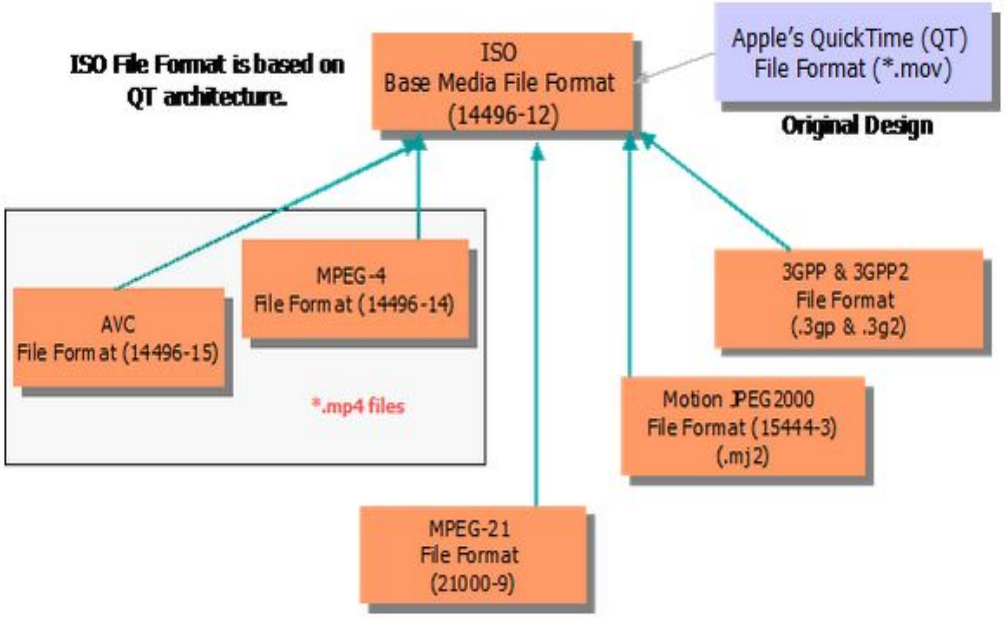
MP4 File Format

Introduction

- MPEG-4 part 14 —> MPEG-4 part 12 —> Quicktime file Format

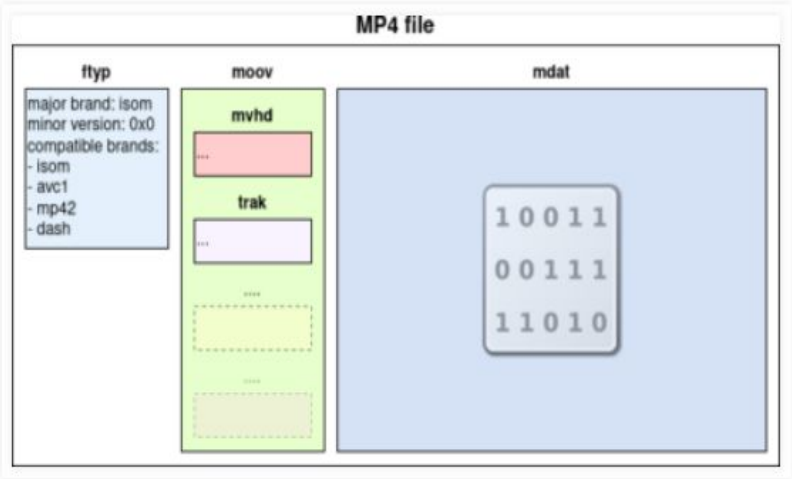


Terms used

- ➔ **Box/Atom-** A object oriented building block
- ➔ **Sample-** Timed unit of media
- ➔ **Chunk-** A group of samples of a single track
- ➔ **Container box-** Contains other boxes as child
- ➔ **Sync Sample-** Does not depend on other samples while decoding

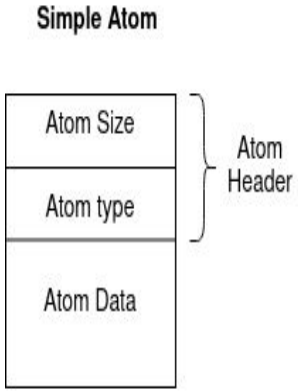
Introduction

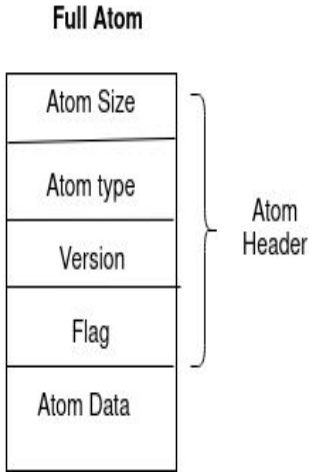
- **Movie -> Tracks -> Samples**
- Each track corresponds to a **elementary stream**
- Each track has one or more **sample descriptions**
- Each sample in the track is tied to description by reference
- **Description** defines how sample is decoded



Atom/Box Structure

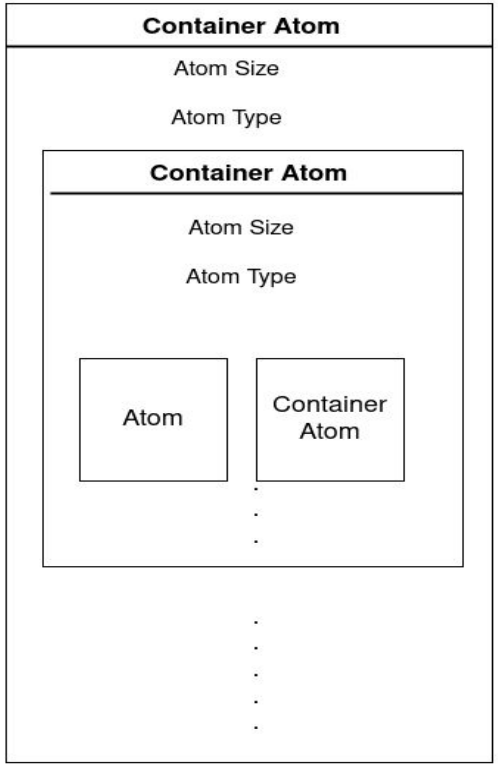
- **Atom Size(4 bytes)** = Header + Data
- **Atom type(4 bytes)** = Uniquely identifies the Atom (Ex- mdat, moov, ftyp…..)
- **Version(1 byte)** Version of the box (0,1)
- **Flag(3 byte)** Indicate presence or absence of fields





Container Atom

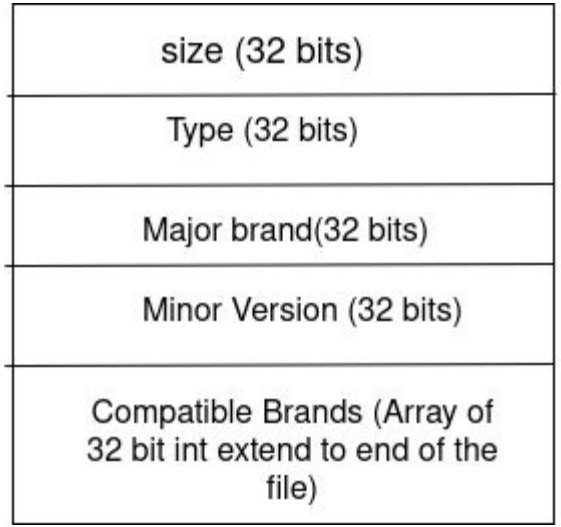
- Contains atoms or container atoms as part of its data



Types of Box/Atom

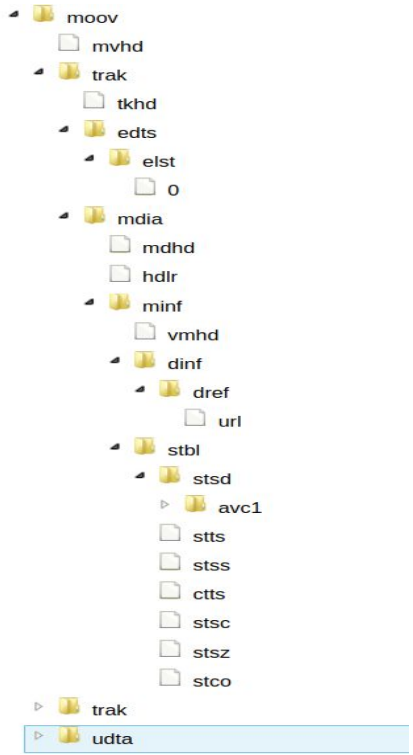
ftyp(file type compatibility) Atom

- Present at beginning of the file
- Distinguish between closely related file types such as QuickTime, MPEG-4 and JPEG-2000
- *Major brand* specify **best use** of the file
- *Minor version* is minor version of major brand



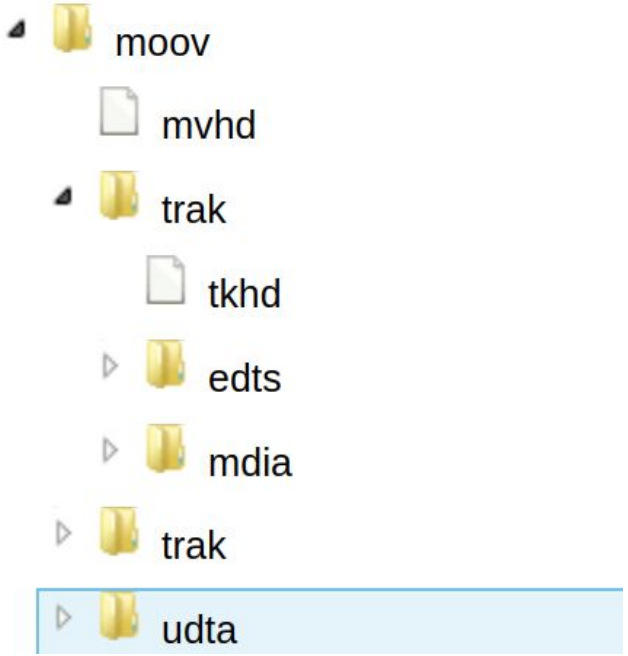
moov (Movie Box)

- Stores metadata for overall presentation
- Metadata includes timing and location information
- It is **Container box**
- Exactly **one** is present



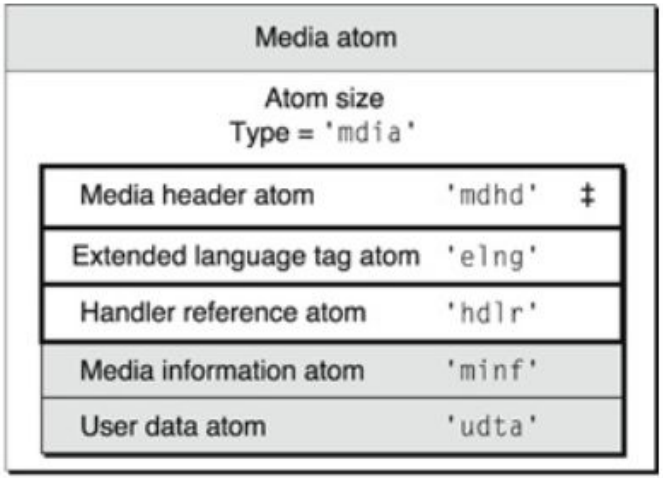
trak (Track Box)

- Inside **moov** atom (one or more can present)
- Contain metadata for a single track
- **Track types:** video, audio, subtitle, hint
- Each track is independent



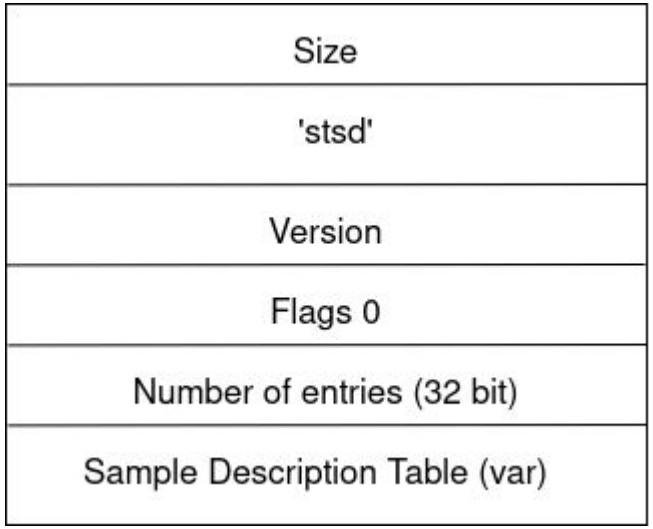
mdia (Media Box)

- Contains references to media data
- Describe and define a track’s media type and sample data
- Exactly one instance inside **trak** box



stsd (Sample Description Atom)

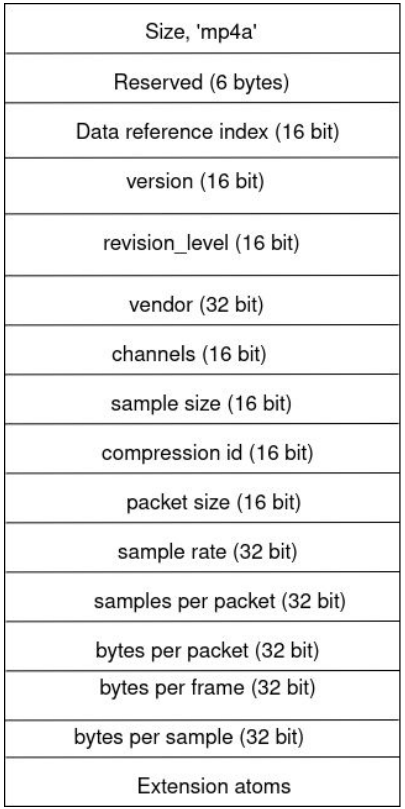
- Gives information about coding type used
- Stores information that allows you to **decode samples** in the media
- Information stored after ‘Number of entries’ is track type specific
- In case of video media sample descriptions are **image description structures**
- Sample description table contain array of sample descriptions



mp4v (sample description for video)

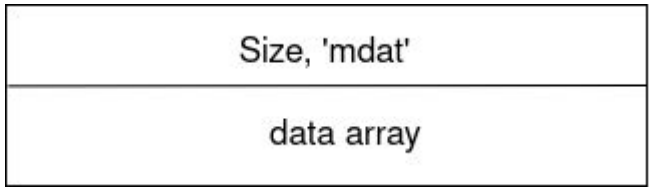
Size, 'avc'
Reserved (6 bytes)
Data reference index (16 bit)
version (16 bit)
revision_level (16 bit)
vendor (32 bit)
temporal quality (32 bit)
spatial quality (32 bit)
width (16 bit)
height (16 bit)
horizontal resolution (32 bit)
vertical resolution (32 bit)
data size (32 bit)
frame count (16 bit)
compressor (fixed size string)
Deapth (16 bit)
color table id (16 bit)
Extension atoms

mp4a (Sample description for audio)



mdat (Media Data Box)

- This box contain **media data**
- In **video track**, this box contain video frames
- In **audio track**, this box contains audio samples
- A presentation may contain zero or more mdat boxes
- The metadata refers to media data by its absolute offset within the file
- mdat can be in same file or other file



Fragmented MP4

- For streaming application, a movie is divided into smaller pieces called **fragments, segments or chunks**
- Instead of downloading full MP4 file, we want to download no more than a few seconds at a time, so that we can play and download at the same time.
- We want adaptive bitrate streaming for mp4, shorter the segment faster we can adapt

File structure for fmp4

- fmp4 stream consist of an initialization segment and sequence of media segment
- Initialization segment is same as unfragmented file
- **moov** now contain additional information to indicate stream is fragmented