/Endianness

Compressed sound file formats

- Lossy and Lossless
- **Lossy compression**:
    - only **an approximation of the original data can be reconstructed after decompression**,
    - How well it approximates the original data **depends on the compression rate**
    - Common lossy formats
        - MP3 (patented!)
        - Vorbis (aka Ogg Vorbis), xiph.org webpage (free and open source) (V1.0 2002)
        - WMA
- **Lossless compression**:
    - **data can be perfectly reconstructed after decompression**
    - Lossless formats
        - Monkey's Audio, WavPack, Apple Lossless, ...
        - FLAC, xiph.org webpage
            - non-proprietary
            - no patent restricted
            - open-source

Audio compression

- Basic task of audio compression?
  - to compress the digital audio data in a way that
    - compression as efficient as possible, i.e., file **size as small as possible**
    - the decoded audio should be **as close as possible to the original** audio before compression
- **Perceptual audio coding** is the technique used to accomplish this
  - research topic since the late 70's, exploded since 1986
  - uses **knowledge from psychoacoustics** to reach the target of efficient, but inaudible, compression
  - is a *lossy* compression technique
    - the decoded file **is not a bit-exact replica of the original** digital audio data

The MPEG and MPEG-1, -2

- MPEG (Moving Pictures Expert Group) has been in charge of developing **generic standards for the coded representation of moving pictures and associated audio**
  - standardization of **compression techniques** for video and audio
- MPEG-1 was the first phase of MPEG work in 1988. Became an ISO standard in 1992
  - Generic coding system with three layers: Layer-1 to Layer-3
    - **Layer-3** is the highest complexity mode: to **provide highest quality at low bit-rates**
- MPEG-2 was the second phase of MPEG (1994). No new coding algorithms, but
  - **coding of multichannel** signals (including 5.1)
  - **coding efficiency at very low rates** (lower sampling frequencies, e.g., 16Khz)
  - it is **backwards compatible** with MPEG-1
  - its main application is digital television
  - named **MPEG-2 Advanced Audio Coding (AAC)**

MPEG-1 Layer-3 audio standard

- MPEG audio standard is **informative instead of normative**
  - minimum amount of **normative elements**:
    - the **data representation** (i.e., format of the compressed audio)
    - the **decoder** (however there is freedom in how to implement it)
- **Encoding** of MPEG audio is **left to the implementer**
  - the standard only gives description of example encoders
  - MPEG audio **encoders can vary in quality**

MPEG-1 Layer-3 characteristics

- **Flexibility**
  - **Different operating modes**:
    - **Single channel**
    - **Dual channel** (two independent channels, e.g., two different language versions of an audio piece)
    - **Stereo** (no joint stereo coding, the two channels are encoded independently)
    - **Joint stereo** (information about differences from each channel is stored in one channel, whilst identical information is stored in the other. Help to reduce bit-rate)
- **Sampling frequency**
  - **MPEG-1**: **32kHz, 44.1kHz, 48kHz**
  - MPEG-2: extends MPEG-1 to half rates: 16kHz, 22.05kHz, 24kHz
  - MPEG-2.5: Fraunhofer-proprietary: 8kHz, 11.05kHz, 12kHz
- **Bit-rate**
  - the MPEG-1 standard defined a range of **bit-rates from 32 kbits/s up to 320k bits/s**
  - MPEG-2 standard extends the bit-rate to 8 kbits/s
  - **selection of bit-rate left to the operator of the audio coder**

MPEG-1 Layer-3 perceptual model

- PCM digitally reproduces the waveform of an incoming signal as accurately as possible

- **Human aural perception works different**, our ears and brain are imperfect and biased

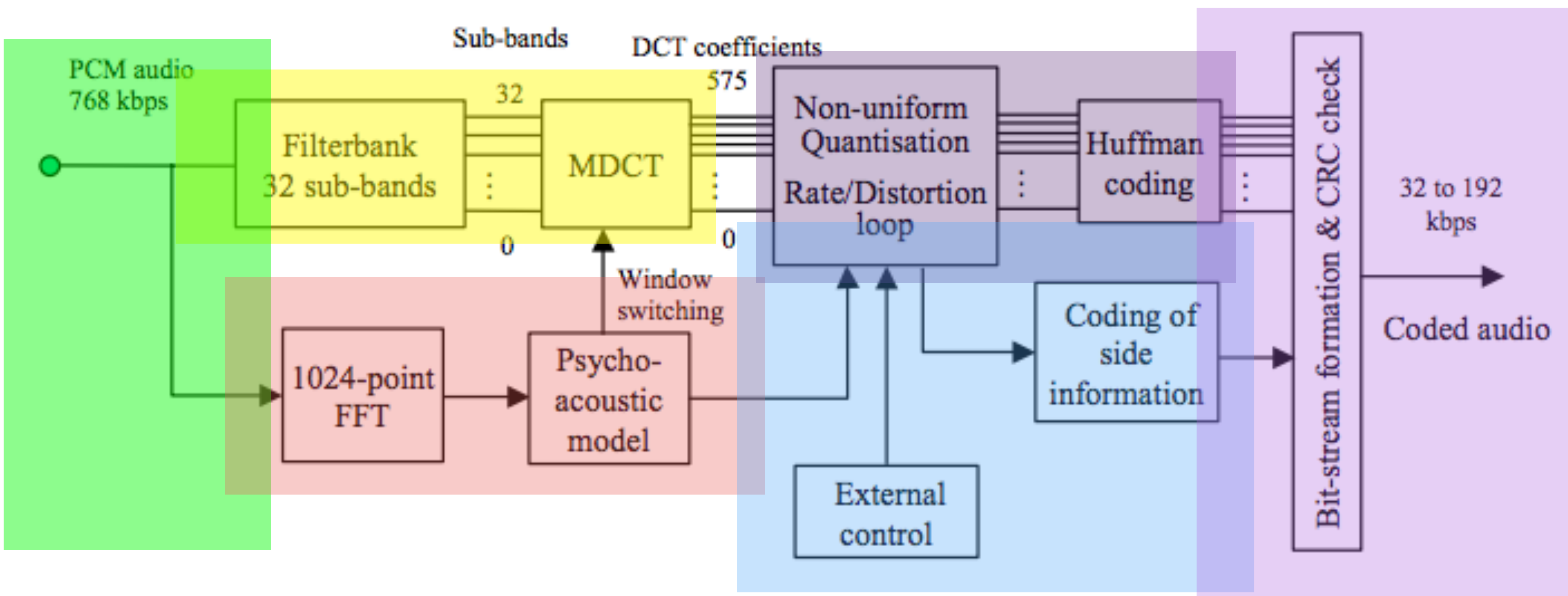
- **Sounds can be masked in frequency and/or time**

- The theory behind perceptual coding uses the idea that **our aural perception** doesn't need the actual bit-to-bit waveform representation, but only the **properties of the waveform that are most important for the listener**, and prioritize the recording of these properties

- In other words, **PCM attempts to capture a waveform *as it is*, an MP3 attempts to capture *as it sounds*** (as we perceive it)

MPEG-1 Layer-3 perceptual model

- Perceptual coding basic concept: **irrelevancy**
  - Certain properties of any given waveform will be not perceived by a human listener, and so it will be **meaningless to store them**
- Perceptual model depends on peculiarity of human auditory perception: **auditory masking**
  - in the temporal domain it is called **temporal masking**: a sudden stimulus makes inaudible other sounds immediately preceding or following the stimulus
  - in the frequency domain, it is called **frequency masking**: the threshold of hearing for one sound is raised by the presence of another sound
- Masking enables perceptual coding to get away much of the data that conventional waveform coding (e.g., PCM) stores
- **Not all irrelevant data is discarded, but fewer bits are assigned to the masked elements than to relevant ones**
- This process introduces distortion, but it will be hopefully confined to the masked zone, and hopefully will be imperceptible on playback.



Taken from Brandeburg, K. 1999. MP3 and AAC explained. In Proceedings of the AES 17th International Conference on High Quality Audio Coding

MP3 Header
MP3 Data
MP3 Header
MP3 Data
+++ Repeated +++
MP3 Header
MP3 Data
MP3 Header
MP3 Data

MP3 header

Bits	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Binary	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	0
Hex	F				F				F				B				A			
Meaning	MP3 Sync Word												Version	Layer		Error Protection	Bit Rate			
Value	Sync Word												1 = MPEG	01 = Layer 3		1 = No	1010 = 160			

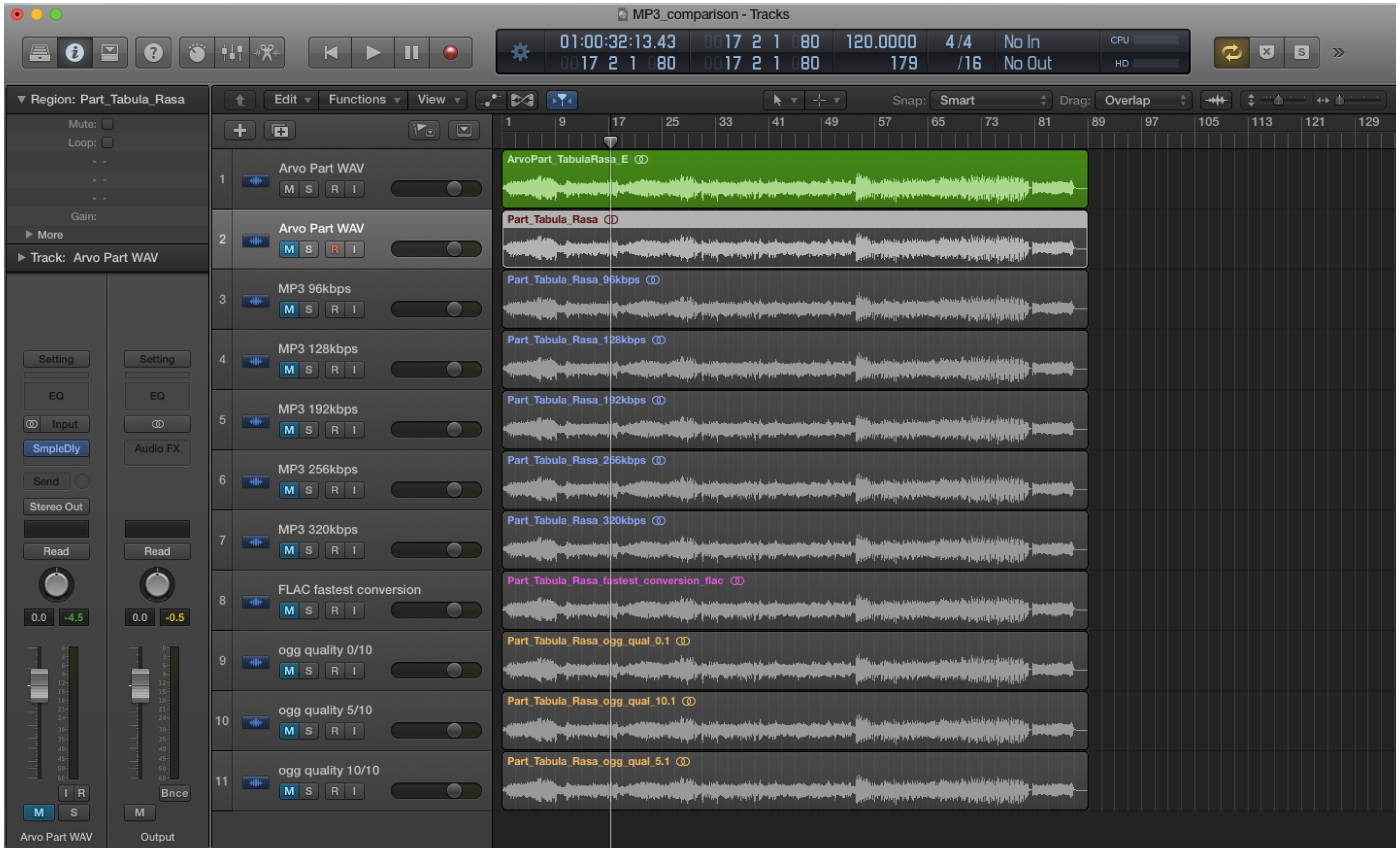
21	22	23	24	25	26	27	28	29	30	31	32
0	0	0	0	0	1	0	0	0	0	0	0
0				4				0			
Frequency		Pad. Bit	Priv. Bit	Mode		Mode Extension (Used With Joint Stereo)		Copy	Original	Emphasis	
00 = 44100 Hz		0 = Frame is not padded	Unknown	01 = Joint Stereo		0 = Intensity Stereo Off	0 = MS Stereo Off	0 = Not Copy-righted	0 = Copy Of Original Media	00 = None	

Taken from http://en.wikipedia.org/wiki/MP3

MPEG-2 AAC

- Same paradigm as MPEG-1 Layer-3, but
  - **higher frequency resolution** (1024 instead of 576 frequency lines)
  - improved joint stereo coding
  - improved Huffman coding
- AAC reaches the same quality as Layer-3 at 70% of the bit-rate
- Do not confuse MPEG-2 with MP2 (MPEG-1 Layer-2). MPEG-2 is standard for video (and associated audio), whose audio can be in a number of formats, including AC3 and AAC

MP3 comparison



BREAK

Introduction to JavaScript

https://mumt301.github.io

Today’s class

- Review of last class and assignment #3
- Sound file formats
- Audio compression
- Introduction to JavaScript
- Assignment #4

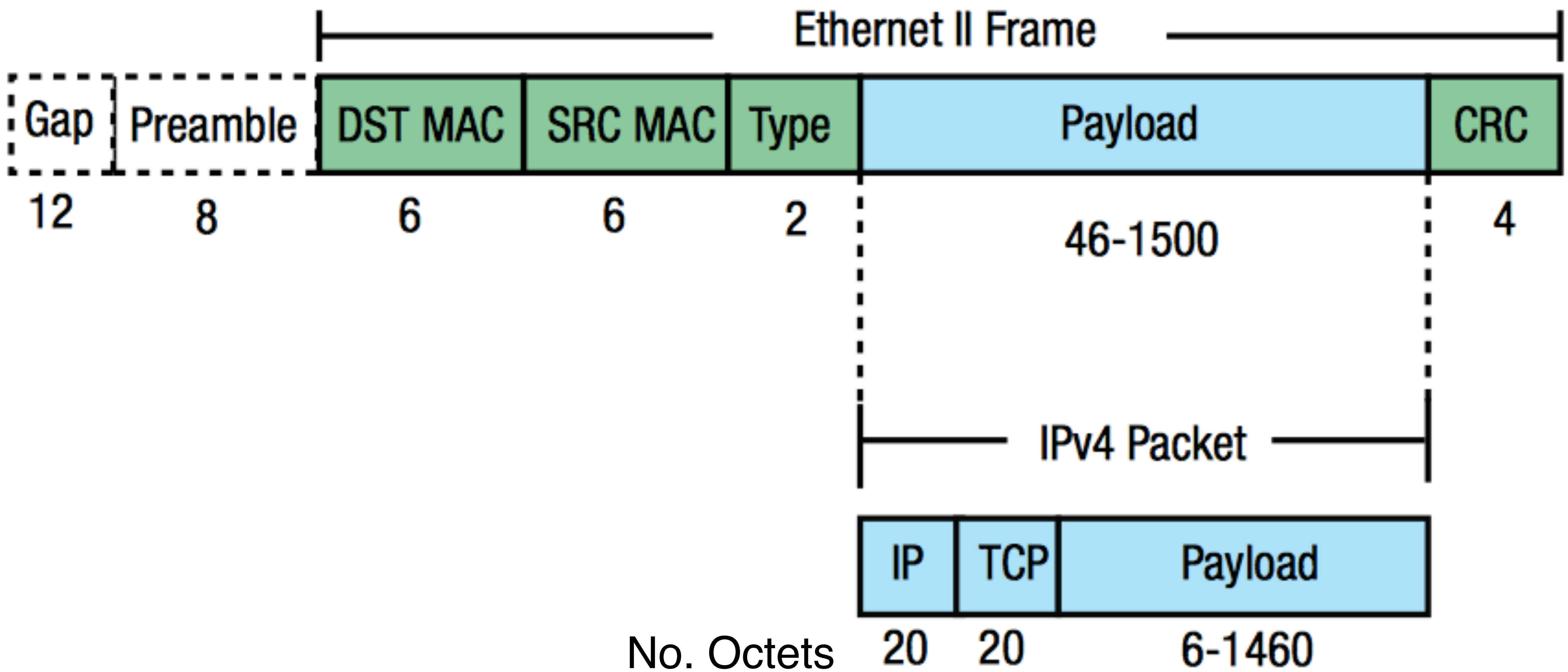
HTTP requests

- Request messages are at the heart of web communications using HTTP
- These messages are sent using URLs (Uniform Resource Locators)

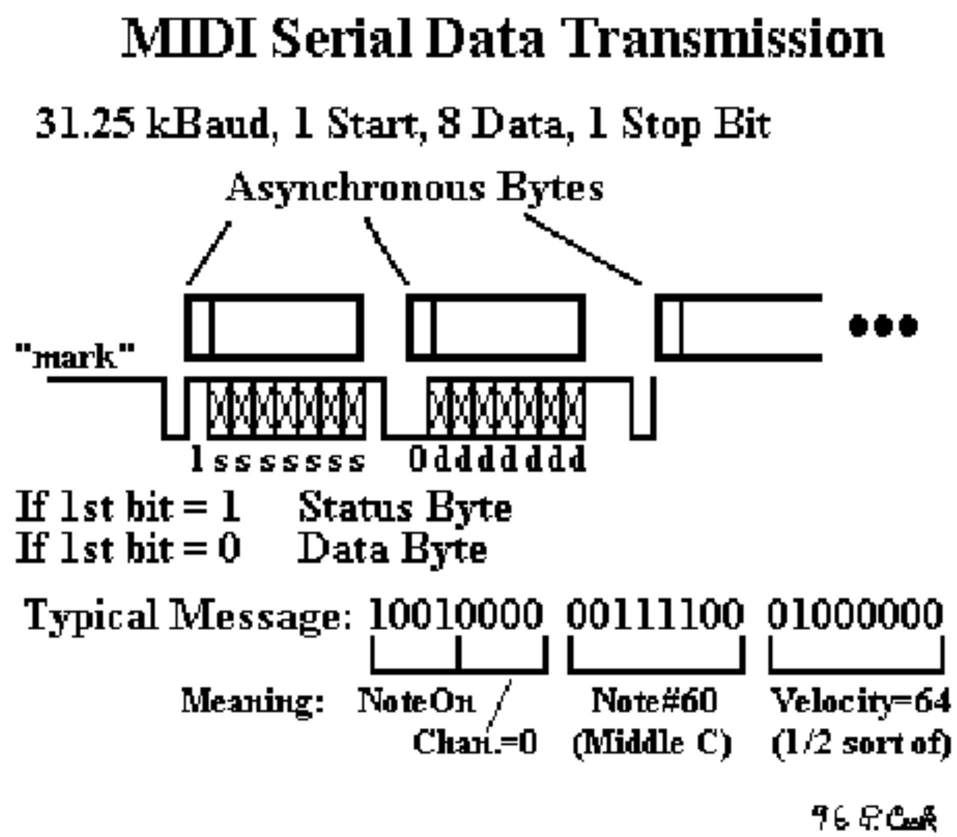
https://www.google.ca/search?q=malcisne+music

Complete Ethernet Packet

Taken from openmicrolab.com



MIDI Message example



The **NOTE ON** message is structured as follows:

- Status byte : 1001 CCCC
- Data byte 1 : 0PPP PPPP
- Data byte 2 : 0VVV VVVV

Pitch value: 60

The **NOTE OFF** message is structured as follows

- Status byte : 1000 CCCC
- Data byte 1 : 0PPP PPPP
- Data byte 2 : 0VVV VVVV

where:

CCCC is the MIDI channel (from 0 to 15)

PPP PPPP is the pitch value (from 0 to 127)

VVV VVVV is the velocity value (from 0 to 127)