http://wireless.poly.edu

An Analytical Foundation for Radio Resource Management applied to Video Streaming

Virgilio Rodriguez

Some discussed work co-authored with D.J. Goodman and Y. Wang

ECE Department
Polytechnic University
Brooklyn, NY
email: vr@ieee.org
http://pages.poly.edu/~vrodri01/research

Outline of research program

- Analytical Core
- Single-user applications: data, image, video
- Decentralized multi-user applications:
  - Game formulation
  - Mechanism design
- Centralized data throughput maximization
  - Without noise
  - With noise and media terminals present

Overview of the framework

- Many radio-resource optimizations share a common analytical core
- This core enables robust and tractable analysis and provides clear answers in fairly general scenarios
- It involves
  - A tractable abstraction of the physical layer
  - A tractable abstraction of the human visual system
  - A fundamental result: maximize $f(x)/x$ with f an "S-curve".
- Problems to which this framework applies:
  - Power and coding rate choice for media files (images, video)
  - Choosing the "right amount" of media distortion
  - Decentralized power control for 3G CDMA
  - Data rate and power allocation for maximal cell throughput when data and media terminals share a CDMA cell

Sample application: Coding rate & power for video streaming

- Each T-secs of video yields "scalable" file (i.e., file can be truncated and decoded; e.g., MPEG-4, SPIHT-3D)
- Energy E is limited!
- File for given segment must be transferred in a deadline of Δ secs.
- Files will be split into small packets for transmission purposes; ECC bits will be added and an ARQ system will be available
- Transferring each file complete $\Rightarrow$ maximal quality per segment BUT short total viewing time with available energy. Transferring few bits per file $\Rightarrow$ long running time BUT low quality per segment.
- Problem: how many bits per file to transfer (where to truncate) AND at which power to transmit?

What do we need to formulate this problem?

In order to formulate and solve this problem we need:

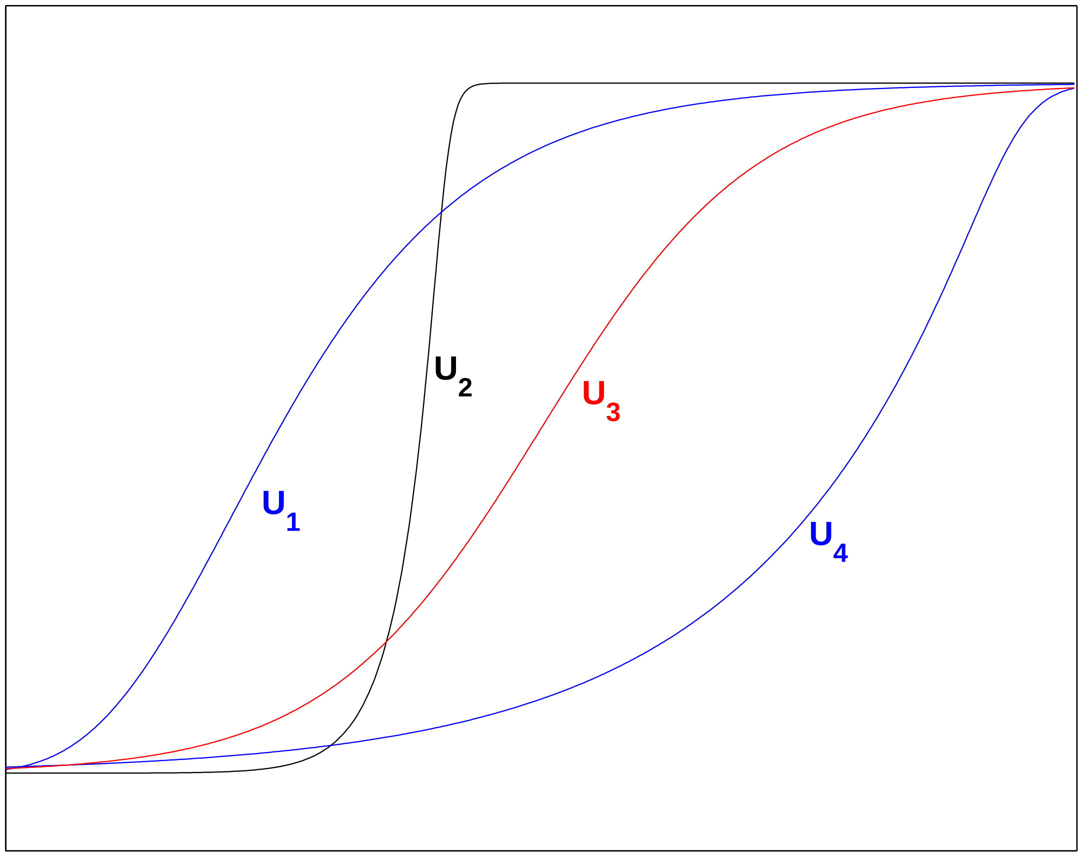
- A function $U(y)$ giving the end-user “perceptual quality” or “utility” of decoded video segment when there are y bits in the corresponding *truncated* file (coding rate).
- A function $f(x)$ giving the probability of successful reception of a data packet when the SIR at the receiver is x.
- A criterion giving an index to be optimized as function of the quality of individual video segments

An abstraction of the physical layer

- For resource-management purposes, a frame success function (FSF) encapsulates the essential information about the physical layer
- The FSF gives probability that a data packet is received successfully as function of SIR at receiver
- It's determined by the details of the physical layer: modulation, diversity, FEC, etc.
- Ex: for Gaussian channel, non-coherent FSK modulation, with packet size M=80, no FEC, independent bit errors, and perfect error detection, the FSF is $f_s(x) = \left[1 - \frac{1}{2}\exp\left(-\frac{x}{2}\right)\right]^{80}$
- If the analysis assumes that all that is known about the FSF is that it is a smooth "S-curve", the analysis de facto accommodates most physical layers of interest

S-curves

- For any physical layer, the function giving the probability that a data packet is received successfully as function of the SIR is an S-curve
- An arbitrary S-curve includes as special cases
  - an arbitrary convex curve
  - an arbitrary concave curve
  - an arbitrary threshold (step)
  - a straight line (almost)



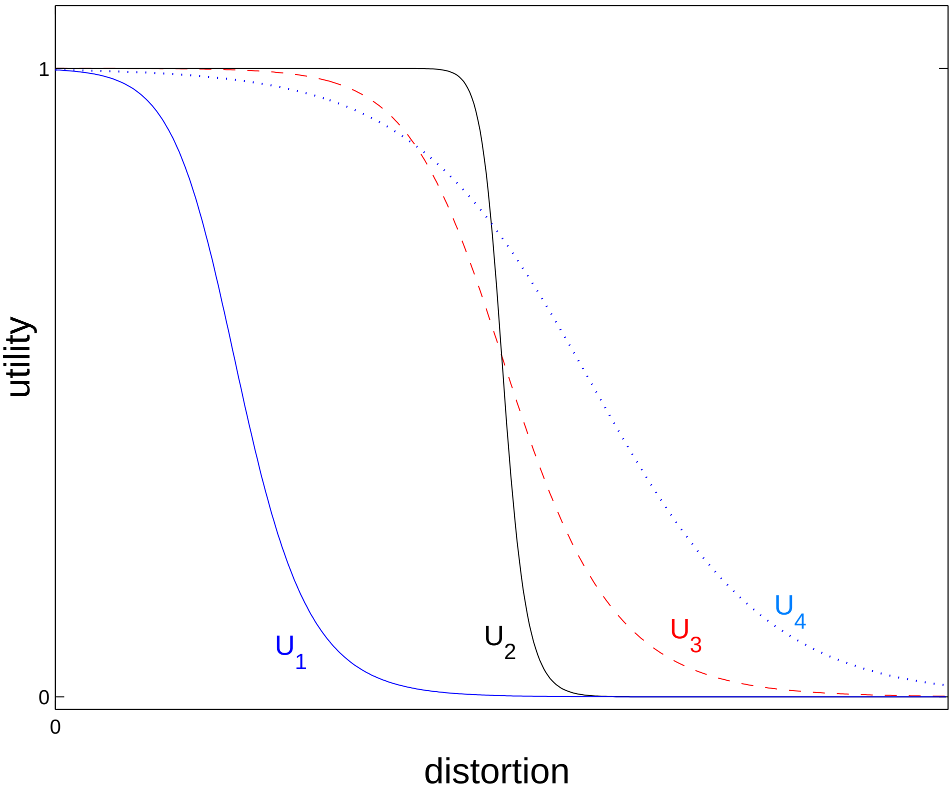
Frame success rate versus SIR

Rate/quality, rate/distortion, quality/distortion

- We want a quality/rate function, $U(y)$. But there is a well developed theory relating rate to distortion.
- Distortion is a very simple measure of the difference between a signal and its copy (e.g., the original vs. the reproduced video segment)
- The perceptual quality of a distorted signal is determined by the human visual system (HVS)
- It is reasonable to assume that the perceptual quality of the reconstructed signal is determined by distortion; i.e., that a function $Q(D)$ that translates distortion into perceptual quality can be found.
- As function of distortion, $Q(D)$ must be decreasing,..., but with which "shape" (convex, concave, linear, step, etc.) ?
- $Q(D)$ can be obtained by psychophysical experiments for a specific user.

An abstraction of the Human Visual System

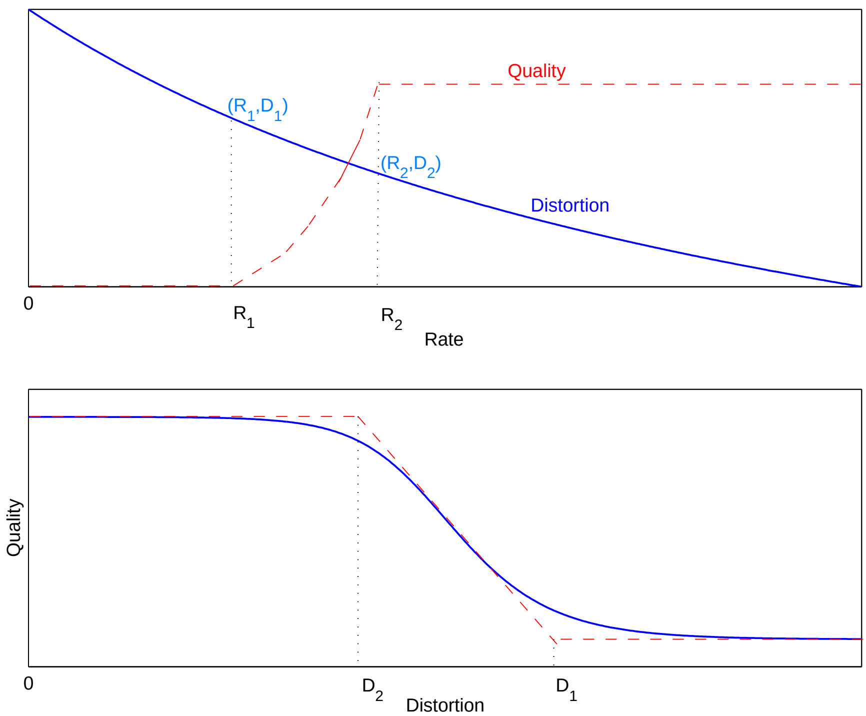
- The resource management literature typically assumes that, up to a level, distortion has no effect on signal quality, but beyond that level it makes the signal useless . This is equivalent to assuming that $Q(D)$ is a step function ("hard threshold").
- By assuming that all that is known about $Q(D)$ is that it is a "reversed" S-curve, the "hard threshold" and many plausible curves ("almost" convex, "almost" concave, "almost" linear, etc.) are contained as special cases.
- For further details on this approach, and how it can be applied to an interesting problem see additional slides and/or a complete paper.



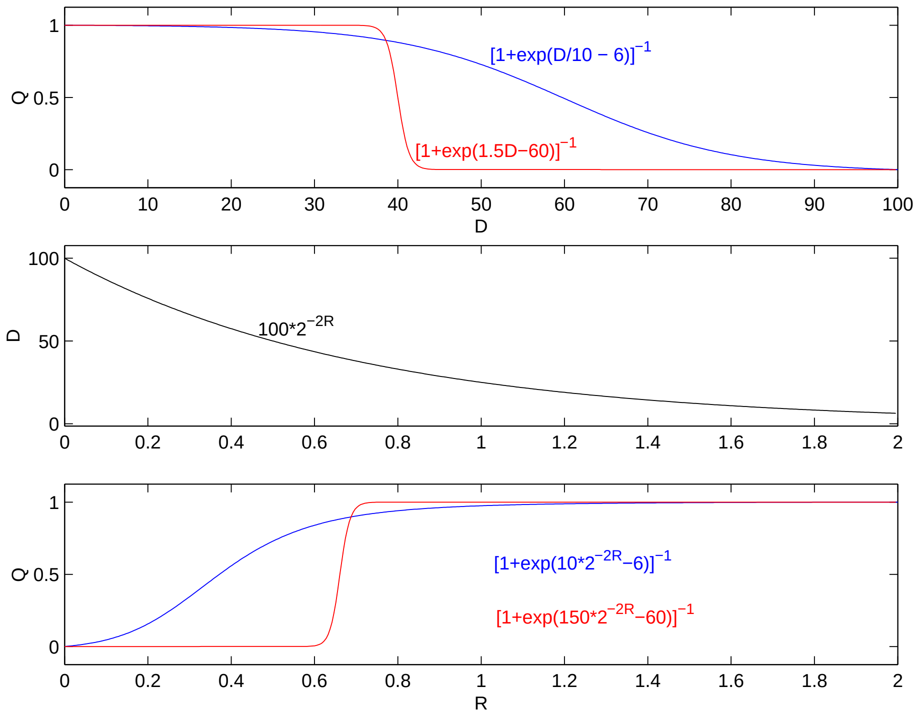
Beauty is in the eyes of the beholder

From rate-distortion to rate-quality theory

- It is generally accepted that the function $D(R)$ giving distortion as a function of the coding rate is decreasing and convex.
- For the memoryless Gaussian source, $D(R) \propto 2^{-2R}$
- Given a quality-distortion function $Q(D)$ and a distortion-rate function $D(R)$ the desired quality-rate function is $Q(D(R))$
- Question: if all that is known about $D(R)$ is that it is decreasing and convex, and all that is known about $Q(D)$ is that it is a "reversed" S-curve, what can be said about $Q(D(R))$?



For $Q(D)$ a reversed S-curve, we expect $Q(D(R))$ to be an S-curve

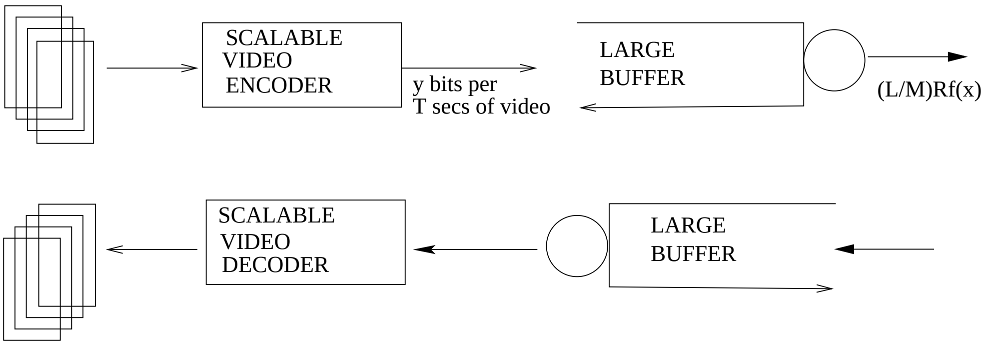


For a memoryless Gaussian source, $D(R) \propto 2^{-2R}$. For the $Q(D)$curves at the top, $Q(D(R))$ are S-curves (bottom).

Wireless scalable video streaming revisited

- Each T-secs of video yields “scalable” file (can be truncated and decoded; e.g., MPEG-4, SPIHT-3D)
- Assume S-curve $u(y)$ gives segment quality from y-bit truncated file
- Energy E is limited!
- File for given segment must be transferred in a deadline of Δ secs.
- Transferring each file complete $\Rightarrow$ maximal quality per segment BUT short total viewing time with available energy. Transferring few bits per file $\Rightarrow$ long running time BUT low quality per segment.
- Problem: how many bits per file to transfer (where to truncate) AND at which power to transmit?
- Criterion: Maximize total utility: $n \times u(y)$ with $n = E/c(y)$ with $c(y)$ the energy cost of successfully transmitting a y-long file in Δ secs.

System for Wireless Scalable Video



Schematic of the wireless transmission of scalably encoded live video. As first approximation, assume channel is "pseudo-deterministic" delivering $(L/M)Rf(x)$ correct information bits per sec. R is the raw bit rate, $f(x)$ the frame-success rate, and L/M the ratio of information bits to the packet size.

Solution Setup

- Maximize total quality : $n \times u(y)$ with $n = E/c(y)$ and $c(y)$ the energy cost of successfully transmitting a y-long file in Δ secs.
- For given y and Δ, $\exists$ a specific SIR $x(y)$ that satisfies

$$\frac{L}{M} R f(x) \Delta = y$$

- There is a specific transmitted power, $P(y)$, that yields $x(y)$
- Thus, the total number of T-sec video segments of quality $u(y)$ that can be transferred with an energy budget of E is $E/(P(y)\Delta)$.
- The total quality viewed is

$$\frac{E}{\Delta} \frac{u(y)}{P(y)}$$

- For fixed E, maximize quality/Joule, and for fixed Δ, max quality/Watt

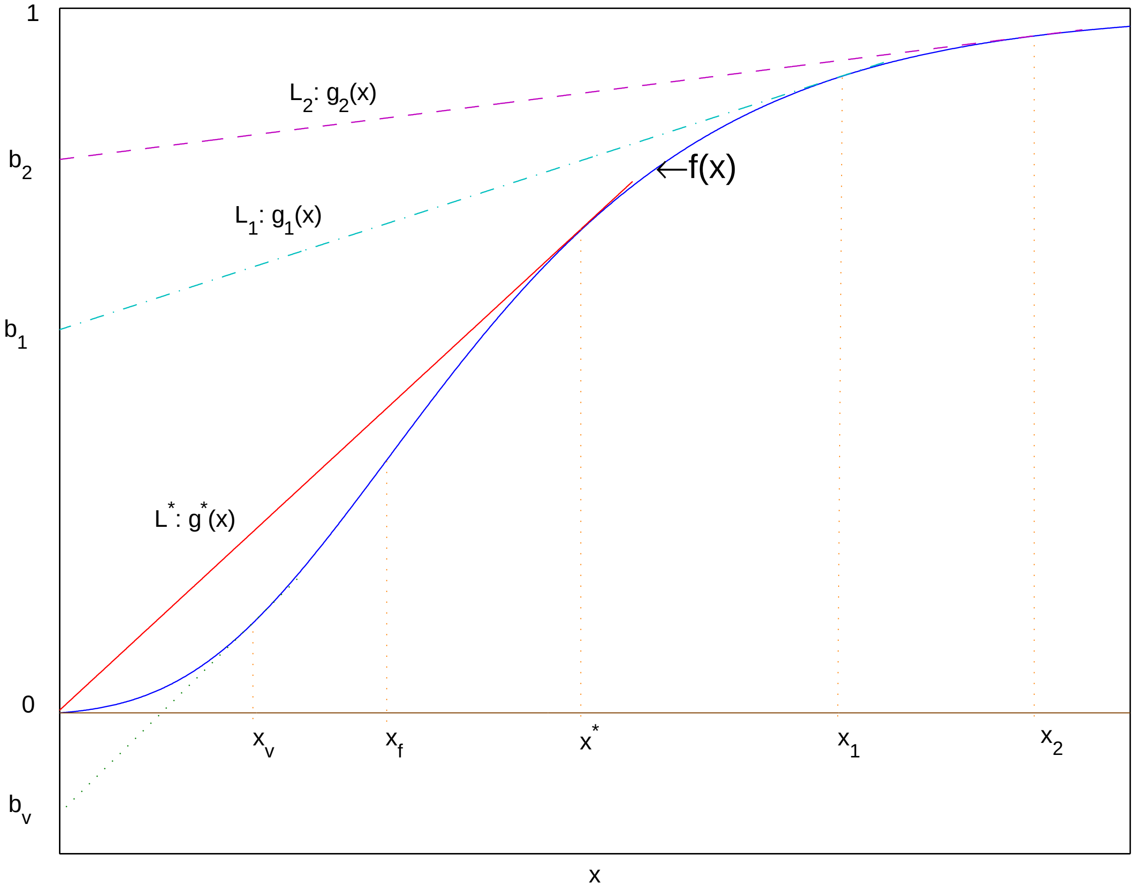
Solution

- $$\begin{array}{rcl} \max\limits_{x,y} \dfrac{u(y)}{x} & & \max\limits_{x} \dfrac{u(Bf(x))}{x} \\ \text{s.t. } y = Bf(x) & \text{OR} & \text{s.t. } 0 \le x \le \bar{x} \\ 0 \le x \le \bar{x} & & \end{array}$$

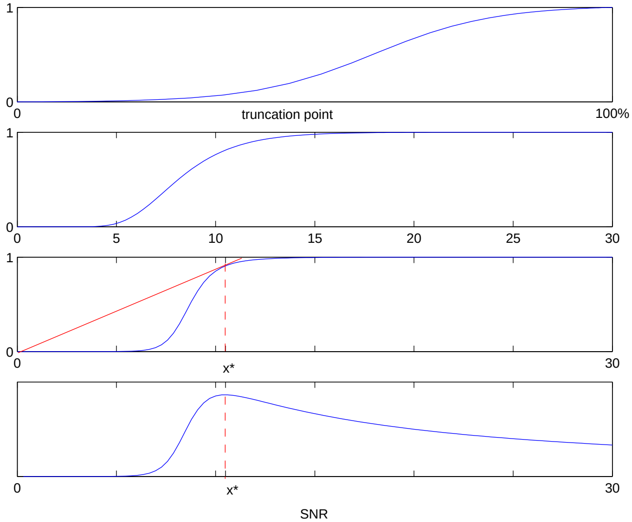
- $B = (L/M)R\Delta$ interpreted as the maximum amount of information bits ("best case scenario") that can be transferred in the deadline Δ.
- u and f are both S-curves. We expect the composite function $h(x) := u(Bf(x))$ to retain the S-shape.
- $\therefore$, we need the solution to max $h(x)/x$ when all that is known about h is that it is an S-curve.

Maximizing S(x)/x

- Maximize $f(x)/x$ where *all that is known* about f is that its graph is an S-curve.
- No functional form (“equation”) is imposed
- Sigmoidness $\Rightarrow$ f “starts out” convex at the origin, and “smoothly” transitions to concave as it approaches a horizontal asymptote
- Maximizer must solve $xf'(x) = f(x)$.
  Solution:
  - always exists
  - is unique
  - can be graphically described by drawing a tangent
- Ratio $f(x)/x$ is quasi-concave (enables application of Debreu’s and other results)



x^* is the unique maximizer of $f(x)/x$



From the top, (i) the S-curve $u(y)$ giving the perceptual quality of a video segment, as a function of the coding rate, (ii) $f(x)$, the probability of successful reception of a packet as a function of the SIR, (iii) the composite function $u(Bf(x)) := s(x)$, (iv) the ratio $s(x)/x$ which the terminal should maximize.

What is a game?

- Preceding analysis involves a single terminal. One way to consider many terminals is by setting up a “game”.
- Game: each of several players chooses a “strategy” in order to maximize a “payoff”.
- Payoffs depend on the choices of ALL players
- Each player is “selfish”
- Key solution concept: Nash equilibrium.
  An allocation (a strategy per player) such that no player would gain by unilaterally changing strategy (“deviating”)
- Nash equilibria are generally “inefficient”

Power control game

- Players: CDMA data transmitting terminals.
- Strategy: transmission power level
- Payoff : number of bits successfully transmitted per unit energy (bits/Joule)
- Signal-to-interference ratio determines bits/Joule
- A Nash equilibrium generally exists
- Equilibrium power levels are “too high”
- Challenge: how to get selfish terminals to choose lower power levels “on their own”
- For further details on this game see this WCNC-03 paper .

Mechanism design for decentralized efficiency

- "mechanism" : a set of procedures, penalties and rewards designed to guide selfish entities toward a desired outcome
- Example of a simple and useful mechanism:
  Vickery's Second Price Auction
  - Each player chooses an amount of money to make a sealed bid for an object, to be won by highest bidder
  - But highest bidder pays second-highest bid
  - Each player's best response is to bid its true valuation of object: "truth-telling" is optimal

The compensation mechanism

- Proposed by Varian in a general context
- Requires a "transferable good", say money, with which agents compensate each other.
- Assuming only 2 terminals, and that terminal 1 interferes with terminal 2 but *not* vice-versa (SIC decoding), it works as follows
  - Terminal 2 declares the amount money (or transferable good) it wishes to charge terminal 1 as compensation for each unit of interference.
  - Terminal 1 (interferer) declares the price it *offers* to pay terminal 2 as compensation.
  - The interferer (#1) must pay penalty if its offered price is different from terminal 2's price

Why does the mechanism work?

- To avoid the penalty, generally the interferer will offer to pay the exact amount terminal 2 wants.
- But why doesn't terminal 2 ask "too much"?
  - If price paid to terminal 2 exceeds its "true cost", then it "makes a profit" per unit of interference.
  - But then, it would want more interference!
  - To get the interferer to produce more, terminal 2 must lower its price.
  - Thus, at equilibrium, terminal 2 price equals its true cost, which is the "fair thing" to do.
- The mechanism also works when both terminals interfere each other, and with many mutually interfering terminals.
- For further details see additional slides and/or an extended abstract.

Discussion

- An analytical foundation for wireless resource management has been discussed, and applied to wireless video streaming.
- At the core are functions about which all that is known is that their respective graphs are S-curves. No “equation” is used.
- The family of S-curves include (i) “mostly” concave, (ii) “mostly” convex, (iii) “mostly” straight, (iv) and “step” “curves”.
- An S-curve can yield a useful abstraction of (i) the physical layer of a wireless communication system and (ii) the human visual system
- With f any S-curve, the unique maximizer of $f(x)/x$ can be easily identified by drawing a tangent, and is crucial in several applications.
- Several interesting problems involving coding rate, data rate and power allocation can be solved by applying this framework.

Discussion (cont'd)

- As an application, a cross-layer allocation of power and coding rate for video streaming was analyzed; and clear and specific answers obtained.
- The terminal should maximize perceptual quality per Joule (or per Watt). The optimal operating point can be identified in the graph of the composite function of two S-curves, one determined by the physical layer, the other by the human visual system.
- A decentralized allocation involving many terminals can be obtained as a "Nash-equilibrium" of a "game", but it is "inefficient".
- "Mechanism design" can lead to decentralized efficiency. A "compensation mechanism" from the economics literature is proposed.
- Full papers and additional slides about several applications can be obtained at http://pages.poly.edu/~vrodri01/research