ECE 594/MAT 594
Multimedia Compression

Jerry D. Gibson

Winter 2003

January 10, 2003

What is Compression?

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

Components of a Compression Problem

- Source
- Rate
- Distortion Measure

Rate Distortion Theory

The minimum rate required to represent a source X with average distortion less than or equal to D is the rate distortion function $R(D)$ defined as

$$R(D) = \min_{f_{\hat{X}|X}(\hat{x}|x) \in P_D} I(X;\hat{X})$$

where

$$I(X;\hat{X}) = \int\int f_X(x) f_{\hat{X}|X}(\hat{x} \mid x) \log \frac{f_{\hat{X}|X}(\hat{x} \mid x)}{f_{\hat{X}}(\hat{x})} dx d\hat{x}$$

Goal of Source Compression

- Represent speech with as few bits as possible with an acceptable loss in quality and intelligibility
- Fix the rate and minimize distortion—*D(R)*
- Fix the acceptable distortion level and minimize the rate—*R(D)*

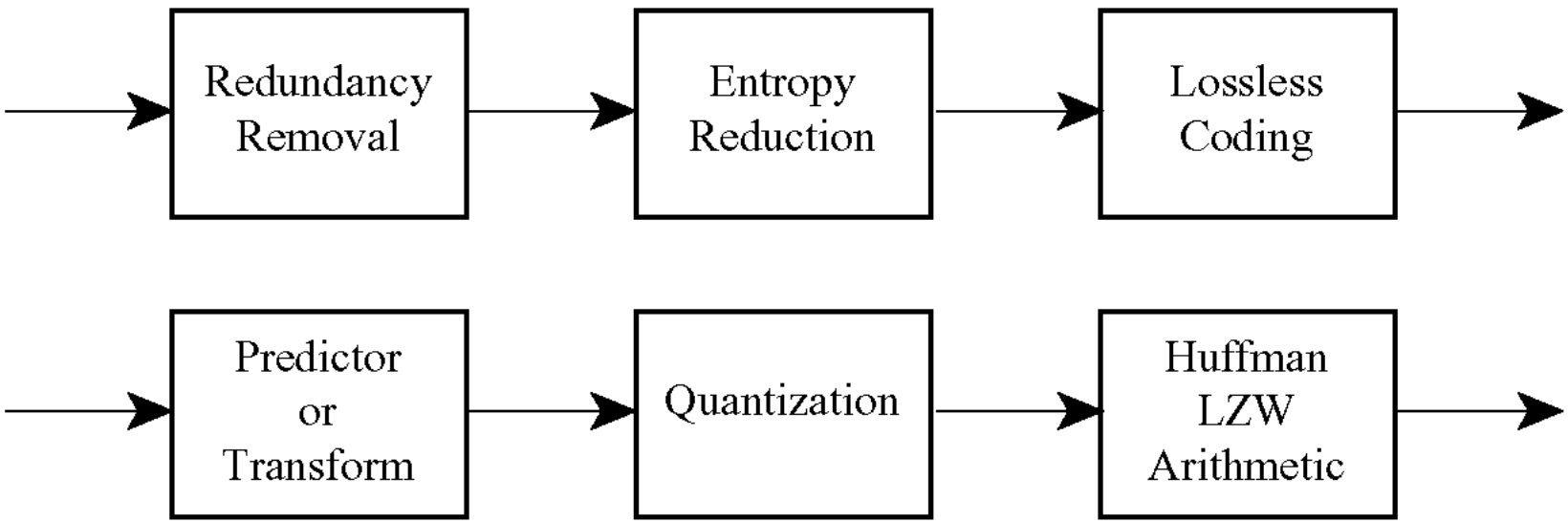
Synonyms for Data Compression

- 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

MAJOR SETS IN DATA COMPRESSION



Speech Coding Categories

- 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

Uncompressed Bit Rates for Speech and Audio

Source	Bandwidth (Hz)	Sampling Rate	Bits Per Sample	Bit Rate
Telephone Speech	200-3400	8000 samples/s	12	96 kbits/s
Wideband Speech	50-7000	16000	14	224 kbits/s
Wideband Audio (2 Channels)	20-20000	44.1 ks/s	16/chan	1.412 Mbits/s (2 channels)

Approximate Bit Rates for Uncompressed Sources

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

Audio Sampling Rates

Application	Bandwidth (kHz)	Sampling Rate (kHz)
Voice telephony	3.2	8
Teleconferencing (audio)	7.0	16
Compact disc (CD) audio	20.0	44.1
Digital audiotape (DAT)	20.0	48

Video Sampling Rates

Format	Lines/Frame × Pixels/Line × Frames/Second =	Sampling Rate (million pixels per second
CIF (videoconferencing)	$360 \times 288 \times 30 =$	3
CCIR (TV)	$720 \times 576 \times 30 =$	12
HDTV	$1280 \times 720 \times 60 =$	60

ITU-T Facsimile Standards

Size	Vertical Resolution (lines/mm)	Horizontal Resolution (pixels/mm)	Lines/ Frame	Pixels/ Line
Normal resolution 20.7 cm (8.27 inches) by 29.2 cm (11.7 inches)	3.85	8	1188	1728
High resolution 20.7 cm (8.27 inches) by 29.2 cm (11.7 inches)	7.7	8	2376	1728

Image and Video Formats

Formats	Usable Horizontal Lines*	Pixels per Line	Total Pixels per Frame	Frames per Second	Required Bandwidt Transmiss Rate
Analog video					
NTSC					
(Americas, Asia)	338	426	150,0005	29.97	4 MHz
PAL (Europe)	411	420	172,0005	25.00	5 MHz
VHS	338	280	95,0005	29.97	<4 MHz
Computer image					
SVGA	1024	768	786,5005	60	—
VGA	640	480	307,0005	60	—

Formats	Usable Horizontal Lines*	Pixels per Line	Total Pixels per Frame	Frames per Second	Required Bandwidth/ Transmissio Rate
Motion picture film					
35mm	(not a raster-		500,000	24	—
16mm	scanned image)		125,000	24	—
Digital video					
QCIF (H.261)	144	176	25,000	15–30	56 kbps–2 M
CIF (H.261)	288	352	100,000	15–30	56 kbps–2 M
HDTV	806	1920	1,550,000	50	140 Mbps
MPEG (constrained set)	345	360	124,000	30	1.5 Mbps and higher

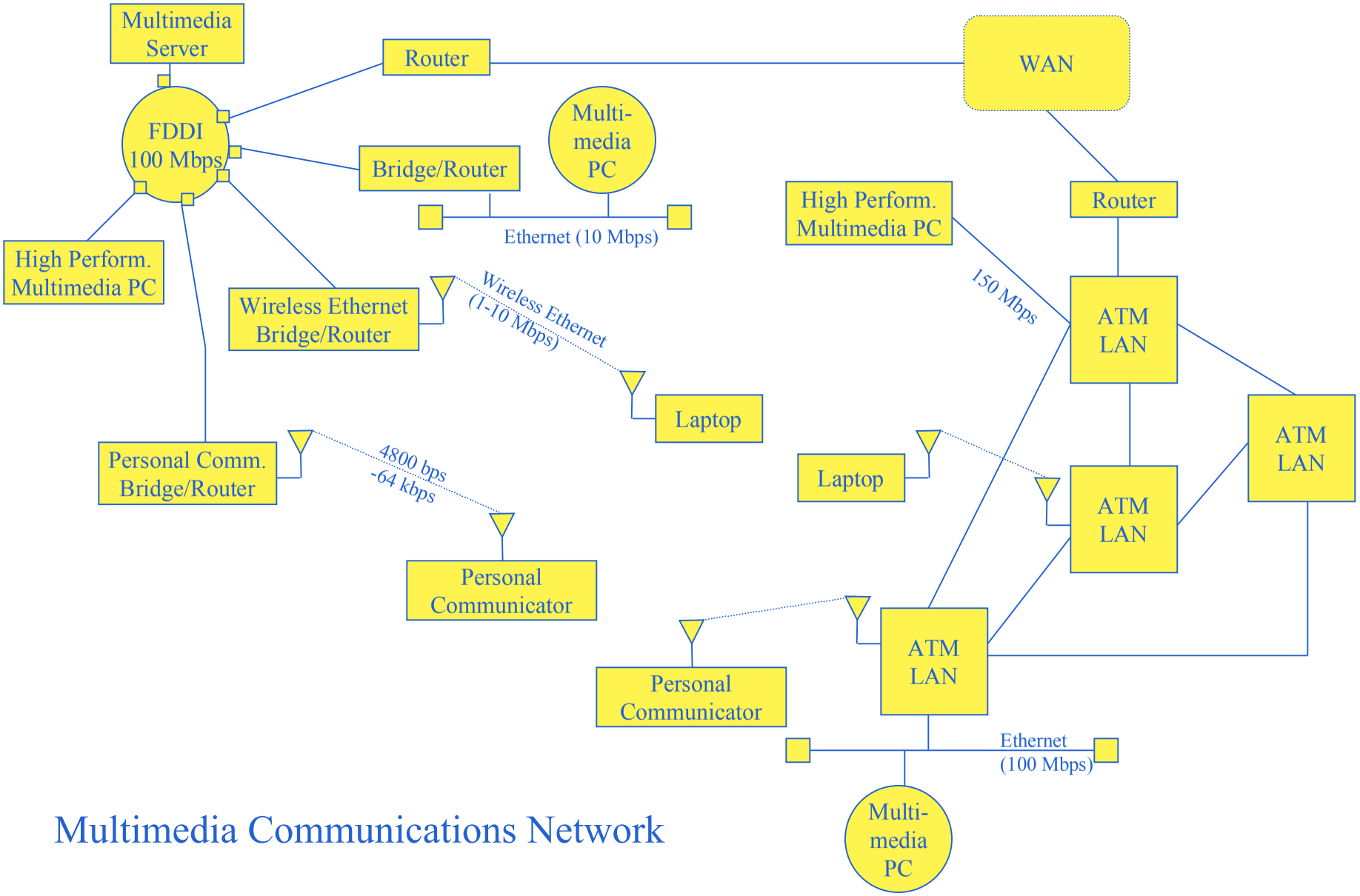
*Eliminates retrace lines and includes the utilization ratio.

H.324 Video Formats

Format	Pixels	H.261	H.263
SQCIF	128×96	optional	required
QCIF	176×144	required	required
CIF	352×288	optional	optional
4 CIF	704×576	n/a	optional
16 CIF	1408×1152	n/a	optional

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



Multimedia Communications Network

January 10, 2003

UCSB

Design Distortion Measures

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

Performance Evaluation

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

Applications of Speech Coding

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

Multimedia Over Networks

- Principal Issues
  - Connectivity or Access
  - Interoperability
  - Performance
  - Implementations
  - Choices

Channels and Platforms

- Heterogeneous Communications Channels
  - Channel Bandwidths
  - Channel Qualities
- Heterogeneous Platforms, Terminals, or Handsets
  - Processing Capability
  - Power Consumption
  - Presentation Capability

User Preferences

- Availability of Services
- Cost
- Choice of QoS

Speech 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

Telephone Bandwidth Speech Coding Standards

Coder	›its/s)	Quality	LD-CELP	16	4.0 MOS	(G.728)	RPE

Telephone Bandwidth Speech Coding Standards (2)

EFR (IS-641)	8	3.8 MOS
EVRC (IS-127-2)	0.8-8.55 (variable)	3.8 MOS
CS-ACELP (G.729)	8	4.0 MOS
CS-ACELP (G.729A)	8	3.75 MOS
MPC-MLQ (G.723.1)	5.3-6.4	3.5 MOS
ACELP (NB-AMR)	4.75-12.2	3.5-4.1 MOS

Narrowband AMR Speech Codec

- NB-AMR was originally standardized for GSM by ETSI in 1998.
- NB-AMR has 8 rates (all ACELP): 4.75, 5.15, 5.9, 6.7, 7.4, 7.95, 10.2 12.2 kbits/s
- 12.2 kbps ---- GSM EFR

  7.4 kbps ---- TDMA136 EFR
- 8 kHz sampling rate

Wideband AMR Speech Codec

- WB-AMR has 9 rates:
  6.6, 8.85, 12.65, 14.25, 15.85, 18.25, 19.25, 23.05, 23.85 kbits/s
- 16 kHz sampling rate
- Separate codec from NB-AMR
- WB-AMR has just been standardized by 3GPP in June, 2001 for GSM/3G systems.

Toll Quality Speech Coding Timeline

- Log PCM 64 kbps 1972
- ADPCM 32 kbps 1984
- SBC 16 kbps 1984
- APC with perceptual weighting 16 kbps 1980
- Multipulse LPC 16 kbps 1982
- Codebook Excited LPC 8 kbps 1995

DRT, DAM, and MOS Scores for Common Speech Coders

Coder	DRT	DAM	MOS
64-kbps PCM (Pulse Code Modulation)	95	73	4.2
32-kbps ADPCM (Adaptive Differential PCM)	94	68	4.0
16-kbps LD-CELP (Low-Delay Code Excited Linear Predictive Coding)	94	70	4.0
4.8-kbps CELP (Code Excited Linear Predictive Coding)	91	65	3.2
2.4-kbps LPC (Linear Predictive Coder—vocoder)	87	54	2.2

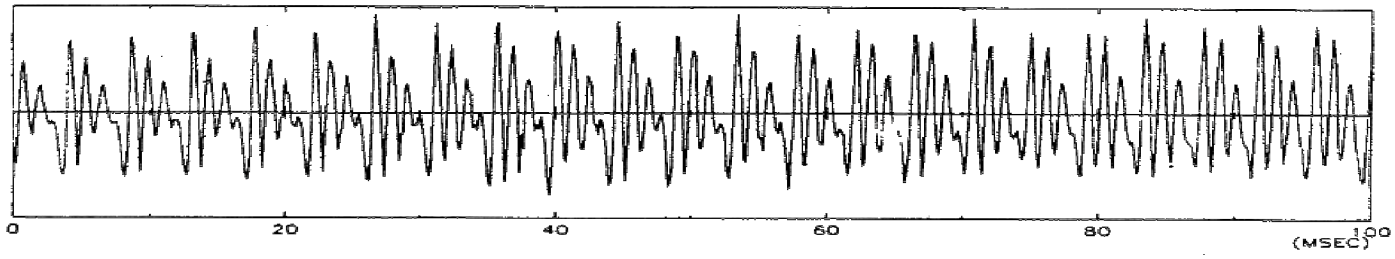
Image/Video Compression Standards

Source	Standard	Rates
Video telephone Px64	ITU-T H.261	56 kbps–2 Mbps
Black-and-white, color, multispectral images	JPEG	0.25–2 bits/pixel
Moving pictures and audio	MPEG-1	1.5 Mbps
Broadcast-quality pictures and audio	MPEG-2	6–10 Mbps
High-quality audio for MPEG	HDTV	64/128/192 kbps per channel
Video	H.263	$\leq$28.8 kbps

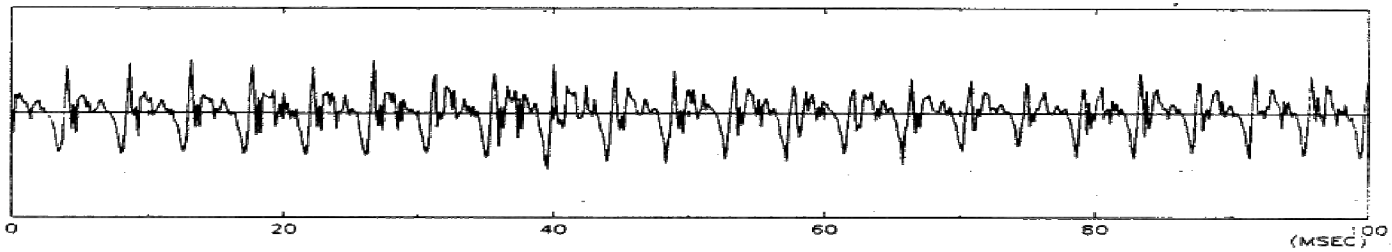
Selected Videoconferencing Standards--Basic Modes

Standard	Network	Video	Audio
H.320 (1990)	ISDN	H.261	G.711
H.323 (1996)	LANs/Internet	H.261	G.711
H.324 (1995)	PSTN	H.263	G.723.1
H.310 (1996)	ATM/B-ISDN	H.262	MPEG-1

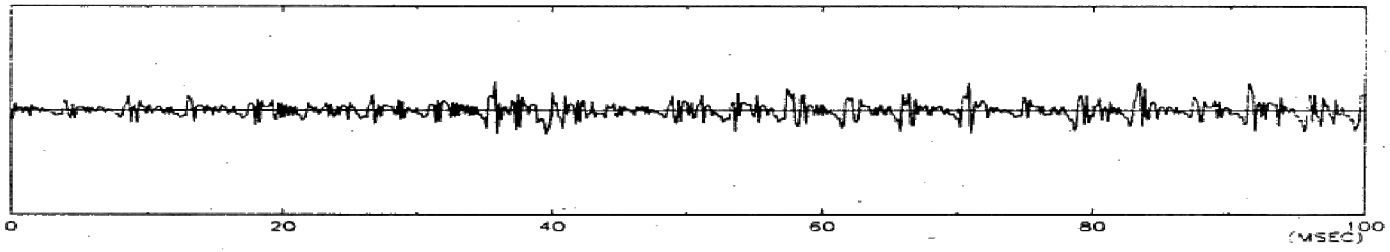
Redundancy Removal From Speech Signals



(a) Speech Waveform



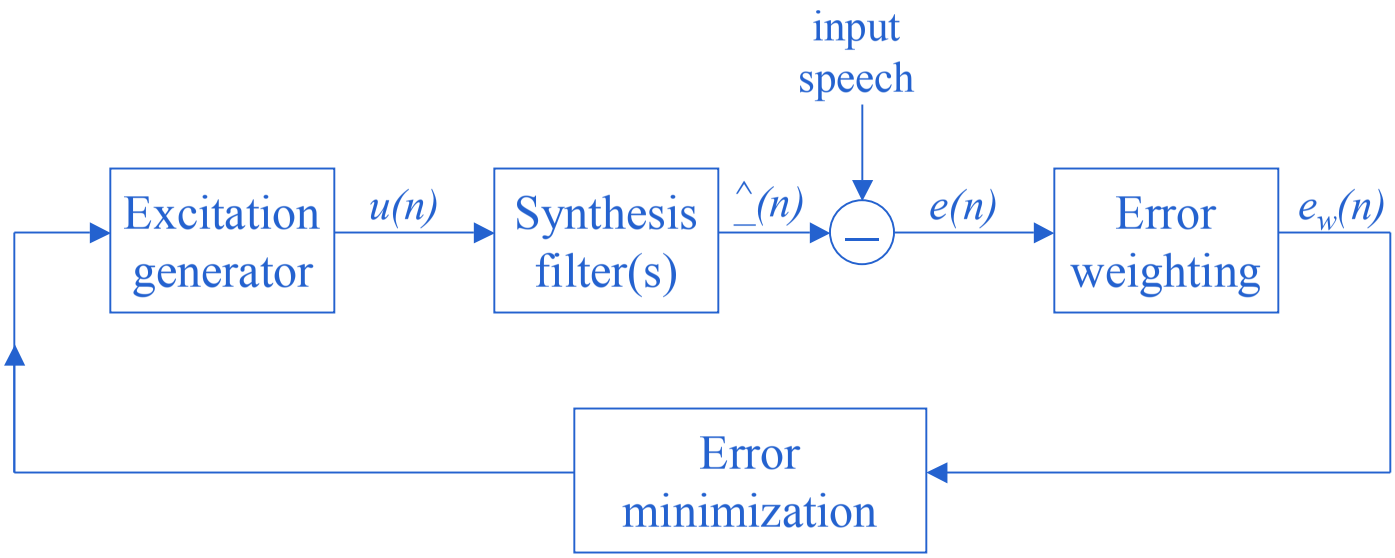
(b) Short term removed signal



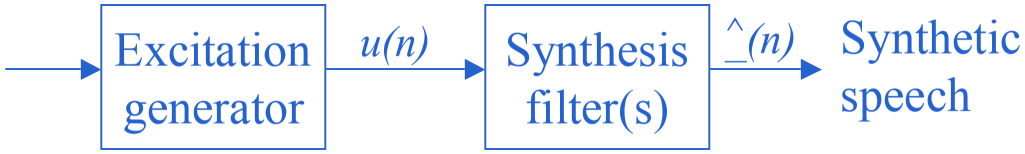
(c) Long and short term removed signal

UCSB

Code-Excited Linear Prediction (CELP) Speech Coding



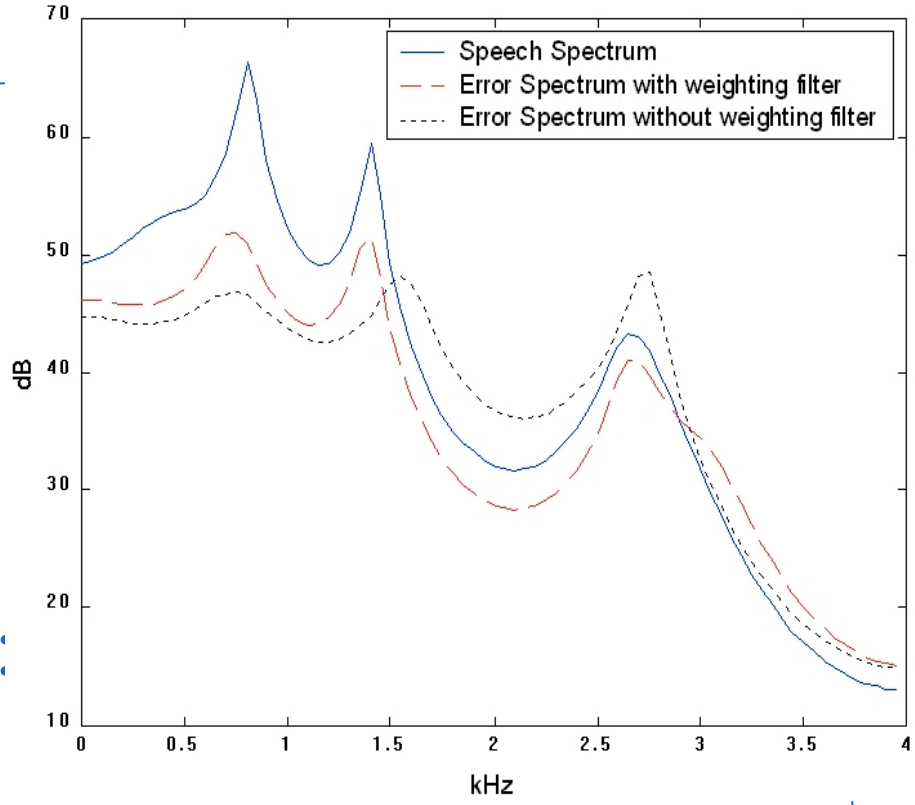
(a) Encoder



(b) Decoder

Speech Coding Techniques

- Algorithms:
  - Parametric coding
  - CELP coding
  - ADPCM coding
  - Transform coding
- Perceptual Masking:
  - Weighting filtering
  - Psychoacoustic modeling



UCSB

Key Advances

- Perceptual Distortion Measures
- Digital Signal Processing
- Analysis-by-Synthesis Structures
  - Codebook Excitation
  - Single Gain for All Pulses

Other Issues in Speech Coding

- Scalable Coding
  - SNR Scalability
  - Bandwidth Scalability
- Variable and Adaptive Multirate Coders
  - Source Controlled for Speech Quality Enhancement
  - Network Controlled for Load Control
  - Joint Source/Channel Coding
- Robustness and Error Concealment

Other Issues in Speech Coding (Cont'd)

- Tandem Coding
- Background Impairments
- Transcoding
- Delay
- Complexity
- Power Consumption

Key Components

- Lossless Coding
  - Huffman and Arithmetic Coding
  - Reversible Variable Length Codes
  - Context Dependent Variable Length Codes
- Unequal Bit Protection
- Side Information (Lots!)
- Signal Processing at the Encoder

More Key Components

- Packetization
- Error Concealment
- Scalability
  - SNR
  - Spatial
  - Temporal
  - Bandwidth

More Key Components

- Variable Rate
  - Source-Driven
  - Network-Driven

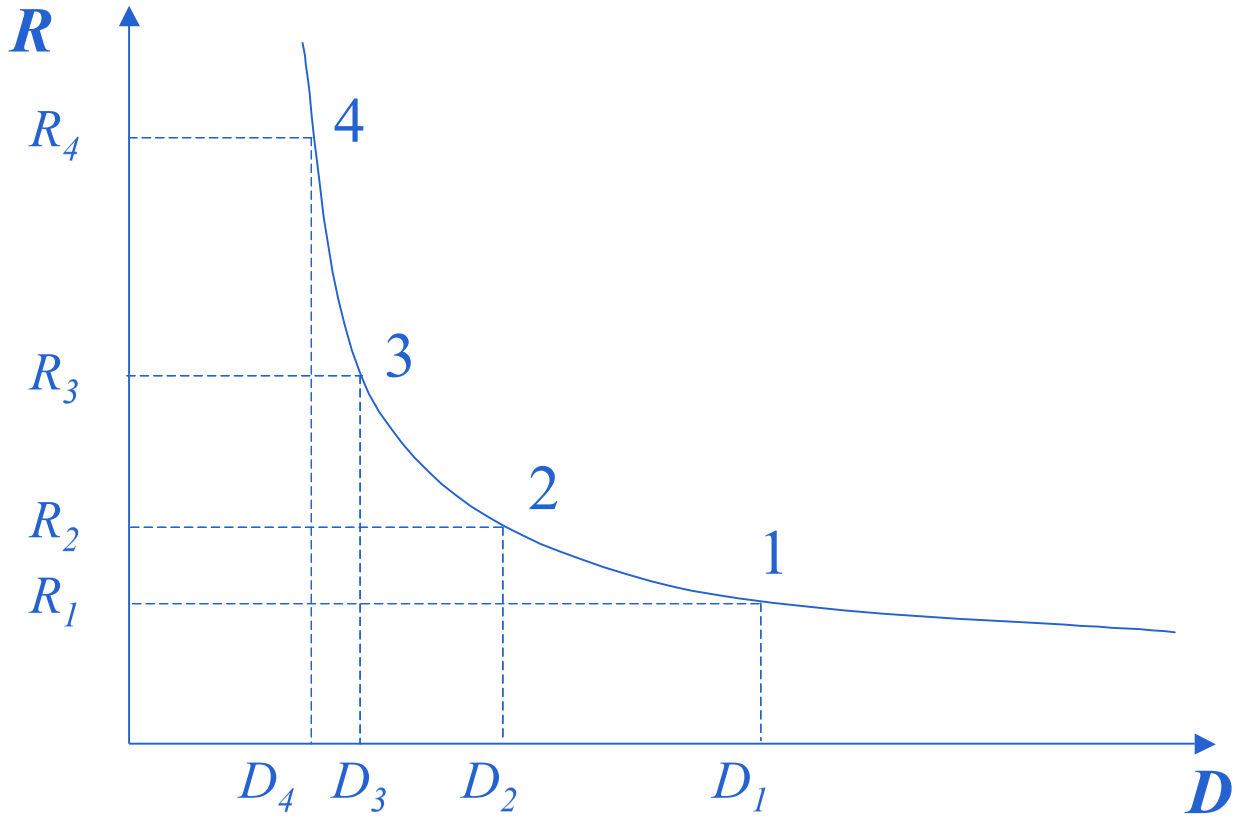
Recent Standards to be Emphasized

- JPEG2000
- H.26L
- MPEG-4
- Multirate Speech and Audio

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

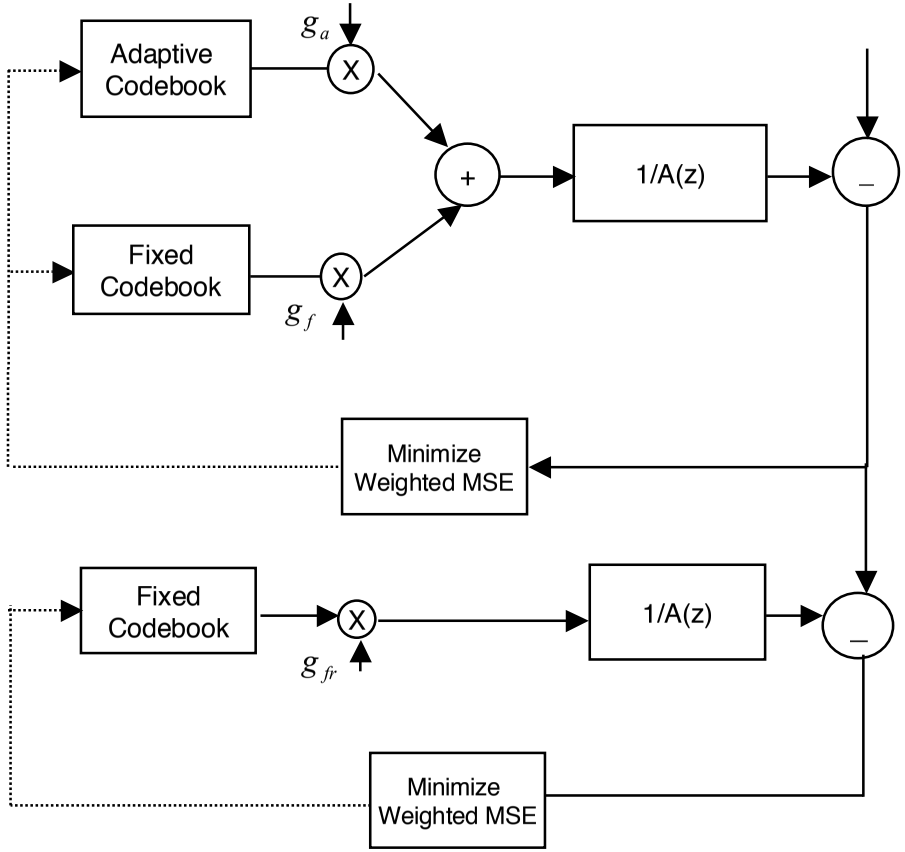
The Successive Refinement Problem



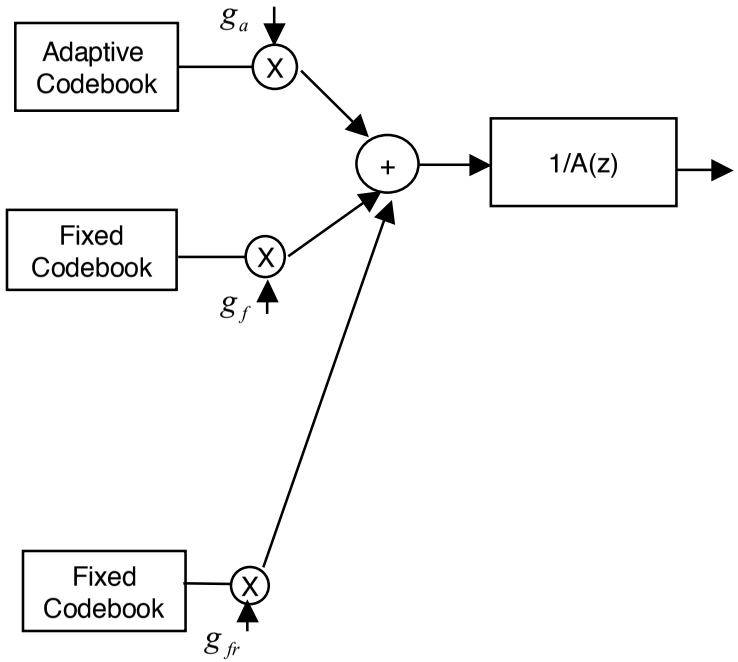
Applications of Scalable Coding

- Multicast transmission over the Internet
- Unequal error protection of the core and enhancement layers
- Multiple layers of quality selectable according to available bandwidth
- Range extension in wireless communications

MPEG-4 CELP Scalable Codec



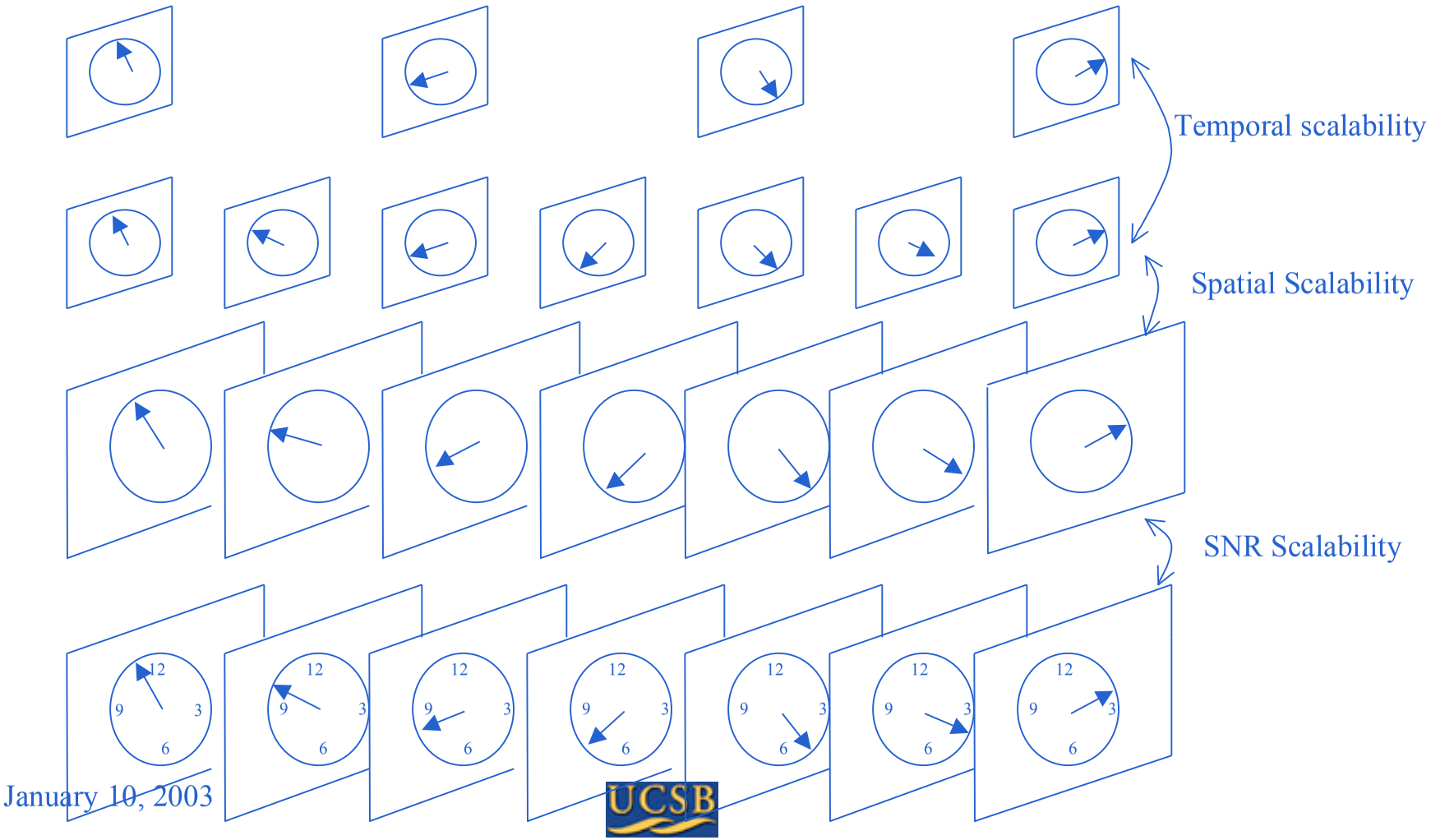
Encoder



Decoder

UCSB

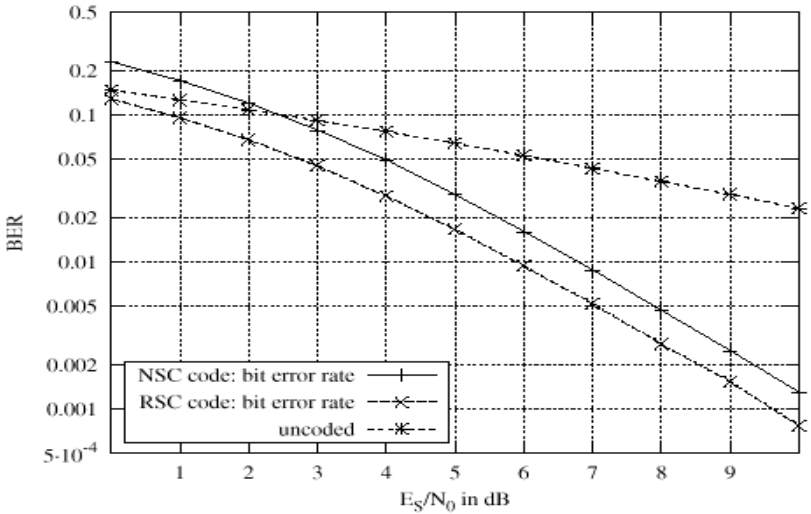
Temporal, Spatial and SNR scalability



UCSB

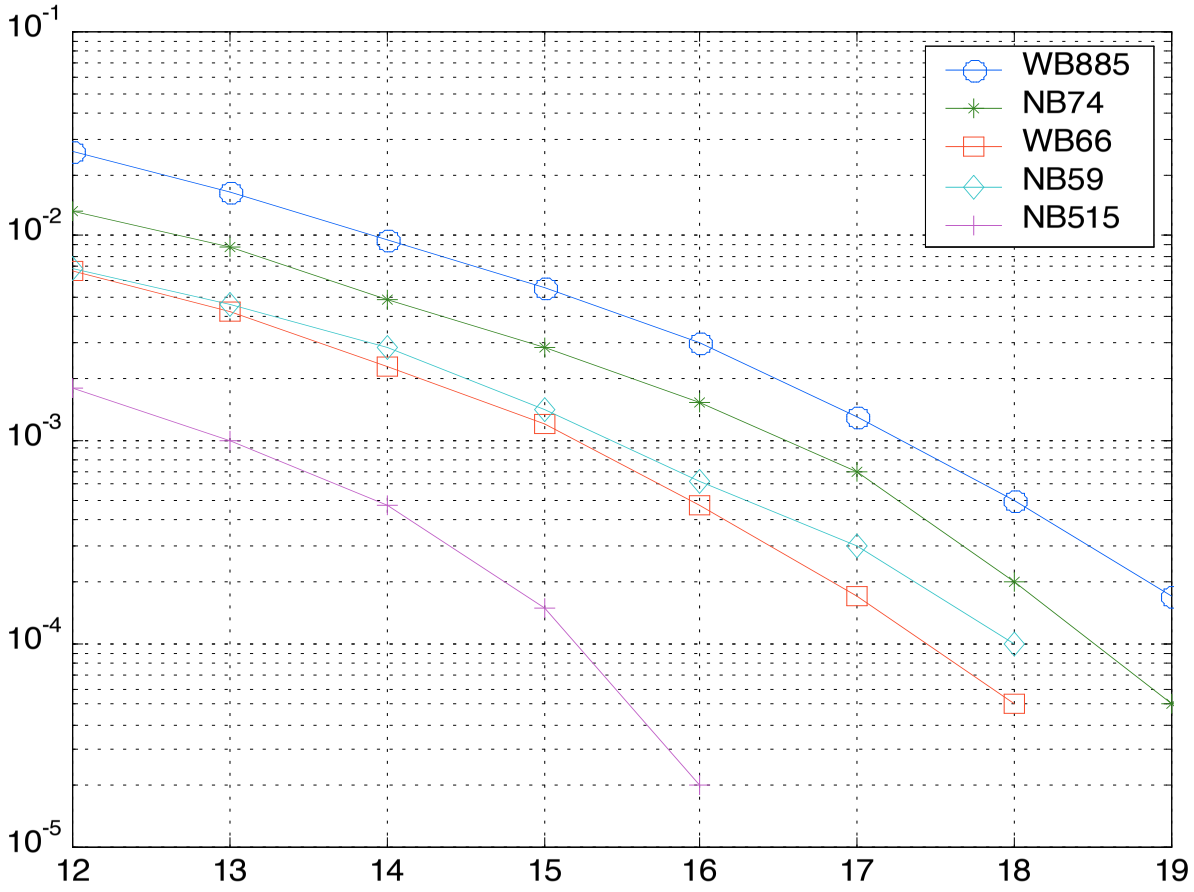
Channel Coding Issues

- Unequal Error Protection with Recursive Systematic Convolutional (RSC) code outperforms the NSC (Hindelang et al), which is applied in 136.



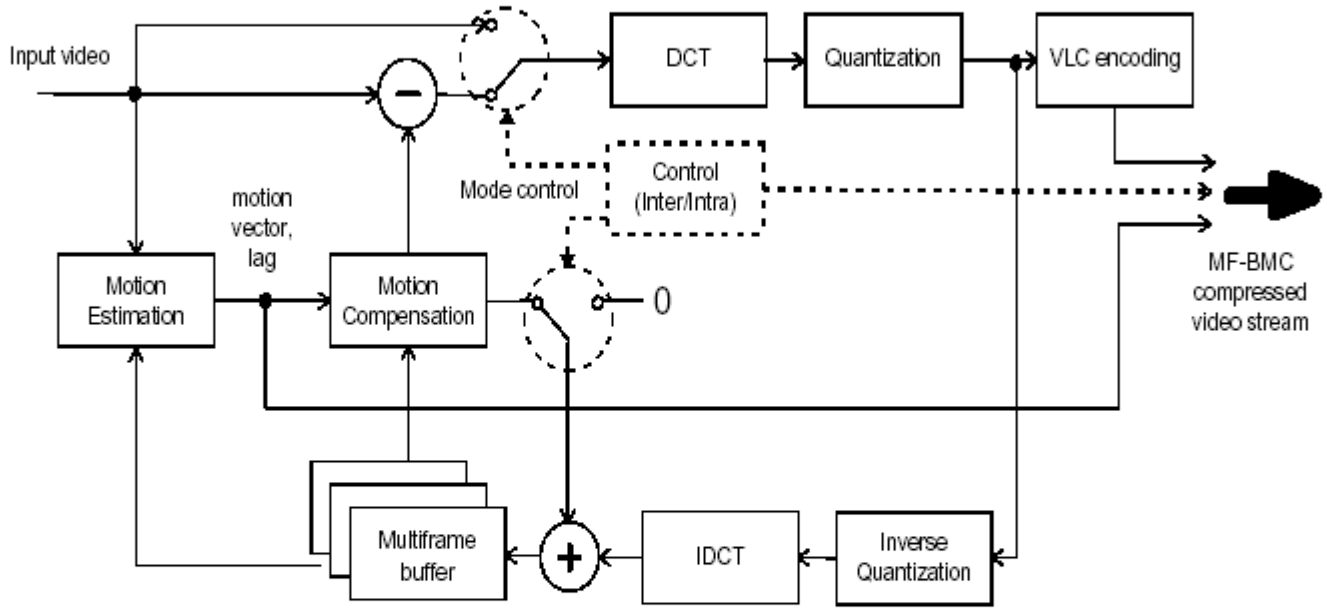
- No-tail-bits convolutional code can save 5 or 6 bits for forward error protection.
- Soft decision channel decoding
- Source-controlled error concealment

Convolutional Decoding Performance



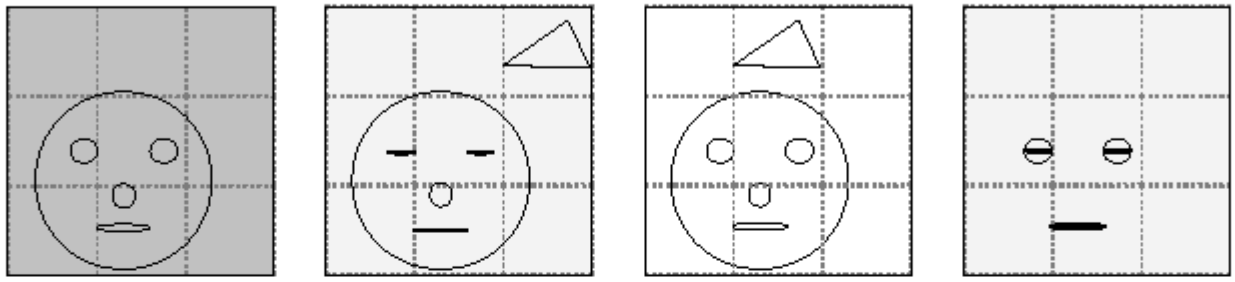
UCSB

Multiframe block motion compensated coder



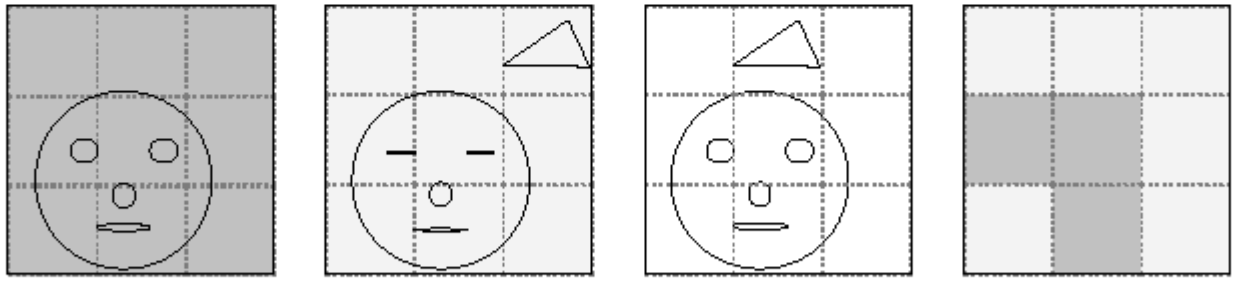
- Based on H.263 - Code based on Telenor R&D H.263 software
- K - Number of previous frames in Multiframe buffer
- lag - Frame number which contains the matching block

Why Multiframe? : Reason 1 - Compression



Frame (k-2) Frame (k-1) Frame k Residual + lag map

(a)



Frame (k-2) Frame (k-1) Frame k Residual + lag map

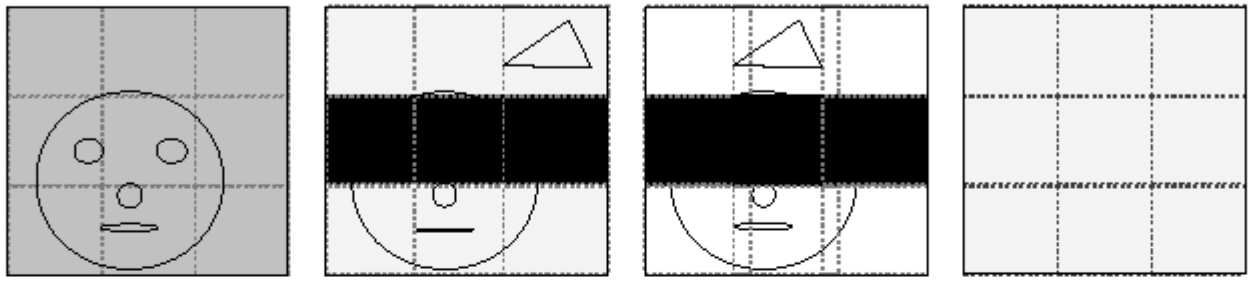
(b)

Results - Compression

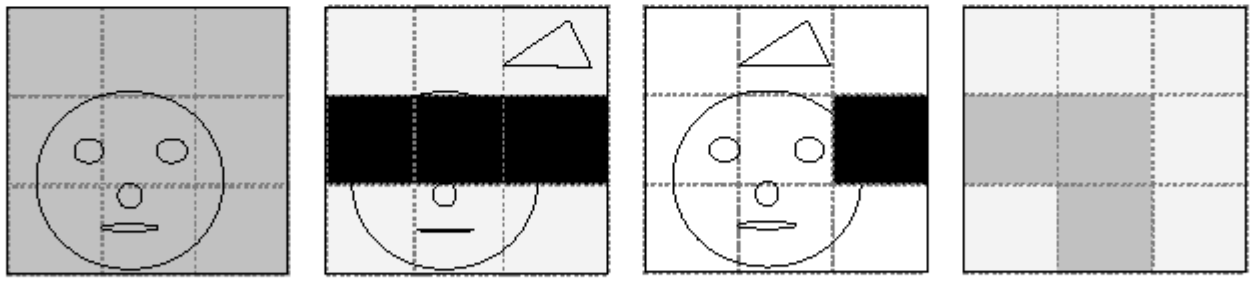
Sequence	max lag (K)	QUANT	Av. DCT bits/frame (Inter)	Av. data rate(Inter) kbps	Average PSNR-Y	Average PSNR-Cb	Average PSNR-Cr
Carphone	1	16	1478.57	32.95	30.25	35.83	36.23
	2	16	1360.32	31.38	30.41	36.23	36.74
	8	16	1249.51	30.76	30.55	36.38	37.05
	8	17	1113.85	28.95	30.37	36.31	36.98
Claire	1	6	1706.50	29.53	39.09	39.75	42.28
	2	6	1682.62	29.29	39.11	39.75	42.28
	8	6	1557.13	27.93	39.19	39.77	42.43
Foreman	1	20	1088.78	32.29	28.60	34.63	34.98
	2	20	1018.41	31.74	28.74	34.83	35.09
	8	20	876.22	31.20	28.97	35.10	35.51
	8	21	807.28	30.26	28.84	35.08	35.38
	8	22	719.60	29.00	28.59	34.87	35.19
Mother and daughter	1	9	1562.49	29.90	33.55	38.53	38.42
	2	9	1522.51	29.47	33.56	38.55	38.45
	8	9	1427.61	28.61	33.65	38.62	38.51

Table 1: Performance of SF-BMC and MF-BMC coders for different test video sequences of QCIF resolution at 12.5 fps.

Why Multiframe? : Reason 2 - Error robustness



(a)

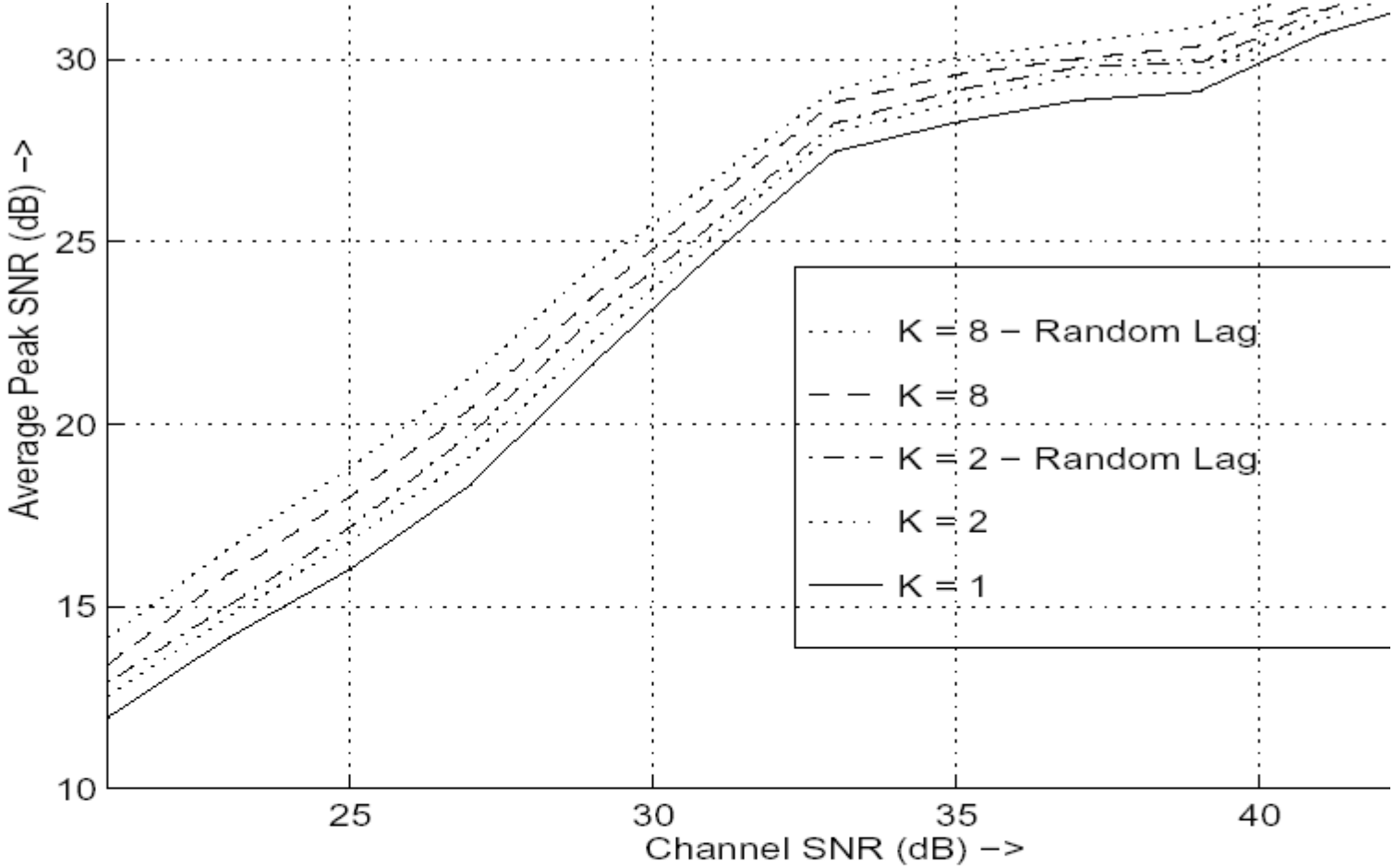


(b)

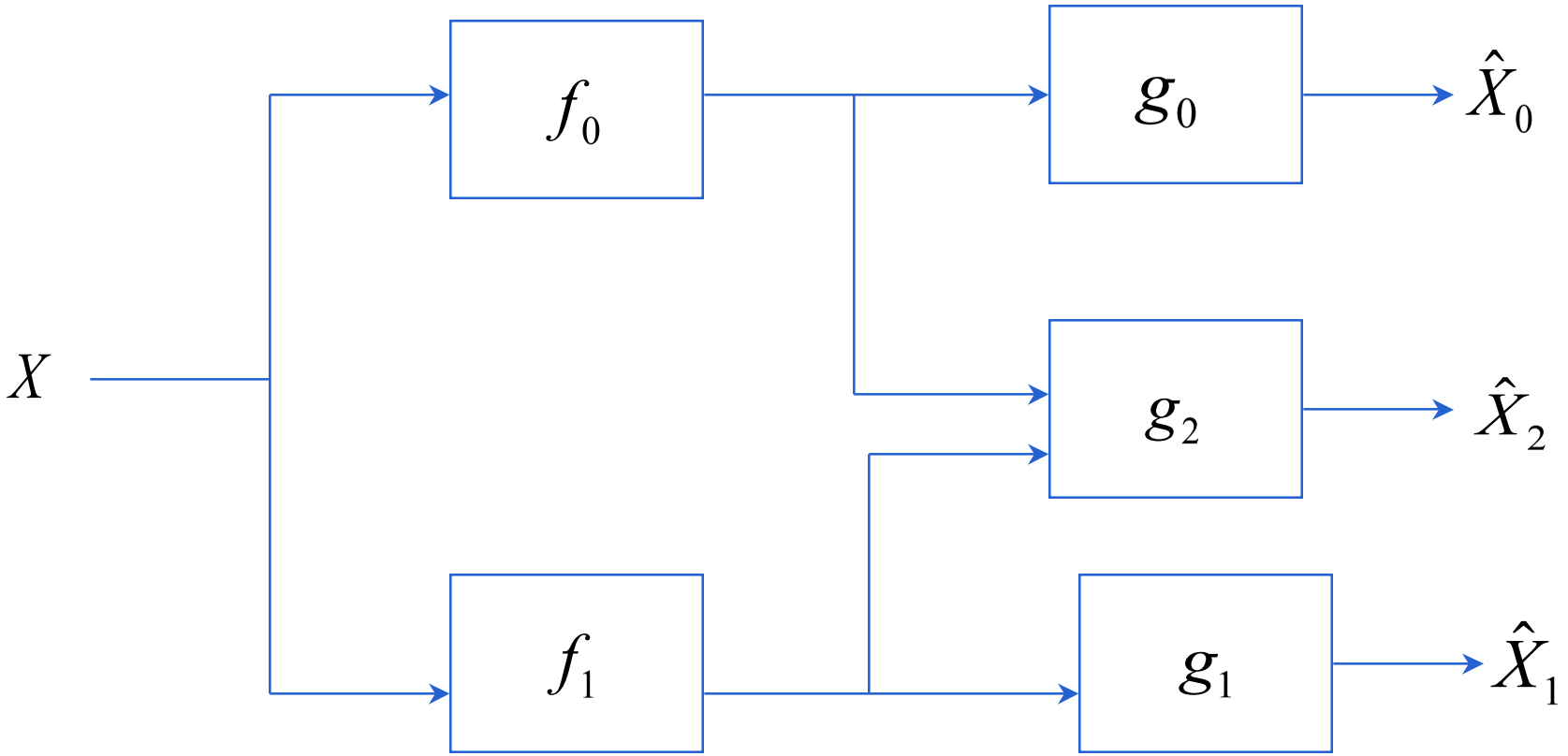
Channel error performance – Bitstreams used

Sequence	max lag (K)	QUANT	Av. data rate(Inter) kbps	Average PSNR
Carphone	1	16	29.42	34.17
	2	16	27.95	34.54
	2 (random lag)	16	28.94	34.54
	8	16	27.32	34.77
	8 (random lag)	16	29.41	34.75
Foreman	1	20	29.57	32.96
	2	20	29.10	33.17
	2 (random lag)	20	29.95	33.27
	8	20	28.45	33.60
	8 (random lag)	20	29.56	33.65

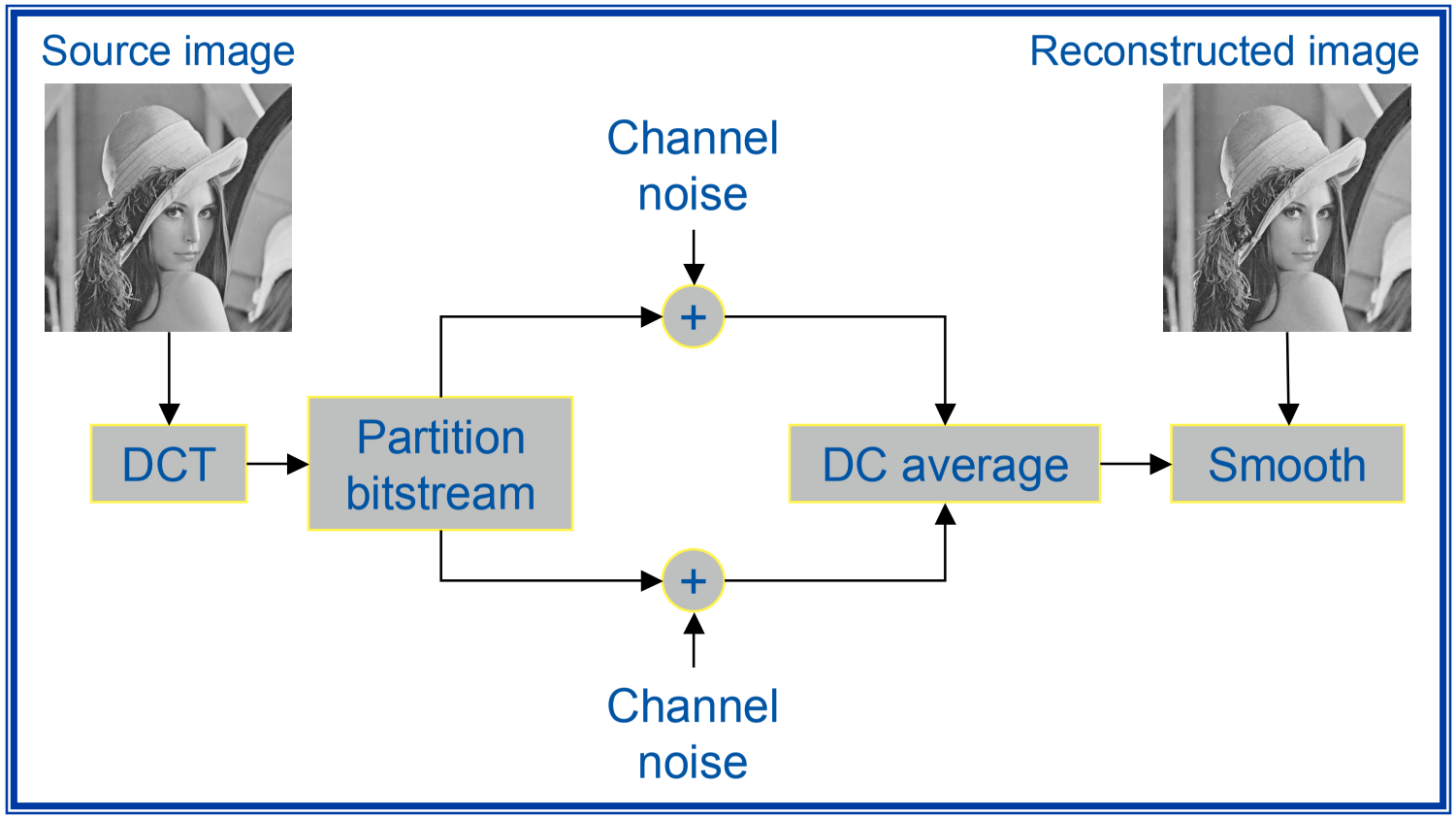
Table 2: SF-BMC and MF-BMC bitstreams used in transmission error performance simulations. Test video sequences (132 frames) are of QCIF resolution at 12.5 fps.


Average Peak SNR (dB) –>
Channel SNR (dB) –>
30
25
20
15
10
25
30
35
40
K = 8 – Random Lag
K = 8
K = 2 – Random Lag
K = 2
K = 1

The Multiple Descriptions Problem



Smoothed description coding



- No rate overhead
- Decent quality at all but highest error rates

Results

- Performance example when p=0.25

No concealment
(PSNR=20.50 dB)

DC averaging only
(PSNR=25.94 dB)

Maximal smoothing only
(PSNR=26.63 dB)

DC averaging + maximal smoothing
(PSNR=28.58 dB)

Results

Goldhill image with p=0.1 at 0.32 bpp. Without (left) and with (right) concealment.

- Comparison of refined SDC with MDC for *lena* image at rate 0.47 bpp