Video Compression Overview

Naser Refaat

Sony BPE

0
101
00110
11010001
11100
001
110
01

Presentation Outline:

Seven Sections

1. **Review of Concepts Basic to Compression**
2. **Definition of Compression**
3. **Stages of Compression**
4. **The Spatial Compression Process**
5. **Why Develop MPEG-2?**
6. **Motion Compensation and Prediction**
7. **MPEG-2 "Formats," Applications, & *Profiles and Levels***

Section 1.
Review of Video Concepts for Compression

Topics

1. **Analog “compression”**
2. **Sampling and Quantization**
3. **ITU-BT-R 601 (a.k.a. CCIR-601)**
4. **4:2:2 Color Space**

Video Compression in Analog

Imagery

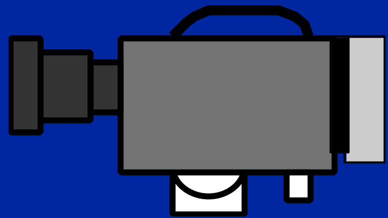


Image Capture

Interlace - 50%

R, G, B - Up to 8 MHz

Y, Pb, Pr

Y, I, Q

Y < 4 MHz

I = - 70% of Y

Q = - 85% of Y

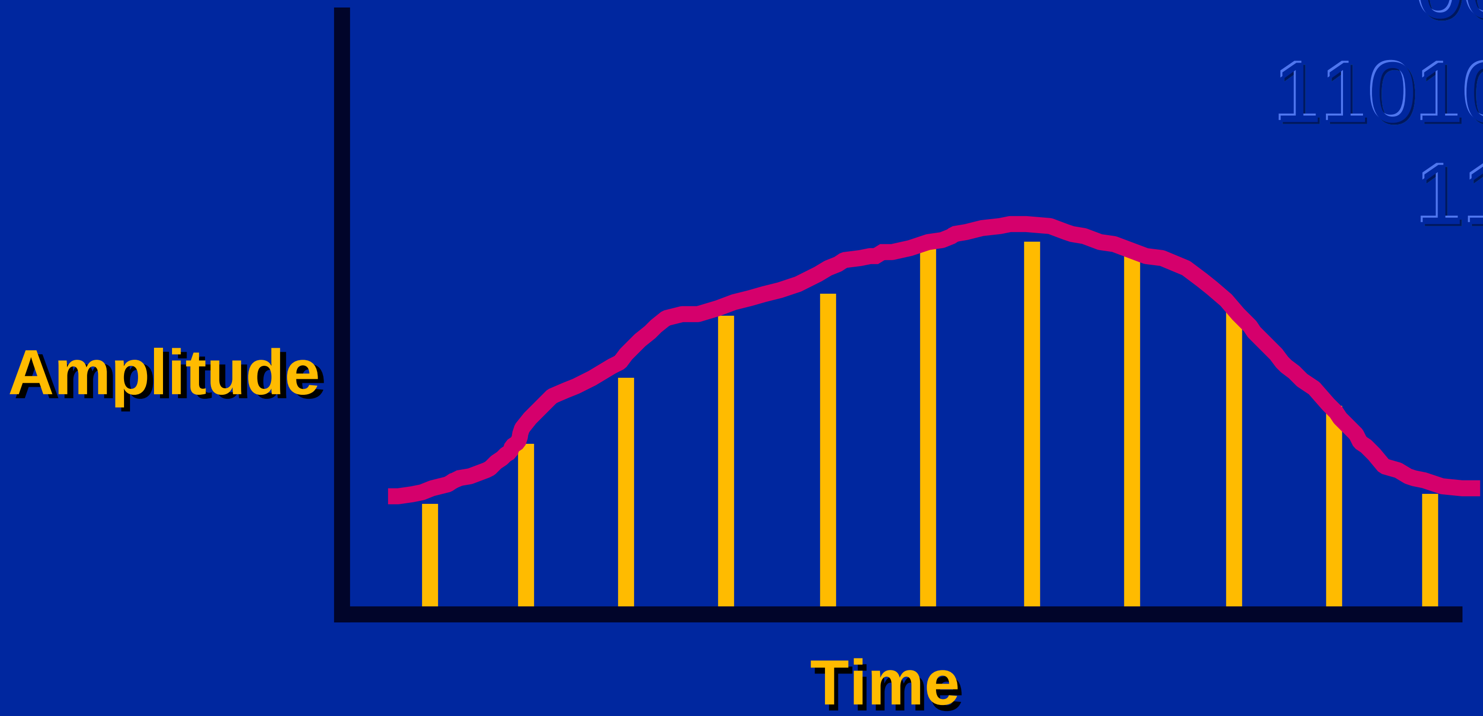
NTSC

Transmission

Display

Review of Digital Basics:

Sampling



0

101

00110

11010001

11100

001

110

01

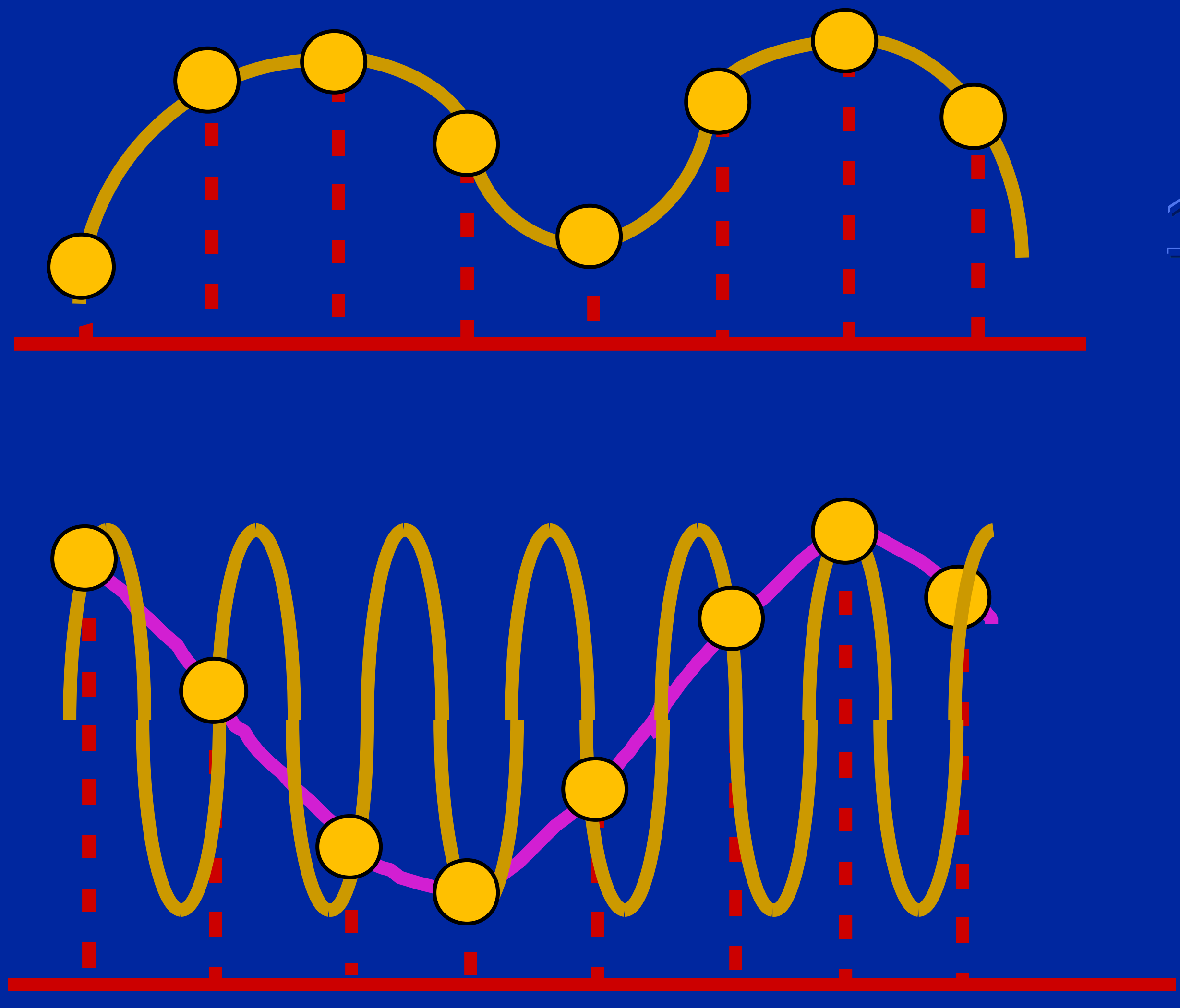
Review of Digital Basics:

Nyquist Criteria

Minimum sampling rate = twice the highest meaningful frequency of the sampled signal

Review of Digital Basics:

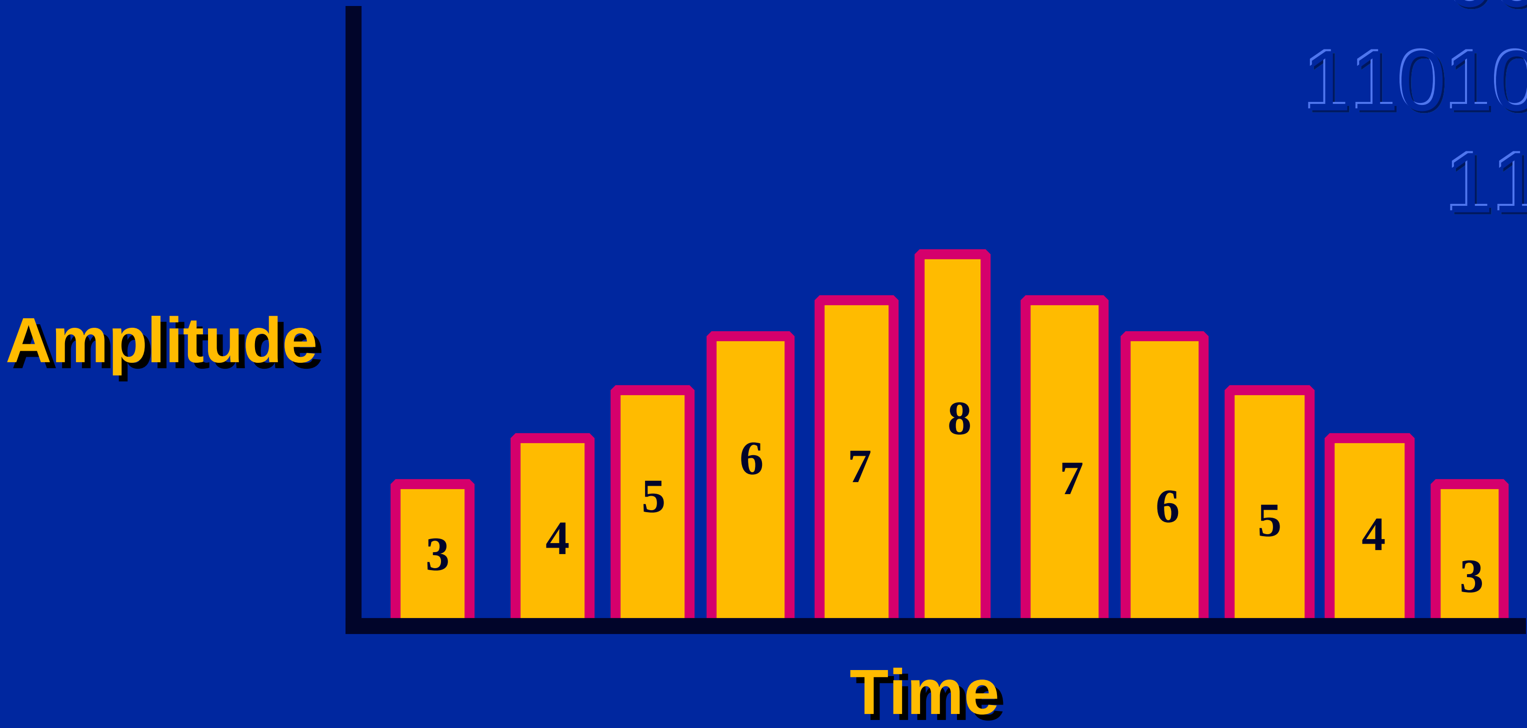
Aliasing



0
101
00110
11010001
11100
001
110
01

Review of Digital Basics:

Quantization



0
101
00110
11010001
11100
001
110
01

Review of Digital Basics:

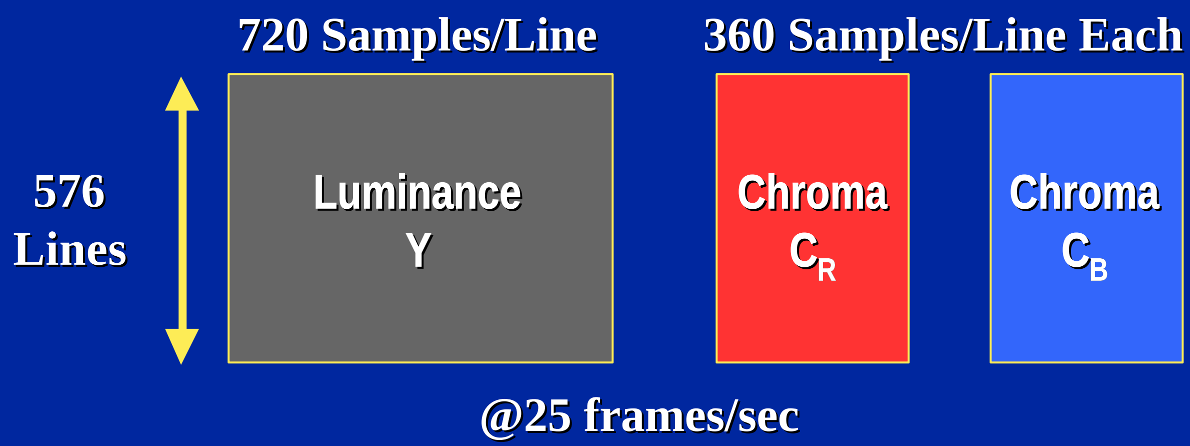
ITU BT-R 601

- International sampling standard for component video
- 4:2:2 sampling structure
  - 13.5 MHz for Y
  - 6.75 MHz for R-Y and B-Y

0
101
00110
11010001
11100
001
110
01

4:2:2 Sampling

Active Picture Bit Rate Reduction (BRR)



8-bit = 165.888 Mbps

10-bit = 207.36 Mbps

Section 2.

Definition of Compression

Topics:

1. What is compression?/Why compress video?
2. Broad Categories of Compression
3. Why is compression possible?
   - Redundancy
   - Limitations of the Human Visual system
4. Spatial Redundancy; Temporal Redundancy
   - New Types of Frames.

0
101
00110
11010001
11100
001
110
01

What is Compression?

- Reduction in the number of bits used to represent a item of data.

- Identification and removal of (visually) redundant information from a collection of data words.

Why Compress Video?

- Digital video sequences are the most demanding form of data
- Uncompressed video impose:
  - Large bandwidth requirements for transmission;
  - Enormous demands on the storage capacity of media.

Storage Requirements

- One hour of 8 bit digital video for
  - 4:2:2 digital component (216 Mbps) requires: 97 GB
  - Digital NTSC (128 Mbps) requires: 58 GB
  - 1125/60 studio HDTV (1.2) Gbps requires: 540 GB

Bit Rates of Common Digital Channels and Interfaces

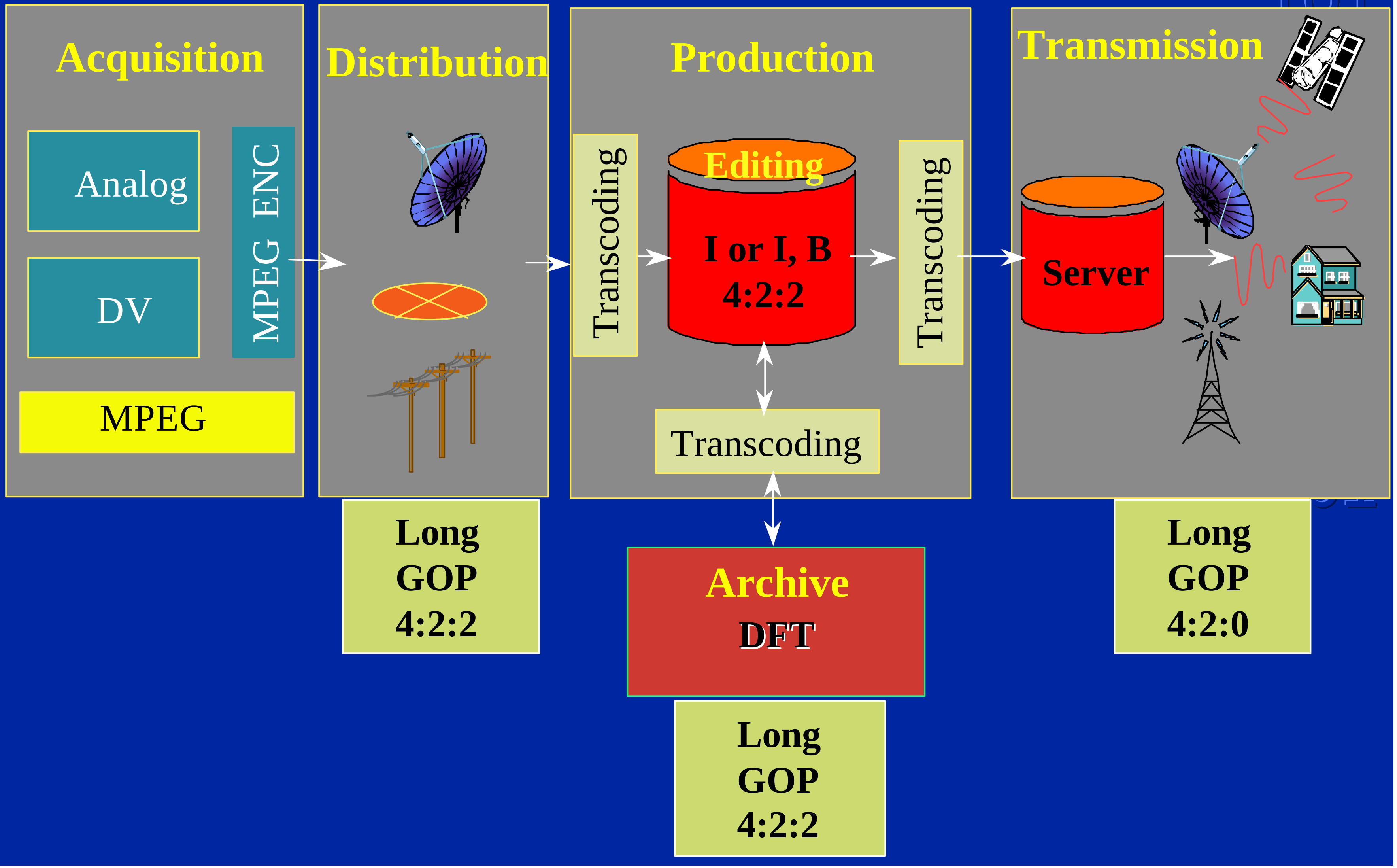
ISDN	144 Kbps
DS-1 (T1)	1.544 Mbps
DS-3 (T3)	44.7 Mbps
RF 6 MHz VHF/UHF channels	19 Mbps (8VSB)
Video Conferencing Lines	56 Kbps to 1.5 Mbps

Uses of Compression in Broadcast

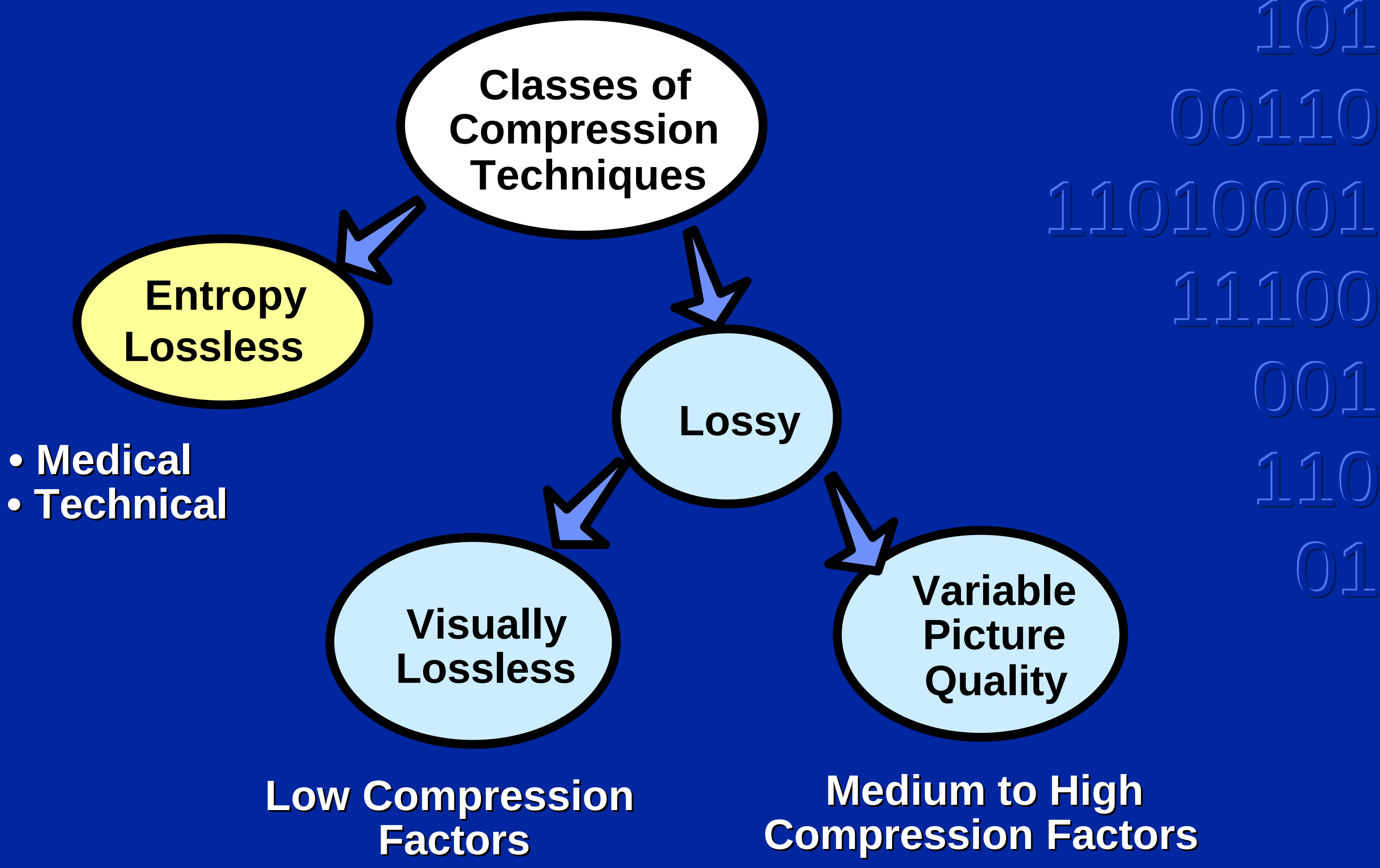
- **Acquisition: Camcorders, Satellites, Telco Networks, Archives, Film.***
  - **No unified Format**
- **Production: Storage and Archives, Film, Non-linear Editing.***
  - **No unified Format**

  **** Analog or Digital, any compression.***
- **Distribution: MPEG-2 MP@ML, 422P@ML (4:2:0 and 4:2:2)**
- **Transmission: NTSC and MPEG-2 MP@ML (4:2:0)**

Broadcast Compression Chain



Broad Categories of Compression Techniques



Entropy Lossless Compression

- The data after compression and decompression is numerically identical to original data.
- It allows only small compression ratios
  - approximately 1.5:1 to 4:1
- With video and audio, this compression ratio highly dependent on:
  - Sampling detail or resolution,
  - Signal noise,
  - Quantisation level.

Visually Lossless Compression

- The decoded image contains errors which are small enough to be classified as invisible.
- Allows for higher compression ratios.
- Errors (artifacts) will depend on the image itself.
- Visually lossless means no *visible* degradations.

0
101
00110
11010001
11100
001
110
01

Lossy Compression

- The decoded image contains errors relative to original image.
- It allows for higher compression ratios.
- Errors (artifacts) may or may not be visually apparent (visually lossless) dependent on the complexity of the visual content.

0
101
00110
11010001
11100
001
110
01

“Quality” and Compression

- Compression quality can only be judged subjectively.
- Use no more than the smallest compression ratio suitable for any particular application.
- Quality varies widely with source material.

0
101
00110
11010001
11100
001
110
01

Why Images Can Be Compressed

- **Images typically contain large amounts of redundant information.**
- **There are two types of redundancy:**
  - **Spatial—(similarities) between neighboring pixels;**
  - **Temporal—between neighboring frames in a video sequence.**

Why Images Can Be Compressed (2)

Compression eliminates redundancy

- between pixels
- between lines
- between frames

0
101
00110
11010001
11100
001
110
01

Limitations of the HVS
(Human Visual System)

- **Spatial properties**
  - **slight changes (errors) in luma are invisible**
  - **larger errors invisible at high luma levels**
  - **high frequency errors are less visible**
  - **distortions are less visible around sharp edges**
  - **limitations of ability to see detail in the areas where saturated colors exist.**

Limitations of the HVS (2)

Temporal Properties

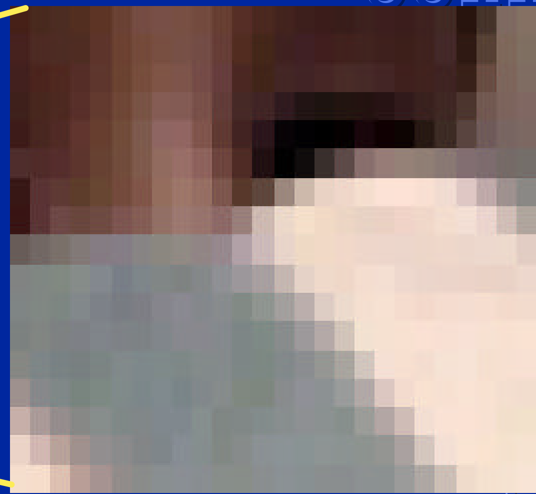
- If movement is drastic (after a scene change), the perceived spatial resolution is reduced immediately after the change
- If moving objects are not tracked by eyes, the loss of perceived spatial resolution is significant

Redundancy:
Impact on the Television Frame

- **Intraframe**
  - **spatial redundancy**
- **Interframe**
  - **temporal redundancy**

0
101
00110
11010001
11100
001
110
01

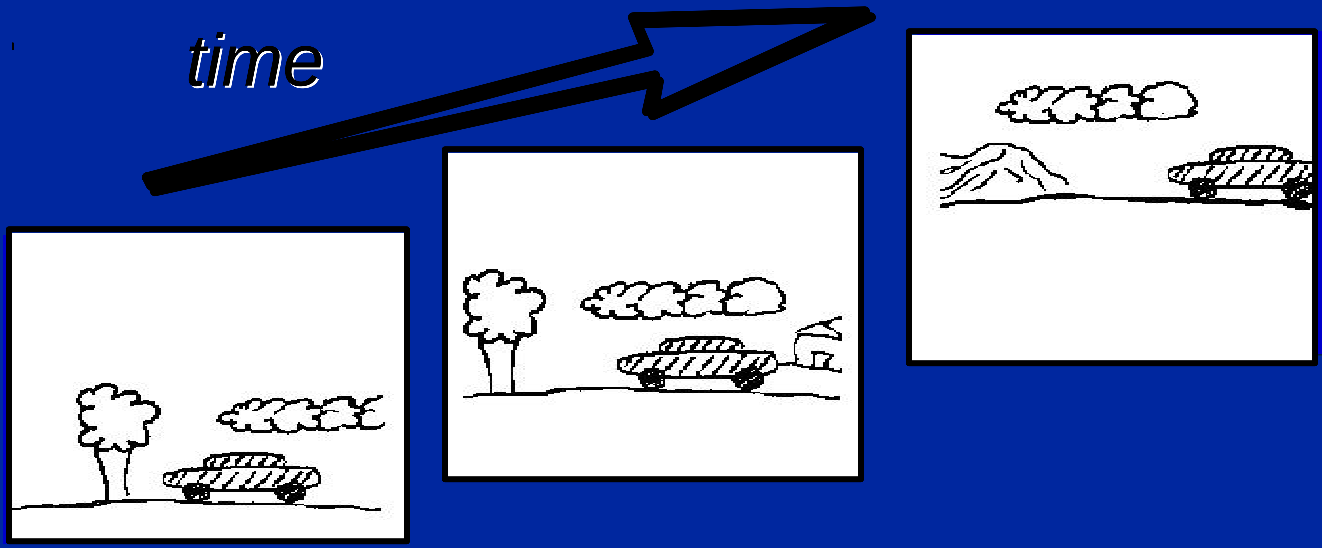
Spatial Redundancy



The smaller the area examined the easier it is to perceive the spatial redundancy.

Many pixels exhibit similar values

Temporal Redundancy



- Object displacements are often not significant between frames
- Many regions of the picture have similar pixel values
- Static regions are fully redundant

Section 3.

Stages of Compression

Topics

1. 4 Stages of Compression of I- frames.
2. Color-Space Sub-sampling
3. Picture Blocking

0
101
00110
11010001
11100
001
110
01

4 Stages of Compression*

0
101
00110
11010001
11100
001
110
01

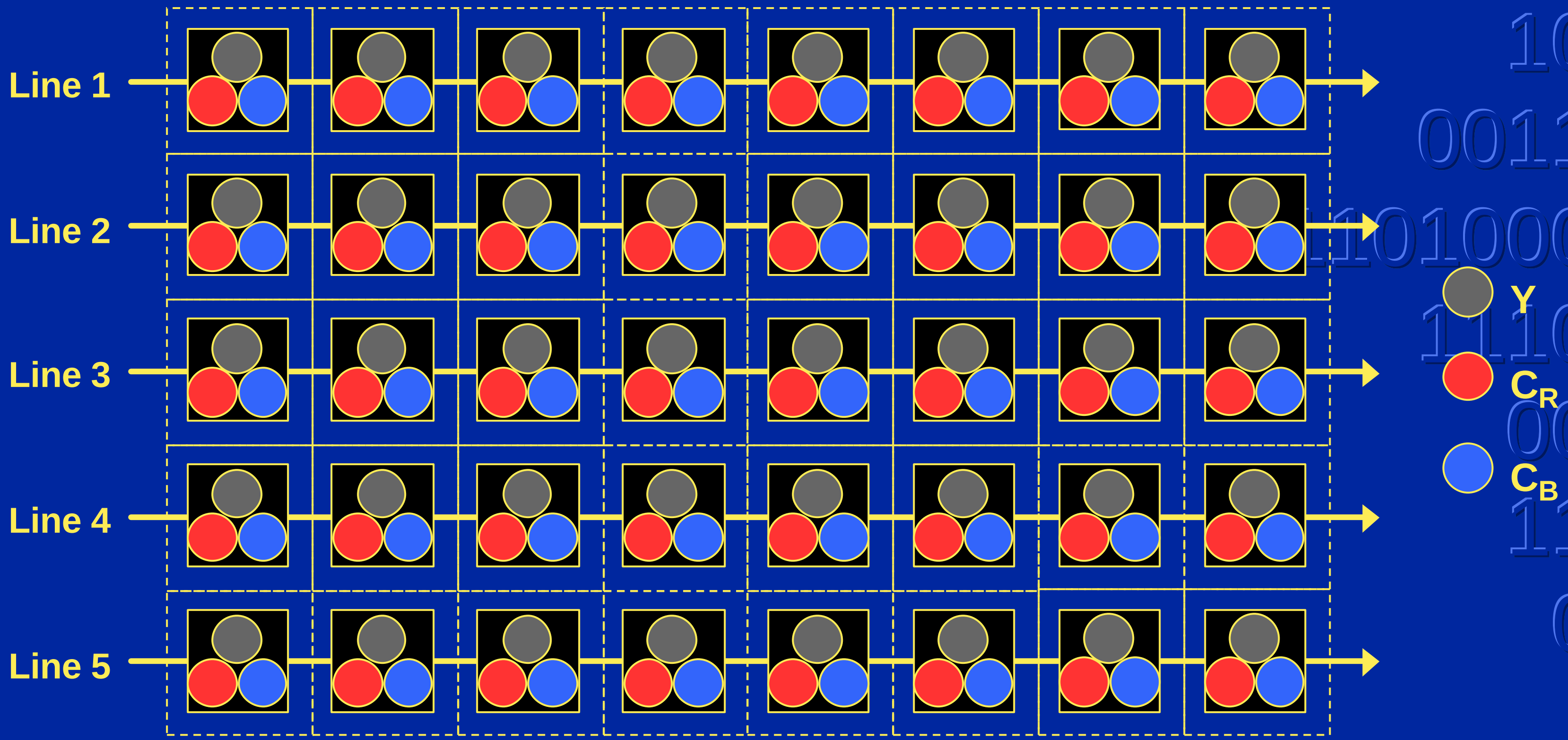
- **Stage 1: Pre-filtering**
  - Restructuring the video's component signals.
- **Stage 2: Picture Blocking: Field or Frame, Slice, Macro- & Micro-Blocking**
- **Stage 3: Algorithm**
  - Fourier Transform using DCT
- **Stage 4: Bit Rate Reduction**
  - Technologies used to exploit the redundancies revealed by DCT

*At this point we are creating I --inter-- frames only.

Stage 1:
Pre-Processing

- A.K.A.: Pre-filtering or Sub-sampling
- The data is sub-sampled to produce the desired preliminary data rate for further processing. Examples are 4:2:2:, 4:2:0, and 4:1:1, each of which is described in the following slides.

4:4:4 Sampling



720 Samples/Line

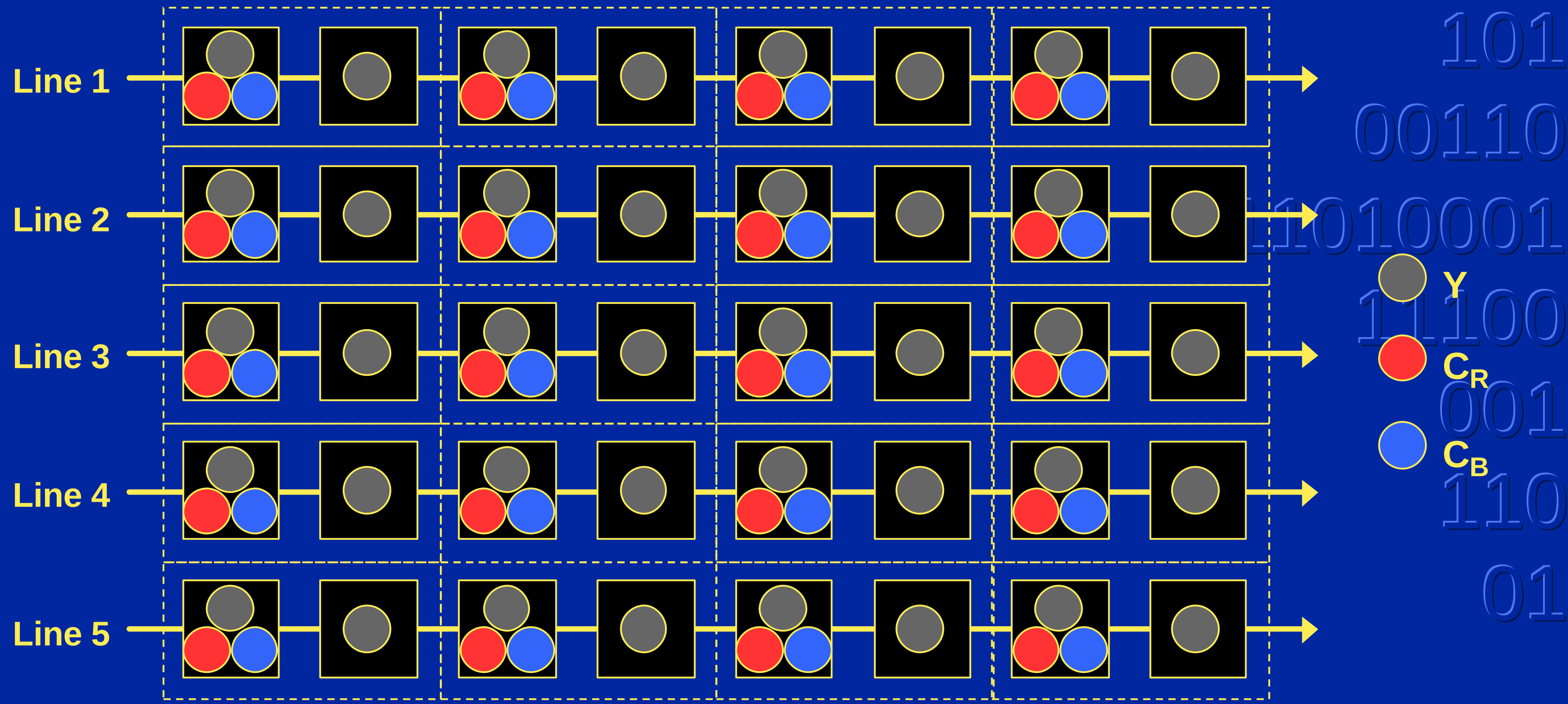
576 Lines — Luminance Y

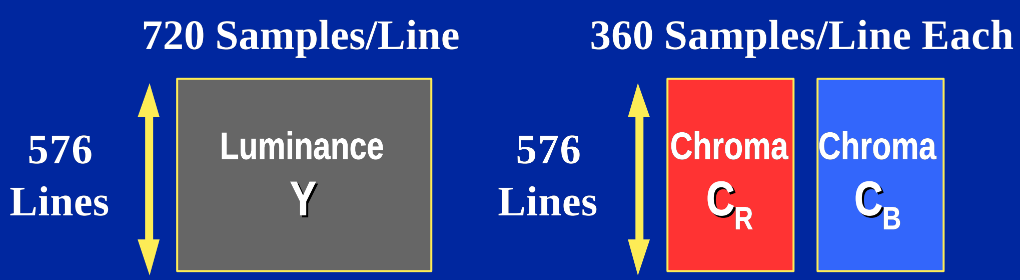
720 Samples/Line Each

576 Lines — Chroma C_R — Chroma C_B

8-bit = 248 Mbps
10-bit = 311 Mbps

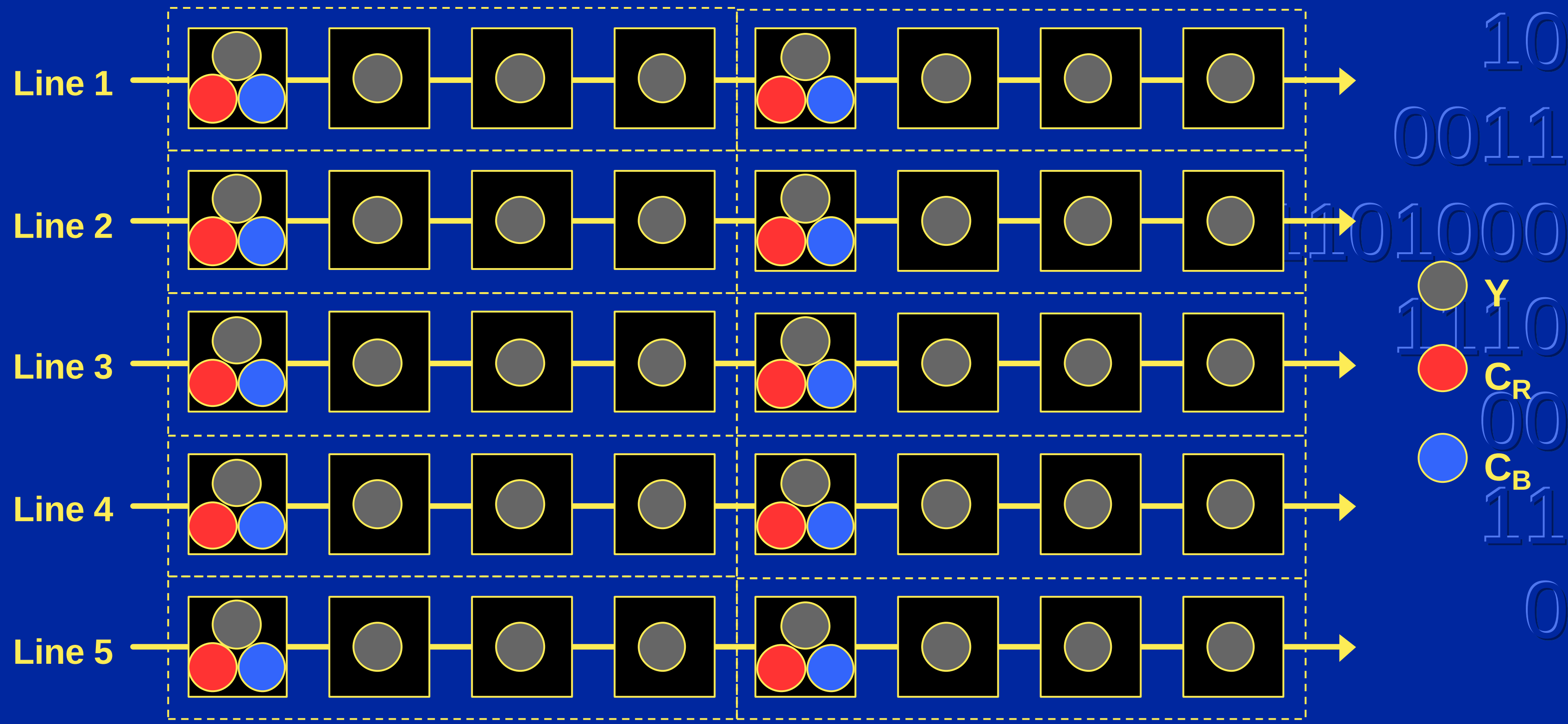
4:2:2 Sampling

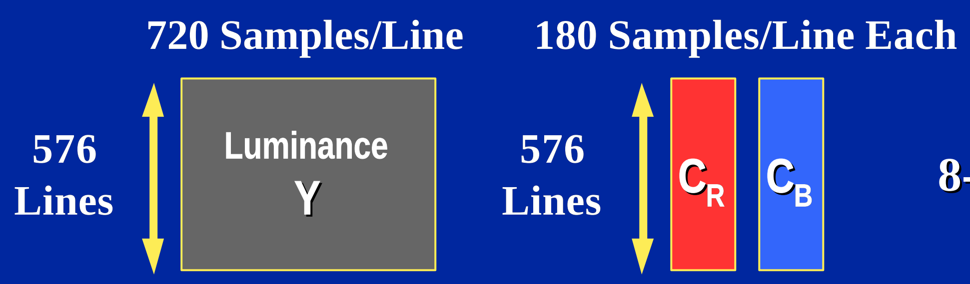




8-bit = 166 Mbps
10-bit = 208 Mbps

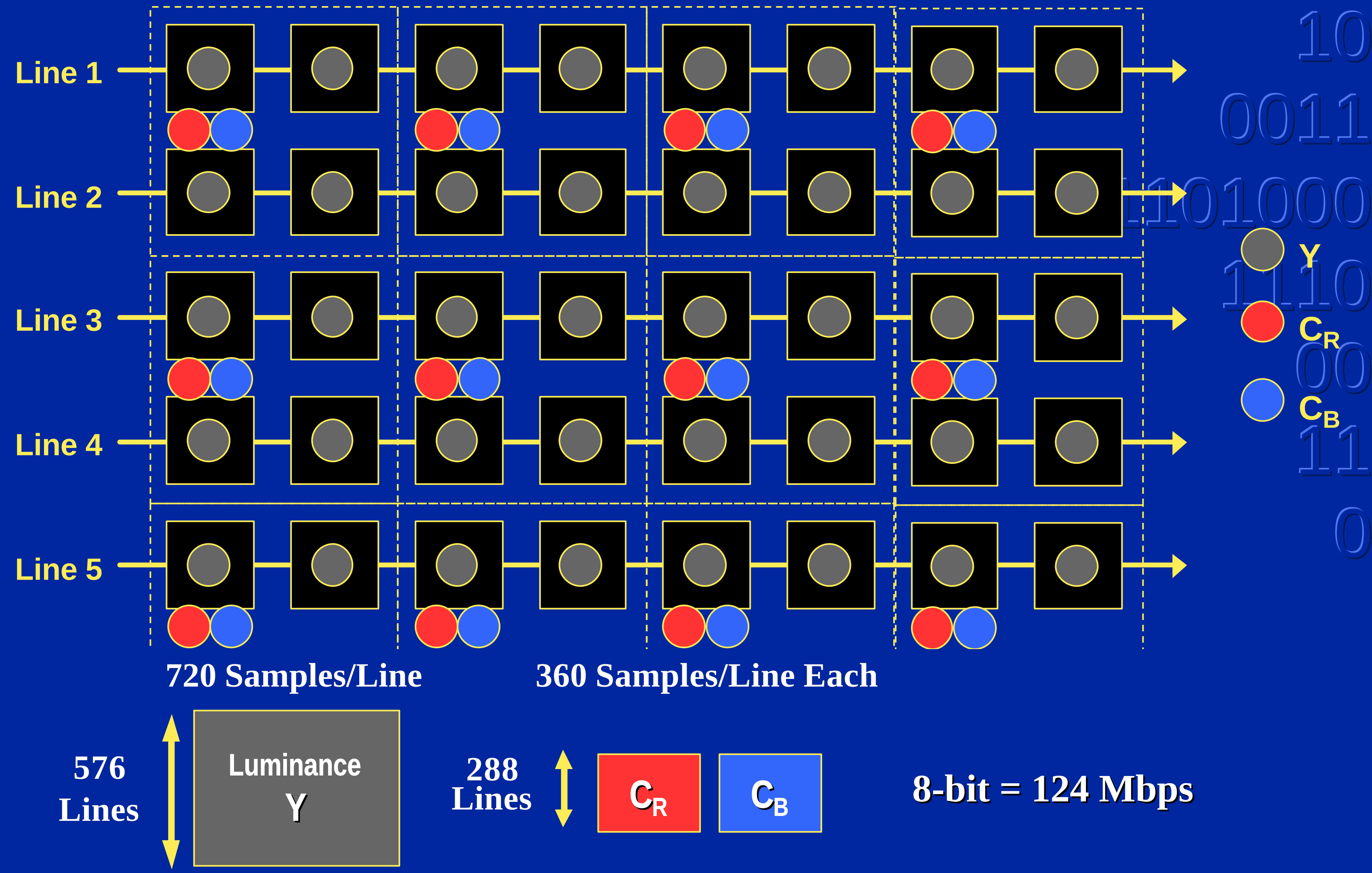
4:1:1 Sampling





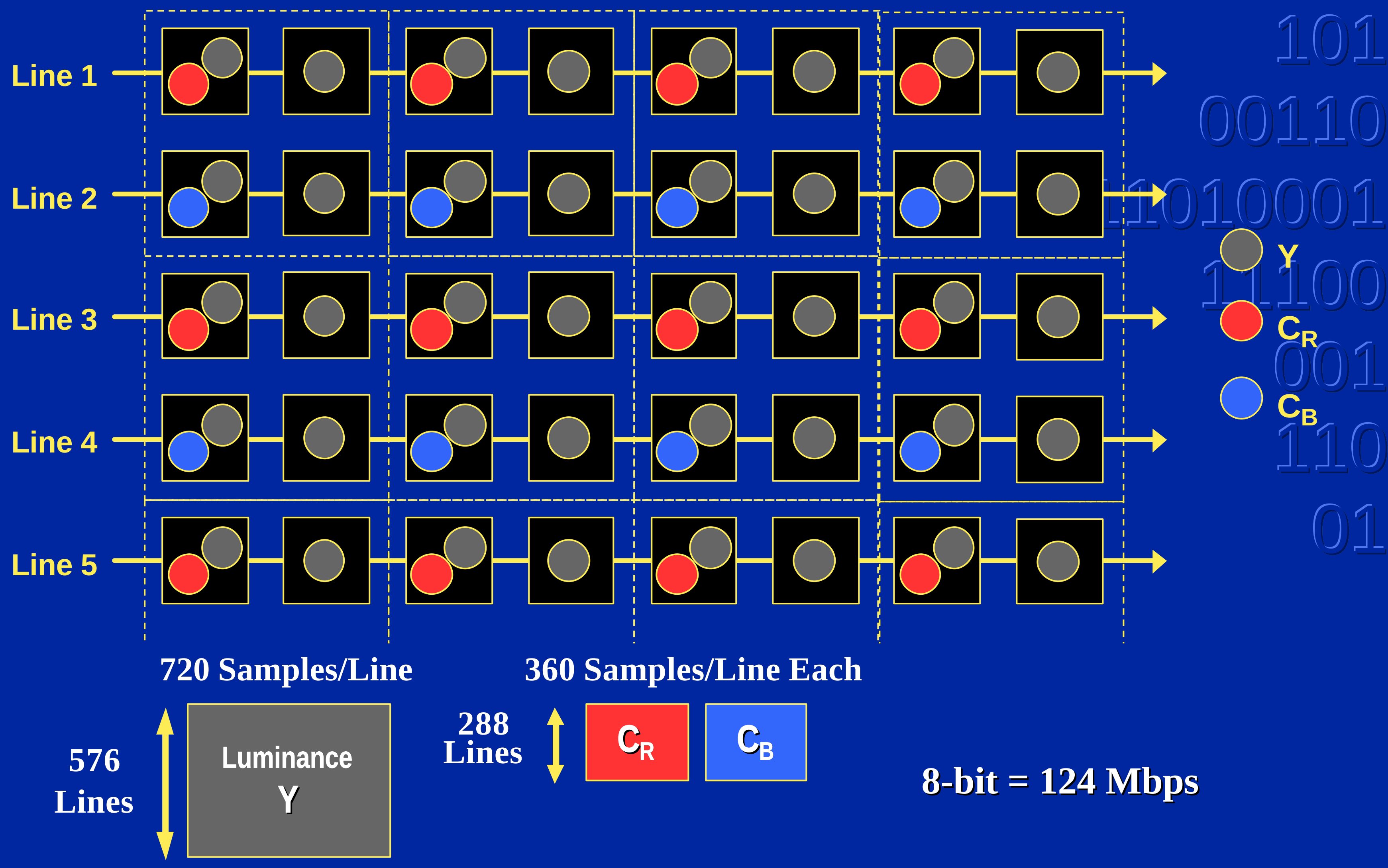
8-bit = 124 Mbps

4:2:0 Sampling

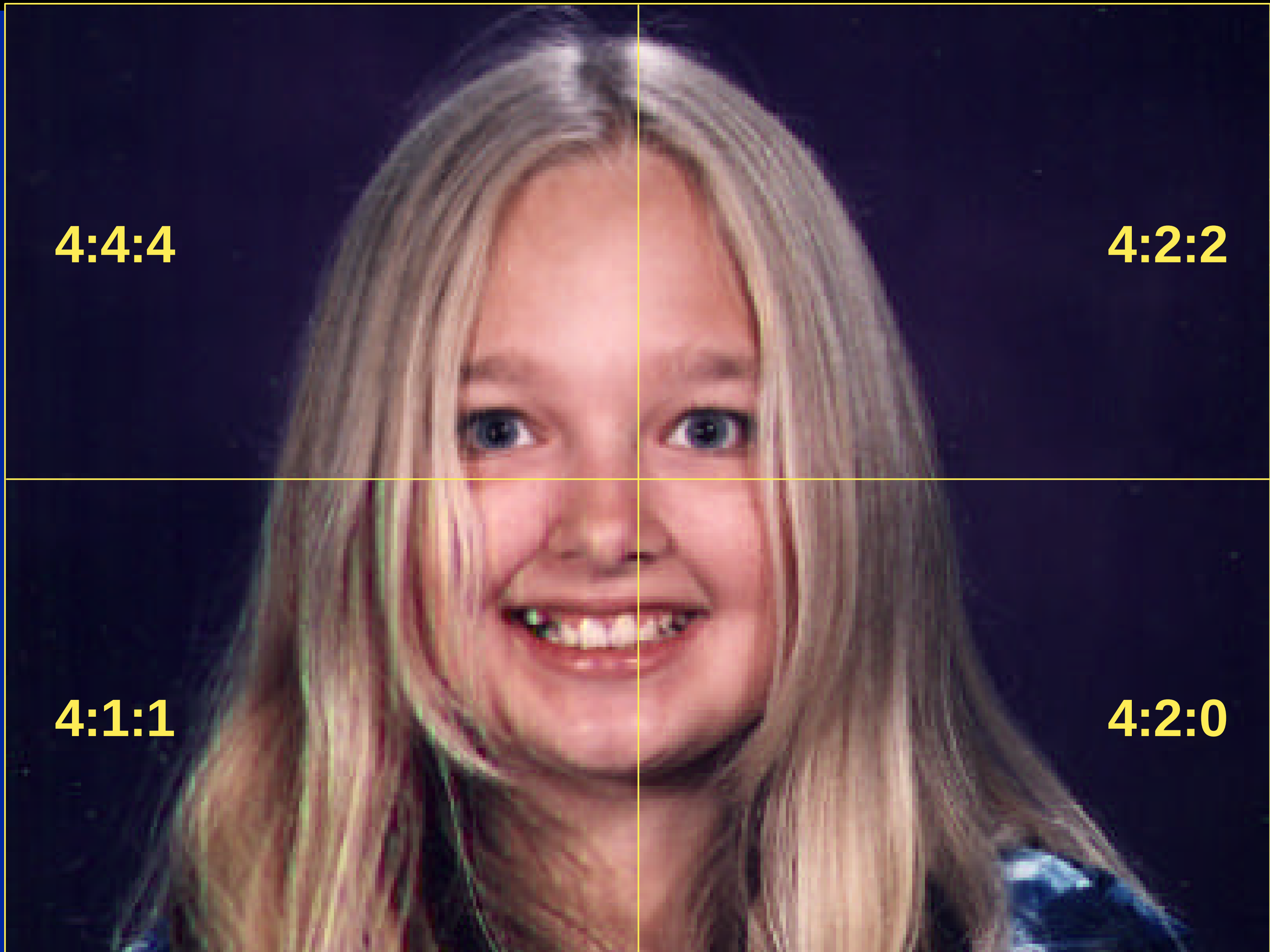


8-bit = 124 Mbps

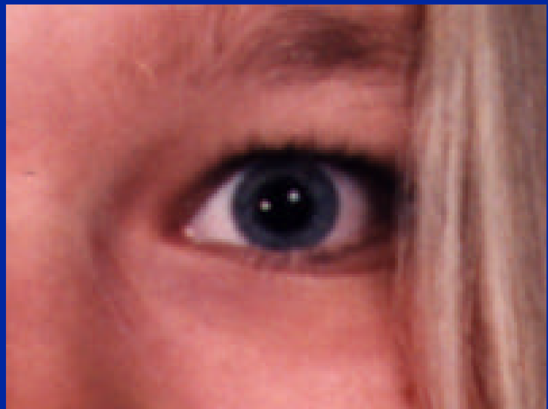
4:2:0 Sampling (625 DVB)



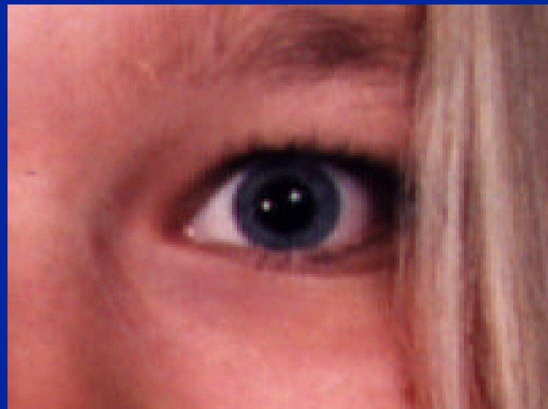
Sampling Examples



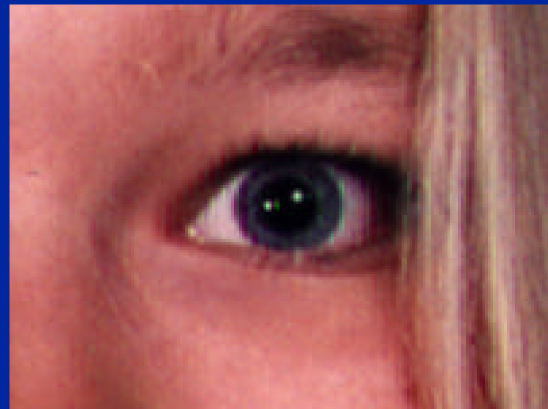
Sampling Examples



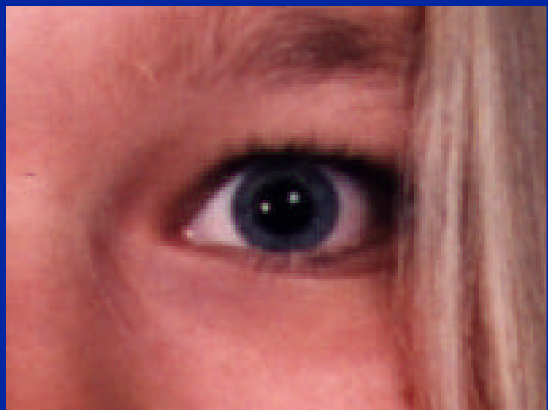
4:4:4



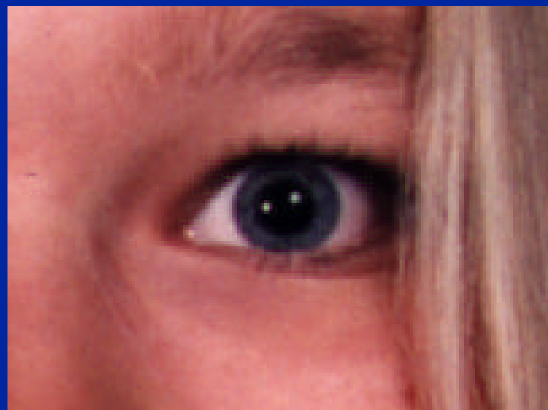
4:2:2



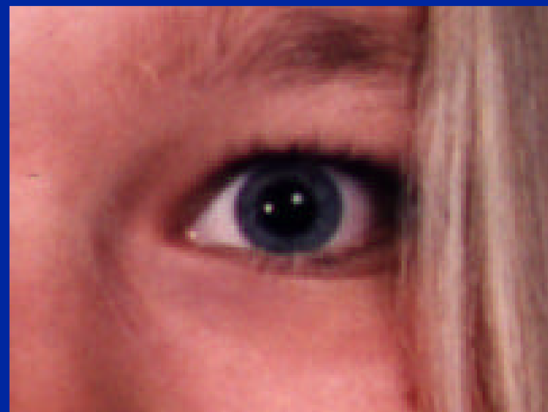
4:1:1



4:2:0
MPEG-1

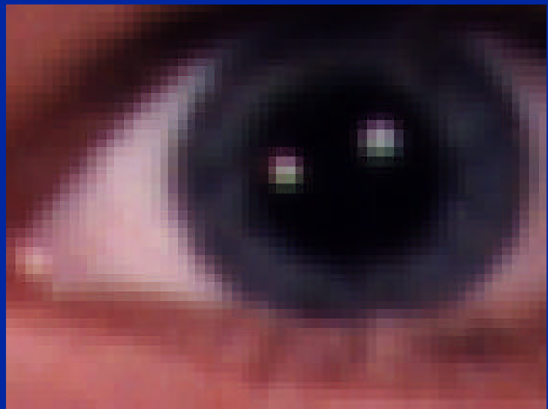


4:2:0
MPEG-2

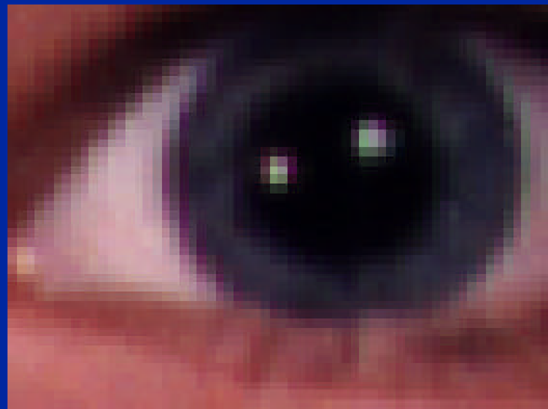


4:2:0
625 DVB

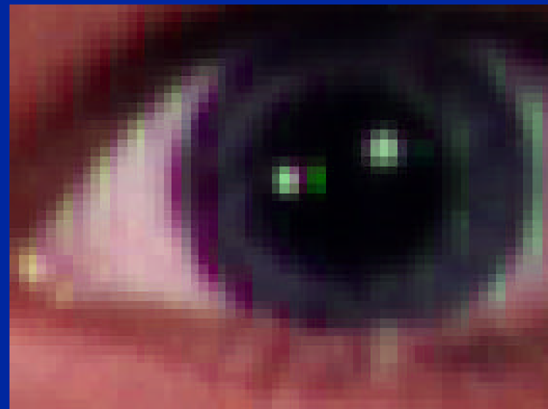
Sampling Examples



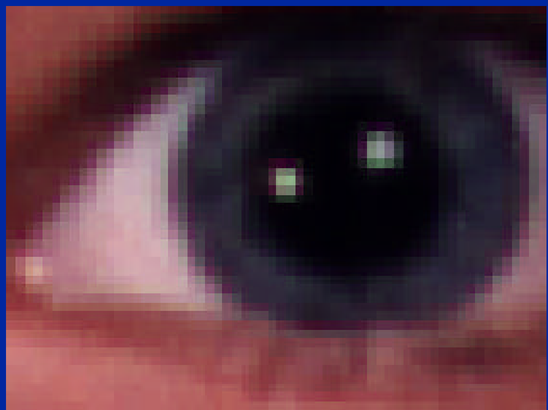
4:4:4



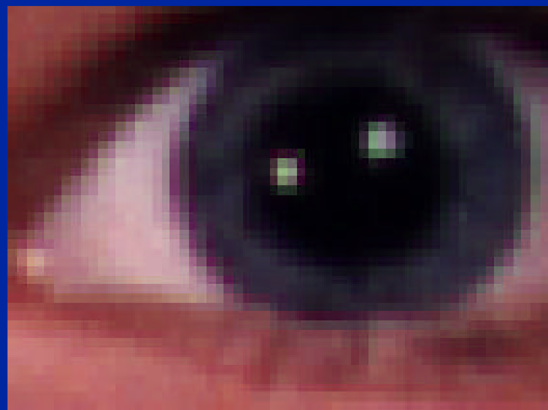
4:2:2



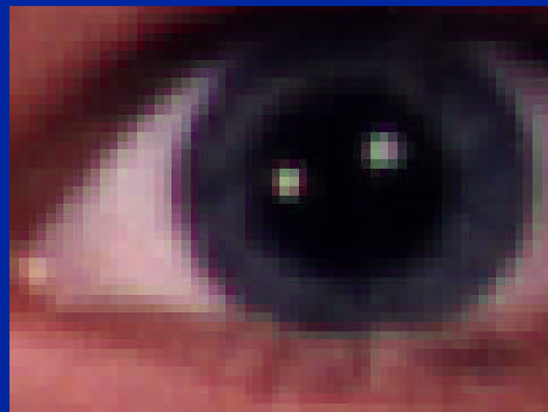
4:1:1



4:2:0
MPEG-1

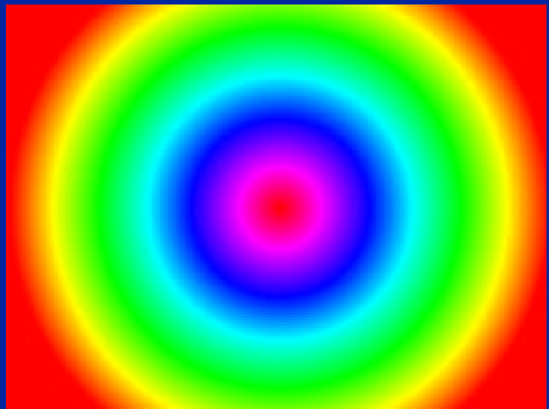


4:2:0
MPEG-2

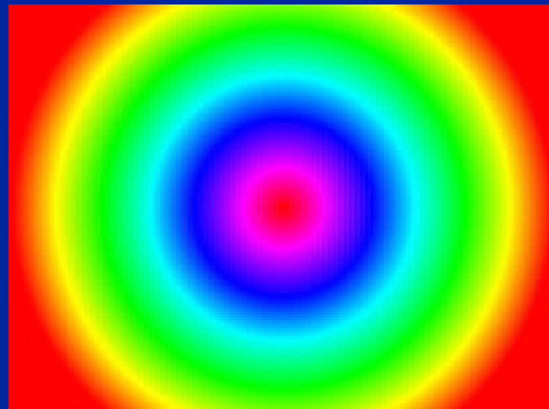


4:2:0
625 DVB

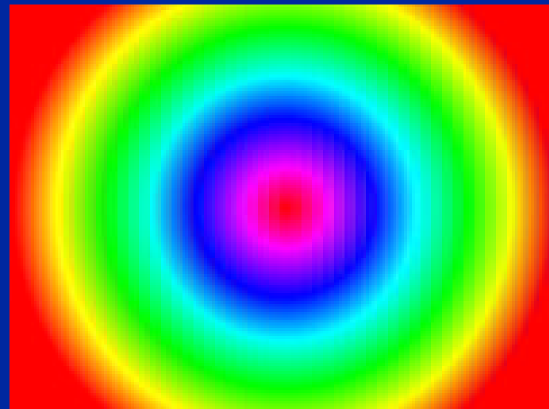
Sampling Examples



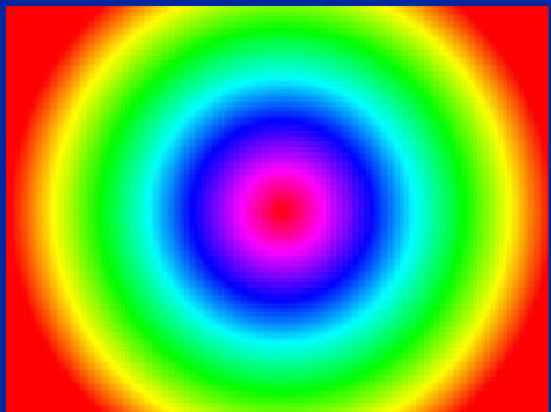
4:4:4



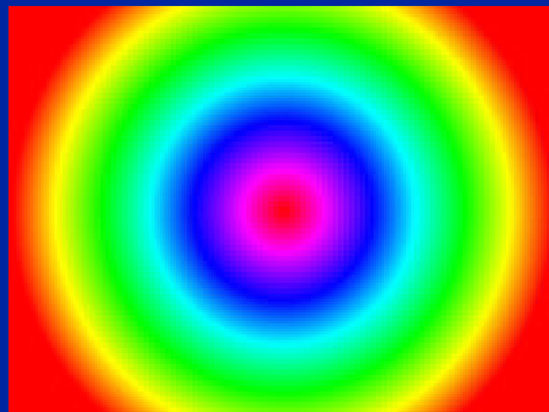
4:2:2



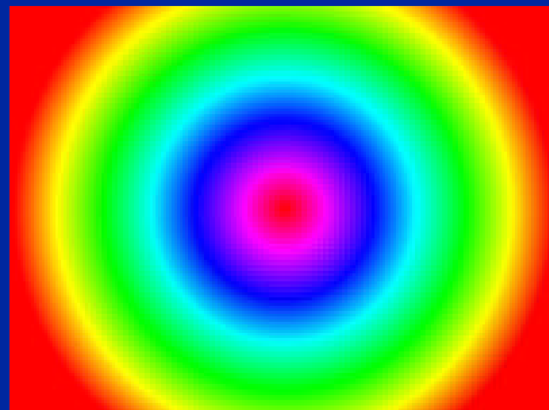
4:1:1



4:2:0
MPEG-1



4:2:0
MPEG-2

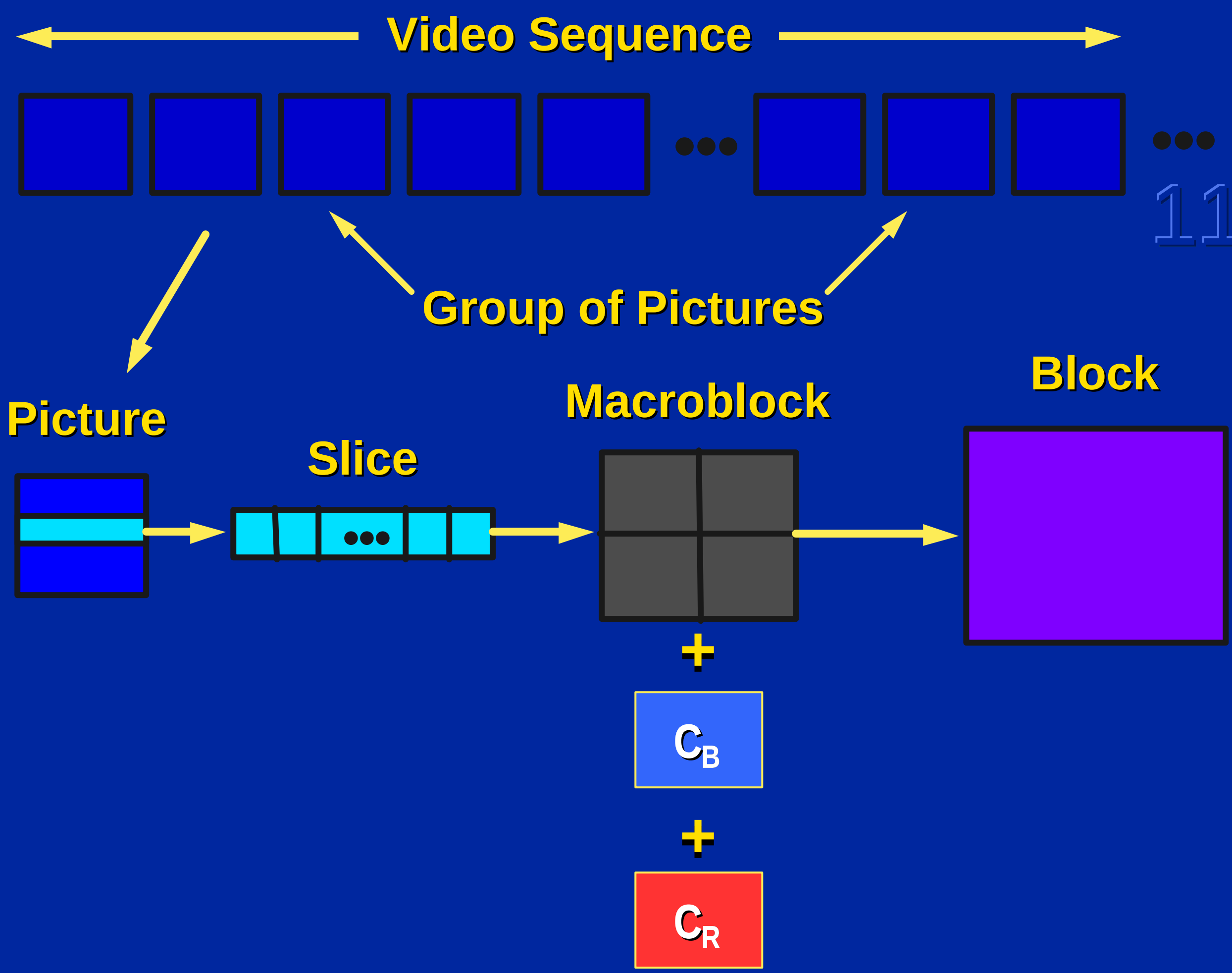


4:2:0
625 DVB

Stage 2:
Field or Frame, Slice, Macro Blocking

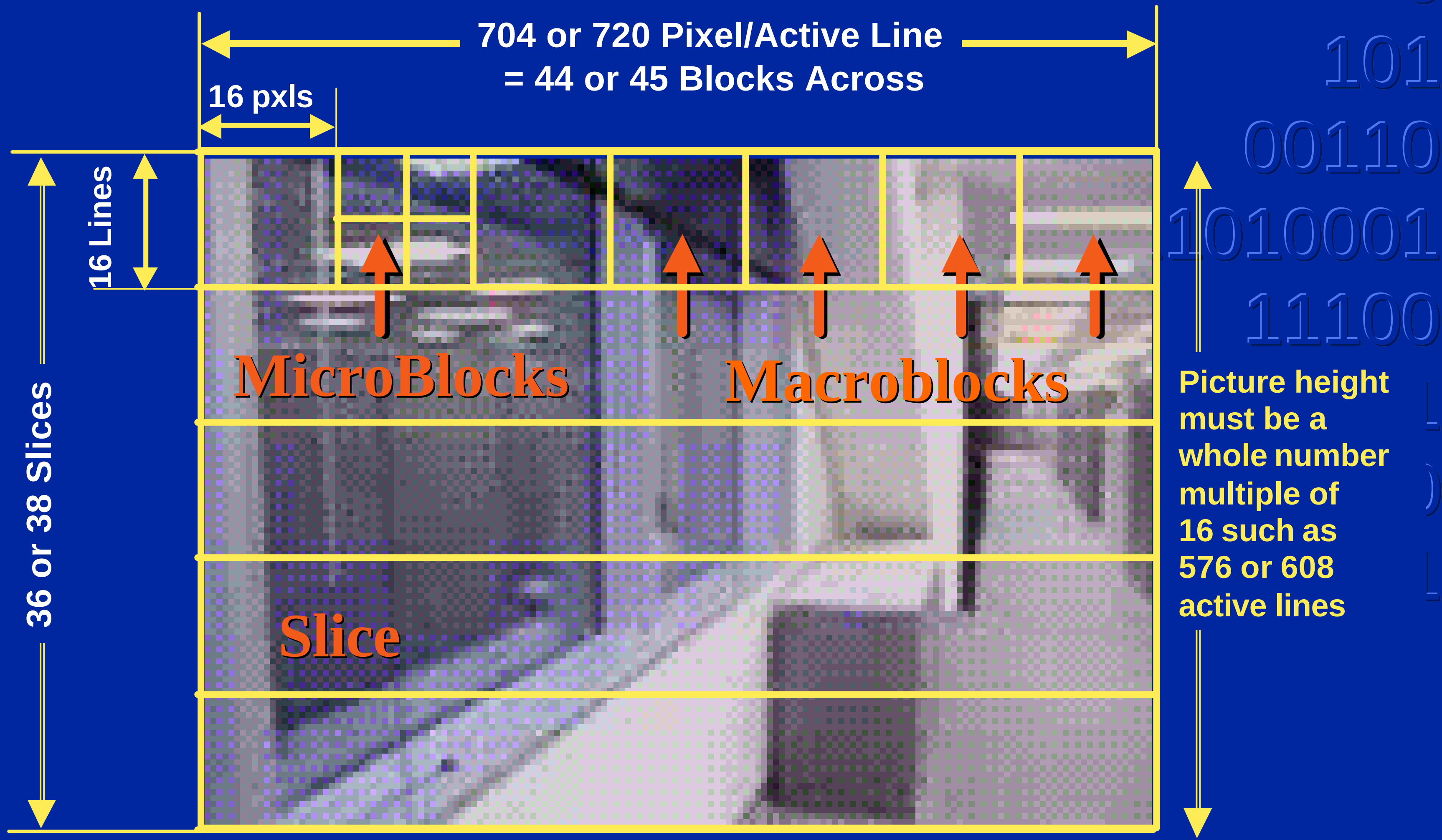
- The picture information is broken into manageable pieces of information to allow easy conversion of data.
- Blocking may be performed in the field or frame mode.
  - Either the field or frame is broken into 16 pixel high (I.e. 16 line thick) slices.
  - These are broken into 16X16 pixel blocks for Y and 2 8X18 pixel blocks for R-Y and B-Y.
  - Grouped together they form a Macroblock.

MPEG Data Hierarchy

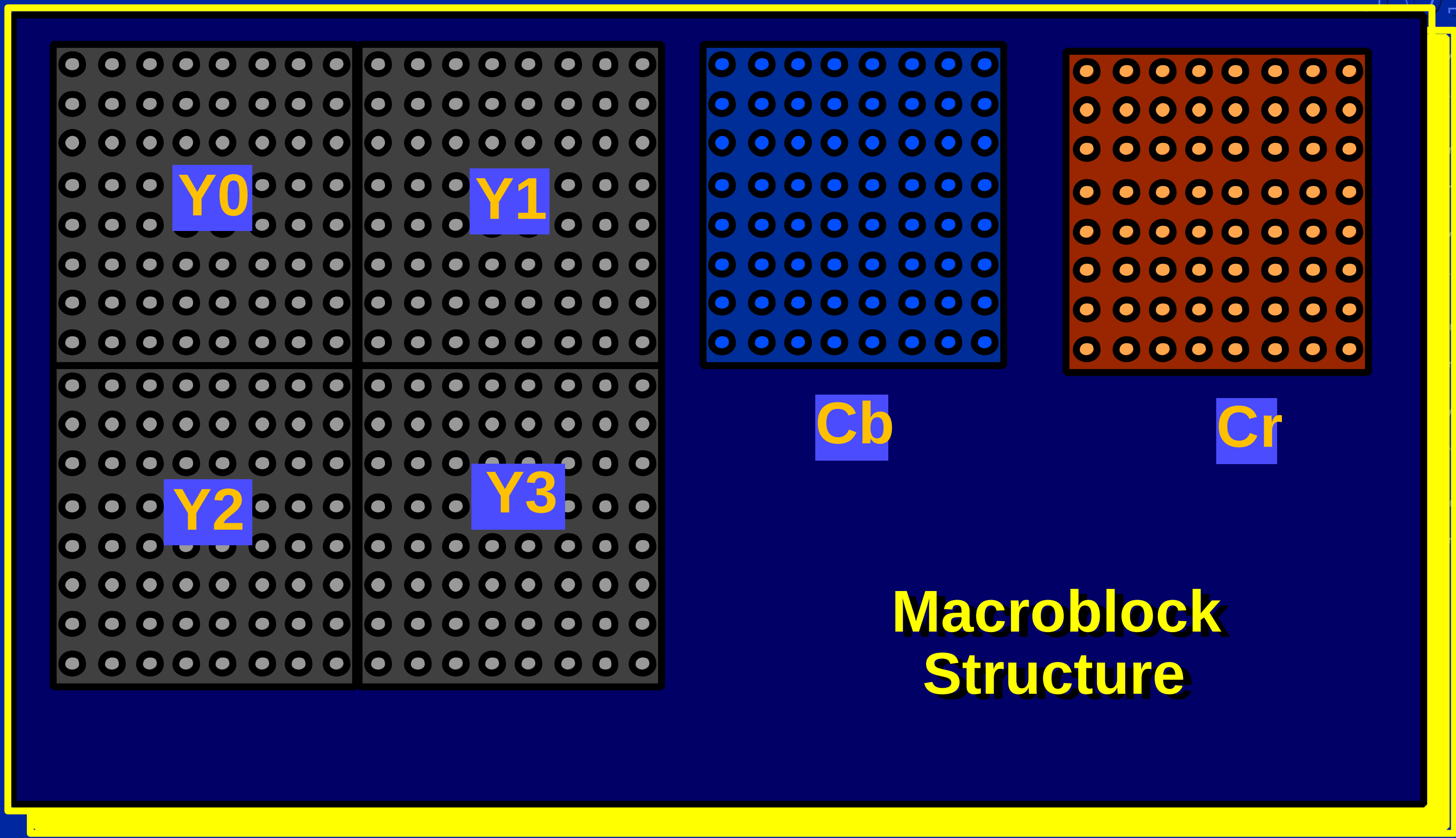


0
101
00110
11010001
11100
001
110
01

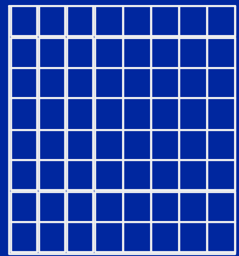
Sub-Division of a Picture into MPEG Macroblocks



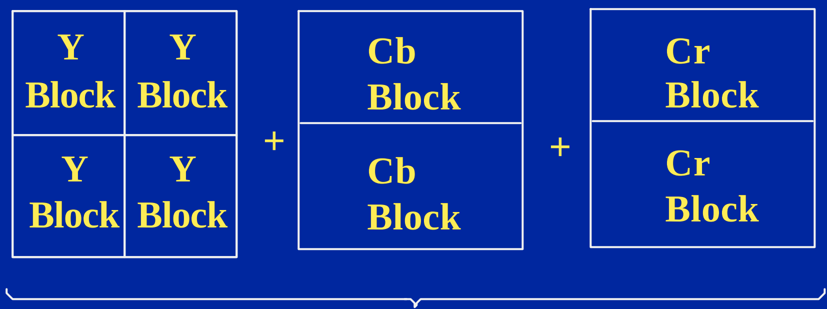
Blocks and Macroblocks



MPEG Block Structure

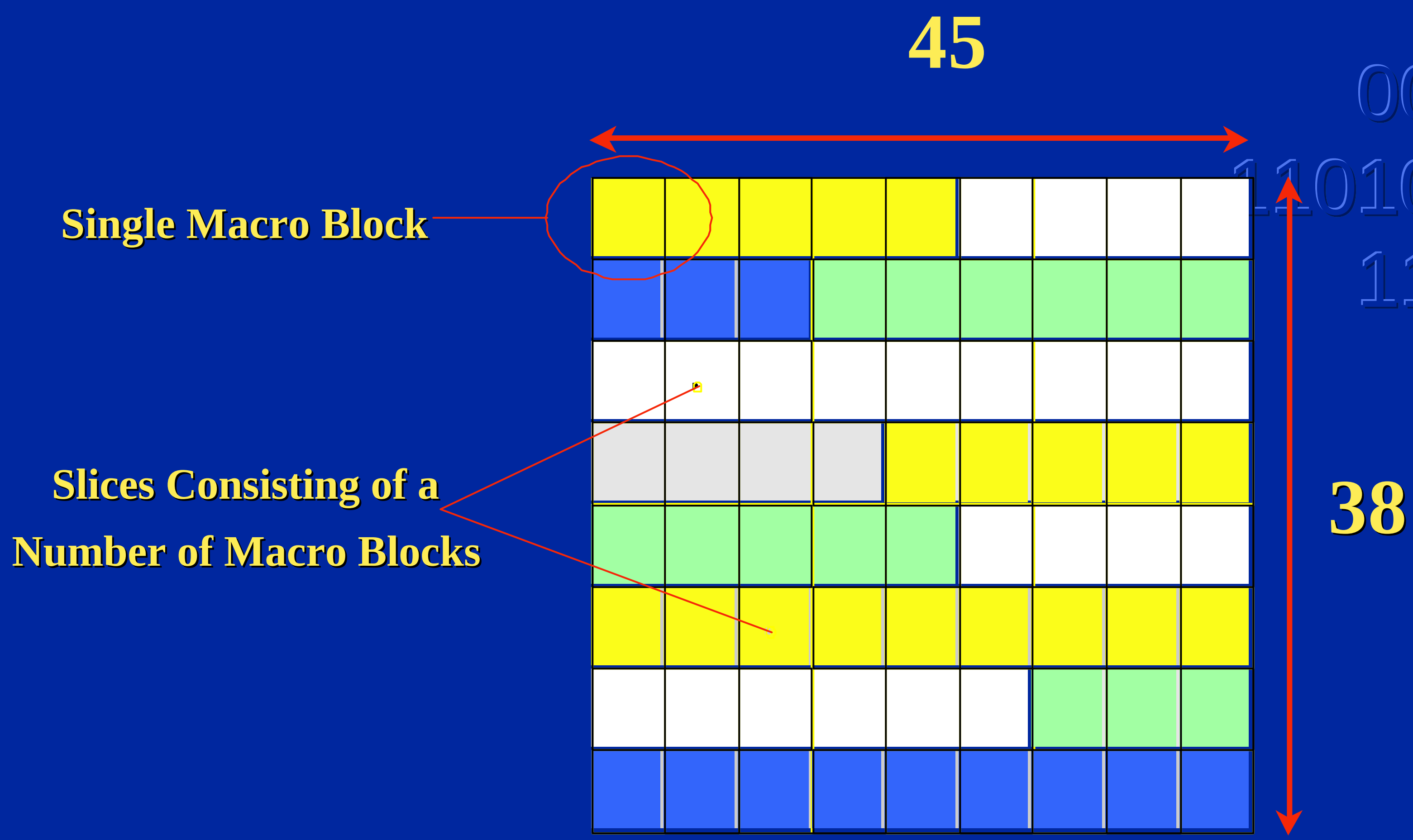


8 x 8 Pixel
Block
Y, Cr or Cb



422P @ ML Macro Block consisting of
Y, Cr & Cb Blocks

MPEG Frame Structure



An MPEG 422P @ ML Frame of 45 x 38 Macro Blocks

0
101
00110
11010001
11100
001
110
01

Section 4.

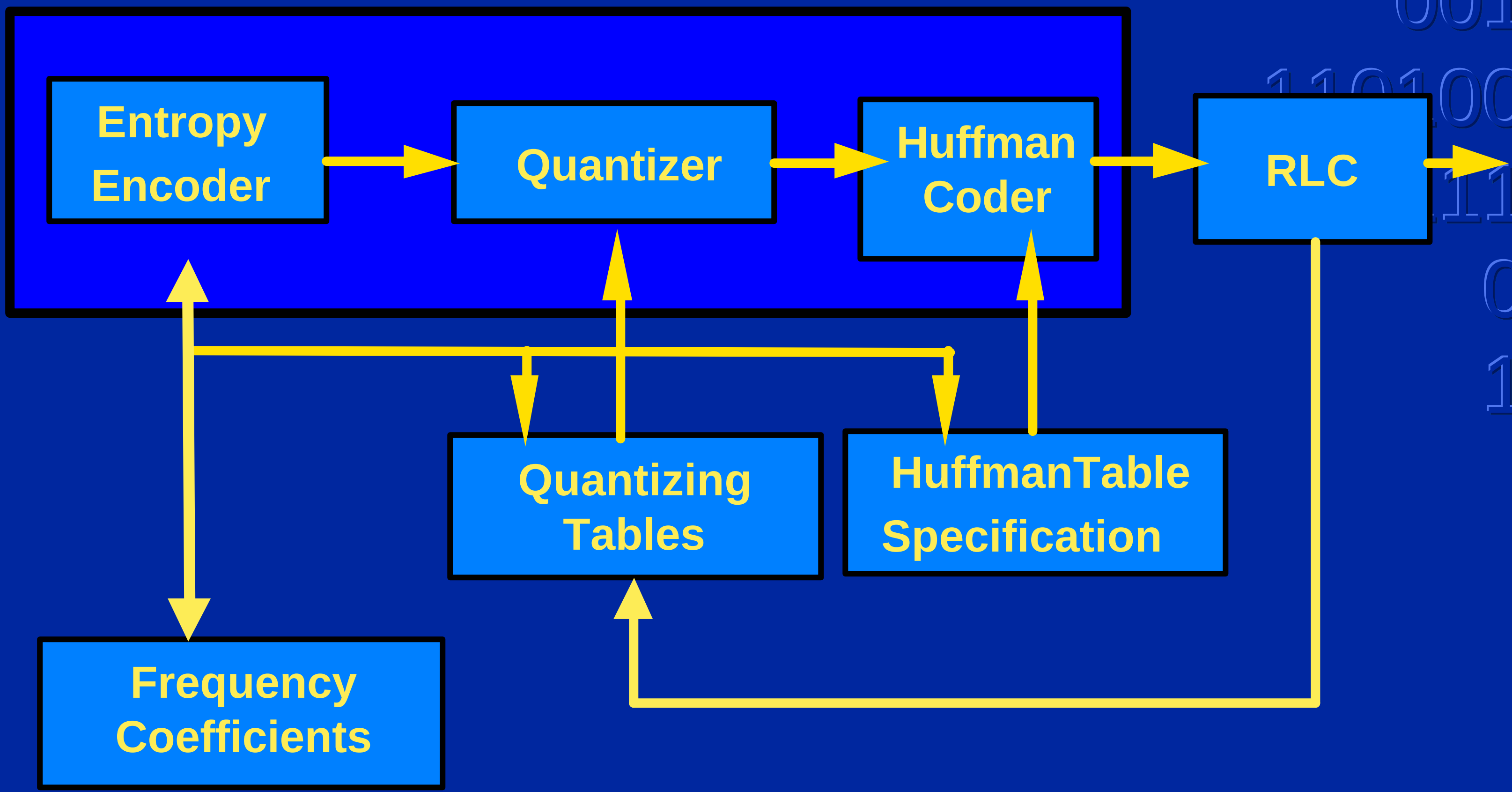
The Spatial Compression Process

Topics:

1. **Transforms**
   - **DCT Example with Basis Functions**
2. **Zig-Zag Ordering**
3. **Quantising**
4. **Entropy Coding**
   - **Using a Huffman Table**

0
101
00110
11010001
11100
001
110
01

DCT-Based Spatial Encoder



Discrete Cosine Transform (DCT)

- Transforms spatial pixel domains to the frequency domain
- Transformed data is more suited to bit-rate reduction techniques
- It takes advantage of limitations of human visual perception in the decoding process

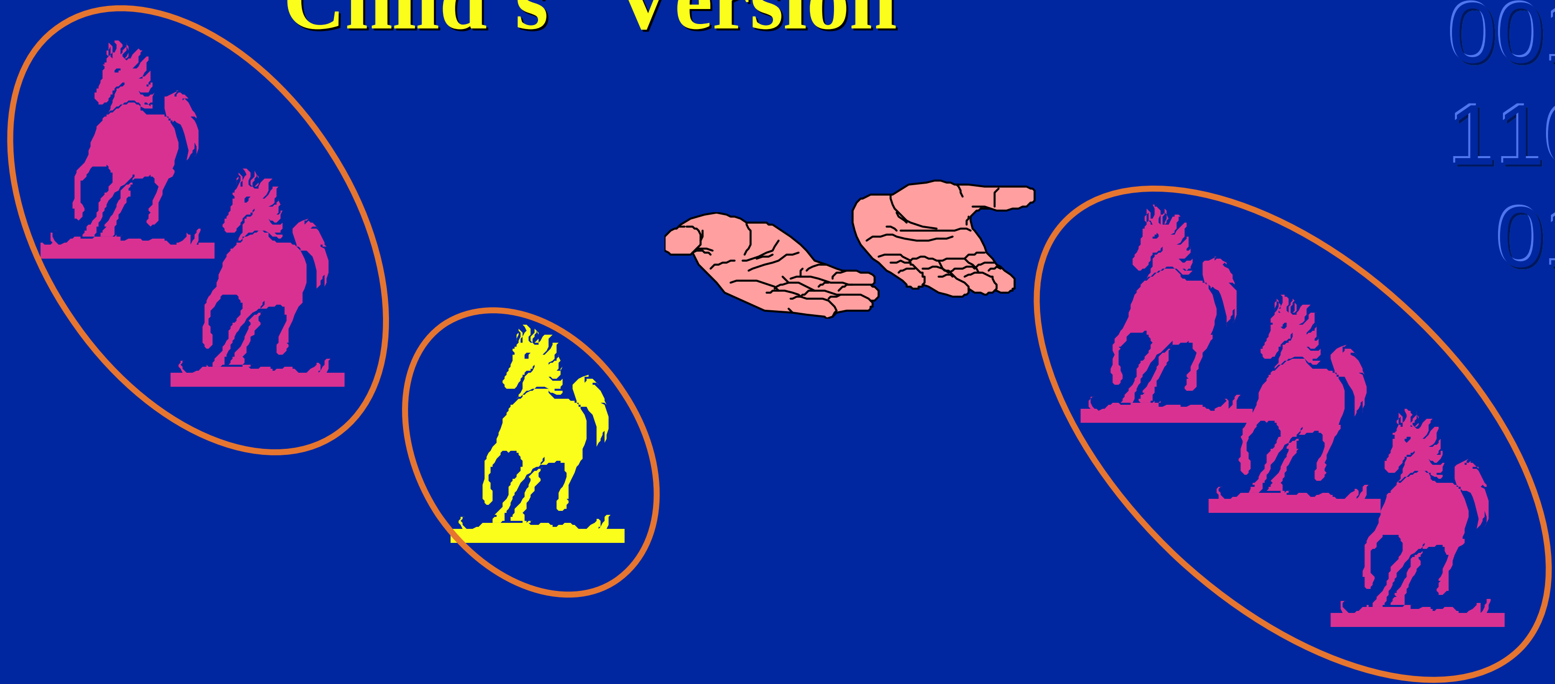
Data Transform

A Simple Number Transform

“Adult” Version: 2+1 = 3

Vs.

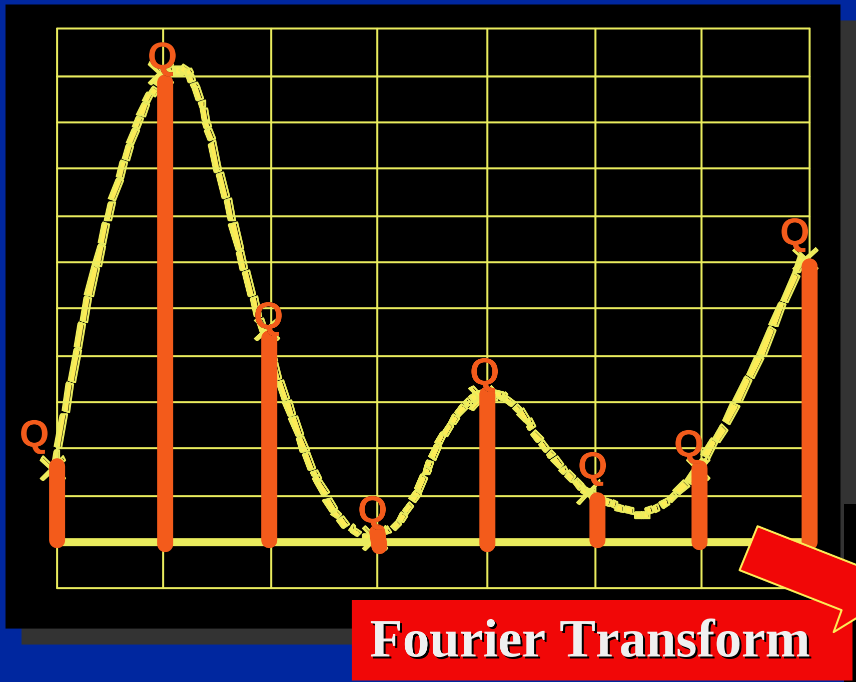
“Child’s” Version

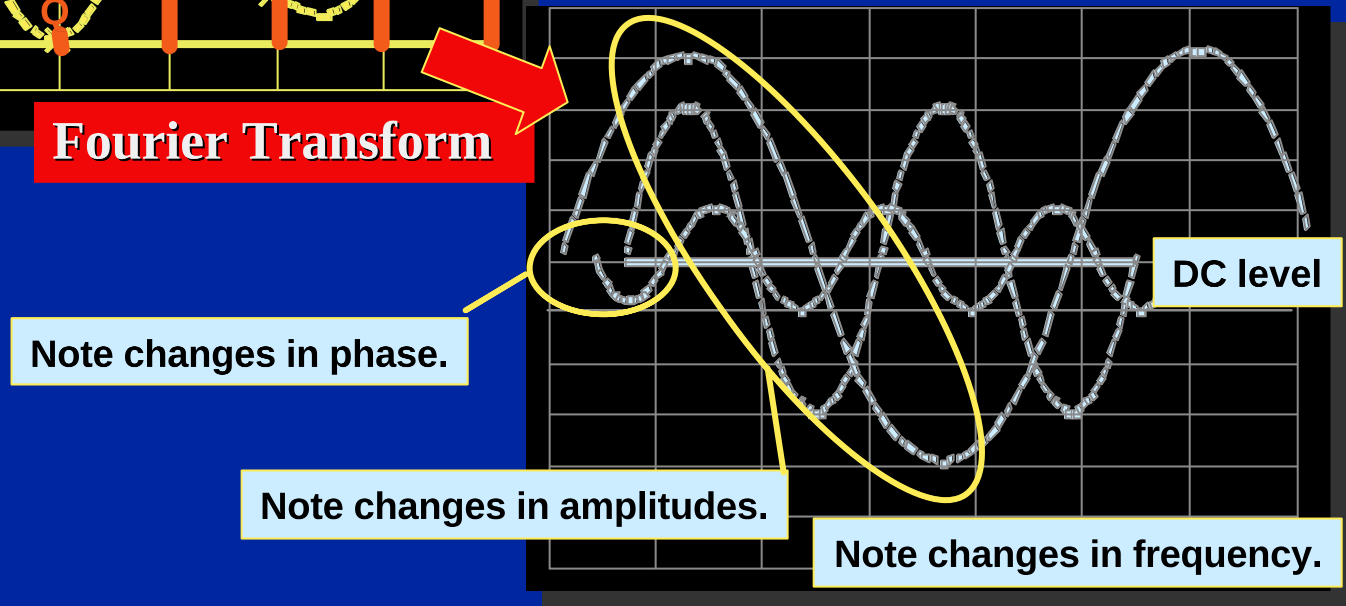


0
101
00110
11010001
11100
001
110
01

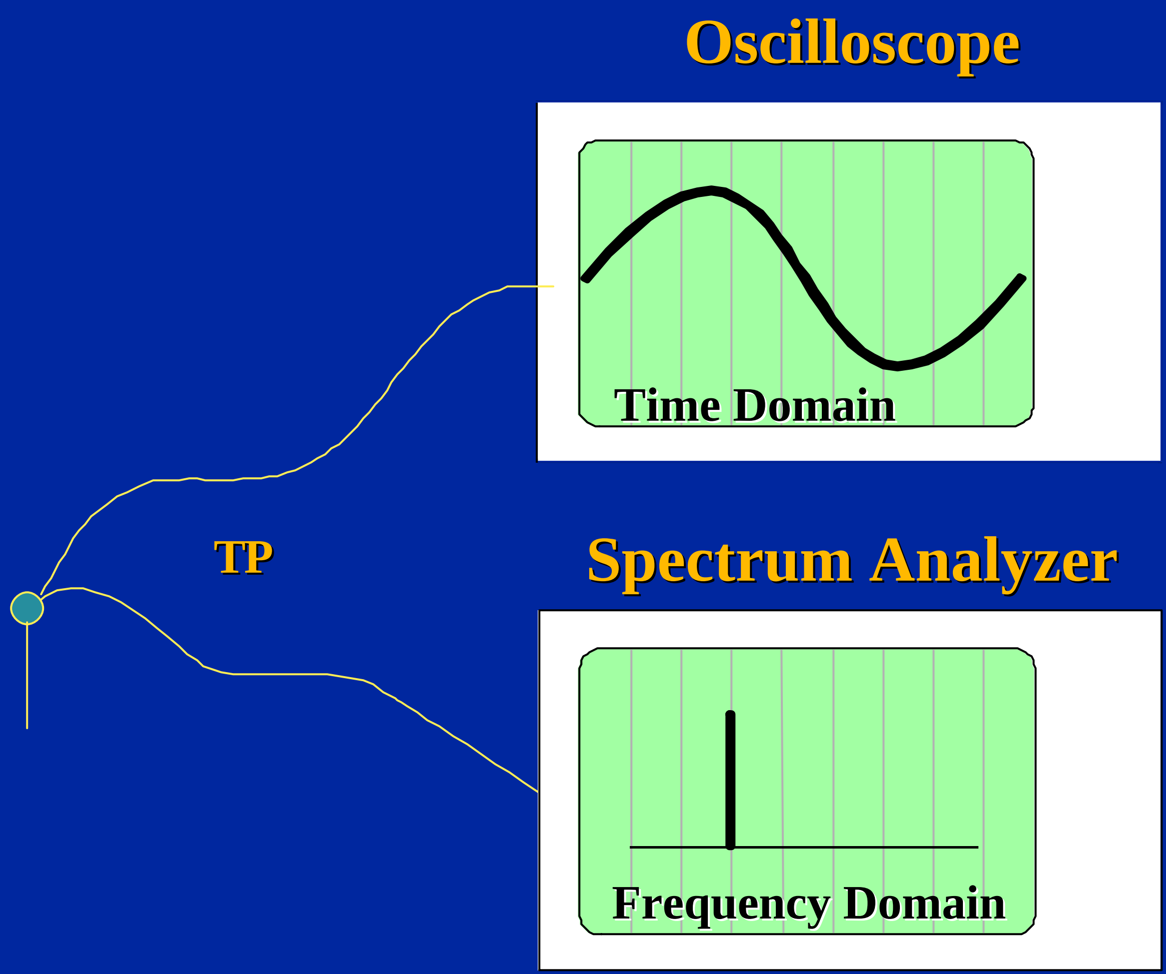
Fourier Transform

Analog Waveform with PAM and Quanta indicated. Note that PAM pulses map the discrete "elevations" of corresponding pixels





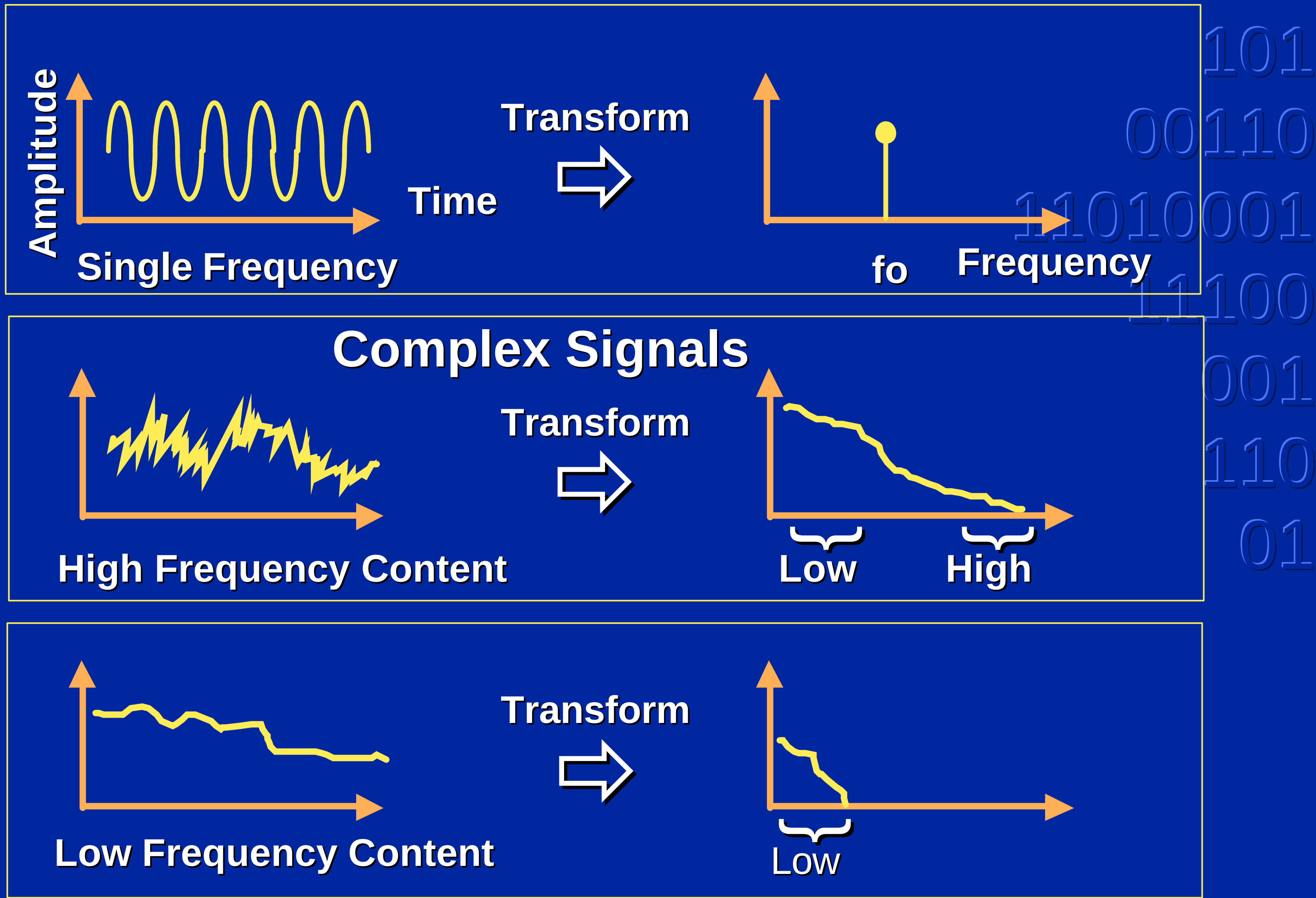
Time and Frequency Domain



Data Values
0.1, 0.2, 0.3, 0.35, 0.3...

Data Values
0, 0, 0, 1, 0, 0, ...

Time vs. Frequency

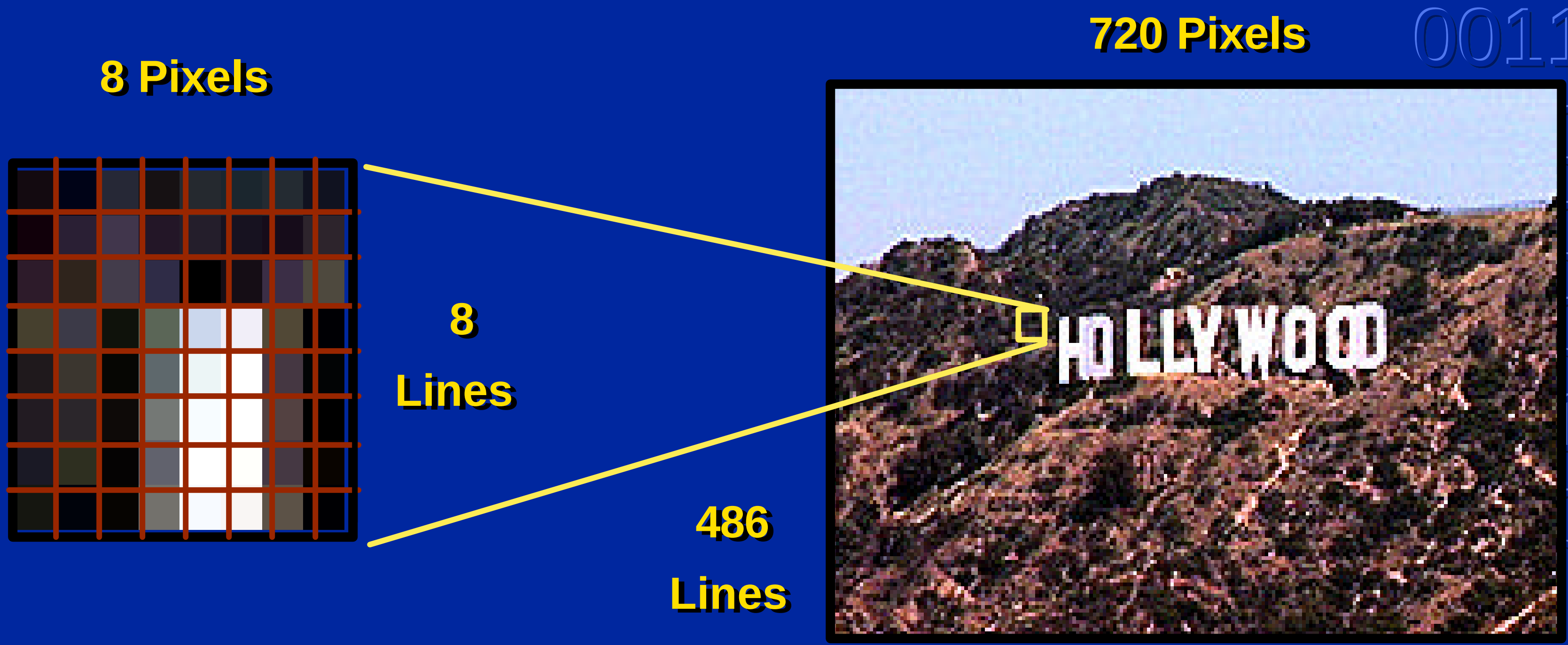


Transform Methods

- Discrete Fourier Transform
  - Mathematically simple
  - Sine function requires two coefficients
    - Generates too much data for compressing video
- Discrete Cosine Transform
  - Mathematically more complex
  - A single coefficient
    - Suitable for compressing video

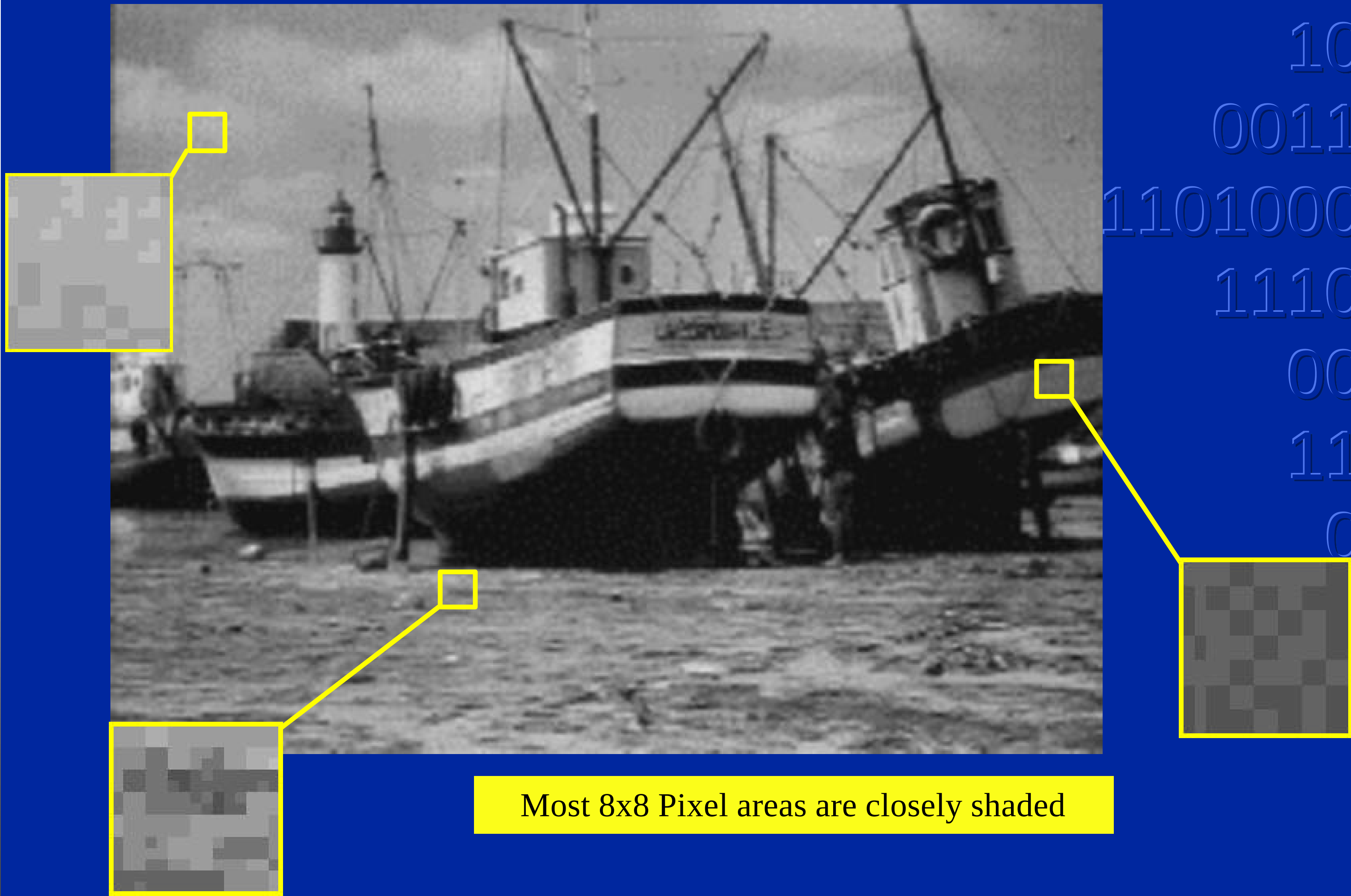
0
101
00110
11010001
11100
001
110
01

DCT 8 x 8 Block of Pixels



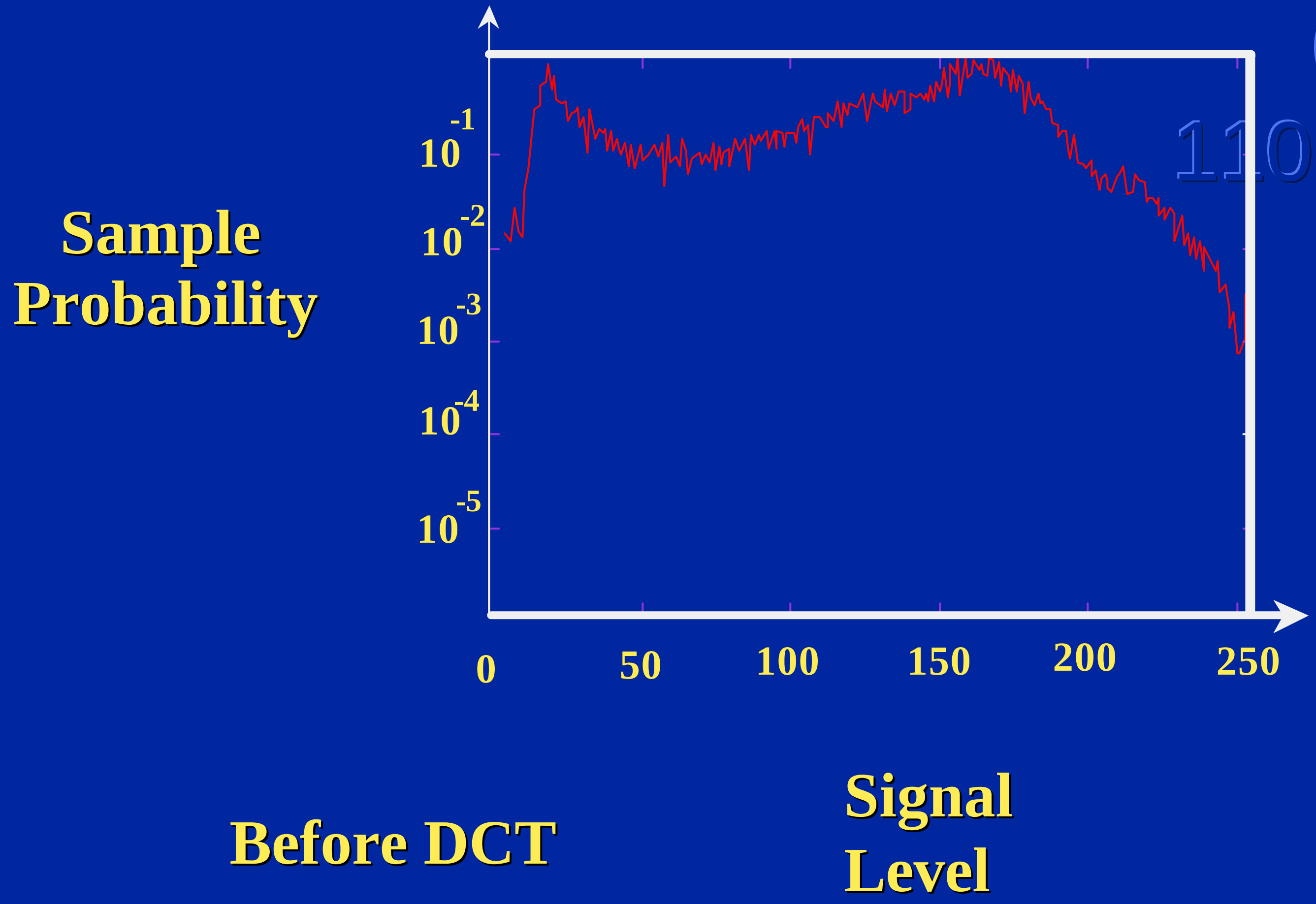
Typical Picture

0
101
00110
11010001
11100
001
110
01



Most 8x8 Pixel areas are closely shaded

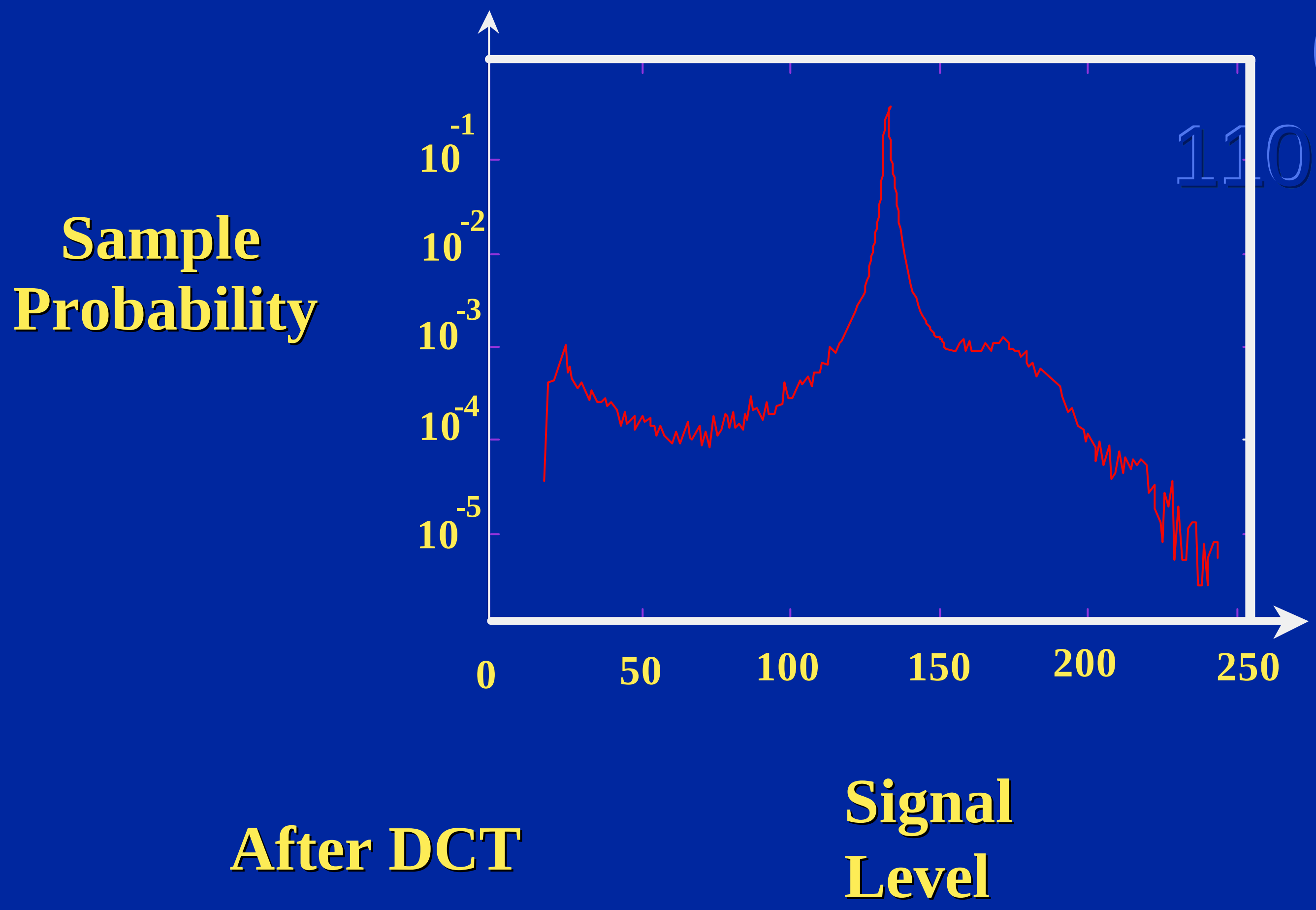
Signal Entropy Before DCT



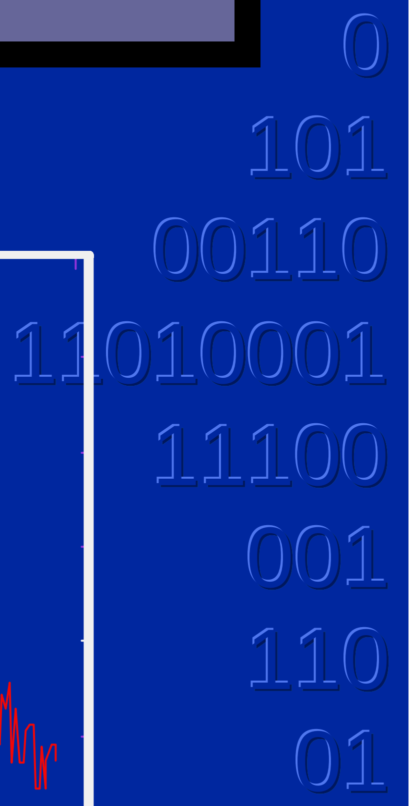
Before DCT

0
101
00110
11010001
11100
001
110
01

Signal Entropy After DCT



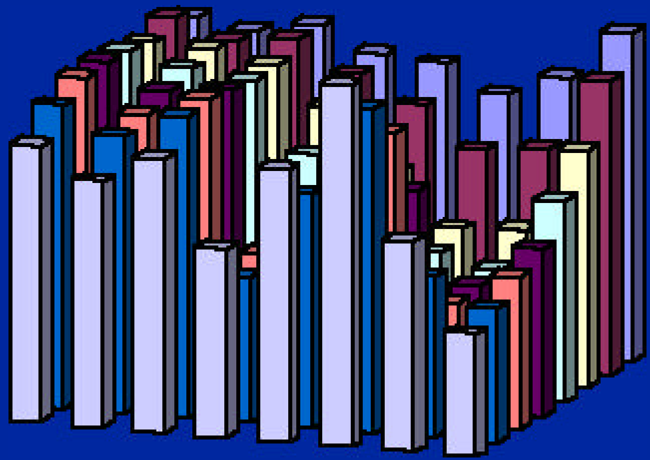
After DCT



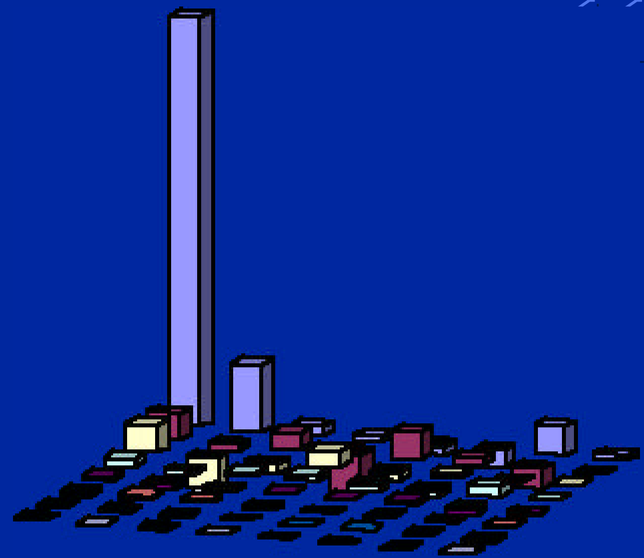
Pixel/Coefficient Distribution

0
101
00110
11010001
1100
001
110
01

Same information in DCT block but more compact



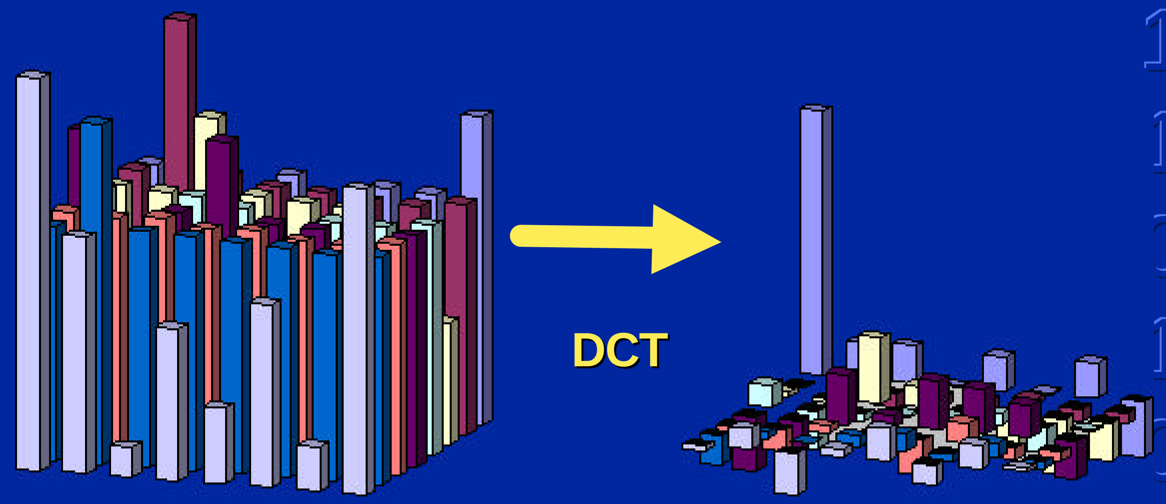
Spatial information carried by all pixels in the block



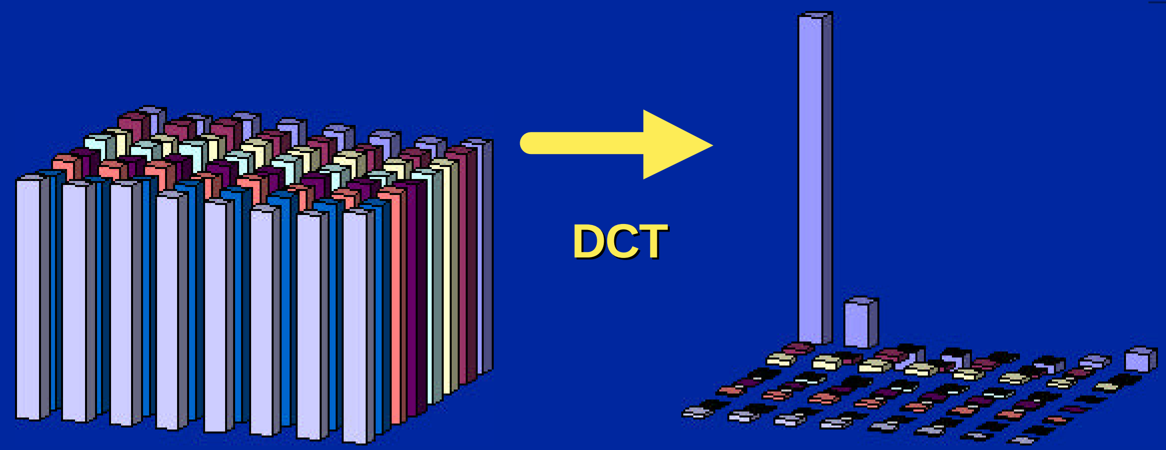
DCT Coding

Pixel/Coefficient Distribution

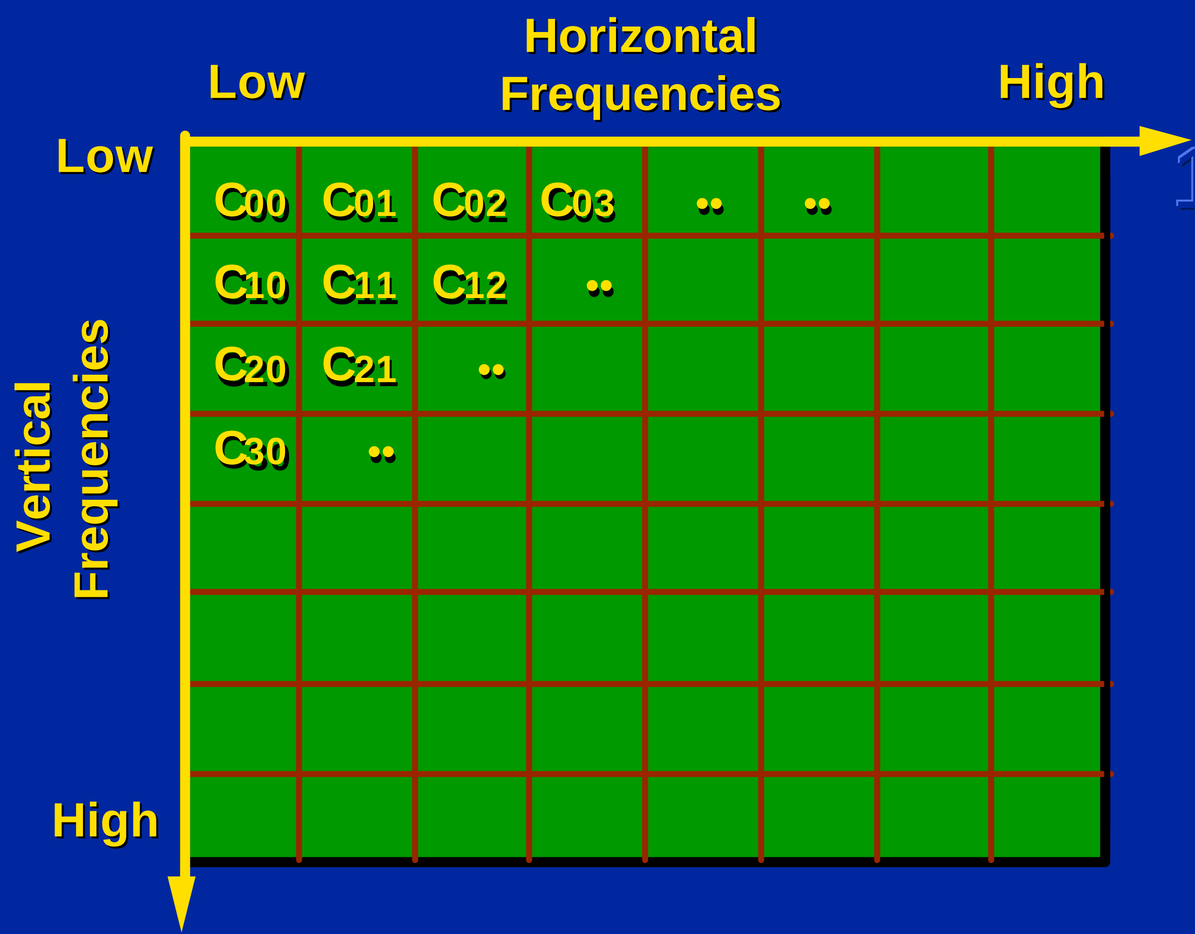
Larger abrupt changes in pixel amplitudes mean more frequency coefficients



Smaller relative amplitude changes mean fewer frequency coefficients

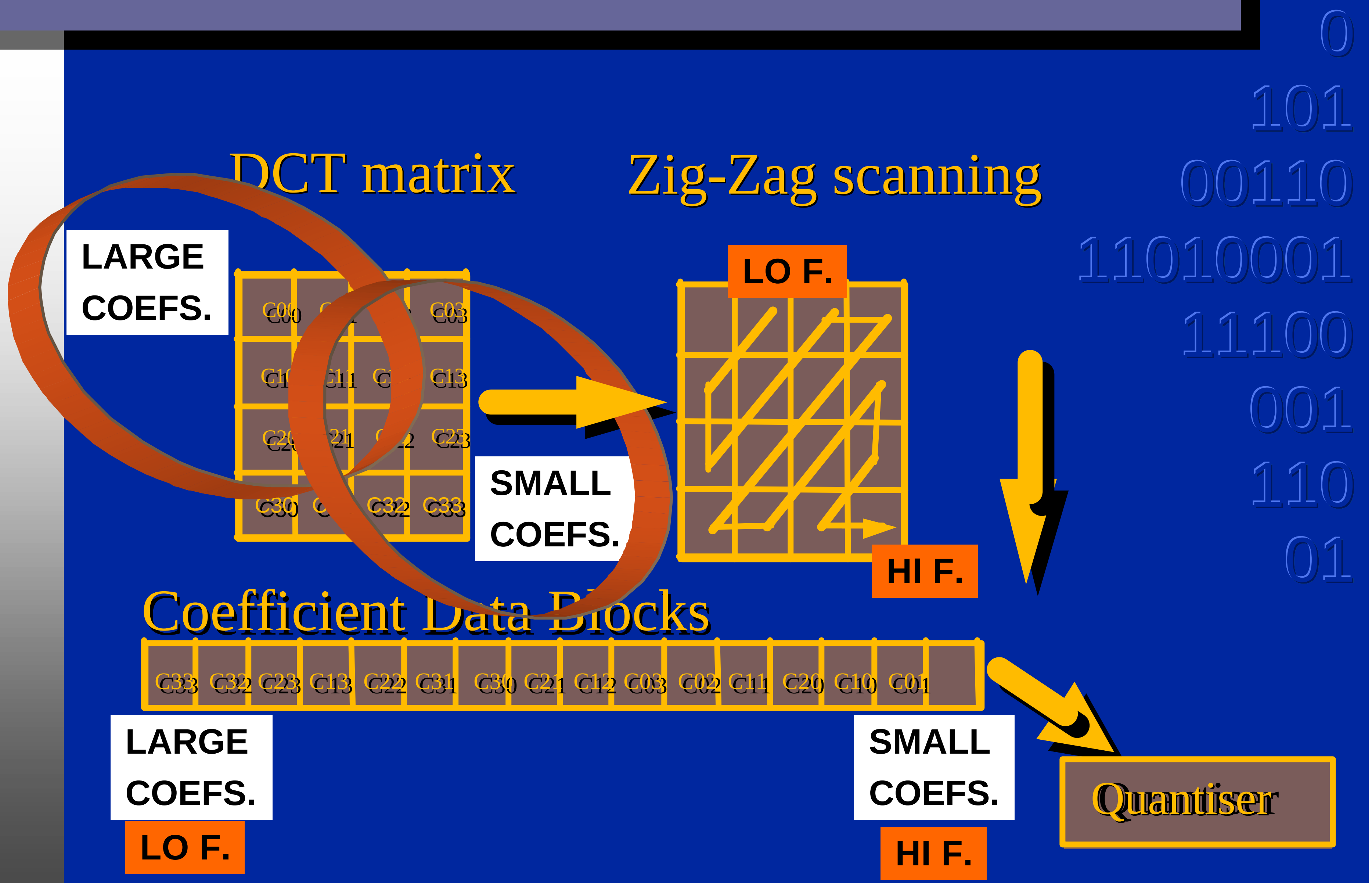


Coefficient Distribution



0
101
00110
11010001
11100
001
110
01

Zig-Zag Scanning and Quantiser



Quantizing Table

Q (u,v) =

16	11	10	16	24	403	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	65	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Quantizing Table

0
101
00110
11010001
11100
001
110
01

DCT Example (Block after Quantizing)

79	0	-1	0	0	0	0	0
-2	-1	0	0	0	0	0	0
-1	-1	0	0	0	0	0	6
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

Zigzag scan data run

79 0 -2 -1 -1 -1 0 0 -1 EOB

0
101
00110
11010001
11100
001
110
01

Entropy

Where's the entropy in this picture?

0
101
00110
11010001
11100
001
110
01

Entropy Coding

- In PCM digital, bit rate set by sampling rate
  - Sampling rate is constant
- Information rate in the signal varies substantially
- Difference between information rate and sampling rate is redundancy.

Entropy

- **Indicates an inactive or static condition.**

- **How to express differences in randomness in terms of measurable qualities. The problem initiated in thermo-dynamics from which the term ultimately derives.**

0
101
00110
11010001
11100
001
110
01

Entropy Coding and the Alphabet

- To set data up for entropy coding, it needs to be structured in a way that likely events are given more importance (and shorter bit codes) than unlikely events. Consider the alphabet:

Alphabetical order

ABCDEFGHIJKLM

NOPQRSTUVWXYZ

Order of Common Usage

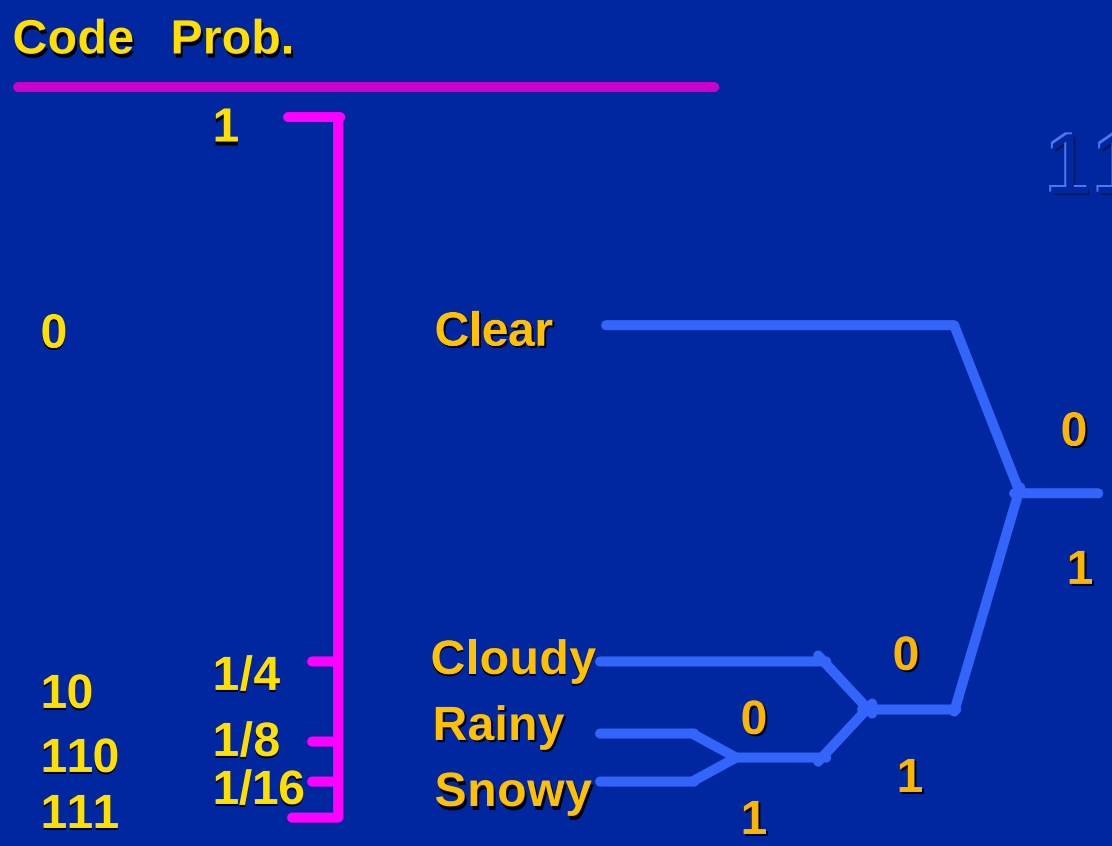
ETAORIHSDLCU

MFPYBWGVJKQZ

Huffman Coding

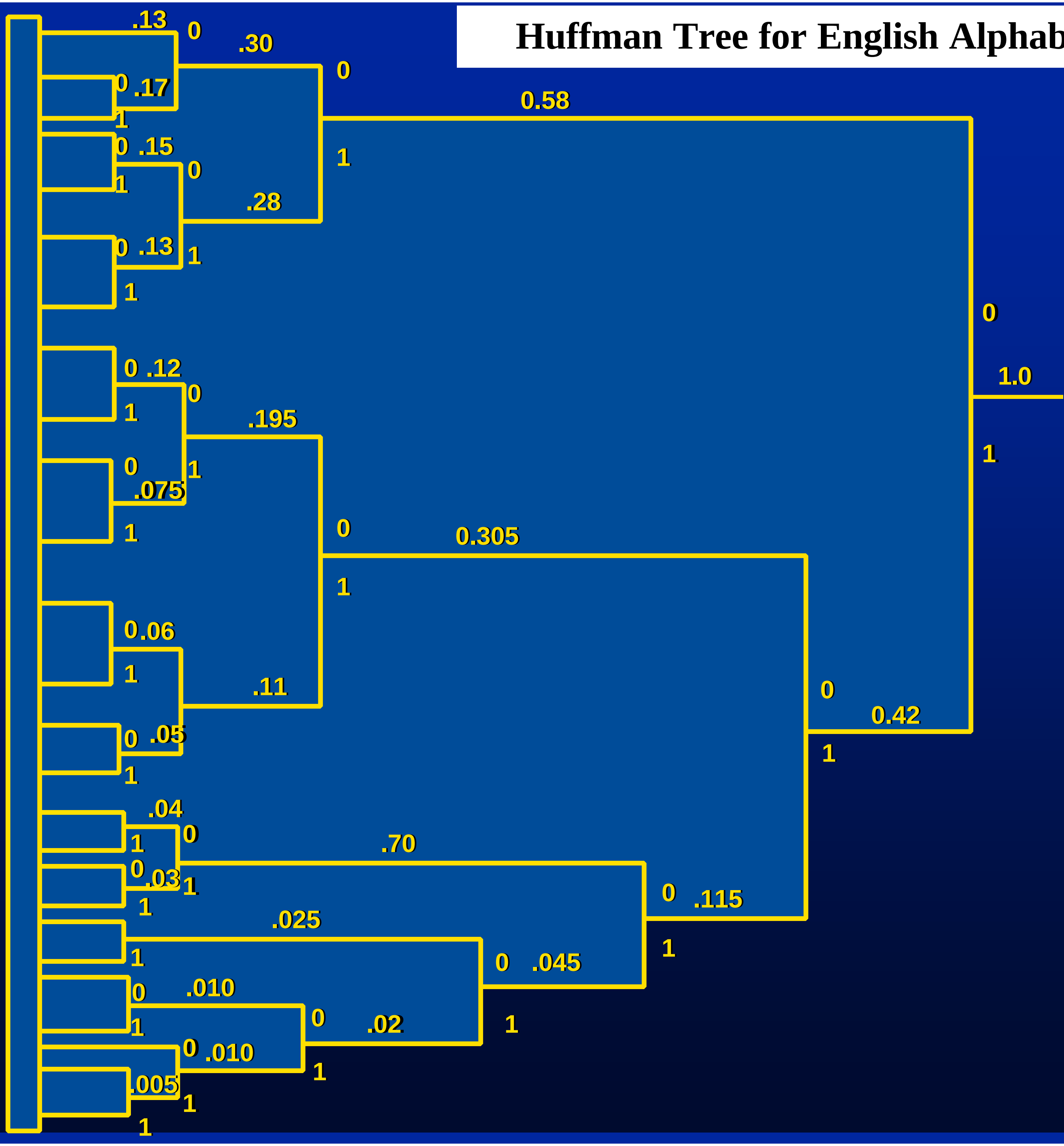
- **Statistical Compression Technique**
  Coding scheme that uses variable length code words to reduce the redundancy and consequently compress the data.

Huffman Coding Tree



Huffman Tree for English Alphabet

000	E	.13
0010	T	.09
0011	A	.08
0100	0	.08
0101	N	.07
0110	R	.065
0111	I	.065
1000	H	.06
10001	S	.06
10010	D	.04
10011	L	.035
10100	C	.03
10101	U	.03
10110	M	.03
10111	F	.02
11000	P	.02
11001	Y	.02
11010	B	.015
11011	W	.015
11100	G	.015
11101	V	.010
111100	J	.005
111101	K	.005
111110	X	.005
1111110	Q	.0025
1111111	Z	.0025



Section 5

Moving Beyond Intra-Frames

Topics

1. Limitations of JPEG.
2. Quality of compression; all comparisons to MPEG-2
3. Introduction to Motion Compensation.
4. MPEG Standards

0
101
00110
11010001
11100
001
110
01

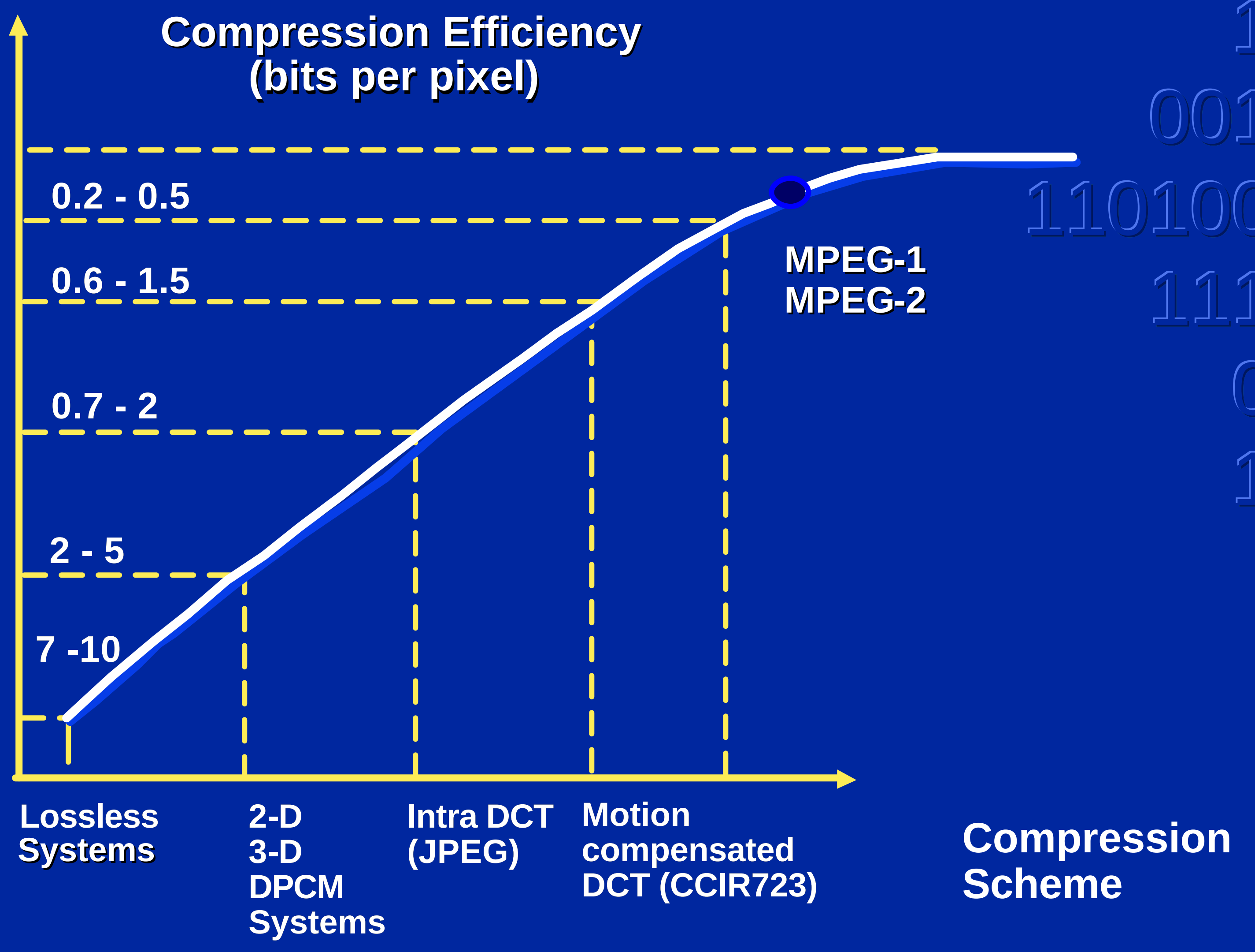
JPEG Quality

Bits/Pixel	Image Quality	Compression Ratio
8.0	Lossless	2:1
1.5	Visually identical to original	11:1
.75	“Good” Quality	22:1
.25	“Useful” Image	64:1
.1	Recognizable Image	160:1

Performance Characteristics

Compression System	Compression Ratio	Storage Cap. (1-hr.)	Picture Quality	Editing Capability	Multigen. Capability
JPEG	2:1	38 GB	Excellent (lossless)	Yes	Yes
JPEG	8:1	7 GB	NTSC (no block)	Yes	Limited
JPEG	22:1	3.5 GB	NTSC (blocking)	Yes	Poor
JPEG	64:1	1.2 GB	Poor	Yes	No
MPEG 1 (1.5 Mbps)	21:1	708.5 MB	VHS	No	No
MPEG 2 (5Mbps)	25:1	2.4 GB	>NTSC	Limited	No
MPEG 2 (10Mbps)	13:1	4.25 GB	Nearly D-1	Limited	Limited

Performance Characteristics



Quality Comparison

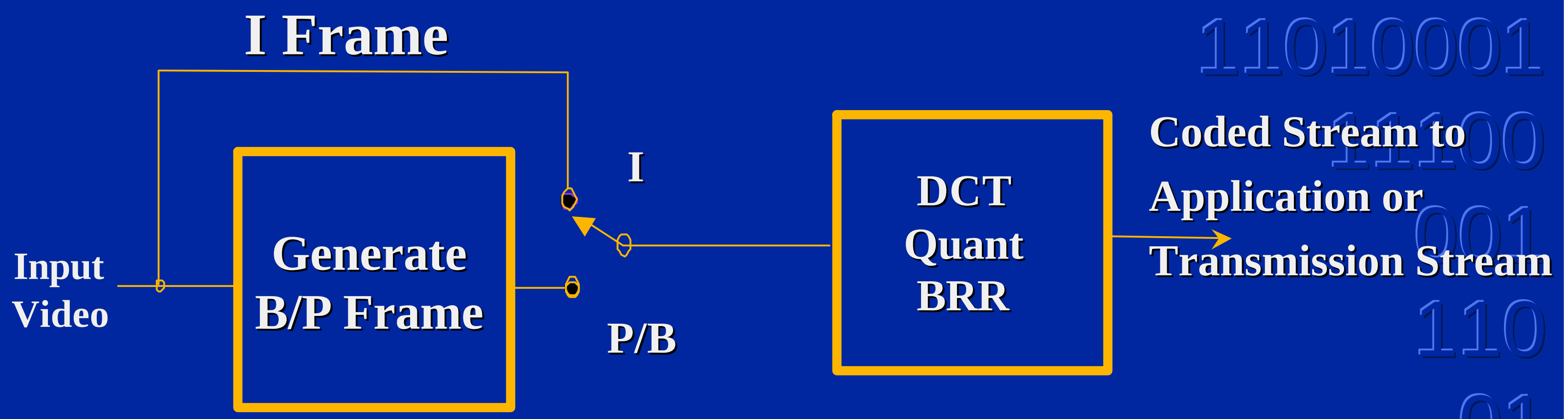
QUALITY	Analog	M-JPEG	DV	MPEG
Acceptable	VHS (or Less)	6 Mbps	DV/DVCAM	1.5Mbps/sec (MPEG-1)
Good	8mm	12 Mbps		3 Mbps
Prosumer	U-matic	20-30 Mbps	25 Mbps	6 Mbps
Production 4:2:2	Beta SP	50+ Mbps	Pro-50 (50 Mbps)	18-50 Mbps

0
101
00110
11010001
11100
001
110
01

Motion Compensated Coding

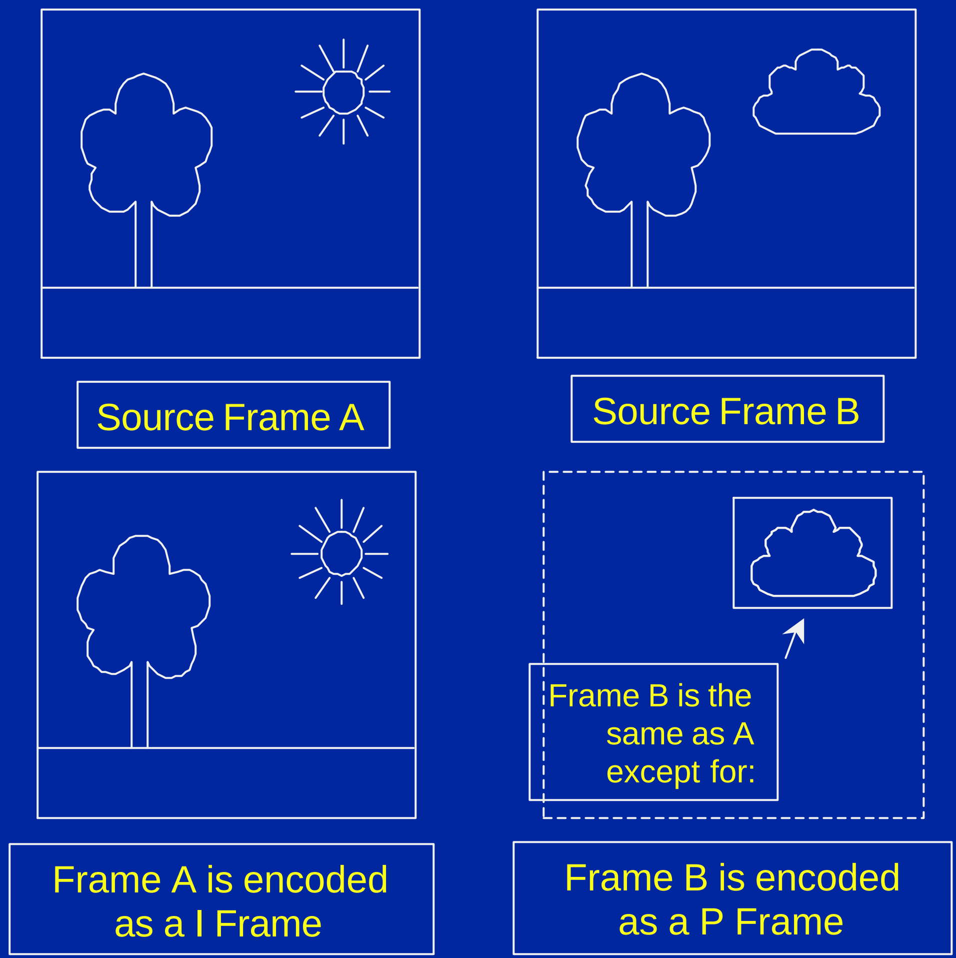
- Data reduction uses both spatial and temporal dimensions.
- Data is reduced across multiple frames
  - Group of Pictures (GOP)
- The larger the GOP the greater compression for equivalent quality.

Inter-Frame Encoding



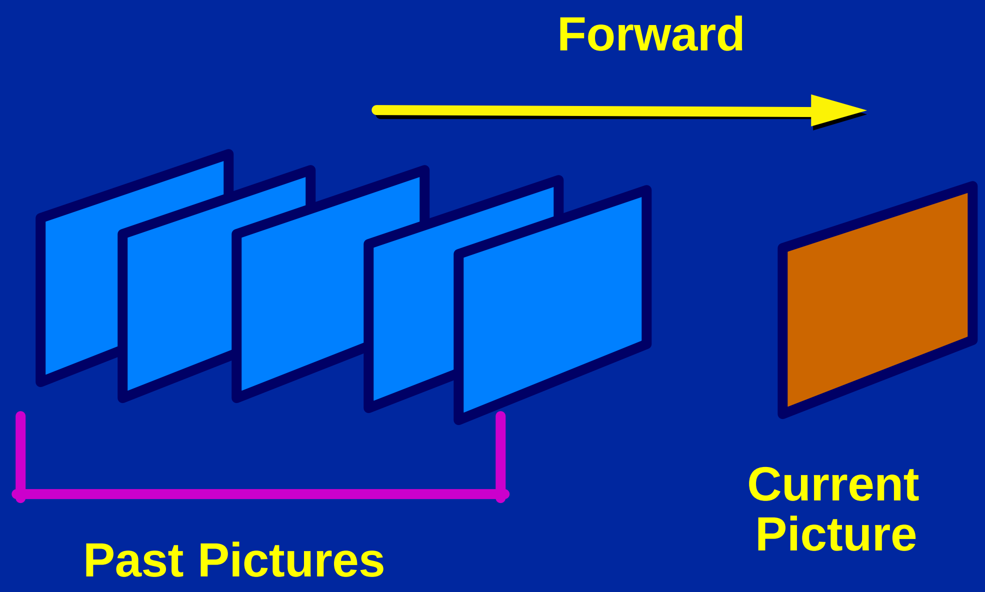
INTER frame Process INTRA frame Process

Prediction Frame



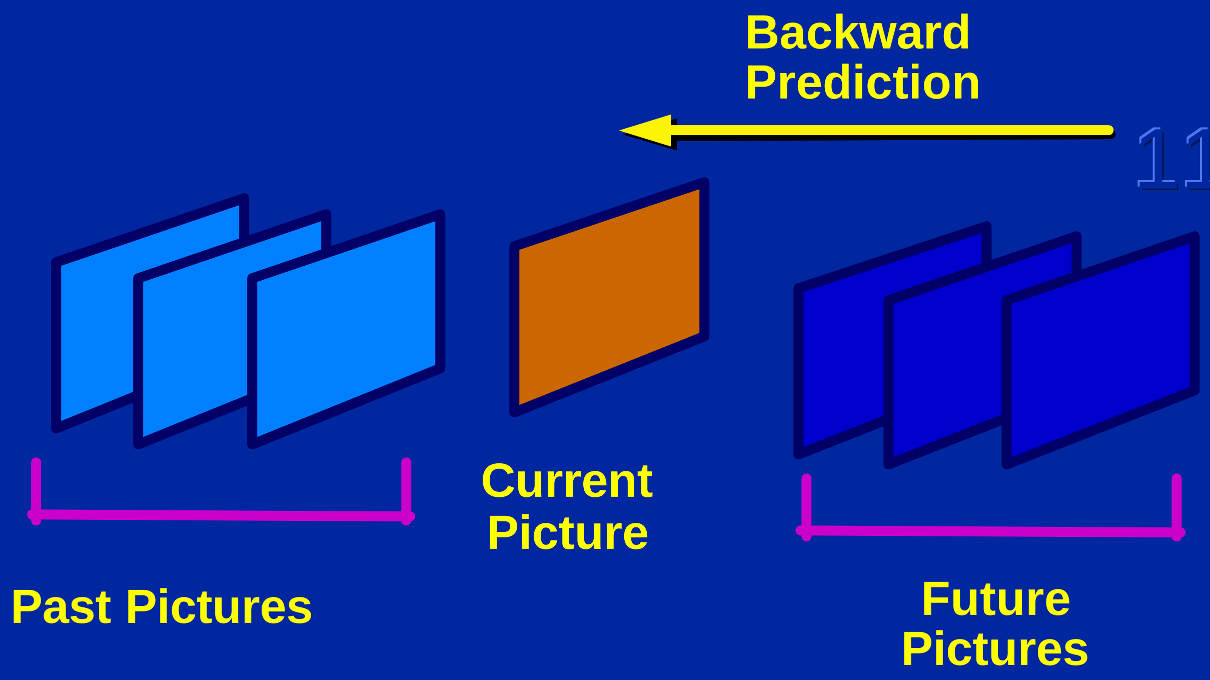
0
101
00110
11010001
11100
001
110
01

What is “Forward Prediction?”



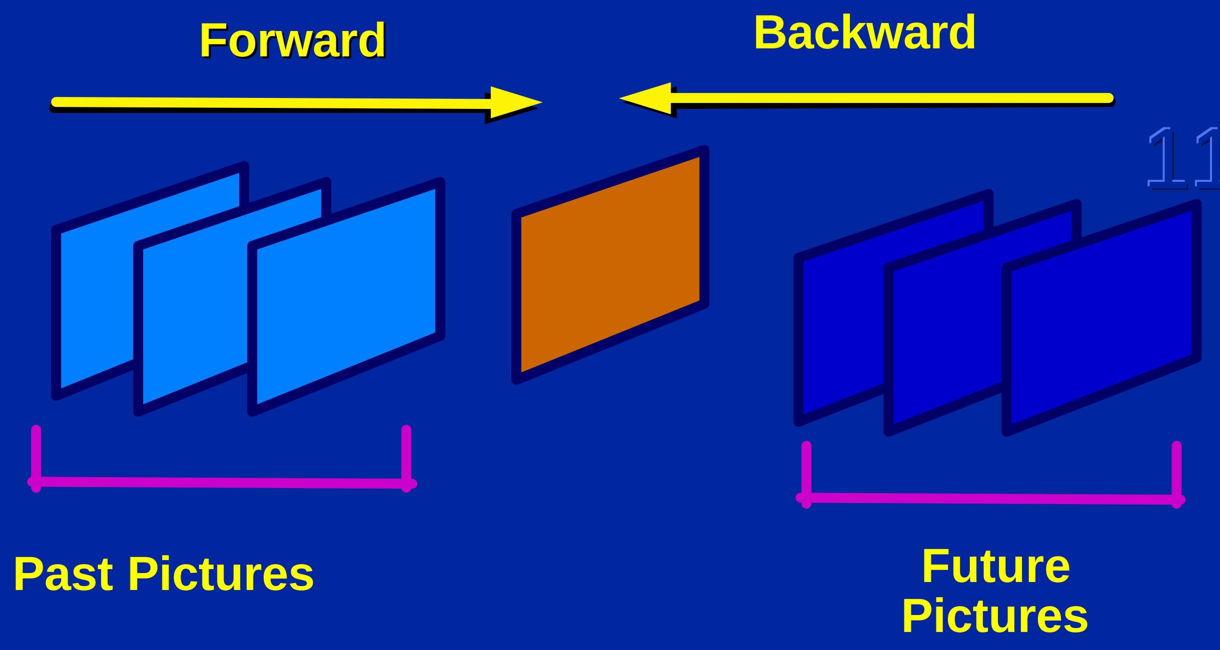
Uses picture information from “Past” pictures to create "Best Guess" for input picture.

What is “Backward Prediction?”



Uses "Future" pictures to create "Best Guess"
for picture to be processed

What is “Bidirectional Prediction?”

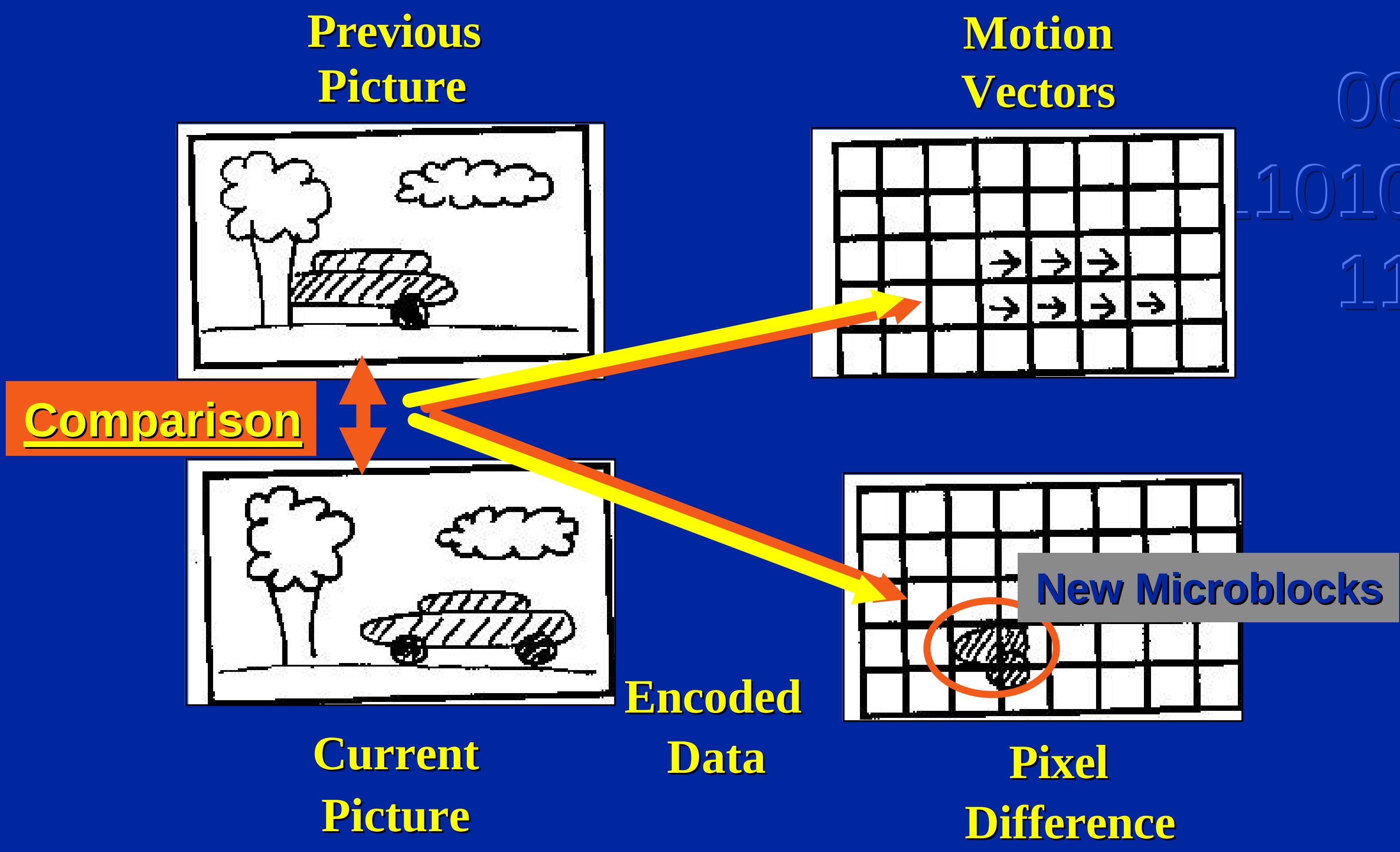


Uses information from "Past" and "Future" pictures to create "Best Guess " of picture to be processed

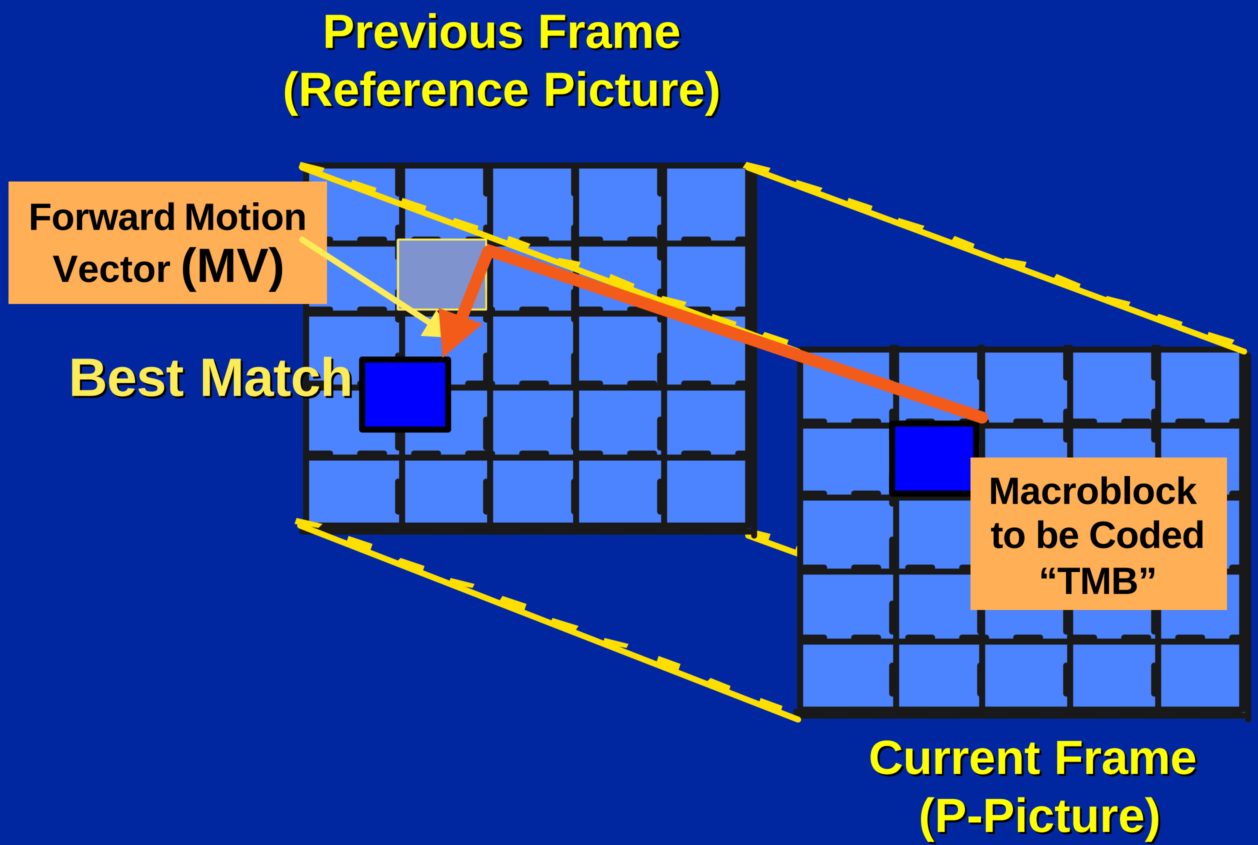
Motion Compensation

- **Object displacements are often not significant between frames**
- **Many regions of the picture have similar pixel values**
- **Static regions are fully redundant**

MPEG Motion Compensation

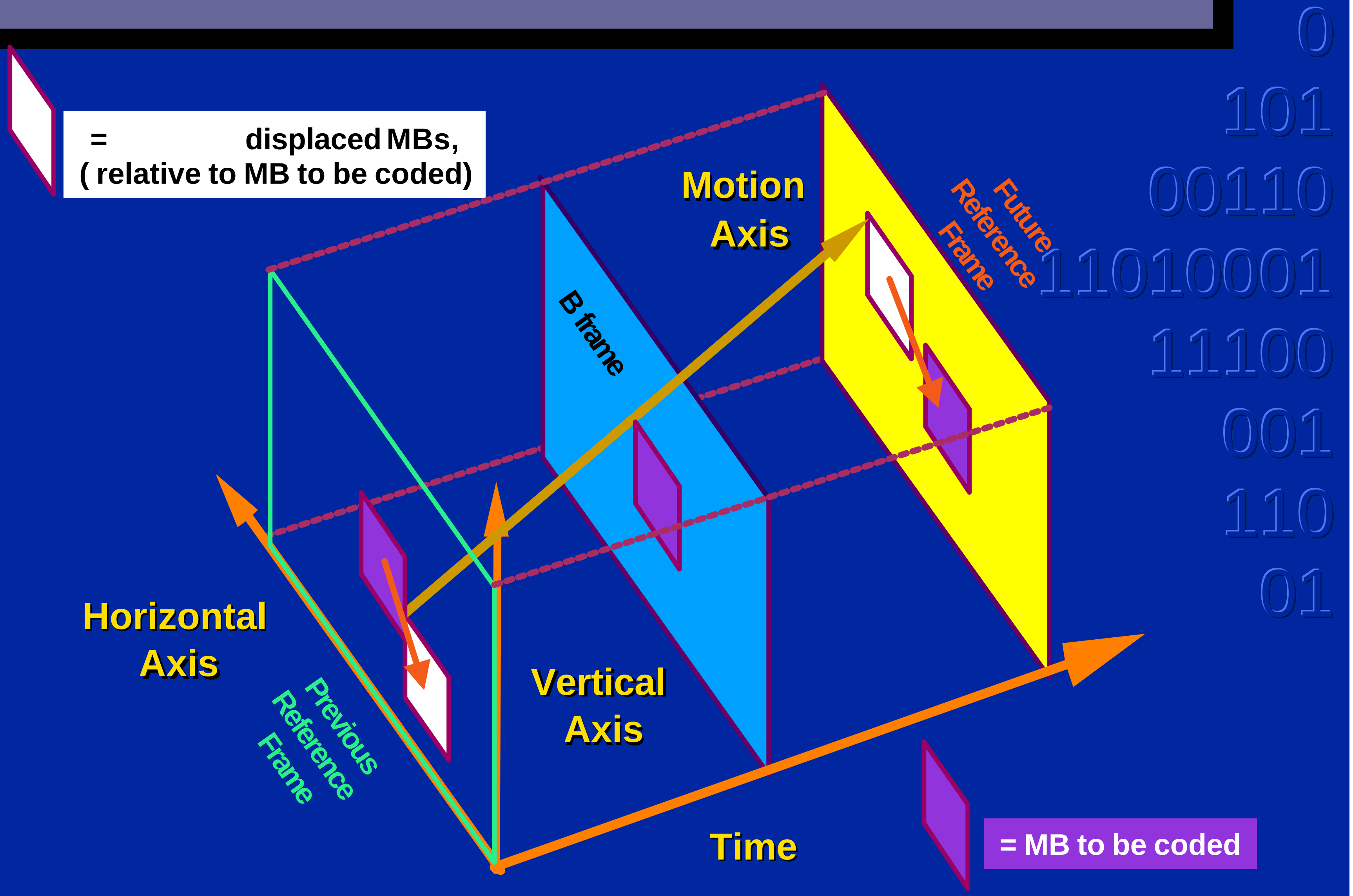


Motion-Compensated Coding



Displaced block in previous frame is used as prediction for block in current frame

MPEG Motion Estimation



Difference Frame

Frame N

Frame N+1

(Frame N) - (Frame N+1)

Disregard

Difference Frame

Frame N

Frame N+1

(Frame N) - (Frame N+1)

Disregard

0
101
00110
10001
11100
001
110
01

Difference Frame

Frame N

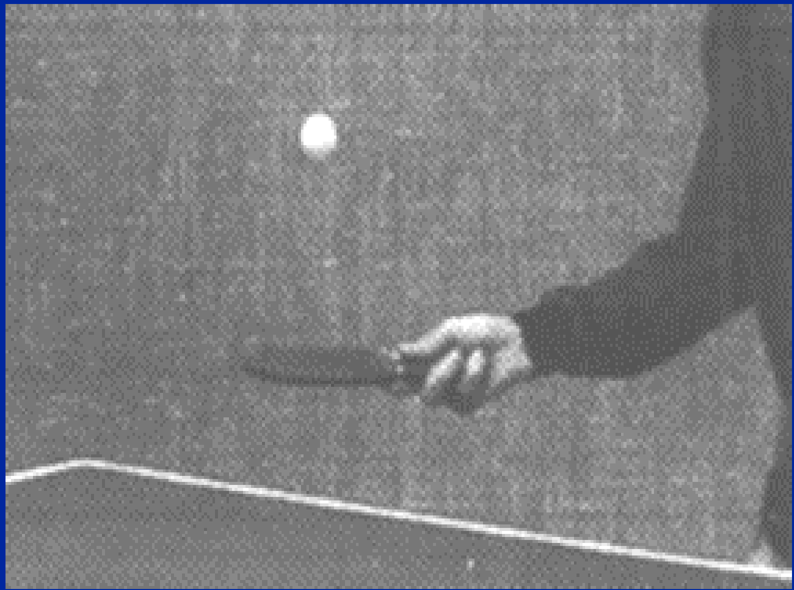
Frame N+1

(Frame N) - (Frame N+1)

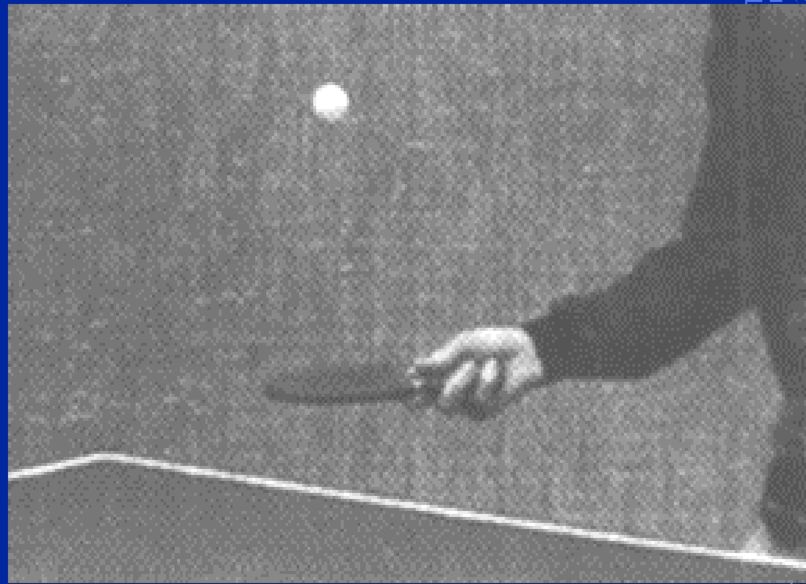
Disregard

0
101
00110
010001
11100
001
110
01

Motion Prediction

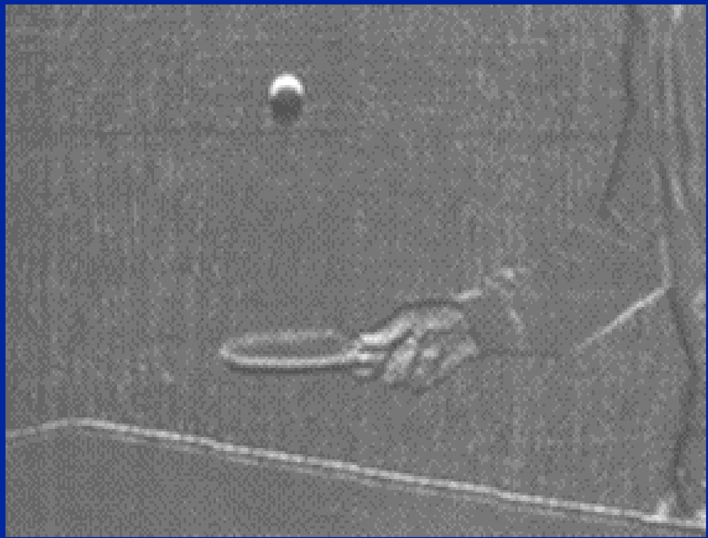


Frame N



Frame N+1

Motion Prediction

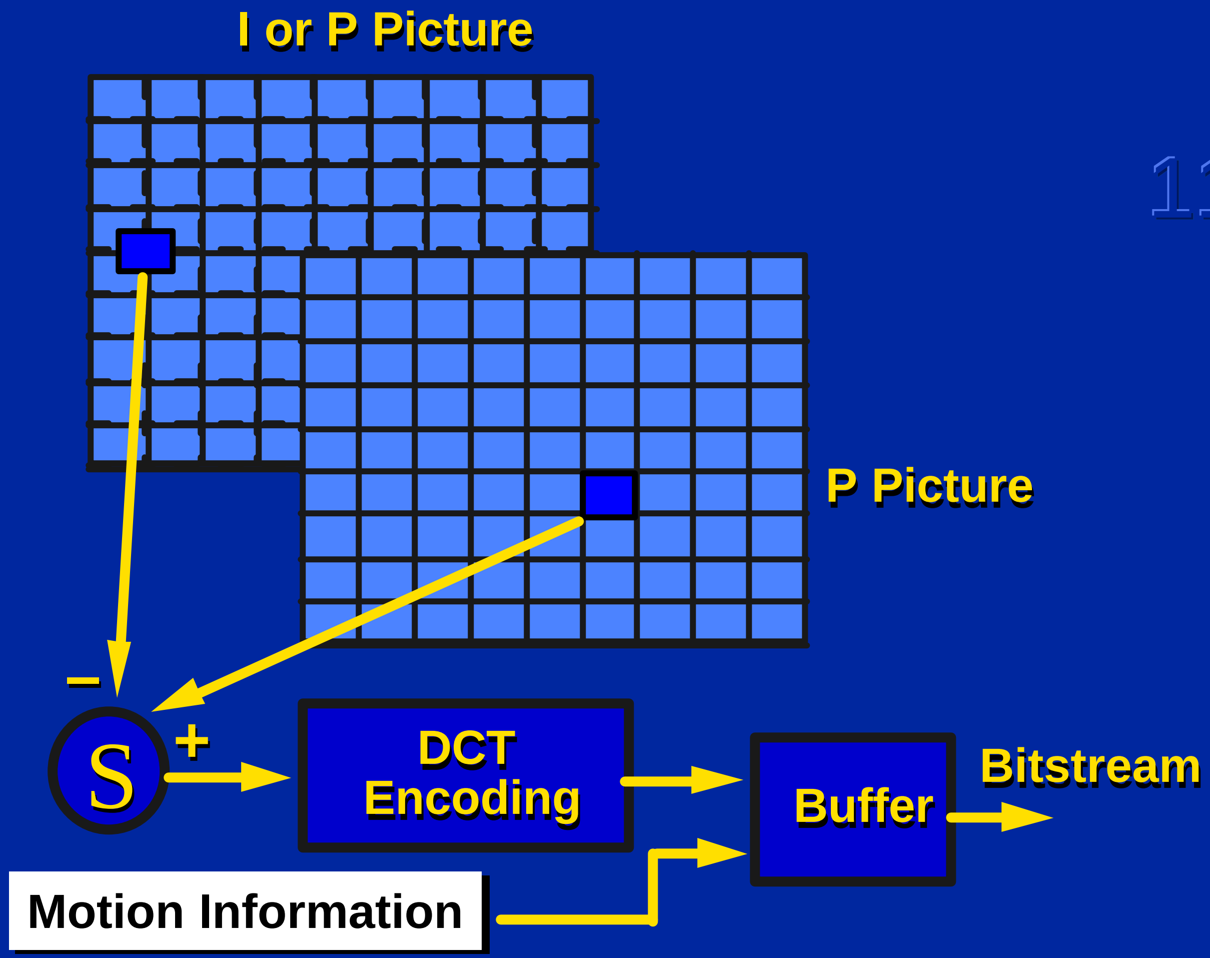


Difference Frame
Without Motion Prediction

Difference Frame
With Motion Prediction

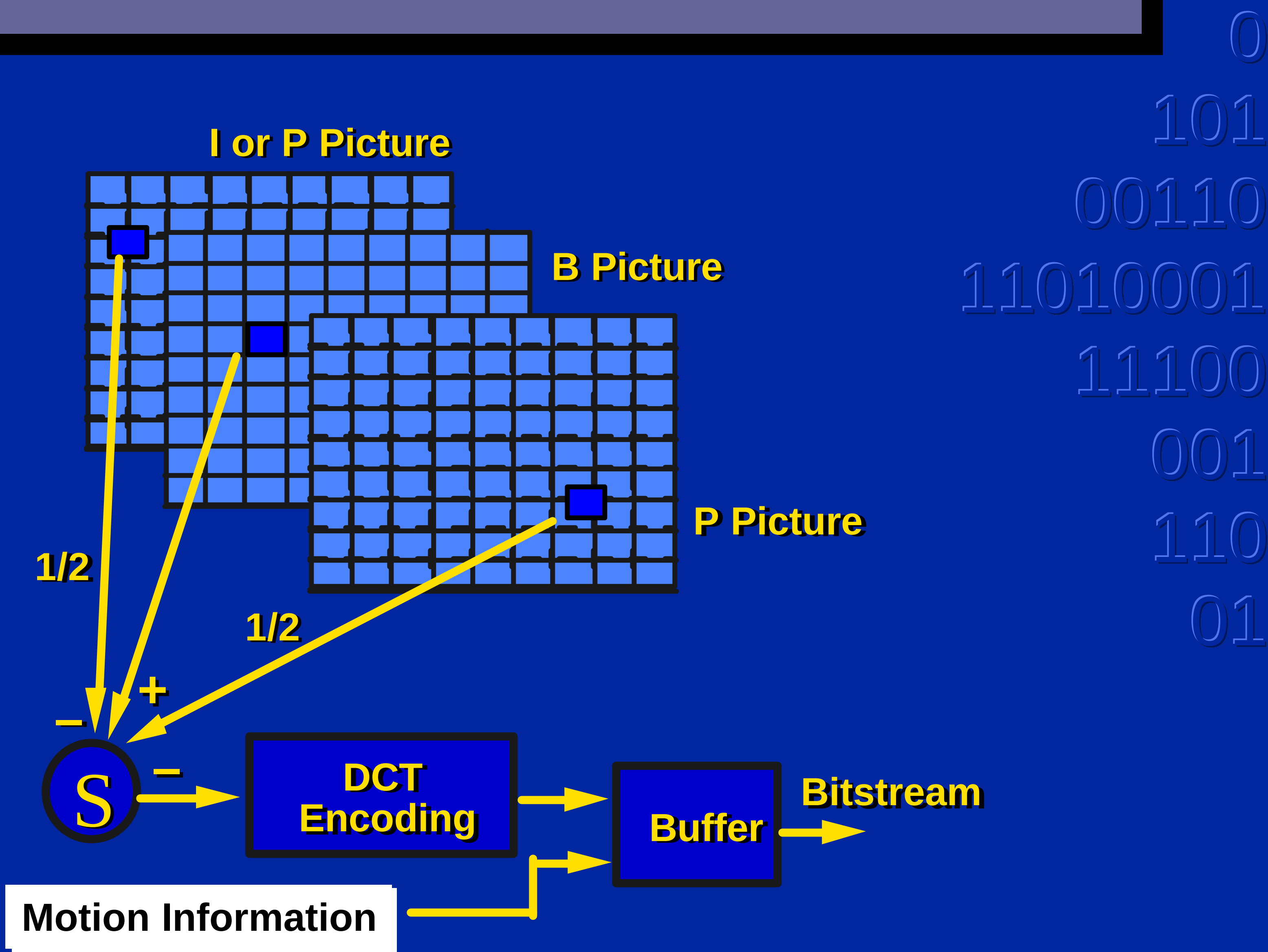
Graphic taken from: “MPEG Video Compression Standard”
Mitchell, Pennebaker, Fogg & LeGall

P-Picture “P” = Predictive Coding

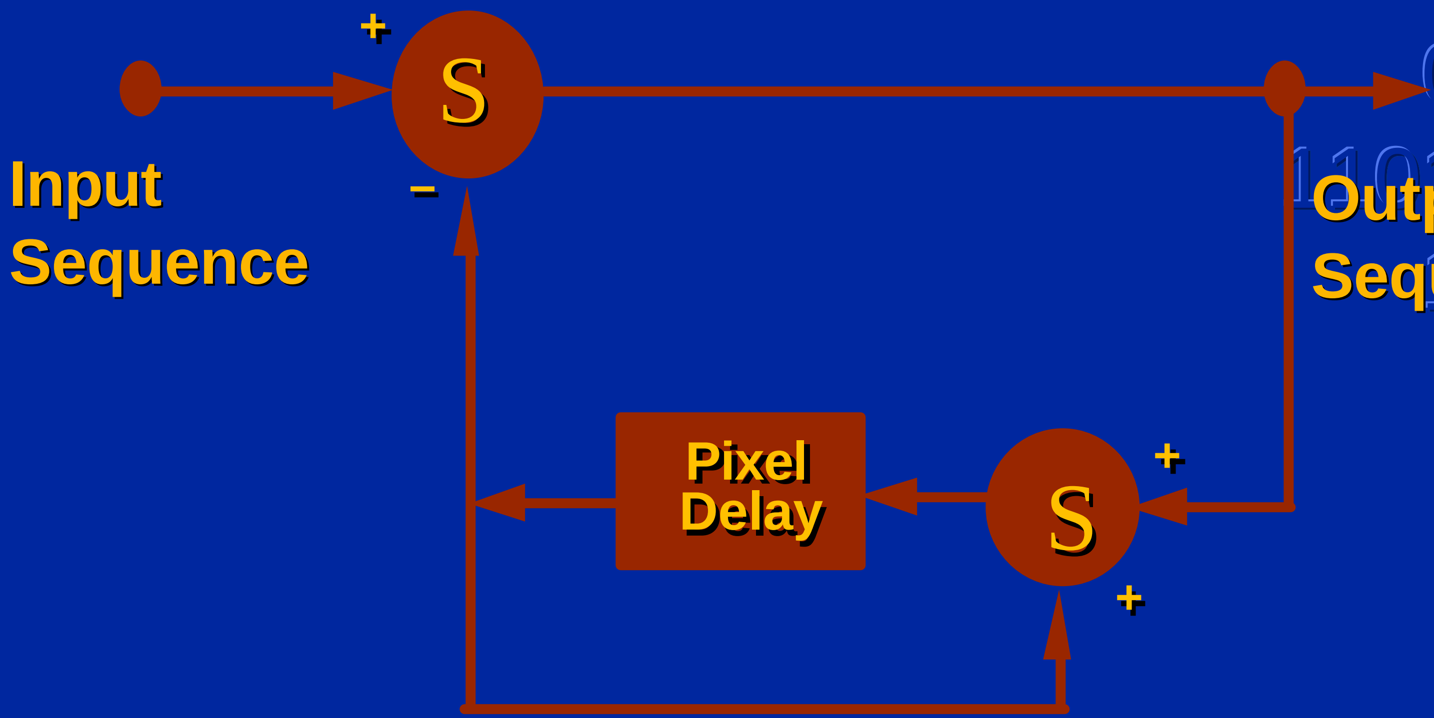


0
101
00110
11010001
11100
001
110
01

B-Picture "B" = Predictive Coding



Differential Encoder



Where does MPEG come In?

- Began in 1988
  - Open process, broad participation
    - European, American, Japanese
  - Goal: World accepted standard
- MPEG -1 specification 11/91
- MPEG-2 specification 11/93

Two Flavors of MPEG

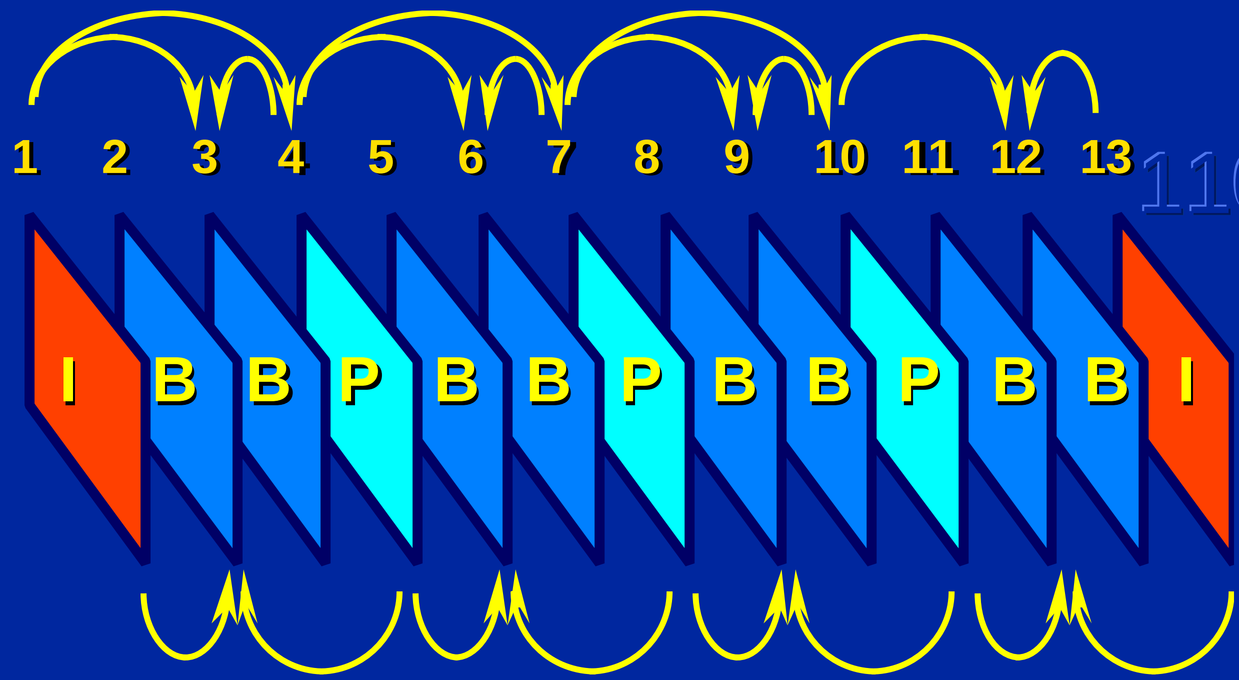
- **MPEG-1**
  - Delivers *VHS* quality
  - CD-ROM and consumer use
  - SIF subsampling 2:1:1
  - 352x240 resolution
- **MPEG-2**
  - better quality
  - several bit-rates and compression ratios
  - *last mile* delivery to users
  - *Profiles* and *Levels* for flexible applications

MPEG-2 Standard

- 1-2 Mbps
  - low resolution “VHS” quality
  - CD-ROM, multimedia
- 3-10 Mbps
  - NTSC, Pal, Cable, DBS: CCIR 601 quality
  - DVD
  - DTV & DVB
  - 4-8 digital TV channels per analog NTSC
- 15-30 Mbps
  - HDTV: terrestrial

0
101
00110
11010001
11100
001
110
01

MPEG Group of Pictures



I : Intra Frame

P: Prediction Frame

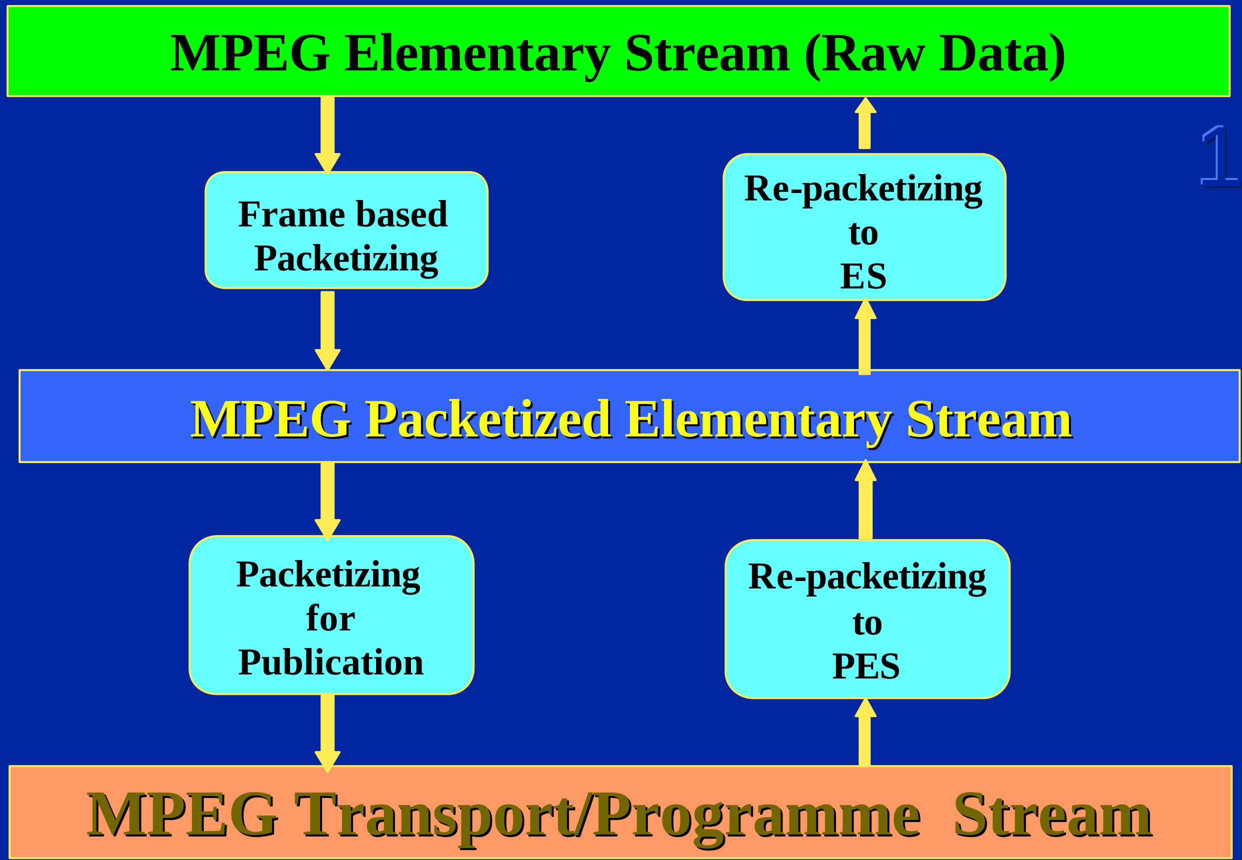
B: Bidirectional Frame

MPEG Stream

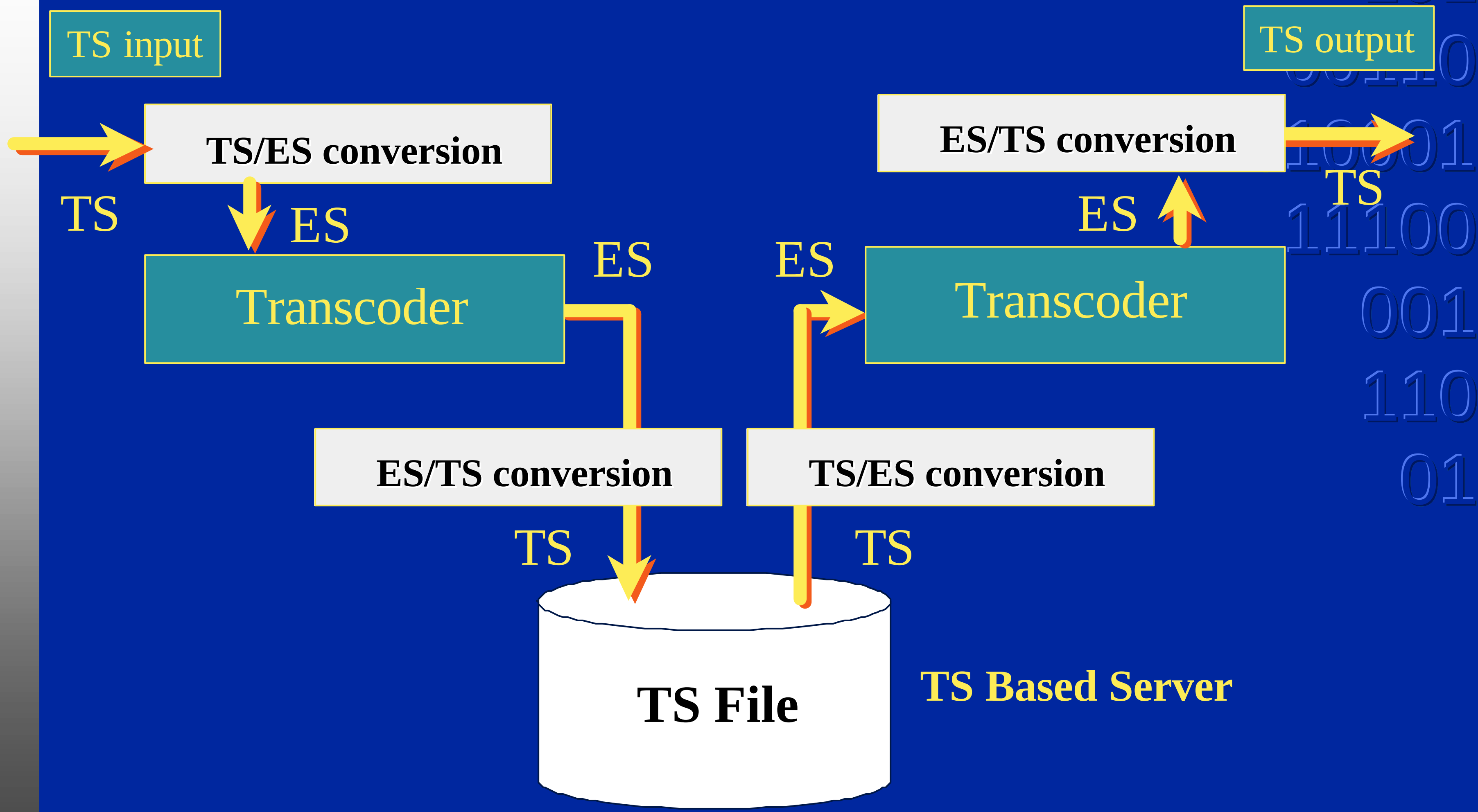
Several Streams for MPEG

- Elementary Stream (MPEG raw data)
  - Over SDTI being standardized by SMPTE
- Packetized Elementary Stream (PES)
  - Frame-based, packetized MPEG stream with time stamp
- Programme Stream
  - An interleave of audio/video packets for use in single-stream applications (e.g. DVD)
- Transport Stream
  - Multi-stream, re-Packetized PES for Transmission
  - DVB-ASI is standardized and mainly used in Europe

MPEG ES and TS



MPEG ES and TS



MPEG ES and TS

- **MPEG Elementary Stream**
  - Less process
    - Less cost
    - Less time delay
  - Variable Speed Operation
  - Suitable for Operation in Station
  - ES over SDTI now standardized by SMPTE
- **MPEG Transport/Programme Stream**
  - Suitable for Publication

0
101
00110
11010001
11100
001
110
01

MPEG-2 Terms

PTS	Presentation Time Stamp
DTS	Decoded Time Stamp
PSI	Program Specific Information
PAT	Program Association Table
PID	Packet Identifier
CAT	Conditional Access Table
NIT	Network Information Table
PCR	Program Clock Reference

MPEG Profiles and Levels

- Standard is generic and broad in application
  - supports a wide range of applications, bit rates, resolutions and services
- Single syntax integrates various applications
- "Profiles" and "Levels" establish subsets of "full syntax"
  - helps avoid requirement for single highly complex full syntax
- These define decoder requirements

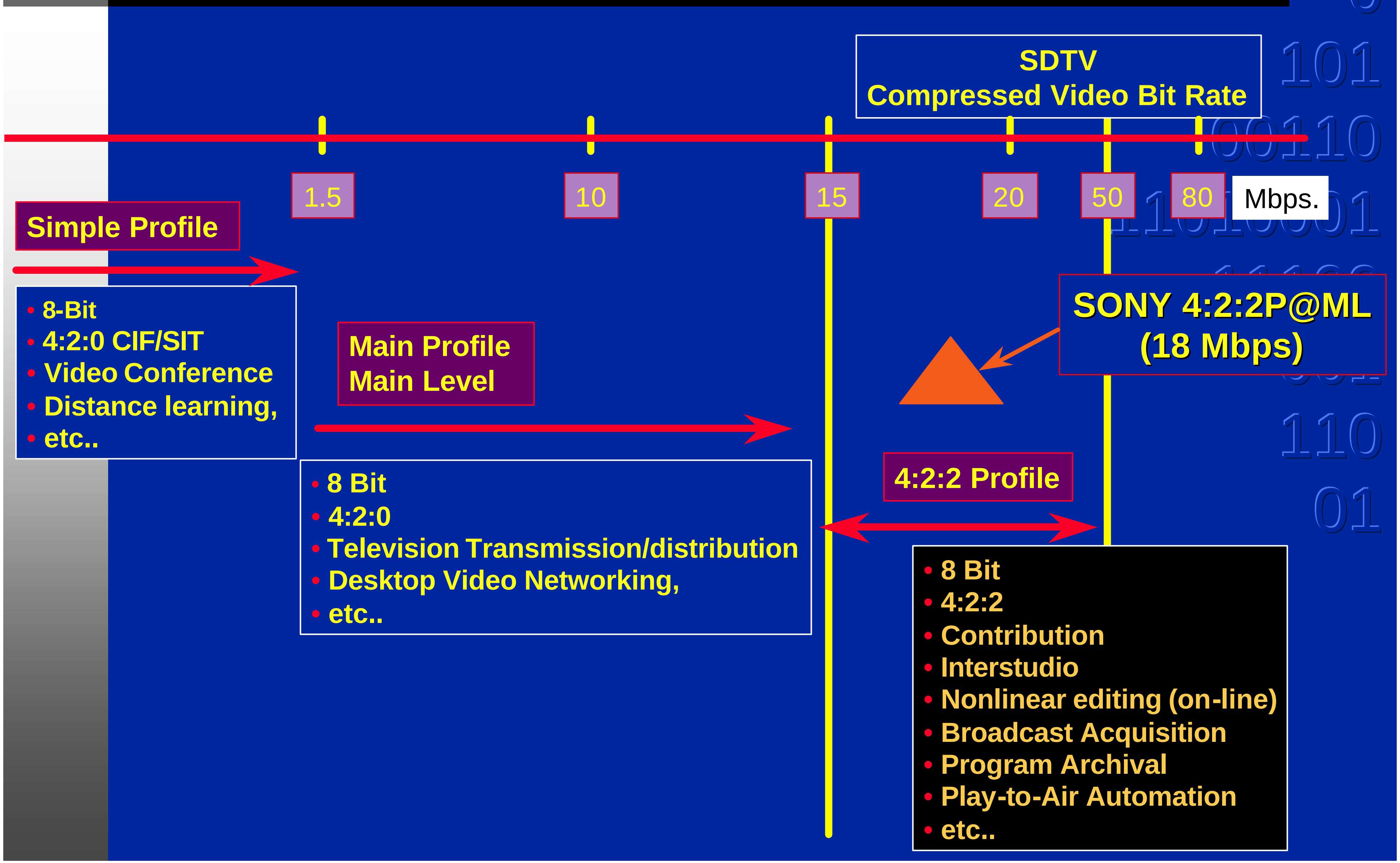
MPEG Profiles "Continued"

LEVEL	Profile: Simple	Main	SNR	Spatial	High	4:2:2
FRAME TYPES →	I & P	I, P & B	I, P & B	I, P & B	I, P & B	I, P & B
CHROMA SAMPLING →	4:2:0	4:2:0	4:2:0	4:2:0	4:2:2 & 4:2:0	4:2:2 & 4:2:0
HIGH Samples/Line Lines/Frames Frames/Sec Max Bit-Rate (Mbps)		1920 1152 60 80			1920 1152 60 100	1920 1088 60 300
HIGH 1440 Samples/Line Lines/Frames Frames/Sec Max Bit-Rate (Mbps)		1440 1152 60 60		1440 1152 60 60	1440 1152 60 80	
MAIN Samples/Line Lines/Frames Frames/Sec Max Bit-Rate (Mbps)	720 576 30 15	720 576 30 15	720 576 30 15	720 576 30 15	720 576 30 20	720 576 30 50
LOW Samples/Line Lines/Frames Frames/Sec Max Bit-Rate (Mbps)		352 288 30 4				

HDTV

SDTV

MPEG-2 Operating Hierarchy



Review/Summary of MPEG Basics

- DCT for spatial compression
- GOP for temporal compression
- Heavy exploitation of temporal redundancy
- “Synthesized” pictures between periodic “anchor” frames

0
101
00110
11010001
11100
001
110
01

But that's just the start......

0
101
00110
11010001
11100
001
110
01