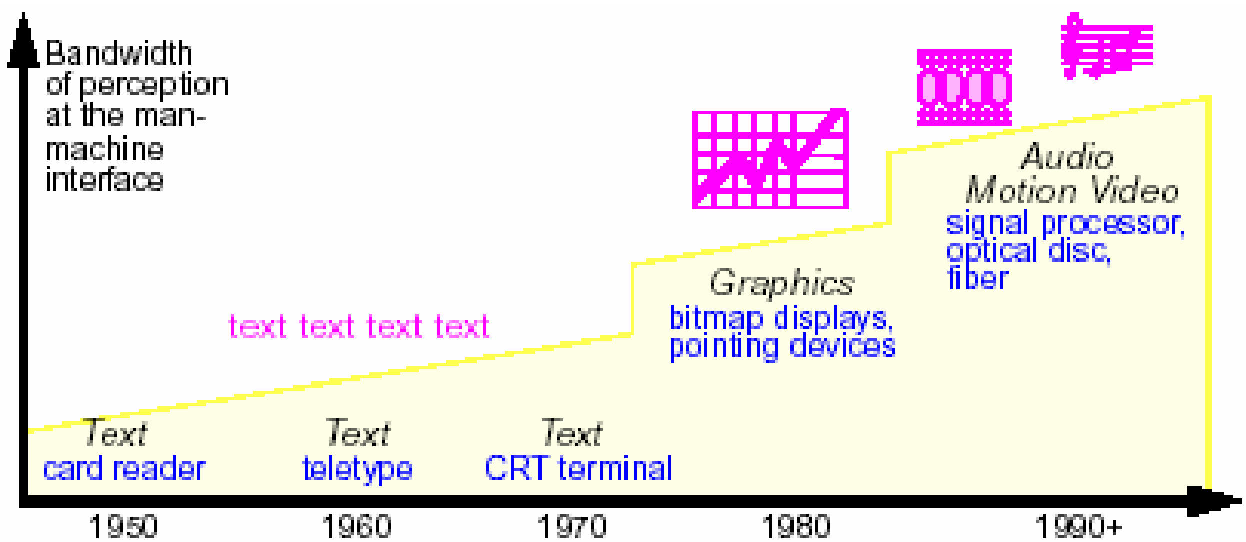
INF 5070 – Media Servers and Distribution Systems

Media and User Behaviour

5/9 – 2005

Media and User Behaviour

- Medium: "Thing in the middle"
  - here: means to distribute and present information
- Media affect human computer interaction



- The mantra of multimedia users
  - Speaking is faster than writing
  - Listening is easier than reading
  - Showing is easier than describing

Dependence of Media

- Time-independent media
  - Text
  - Graphics
  - *Discrete* media

- Time-dependent media
  - Audio
  - Video
  - *Continuous* media

- Interdependant media
  - *Hyper*media
  - *Multi*media

- "Continuous" refers to the user's impression of the data, not necessarily to its representation

Dependence of Media

- Defined by the *presentation* of the data, not its *representation*

- Discrete media
  - Text
  - Graphics
  - Video stills (image displayed by pausing a video stream)
- Continuous media
  - Audio
  - Video
  - Animation
  - Ticker news (continuously scrolling text)
- Multimedia
  - Multiplexed audio and video
  - Subtitled video
  - Video conference

Demand for ***Quality of Service***

- Multimedia approach
  - If you can’t make it, fake it
- Translation
  - Present real-life quality
  - If not possible, save resources where it is not recognizable
- Requirement
  - Know content and environment
  - Understand limitations to user perception
  - If these limitations must be violated, know least disturbing saving options

Media

- Codecs (coders/decoders)
  - Determine how information is represented
  - Important for servers and distribution systems
    - Required sending speed
    - Amount of loss allowed
    - Buffers required
    - …
- Formats
  - Determine how data is stored
  - Important for servers and distribution systems
    - Where is the data?
    - Where is the data about the data?

User Behaviour

- Formalized understanding of
  - users' awareness
  - user behaviour

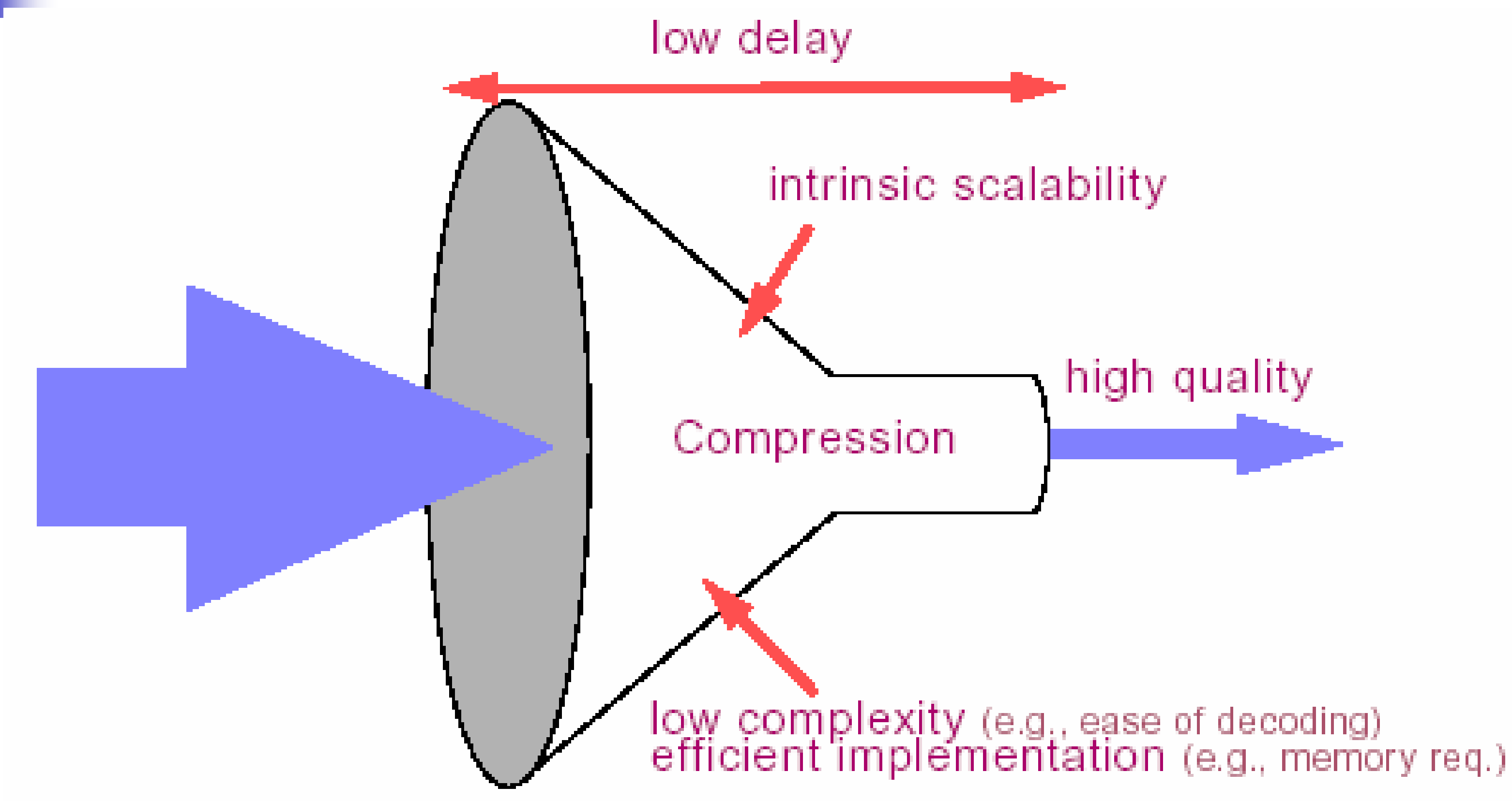
- Achieve the *best* price/performance ratio
- Understand actual resource needs
  - achieve higher compression using lossy compression
  - potential of trading resources against each other
  - potential of resource sharing
  - relax relation between media

Applications of User Modelling

- Encoding Formats
  - Exploit limited awareness of users
    - JPEG/MPEG video and image compression
    - MP3 audio compression
  - Based on medical and psychological models
- Quality Adaptation
  - Adapt to changing resource availability
    - no models - need experiments
- Synchronity
  - Exploit limited awareness of users
    - no models - need experiments
- Access Patterns
  - When will users access a content?
  - Which content will users access?
  - How will they interact with the content?
    - no models, insufficient experiments - need information from related sources

Coding for distribution

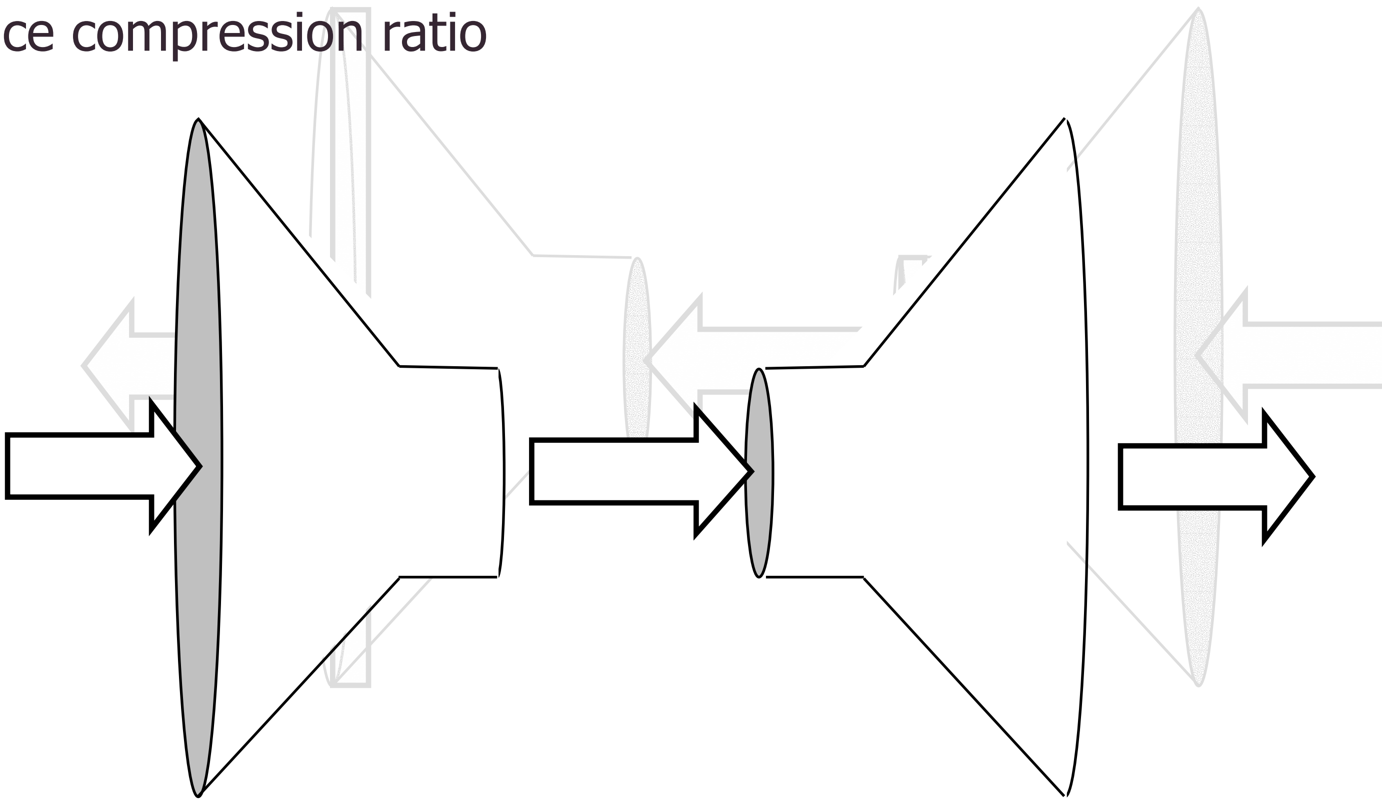
Compression – General Requirements



- *Dependence* on application type:
  - *Interactive* applications (dialog mode)
  - *Non-interactive* applications (retrieval mode)

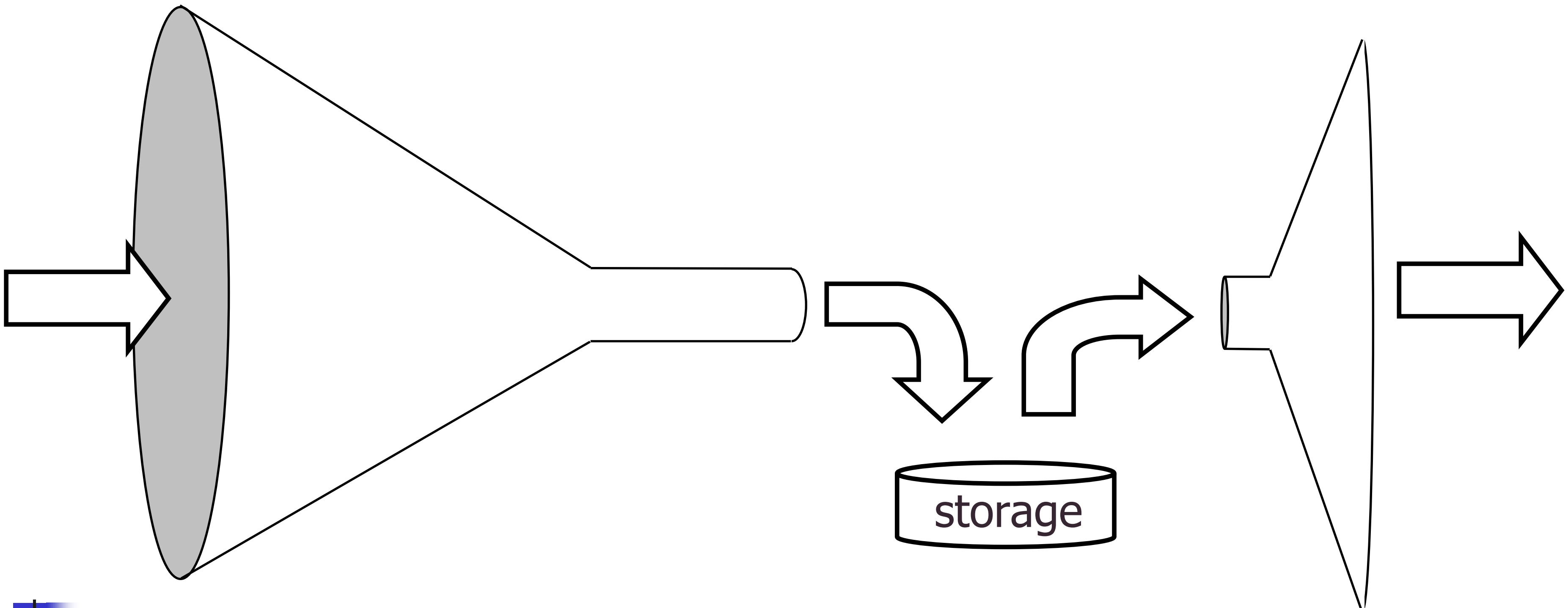
Compression – General Requirements

- Interactive applications
  - Focus on
    - Low delay
    - Low complexity
    - Symmetry
  - Sacrifice compression ratio

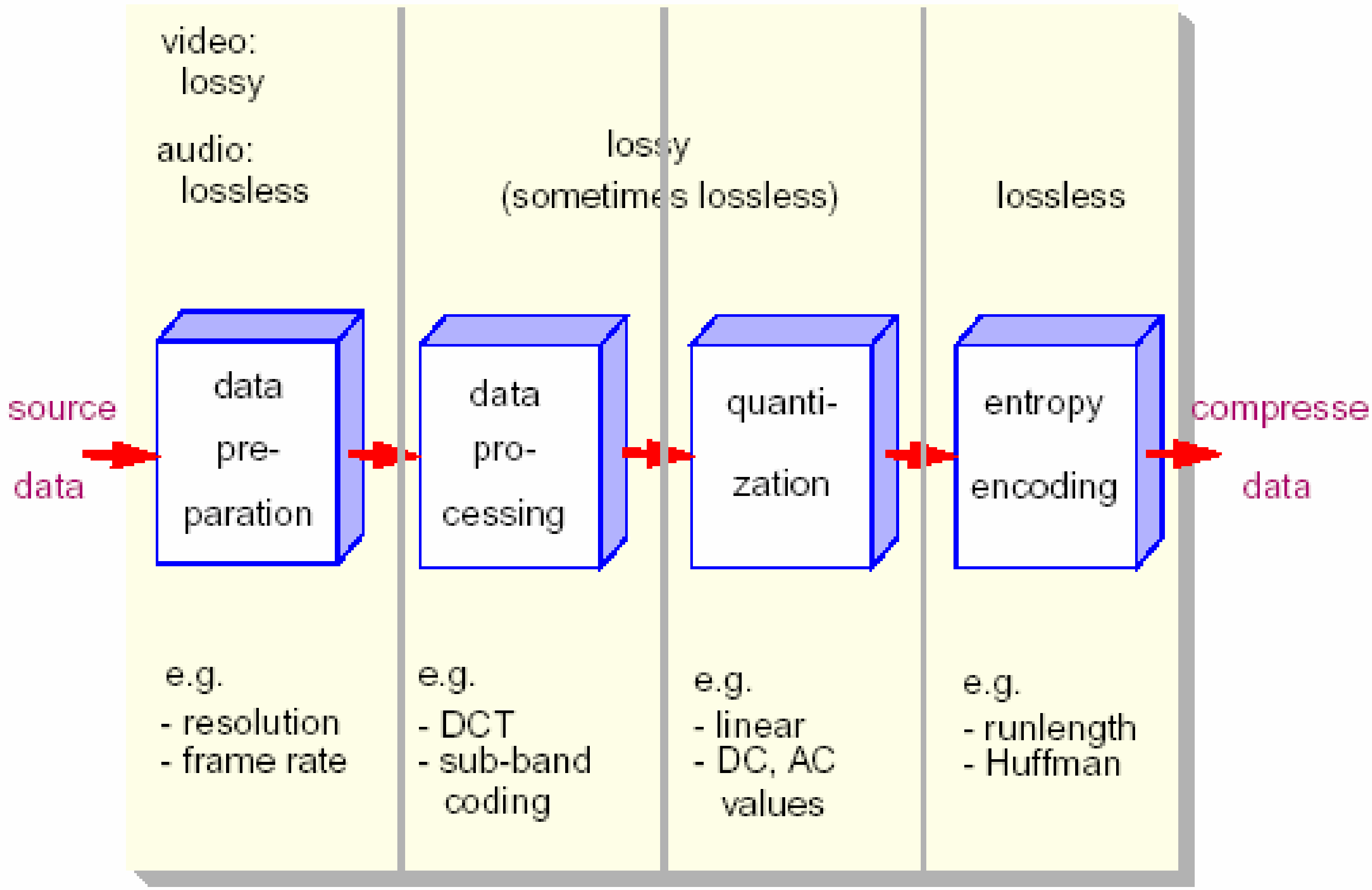


Compression – General Requirements

- Non-nteractive applications
  - Focus on
    - High compression
    - Low complexity on receiver side
    - Low delay on receiver side
  - Accept asymmetry

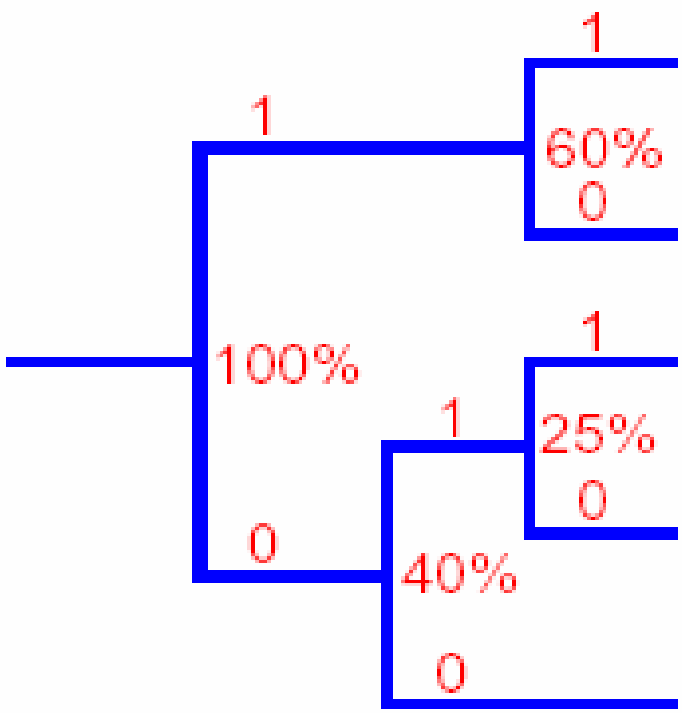


Basic Encoding Steps



Huffman Coding

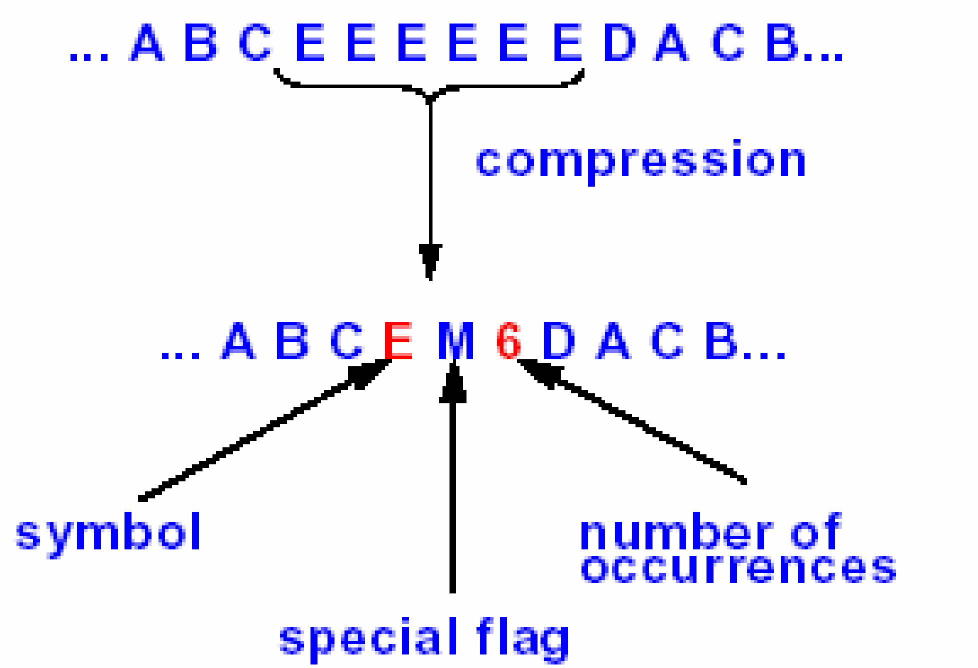
- Assumption
  - Some symbols are more frequent than others
- Example
  - Given: A, B, C, D, E
  - Probability to occur: p(A)=0.3, p(B)=0.3, p(C)=0.1, p(D)=0.15, p(E)=0.15



30%	A	11
30%	B	10
10%	C	011
15%	D	010
15%	E	00

Run-Length Coding

- Assumption
  - Long sequences of identical symbols
- Example



Bit-Plane Coding

- Assumption
  - Even longer sequences of identical bits
- Example

```
10,0,6,0,0,3,0,2,2,0,0,2,0,0,1,0, … ,0,0 (absolute)
 0,x,1,x,x,1,x,0,0,x,x,1,x,x,0,x, … ,x,x (sign bits)

 1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0, … ,0,0 (MSB   → 8)
 0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0, … ,0,0 (MSB-1 → 4)
 1,0,1,0,0,1,0,1,1,0,0,1,0,0,0,0, … ,0,0 (MSB-2 → 2)
 0,0,0,0,0,1,0,0,0,0,0,0,0,0,1,0, … ,0,0 (MSB-3 → 1)
 (0,1)
 (2,1)
 (0,0)(1,0)(2,0)(1,0)(0,0)(2,1)
 (5,0)(8,1)
```

End of plane

No 0s before a 1

➢ Up to 20% savings over run-length coding can be achieved

JPEG

- "JPEG": Joint Photographic Expert Group

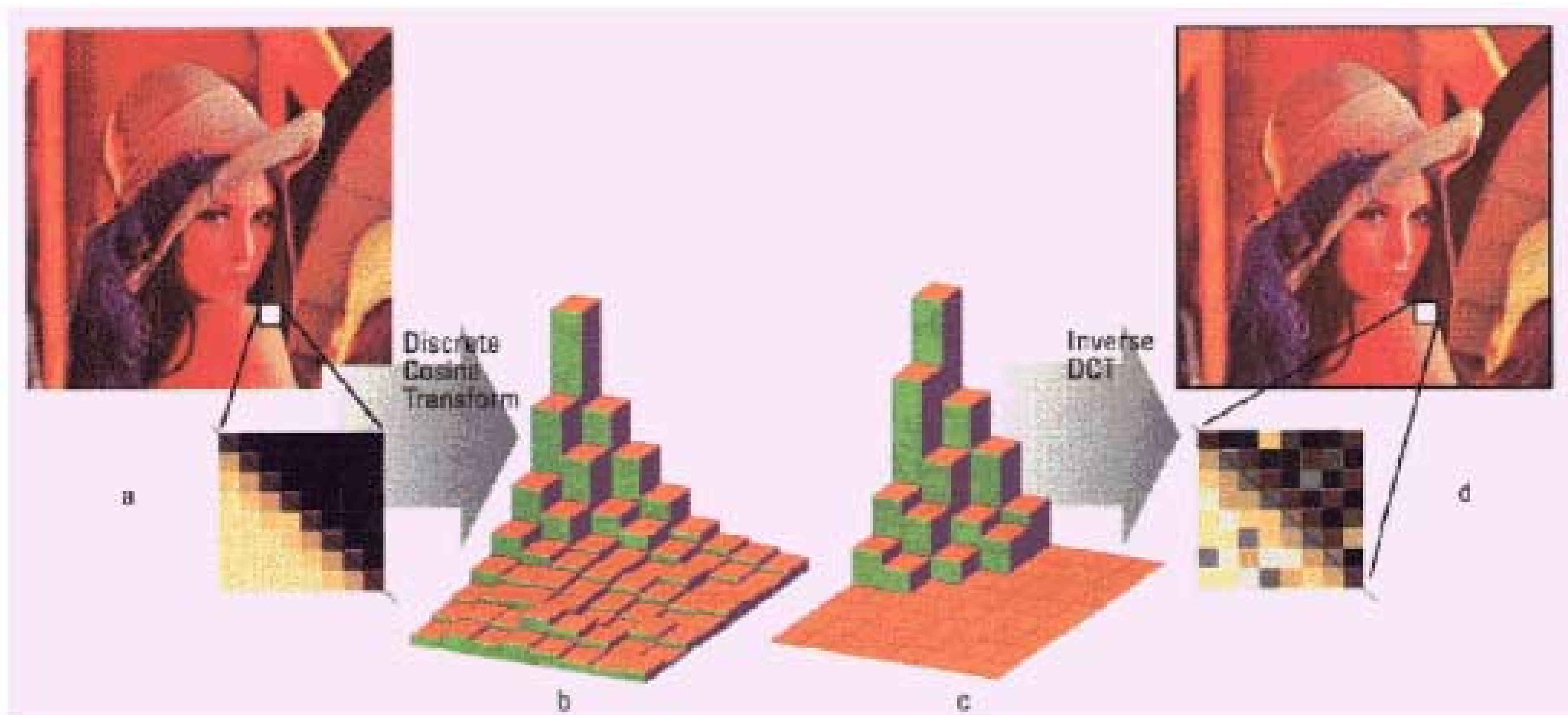
- International Standard:
  - For digital compression and coding of continuous-tone still images
  - Gray-scale and color

- Compression rate of 1:10 yields reasonable results
  - Lossless mode: reasonable compression rate approx. 1:1.6

- Independence of
  - Image resolution
  - Image and pixel aspect ratio
  - Color representation
  - Image complexity and statistical characteristics

JPEG – Baseline Mode: Quantization

- Use of quantization tables for the DCT-coefficients
  - Map interval of real numbers to one integer number
  - Allows to use different granularity for each coefficient



Motion JPEG

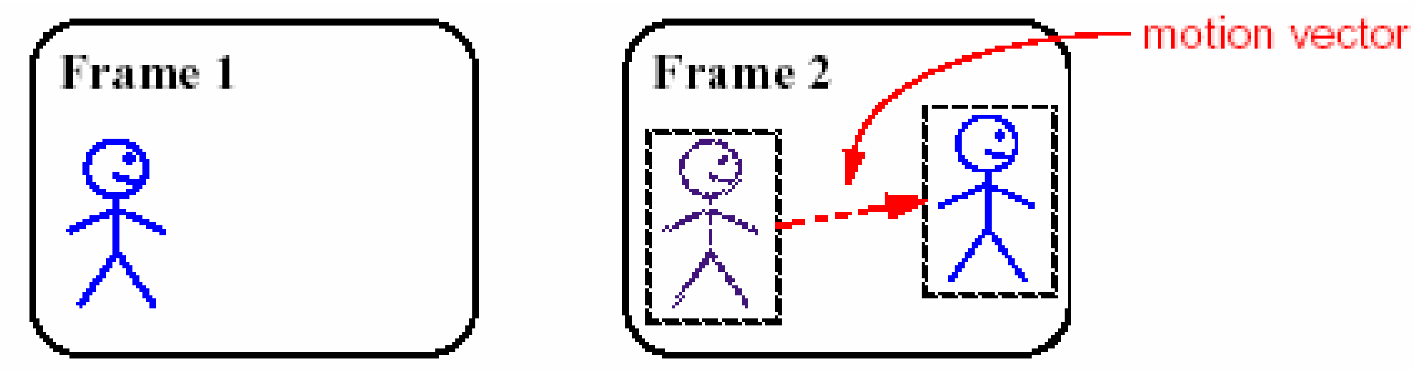
- Use series of JPEG frames to encode video

- Pro
  - Lossless mode – editing advantage
  - Frame-accurate seeking – editing advantage
  - Arbitrary frame rates – playback advantage
  - Arbitrary frame skipping – playback advantage
  - Scaling through progressive mode – distribution advantage
  - Min transmission delay = 1/framerate – conferencing advantage
  - Supported by popular frame grabbers

- Contra
  - Series of JPEG-compressed images
  - No standard, no specification
    - Worse, several competing quasi-standards
  - No relation to audio
  - No inter-frame compression

H.261 (px64)

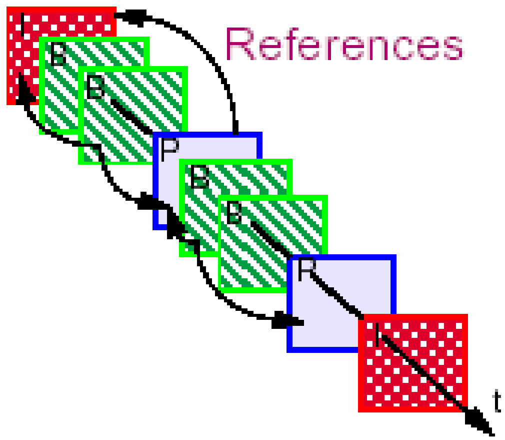
- International Standard
  - Video codec for video conferences at p x 64kbit/s (ISDN):
    - Real-time encoding/decoding, max. signal delay of 150ms
    - Constant data rate
- *Intraframe coding*
  - DCT as in JPEG baseline mode
- *Interframe coding*, motion estimation



  - Search of similar macroblock in previous image and compare
    - Position of this macroblock defines motion vector
    - Difference between similar macroblocks

MPEG (Moving Pictures Expert Group)

- International Standard:
  - Compression of audio and video for playback (1.5 Mbit/s)
  - Real-time decoding
- Sequence of I-, P-, and B-Frames:



- I-Frames (Intracoded)
- P-Frames (Predictive Coded)
- B-Frames (Bidirectionally Coded)

Example: I B B P B B P B B I B B P B B P B B ...

- Random access
  - at I-frames
  - at P-frames: i.e. decode previous I-frame first
  - at B-frame: i.e. decode I and P-frames first

MPEG-2

- From MPEG-1 to MPEG-2
  - Higher data rates
    - MPEG-1: about 1.5 MBit/s
    - MPEG-2: 2-100 MBit/s

- Use cases
  - Program Stream
    - DVD
    - for post-processing, storage
  - Transport Stream
    - DVB-T: Terrestrial
    - DVB-S: Satellite
    - DVB-C: Cable

- Scaling
  - Signal to Noise Ration (SNR) scaling - progressive compression error correcting codes
  - Spatial scaling - several pixel resolutions
  - Temporal scaling - frame dropping

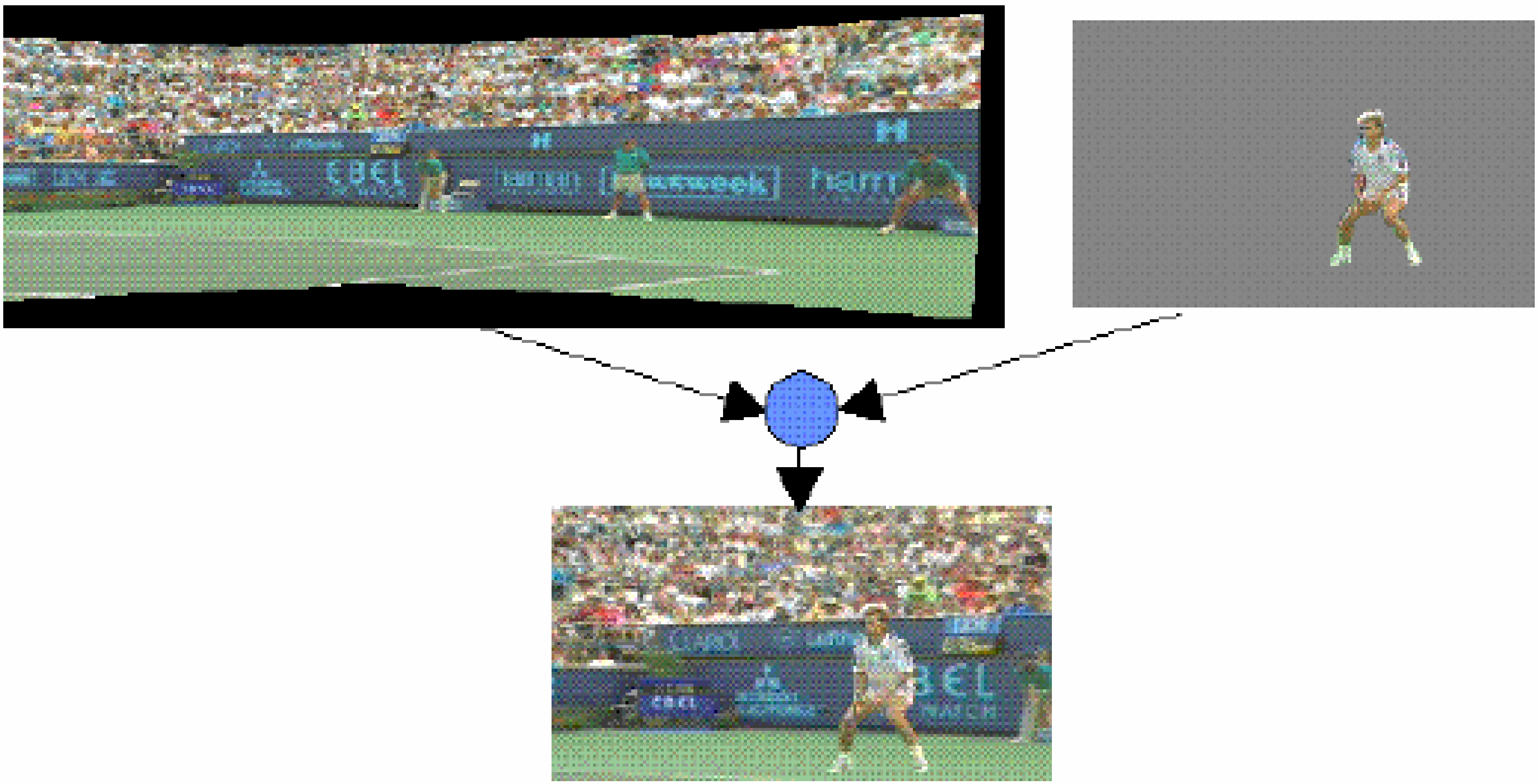
MPEG-4

- MPEG-4 originally
  - Targeted at systems with very scarce resources
  - To support applications like
    - Mobile communication
    - Videophone and E-mail
  - Max. data rates and dimensions (roughly)
    - Between 4800 and 64000 bits/s
    - 176 columns x 144 lines x 10 frames/s
- Further demand
  - To provide enhanced functionality to allow for analysis and manipulation of image contents

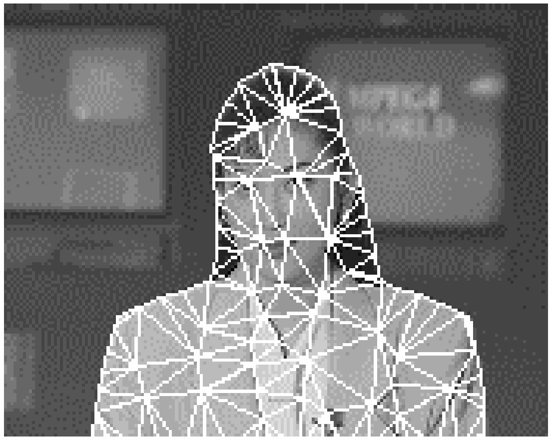
MPEG-4: Scope

- Definition of
  - „System Decoder Model“
    - specification for decoder implementations
  - Description language
    - binary syntax of an AV object’s bitstream representation
    - scene description information
  - Corresponding concepts, tools and algorithms, especially for
    - content-based compression of simple and compound audiovisual objects
    - manipulation of objects
    - transmission of objects
    - random access to objects
    - animation
    - scaling
    - error robustness

MPEG-4: Example of a Composition



MPEG-4: Synthetic Objects



- Visual objects
  - Virtual parts of scenes
    - e.g. virtual background
  - Animation
    - e.g. animated faces
- Audio objects
  - „Text-to-speech"
    - speech generation from given text and prosodic parameters
    - face animation control
  - „Score driven synthesis"
    - music generation from a score
    - more general than MIDI
  - Special effects

Multimedia File Formats

Overview

- File formats
  - Define the storage of media data on disks
  - Specify synchronization
  - Specify timing
  - Contain metadata

- They allow
  - Interchange of data without interpretation
    - Copying
    - Platform independance
  - Management
  - Editing
  - Retrieval for presentation

- Needed for all asynchronous applications

File Format Examples

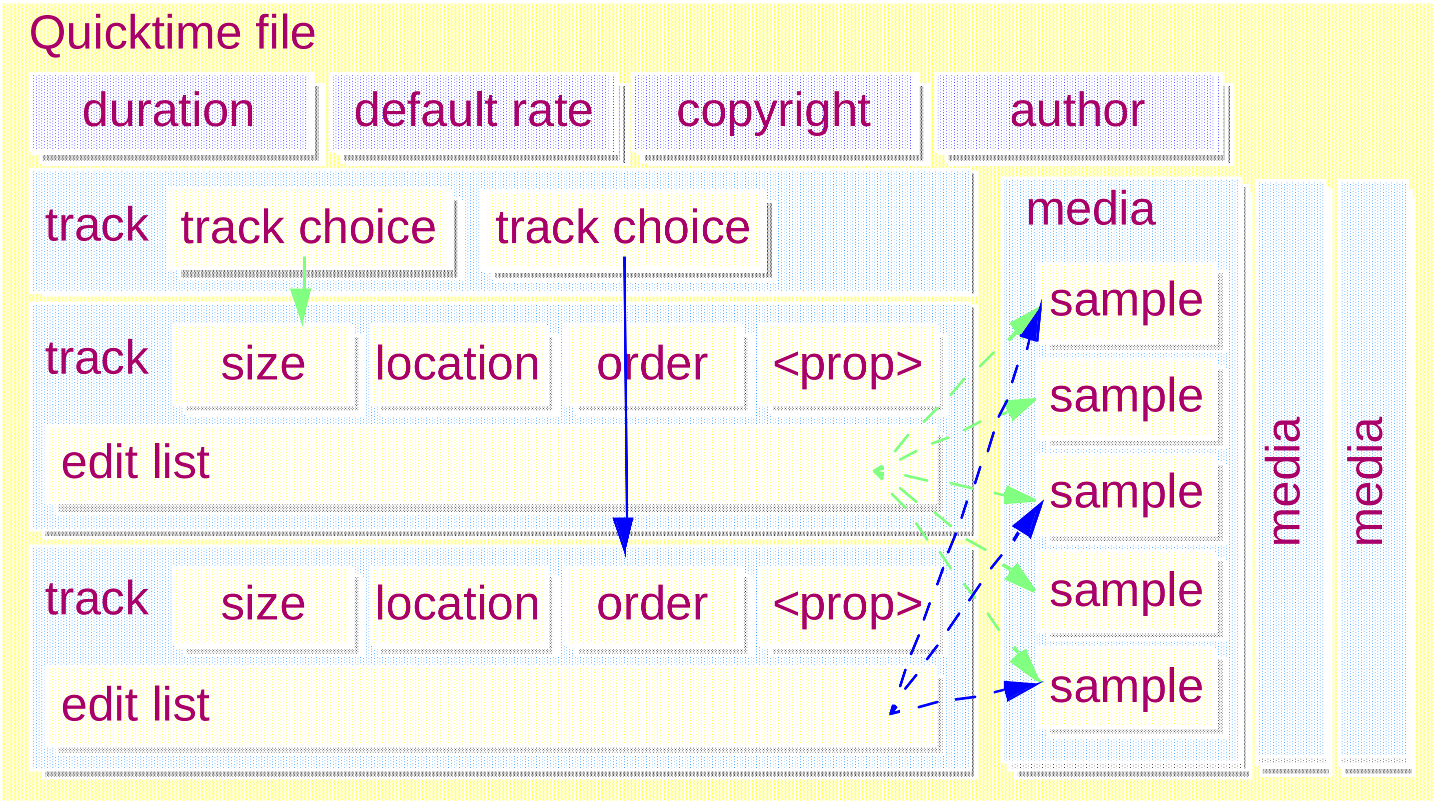
No file formats	Stored wire formats	Generic file formats
H.261 H.263 (Motion JPEG)	DVI MPEG-1 MPEG-2	MPEG-4 Quicktime AVI / Windows Media Real Video

- Streaming format
  - File format and wire format are identical
  - MPEG-1, DVI
- Streamable format
  - File format specifies wire format(s)
  - MPEG-4, Quicktime, Windows Media, Real Video

Stored Motion JPEG

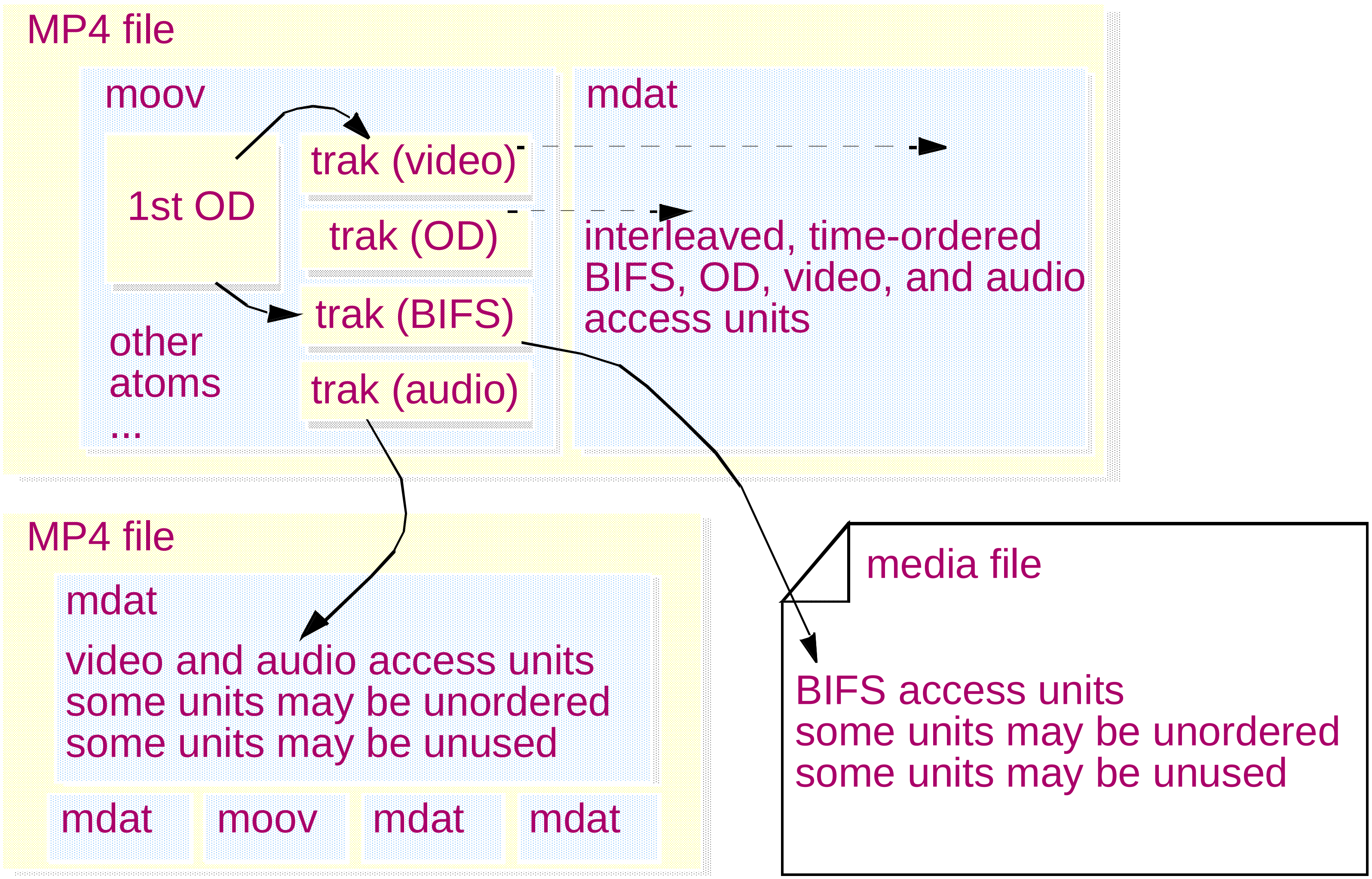
- Motion JPEG Chunk File Format (UC Berkeley)
  - Specifies entire clip's length in s+ns
  - Contains sequence of images
  - Each image in Independent JPEG Group's JFIF format
- AVI MJPEG DIB (Microsoft)
  - Supports audio interleaving
  - Time-stamped data chunks
  - One frame per AVI RIFF data chunk
  - Hack for file size > 1GB
- Quicktime (Apple)
  - Dedicated tracks for interleaving and timing
  - One frame per field
  - Several fields per sample
  - Formats A: full JFIF images, B: QT headers and data only

Quicktime File Format



- Run-time choice of tracks
  - availability of codecs
  - bandwidth
  - language

MPEG-4 File Format



Other File Formats

- Real Video
  - Not published – no source included in Helix
  - Supports various codecs
  - Supports various encoding formats per file
  - Supports dynamic selection
  - Supports dynamic scaling ("stream thinning")
- AVI
  - AVI is published
  - Uses Resource Interchange File Format (RIFF)
  - Supports various codecs
- ASF / Windows Media File Format
  - Submitted as MPEG-4 proposal (but refused)
  - ASF files can include Windows binary code
  - ASF is patented in the USA

Network-aware coding

Network-aware coding

- Adapt to reality of the Internet
  - Content
    - Is created once, off-line
    - Is sent many times, under different circumstances
  - No guarantees concerning
    - Throughput
    - Jitter
    - Packet loss
  - Sending rate
    - Must adhere to rules
    - Often: don’t send more than TCP would

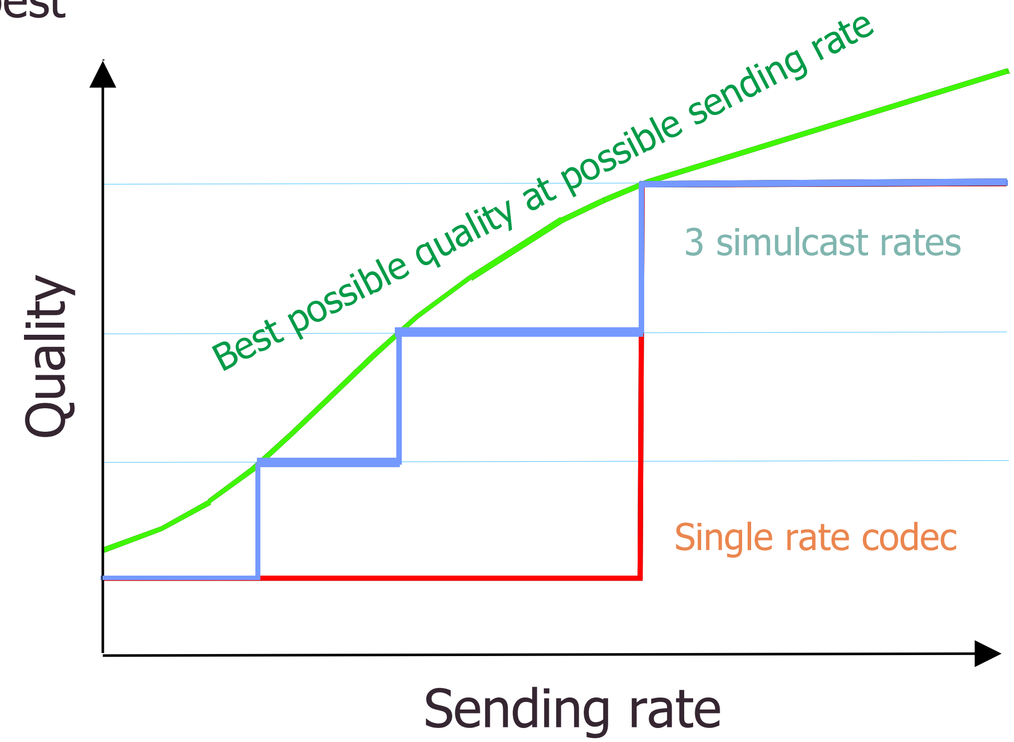
  - Can’t send at the best available encoding rate

Approaches

- Simulcast
- Scalable coding
  - SNR Scalability
  - Temporal Scalability
  - Spatial Scalability
  - Fine Grained Scalability
- Multiple Description Coding

Simulcast

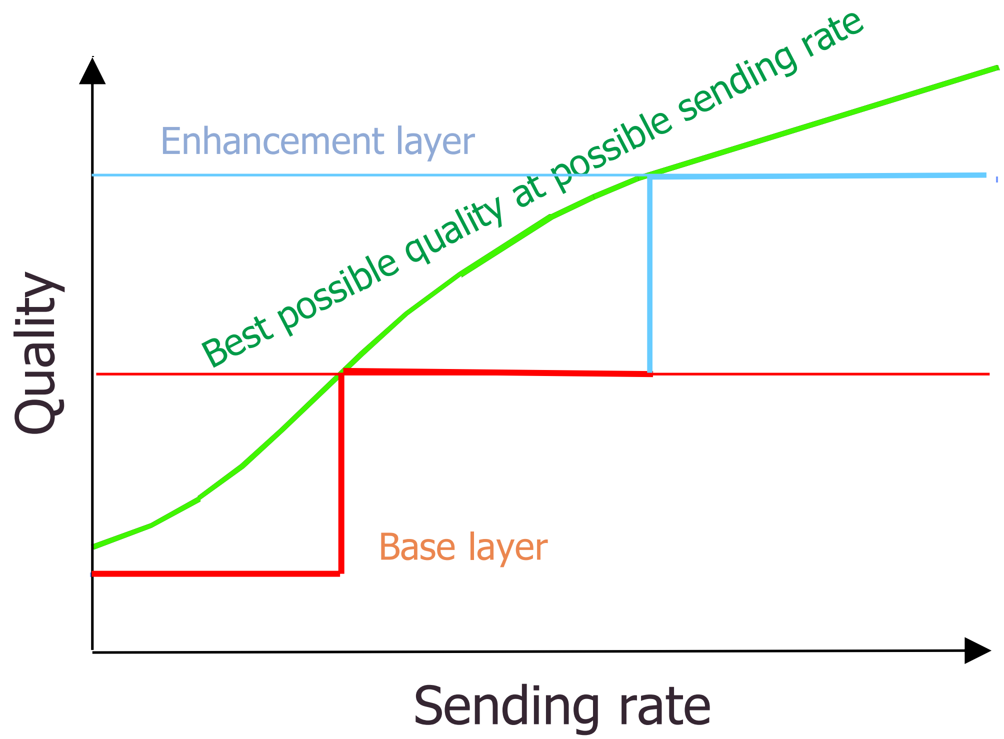
- Choose a set of sending rates
  - During content creation
    - Encode content in best possible quality below that sending rate
  - During transmission
    - Choose version with the best admissable quality



Scalable coding

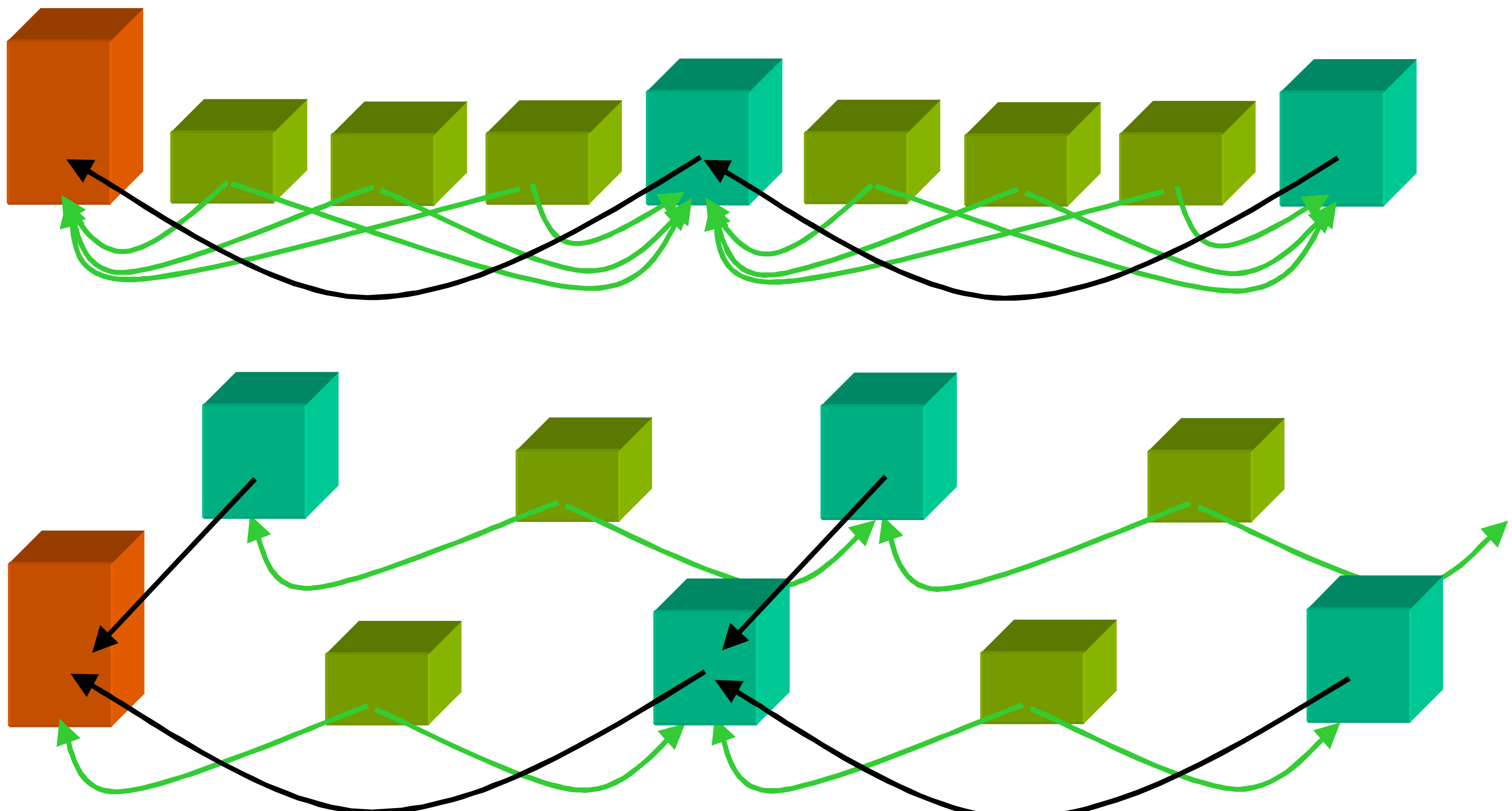
- Typically used as
  Layered coding

- A base layer
  - Provides basic quality
  - Must always be transferred
- One or more enhancement layers
  - Improve quality
  - Transferred if possible



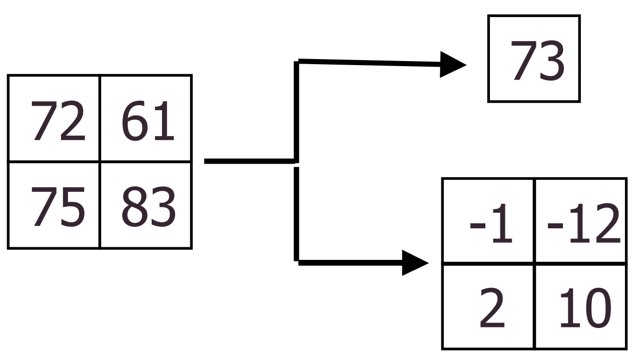
Temporal Scalability

- Frames can be dropped
  - In a controlled manner
  - Frame dropping does not violate dependencies
  - Low gain example: B-frame dropping in MPEG-1



Spatial Scalability

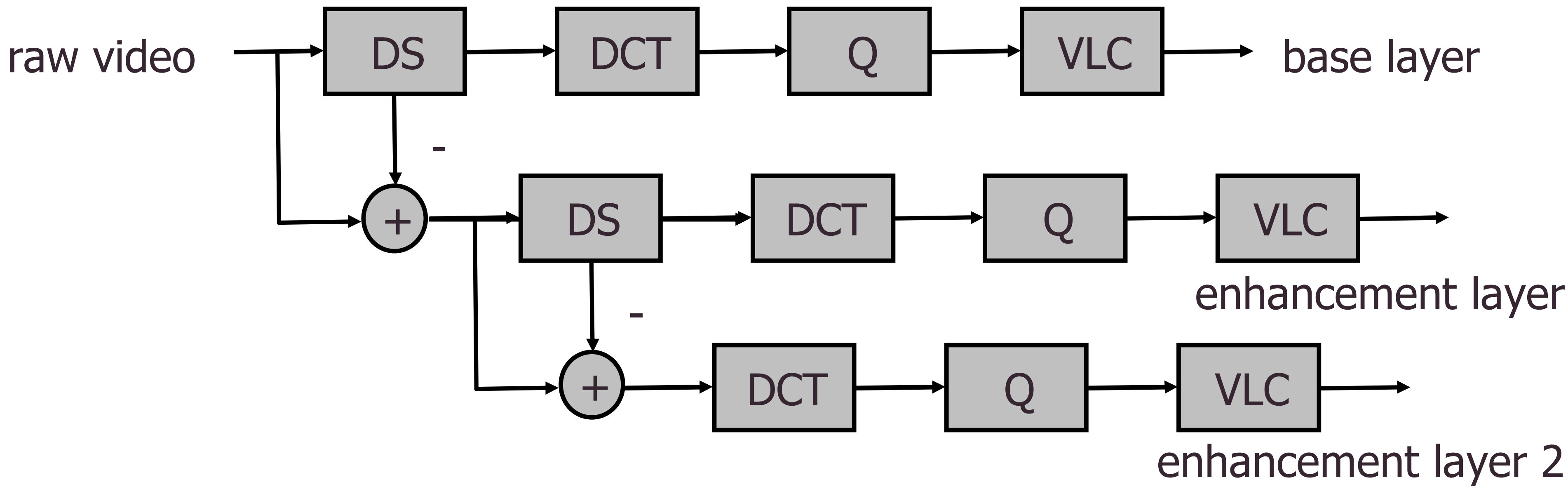
- Idea
  - Base layer
    - Downsample the original image (code only 1 pixel instead of 4)
    - Send like a lower resolution version
  - Enhancement layer
    - Subtract base layer pixels from all pixels
    - Send like a normal resolution version
  - If enhancement layer arrives at client
    - Decode both layers
    - Add layers



Base layer
Less data to code

Enhancement layer
Better compression due to low values

Spatial Scalability



DS - downsampling

DCT – discrete cosine transformation

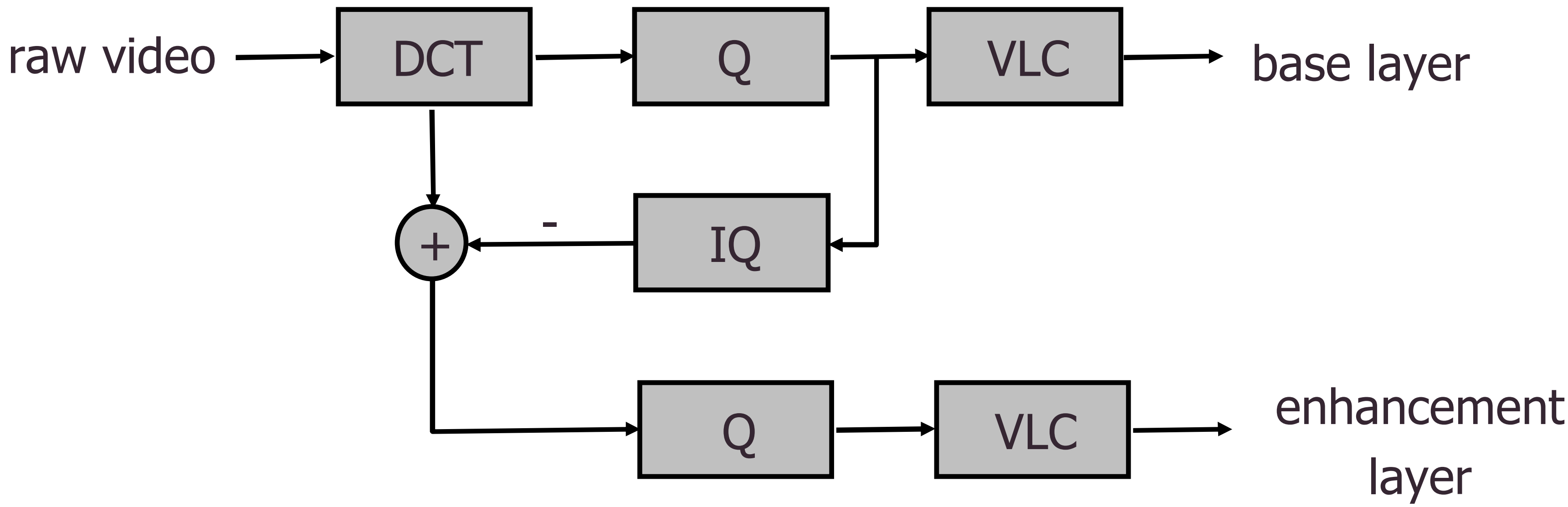
Q – quantization

VLC – variable length coding

SNR Scalability

- SNR – signal-to-noise ratio
- Idea
  - Base layer
    - Is regularly DCT encoded
    - A lot of data is removed using quantization
  - Enhancement layer is regularly DCT encoded
    - Run Inverse DCT on quantized base layer
    - Subtract from original
    - DCT encode the result
  - If enhancement layer arrives at client
    - Add base and enhancement layer before running Inverse DCT

SNR Scalability



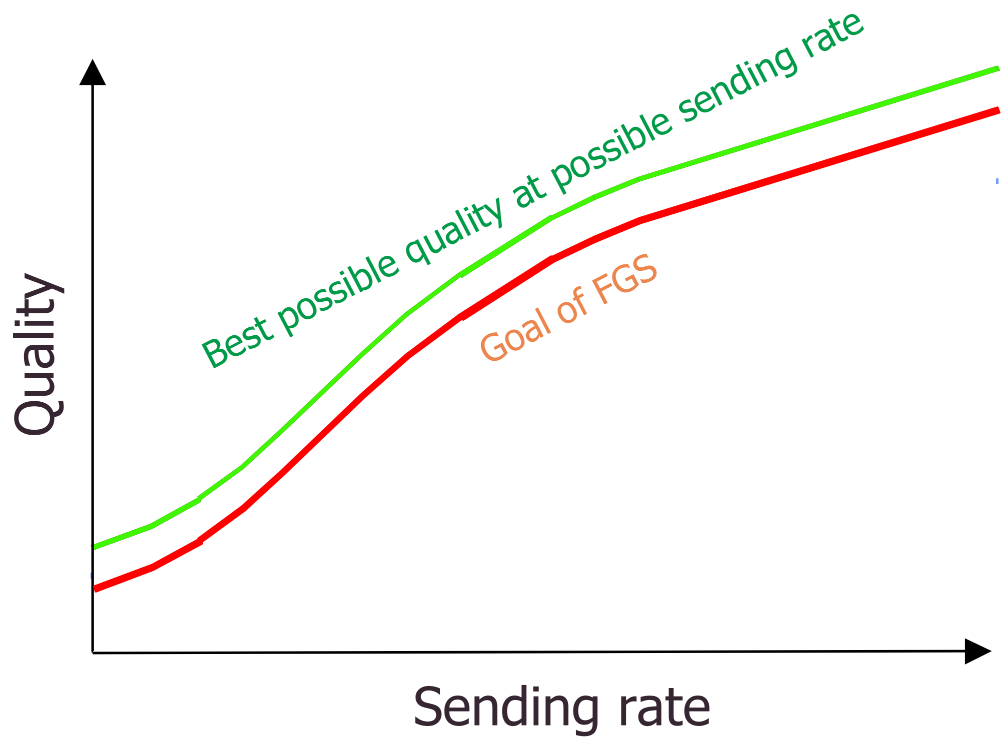
DCT – discrete cosine transformation
Q – quantization
IQ – inverse quantization
VLC – variable length coding

Fine Grained Scalability

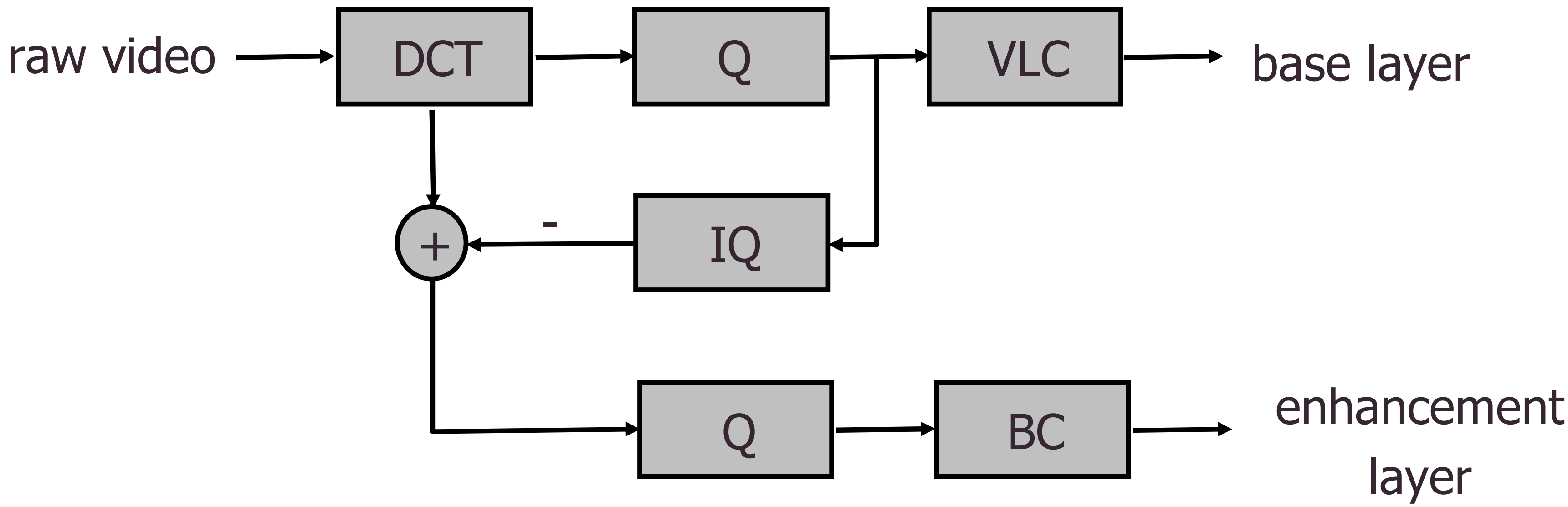
- Idea
  - Cut of compressed tail bits of samples
- Base layer
  - As in SNR coding
- Enhancement layer
  - Use bit-plane coding for enhancement layer instead of run-level coding
  - Cut tail bits off until data rate is reached

Fine Grained Scalability

```
MSB    (0,1)
MSB-1  (2,1)
MSB-2  (0,0)(1,0)(2,0)(1,0)(0,0)(2,1)
MSB-3  (5,0)(8,1)
…
```



Fine Grained Scalability



DCT – discrete cosine transformation

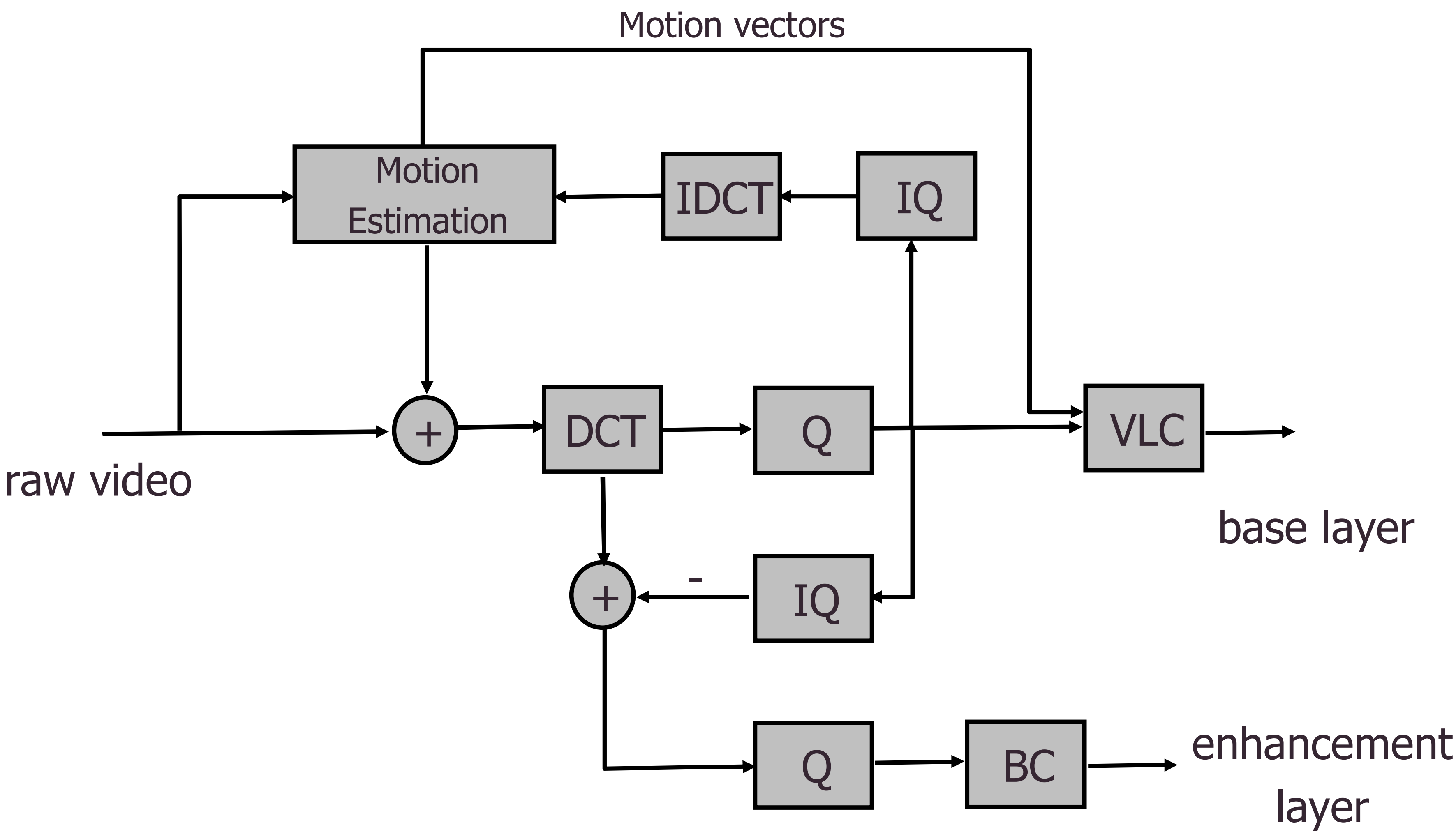
Q – quantization

IQ – inverse quantization

VLC – variable length coding

BC – bitplane coding

Fine Grained Scalability



Multiple Description Coding

- Idea
  - Encode data in two streams
  - Each stream has acceptable quality
  - Both streams combined have good quality
  - The redundancy between both streams is low

- Problem
  - The same relevant information must exist in both streams
  - Old problem: started for audio coding in telephony
  - Currently a hot topic

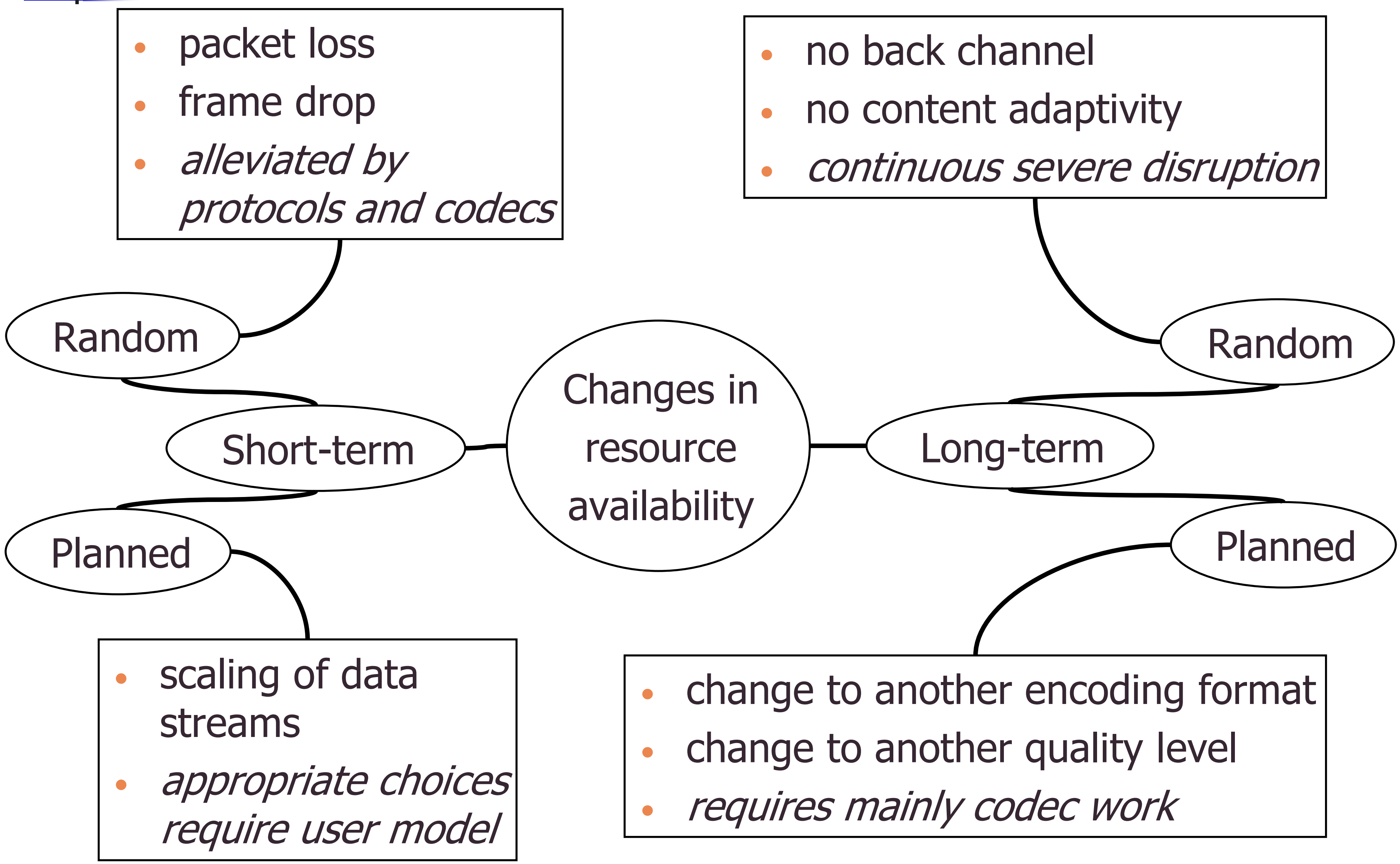
User Perception of Quality Changes

Quality Changes

- Quality of a single stream
  - Issue in Video-on-Demand, Music-on Demand, ...
  - Not quality of an entire *multi*media application

- Quality Changes
  - Usually due to changes in resource availability
    - overloaded server
    - congested network
    - overloaded client

Kinds of Quality Changes



Planned quality changes

- Video: Short-term changes
  - Use scalable encoding
  - Reduce short-term fluctuation by prefetching and buffering

- Scalable encoding
  - Non-hierarchical
    - encodings are more error-resilient
  - Hierarchial
    - encodings have better compression ratios

- Scalable encoding
  - Support for prefetching and buffering is an architecture issue
  - Choice of prefetched and buffered data is not

Planned quality changes

- Video: Short-term changes
  - Use scalable encoding
  - Reduce short-term fluctuation by prefetching and buffering

- Short-term fluctuations
  - Characterized by
    - frequent quality changes
    - small prefetching and buffering overhead
  - Supposed to be very disruptive

- See for yourself: subjective assessment

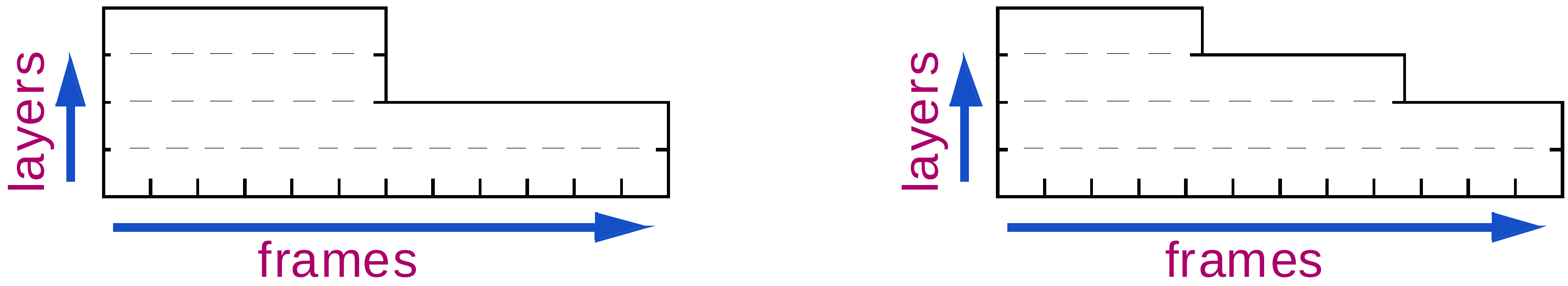
Subjective Assessment

- A test performed by the Multimedia Communications Group at TU Darmstadt
- Goal
  - Predict the most appropriate way to change quality
- Approach
  - Create artificial drop in layered video sequences
  - Show pairs of video sequences to testers
  - Ask which sequence is more acceptable
- Compare two means of prediction
  - Peak signal-to-noise ratio (higher is better)
    - compares degraded and original sequences per-frame
    - ignores order
  - Spectrum of layer changes (lower is better)
    - takes number of layer changes into account
    - ignores content and order

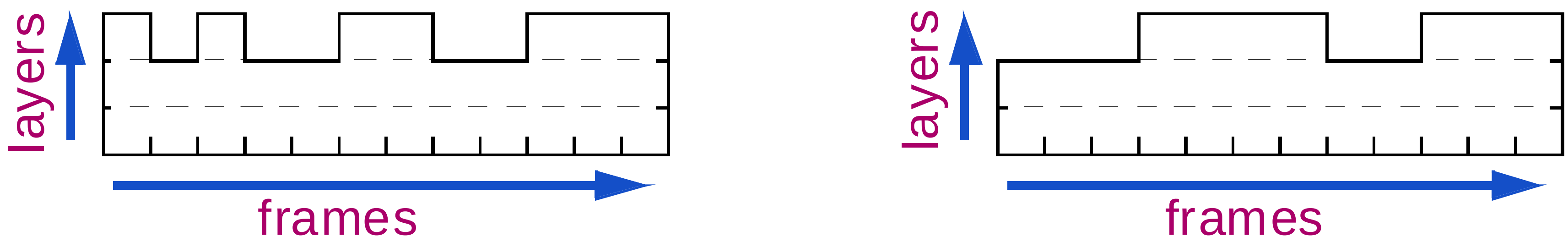
Subjective Assessment

Subjective Assessment

- Used SPEG (OGI) as layer encoded video format



amplitude of layer variation

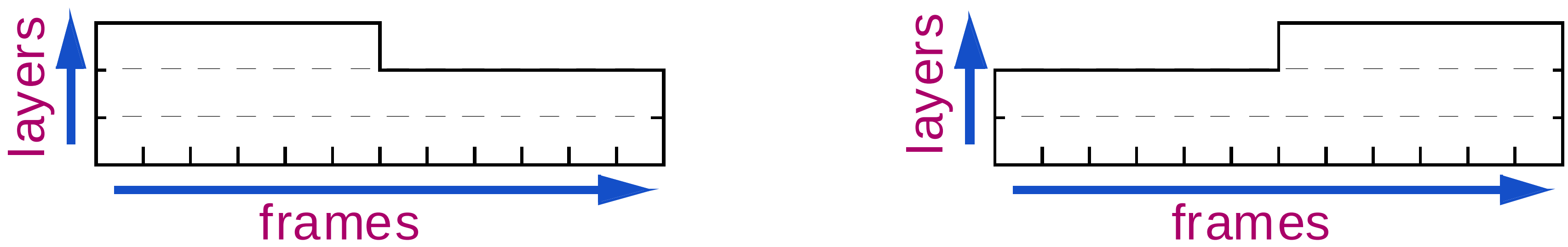


frequency of layer variation

Subjective Assessment

- What is better?

First gap first or lowest gap first?



Early or late high quality?

Subjective Assessment

- How does the spectrum correspond with the results of the subjective assessment?
- Comparison with the peak signal-to-noise ratio

#	Metric \ Clip	Farm 1		Farm 2		M&C1	
		ts 1	ts 2	ts 1	ts 2	ts 1	ts 2
1	Subjective assessment	0.35		0.55		0.73	
2	PSNR (higher is better)	62.86	49.47	61.46	73.28	63.15	52.38
3	Spectrum (lower is better)	2	2	6.86	4	2	1

#	Metric \ Clip	M&C3		M&C4		T-Tennis3	
		ts 1	ts 2	ts 1	ts 2	ts 1	ts 2
1	Subjective assessment	1.18		1.02		2.18	
2	PSNR (higher is better)	48.01	25.08	49.40	26.95	66.02	63.28
3	Spectrum (lower is better)	2	0	2	0	0.5	0.5

- According to the results of the subjective assessment the spectrum is a more suitable measure than the PSNR

Subjective Assessment

- Conclusions
  - Subjective assessment of variations in layer encoded videos
  - Comparison of spectrum measure vs. PSNR measure
    - Observing spectrum changes is easier to implement
    - Spectrum changes indicate user perception better than PSNR
    - Spectrum changes do not capture all situations
- Missing
  - Subjective assessment of longer sequences
  - Better heuristics
    - "thickness" of layers
    - order to quality changes
    - target layer of changes

User Model for Access Patterns

Short-term VoD model

Modeling for Video-on-Demand

- Video-on-demand systems
  - Objects are read-only
  - Hierarchical distribution system is the rule
  - Commercial VoD
    - Objects are generally consumed from start to end
    - Repeated consumption is rare
- Simulation approach
  - No real-world systems exist
  - Similar real-world situations can be adopted

Modeling

- User behaviour
  - The basis for simulation and emulation
    - In turn allows performance tests
  - Separation into
    - Frequency of using the VoD system
    - Selection of a movie

- User Interaction
  - Models exist
    - But are not verified

- Selection of a movie
  - Dominated by the access probability
  - Should be simulated by realistic access patterns

Model for Large User Populations

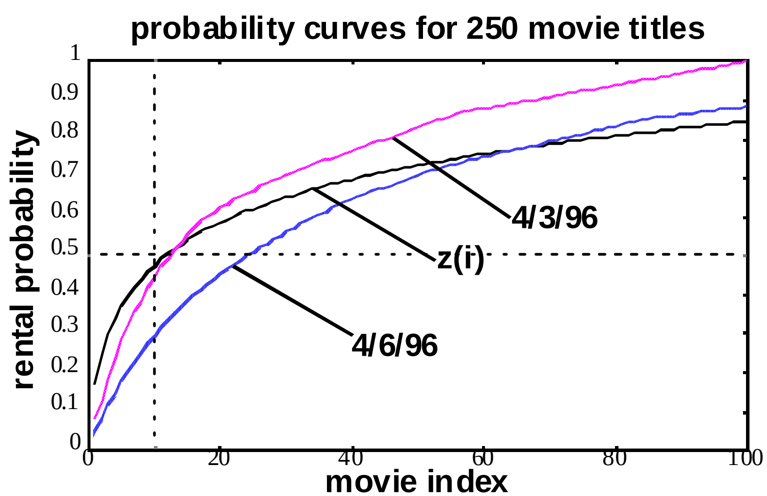
- Zipf Distribution

$$z(i) = \frac{C}{i^{\xi}}, C = 1 / \left(\sum_{j=1}^{N} 1 / j^{\xi} \right)$$

- Verified for VoD by A. Chervenak
  - N - overall number of movies
  - ξ – skew factor
  - i - movie i in a list ordered by descreasing popularities
  - z(i) - hit probability

Comparison with the Zipf Distribution

- Well-known and accepted model
- Easily computable
- Supports the earliest researchers' 90:10 rule-of-thumb



Comparison with two days from a movie rental shop

Problems of Zipf

- Does not work in hierarchical systems
  - Access to independent caches beyond first-level are not described
- Not easily extended to long-term model
  - Is timeless
  - Describes a snapshot situation
- Optimistic for the popularity of most popular titles

User Model for Access Patterns

Long-term VoD model

Long-Term Model

- Model should represent movie life cycles
  - To reflect the aging of titles
  - To observe movement of movies through a hierarchy of servers
  - To make observations with respect to a single movie
  - To support the idea of pre-distribution

- Model should work for large and small user populations
  - To allow variations in client numbers
  - To prevent from built-in smoothing effects

- Model can not be trace-driven
  - The number of movies is too small
  - The observation time is too short
  - The user population size is not variable
  - One title can not be re-used without similarity effects

Using Existing Models

- Use of existing access models ?
  - Some access models exist
  - Most are used to investigate single server or cluster behavior
  - Real-world data is necessary to verify existing models

- Optimistic model
  - Cache hit probabilities are over-estimated
  - Caches are under-dimensioned
  - Network traffic is higher than expected

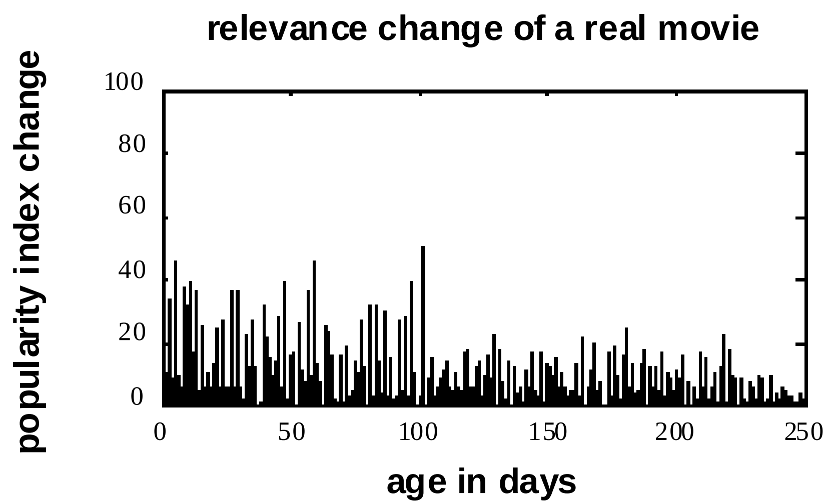
- Pessimistic model
  - Cache hit probabilities are under-estimated
  - Cache servers are too large or not used at all
  - Networks are overly large

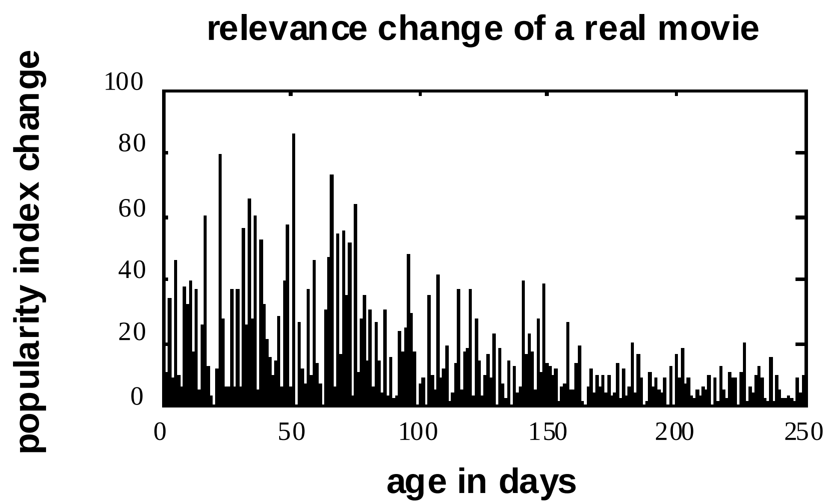
Approaches to Long-term Development

- Simple models for long-term studies
  - Static approach
    - No long-term changes
    - Movie are assumed to be distributed in off-peak hours
  - CD sales model
    - Smooth curve with a single peak
    - Models the increase and decrease in popularity
  - Shifted Zipf distribution
    - Zipf distribution models the daily distribution
    - Shift simulates daily shift of popularities
  - Permutated Zipf distribution
    - Zipf distribution models the daily distribution
    - Permutation simulates daily shift of popularities

Verification: Zipf Variations

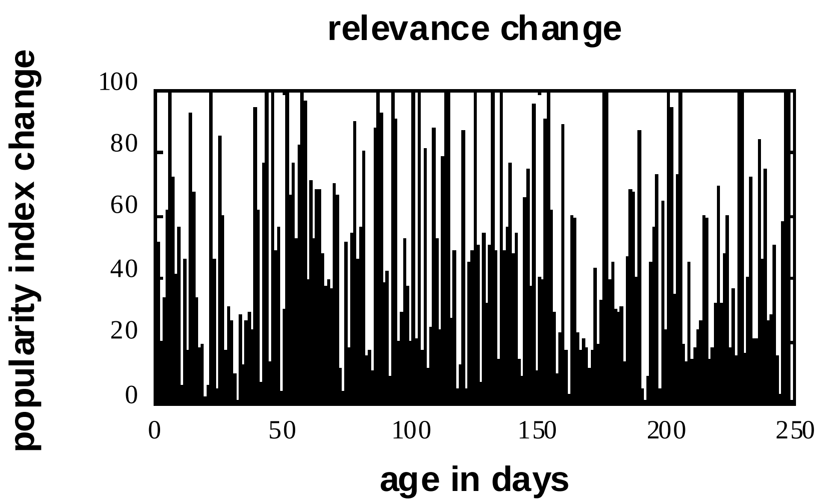
- Rotation model for day-to-day relevance changes

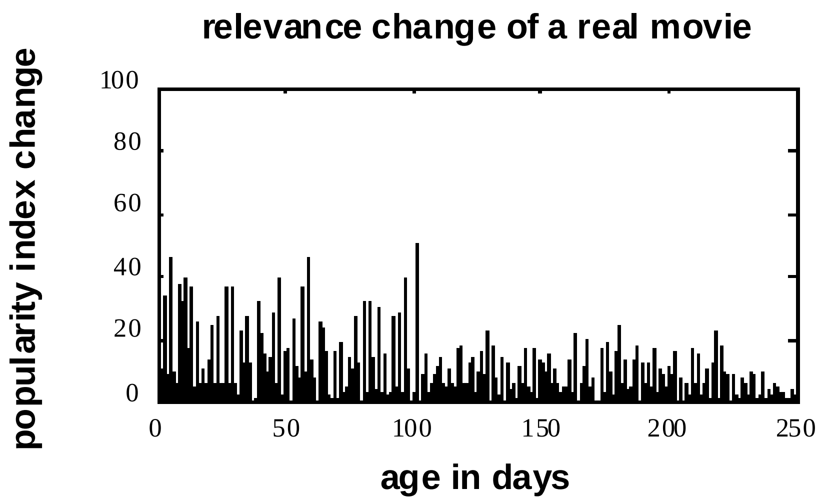


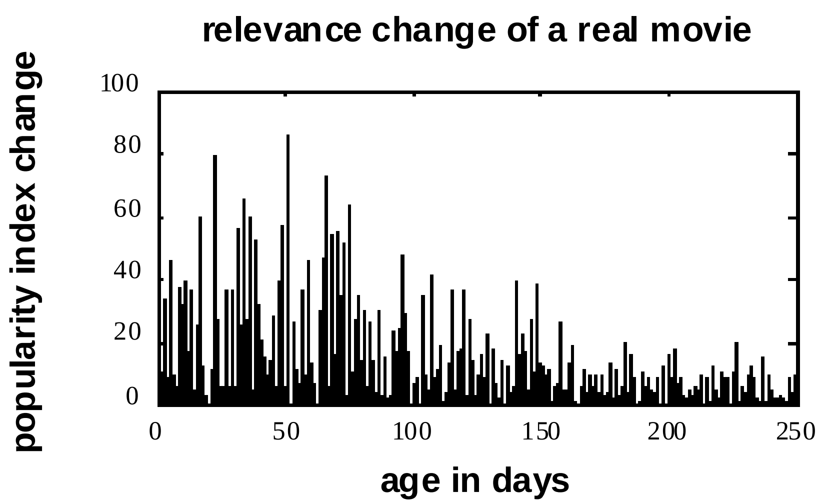


Verification: Zipf Variations

- Permutation model for day-to-day relevance changes







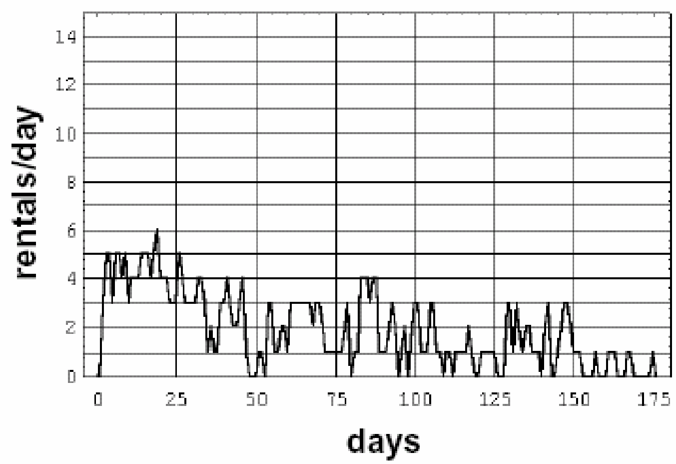
Existing Data Sources for Video-on-Demand

- Movie magazines
  - Data about average user behaviour
  - Represents large user populations
  - Small number of observation points (weekly)
- Movie rental shops
  - Actual rental operations
  - Serves only a small user population
  - Initial peaks may be clipped
- Cinemas
  - Actual viewing operations
  - Serves only a small user population
  - Few number of titles
  - Short observation periods

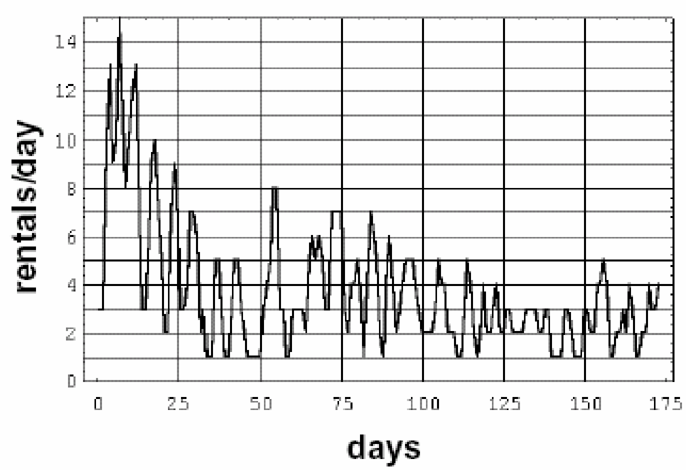
Verification: Small and Large User Populations

Rental shop

Highlander 3

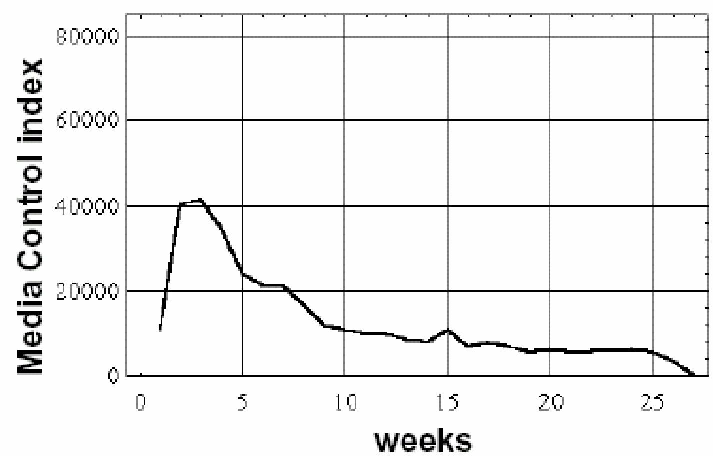


The Lion King

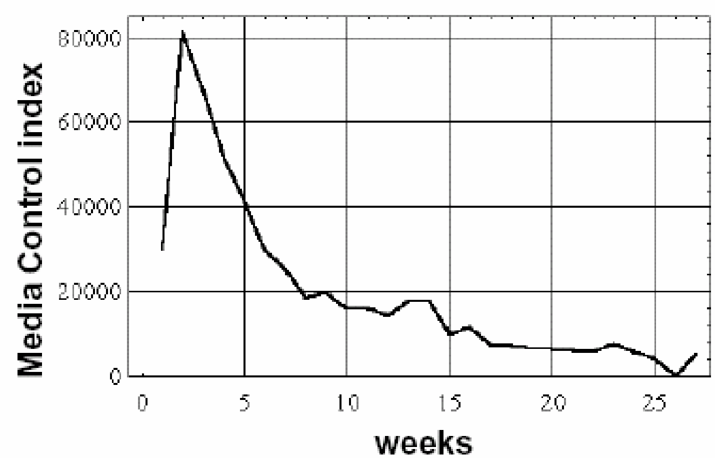


Movie magazine

Highlander 3



The Lion King



Verification: Small and Large User Populations

- Similarities
  - Small populations follow the general trends
  - Computing averages makes the trends better visible
  - Time-scale of popularity changes is identical
  - No decrease to a zero average popularity

- Differences
  - Large differences in total numbers
  - Large day-to-day fluctuations in the small populations

- Typical assumptions
  - 90:10 rule
  - Zipf distribution models real hit probability

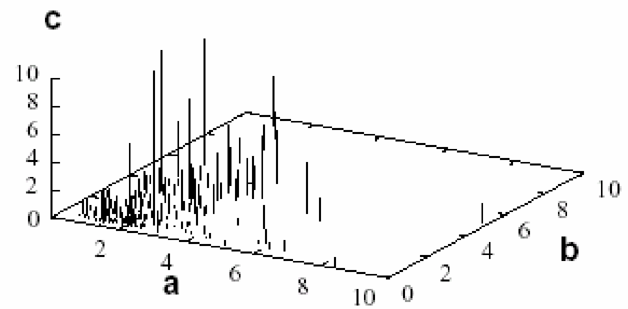
New Model: Movie Life Cycle

Product life cycle function

$$RP(t) = a \times e^{\left(\frac{2}{\sqrt{10}} - \frac{t}{10 \times b} - \frac{b}{t}\right)} + c$$

- a - peek relevance
- b - steepness of decline
- c - remaining popularity
- RP(t) - popularity at time t

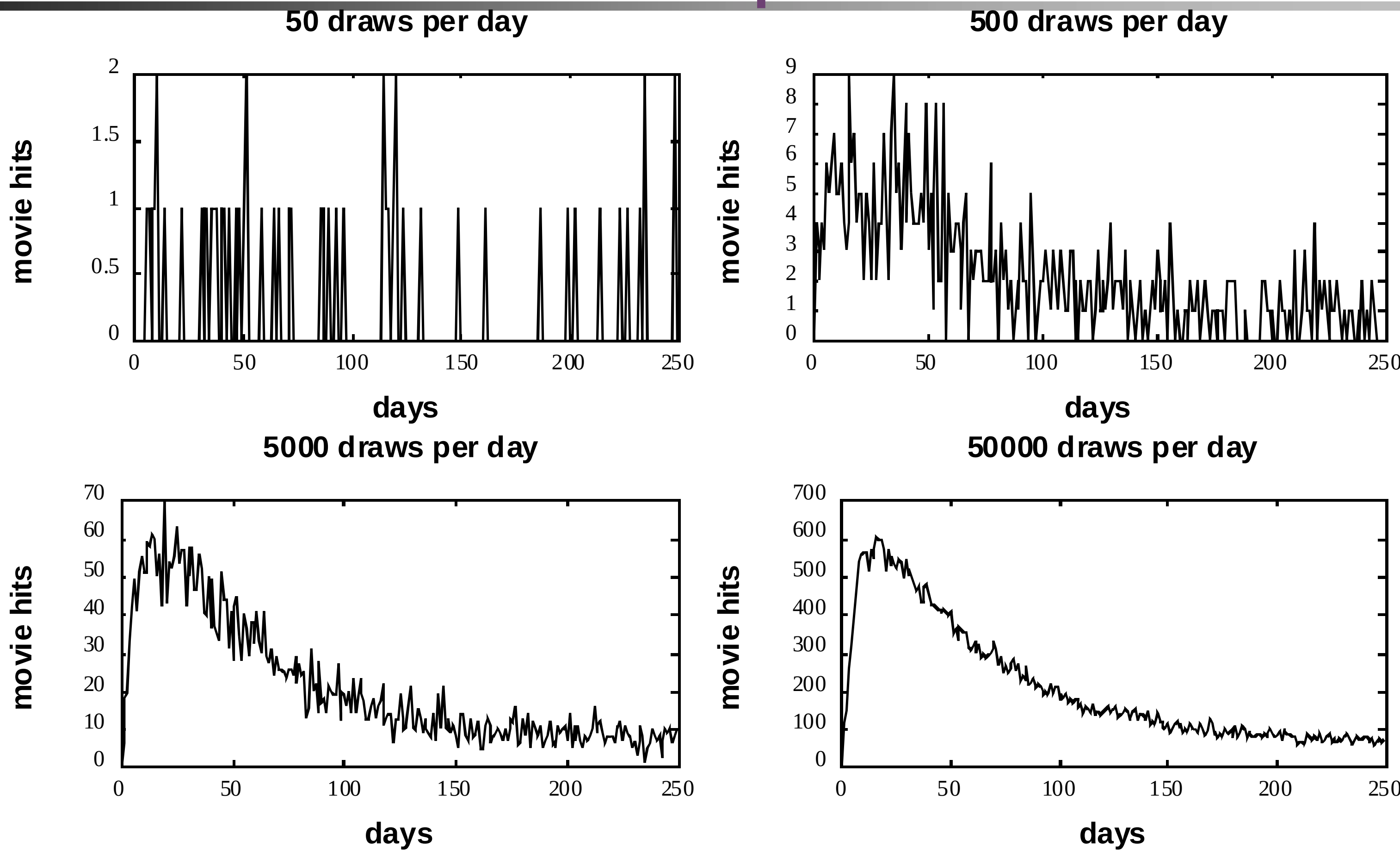
fitted parameters



- Least square fitting to real data converged for all movie older than 10 days
- Small observed coefficient of correlation between parameters (< 0.1)

- Characteristics
  - Quick popularity increase
  - Various top popularities
  - Various speeds in popularity decrease
  - Various residual popularity

New Model: User Population Size



- Smoothing effect of larger user populations
- Day-to-day relevance changes
- Probability distribution of all movies by „new releases“

Problems with Data Sources

- Lack of additional real-world data
  - No verification data for medium-sized populations available
- Missing details
  - Genres
    - Popularity rise and decline depends on genres
    - Single users' behaviour can be predicted
  - Single day probability variations
    - Children's choices at daytime, adults´ choices at night
  - Regional popularity differences
    - Ethnic groups
    - Regional information
  - Comebacks
    - Sequels inspire comebacks
- Detail overload
  - Simplifications are required for large simulations

Video Access Modeling

- Simple Zipf models are not suited for simulation of server hierarchies
- Trace-driven simulation can not be used
- Our model is sufficient for general investigation on caching
  - Long-term movie life cycles can be modeled nicely
  - Optimistic assumptions due to smoothness are removed
  - Variations in movie behavior are supported
  - Day-to-day popularity changes are realistic
- It is not sufficient yet for advanced caching mechanisms
  - Single-day variations are missing
  - Genres are missing

User Model for Access Patterns

Interactive VoD model

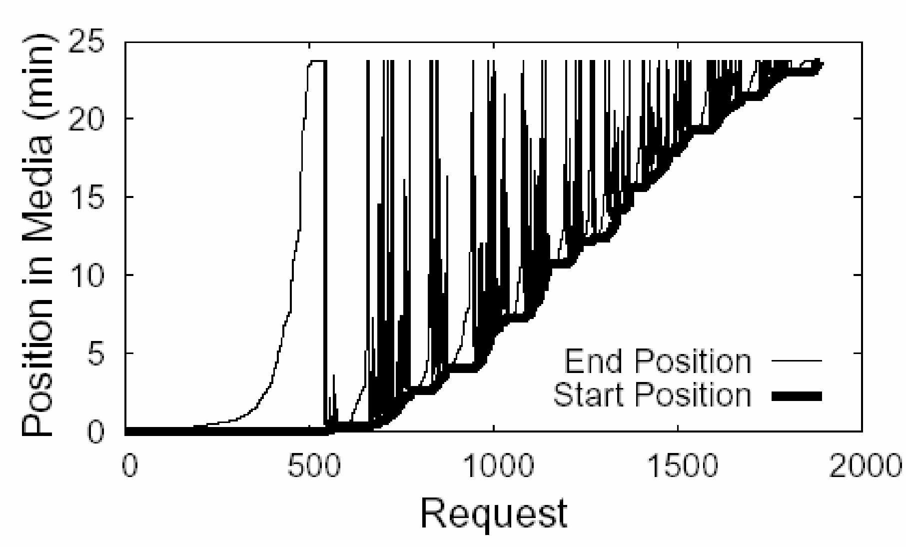
Interactive VoD Modeling

- Non-interactive models
  - Allow resource planning in network and server
  - Are realistic for watching movies

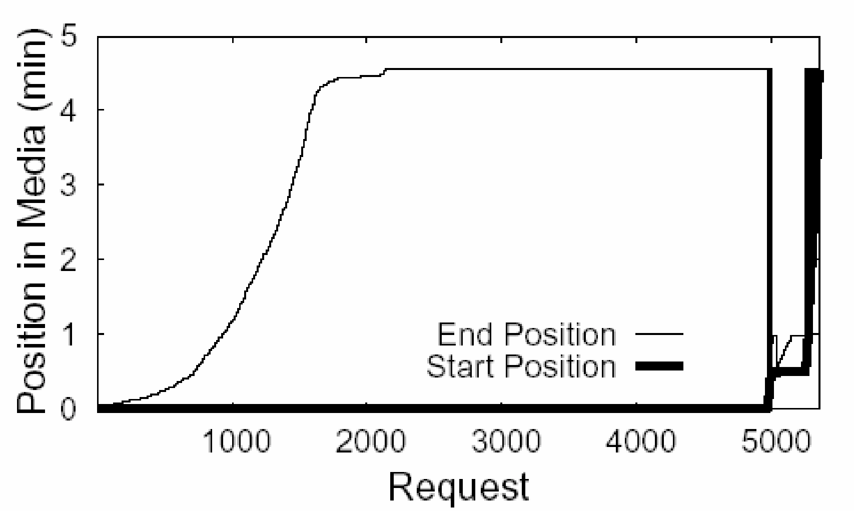
- Higher interactivity in
  - Editing applications
    - Cutting
  - Browsing applications
    - Shopping
    - Web surfing
  - Interactive applications
    - Embedded in virtual reality
    - E-learning

Interactive Models

- High interactivity
  - Typical for long e-learning movies
  - >=3 requests per session
  - Duration <20% of media length
  - Average start position between 30% and 60% of media length
  - <30% begin at the start

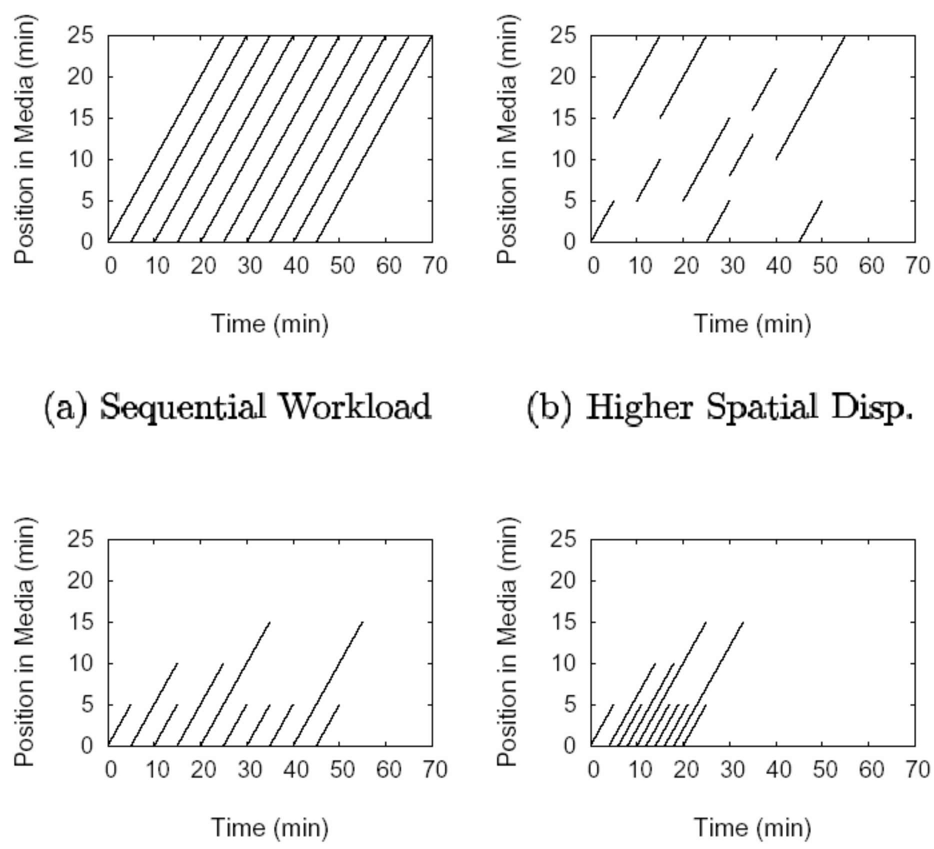


- Low interactivity
  - Typical for short clips
  - Duration >>20% of media length
  - Most begin at the start
  - 1 or 2 request per session



Interactive Models

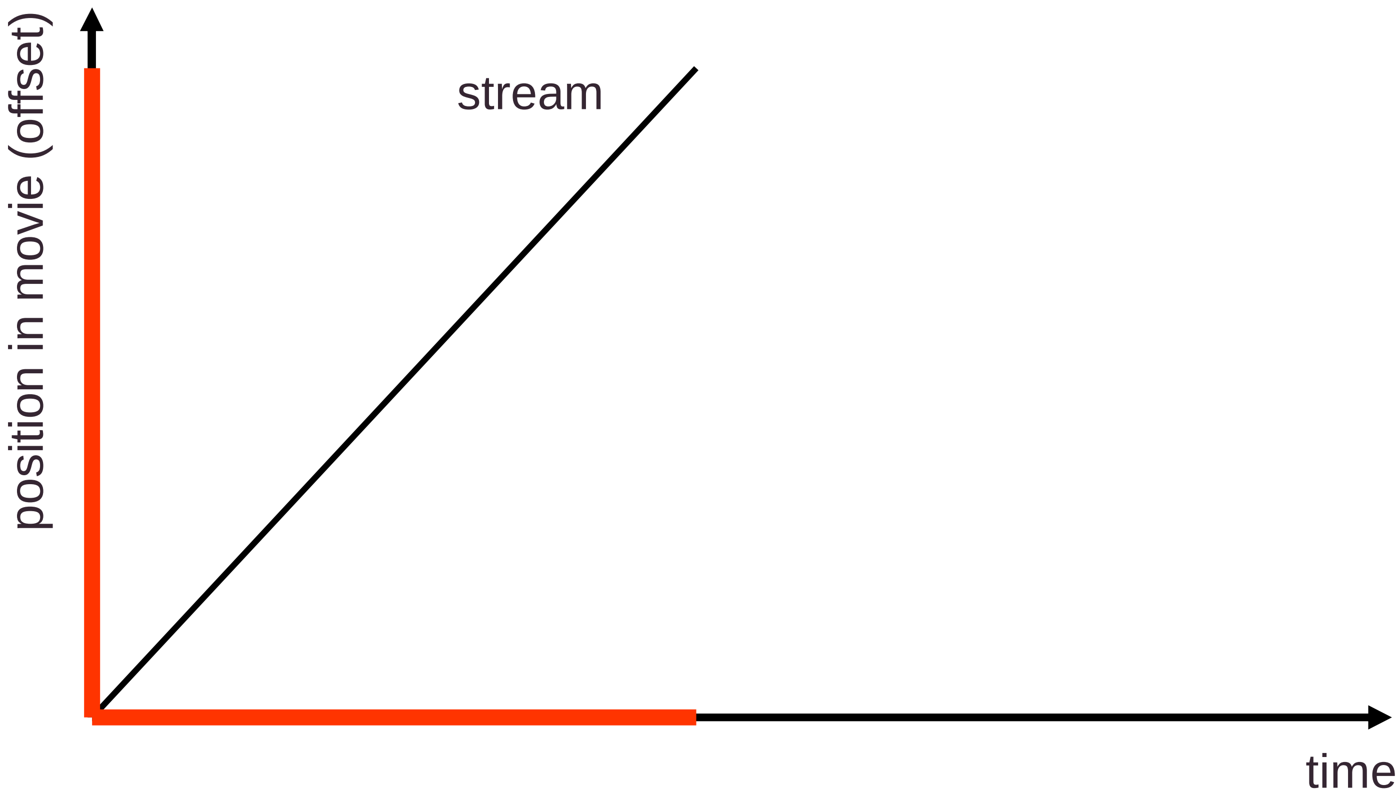
- Rocha et al.'s interactive model
  - Simulation based on
    - Temporal dispersion
    - Spatial dispersion
- Spatial dispersion
  - Higher when requests have more data in common
- Temporal dispersion
  - Lower when number of interactive requests is higher
- These two variables do not define behaviour completely: Where do requests start?
- Application-dependent choice
  - highly interactivity
  - low interactivity



(a) Sequential Workload (b) Higher Spatial Disp.
(c) Lower Spatial Disp. (d) Low Temporal Disp.

[Rocha et al. 2005]

Graphics Explained



- Y - the current position in the movie
  - the temporal position of data within the movie that is leaving the server
- X - the current actual time

User Model for Synchrony

Synchronization

- Temporal Relations
  - Intra-object Synchronization
    - Intra-object synchronization defines the time relation between various presentation units of one time-dependent media object
  - Inter-object Synchronization
    - Inter-object synchronization defines the synchronization between media objects
  - Skew
    - Deviation between intended and actual time relation

- Relevance of inter-object synchronization
  - Hardly relevant in NVoD systems: only intra-object sync. required
  - Somewhat relevant in conferencing systems
  - Very relevant in games
  - Relevant in multi-object formats: MPEG-4, Quicktime

- Inter-object synchronization example: Lip synchronization
  - Tight coupling of audio and video streams
  - Limited skew acceptable
  - Main problem of the user model: permissible skew

Synchronization Requirements – Fundamentals

- 100% accuracy is not required, i.e., skew is allowed

- Skew depends on
  - Media
  - Applications

- Difference between
  - Detection of skew
  - Annoyance of skew

- Explicit knowledge of skew
  - Alleviates implementation
  - Allows for portability

Experimental Set-Up

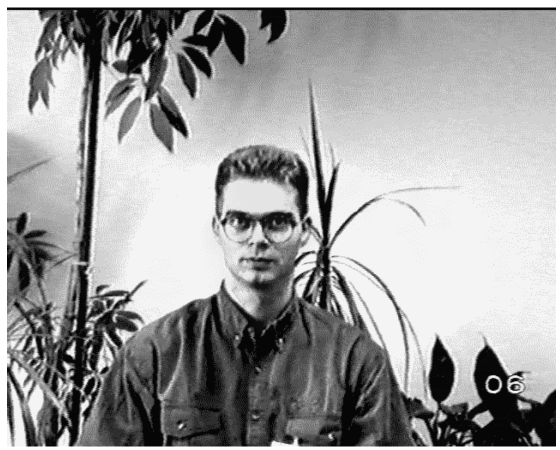
- Experiments at IBM ENC Heidelberg to quantify synchronization requirements
  - Audio/video synchronization, audio/pointer synchronization

- Selection of material
  - Duration
    - 30s in experiments
    - 5s would have been sufficient
  - Reuse of same material for all tests
  - Introduction of artificial skew

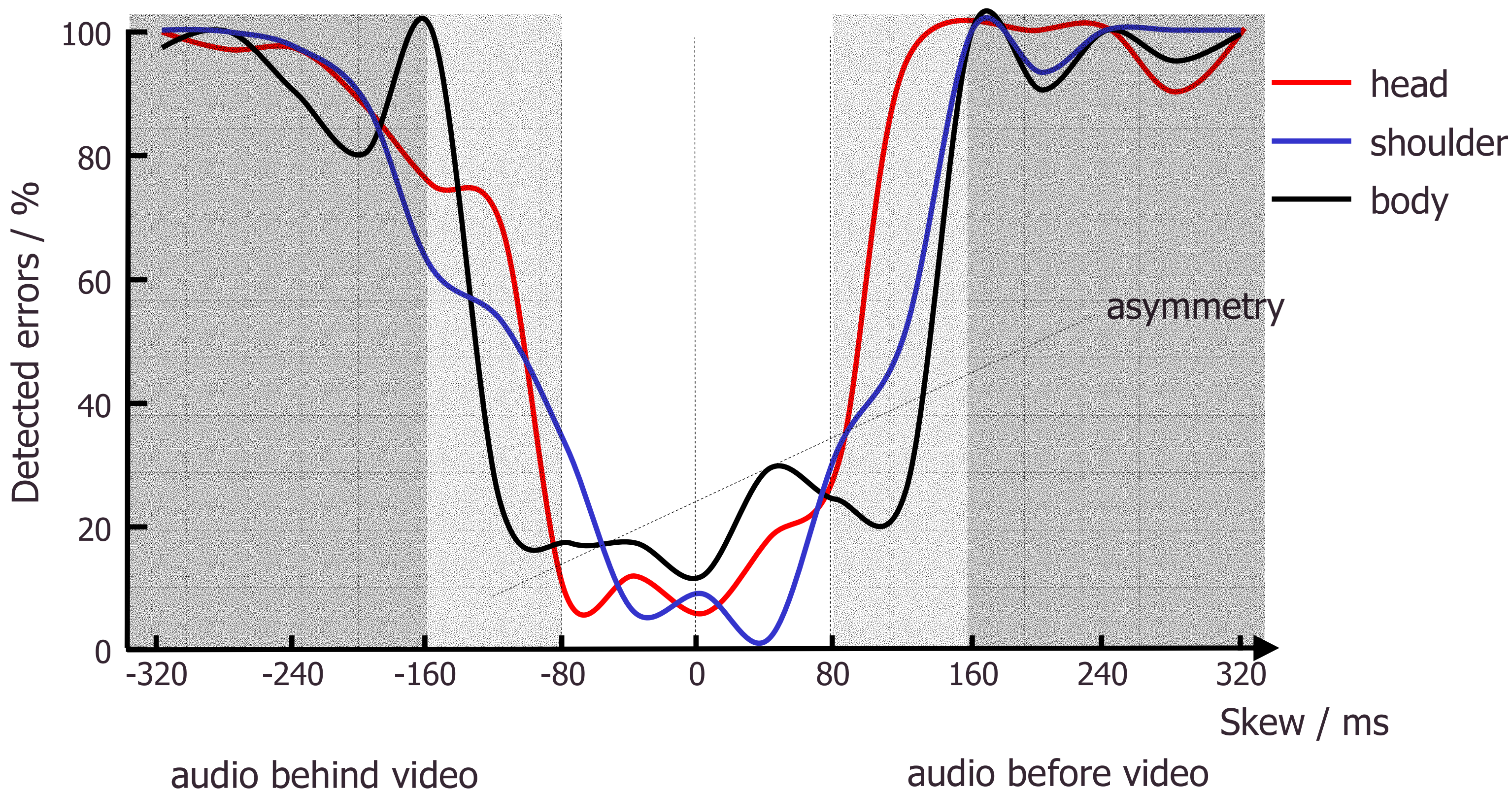
- Experiments
  - Large set of test candidates
    - Professional: cutter at TV studios
    - Casual: every day "user"
  - Awareness of the synchronization issues
  - Set of tests with different skews lasted 45 min

Lip Synchronization: Major Influencing Factors

- Video
  - Content
    - Talking head
    - Still background
  - View mode
    - head view
    - shoulder view
    - body view

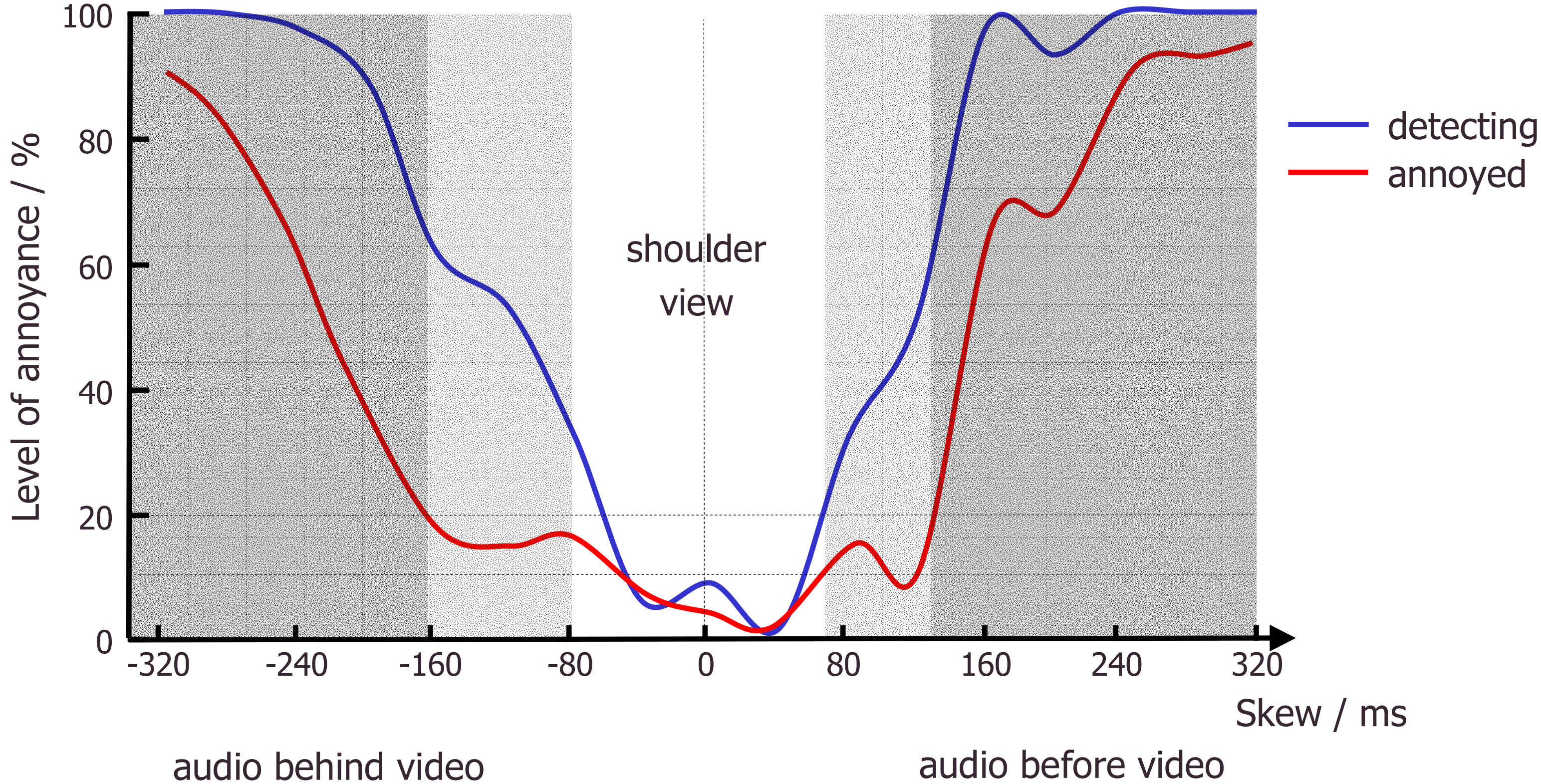


Lip Synchronization: Level of Detection



- Areas
  - In sync QoS: +/- 80 ms
  - Transient
  - Out of sync

Lip Synchronization: Level of Annoyance



- Some observations
  - Asymmetry
  - Additional tests with long movie
    - +/- 80 ms: no distraction
    - -240 ms, +160 ms: disturbing

Quality of Service of Two Related Media Objects

Media		Mode, application	QoS
Video	Animation	Correlated	+/- 120 ms
	Audio	Lip synchronization	+/- 80 ms
	Images	Overlay	+/- 240 ms
		No overlay	+/- 500 ms
	Text	Overlay	+/- 240 ms
		No overlay	+/- 500 ms

Quality of Service of Two Related Media Objects

Media		Mode, application	QoS
Audio	Animation	Event correlation	+/- 80 ms
	Audio	Tightly coupled (stereo)	+/- 11 µs
		Loosely coupled (dialog mode with various participants)	+/- 120 ms
		Loosely coupled (background music)	+/- 500 ms
	Image	Tightly coupled (music with notes)	+/- 5 ms
		Loosely coupled (slide show)	+/- 500 ms
	Text	Text annotation	+/- 240 ms
	Pointer	Audio related to shown item	-500 - +750 ms

Summary

Summary

- Storage and distribution system must support
  - *Discrete* media such as text and graphics
  - *Continuous* media such as audio and video
  - Interrelated *Multiplexed* media

- *Encoding Format* and *File Format* must be distinguished
  - Separation of file format and wire format
  - Streamable files vs. streaming format

- Trend towards
  - Formats that *define presentation* environments
  - *Interaction* of encoding format and application
  - *Interaction* of client and server

- *Influence on Distribution Systems?*

Summary

- User modeling helps achieving a good price/performance ratio for multimedia systems

- User modeling allows cheating

- Examples seen
  - Modeling quality assessment of layered video
  - Modeling audio/video synchronization
  - Modeling video access probability

References

- Ralf Steinmetz, Klara Nahrstedt: *Multimedia Fundamentals, Volume I: Media Coding and Content Processing* (2nd Edition), Prentice Hall, 2002, ISBN 0130313998
- Touradj Ebrahimi (Ed.), Fernando Pereira, *The MPEG-4 Book*, Prentice Hall, 2002, ISBN 0130616214
- Weiping Li, *Overview of Fine Granularity Scalability in MPEG-4 Video Standard*, IEEE Transactions on Circuits and Systems for Video Technology, 11(3), Mar. 2001
- Vivek K. Goyal, *Multiple Description Coding: Compression Meets the Network*, IEEE Signal Processing Magazine, Sep. 2001
- Ann Chervenak: *Tertiary Storage: An Evaluation of New Applications*, PhD thesis, University of California, Berkeley, 1994
- Carsten Griwodz, Michael Bär, Lars Wolf: *Long-Movie Popularity Models in Video-on-Demand Systems*, ACM Multimedia, Seattle, WA, USA, Nov. 1997
- Charles Krasic, Jonathan Walpole: *Priority-Progress Streaming for Quality-Adaptive Multimedia*, ACM Multimedia Doctoral Symposium, Ottawa, Canada, Oct. 2001

- Ralf Steinmetz, Klara Nahrstedt: *Multimedia Fundamentals, Volume I: Media Coding and Content Processing* (2nd Edition), Prentice Hall, 2002, ISBN 0130313998

- Michael Zink, Oliver Künzel, Jens Schmitt, Ralf Steinmetz: *Subjective Impression of Variations in Layer-Encoded Videos*, IWQoS, Monterey, CA, USA, Jun. 2003
- Michael Zink, Jens Schmitt, and Carsten Griwodz. Layer-Encoded Video Streaming: A Proxy's Perspective. In IEEE Communications Magazine, Vol. 42, No. 8, August 2004