Recent Advances in Video Transmission: Challenges and Opportunities

Aggelos K. Katsaggelos

Professor
Northwestern University
Department of EECS
Evanston, IL 60208
www.eecs.northwestern.edu/~aggk

Plethora of exciting applications!

- Freedom of Multimedia Access (FOMA)
- Exciting Applications
- TiVo2Go
- Sales of phone digital cameras surpassed sales of digital cameras in 2001
- Seamless mobility, Ubiquitous access, UMA/UME

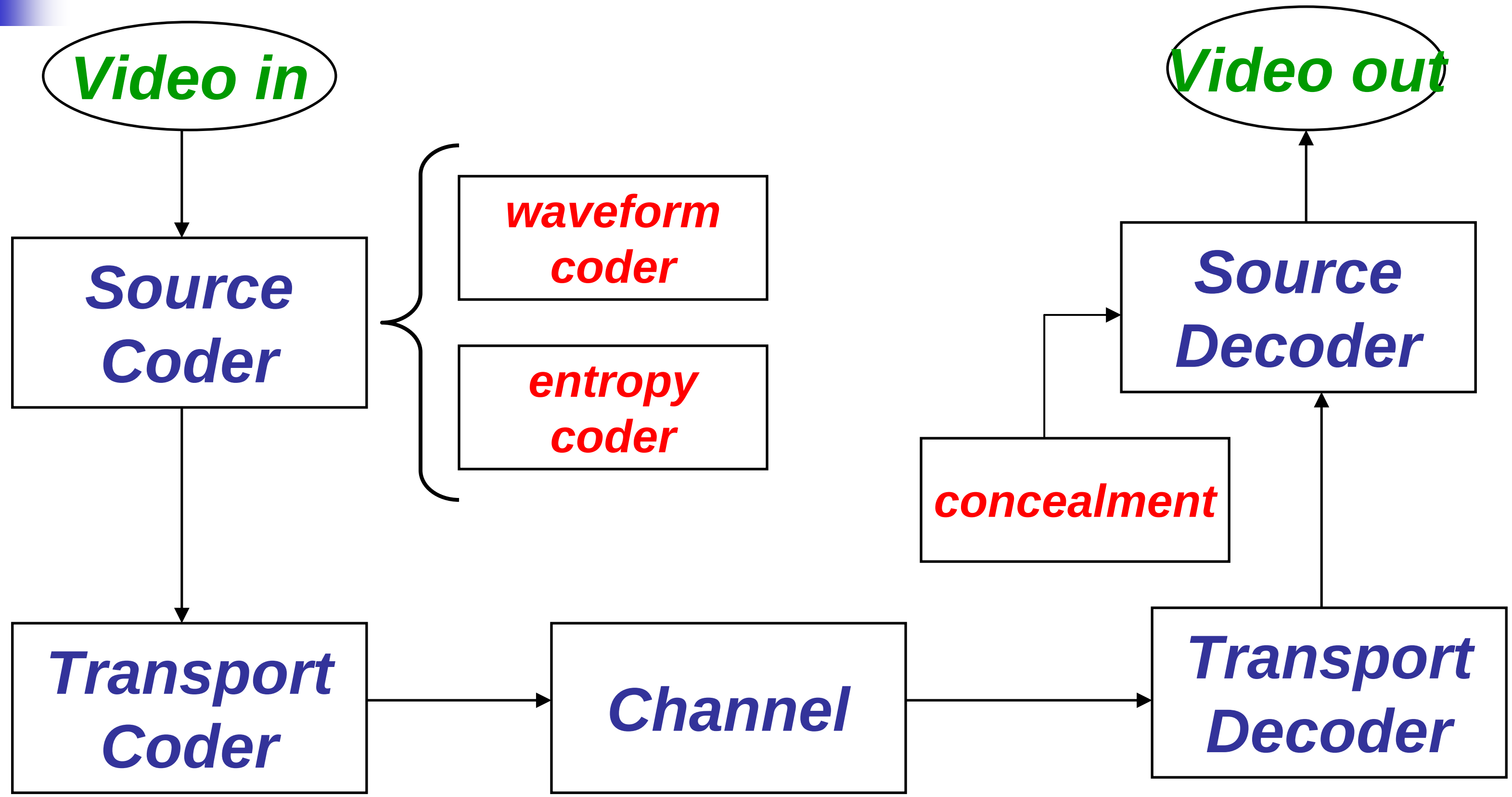
Video is the protagonist!

The Five+ Sense Communication Interface	Contribution to Human Perception
Visual	60%
Auditory	20%
Tactile	15%
Taste	3%
Smell	2%
Existance/Emotion	?%

Talk Objectives

- Describe recent advances in video transmission
- Towards this goal, describe in some detail the building blocks of a video transmission system, and provide specific examples
- Describe “where are things going”, challenges and opportunities

Video Communications System



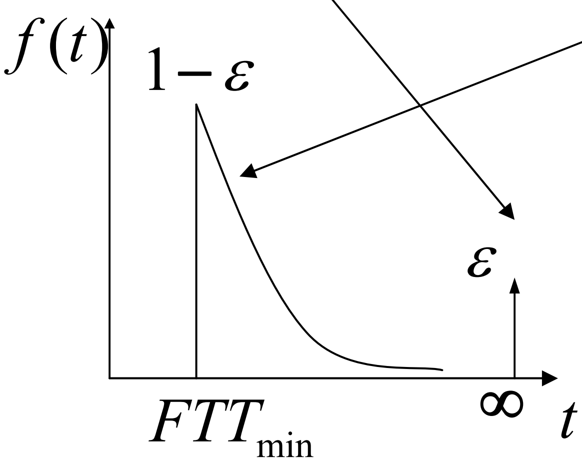
Channel coding
packetization
modulation

Channel Models

- Network model :

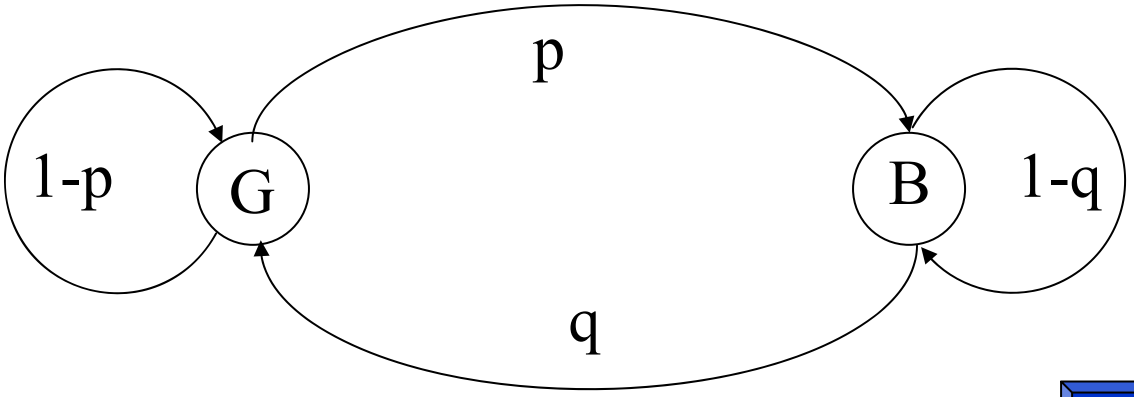
  an independent time-invariant packet erasure channel + random delays

$$\rho^k = \underbrace{\varepsilon} + (1-\varepsilon)\underbrace{P\{\Delta T_n(k) > \tau\}}$$



- Packet delay
  - Exponential distribution: fast decaying tail
  - Gamma distribution
  - Pareto distribution: slowly decaying (heavy) tail
- Packet Loss
  - Bernoulli
  - 2-state Markov (Gilbert)
  - High-order Markov

Gilbert Model



G: Good State, B: Bad State

While in B, error prob=1-h

burst length=1/q*1000/bitrate msec

Overall BER=(1-h)*p/(p+q)

Gilbert Model

BER = $5x10^{-3}$

Burst Length = 2.5ms

Bad BER h = 0.5

Error Resiliency Challenges

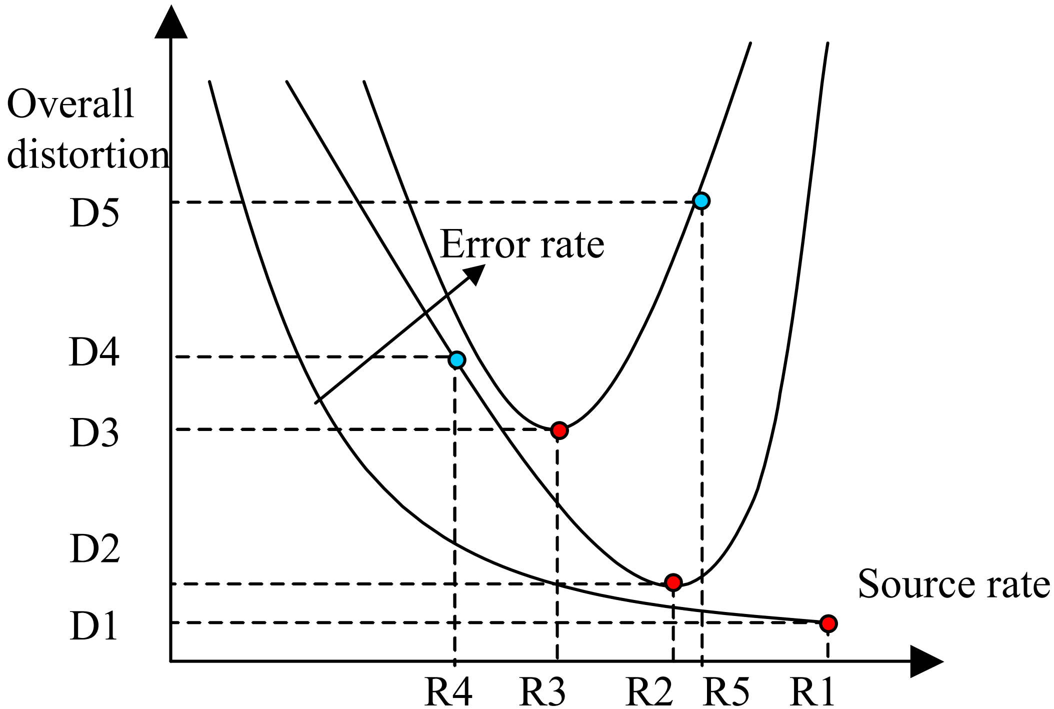
- Non-Robust Nature of (VLBR) Video Coding
  - Highly Predictive
  - Variable Length Codes (requires resynchronization)
- Broad Error Conditions
  - Random Bit Errors
  - Burst Errors
  - Packet Loss Errors
- Low Delay (and additional constraints on resources)
  - Interleaving can be a problem

Considerations

Shannon’s Separation Theorem

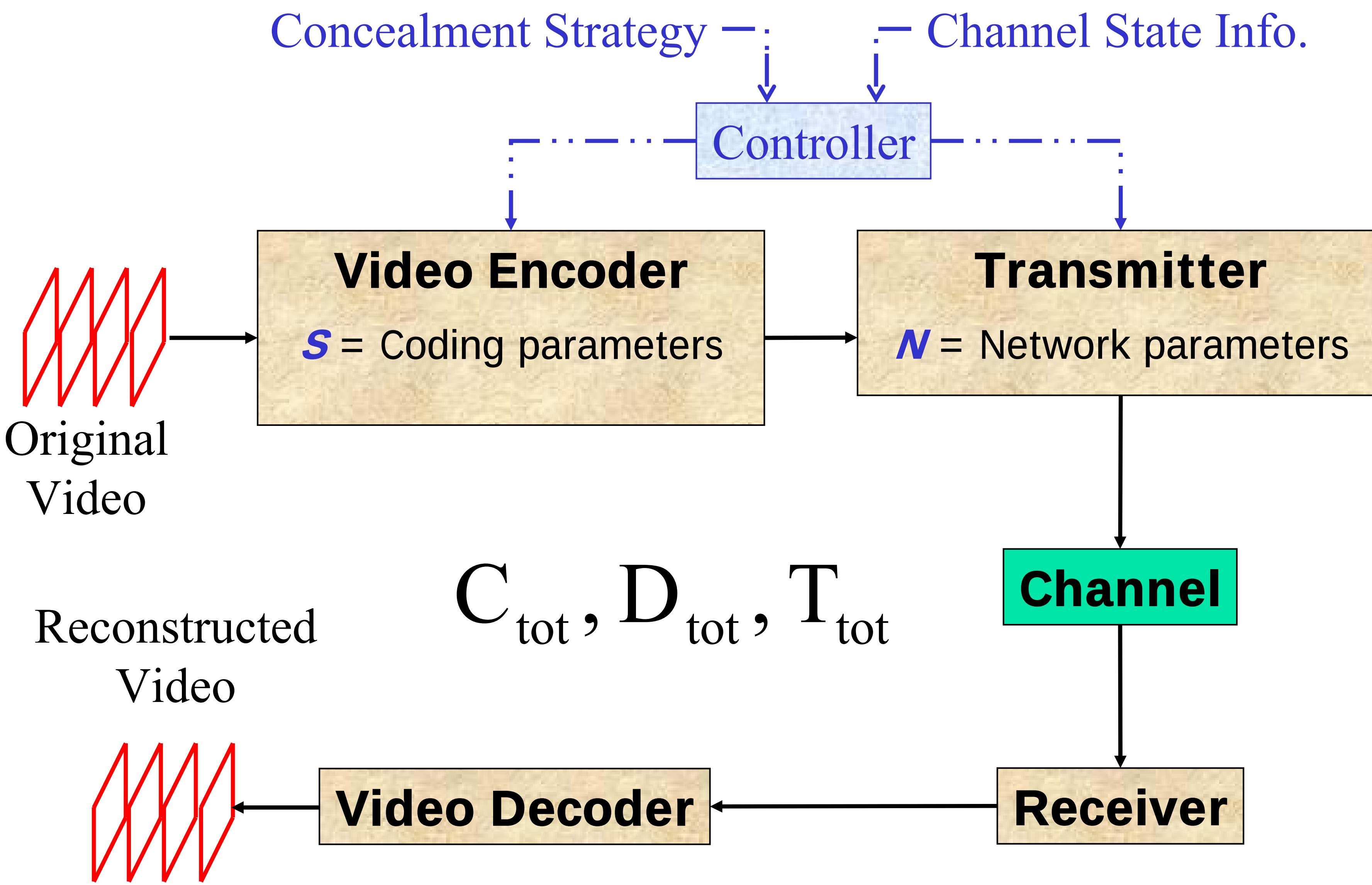
- Joint-Cross Channel Coding
- Network Adaptive Video Transmission
- Dynamic Resource Allocation
- Cross Layer Optimization
- Resource Allocation Based on “Importance of Content”
  - Protection and Energy are given to regions that are difficult to conceal
  - Regions of Interest / Object oriented coding

Joint Source-Channel Coding



The total bandwidth for source and channel rates is the same for the three curves

System Model



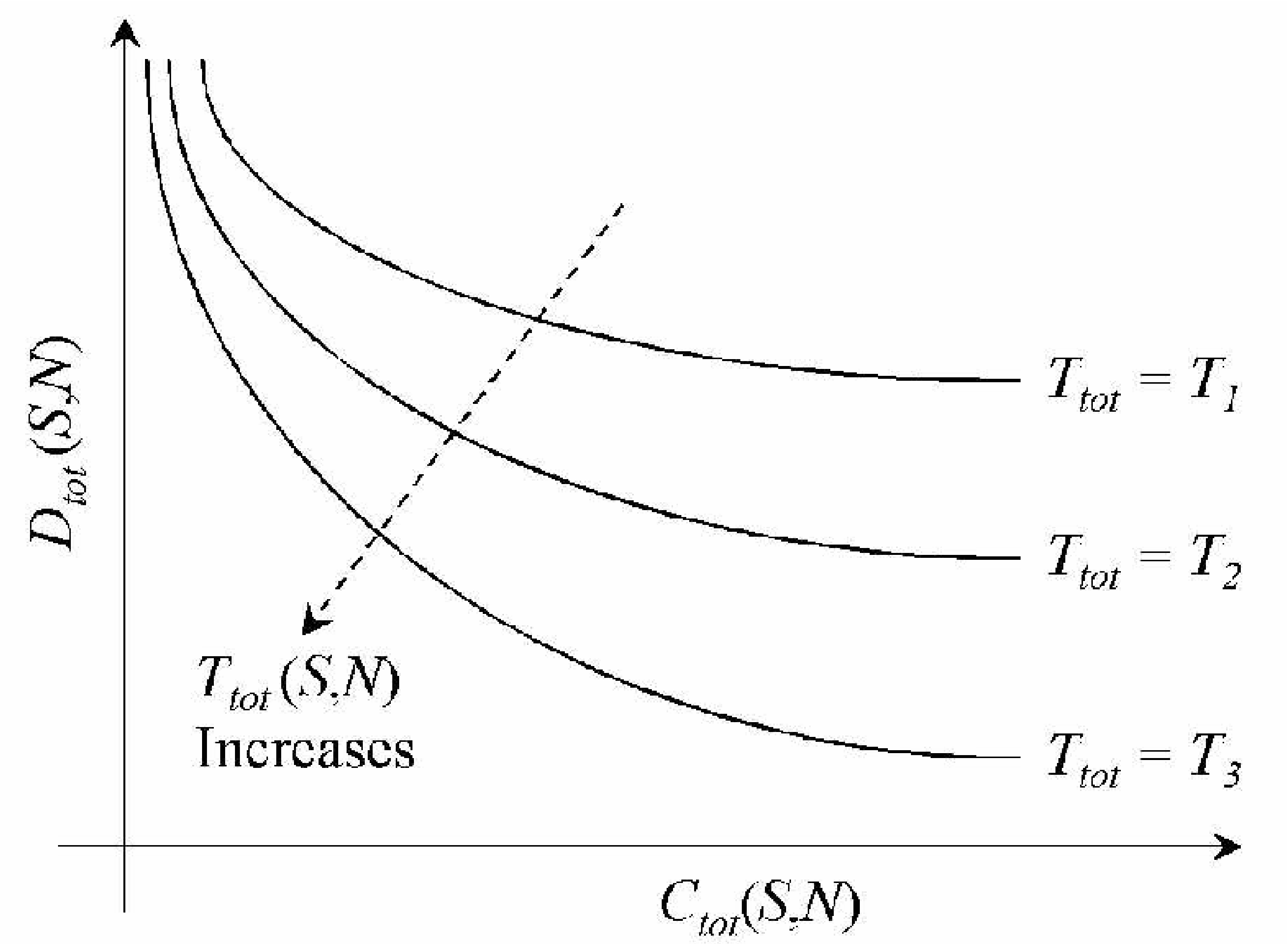
$$C_{tot}, D_{tot}, T_{tot}$$

Examples of cost/constraints

- Cost
  - Transmission (computation) power
  - $$ for DiffServ
- Network Parameters
  - Scheduling
  - Transmission rate
  - Probability of packet loss
  - FEC
  - ARQ (fast, hybrid)

A. K. Katsaggelos et al, “Advances in Efficient Resource Allocation for Packet-Based Real-Time Video Transmission”, *IEEE Proc., Jan.* 2005.

TRADE-OFFS



Resource-Distortion Optimization Framework

- Goal: Minimize transmission cost while limiting the end-to-end distortion and delay.

$$\min \quad \mathrm{C}_{\mathrm{tot}}(S, N)$$ Transmission Cost

$$s.t.:$$

$$\mathrm{D}_{\mathrm{tot}}(S, N) \leq \mathrm{D}_0$$ End-to-End Distortion Constraint

$$\mathrm{T}_{\mathrm{tot}}(S, N) \leq \mathrm{T}_0$$ Transmission Delay Constraint

Expected Distortion

Depends on coding parameters for the current packet

Depends on concealment scheme (μ and ρ for other packets)

$$E\left[D^k\right] = \left(1-\rho^k\right) E\left[D_R^k(\mu^k)\right] + \left(\rho^k\right) E\left[D_L^k\right]$$

- What Affects the Expected Distortion?
  1. Source coding
  2. Probability of loss in the channel
  3. Error concealment

End-to-End Frame Distortion

Expected Distortion for the kth packet

Average Expected Distortion →

$$D_{tot} = \frac{1}{K}\sum_{k=1}^{K} \underbrace{E\left[D^{k}\right]}$$

Maximum Expected Distortion →

$$D_{tot} = \max_{k=[1,\ldots,K]}\left\{ E\left[D^{k}\right]\right\}$$

Vector of Expected Distortions →

$$D_{tot} = \left[E\left[D^{1}\right], E\left[D^{2}\right], \ldots, E\left[D^{K}\right]\right]^{T}$$

Transmission Cost

- Wireless = Transmission energy

Source Coding Parameters

Transmission Power

Total Energy

$$E_{tot} = \sum_{k=1}^{K} E^k = \sum_{k=1}^{K} B^k(\mu^k) \frac{P^k}{R^k}$$

Packet index

Number of Bits

Transmission Rate

Cost per bit

- DiffServ = Pricing

Total Cost

$$C_{tot} = \sum_{k=1}^{K} C^k = \sum_{k=1}^{K} B^k(\mu^k) c^k$$

Power vs. Probability of Loss

Transmission power for the kth packet → $P^k = g\left(\rho^k\right)$ ← Probability of loss for the kth packet

- Empirical measurements or analytical models can be used by the transmitter to obtain the function $g(\bullet)$
- Example: (outage probability)
  - Narrowband slowly fading channel with AWGN and i.i.d channel fading per packet (L. Ozarow et al, *VT* '94)

$$P^k = g\left(\rho^k\right) = \frac{-C}{\ln(1-\rho^k)} \quad \text{where} \quad C = \frac{N_0 W}{E[H]}\left(2^{R/W} - 1\right)$$

R = channel rate; $E[H]$ = expected channel fade; $N_0 W$ = noise power; W = bandwidth

Minimum Energy Approach

- Goal: Use the minimum energy to achieve an acceptable level of quality and delay.

$$\underset{\{\mu^k, P^k\}}{\text{minimize}} \qquad \mathrm{E}_{\mathrm{tot}} = \sum_{k=1}^{K} \frac{B^k(\mu^k)}{R} P^k$$

Total Transmission Energy for frame

$$s.t.:$$

$$E\left[D^k\right] = \begin{cases} D_0 & \forall k: E\left[D_R^k(\mu^k)\right] \le D_0 \le E\left[D_L^k\right] \\ E\left[D_L^k\right] & \forall k: D_0 > E\left[D_L^k\right] \end{cases}$$

Maximum Expected Distortion Constraint per packet

$$and$$

$$T_{tot} = \sum_{k=1}^{K} \frac{B^k(\mu^k)}{R} \le T_0$$

Delay Constraint for frame

Min Energy Solution

- Coupling Power to Source Coding Parameters
  - For the distortion constraint
  - Assumption: spatially causal concealment strategy

$$\underset{\{\mu^k\}}{\text{minimize}} \qquad J_{tot} = \sum_{k=1}^{K} J^k(\mu^k)$$

Energy

Delay

$$J^k(\mu^k) = \begin{cases} \overbrace{\dfrac{B^k(\mu^k)}{R} g\left(\dfrac{E\left[D_L^k\right] - D_0^k}{E\left[D_L^k\right] - E\left[D_R^k(\mu^k)\right]}\right)}^{\text{Energy}} + \lambda \overbrace{\dfrac{B^k(\mu^k)}{R}}^{\text{Delay}} & \text{if } E\left[D_R^k\right] \le D_0^k \le E\left[D_L^k\right] \\ 0 & \text{if } D_0^k > E\left[D_L^k\right] \end{cases}$$

Generalized Skip

Experimental Results

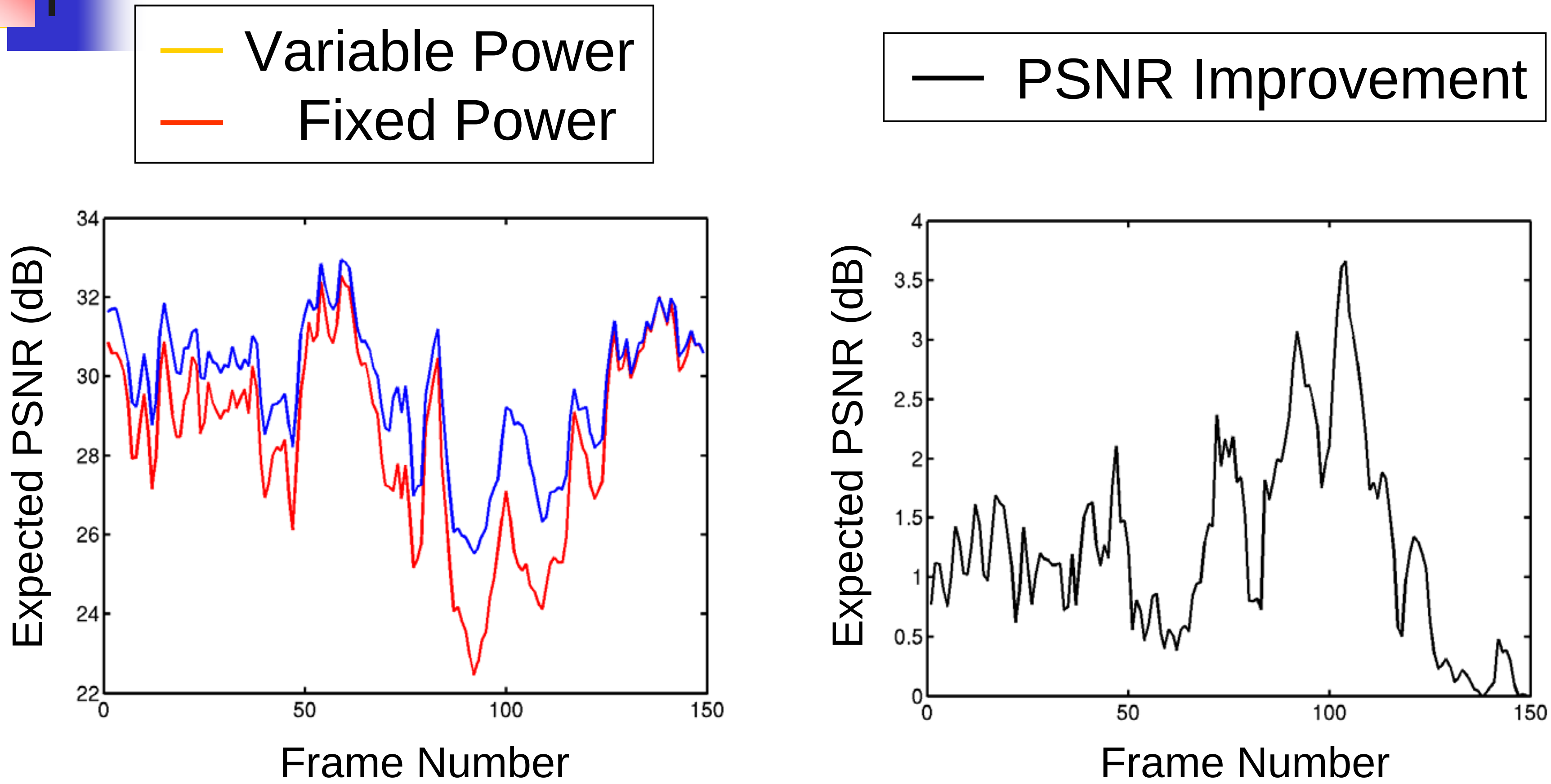
<u>Compare:</u>

- Variable Power Approach
  - Joint Source Coding and Power Allocation
  - min $E[D(S,P)]$; s.t.: $B_{tot}(S) \leq B_{\max}$, $E_{tot}(S,P) \leq E_{\max}$
- Fixed Power Approach
  - Independent Source Coding and Power Allocation
  - min $E[D(S)]$; s.t.: $B_{tot}(S) \leq B_{\max}$

<u>Setup:</u>

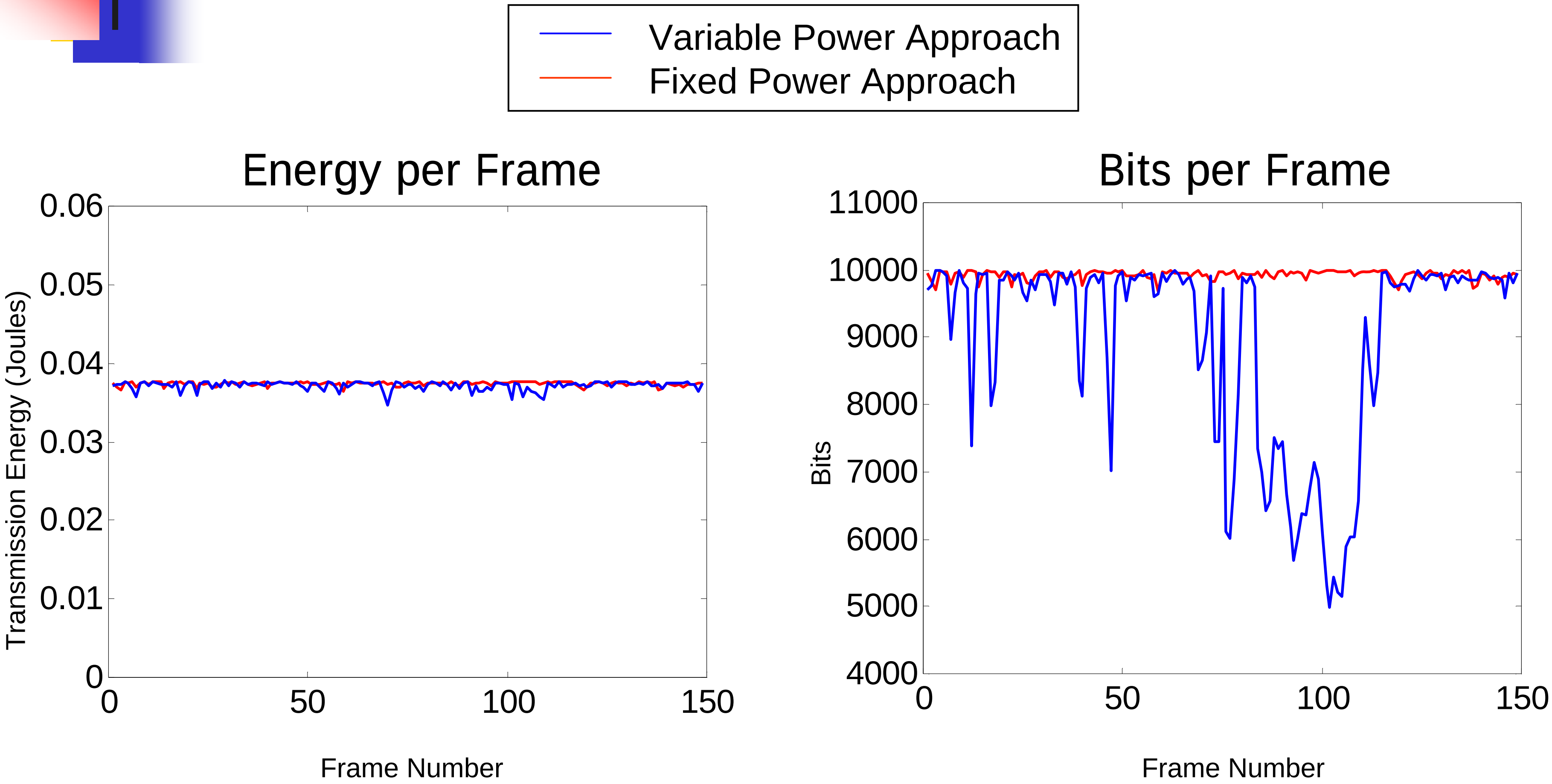
- Fixed Delay Constraint ($T_{\max}$ = 33 ms)
- Fixed Transmission Rate (R = 300 kbps)
- Packetization: One MB per packet
- Concealment: Based on Neighboring MB to the left

Expected Distortion Per Frame



“**Foreman**” Sequence: transmission rate = 300 kbps ; avg. prob. of error = 0.20

Energy and Bits per Frame



Visual Comparison: Slow Motion

Fixed Power Approach

Variable Power Approach

- Same energy and delay constraints per frame
- Approaches differ in source coding and power allocation

Visual Comparison: Real-Time

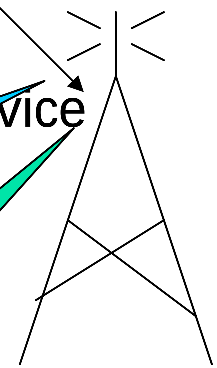
Fixed Power Approach

Variable Power Approach

- How do spatio-temporal artifacts affect the *perceived* video quality?

Another Application

- General Problem Definition
  - Transmit multiple pre-encoded video sequences
  - To multiple users
  - Over HSDPA
  - Find optimal distribution of resources at base station
  - Such that each user receives a reasonable quality of service
- Limited Resources (limits achievable data rates)
  - Transmission power
  - Number of spreading codes (Bandwidth)
- Quality of Service Measures
  - End-to-end distortion of video sequence
  - Transmission delay (stringent requirement for real-time applications)

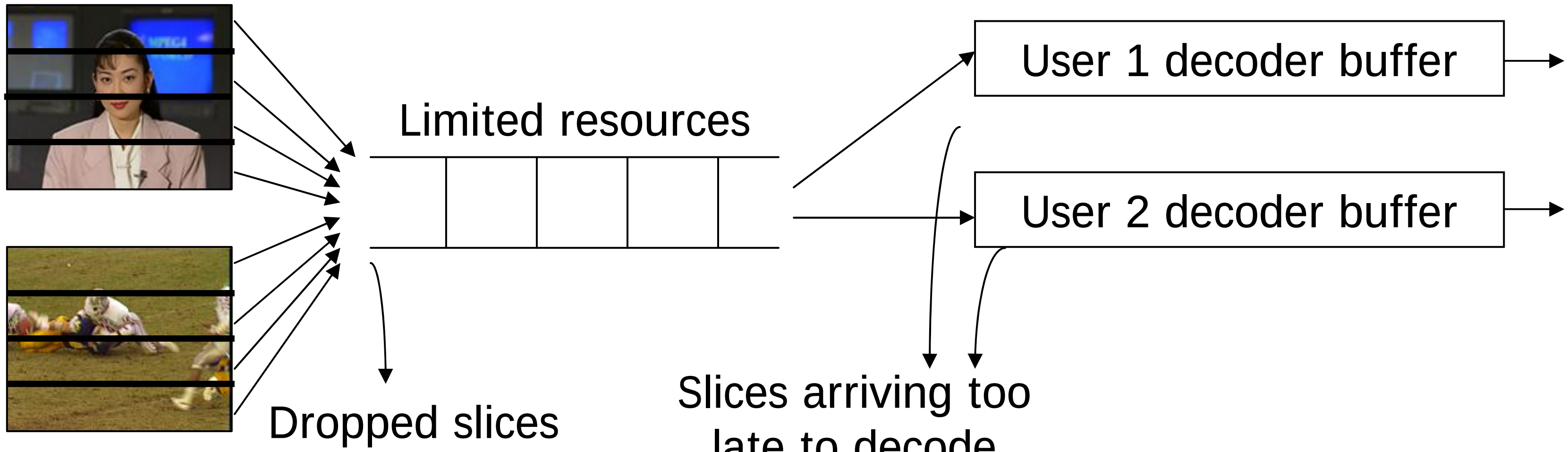


Downlink Packet Scheduling

- “Dumb” Method
  - Round Robin
- State of the Art
  - Basic Idea: Allocate resources to users with better channel quality
  - Maximum Throughput Methods
  - Proportionally Fair Methods
    - Fairness criterion based on current average throughput
  - Gradient Based Scheduling
    - Maximize rate to users that will gain the most, subject to channel conditions
- Scheduling for Streaming Video
  - Minimize queue length (delay of head-of-line packet)
  - Current work does not consider rate-distortion trade-offs for individual video packets

Video Transmission Assumptions

- Frames are split into independently decodable slices
- Video will be viewed in real-time (slices from the current frame must be received by the decoder before it finishes decoding the previous frame)
- Achievable data rates may not be sufficient to transmit every slice of every frame to all the users within the real-time constraints



Formulation

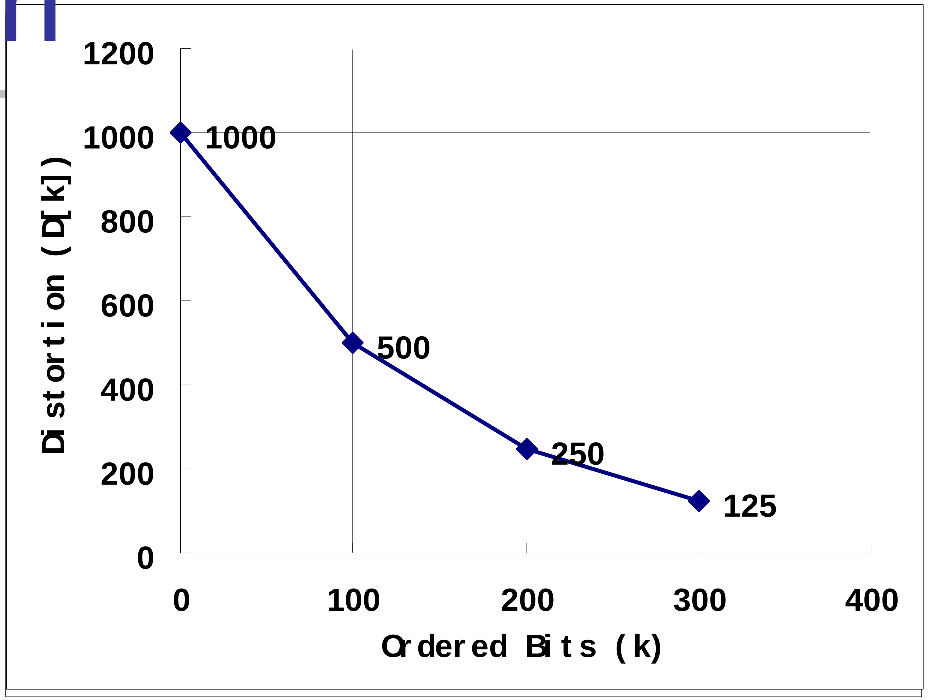
- Key Idea
  - Order slices by amount of distortion reduction
  - Define Slice Utility in terms of distortion reduction
- Assumptions
  - Channel states may change by an order of magnitude over one frame's duration
    - Feedback is available every 2ms in HSDPA
    - Optimization performed every time slot (2ms)
  - Simple error concealment
    - Copying MB from the same position in previous frame
    - Complex error concealment adds dependencies between slices
      - complicates slice ordering
    - Will eventually include complex error concealment

Utility Function

- Utility Per User Based on Slice Ordering
  - $D[k]$:= Distortion given k slices are received
  - $D[M]$ = Minimum distortion

    where, M = total # of slices

$$U[k] = \left(D[M] - D[k]\right)$$



Distortion vs bits after slice ordering

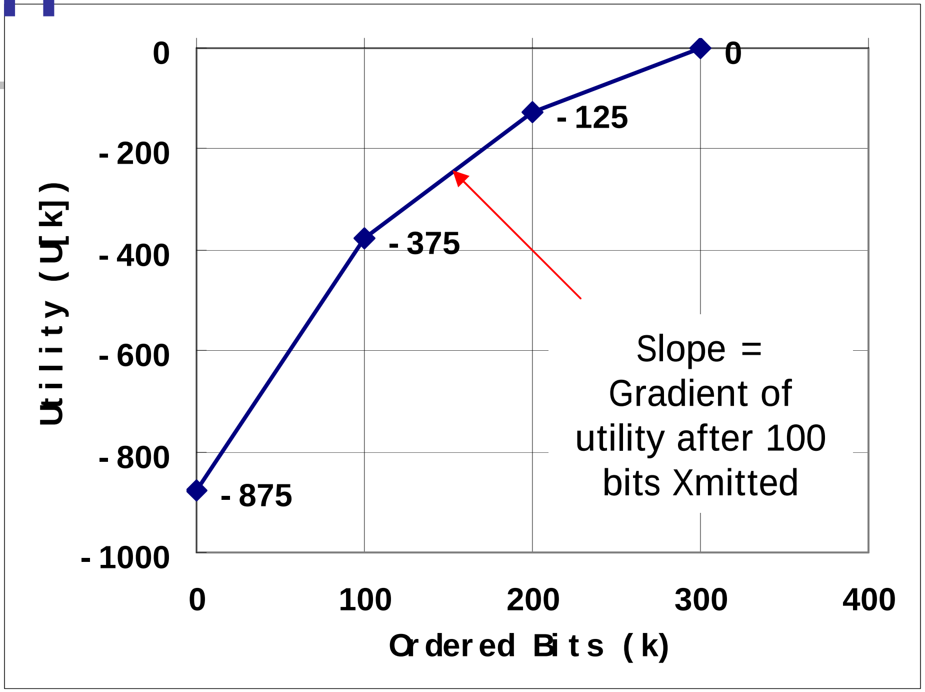
Utility Function

- Utility Per User Based on Slice Ordering
  - $D[k]$:= Distortion given k slices are received
  - $D[M]$ = Minimum distortion

    where, M = total # of slices

$$U[k] = \left(D[M] - D[k]\right)$$

$$u[k] = U'[k] = \frac{D[k] - D[k+1]}{B_{k+1}}$$



Utility vs bits after slice ordering

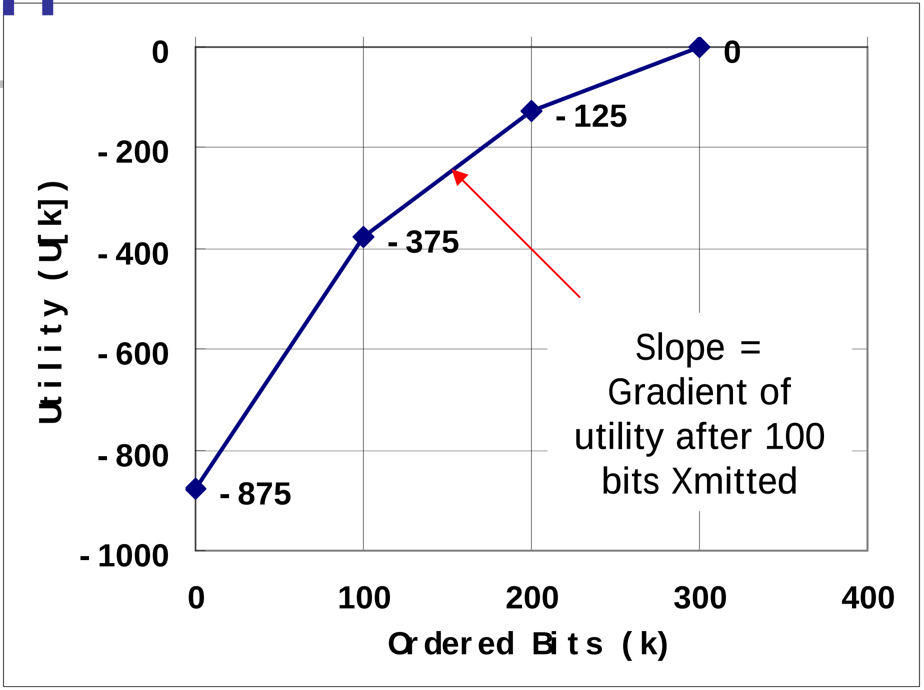
Utility Function

- Utility Per User Based on Slice Ordering
  - $D[k]$:= Distortion given k slices are received
  - $D[M]$ = Minimum distortion

    where, M = total # of slices

$$U[k] = \left(D[M] - D[k]\right)$$

$$u[k] = U'[k] = \frac{D[k] - D[k+1]}{B_{k+1}}$$



Utility vs bits after slice ordering

- Gradient Based Scheduling – Maximizes rates per user weighted by gradient of utility function

$$\max \sum_{i=1}^{K} w_i u_i \cdot r_i$$

K : # of users
r_i : rate per user
w_i : additional weighting

HSDPA System Constraints

- K users
- Maximum transmission power: → $$\sum_{i=1}^{K} p_i \le P$$
  - p_i := power per user
- Number n_i of codes per user: → $$n_i \le N_i$$
- Total number of spreading codes: → $$\sum_{i=1}^{K} n_i \le N$$
- Achievable rates: $$r_i = n_i \Gamma\left(\xi_i \cdot SINR_i\right)$$
  - $SINR_i := \dfrac{p_i}{n_i} e_i$
  - Where e_i := channel state (SINR per unit power)
  - Γ := Shannon capacity assuming Gaussian noise channel
  - $\xi_i \in (0,1]$:= gap from capacity

Problem Definition

- Maximize sum of user rates weighted by utility gradients:

$$V^* = \max_{\mathbf{n},\mathbf{p}} \sum_{i=1}^{K} w_i u_i \cdot n_i B \log\left(1 + \frac{p_i e_i}{n_i}\right)$$

$$\text{subject to : } \sum_{i=1}^{K} n_i \leq N$$

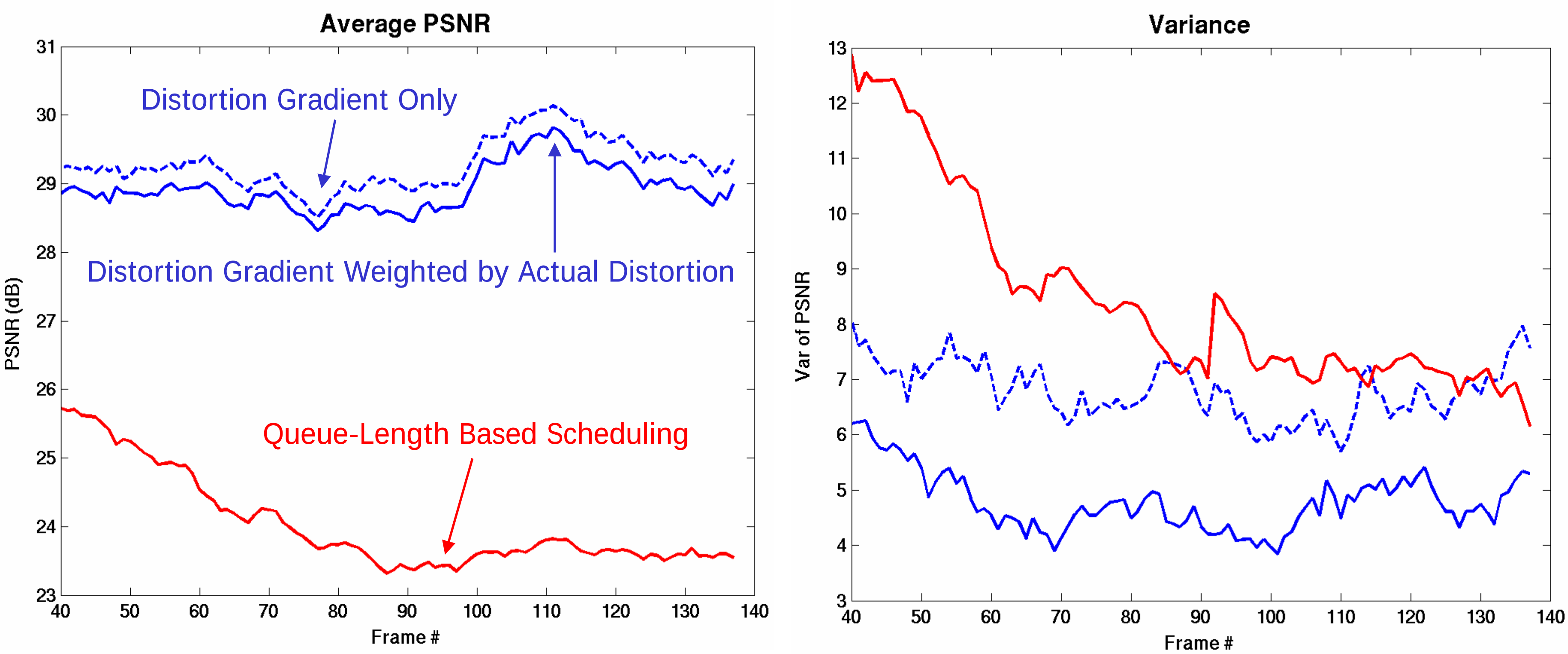
$$\text{and } \sum_{i=1}^{K} p_i \leq P$$

$$\text{additional constraint : } n_i \leq N_i$$

Simulation I : Similar Fading Stats

- HSDPA System
  - # of Users (K) = 12
  - Total # of spreading codes (N) = 15
  - Max codes per user (N_i) = 5
  - Total Power (P) = 9.9W
  - Max SINR = 1.59
  - Simulated uncorrelated Raleigh fading channels, with same average channel conditions, for each user
- Video Source
  - H.264 encoded at 256kbps
  - QCIF (176x144) {foreman, mother and daughter, carphone, news, silent, 0…139, 150…289}
  - 11 MBs per slice

Simulation I Results



Demo 1

- 6 users
- P = 7W

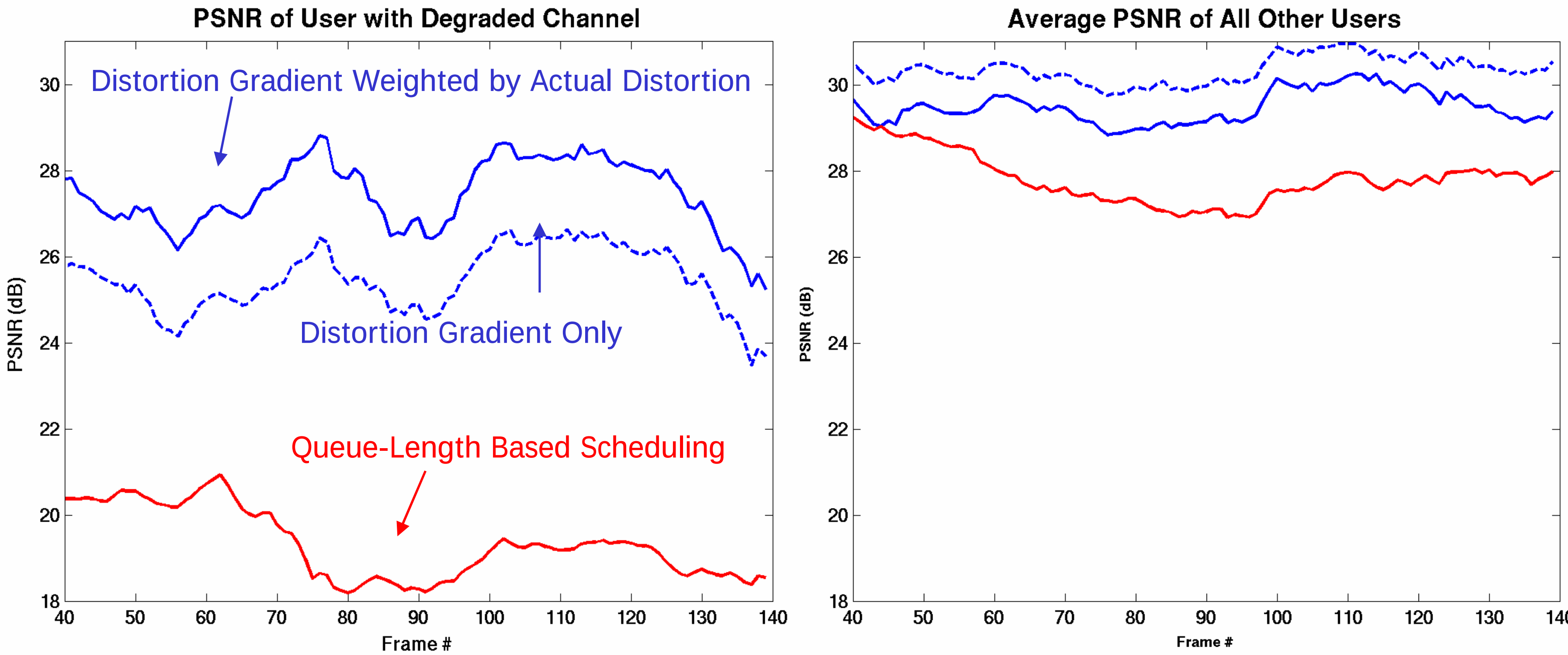
Weighted Distortion Gradient Metric

Queue-based Utility Metric

Simulation II : One Degraded User

- HSDPA System
  - # of Users (K) = 10
  - Total # of spreading codes (N) = 15
  - Max codes per user (N_i) = 5
  - Total Power (P) = 9.9W
  - Max SINR = 1.59
  - Simulated uncorrelated Raleigh fading channels, with one user having lower average channel quality
- Video Source
  - H.264 encoded at 256kbps
  - QCIF (176x144) {foreman, mother and daughter, carphone, news, silent, 0...139, 150…289}
  - 11 MBs per slice

Simulation II Results



- The weighted distortion gradient tends to be fair to all users while not significantly sacrificing overall quality

Demo 2

Reception Quality for User with Degraded Channel

Distortion-based Utility

Queue length-based Utility

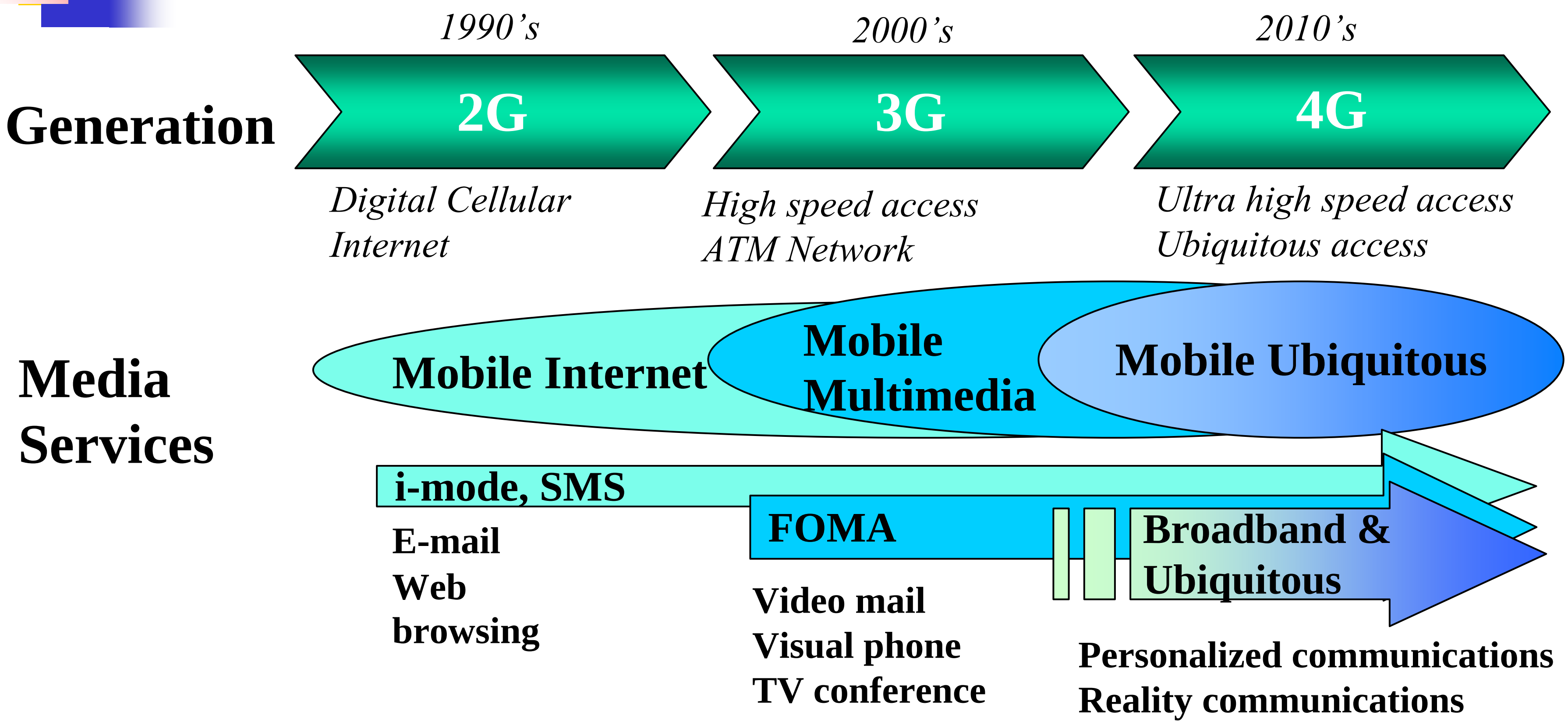
Closing Remarks

- Rich topic at the intersection of multimedia signal processing, communications, and networking
- Increasing number of applications
- Touched the tip of the iceberg

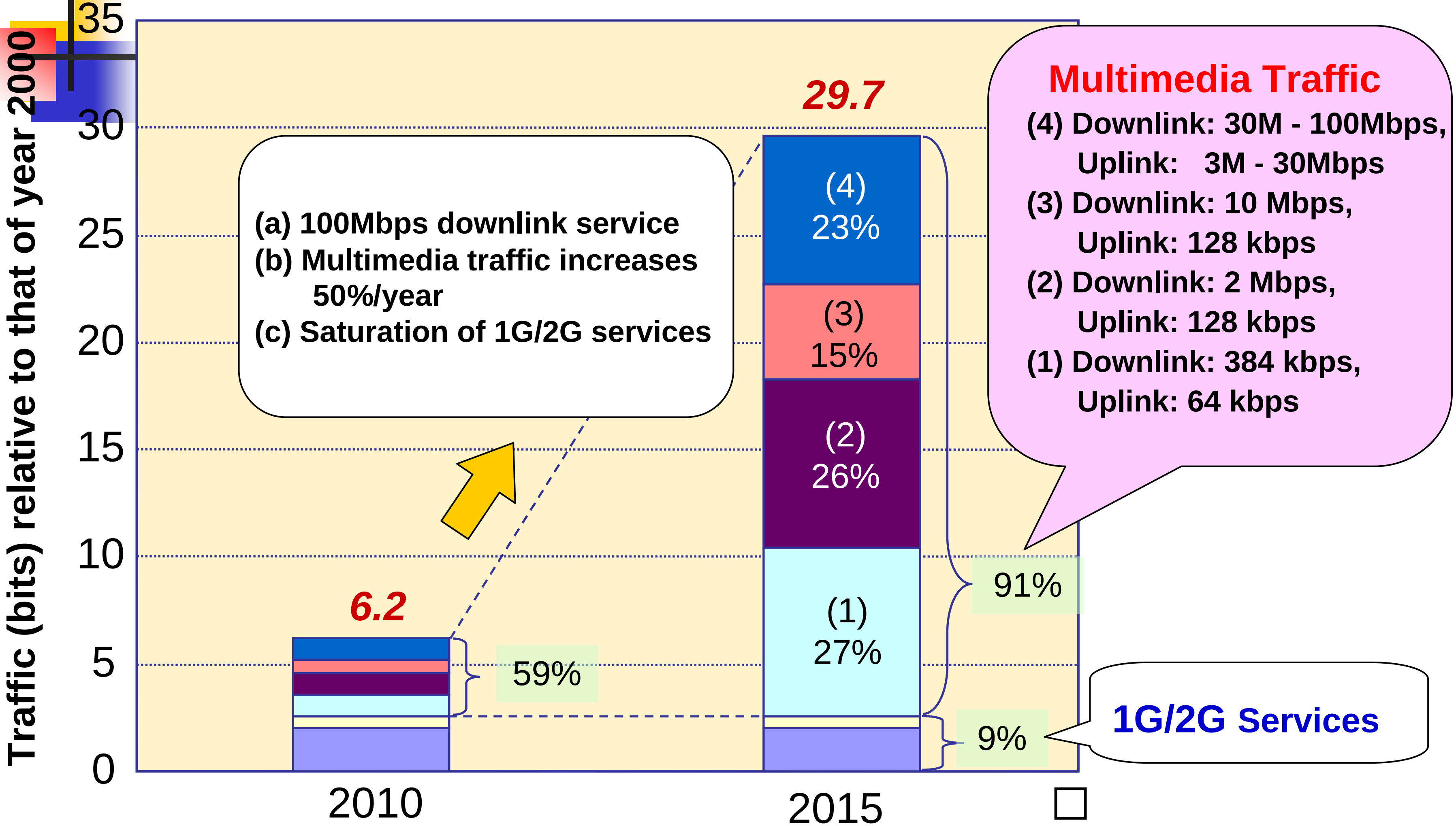
Future Research Directions

- Multi-hop, Image/Video Sensor Networks
- Multi-user, up- and down-link
- Hybrid networks
- More sophisticated concealment methods
- Distortion metrics, utility functions
- Scalable coding, MDC
- Data hiding appraoches

Network Evolution



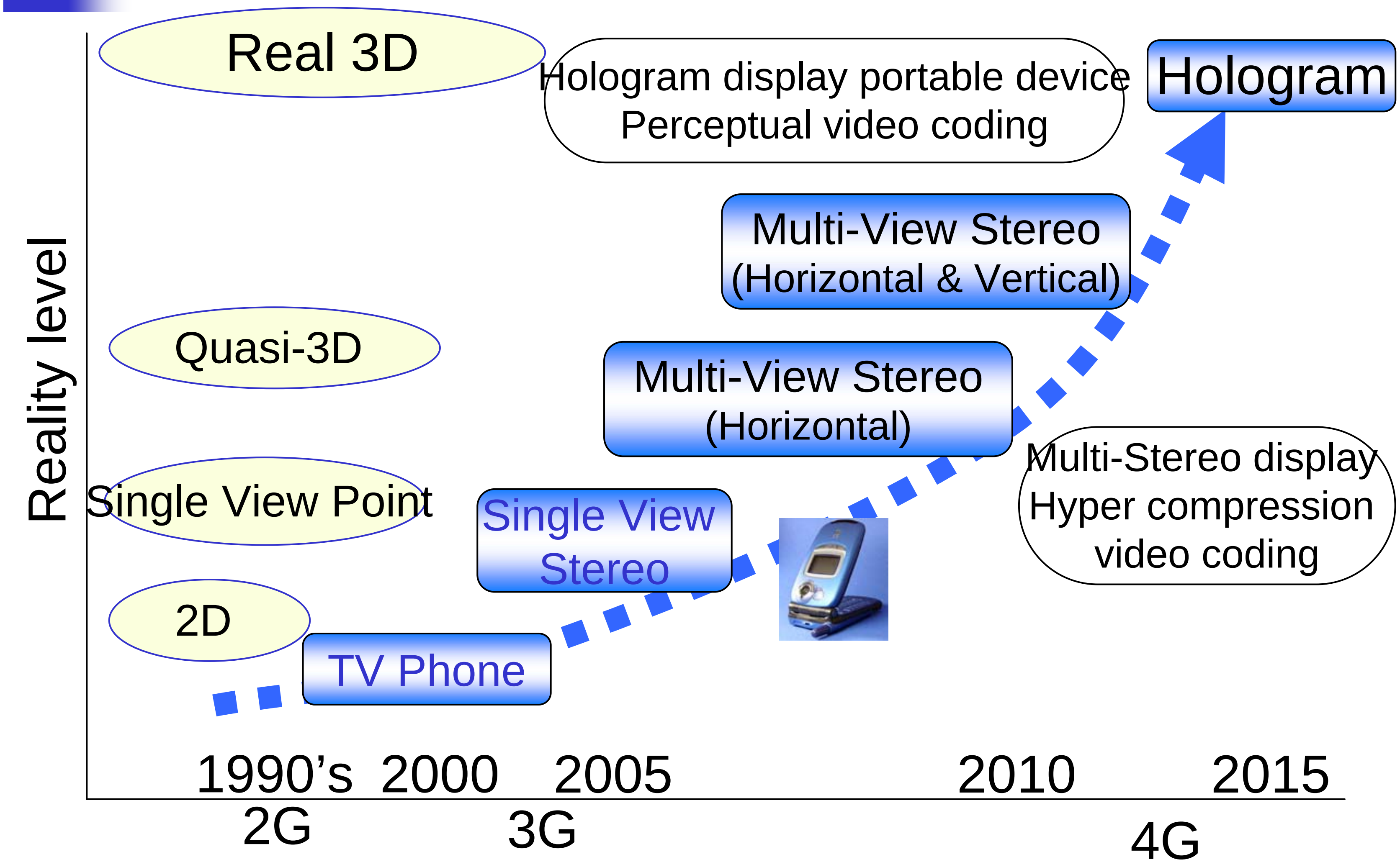
Forecast on Future Multimedia Traffic



[Forecast from ITU-R TG- 8/1 for Asia]

[Forecast from Telecommunication Council of Japan]

Reality Visual Communications



Requirements for Future Mobile Networks

Media	Transmission speed	Delay	Connection Latency	Terminal capabilities
Speech/ 3D Audio	< 1 Mbps	<50ms	<1sec	• 3D sound field control • High efficiency loud speakers
Video/ 3D video	10Mbps (2D video) ~ 30Gbps (3D video)	<50ms	<1sec	• Real time hologram
Enhanced Reality	< 1Mbps	<< 50ms Should be predictable	N/A	•Eyeglass display •3D and multimodal UI
Five senses communications	< 1Mbps	<50ms	N/A	•Five sense sensors
Tele-existence	<10Mbps (Robotic I/F) < 1Gbps (Virtual avatar) < 100Mbps (Alter-ego existence)	< 10ms < 30ms < 5ms (Small and known jitter)	<1sec	• Alter-ego robot

Additional Work

- Cross Layer Design
- Resource-Distortion Optimization Framework (A. K. Katsaggelos et al, “Advances in Efficient Resource Allocation for Packet-Based Real-Time Video Transmission”, *IEEE Proc., Jan.* 2005)
- JSCC (L. Kondi, F. Ishtiaq, and A. K. Katsaggelos, "Joint Source-Channel Coding for Motion-Compensated DCT-Based Scalable Video," *IEEE Trans. IP*, vol. 11, pp. 1043-1052, Sept. 2002)
- Energy Efficient Wireless Video Streaming (Y. Eisenberg, C. Luna, T. Pappas, R. Berry, A. K. Katsaggelos, “Joint Source Coding and Transmission Power Management for Energy Efficient Wireless Video Communications”, *IEEE CASVT*, pp. 411-424, June 2002)
- Data Rate Adaptation and Scheduling (C. Luna, Y. Eisenberg, R. Berry, T. Pappas, A. K. Katsaggelos, “Joint Source Coding and Data Rate Adaptation for Energy Efficient Wireless Video Streaming”, *IEEE JSAC*, pp. 1710-1720, Dec. 2003)
- DiffServ (F. Zhai, C. Luna, Y. Eisenberg, R. Berry, T. Pappas, A. K. Katsaggelos, “Joint Source Coding and Packet Classification for Real Time Video Transmission over DiffServ Networks”, *IEEE Trans. Multimedia*, to appear).
- Variance-Aware Distortion Estimation (Y. Eisenberg, F. Zhai, C. Luna, T. Pappas, R. Berry, A. K. Katsaggelos, “Variance-Aware Distortion Estimation for Wireless Video Communications”, *IEEE Image Processing, to appear*)

Additional Work (cont'ed)

- Hybrid (FEC, retransmission) Rate Control (F. Zhai, Y. Eisenberg, T. Pappas, R. Berry, A. K. Katsaggelos, "Rate Distortion Optimized Hybrid Error Control for Real-Time Packetized Video Transmission", *IEEE Trans. Image Processing*, to appear).
- Hybrid Intra/Inter FEC
- Hybrid (wired/wireless) Networks
- Object-based Encoded Video Transmission (*CSVT 2006*)
- FGS Scalable Video Transmission *(ICC '04)*
- Correlated Fading, Channel Mismatch Sensitivity (*ICIP'04*)