MPEG-2:

The basics of how it works

HEWLETT PACKARD

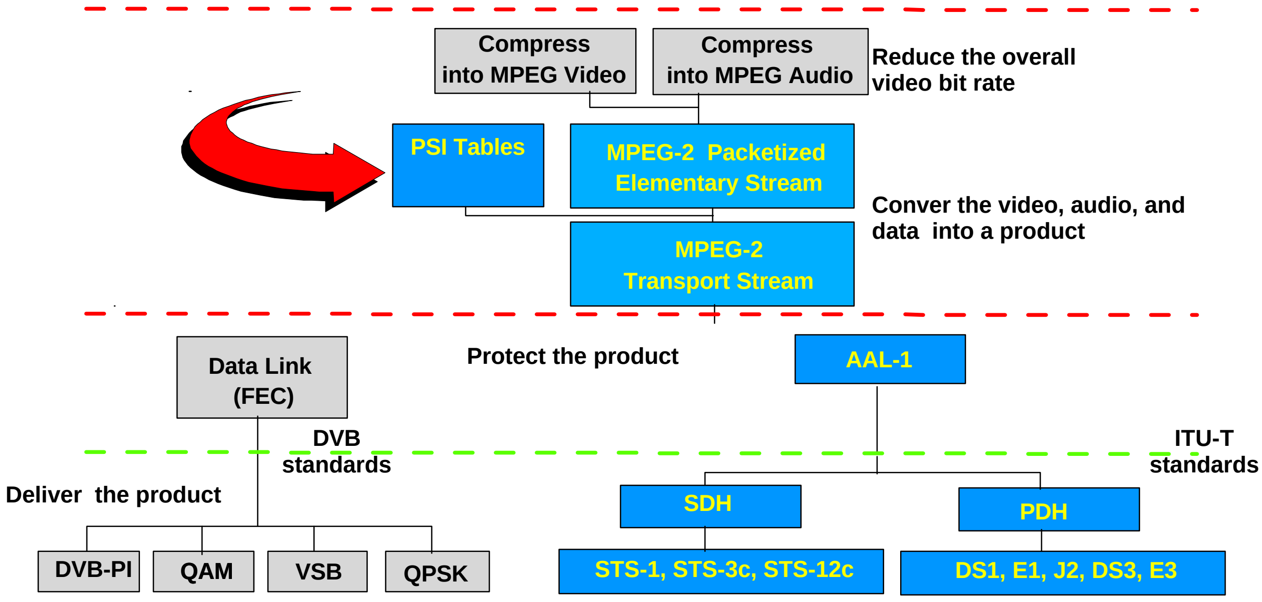
This slide & speaker notes based tutorial, covers the very basics of how MPEG-2 works. That is how a high bit rate digital video signal is taken from a studio camera. Compressed down to a low enough bit rate, so that it can make economic use of available transmission bandwidth. And converted into a form that a consumer set-top box can decode.

The tutorial covers the creation of the elementary stream, the packetised elementary stream, and the transport stream multiplex. It discusses the use of Programme Specific Information, and also the extensions created by the DVB, known as the DVB Service Information.

It also tries to give a feel for the fragility of the MPEG-2 transport stream. This is of major importance to service providers, wishing o guarantee quality of service.

MPEG-2 transport stream

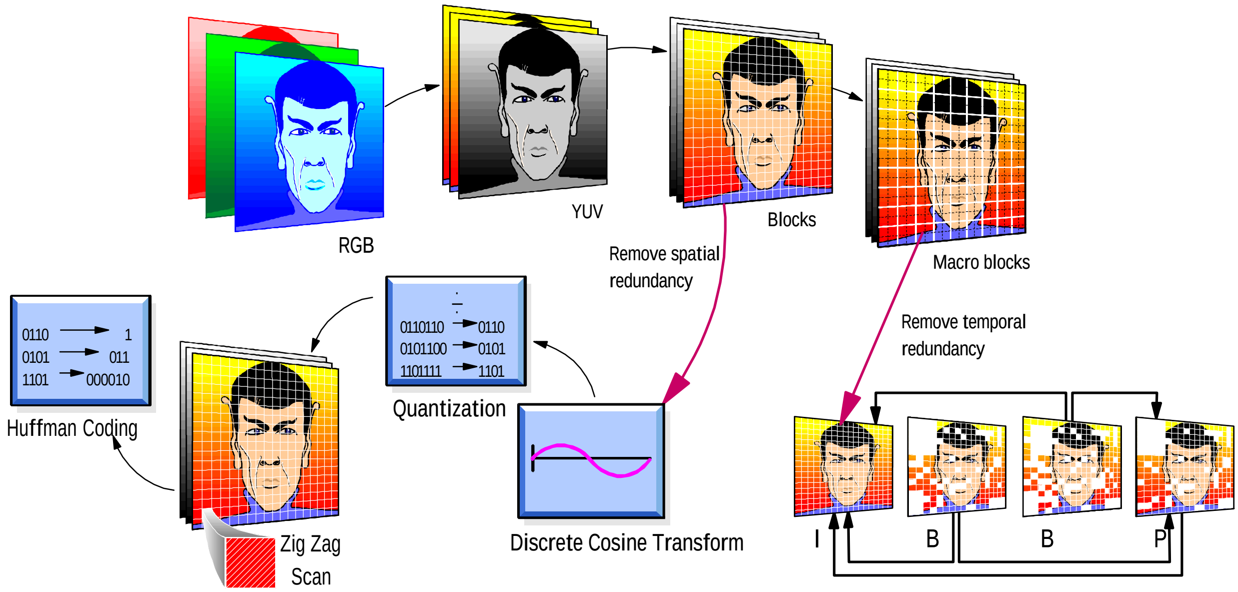
Protocol stack



- This slide shows the protocol levels that are used to transport the digital video signal, from the studio to the consumer.
- We are interested in the section between red dashed lines. Transport stream protection is covered in another tutorial.
- The protected transport stream is then fed into whichever physical channel is used to deliver the video signal to the consumer.
- Note this architecture is point to multi-point, constant bit rate, broadcast orientated: which is why only AAL-1 is shown, not AAL-5.

MPEG compression of video:

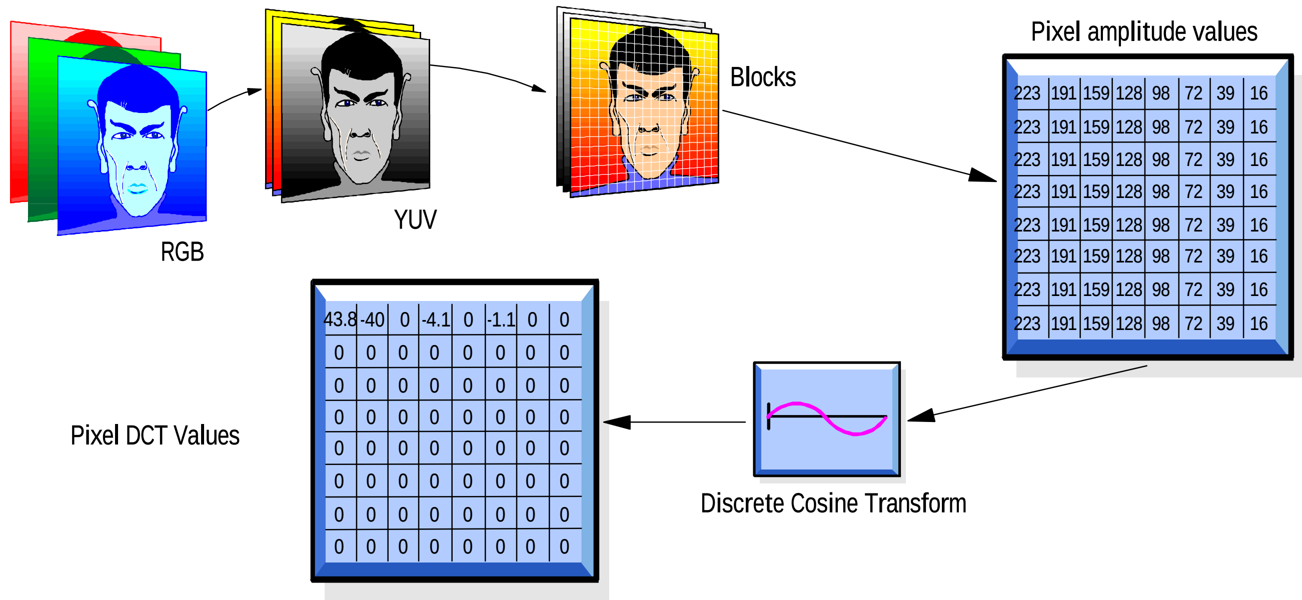
Spatial & temporal redundancy



- Video compression relies on the eye's inability to resolve High Frequency color changes, and the fact that theres a lot of redundancy within each frame and between frames.
- The Discrete Cosine Transform is used, along with quantization and Huffmann coding; to predict a pixel value from all adjacent pixel values, and minimize the overall bit rate.
- This generates the Intra-frames (I-frames).
- Prediction & motion compensation, predicts the value of pixels in a frame, from the information in adjacent frames.
- Audio compression makes use of the fact that, high power tones tend to blot out lower power adjacent tones. So if you can't hear it, don't transmit it.

Spatial redundancy:

Pixel coding using the DCT



Pixel amplitude values

223	191	159	128	98	72	39	16
223	191	159	128	98	72	39	16
223	191	159	128	98	72	39	16
223	191	159	128	98	72	39	16
223	191	159	128	98	72	39	16
223	191	159	128	98	72	39	16
223	191	159	128	98	72	39	16
223	191	159	128	98	72	39	16

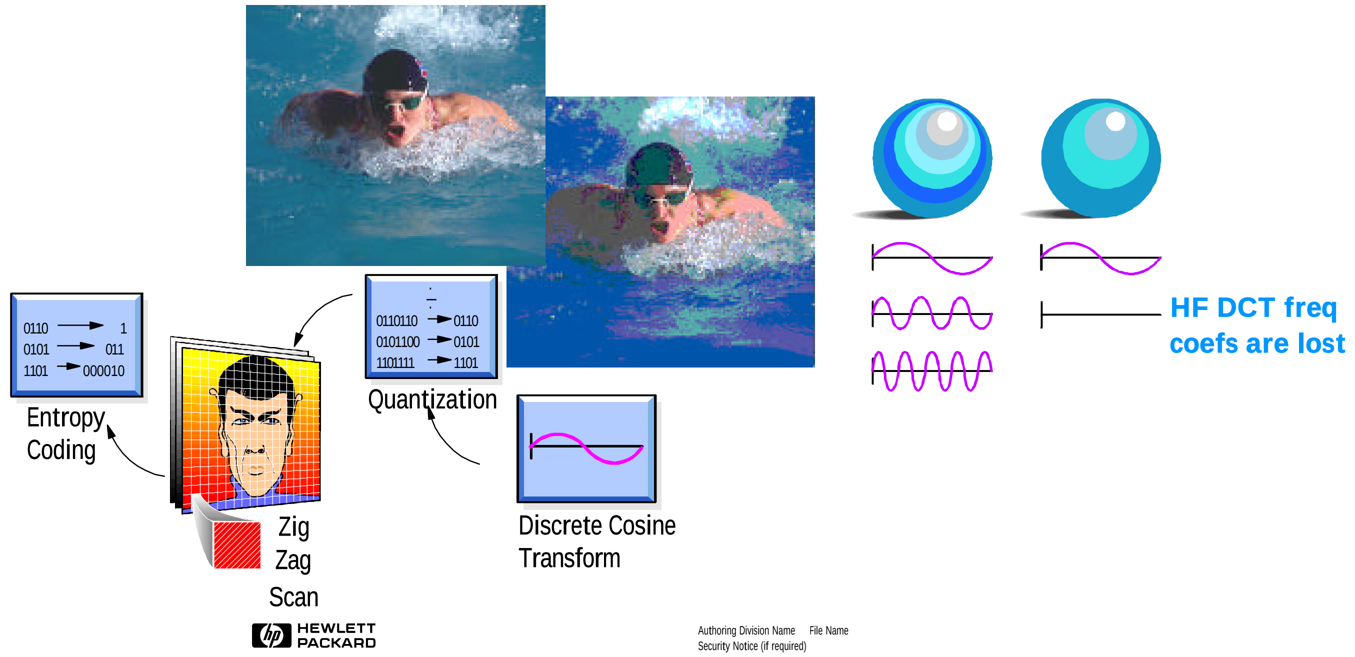
Pixel DCT Values

43.8	-40	0	-4.1	0	-1.1	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

- The first stage is to create an I-frame, subsequent frames in a group of frames will be predicted from this frame.
- As the eye is insensitive to HF color changes, we convert the R,G,B signal into a luminance (how bright the picture is) and two color difference signals. We can remove more U,V information than Y.
- Each pixels DCT is calculated from all other pixel values, so taking 8x8 blocks reduces the processing time.
- The top left pixel in a block is taken as the dc datum for the block.
- DCT's to the right of the datum are increasingly higher horizontal spatial freqs. DCT's below are higher vertical spatial frequencies.
- Using an Inverse DCT we could reconstruct each pixel's value in the 8x8 block. The DCT is a lossless and reversible process.
- Its the next stage which introduces compression.
- Note that the smaller the difference between one pixel and its adjacent pixels, the smaller its DCT value.
- In the example shown, a greyscales 8x8 pixel values are reduced to one row of DCT's. With all other values going to zero

Spatial redundancy:

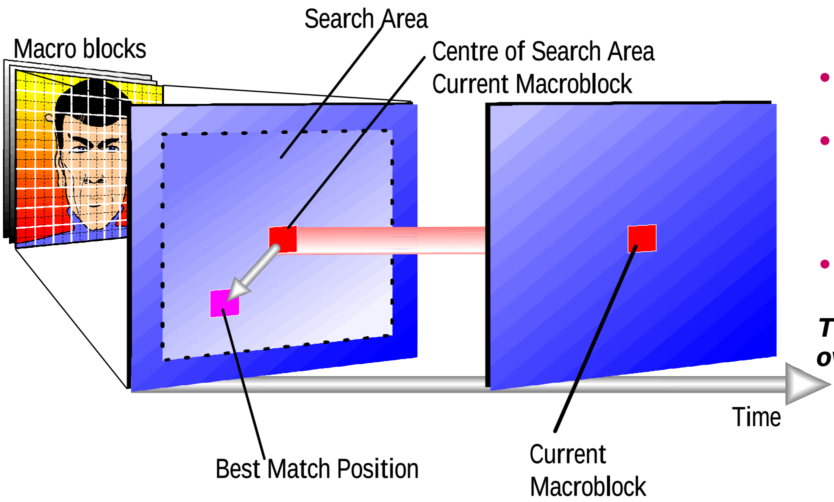
Quantization & Entropy coding



- The higher the DCT frequency, the higher the Quant Matrix value its divided by. This makes many coeficientss go to zero.
- The fixed value scale factor reduces even more of the DCT's to zero.
- The next stage is to increase the number of zero's in the run of bits into the entropy coder. This is done by zig-zag scanning the 8x8 pixel block DCT values and helps the entropy coder do its job.
- Entropy coding essentially sizes coeficients by how often they occur.
- The more a coeficient occurs, the smaller a binary value its given.
- Since in any frame your going to get a large number of identical 8x8 blocks, your reducing the overall binary data rate.
- To summarize then, quantization makes many higher frequency DCT values go to zero. Entropy coding removes duplication of DCT's, assigning each DCT position with a pointer to its value.
- **This all has a cost. Thats shown in the pictures above: the upper picture is unquantized, the lower one quantized.**

Temporal redundancy:

Inter-frame prediction & motion estimation



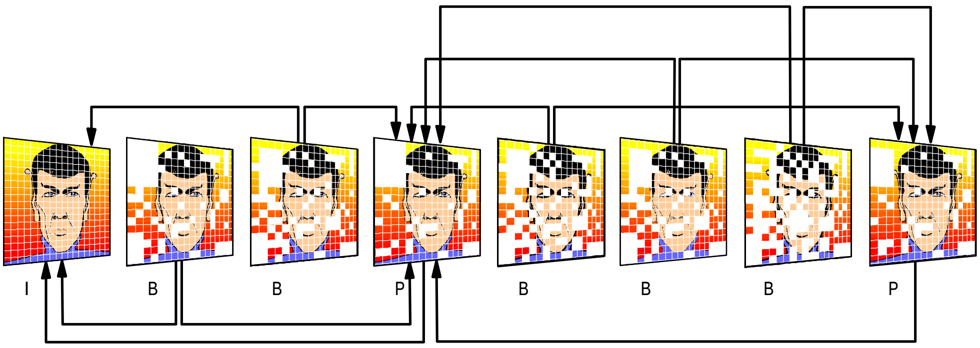
- **If adjacent frame macroblocks same, don't retransmit block.**
- **Search to see if same block exists somewhere else: if it does, just transmit its coordinates (motion vectors).**
- **Only Intra-code macroblocks which are completely new.**

This process really drops the overall bit rate

- This is were the real bit rate reduction kicks in. As we'll cover in the next slide, there are three different frame types.
- By just doing spatial redundancy on a frame you create an I frame. This has all the information necessary to decode the picture.
- The next stage is to look at the next frame to this and see how similar it is. You can do three things to minimize this frames bit rate.
- Firstly, look to see if the macroblock in the same position in the next frame hasn't change. If it hasn't, don't do any coding, Just transmit that its the same.
- The next stage is to search around in the I-frame and see if this macro-block exists, but its in a different place. If so transmit motion vectors for its old location.
- Only if its completely new, do you go for the complete intra-coding process.
- **This really reduces the overall bit rate from frame to frame.**
- **But note if you kept predicting each frame from the last, it would only take a little error, and the whole process would fast start to unravel.**
- **Thats why there are three different frame types, and a specific frame transmit process.**

Putting it all together

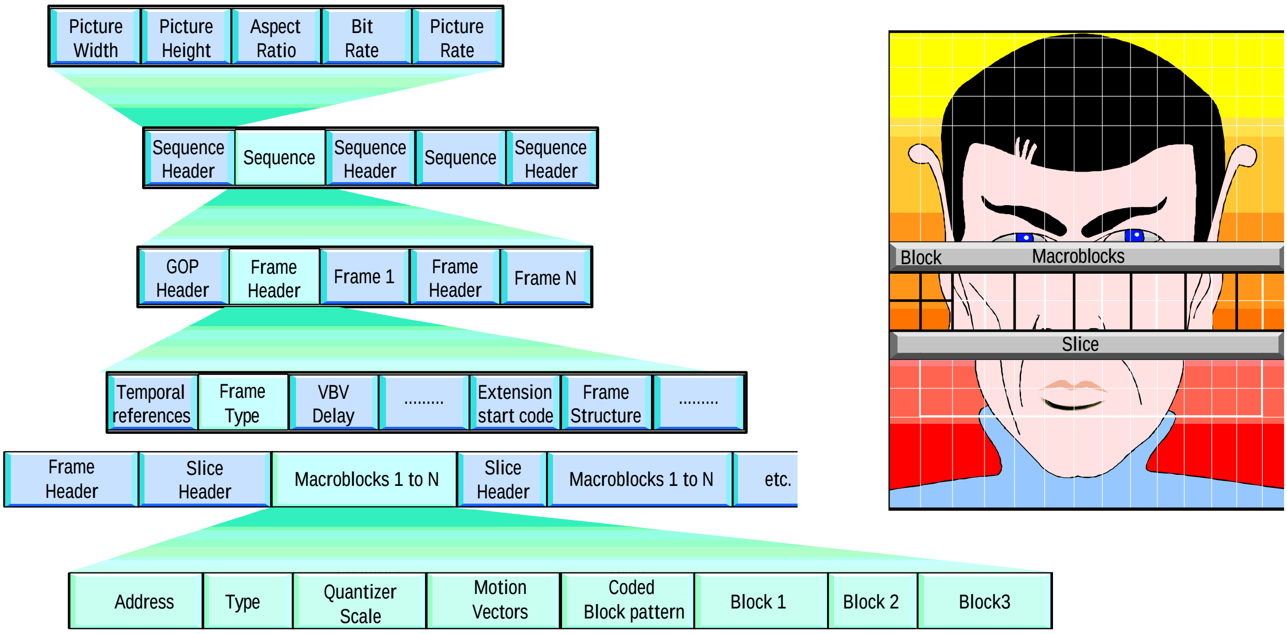
I, P, B Frames & avoiding "MPEG heat death"



- **I-frames: contain full picture information**
- **P-frames: predicted from past I or P frames**
- **B-frames: use past and future I or P frames**
- **Transmit I frames every 12 frames or so.**

- The Intra Frames contain full picture information. These are your lifeline, if errors occur, or the decoder loses a frame. Without periodic transmission of these the whole process falls apart. But the I-frames are the least compressed.
- Predicted (P) Frames are predicted from past I, or P frames,
- Bi-directional predicted frames offer the greatest compression and use past and future I & P frames for motion compensation. But they are the most sensitive to errors.
- The encoder will cycle through each frame and decide whether to do I, P, or B coding. The order will depend on the application. But roughly every twelve frames, an I-frame is created.
- If the encoder didn't do this, any small errors would build up and the MPEG compressor would rapidly descend into an electronic form of Entropic "heat death".
- The process detailed in the last few slides does the real work. But a decoder needs additional information to reconstruct the frames.

Building the Elementary Stream



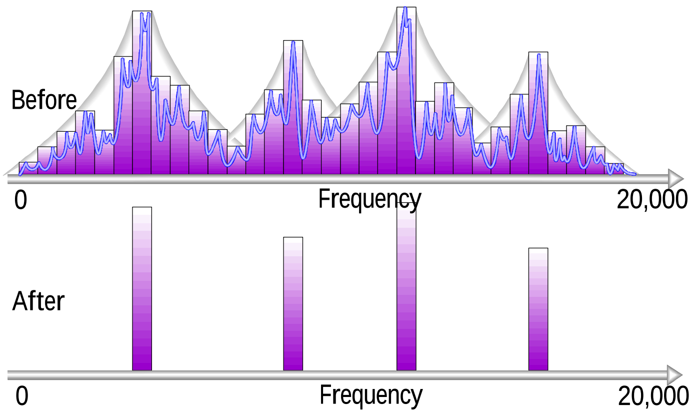
HEWLETT PACKARD

Authoring Division Name File Name
Security Notice (if required)

- This slide shows how the actual blocks, slices, frames etc. are all put together to form the elementary stream.
- Along with the actual picture data, header information is required to reconstruct the I, B, P frames. This header structure is shown.
- Each slice will contain a header detailing its contents & location.
- Each frame will have a header, and each group of I, B, P frames, known as a Group Of Pictures (GOP) will have a header.
- The next stage is to take this ES and convert it into something that can be transmitted and decoded at the other end.
- At this stage, the elementary stream is a continual stream of encoded video frames. Though all the data required to reconstuct frames exists here. No timing information or systems data is contained
- Thats the job of the MPEG-2 multiplexer
- First a few words on what we do with the audio signal associated with the video

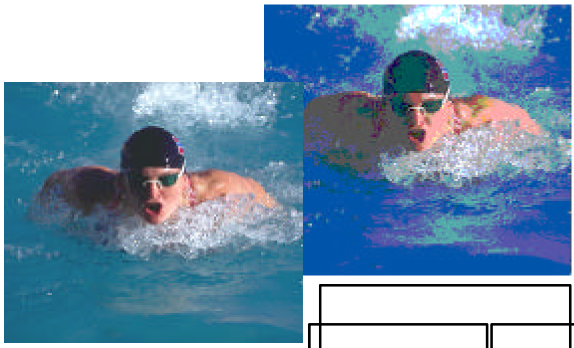
Sub-Band Masking

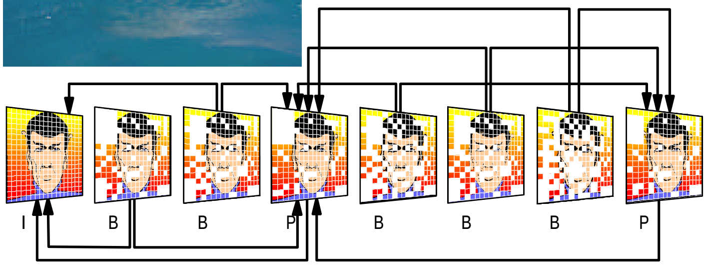
- System relies on masking of tones either side of hi-power tone.
- If you can't hear it, don't code it.
- Sub-band masking takes 3 MBit/s & reduces it to about 200 Kbit/s.
- Two types of audio coding: exists MPEG-2 & Dolby AC3
- US use AC3, DVB uses MPEG-2.



MPEG compression: summary

Reduced bit rate, increased error susceptibility



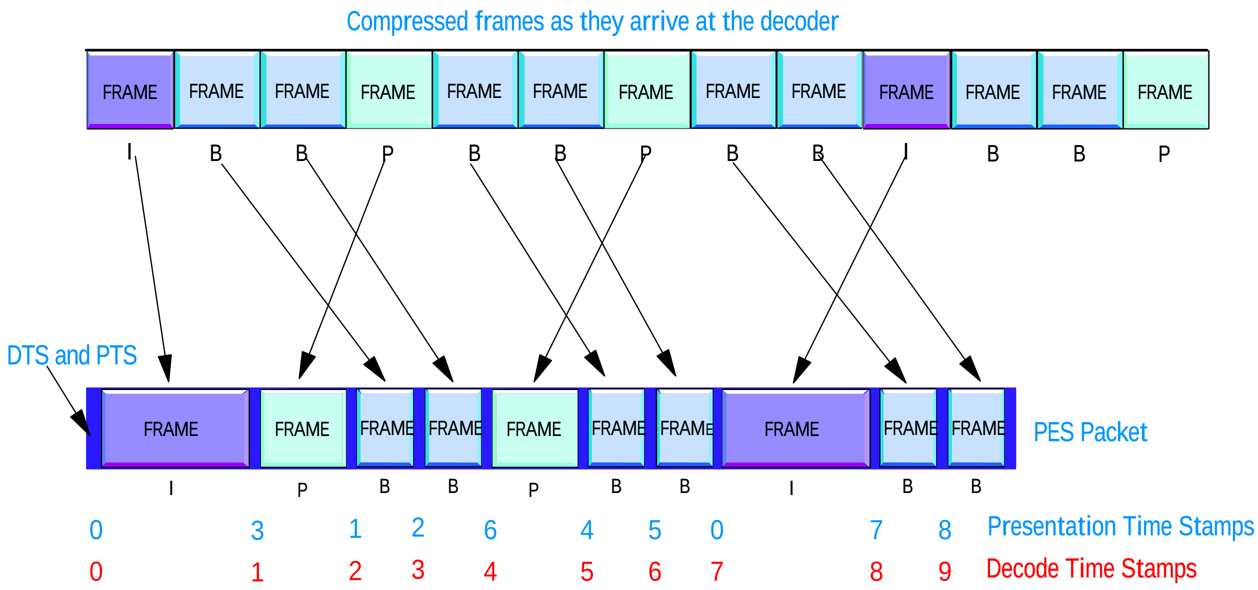


But how do we know what channel we've compressed?

- This process gets rid of a huge amount of redundant information.
- But that means we've lost information we could use if errors occur.
- The compression process may make digital tv economically viable, but its already caused a problem for the decoder, if errors are injected into the data: Quantization & motion estimation ensure, these errors quickly propagate.
- Unfortunately it gets worse.
- There's no point doing all this compression alone. Any decoder has to know what information you've compressed, what channel it is, is there any special data contained. And a lot of other housekeeping information.
- **Thats why you need a systems layer. Thats what we'll describe next.**
- **The QoS problem is already getting big due to compression. But at least at this level you only lose some frames. At the systems level, you will lose many frames if errors occur.**

Ordering frames for decoding:

The PTS & DTS.

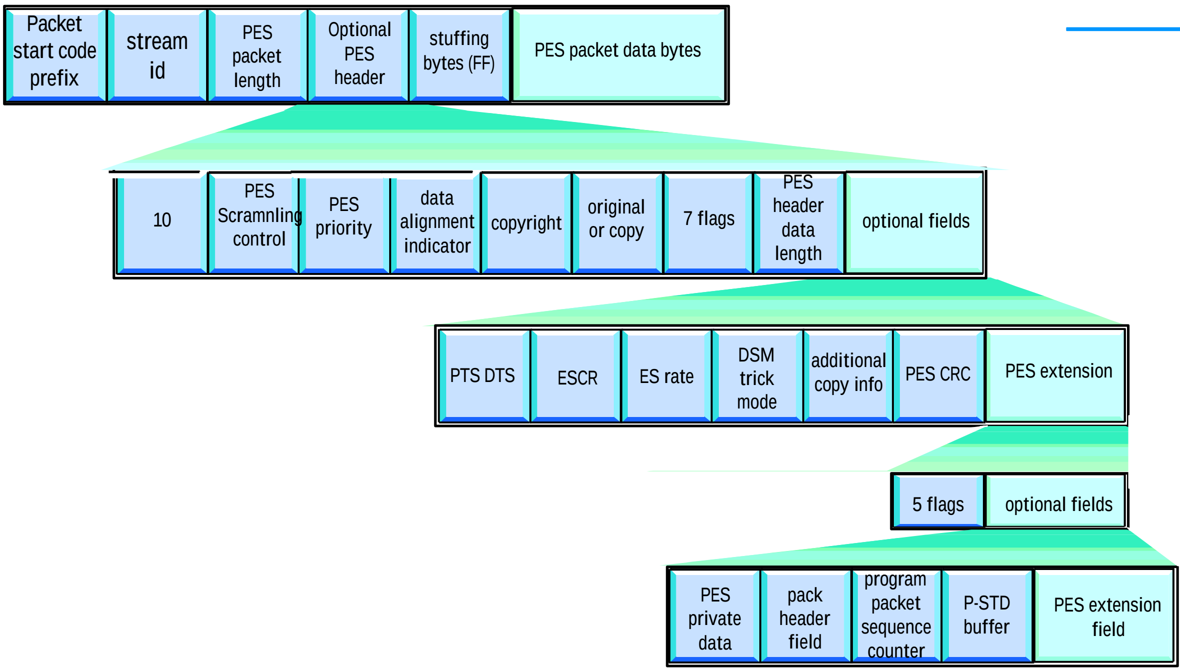


HEWLETT PACKARD

Authoring Division Name File Name
Security Notice (if required)

- In order for a decoder to reconstruct a B-frame from the preceding I and following P frames, both these must arrive first.
- So the order of frame transmission must be different to the order they appear on the tv screen. As shown in the slide above.
- But for this to work, the decoder must also know at what time it should show the frames. That its their order in time.
- The Decode Time Stamp tells the decoder when to decode the frame.
- The Presentation Time Stamp tells the decoder when to display the frame.
- In addition to knowing at what time decode and presentation should occur, a clock must be embedded, to allow a time reference to be created.
- The PTS and DTS are added to the Packetised Elementary Stream, whilst the clock, known as the Programme Clock Reference (PCR) is contained in the Transport stream.
- The PCR will be discussed shortly.

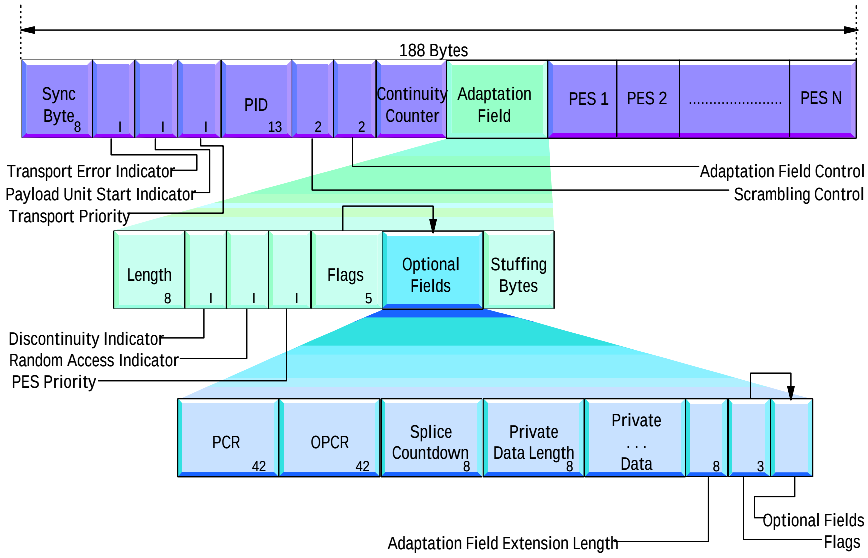
The Packetised Elementary Stream



- This slide shows the PES structure.
- The PES header contains information about the content of the PES packet data bytes. Allowing a decoder to make sense of the packets.
- The PES packets can be of variable length, typically upto 64 kbytes, but they can be longer.
- One of the most important parts of the structure, are the PTS and DTS, these allow the decoder to reconstruct the video stream from the I, B, P frames sent by the encoder.
- The PTS & DTS process was discussed in the previous slide.
- The important thing to note is that if information carried in the header is corrupted, the entire PES packet will be lost.
- As seen, the higher up the MPEG protocol stack we go, the greater the potential damage that errors will cause.
- The next stage is to multiplex many PES together to create a stream of many tv programme events.
- Since at the moment, this is simply a raw, but decodable single elementary stream: without associated audio, or data. And no infomation which tells the consumer what it contains.

MPEG-2 Transport Stream

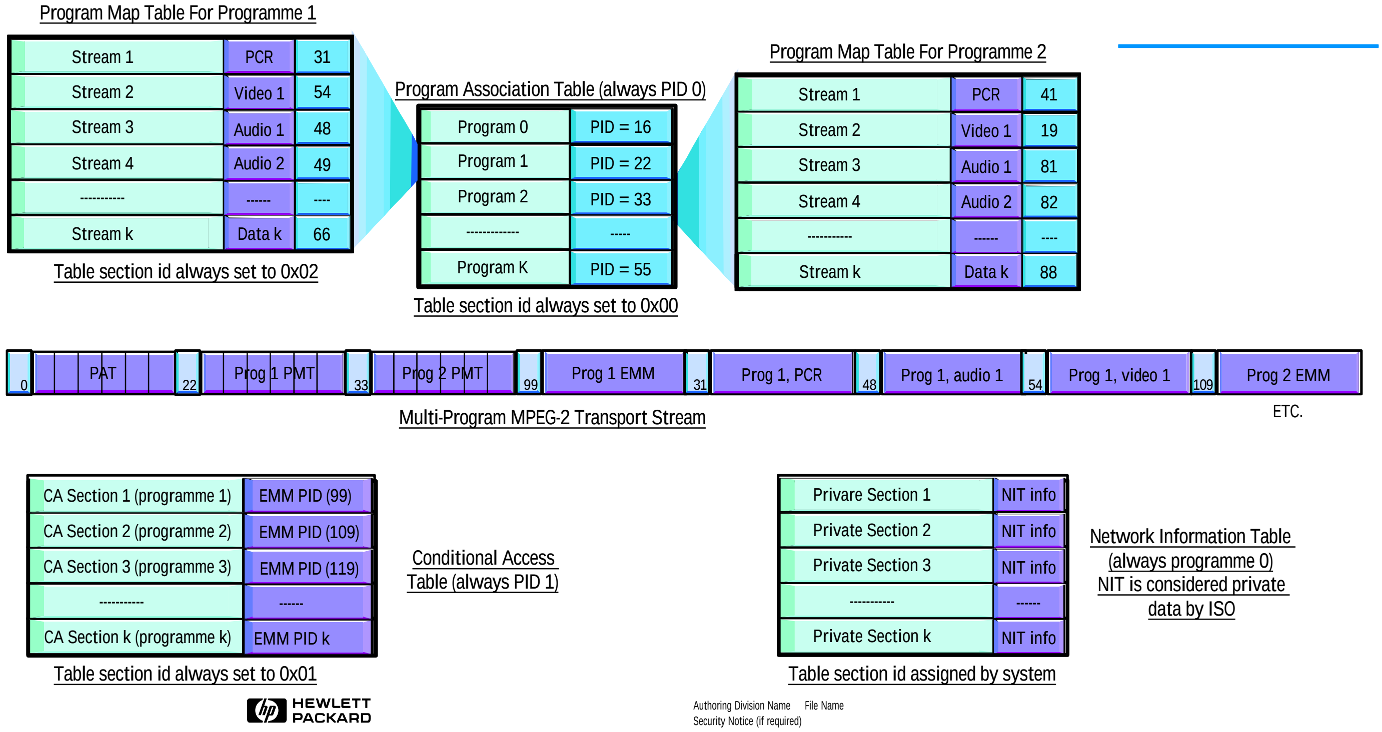
multiplexing many programs



HEWLETT PACKARD

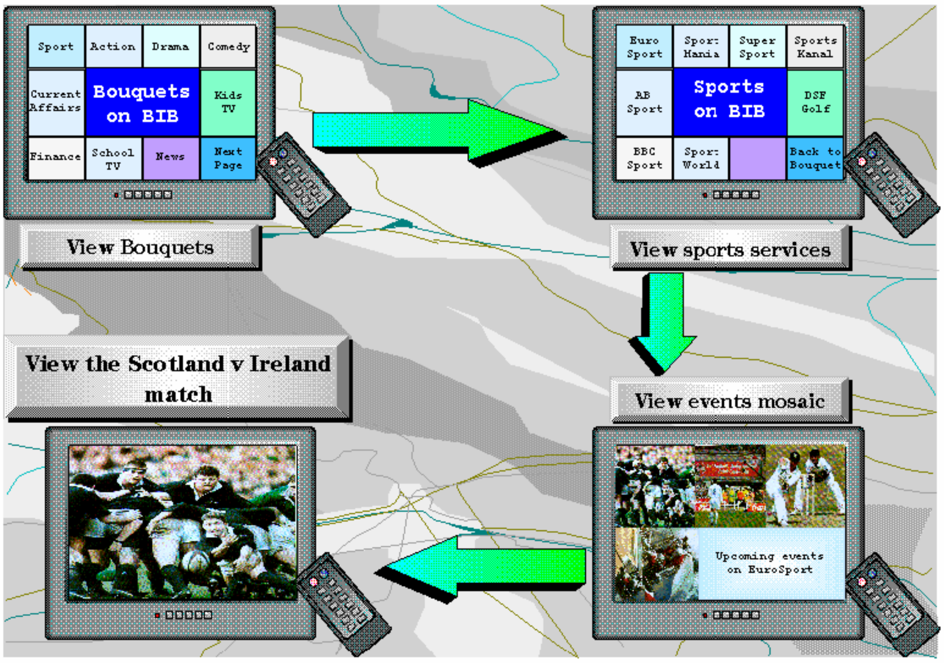
- This is a blowup of the Transport Stream packet structure
- The TS takes the variable length PES, and chops it up into fixed length packets: this is required for most transmission systems.
- Note this information is on top of the syntax information contained in the PES header. Perhaps you begin to get a feel for the error potential here. The TS allows the multiplexing of many PES.
- The key features are as follows:
- Sync Byte, sets the start of a TS packet and allows transmission synchronization.
- Transport Error indicator: indicates the packet is errored (block error testing)
- PID: Packet IDentifier, is the channel identifier It contains all the navigation information required to find, identify and reconstruct programmes. PID values are contained in the PSI tables.
- PCR, the programme clock reference: provides 27 MHz clock recovery information.
- **How does the decoder know which packets containf the PES, which makes up the tv event to decode?**

PSI Table structure



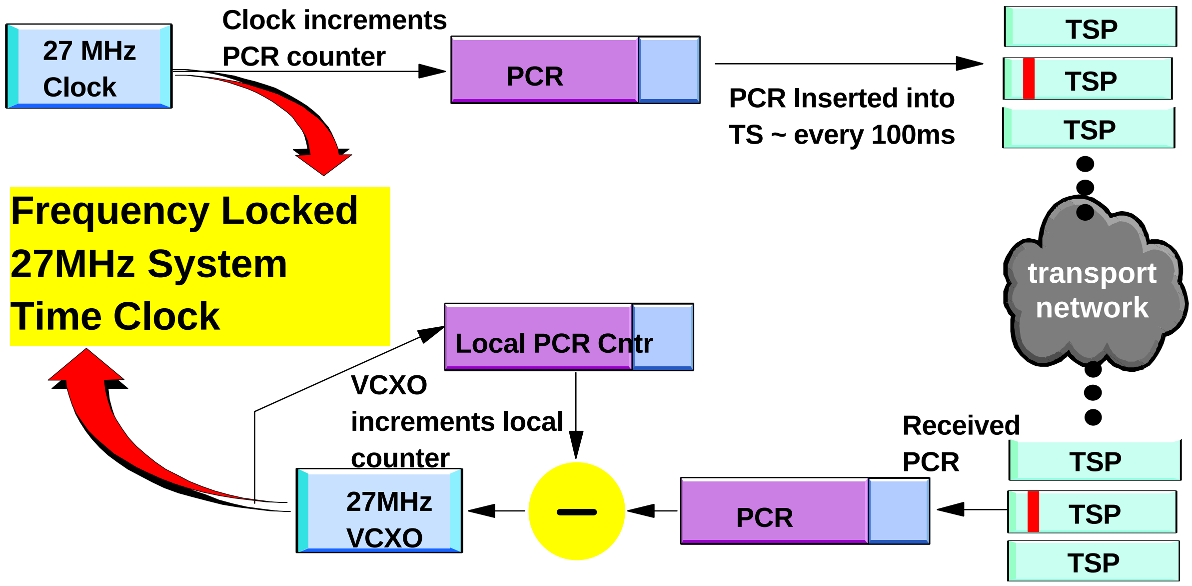
- To reconstruct the PES, the PSI uses a series of identifiers known as the Programme Identifiers, or PID's.
- Once the programme to be decoded is known, the decoder searches for PID=0 - the Programme Association Table PID.
- The PAT contains the PID's of all the PMT's for the programmes contained. We assume programme one has been chosen for decode.
- The PMT for programme 1 is identified via its PID (22), extracted from the transport stream packets containing it, and decoded.
- Prog 1's PMT contains all the PID's for Prog1's video, audio and data packets. These must be put together to reconstruct the PES.
- Prog 1's timing info, required for decode, is contained in a transport packet, identified by the PCR PID (31). Each prog has a PCR.
- PID zero is always used to identify the CAT. This is needed to find out whether the consumer is allowed to decode and view prog 1.
- The CAT contains all the PID's identifying the EMM's for all progs.
- The NIT, contains information about the user-selected service. Such as channel freq's transponder numbers etc. The NIT is always associated with prog 0's PID. See DVB-SI for more about the NIT.

DVB-SI: program guides



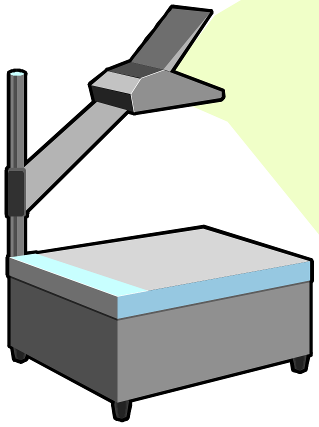
- WIthin the PSI structure, is a table known as a private table. This was created by MPEG, so that service providers could create their own extentions to the MPEG-2 PSI.
- This has been used by the DVB and the ATSC, to create what are called Service Information tables. We discuss the DVB-SI only.
- These are used to carry information about what tv events are contained in the transport stream: via the Event Information Table.
- What services are being carried, via the Service Desciption Table.
- What groups of services, with common themes exist: via the Bouquet Association Table.
- And what are the physical parameters of the network carrying a transport stream: via the Network Information Table.
- The tables are highly complex constructs. And like the PSI, are protected by a CRC-32.
- The SI are used to create Electronic Programme Guides (EPG's). These help the consumer find the programme they want to watch.
- The example shows the process by which a user goes from the display of bouquet's all the way to viewing a Rugby match.

MPEG Technology
Program Clock Reference - PCR



- PCR is a clock recovery mechanism for MPEG programs. When a program is encoded, a 27MHz System Time Clock (STC) drives the encoding process. When the program is decoded (or remultiplexed), the decoding process must be driven by a clock which is locked to the encoder's STC. The decoder uses the PCR to regenerate a local 27MHz clock.
- When a program is inserted into the Transport Stream a 27MHz timestamp is inserted - the PCR. At the decoder end, it uses a Voltage Controlled Oscillator (VCXO) to generate a 27MHz clock. When a PCR is received, it is compared to a local counter which is driven by the VCXO, and the difference is used to correct the frequency of the VCXO to ensure that the 27MHz clock is locked to the PCR.
- The PCR field is a 42 bit field in the adaptation field of the Transport Stream. The PCR field consists of a 9 bit part that increments at a 27MHz rate and a 33 bit part that increments at a 90kHz rate (when the 27MHz part rolls over).

Summary: *"the MPEG-2 transport stream is efficient, but fragile"*



- Corrupt ES headers lose frames
- Corrupt PES header, lose whole ES
- TS carries many PES & own hdr (4byte)
- Hdr contains info on navigation thro PES (PID &PSI). Corrupt this, lose many PES
- Corrupt SI & EPG: lose pay-per view
- PCR provides timing clock. PCR errors means you lose timing & PTS/DTS.

"theres no margin for error, the TS needs protection"

- The elementary stream is taken and formed into a Packetised Elementary Stream. This has a header which allows ES decode.
- The TS carries many PES, along with the PSI, SI and a header.
- If the TS header is corrupted, then many PES will be lost.
- If the PSI, or SI are corrupted, then it may be impossible for the consumer to find out what is contained in the transport stream.
- To conclude, though MPEG-2 allows the economic transmission of many different tv events, along with associated service information, it does so by having very little margin for error.
- Due to this, the protection of the transport stream, and the ability to verify its integrity is vital.
- The problem with tv broadcast is any errors are glaringly obvious, and directly affect the consumer. This means the potential for service provider conflict is high.