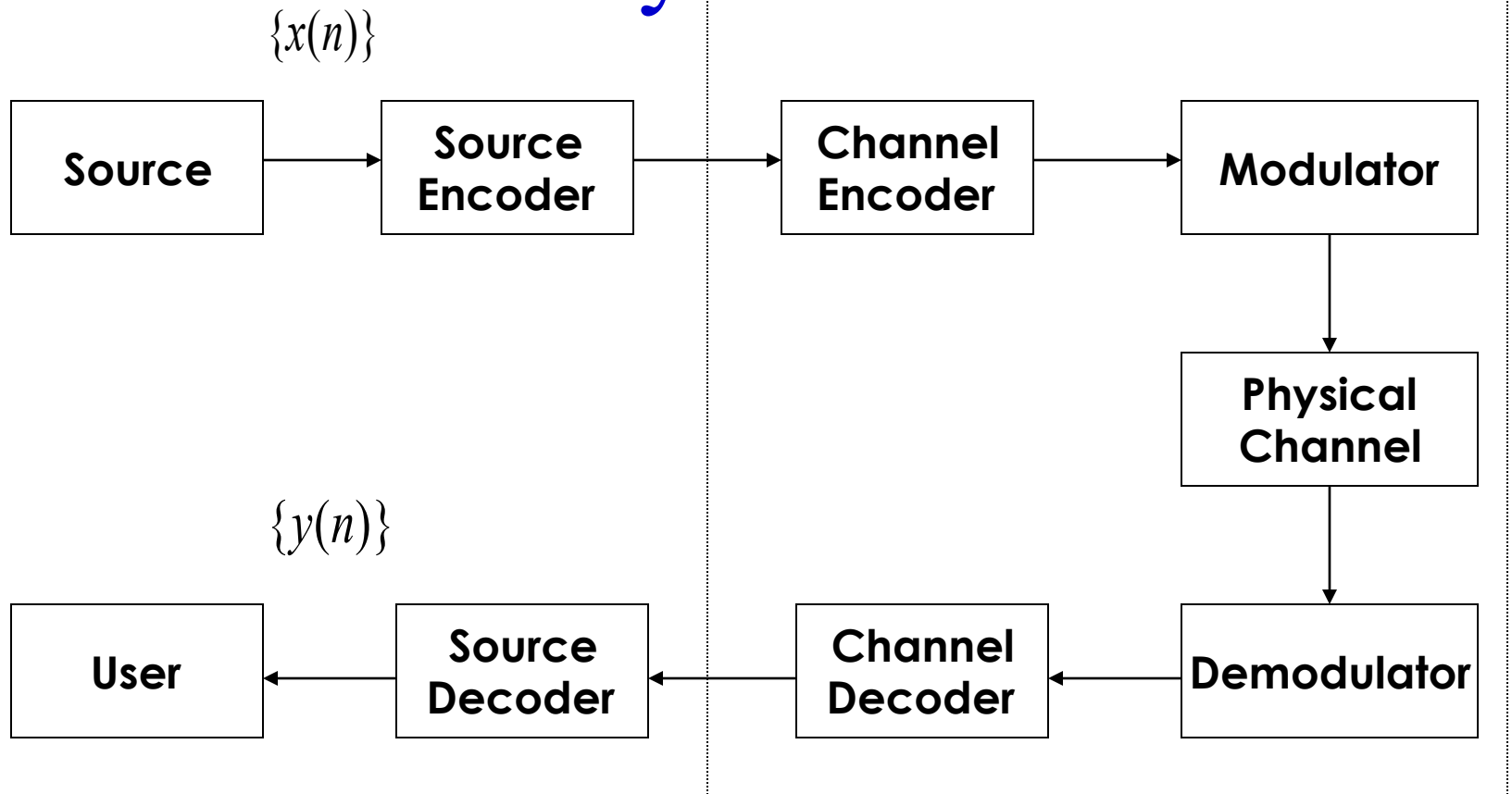


ECE 242 Signal Compression

Introduction to Speech, Audio, Still
Image, and Video Compression

Jerry D. Gibson

Typical Digital Communication System



What is Compression?

- Represent a source in digital form with as few bits as possible while still providing an acceptable reproduction of the original

Synonyms for Data Compression

- Signal Compression
- Signal Coding
- Source Coding
- Source Coding with a Fidelity Criterion
- Lossy (Noisy) Source Coding
- Lossless (Noiseless) Source Coding
- Data Compaction

More Synonyms

- Redundancy Removal
- Bandwidth Compression

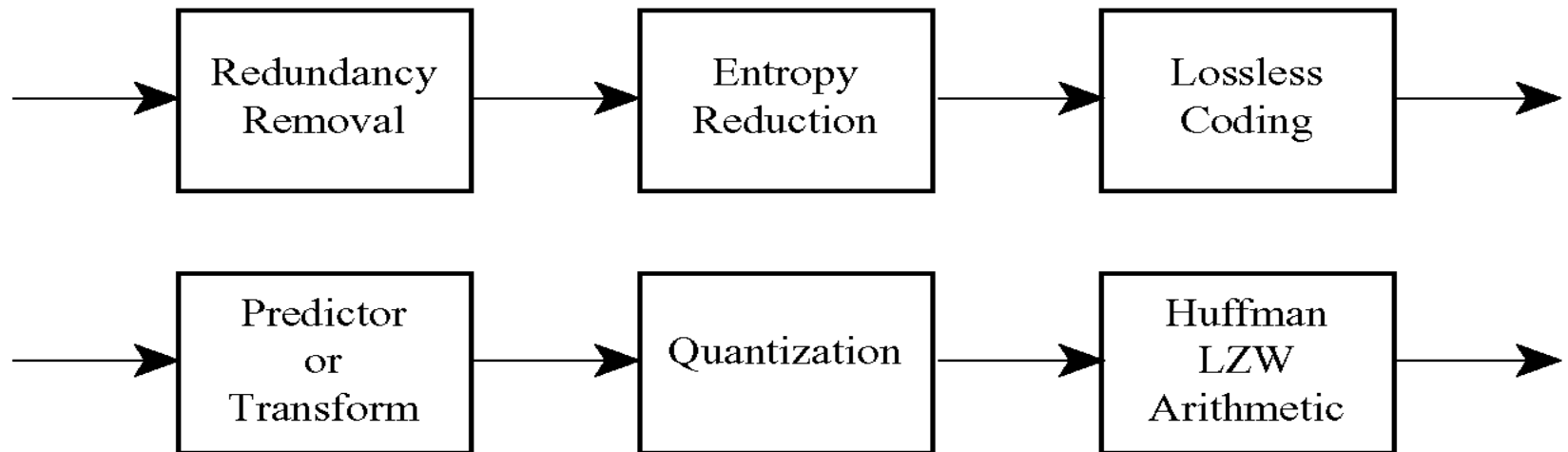
Components of a Compression Problem

- Source
- Rate
- Distortion Measure
- And Complexity

Rate, Bits, and Bandwidth

- Required network bandwidth set by transmitted bit rate
- Bit Rate in *bits/sec* = *bits/sample* x *samples/sec*
- Sampling rate determined by source bandwidth

Major Steps in Data Compression



Speech and Audio Coding Bandwidths

- Narrowband Speech — 200 to 3400 Hz
- Wideband Speech — 50 to 7000 Hz
- Wideband Audio — 20 to 20,000 Hz
 - Number of channels
 - Stereo
 - Five channel surround

Approximate Bit Rates for Uncompressed Sources

| | |
|-----------------------------------|---|
| Telephony (200–3400 Hz): | $8000 \text{ samples/second} \times 12 \text{ bits/sample} = 96 \text{ kbps}$ |
| Wideband speech (50–7000 Hz): | $16,000 \text{ samples/second} \times 14 \text{ bits/sample} = 224 \text{ kbps}$ |
| Wideband audio (20–20,000 Hz): | $44,100 \text{ samples/second} \times 2 \text{ channels} \times 16 \text{ bits/sample} = 1.412 \text{ Mbps}$ |
| Images: | $512 \times 512 \text{ pixel color image} \times 24 \text{ bits/pixel} = 6.3 \text{ Mbits/image}$ |
| Video: | $640 \times 480 \text{ pixel color image} \times 24 \text{ bits/pixel} \times 30 \text{ images/second} = 221 \text{ Mbps}$ |
| HDTV: | $1280 \times 720 \text{ pixel color image} \times 60 \text{ images/second} \times 24 \text{ bits/pixel} = 1.3 \text{ Gbps}$ |

Networks and Network Services

| | |
|--------------------|---|
| POTS | 28.8-56 Kbits/s |
| ISDN | 64-128 Kbits/s |
| ADSL | 1.544-8.448 Mbits/s (downstream) 16-640 Kbits/s (upstream) |
| VDSL | 12.96-55.2 Mbits/s |
| CATV | 20-40 Mbits/s |
| OC-N/STS-N | N x 51.84 Mbits/s |
| Ethernet | 10 Mbits/s |
| Fast Ethernet | 100 Mbits/s |
| Gigabit Ethernet | 1,000 Mbits/s |
| FDDI | 100 Mbits/s |
| 802.11(wireless) | 1, 2, 5.5, 11, and 22 Mbits/s in 2.4 GHz band |
| 802.11 a(wireless) | 6-54 Mbits/s in 5GHz band |

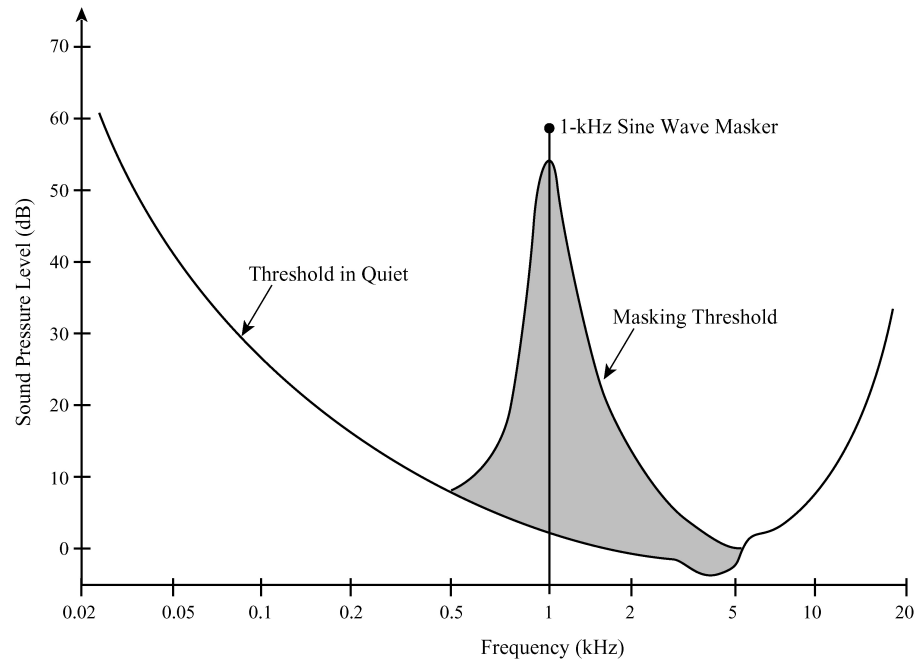
Reduce Source Bit Rates but Keep Quality

- Source (Speech, Audio, Still Images, Video)
Compression
- What is Compression?
- **Goal: Represent a source in digital form with as few bits as possible while still providing an acceptable reproduction of the original**

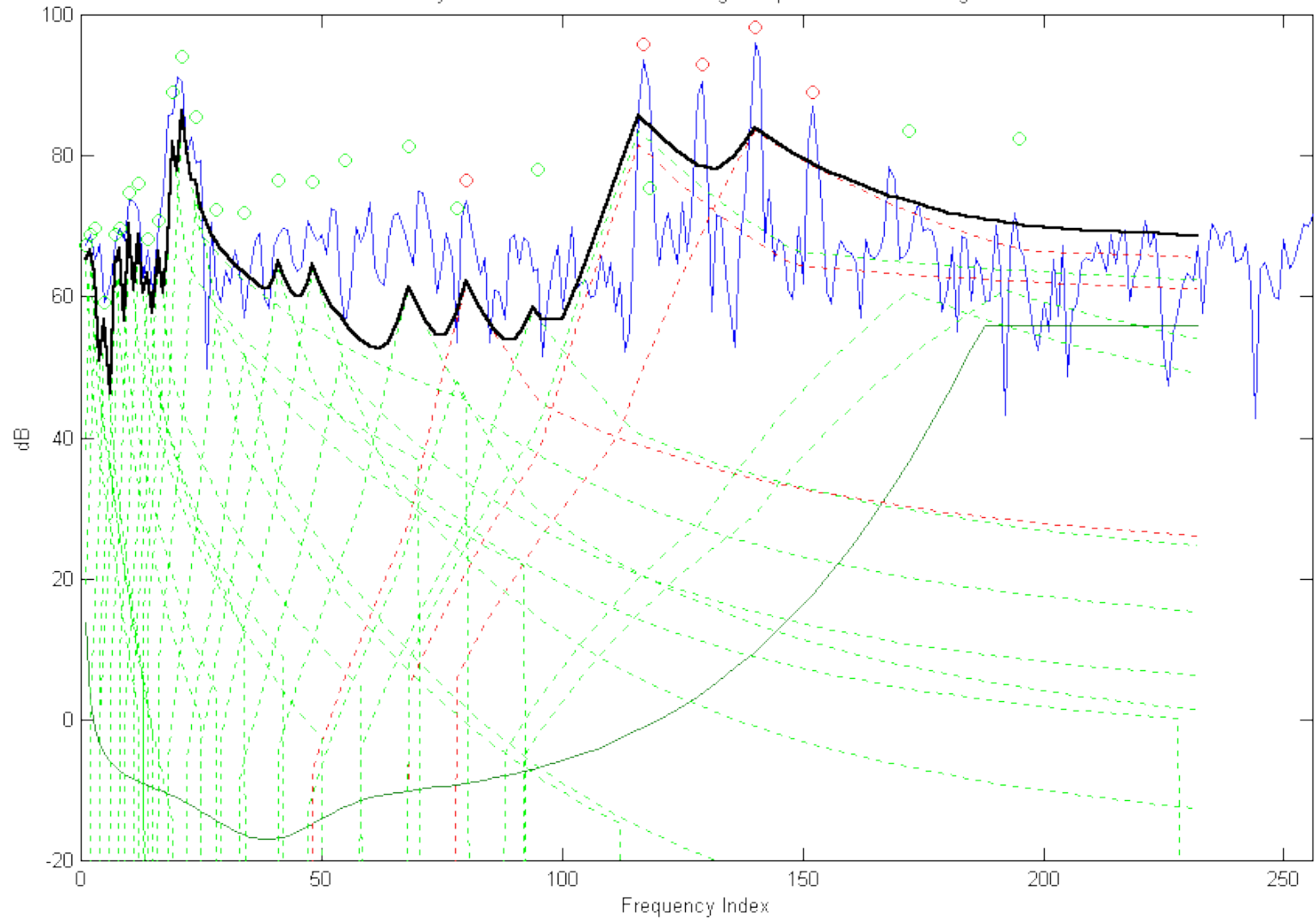
Design Distortion Measures

- Mean Squared Error
 - Mathematically Tractable
 - Not Necessarily Perceptually Meaningful
 - Important for Initial Rankings
- Frequency-Weighted Squared Error
- Perceptually-Based Distortion Measures

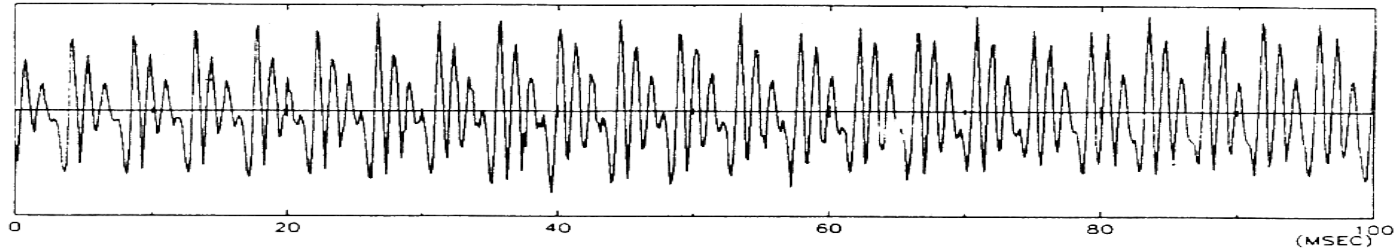
THRESHOLD IN QUIET AND MASKING THRESHOLD



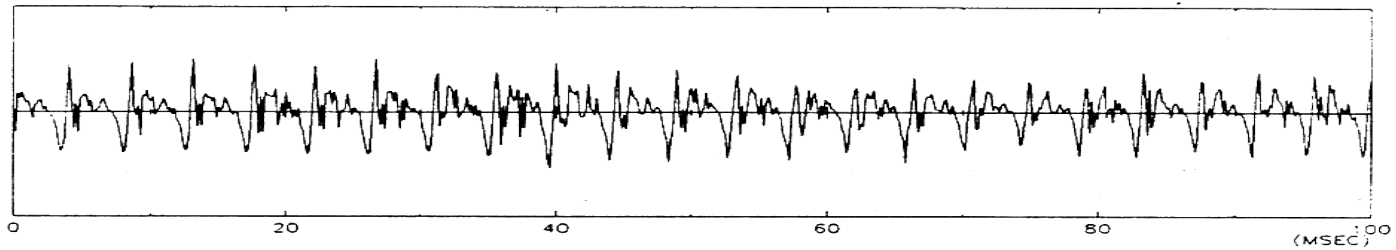
MPEG-2 Psychoacoustic Model 1: Masking Components and Masking Threshold



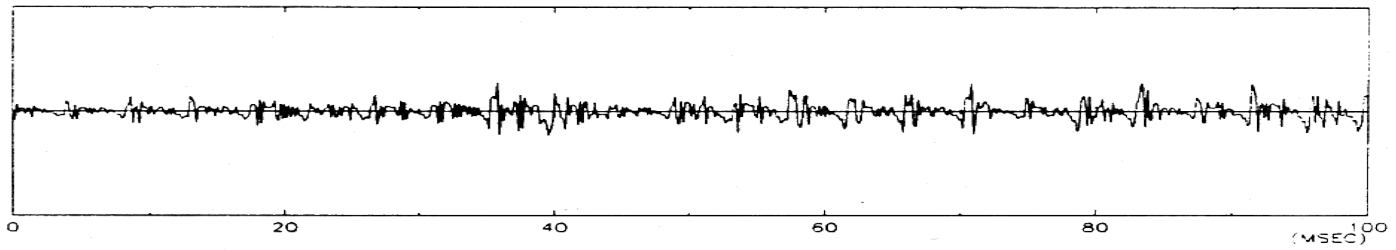
Redundancy Removal From Speech Signals



(a) Speech Waveform

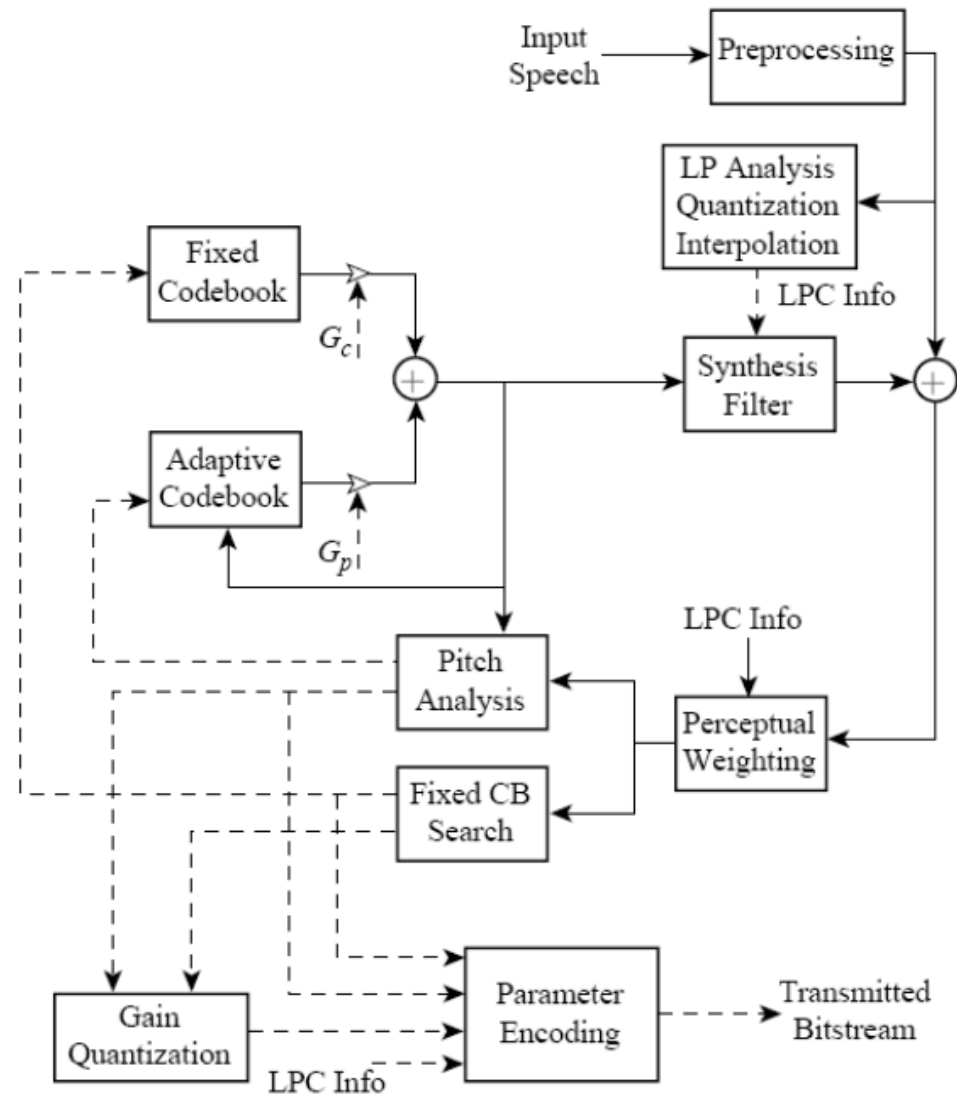


(b) Short term removed signal

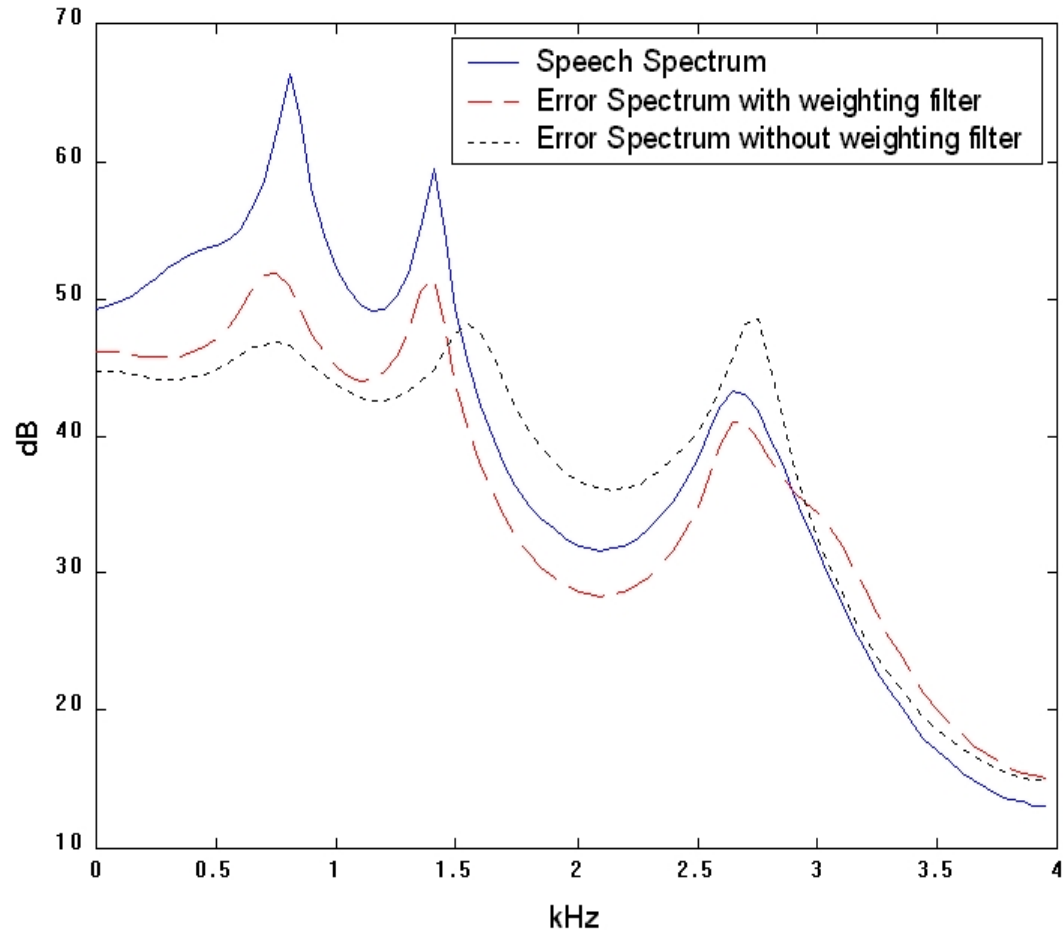


(c) Long and short term removed signal

Code Excited Linear Prediction



CELP Perceptual Weighting



Performance Evaluation

- Speech
 - Listening Tests, including
 - MOS
 - DRT
 - DAM
 - PESQ-MOS
- Audio—Listening tests—transparency
- Images and Video--Viewing

Applications of Speech Coding

- Wireline Telephony
- Videoconferencing
- Digital Cellular
- IP Telephony
- Voice Mail
- Speech Storage

Speech and Audio Coding Standards

➤ Narrowband speech

- GSM-AMR, G.729, G.723, G.728, IS-127(EVRC), IS-96(QCELP), IS-95(VSELP)
- G.711(PCM), G.721(ADPCM), G.726(ADPCM)
- LPC-10, MELP,...

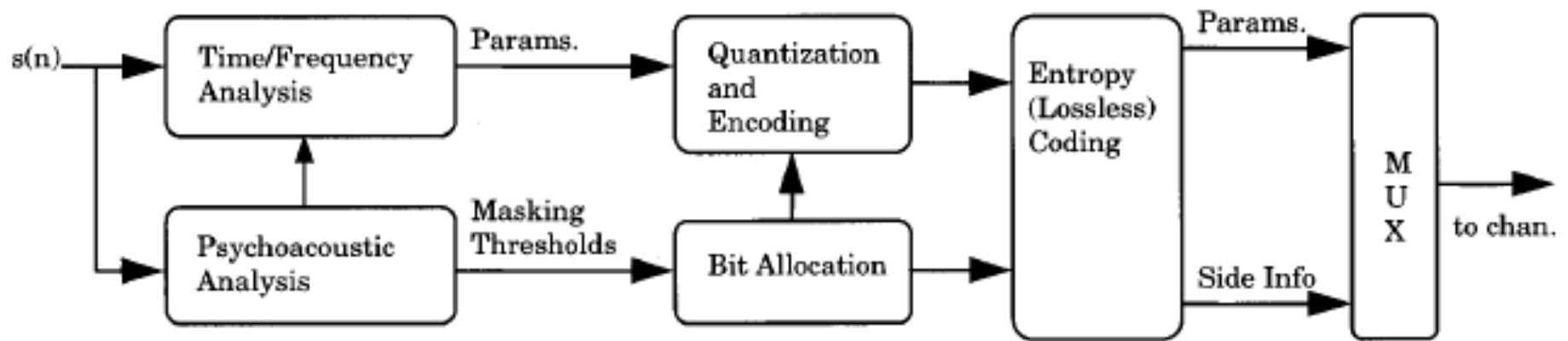
➤ Wideband speech

- G.722 (ADPCM)
- G.722.1 (Transform)
- AMR-WB (CELP)

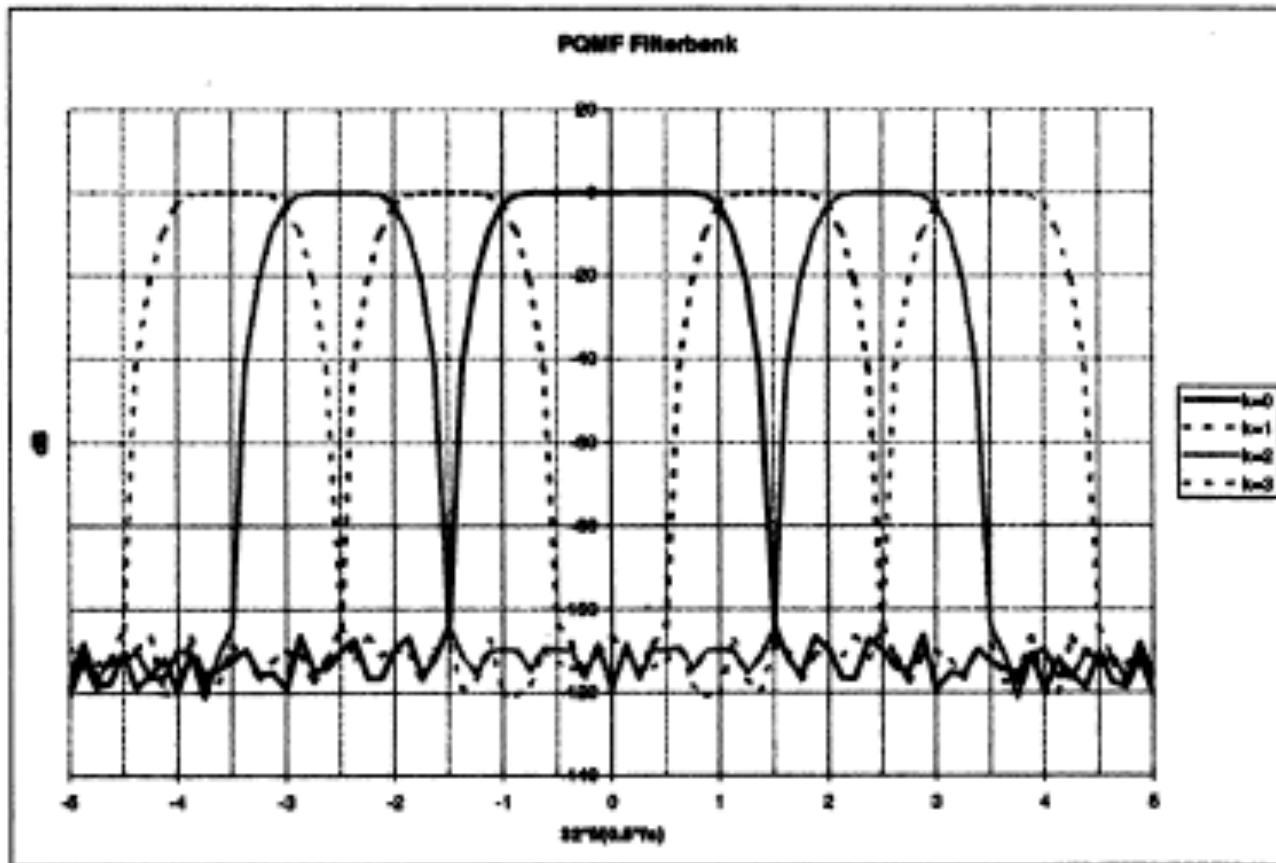
➤ Wideband audio

- MPEG-1,2,4
- Philips PASC
- Sony ATRAC
- DOLBY AC-3

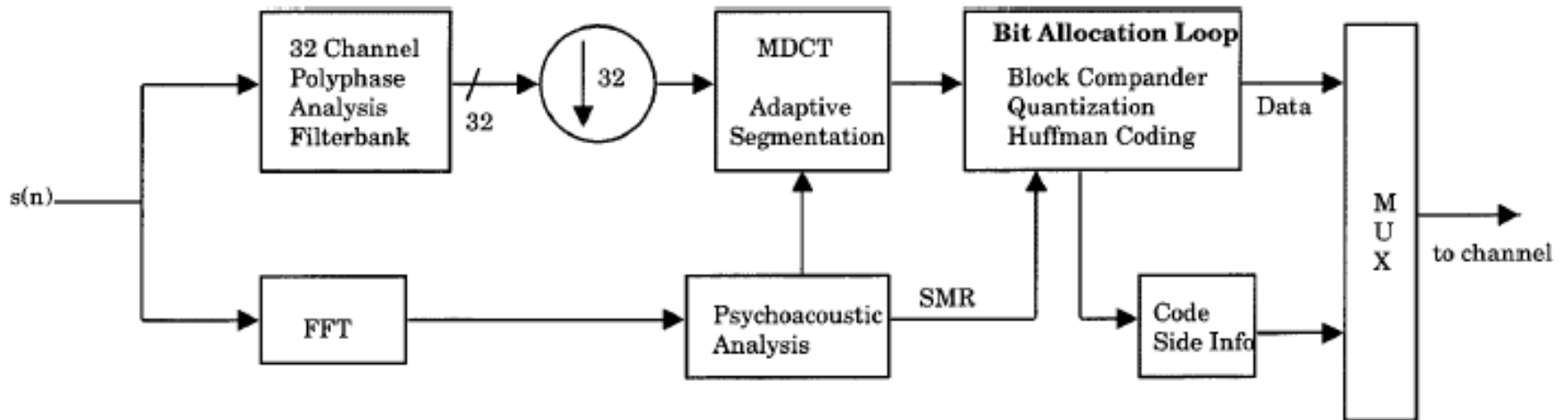
Generic Audio Coding Method [2]



Frequency Response of the MPEG Audio First Four Bands [1]



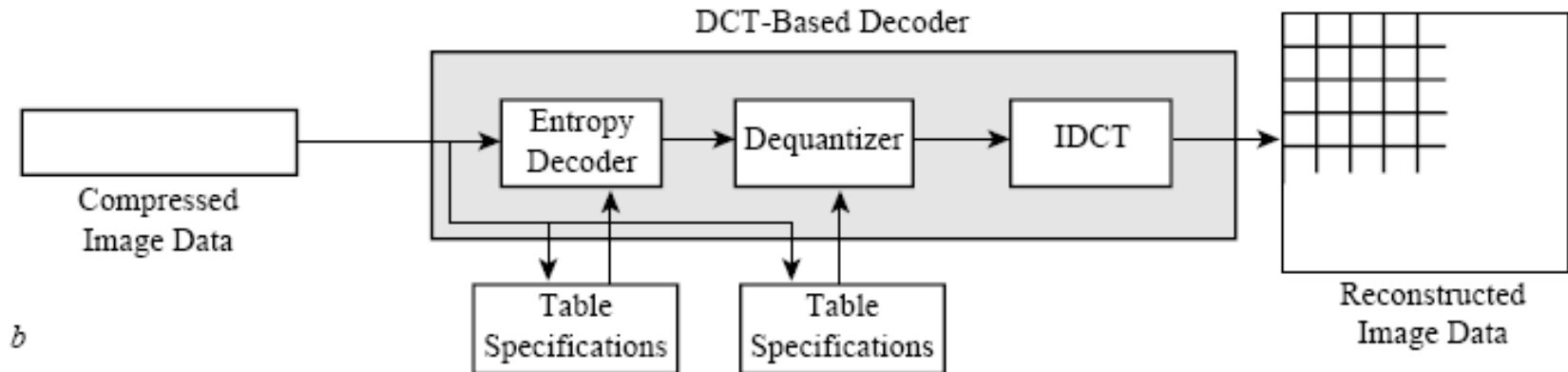
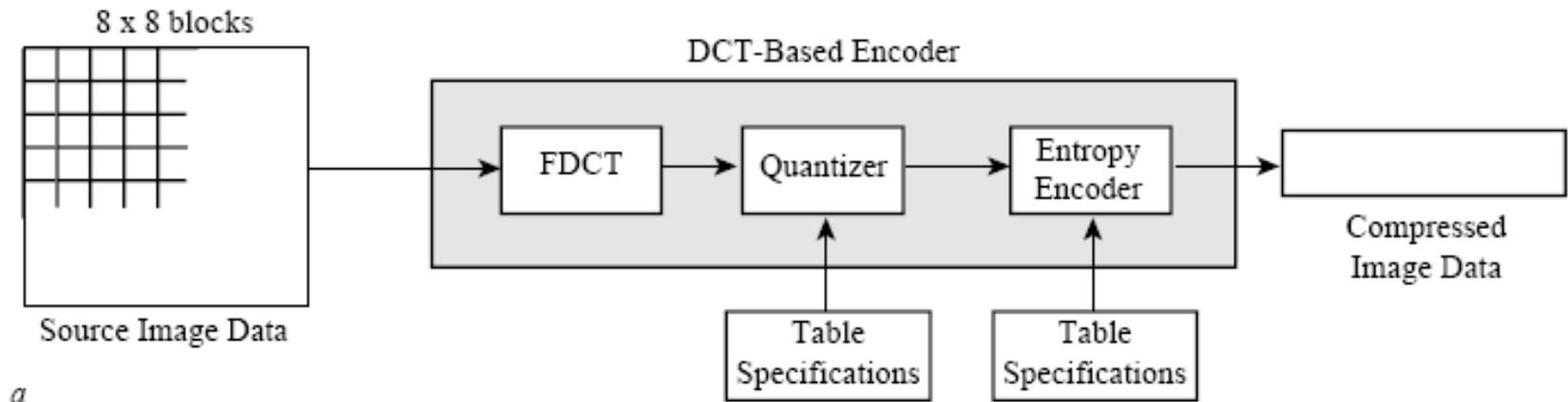
MPEG-1 Layer III Encoder [2]



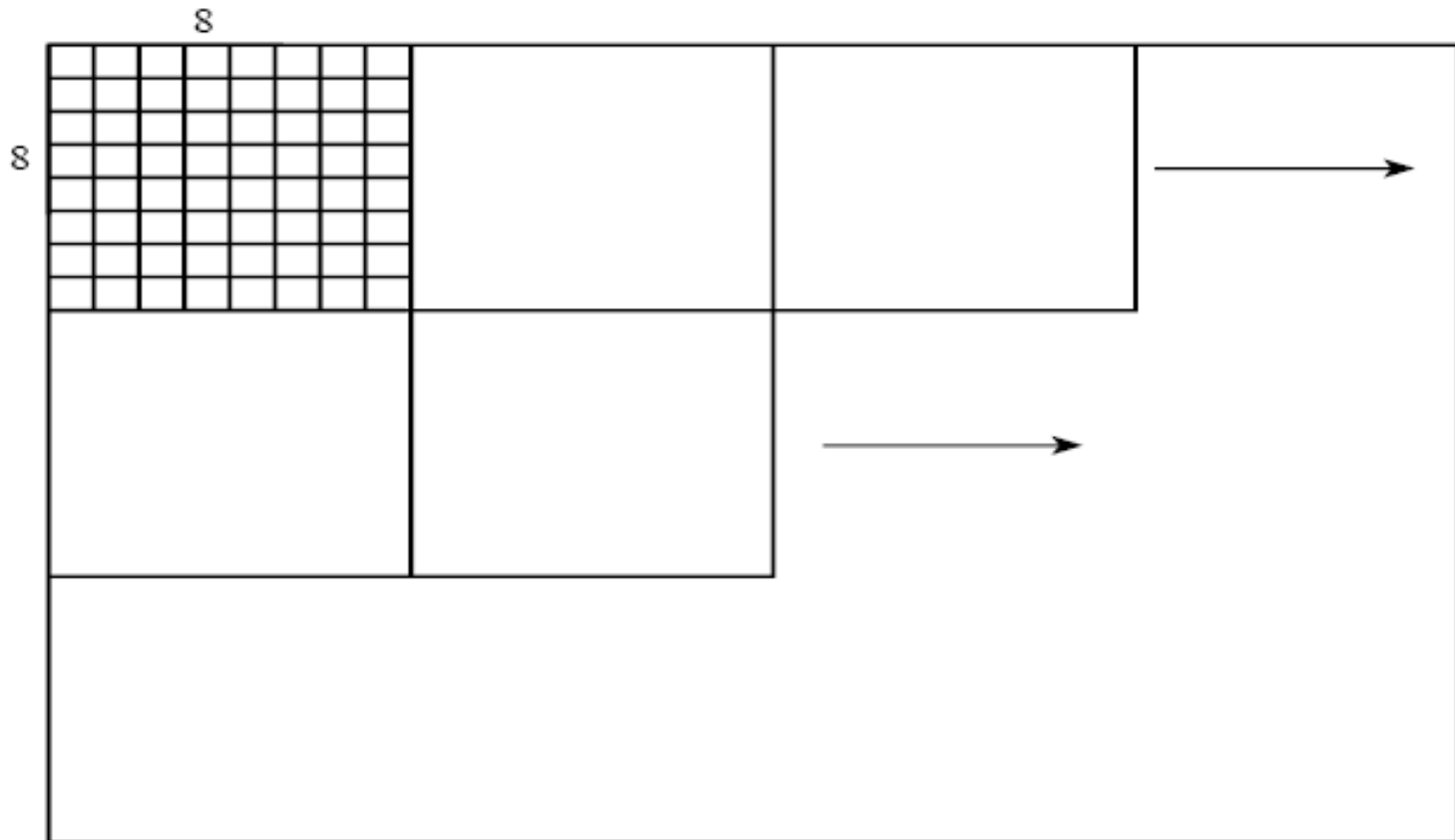
Still Image and Video Compression Standards

- JPEG, 0.25-2.0 bits/pixel
- JPEG 2000
- MPEG-2, 4-10 Mbps
- VC-1
- AVC/H.264

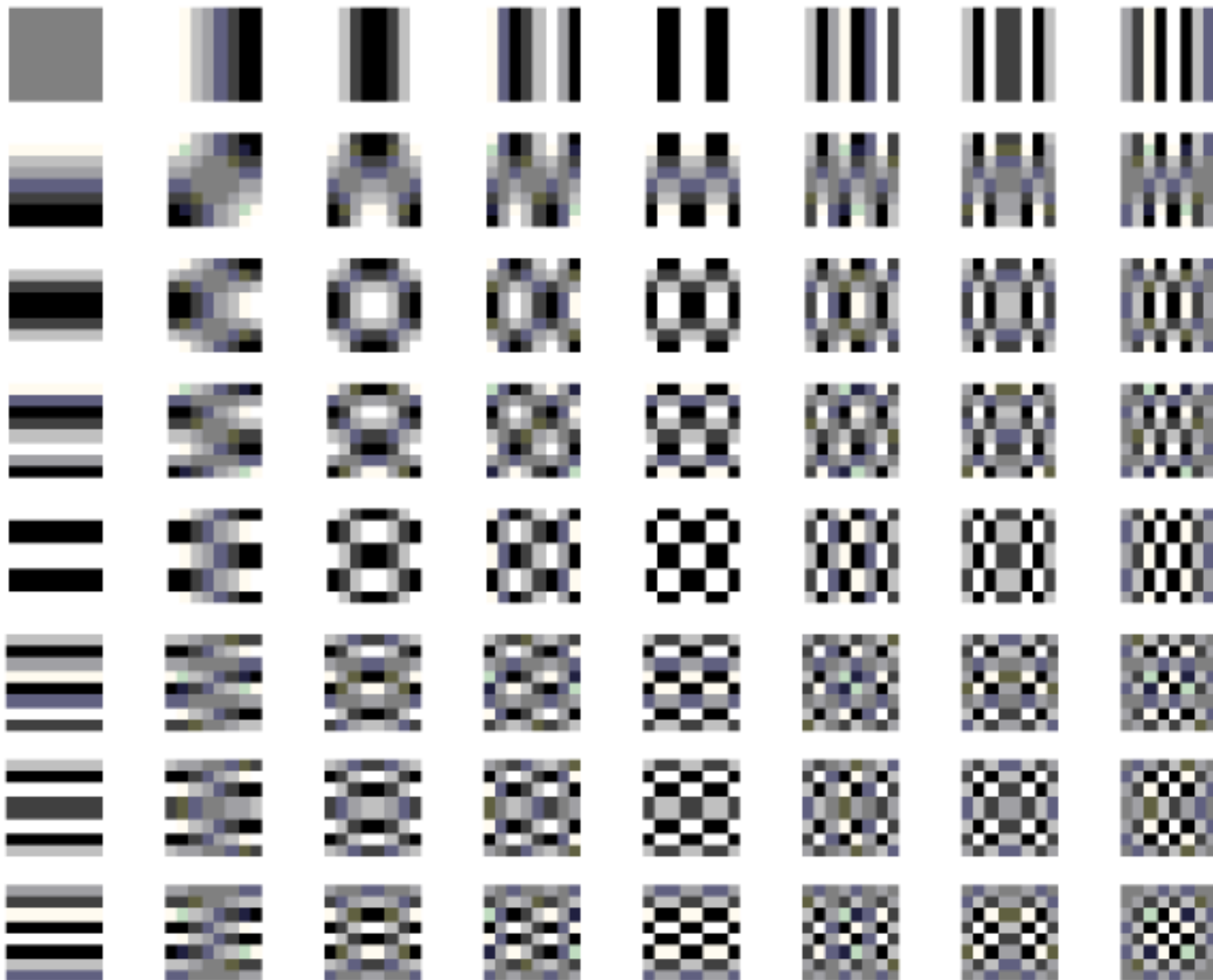
JPEG Encoder and Decoder



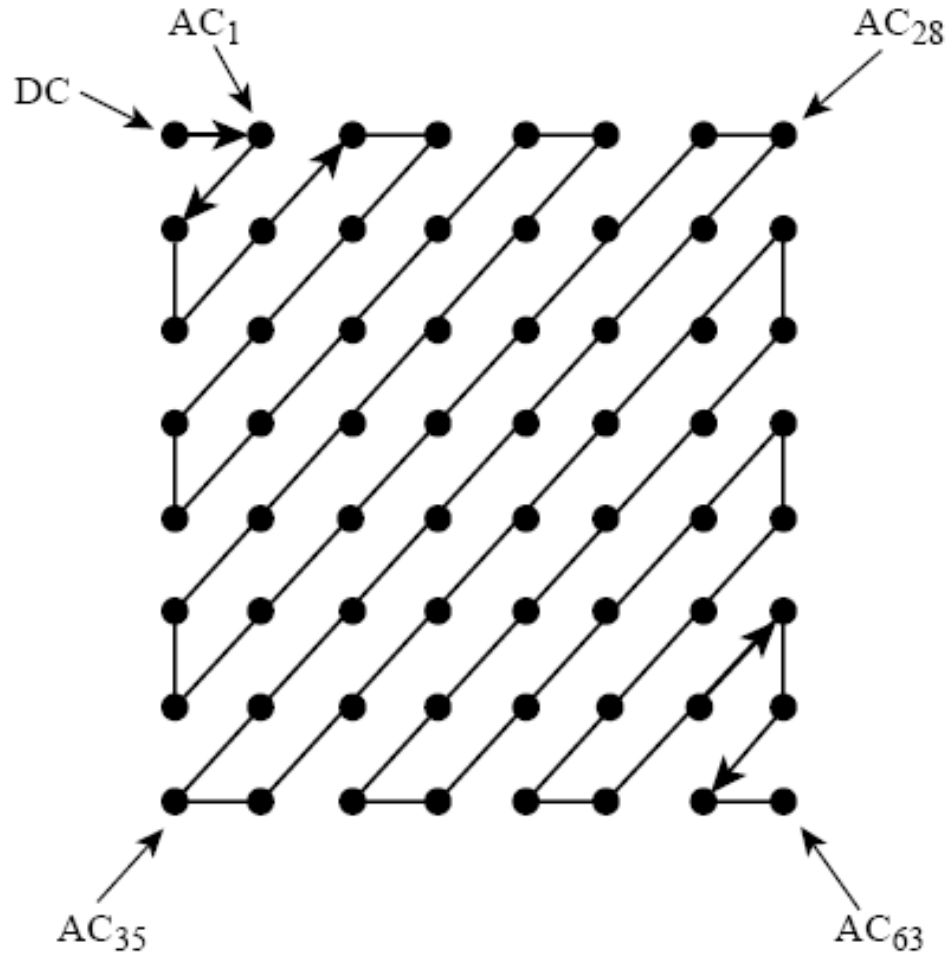
Partitioning into 8 by 8 Blocks



2D-DCT Basis Functions

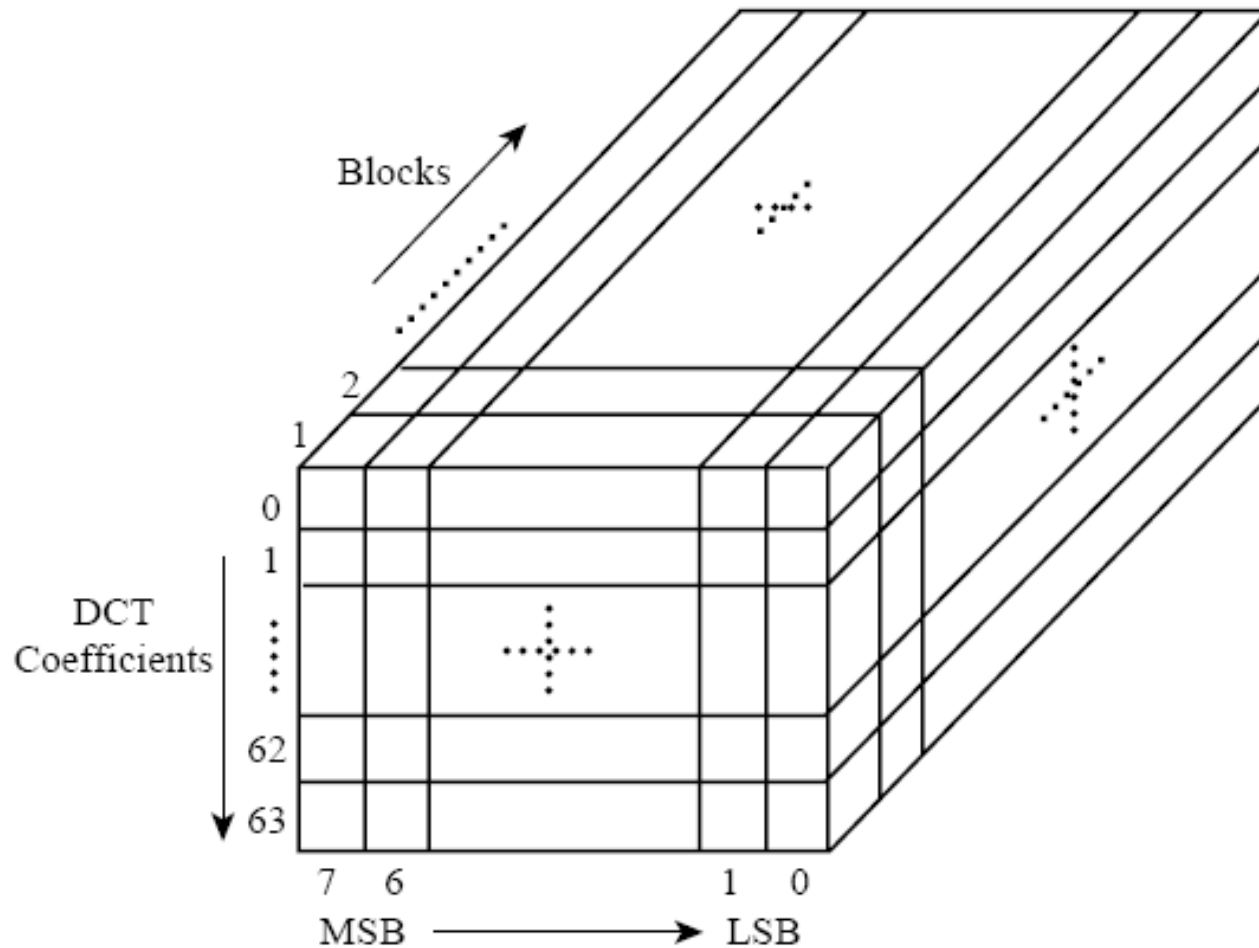


Zig-Zag Coefficient Ordering



Zigzag Order

Sequential Lossy Encoding

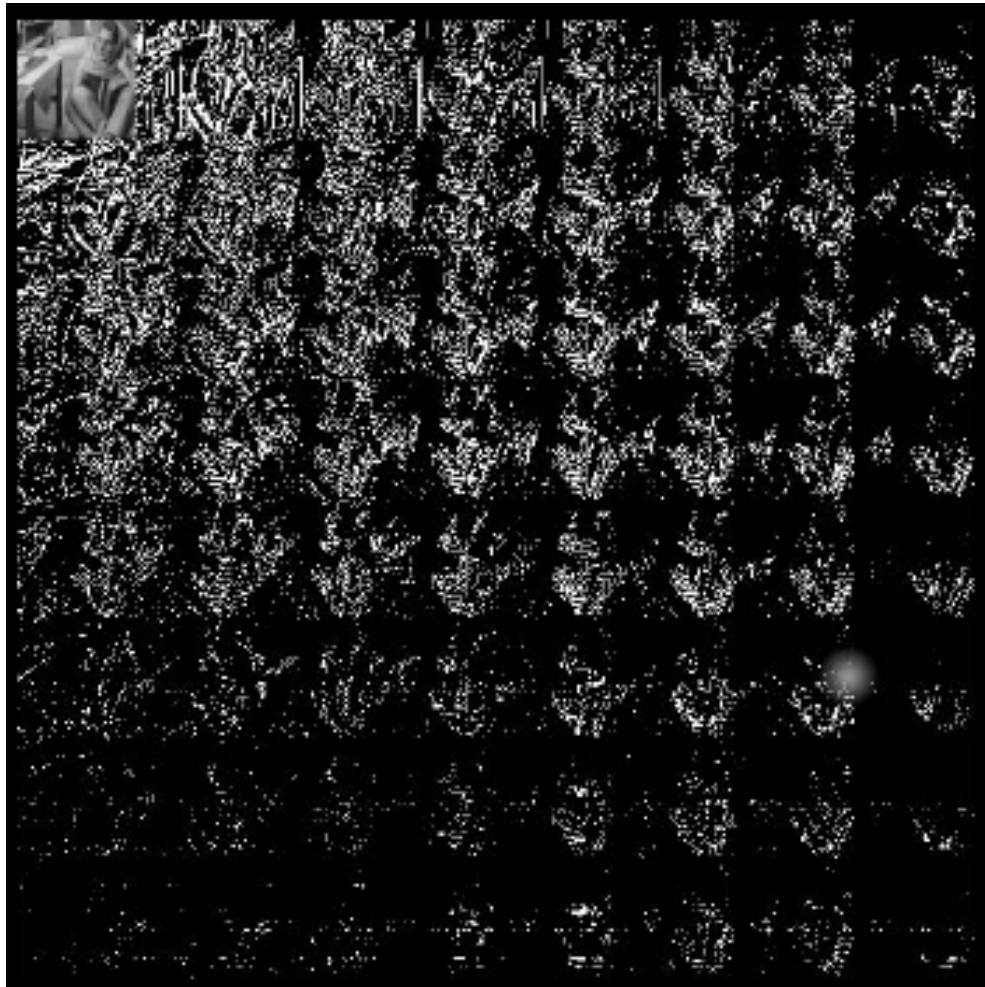


Luminance Quantization Table

Luminance quantization table

| | | | | | | | |
|----|----|----|----|-----|-----|-----|-----|
| 16 | 11 | 10 | 16 | 24 | 40 | 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 | 64 | 81 | 104 | 113 | 92 |
| 49 | 64 | 78 | 87 | 103 | 121 | 120 | 101 |
| 72 | 92 | 95 | 98 | 112 | 100 | 103 | 99 |

Images Reconstructed from the Same Coefficient Taken from Each Block



JPEG2000 Features

- Superior low bit-rate performance—below 0.25 bpp for highly detailed gray-scale images
- Lossless and lossy compression—lossless compression in the course of progressive decoding
- Progressive transmission by increasing pixel accuracy or spatial resolution

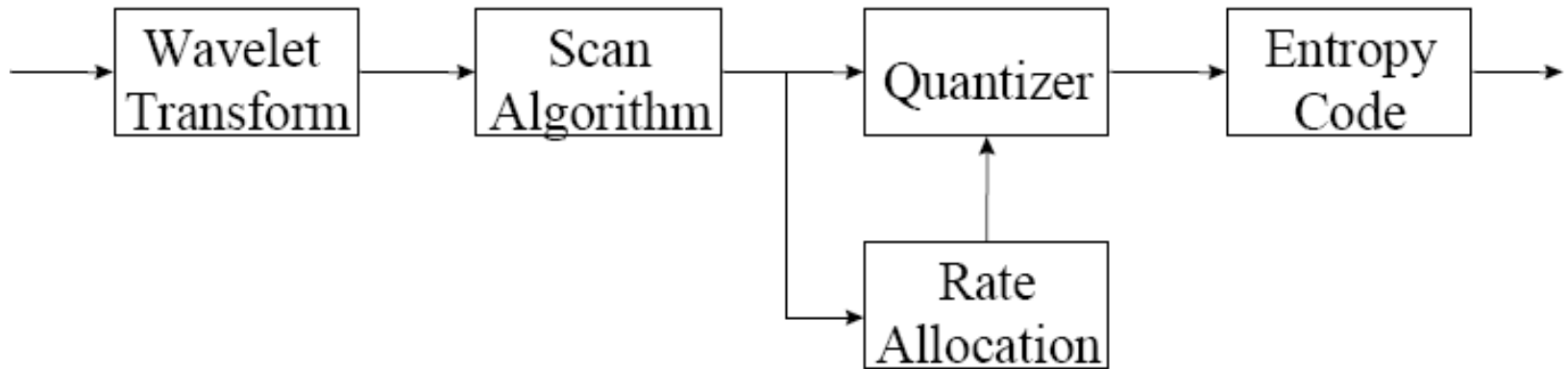
JPEG2000 Features (cont' d)

- Region-of-interest (ROI) coding
- Random codestream access and processing
- Robustness to bit-errors
- Open architecture—a decoder is only required to implement the core tool set and a parser that understands the codestream
- Content-based description

JPEG2000 Features (cont' d)

- Continuous-tone and bi-level compression
—compress and decompress images with various dynamic ranges (1 to 16 bit) for each color component

Basic Encoding Diagram



Basic Encoding Steps

- Pre-processing of the image
- The image is decomposed into components
- The image/components are decomposed into tiles
- Tiling refers to partitioning the image into rectangular non-overlapping blocks, called tiles, which are compressed independently as if they are independent images

Basic Encoding Steps (cont' d)

- A wavelet transform is applied on each tile
- Each tile is decomposed into different resolution levels
- The decomposition levels are made up of subbands of coefficients that describe the frequency characteristics of local areas of the tile-component
- The subbands of coefficients are quantized and collected into rectangular arrays of code blocks

Basic Encoding Steps (cont' d)

- The bit-planes of the coefficients in a cod-block are entropy coded
- ROI' s can be encoded in higher quality than the background
- Markers are added in the bitstream for error resilience
- The codestream has a main header that describes the original image and the various decomposition and coding styles

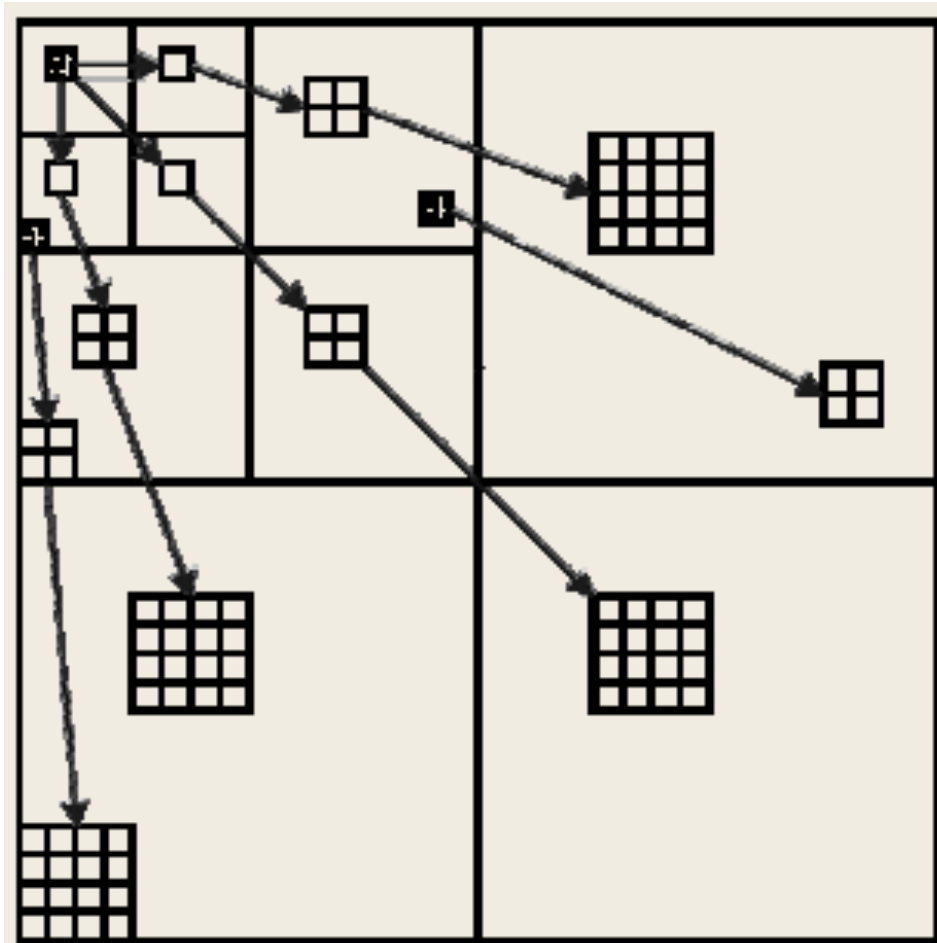
Three-Level 2-D Wavelet Transform



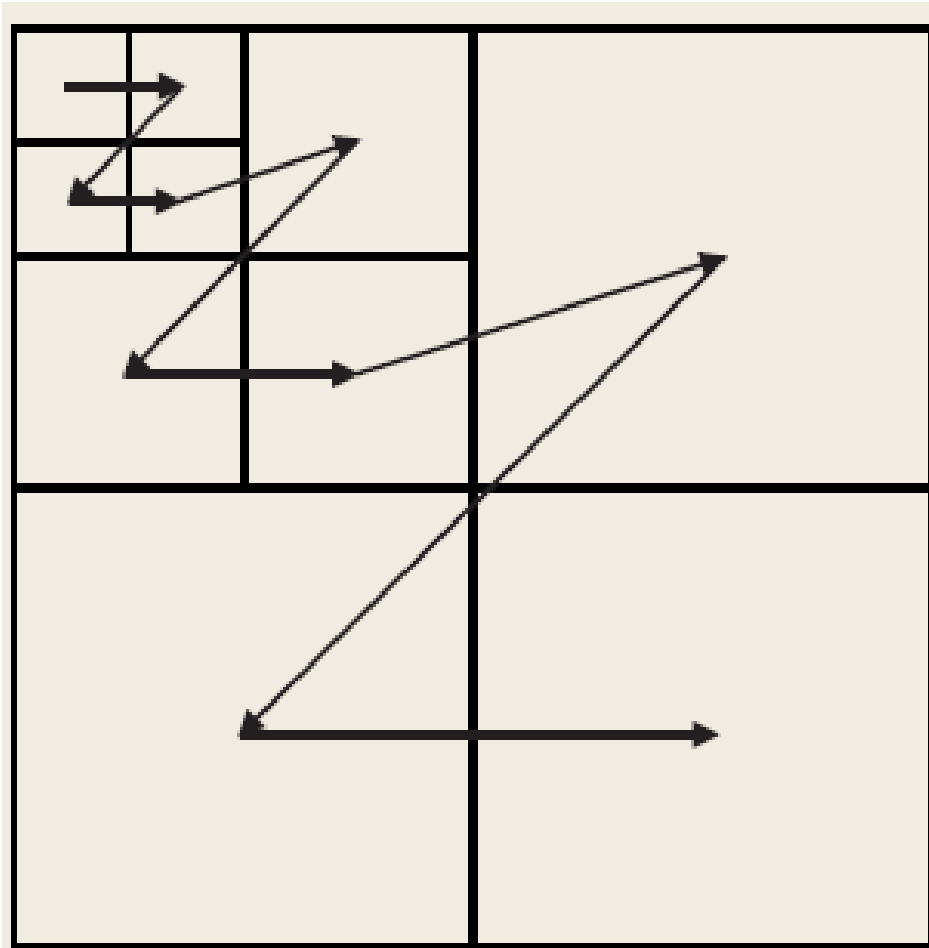
Subband Labeling

| | | | |
|--------|--------|--------|--------|
| LL_3 | HL_3 | HL_2 | HL_1 |
| LH_3 | HH_3 | | |
| LH_2 | HH_2 | | |
| LH_1 | | HH_1 | |

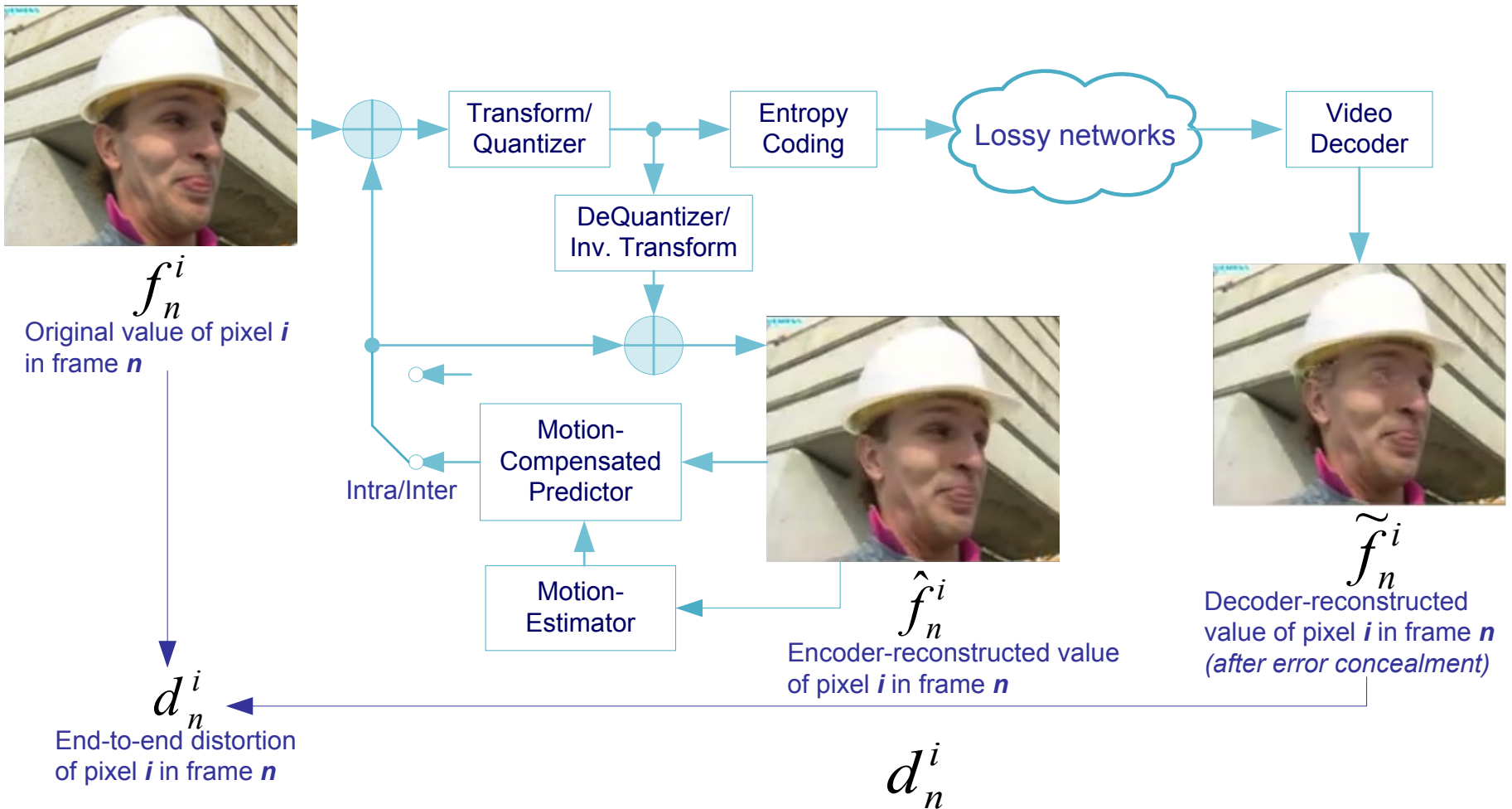
Assumed Relationship Between Quadtree Coefficients



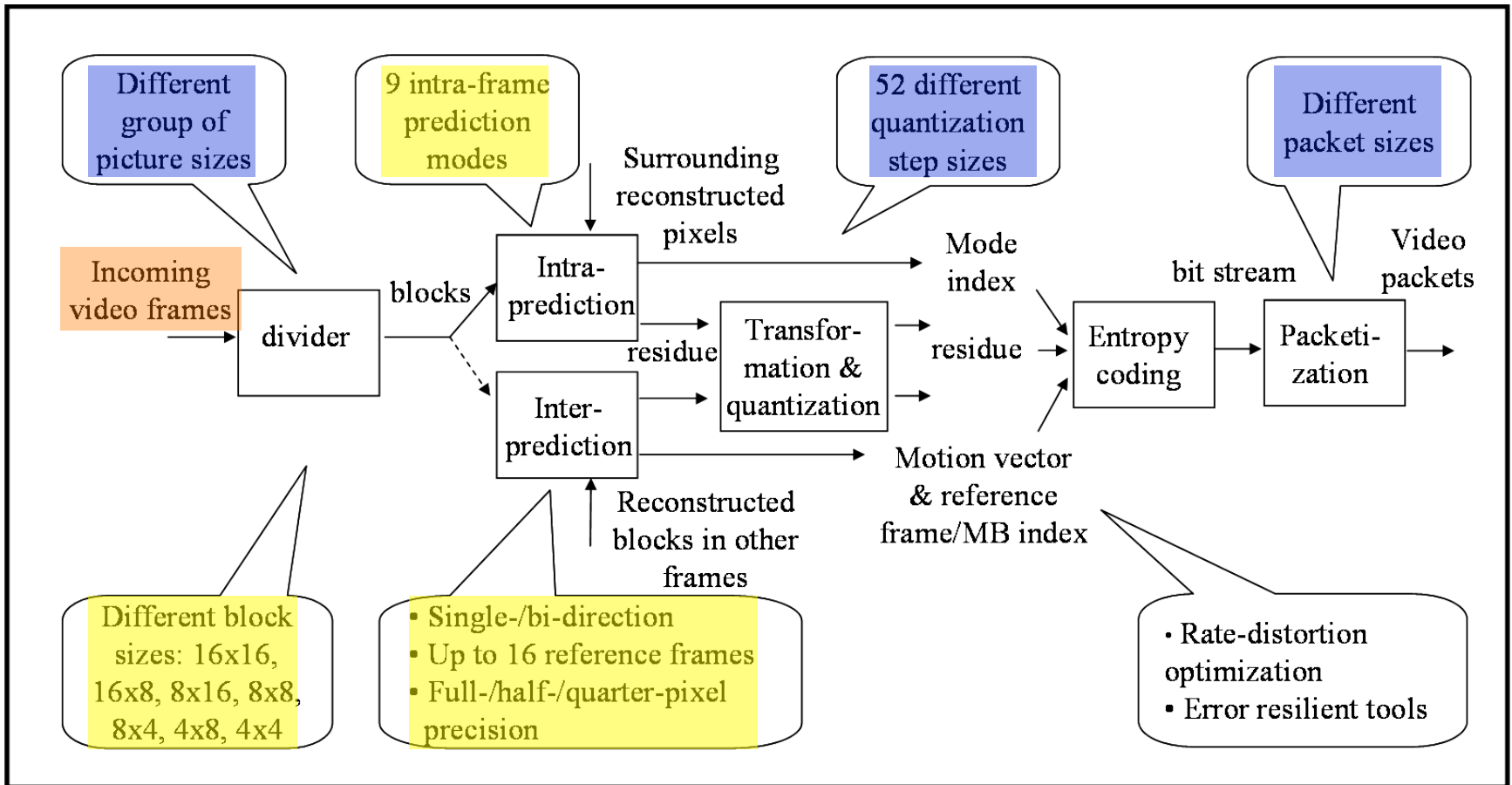
The Scanning Order for Dominant Passes of the EZW Algorithm



Classical Video Coding



H.264 Encoder



Video Codec Parameters and Performance for Different Videos

| Video | silent.cif | | paris.cif | | stefan.cif | |
|----------------------------------|------------------|-------|----------------|-------|------------------|-------|
| Typical application | video conference | | news broadcast | | sports broadcast | |
| QP | 26 | 30 | 26 | 30 | 26 | 30 |
| Avg psnr | 36.69 | 34.22 | 36.59 | 33.45 | 36.69 | 33.47 |
| Bit rate (kbps) | 169.5 | 97.8 | 373.5 | 218.9 | 1396.8 | 404.6 |
| I frame size (bytes) | 13945 | 8826 | 19886 | 14390 | 30432 | 15978 |
| Average of P frame size (bytes) | 1272 | 725 | 2924 | 1683 | 11429 | 3230 |
| Variance of P frame size (bytes) | 412 | 254 | 322 | 219 | 1544 | 625 |

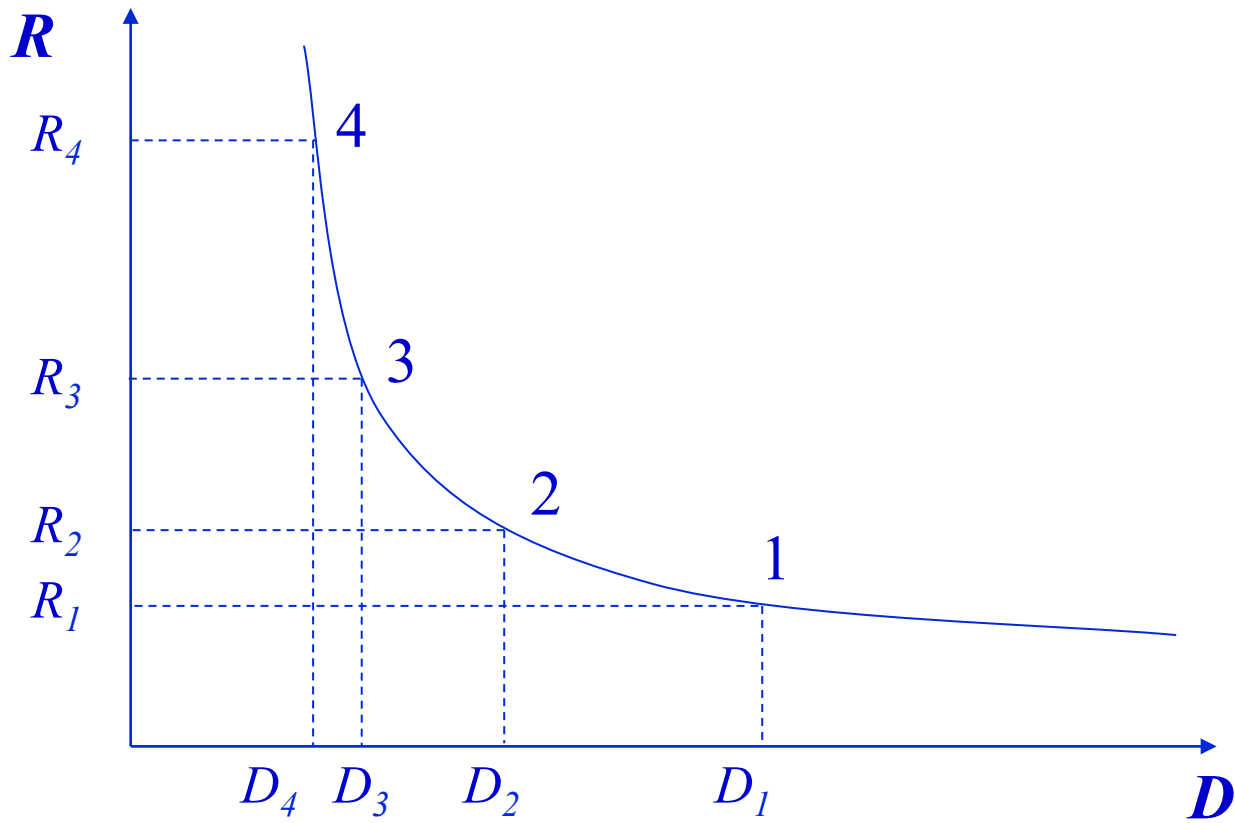
Key Functionalities for Lossy Compression

- Error Concealment
- Scalability
 - SNR
 - Spatial
 - Temporal
 - Bandwidth
- Multiple Descriptions

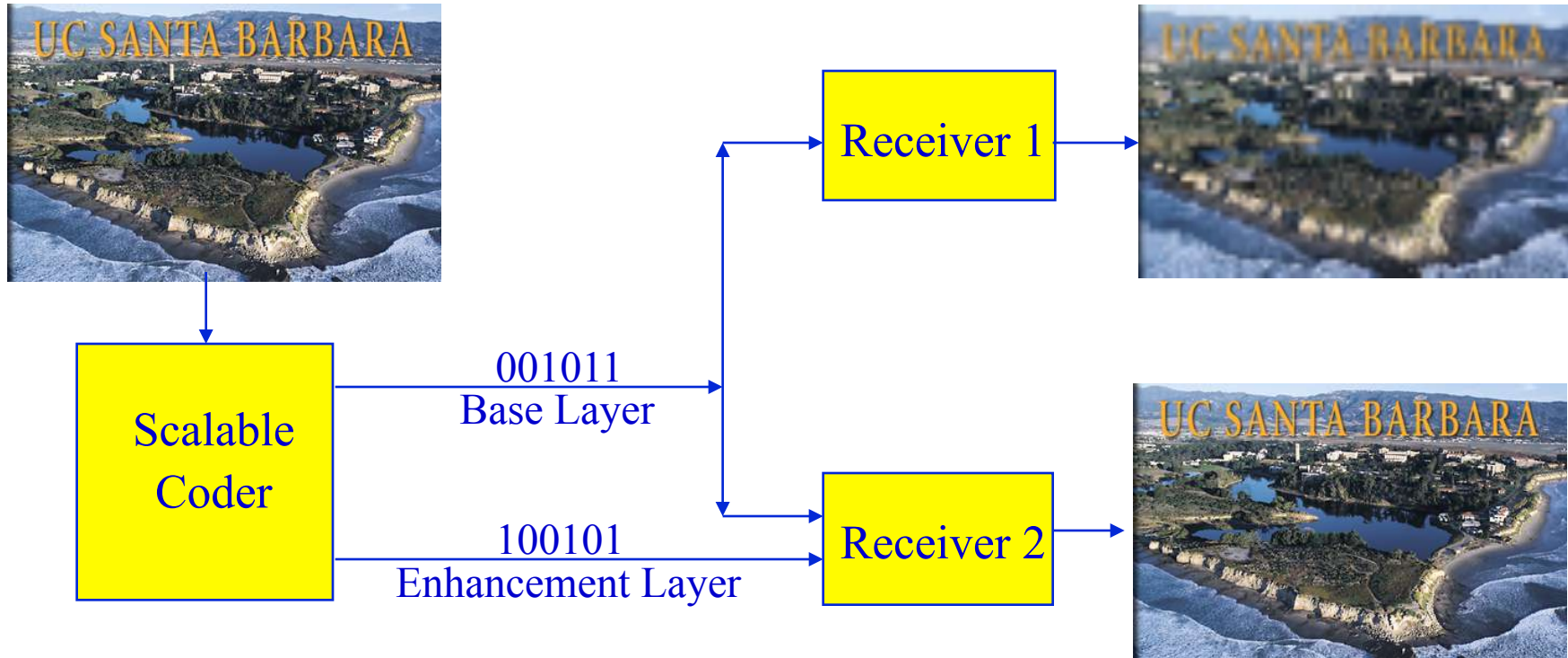
Scalable Coding

- Sometimes denoted as layered coding, embedded coding, or variable rate coding
- Scalable Coding consists of a core coder at the lowest bit rate plus one or more enhancement layers
- Quality improvement is achieved by sending only an incremental bit rate above the core layer
- Speech Scalable Coding: SNR scalability, Bandwidth scalability

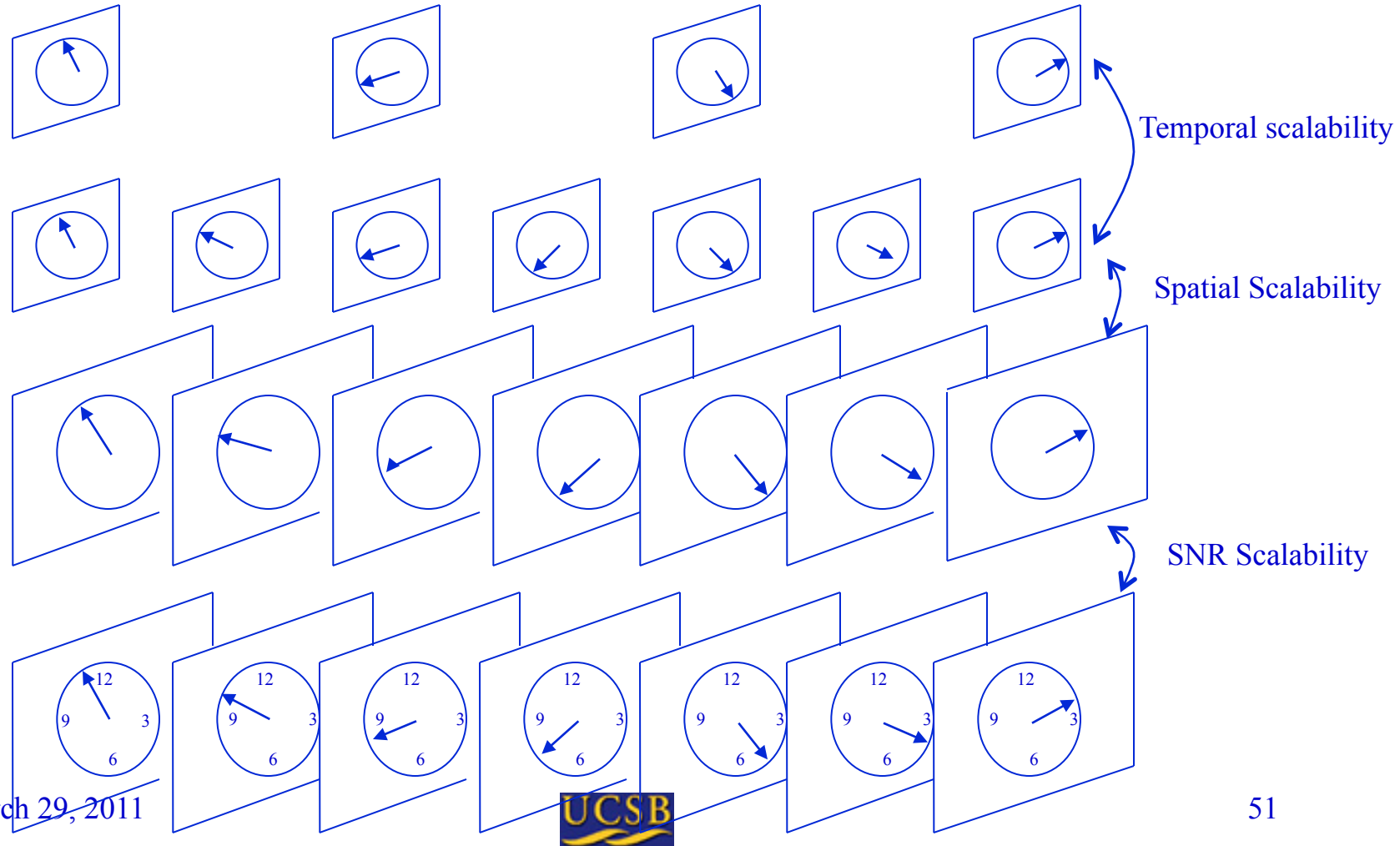
The Successive Refinement Problem



Scalable Coding



Temporal, Spatial and SNR scalability



Multiple Descriptions Coding

