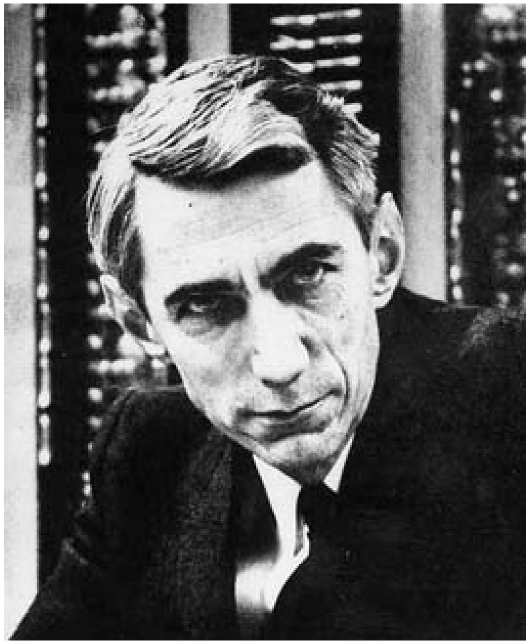
Information Theory:
Recent Advances and Future Challenges

Venkat Anantharam
EECS Department
University of California Berkeley

March. 27, 2006

Claude Elwood Shannon (1916 -2001)



- *A Mathematical Theory of Communication*, Bell System Technical Journal, 1948, called "The Magna Charta of the Communication Age" in an appreciation in the U.S. Congress on his death in 2001.

Entropy

-

$$X \sim (p_1, p_2, \ldots, p_M)$$

$$H(X) = -p_1 \log p_1 - p_2 \log p_2 - \ldots - p_M \log p_M$$

- Similarly:

$$H(X_1, \ldots, X_n)$$

Information sources

- For *syntactic purposes* each information source has an entropy rate
-

La musique souvent me prend comme une mer!
Vers ma pâle étoile,
Sous un plafond de brume ou dans un vaste éther,
Je mets à la voile;
⋮

-

$$H(\text{ Baudelaire }) = ??$$

Multiple sources

- These reveal information about each other.
-

$$H(Y \mid X, Z, W)$$

  or

$$H(X, Y \mid A, W) - H(X \mid A, W)$$

  etc.

- The mutual information is *symmetric*

$$I(X \wedge Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X)$$

Compression to the entropy rate

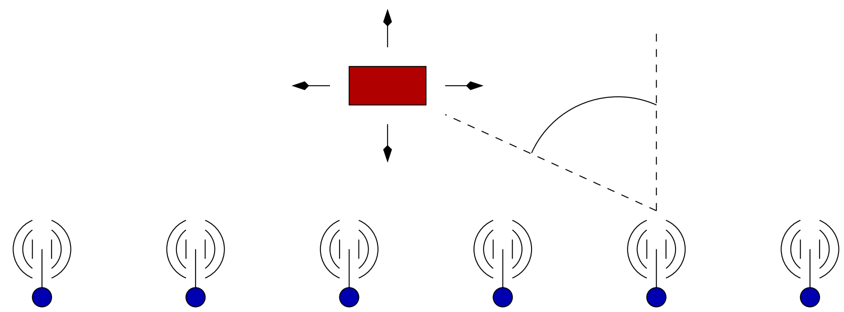
- Some popular techniques:
  - Huffman coding
  - Arithmetic coding
  - Lempel-Ziv (LZ '77, LZ '78, . . .)
  - Lempel-Ziv-Welch (LZW)
  - Context tree weighting
  - Burrows-Wheeler transform
  - . . .

  Several of these techniques are universal, i.e. they do not assume any prior knowledge of the source statistics.

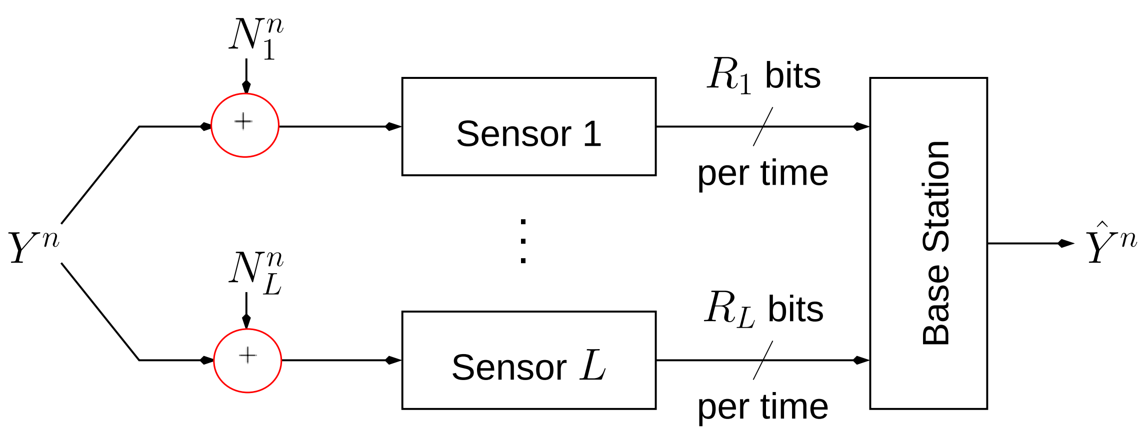
- Many of these are widely deployed in many products.

Sensor networks

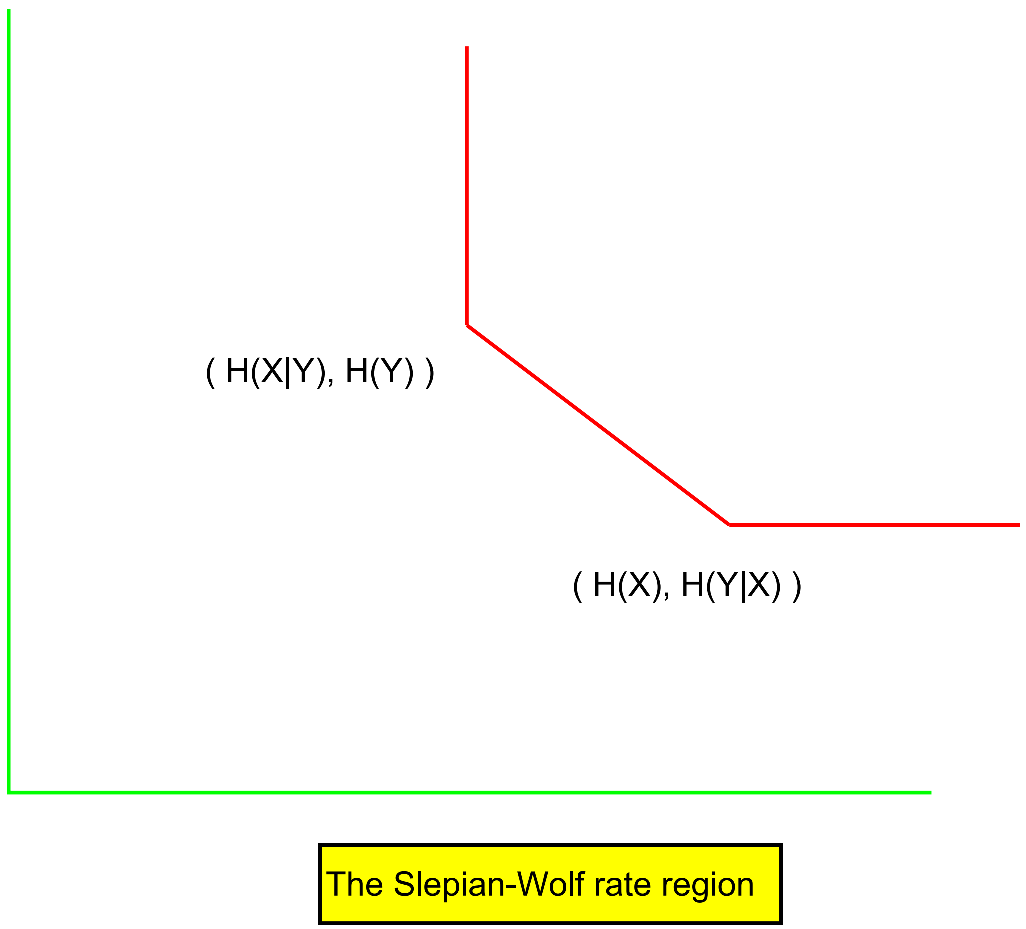
- A tracking problem:



- A schematic view:

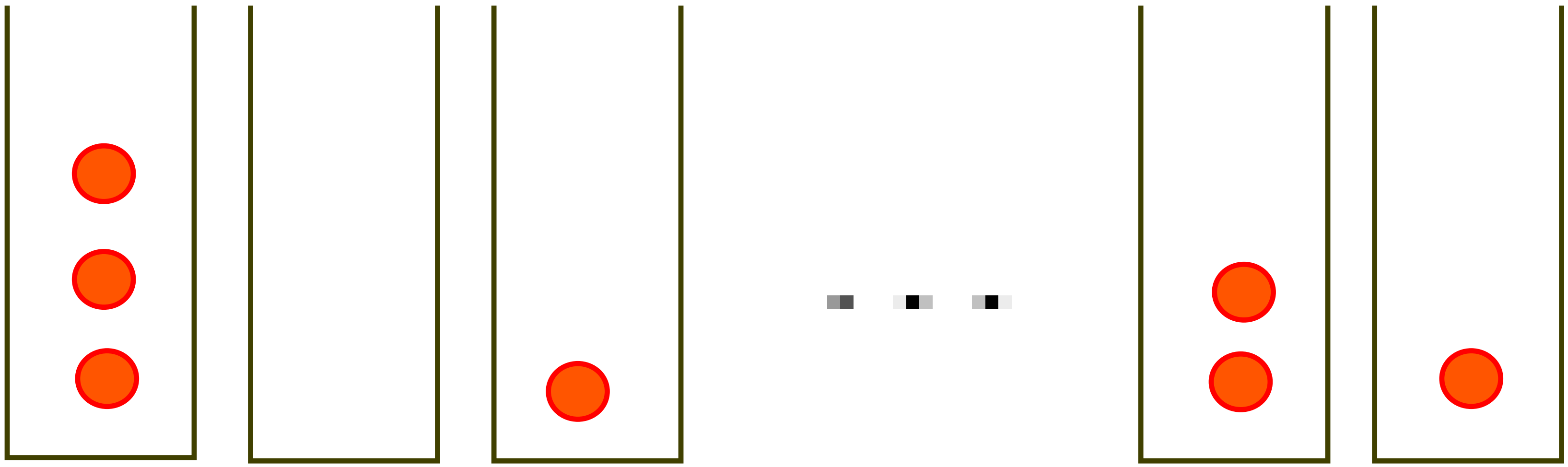


Distributed compression



- Compression is possible at a total rate of $H(X, Y)$ even though the sources are distributed.

The “direct” part of Information Theory



Illustrating the Slepian-Wolf ``binning'' strategy

- There are as many bins as there are typical Y-sequences.
- The typical X-sequences are “randomly” distributed among these bins.

Lossless versus lossy

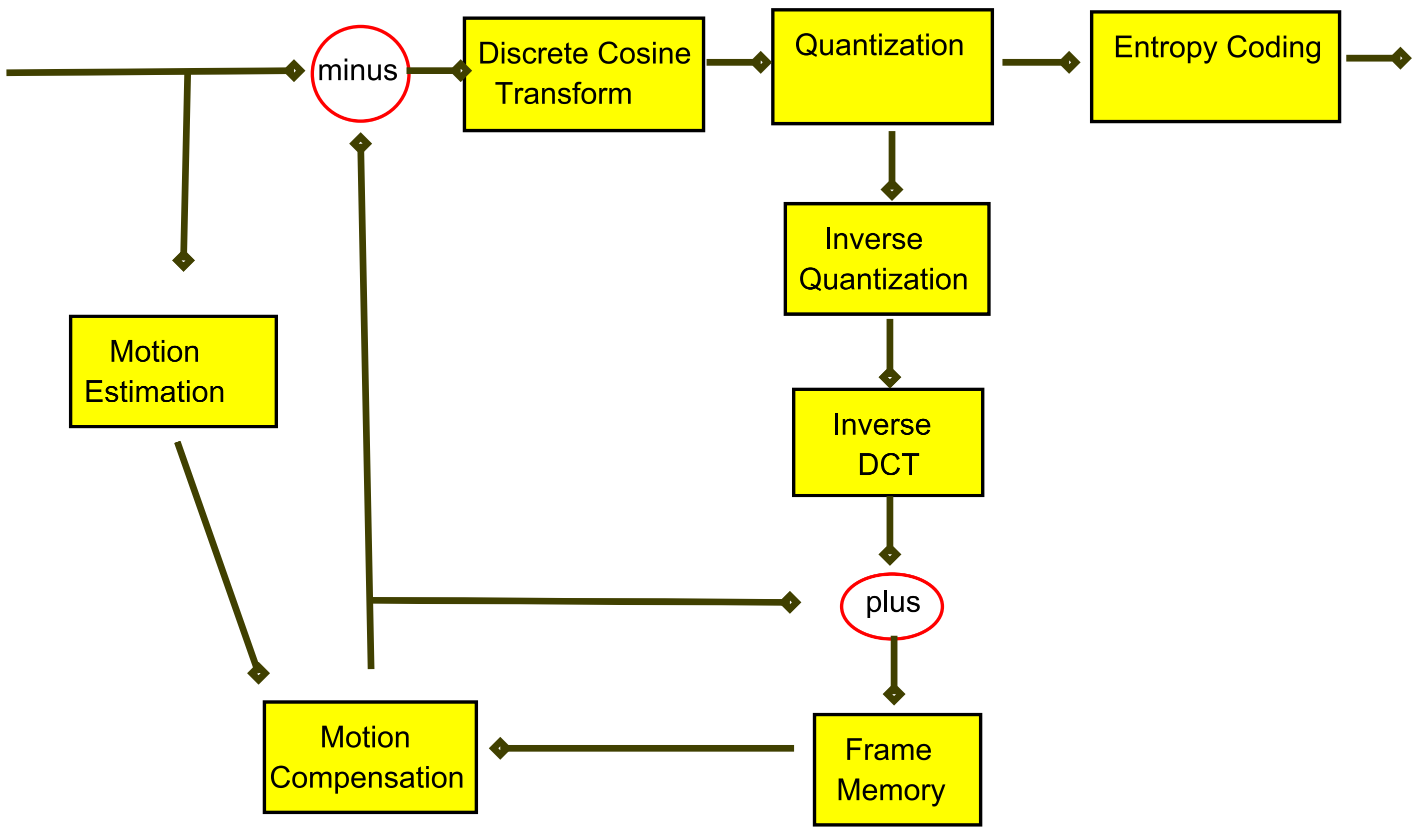
- For lossless compression the syntactic point of view is appropriate
  (Data, mission-critical information, Kolmogorov descriptions, . . .)

- For lossy compression, more intangibles enter the story: human factors engineering, empirical techniques, . . .
  (audio, video, imaging, multimedia, . . .)

Audio and Video compression standards

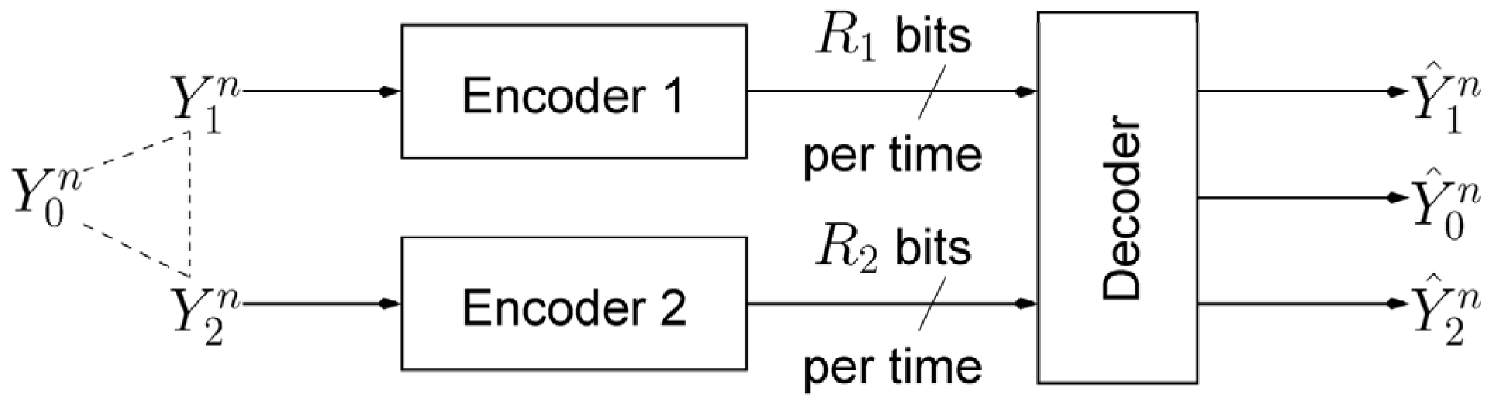
- Some of the many standards:
  - JPEG
  - JPEG 2000
  - MPEG
  - MPEG-2
  - H.261
  - MP3
  - . . .
- Many of these are widely deployed in many products.

MPEG-2



- Redundancy is removed at many stages:
  spatial redundancy (I-frames); redundancy across time (P-frames and B-frames);
  and the entropy coding.
- The standard has a “human factors” part and a “syntactical compression” part.

Lossy distributed source coding (1978)

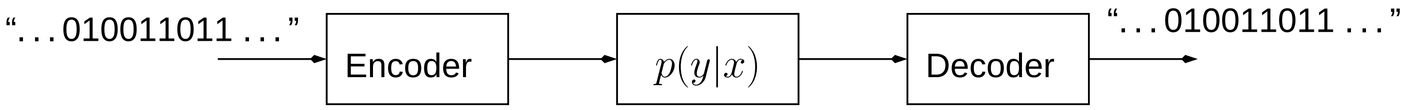


$$E\left[\frac{1}{n}\sum_{t=1}^{n} d_i(Y_i^n(t), \hat{Y}_i^n(t))\right] \leq D_i \text{ for } i = 0, 1, 2.$$

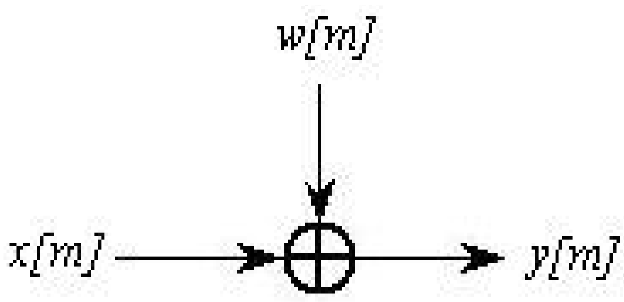
- Find the set of achievable (R_1, R_2) for given (D_0, D_1, D_2).
- The rate region of the quadratic Gaussian two-terminal source-coding problem
  Aaron B. Wagner, Saurabha Tavildar, and Pramod Viswanath. Preprint, arxiv:cs.IT 2005

The channel coding problem

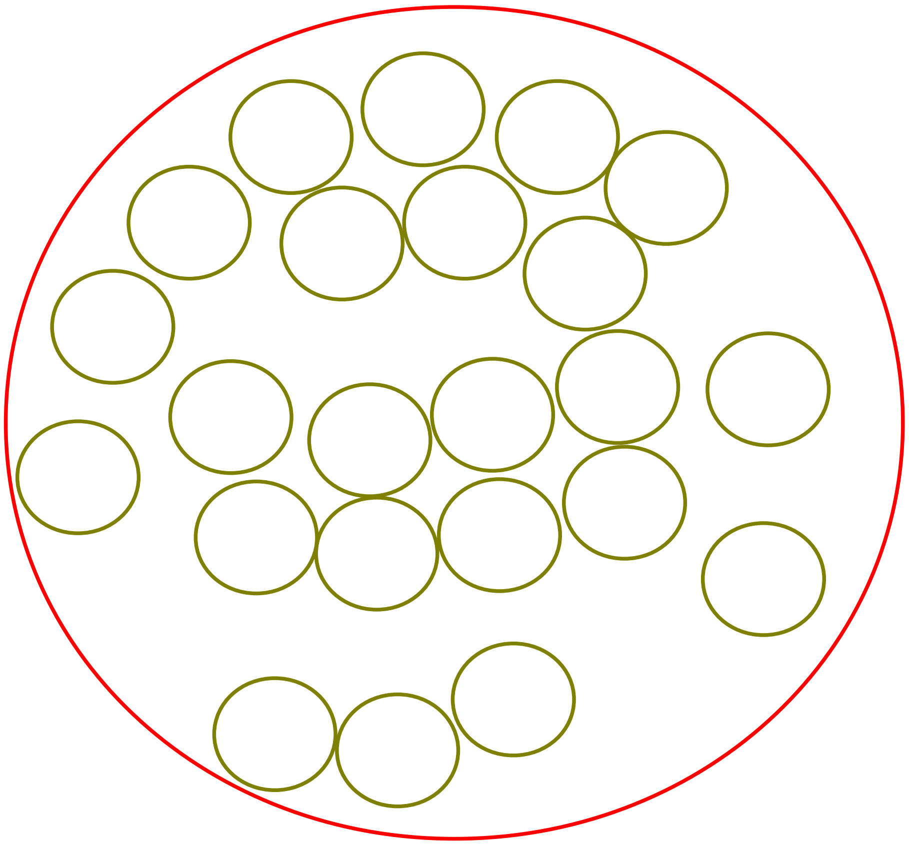
- The view at the level of symbols:



- The analog view:

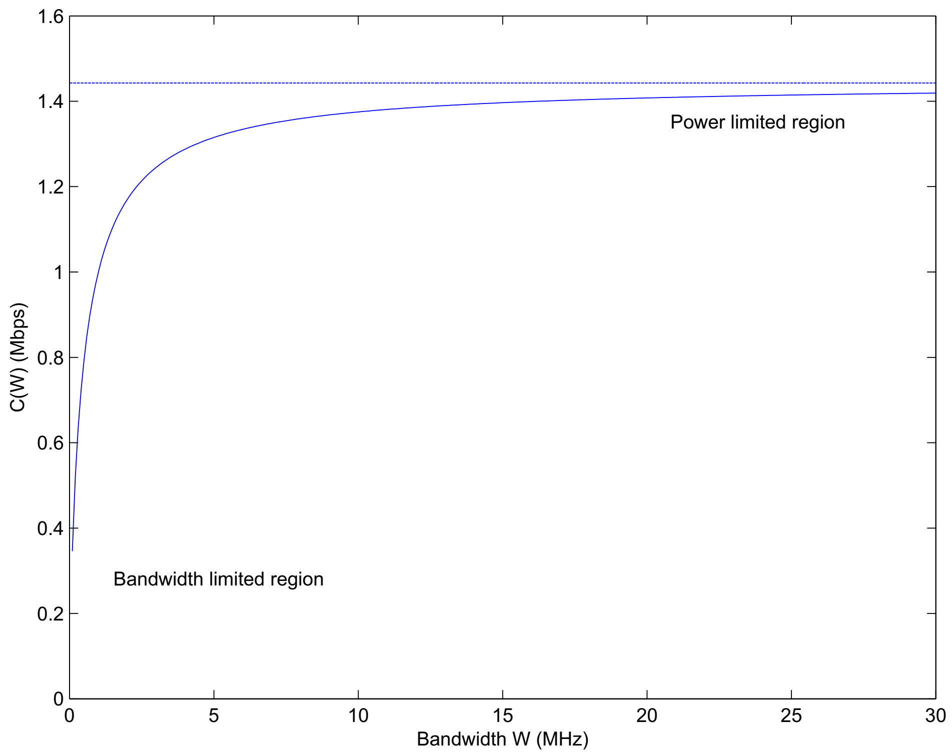


Coding as bin packing



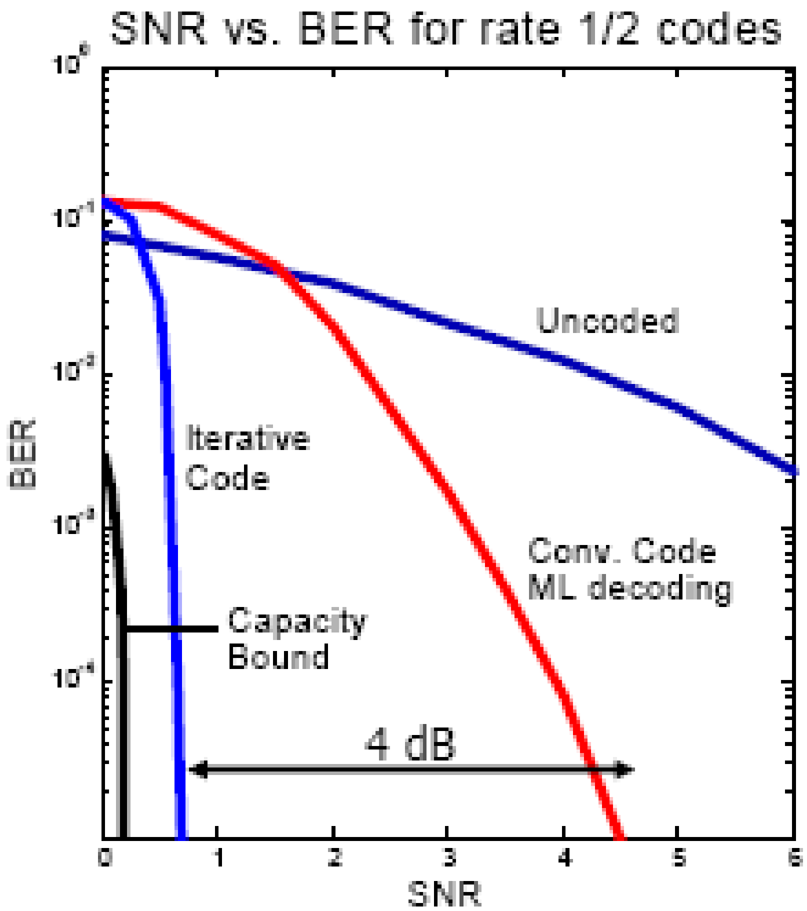
Illustrating the problem of coding for the AWGN channel

The capacity of the AWGN channel



- $C(W) = W \log(1 + \frac{P}{N_0 W})$ plotted for $\frac{P}{N_0} = 10^6$.
- Even at infinite bandwidth one can only transmit at finite rate $\frac{P}{N_0} \log_2 e$.

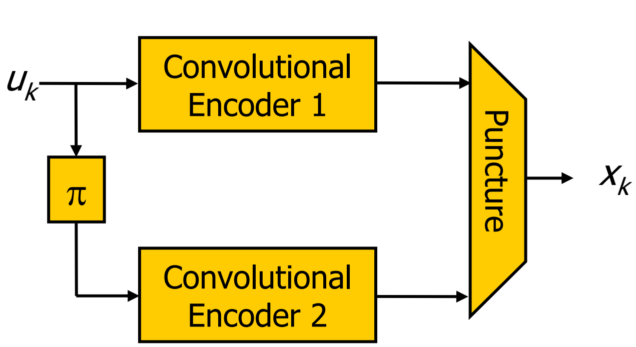
The “direct” part of achieving capacity



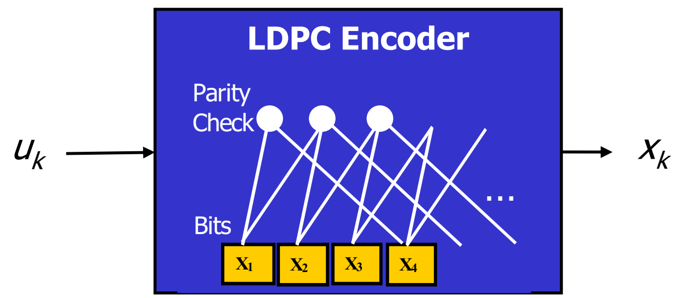
Year	Rate ½ Code	SNR Required for BER < 10^{-5}
1948	SHANNON	0dB
1967	(255,123) BCH	5.4dB
1977	Convolutional Code	4.5dB
1993	Iterative Turbo Code	0.7dB
2001	Iterative LDPC Code	0.0245dB

- The values in the table are relative to the Shannon limit at rate $\frac{1}{2}$

Turbo and LDPC



Concatenated convolutional schemes (Turbo convolutional)



Low Density Parity Check (LDPC) codes

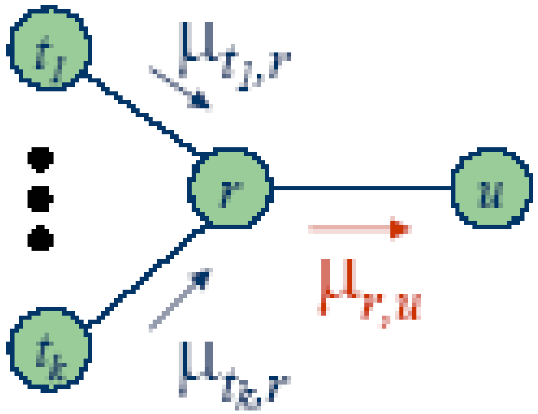
- Near Shannon limit error correcting codes: Turbo codes

  C. Berrou, A. Glavieux, and P. Thitimajshima IEEE-ICC 1993.

- Low density parity check codes R. G. Gallager M.I.T. Press 1963

Message passing algorithms

- Define 'messages' $\mu_{r,u}(x_{u\cap r})$ for each edge (r, u) of the independency graph. Initialize to 1.



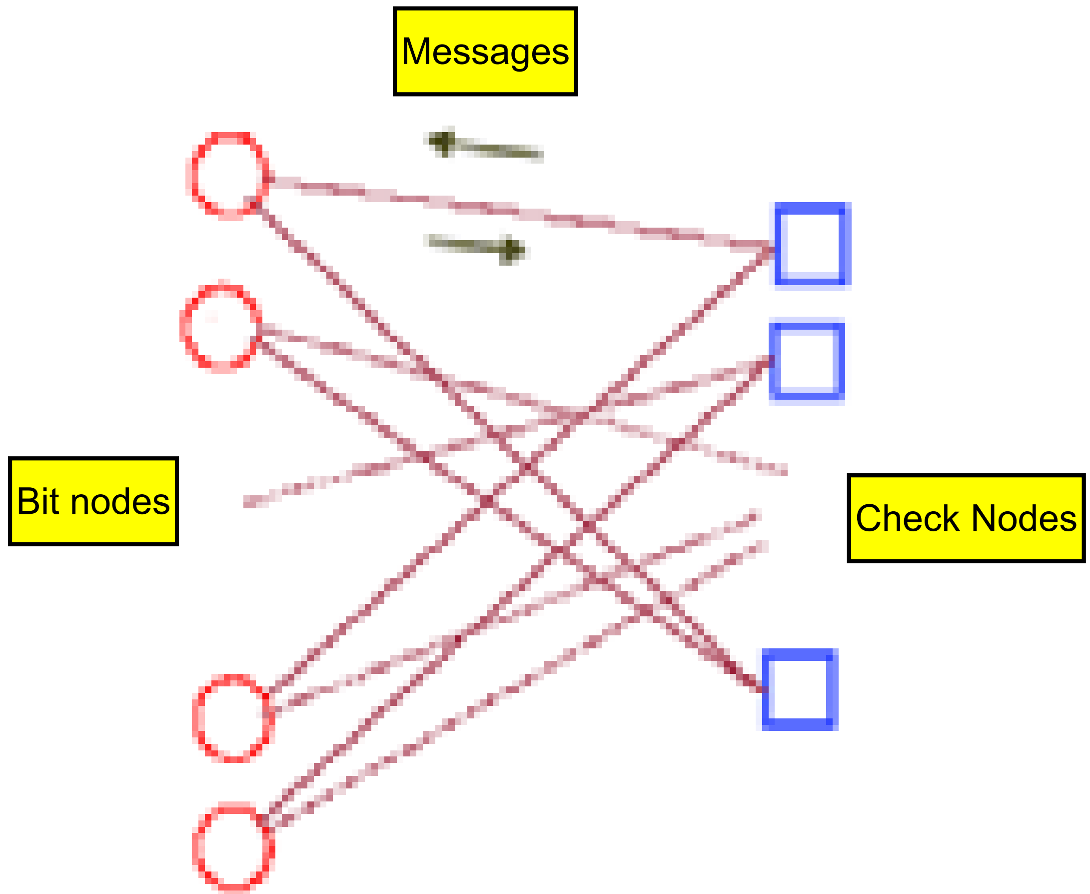
- Update messages as:

$$\mu_{r,u}(x_{u\cap r}) \equiv \sum_{x_r \setminus x_u} \alpha_r(x_r) \prod_{t \in N(r)\setminus\{u\}} \mu_{t,r}(x_{r\cap t})$$

- Define 'Beliefs' $$b_r(x_r) \equiv \alpha_r(x_r) \prod_{t \in N(r)} \mu_{t,r}(x_{r\cap t})$$

Message passing algorithms for LDPC Codes

- What the message passing algorithm looks like:

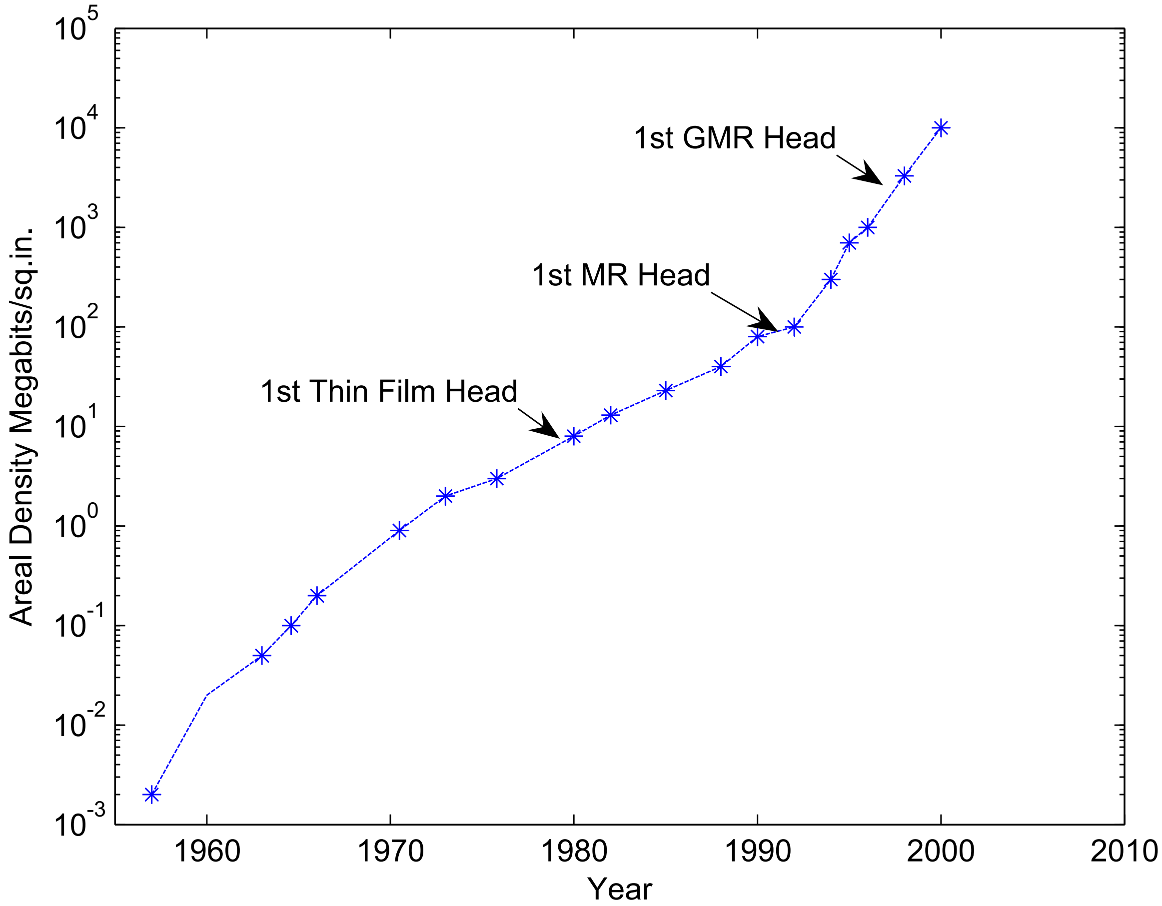


And now a word from our sponsors ...

Information theory has played a big role in some of the key technological trends of recent decades.

- Data compression and multimedia compression as already discussed (compact disks; DVDs; Ipods;)
- The growth in rates of information access (modem standards; DSL; Gigabit Ethernet over copper; . . .)
- The super-Moore's law improvements in magnetic recording.
- The exploration of deep space.
- The explosive growth of cellular wireless communication.

Areal density in magnetic recording



- Run length limited codes.
- Partial response maximum likelihood signalling.
- Media noise.

Space: the final frontier

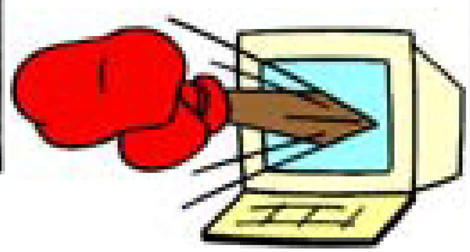
Mission Name	Year	Compression	Coding	Information Rate
Mariner 4	1965	None	None	8.33 bps
Viking	1976	None	Biorthogonal code	3 Kbps
Mars Global Surveyor	1997	2:1 lossless	Conv. + RS Conc. code	128 Kbps
Mars Rover	2004	12:1 lossy	Conv. + RS Conc. code	168 Kbps
Mars Reconn. Orbiter	2006	2:1 lossy	Turbo code	12 Mbps

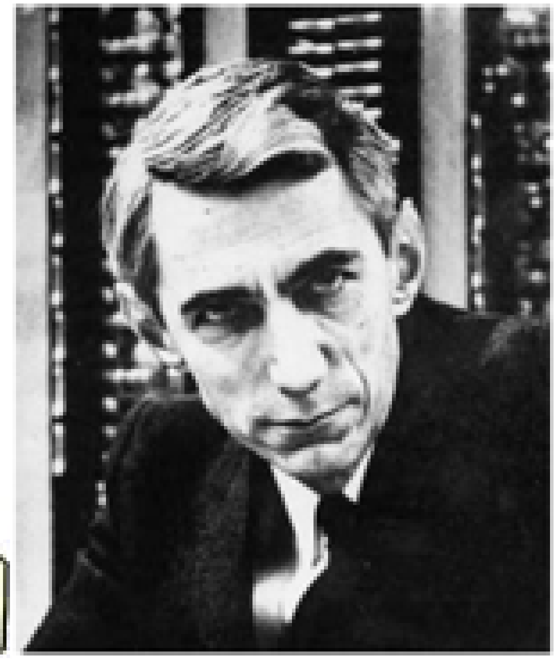
- For more information see the Shannon lecture of Robert J. McEliece: http://www.systems.caltech.edu/EE/Faculty/rjm/papers/ShannonLecture.pdf

The star cluster NGC 346

- Courtesy of NASA, STScI, and the Hubble Space Telescope.

Newton vs. Shannon

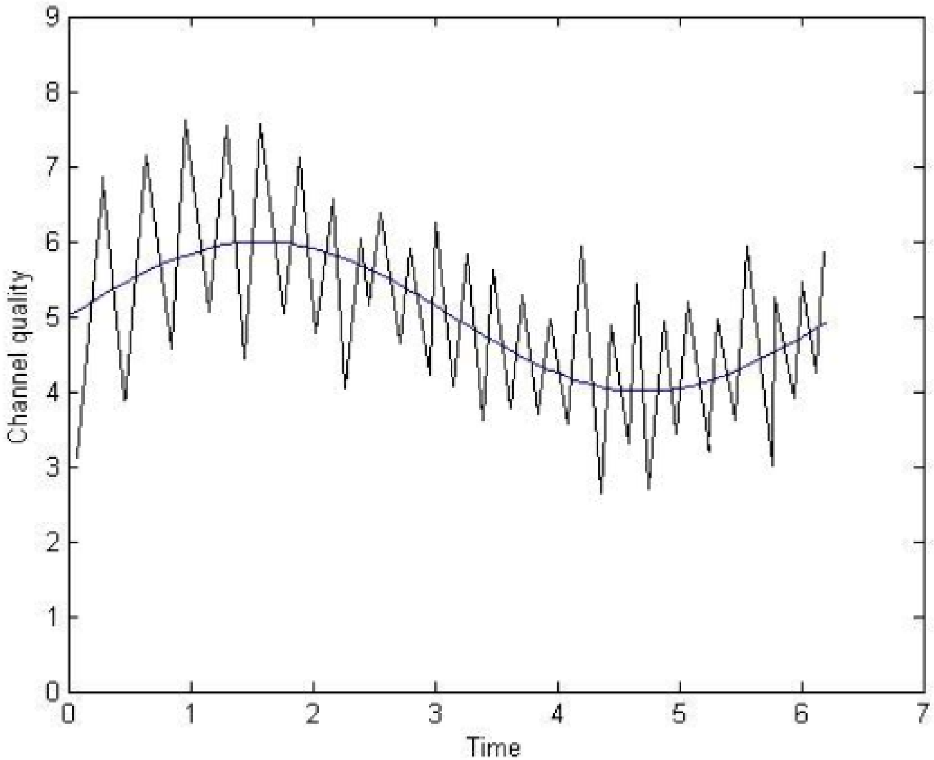




- McEliece attributes 21 % of the increase in data rate to Shannon (source and channel coding) and 79 % of the increase to Newton (antenna aperture, transmission frequency, power)

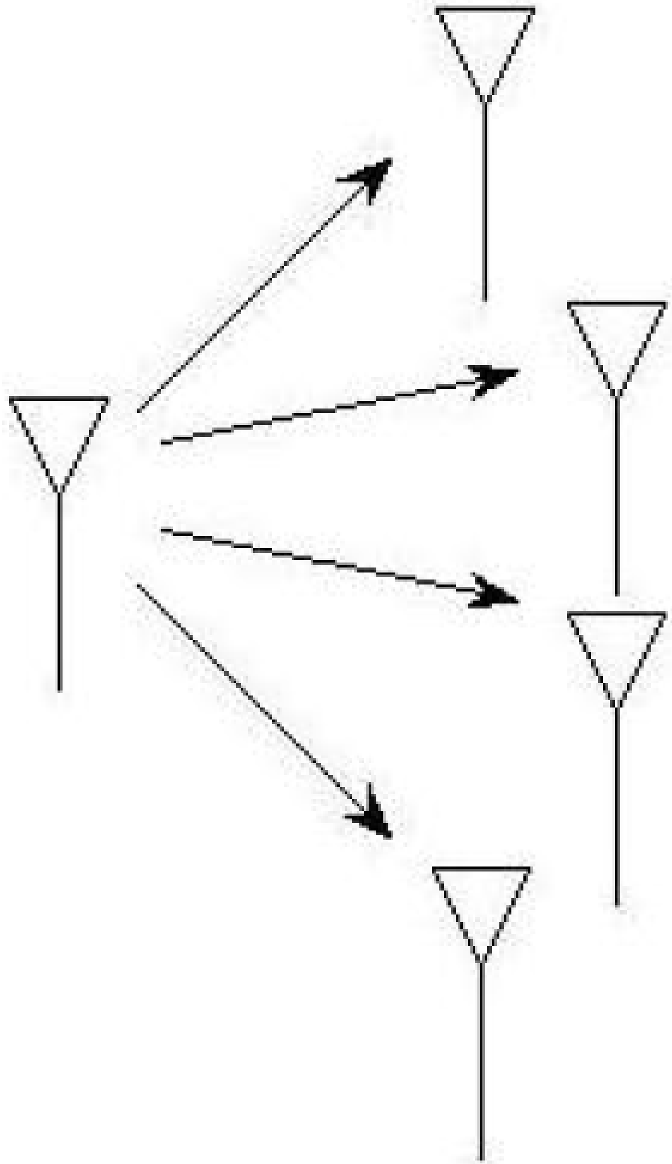
Fading channels

$$y[m] = h[m]x[m] + w[m]$$



- Fading can be slow or fast relative to the delay requirement.
- In a fast fading channel a symbol to be transmitted can be interleaved across multiple channel states.
- In a slow fading channel a symbol to be transmitted sees a different environment in different fading states.

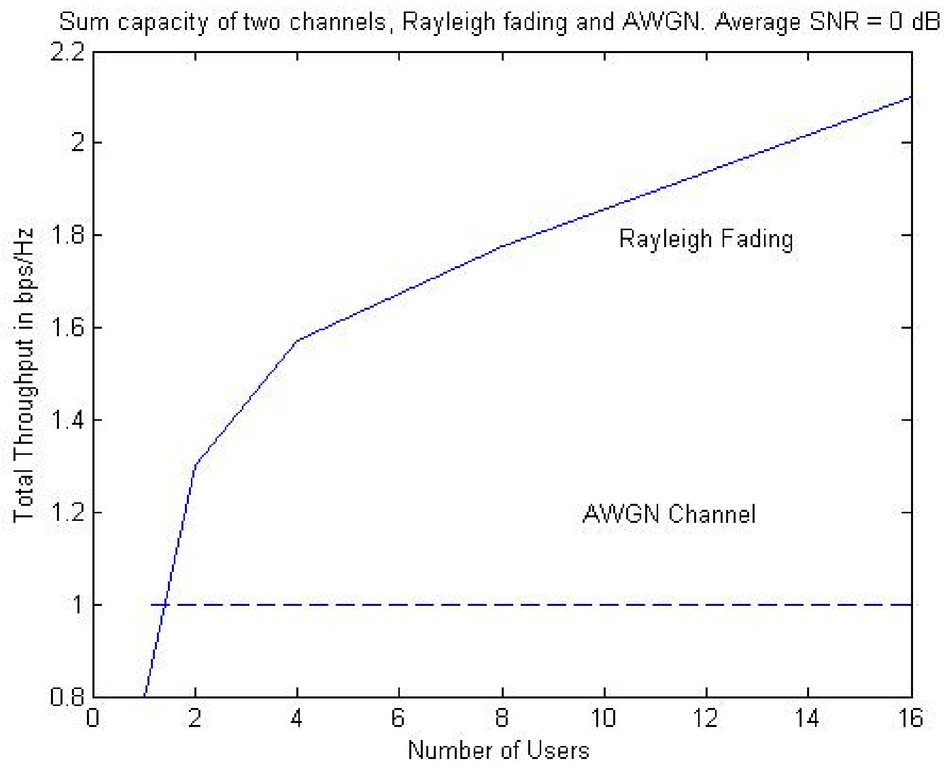
The wireless downlink



Broadcast Channel

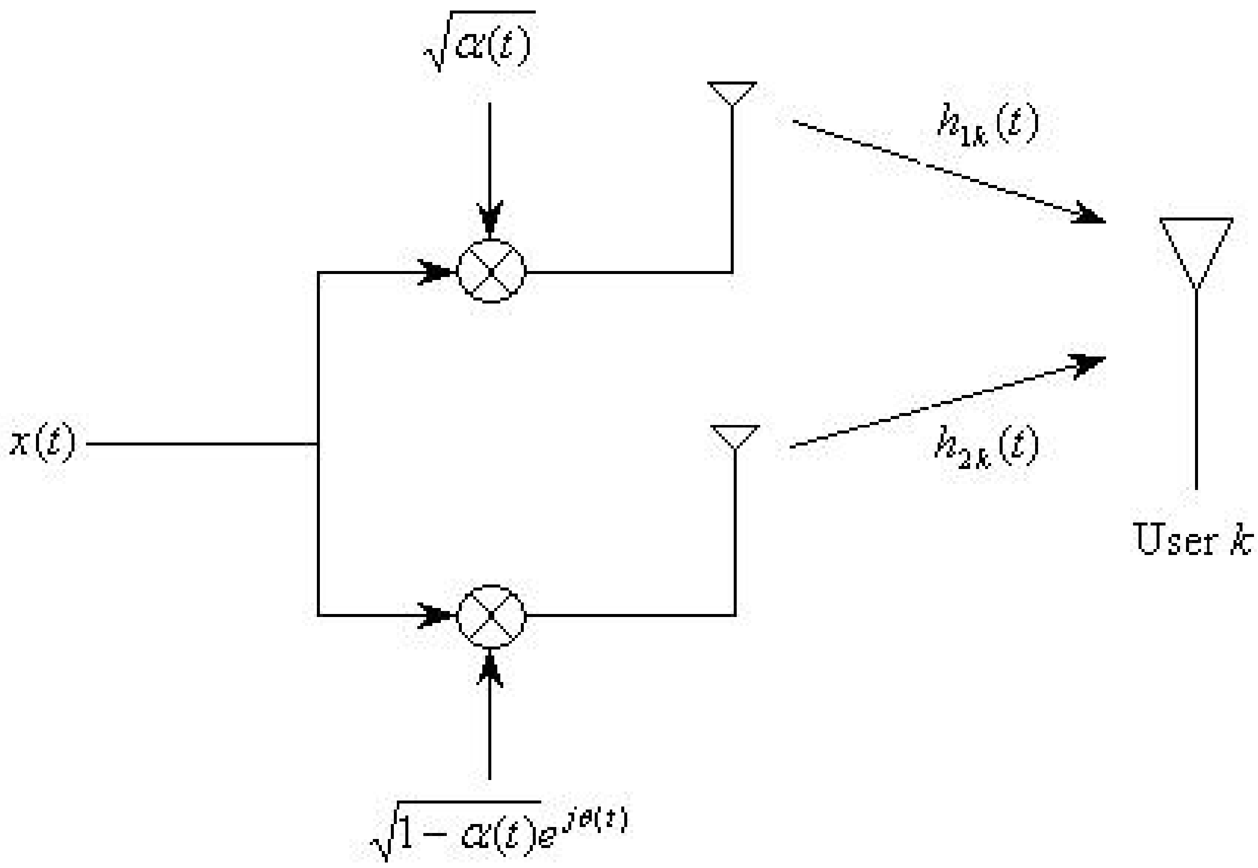
Multiuser diversity

- By scheduling to the strong users one has multiuser diversity.



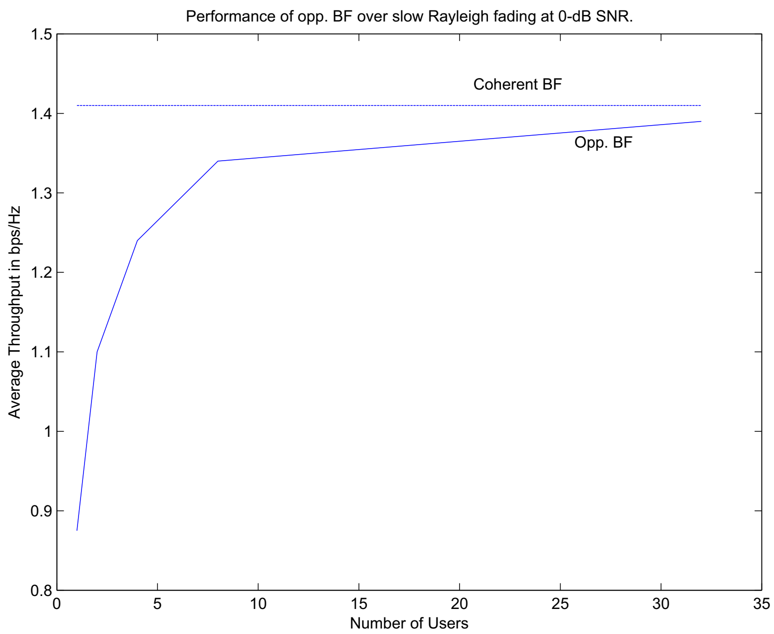
- Assumes each user feeds back the SNR of its channel to the base station.

Opportunistic beamforming



The same signal is transmitted over the two antennas with time-varying phase and powers.

Perfomance of opportunistic beamforming



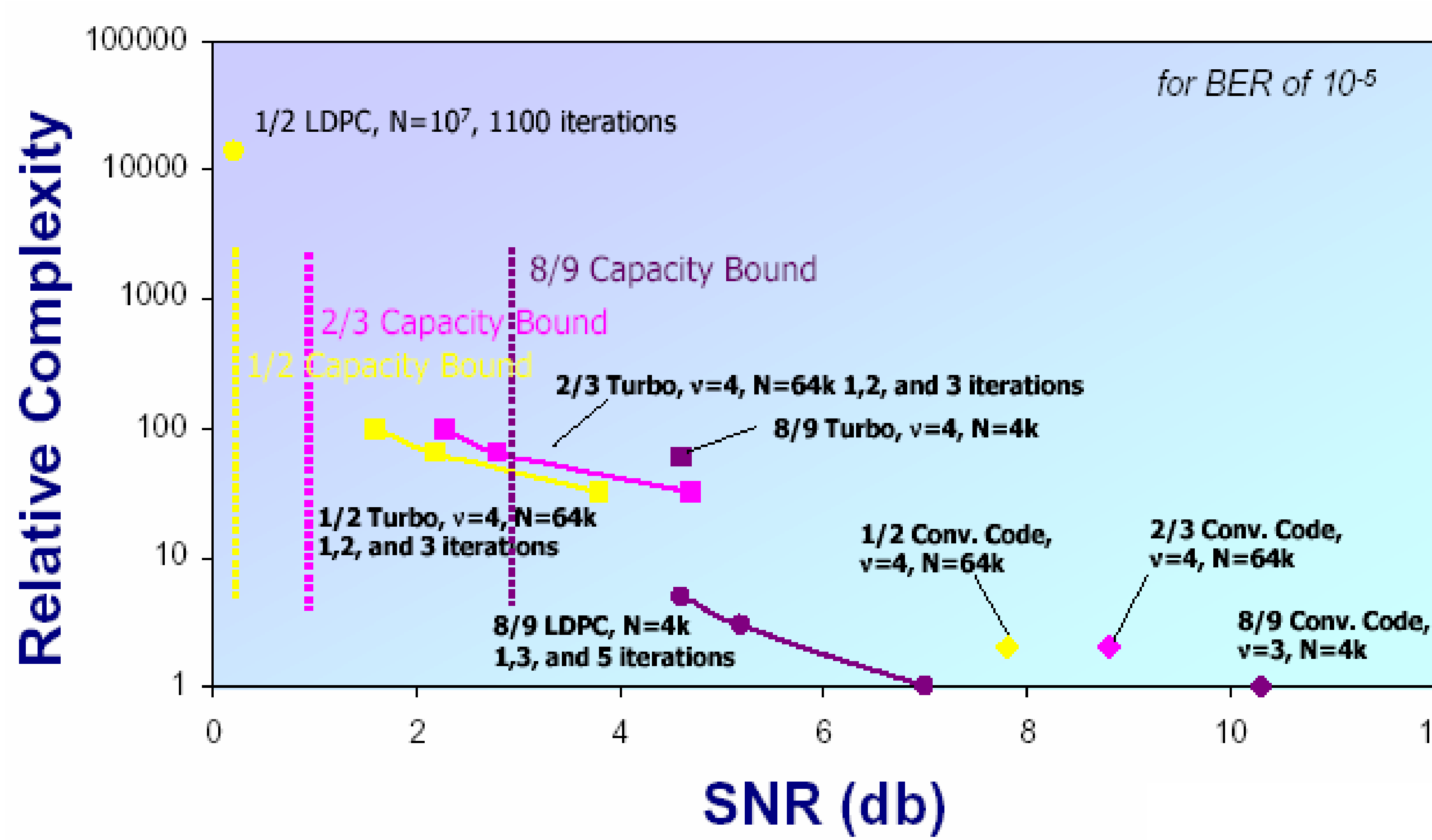
- Opportunistic beamforming using dumb antennas Pramod Viswanath, David N. C. Tse, and Rajiv Laroia IEEE-IT 2002.

- The system requirements should be contrasted with those needed for space-time codes.

Allons vers l'avenir!

- Computational complexity is still an issue.
- Coding in the deep bit error regime.
- Unreliability in the deep submicron regime.
- Core problems in multiuser information theory.
- Incentive issues with multiple players.
- Spatial information theory
- Revisiting information-theoretic security.
- Real-time information theory.
- Quantum information theory.
- Information theory and cognition.
- . . .

Computational complexity of decoders



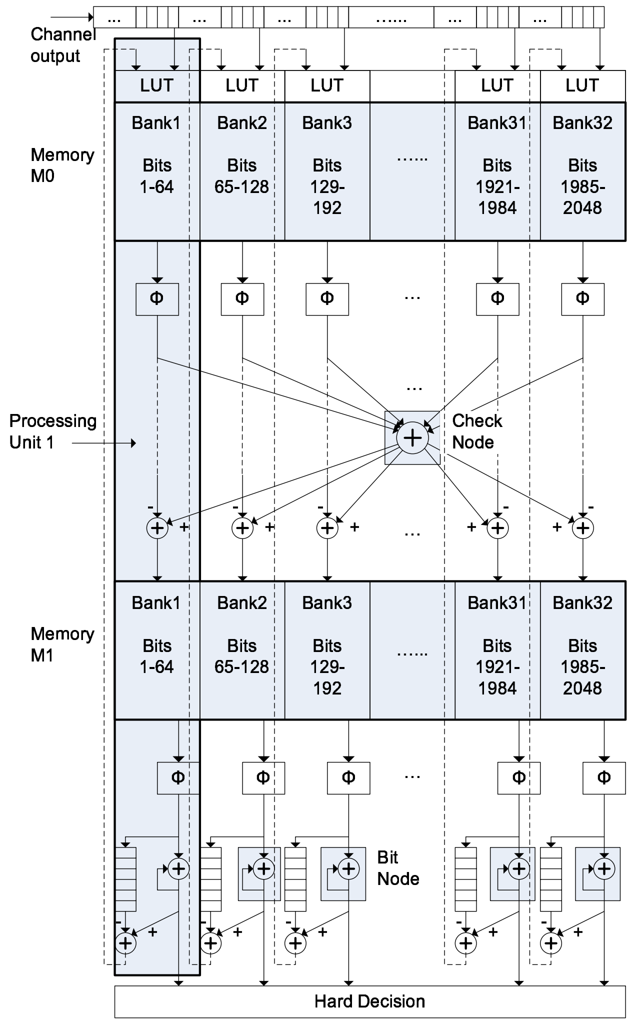
- Picture courtesy of Engling Yeo.

Deep BER performance of error control codes

- For some important applications it pays to focus on very stringent bit error rate requirements
  (magnetic recording, transcontinental fiber optic communication)

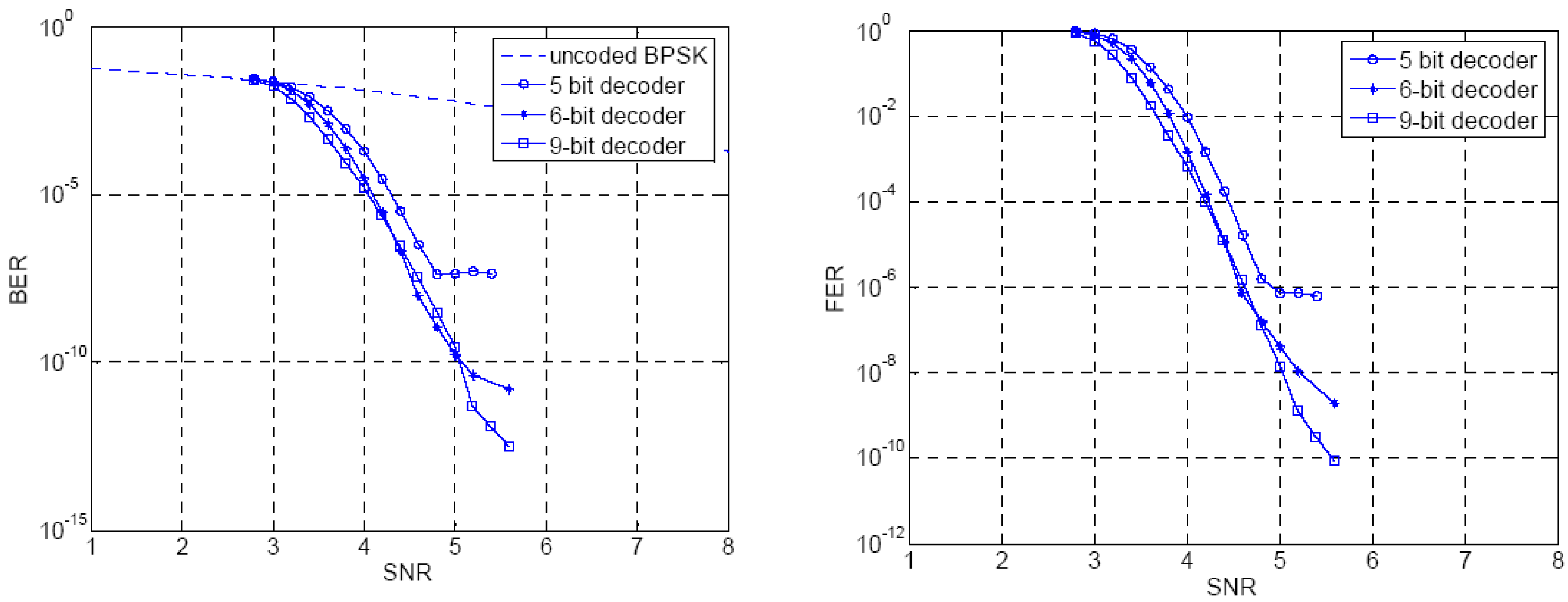
- Even the best known codes have an error floor in the deep BER regime.

(2048,1723) RS-LDPC decoder on an FPGA platform.



$$\Phi(x) = -\log\left(\tanh(\tfrac{x}{2})\right),\ x \geq 0.$$

Statistics from deep BER emulation

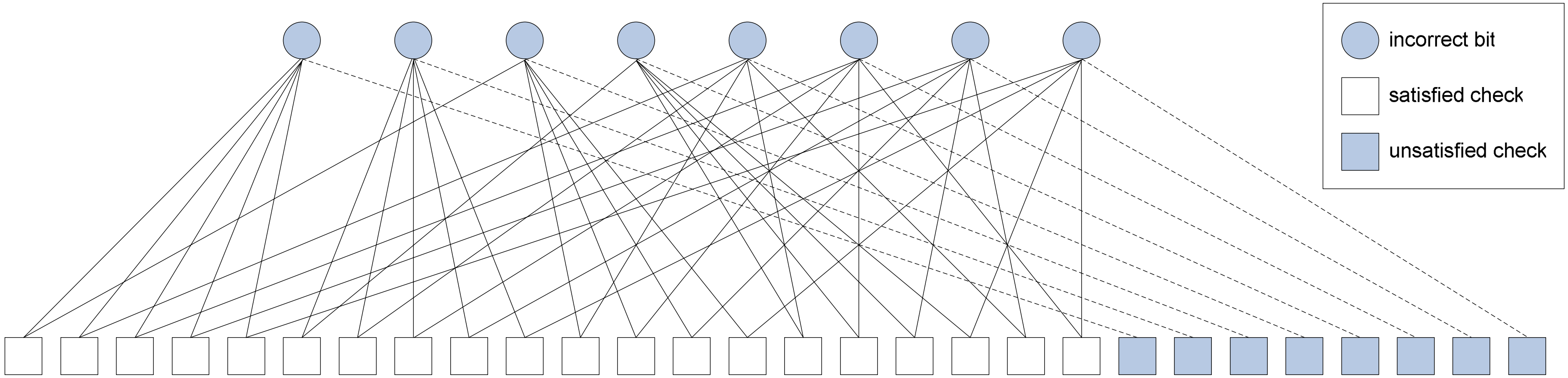


- Investigation of Error Floors of Structured LDPC Codes by Hardware Emulation

  Zhengya Zhang, Lara Dolecek, Borivoje Nikolic, VA, and Martin Wainwright

  Preprint 2006

Absorbing sets



- The emulation reveals specific non-codeword patterns of 8 bit nodes and 28 check nodes that absorb the decoding iteration.
- Eliminating these systematically should improve the deep BER performance.

Information theory in the deep submicron regime

The challenge: using information theory to reliably move information around the chip in a low power high interference environment

- A quick history of integrated circuits:

Decade	Technology	Line width
1940's	Invention of the Transistor	
1950's	Invention of the Integrated Circuit	
1960's	Small/Medium scale integration (SSI/MSI)	
1970's	Large scale integration(LSI)	10 microns
1980's	VLSI	2 microns
1990's -now	CMOS	1 micron -100 nanometers

- The deep submicron regime starts at 0.35 micron line widths
- Fabrication at 0.13 micron line widths is already considered routine
- Supply voltages are dropping because of power constraints

(from roughly 3.3V in 1995 to roughly 1V today)

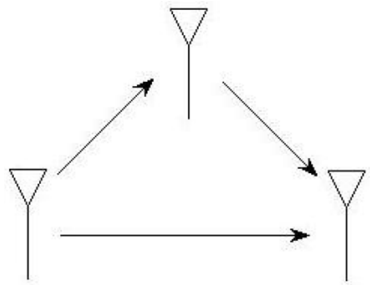
- Line widths are already pushing past 90 nanometers.

Design at the end of the Silicon Roadmap Jan M. Rabaey
Keynote address, Design Automation Conference 2005

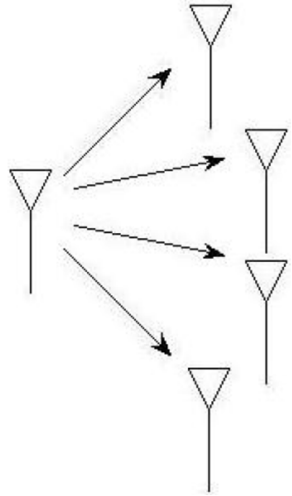
Coding for System-on-Chip Networks: A Unified Framework
S. R. Sridhara and N. R. Shanbhag IEEE-VLSI 2005

Open problems in multiuser information theory

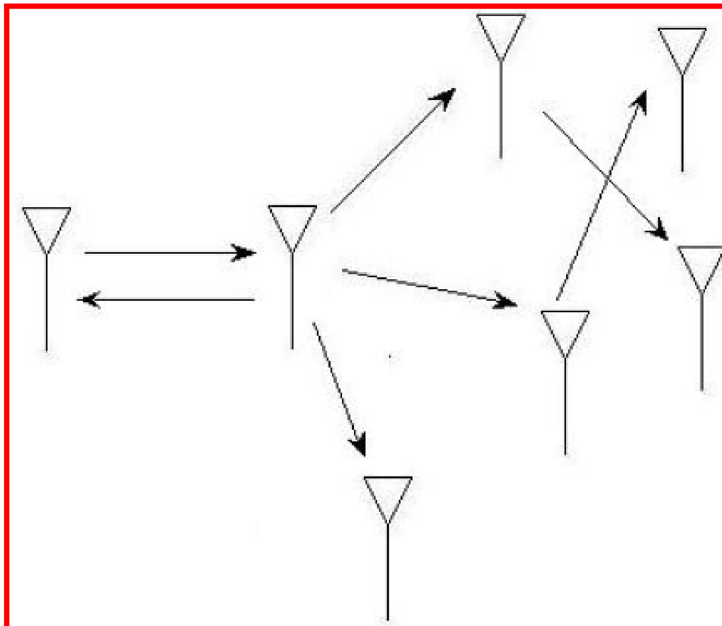
- Most of the core problems of multiuser information theory are still open



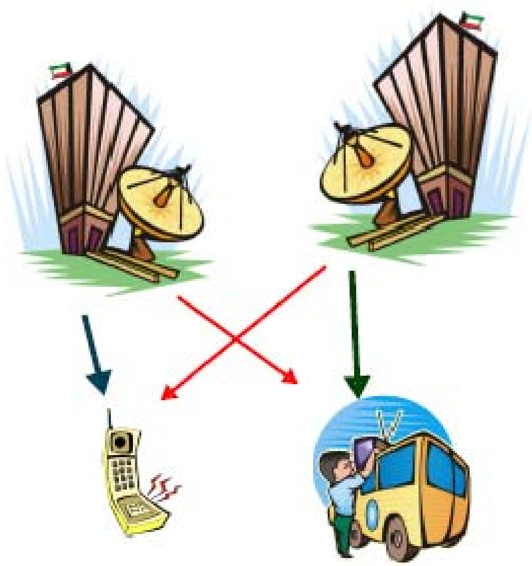
The relay channel



The broadcast channel



The capacity region of networks

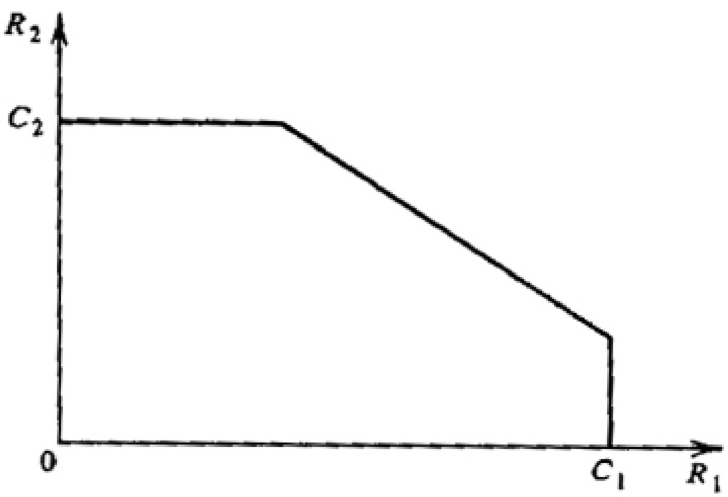


The interference channel

Incentive issues in Information Theory

- The rules of communication over the shared medium should be rational
- Non-cooperative (self-centered) agents: Nash equilibrium strategies
- Cooperative (coalition-forming) agents: social choice issues
- Example:

Assume $P_1 \geq P_2 \geq \ldots \geq P_M$



Gaussian multiple-access capacity region

$$\Phi_i = \frac{1}{i}\left[C(iP_i + \Sigma_{j=i+1}^{M} P_j, \sigma^2) - \Sigma_{j=i+1}^{M} \Phi_j\right]$$

The unique envy-free allocation of greedy users in a slow fading wireless uplink

Here $C(P, \sigma^2) = \frac{1}{2}\log(1 + \frac{P}{\sigma^2})$

A Game-theoretic look at the Gaussian Multiple-access Channel
Richard J. La and VA DIMACS 2003

Spatial Information Theory: Multiple Spatial Phases

Communication from a node to another in an ad-hoc network requires high enough signal-to-interference-and-noise ratio

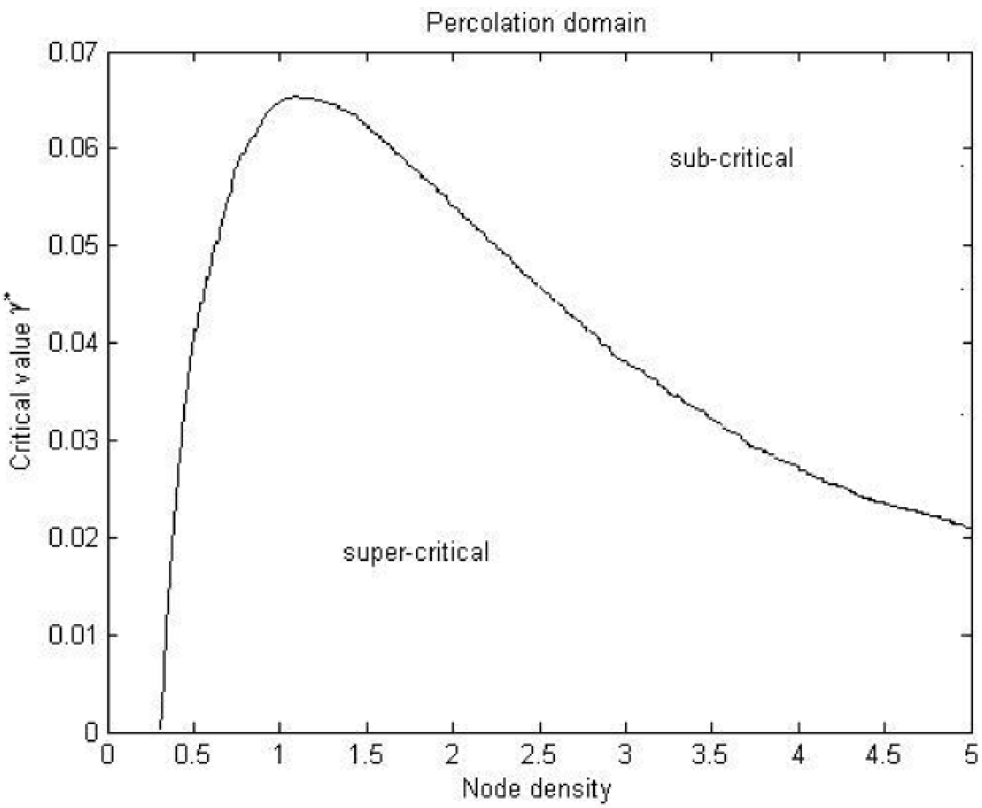
Fix β

As γ increases past a threshold many physically important quantities (connectivity, etc.) undergo a discontinuous transition

Related to phenomena of statistical physics

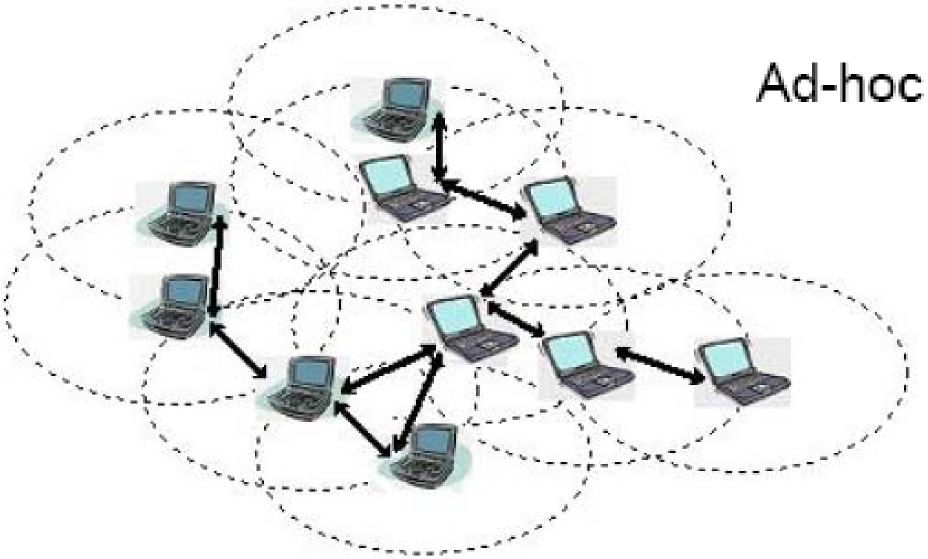
Node i can transmit to node j iff

$$\frac{P_i L(x_j - x_i)}{N_0 + \gamma \sum_{k \neq i,j} P_k L(x_j - x_k)} \geq \beta$$



Impact of Interferences on Connectivity in Ad-hoc Networks
Olivier Dousse, François Baccelli, and Patrick Thiran
IEEE/ACM-Networking 2005

Spatial Information Theory: Spatial capacity notions



Ad-hoc Network

A bit may not be a bit in a spatially extended network: bits that are moved further may be worth more

Transport capacity: a distance-weighted sum of rates

For several fading models the transport capacity of a network of n nodes is $\Theta(n)$

Important problems:

- Broader range of fading scenarios?
- Delay?
- Constants in the scaling?
- Other notions of spatial capacity?
- Incentive issues?

The study of such spatially extended capacity notions is in its infancy

The transport capacity of wireless networks over fading channels

F. Xue, L.-L. Xie, and P. R. Kumar IEEE-IT 2005

Communicating a secret

Assume $X \amalg (U, V, W)$

Require $X \amalg (T, W)$

Eve
has W

Alice
has U
wants to send X

sends $T = F(X, U)$

Bob
has V
recovers $X = G(T, V)$

Shannon's negative result

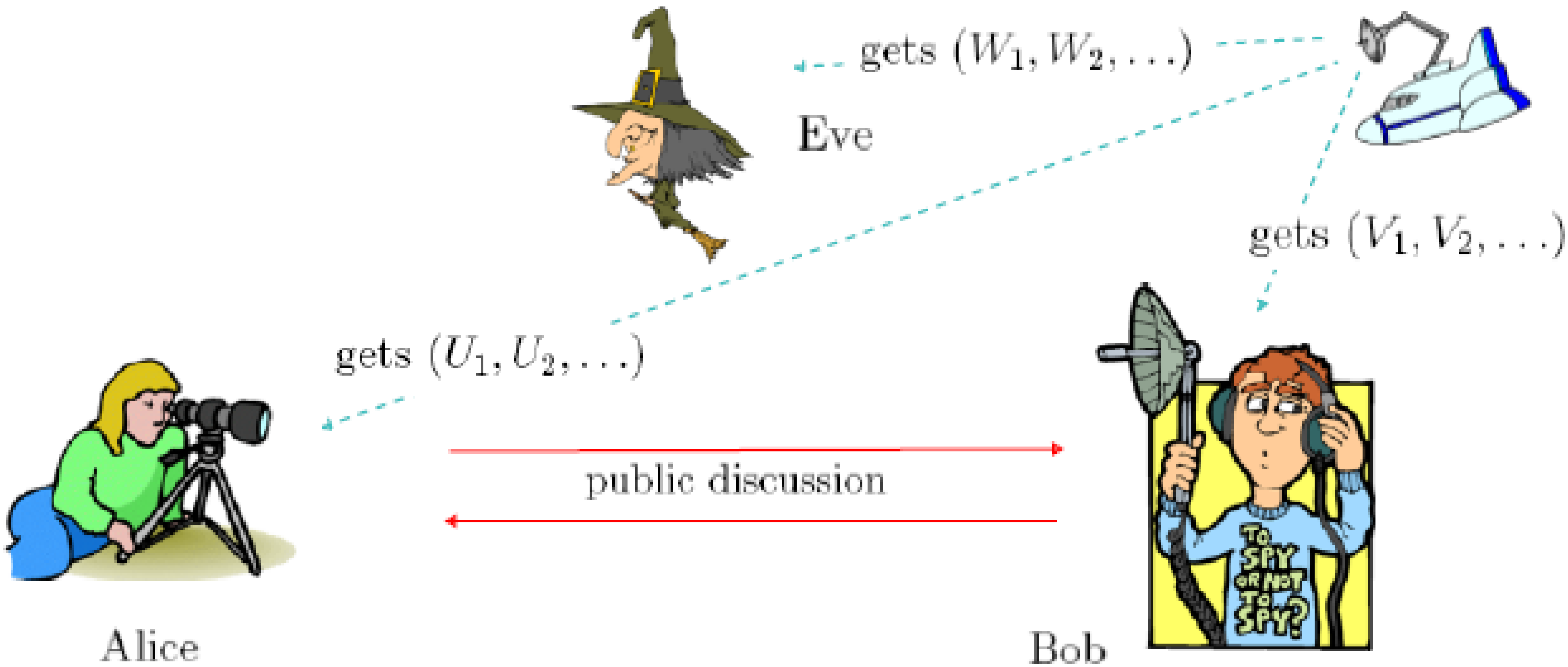
Shannon tells us :

There must exist a random variable K such that

- $K = g_A(U)$
- $K = g_B(V)$
- $K \amalg W$
- $H(K) \geq H(X)$

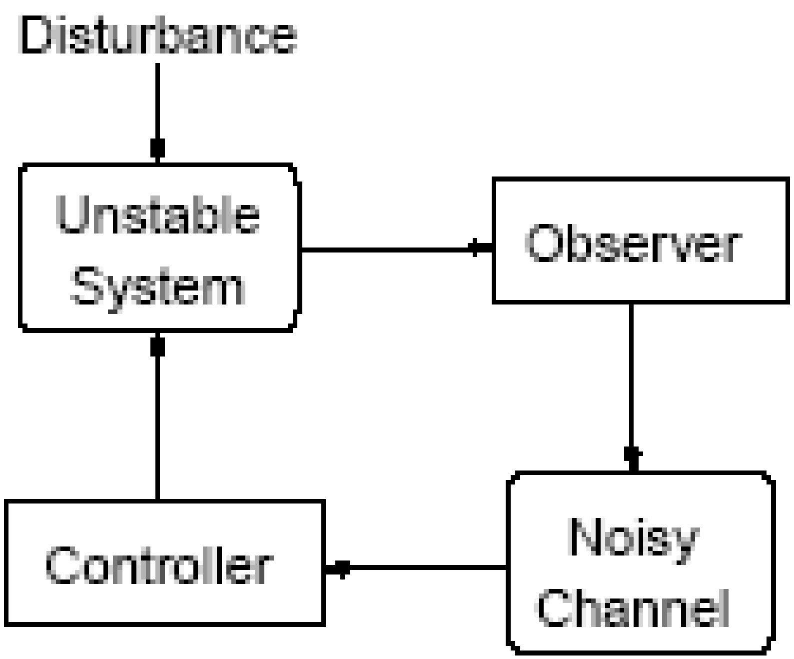
Apparently Alice and Bob must already have a big enough one-time pad.

Is information theoretic security dead?



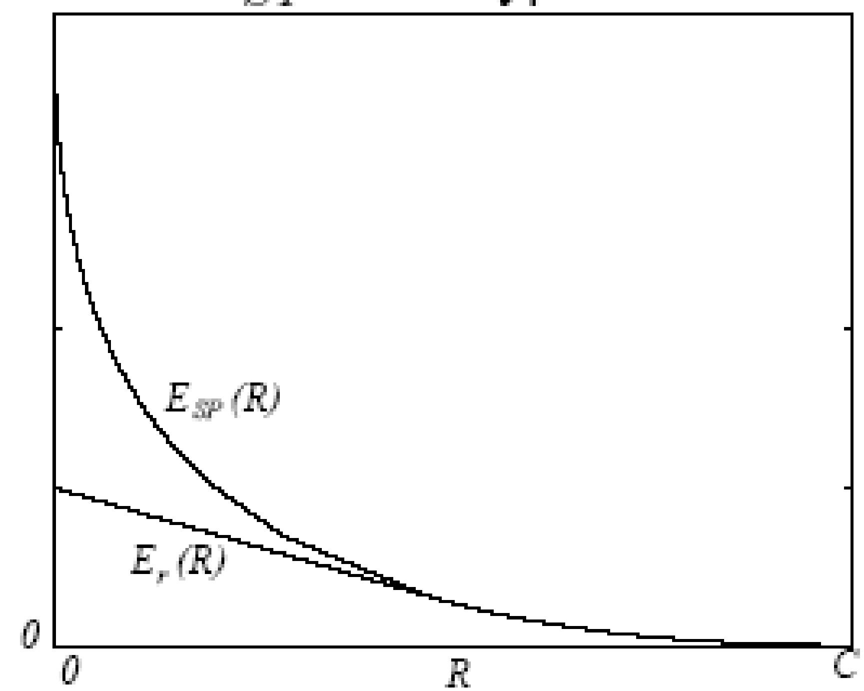
Secret key agreement by public discussion from common information U. M. Maurer IEEE-IT 1993

Real time information theory



The communication should be accurate enough to avoid the growth of error under the dynamics

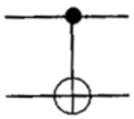
E_r and E_{SP} for a Typical Channel



Related to the theory of error exponents

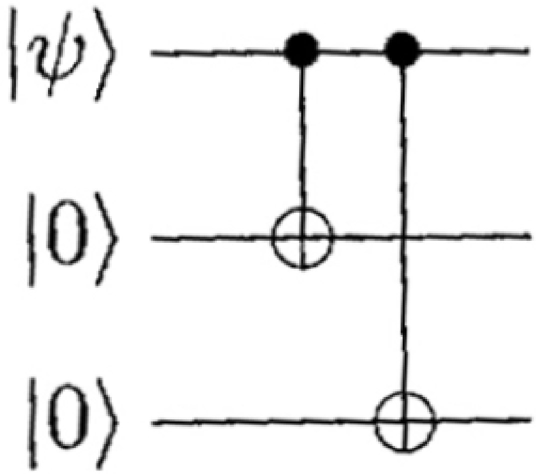
The necessity and sufficiency of anytime capacity for control over a noisy communication link,
Anant Sahai IEEE CDC 2004

Quantum information theory

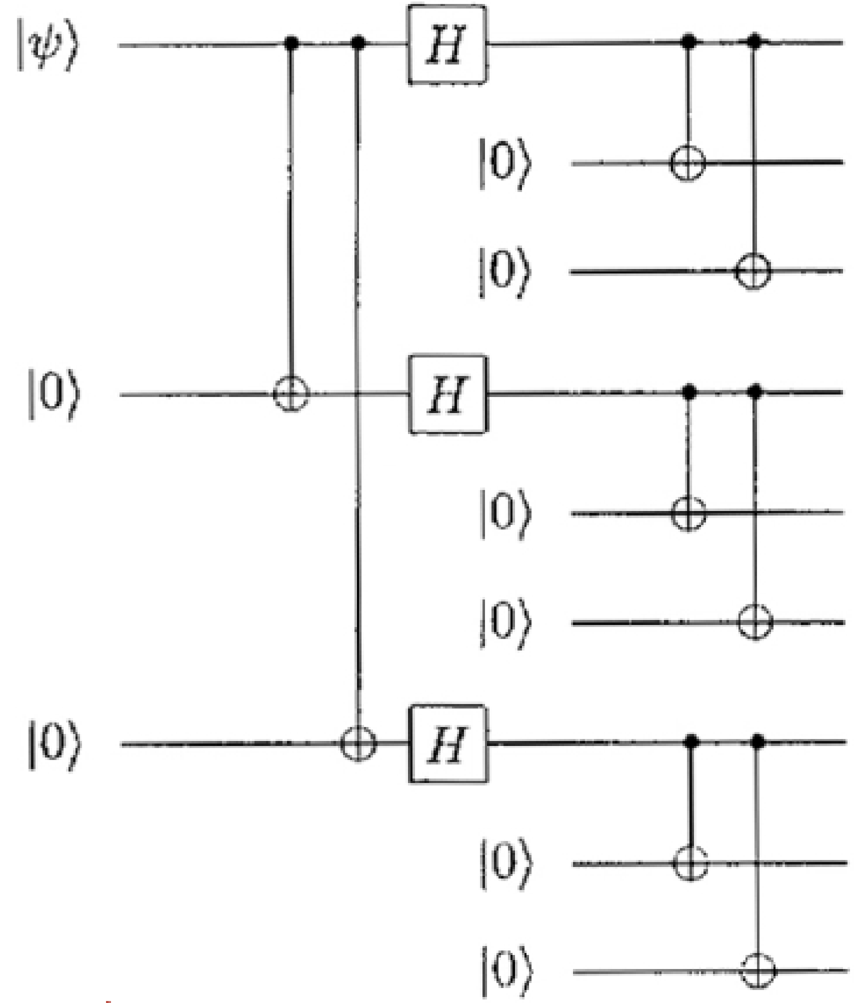


represents a controlled NOT gate

H represents the Hadamard transform $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$



With this code we can recover from a single qubit flip



The Shor code can recover from arbitrary single qubit errors

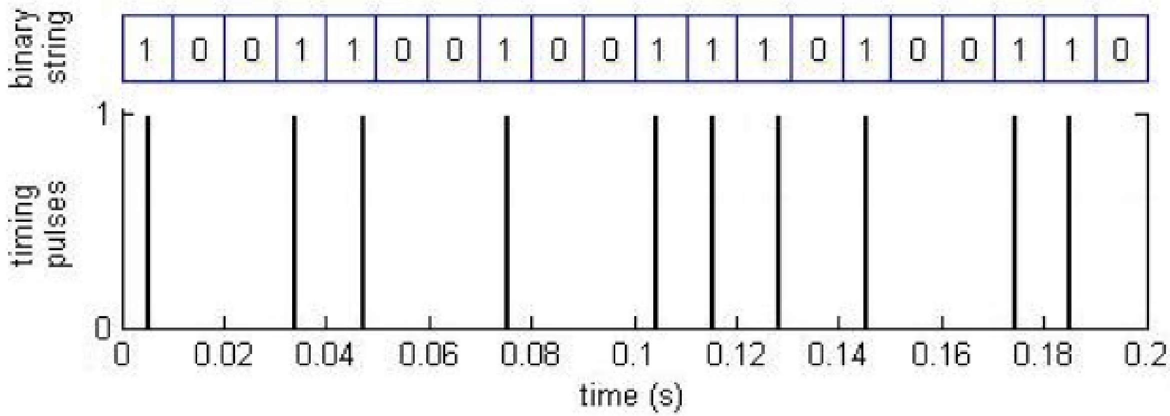
Quantum Computation and Quantum Information
Michael A. Nielsen and Isaac L. Chuang

Information Theory and the Brain

The information processing techniques of the brain are almost completely unknown to us.

Several experiments have empirically computed the mutual information between external stimuli and signals in the brain:

Spikes F. Rieke, D. Warland, R.R.v. Steveninck, and W. Bialek M.I.T. Press 1997



Some believe that information is conveyed by the timing of neural spikes

Some believe the need for an information theory of chemical signalling at neural synapses:

Living Information Theory
Shannon Lecture Toby Berger
IEEE-ISIT, Lausanne 2002

Tu n’as pas fini?

Tu n’as pas fini?