USC Viterbi
School of Engineering

2009 Viterbi Lecture
USC Viterbi School of Engineering
26 February 2009

Codes and Coin Flips: Compression and Modeling

Robert M. Gray
Department of Electrical Engineering
Stanford, CA 94305

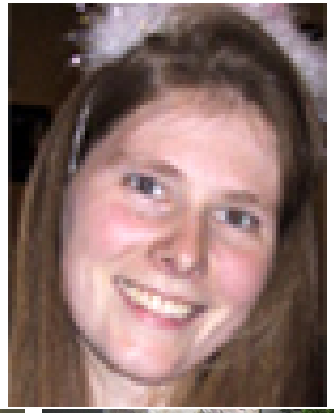
rmgray@stanford.edu http://ee.stanford.edu/~ gray/

Begin with the important stuff: Thanks to

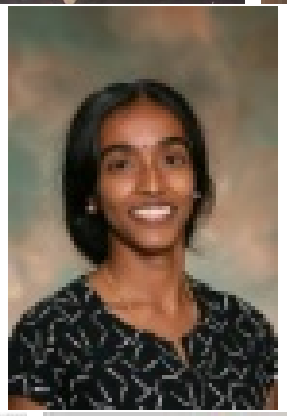
my USC dissertation committee

my students: *for making me look good through the years*

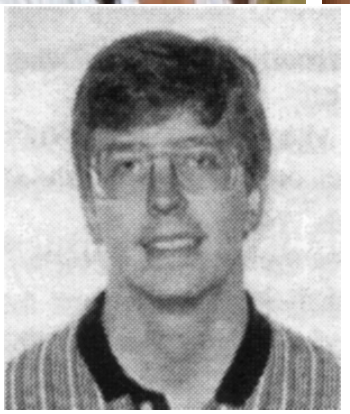




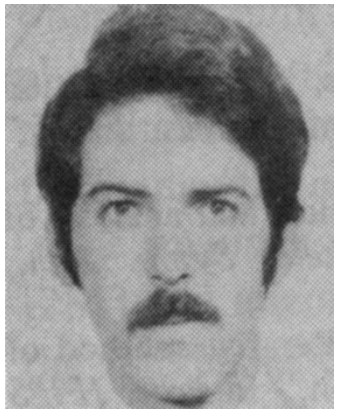










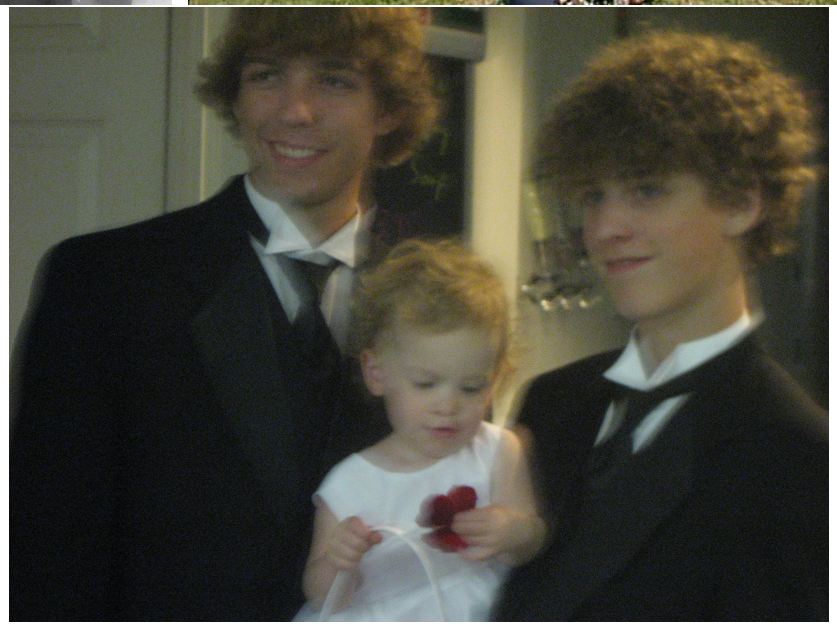




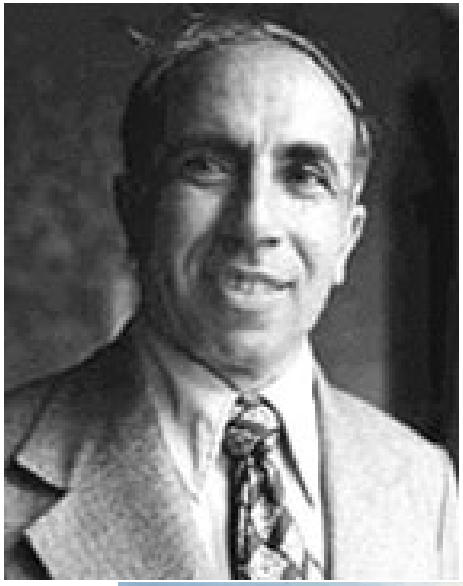
+ many more grandstudents of Bob Scholtz

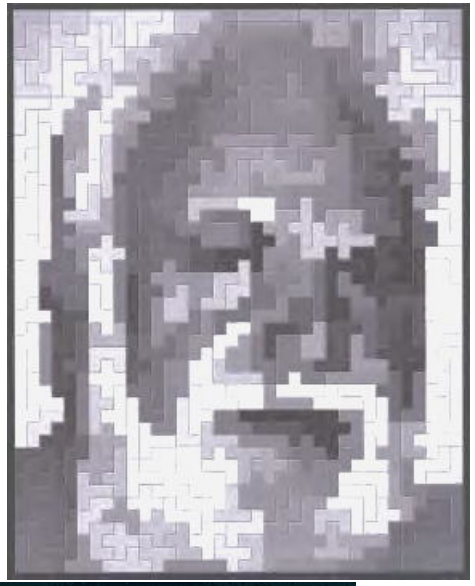
Total: 52, including 15 women (7 of whom are now professors)

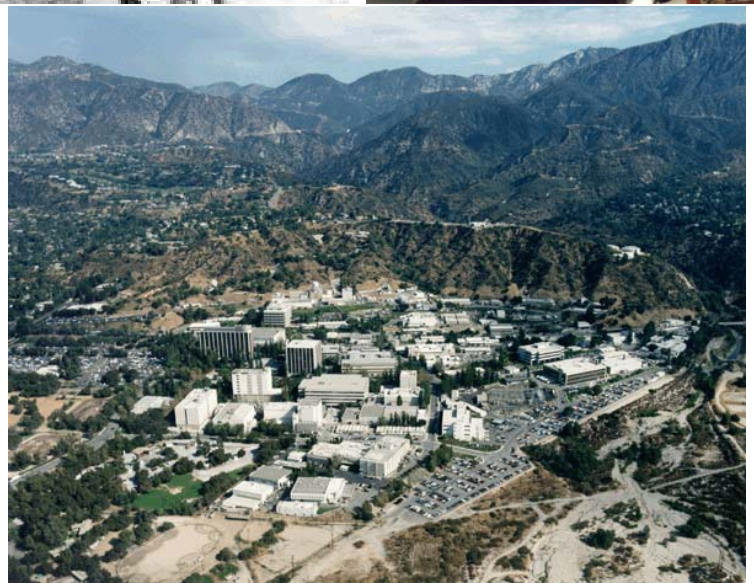
my family

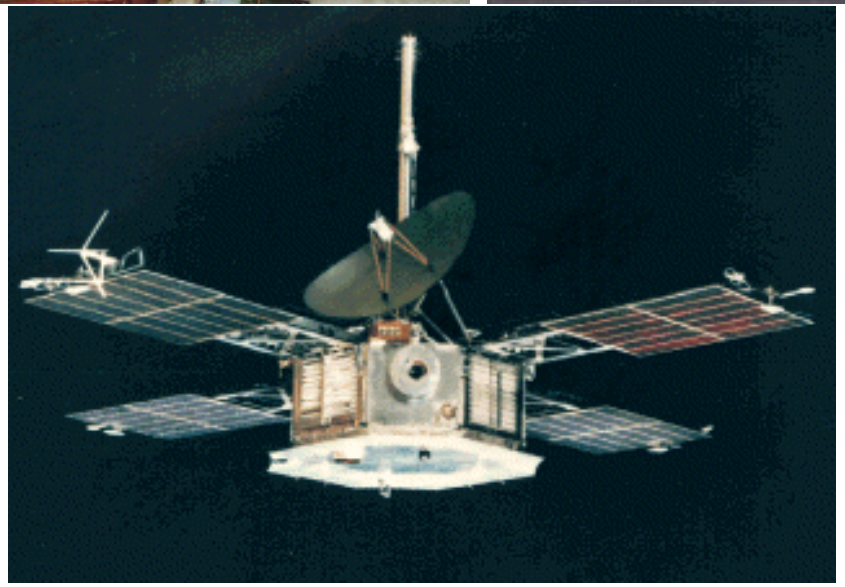


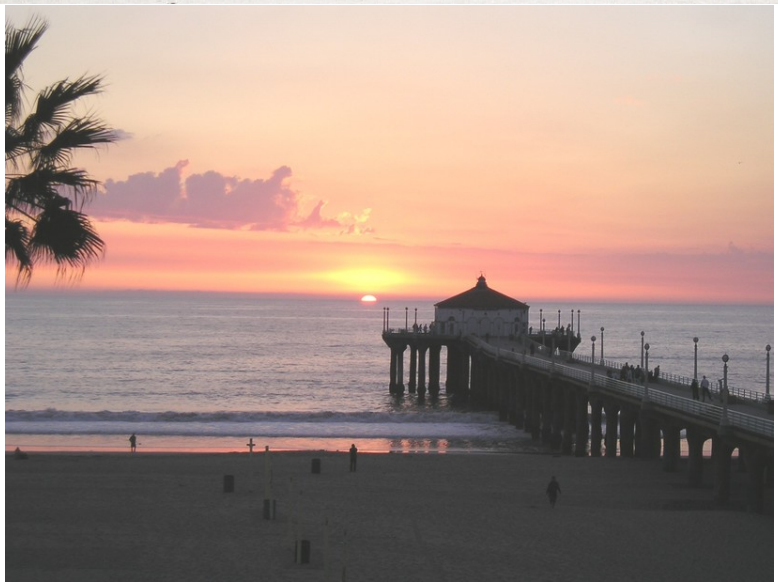
Why did I come to USC?





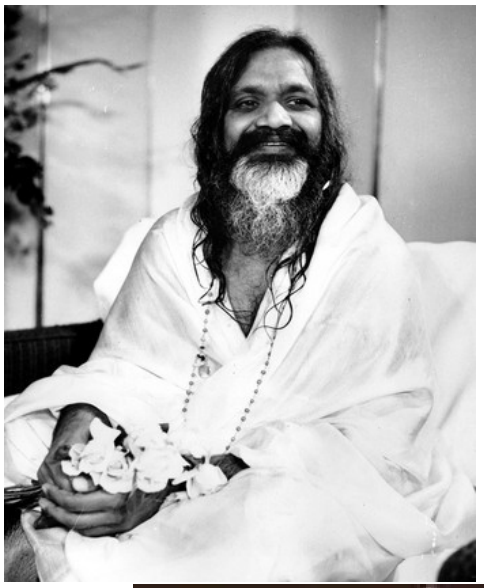






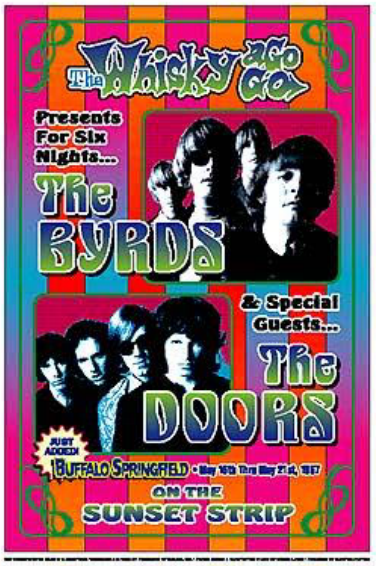

ARROWHEAD

LA 1966-1969



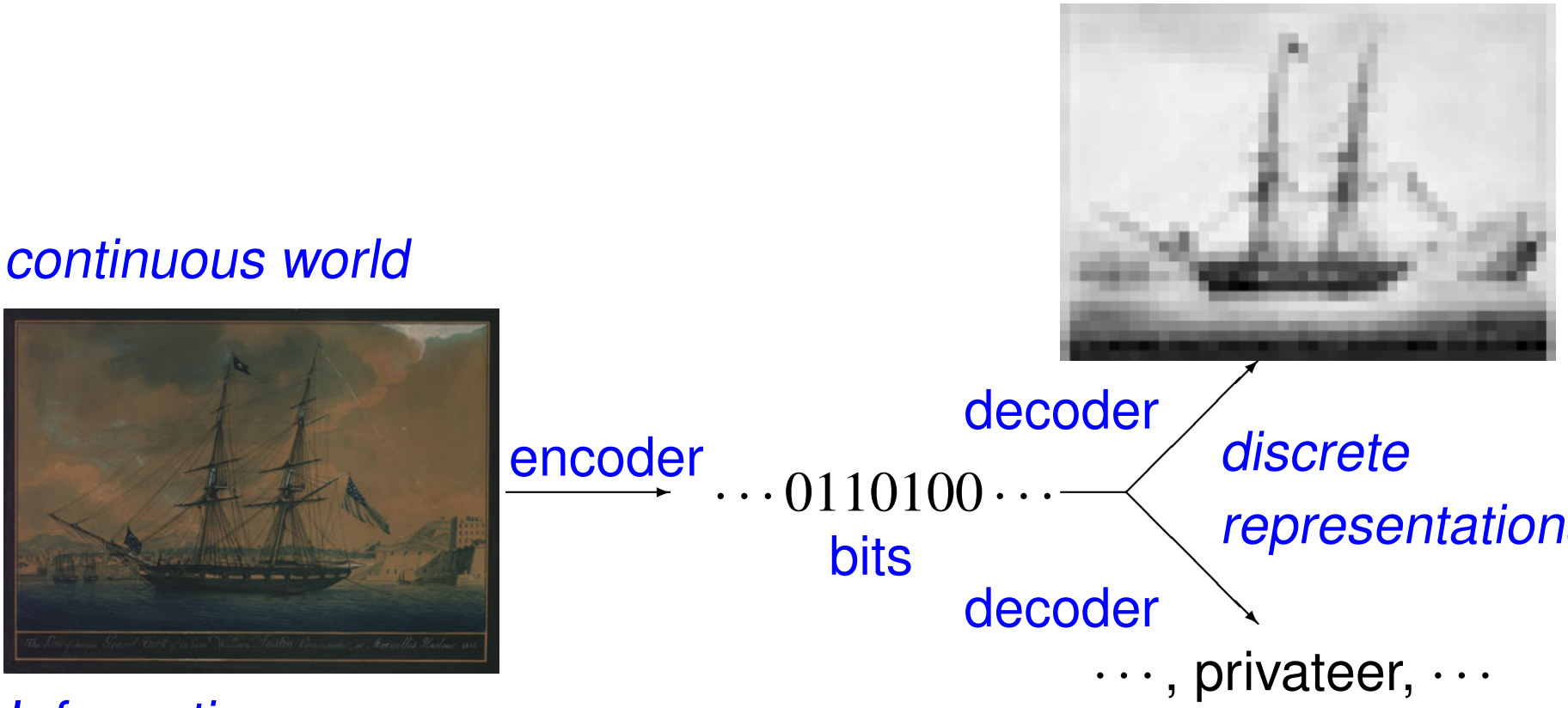

GENE
McCARTHY
for PRESIDENT in '68


STOP


The Whisky a go go
Presents
For Six
Nights...
The
BYRDS
& Special
Guests...
The
DOORS
JUST
ADDED!
BUFFALO SPRINGFIELD
ON THE
SUNSET STRIP

. . . four decades later

Communication and Bits



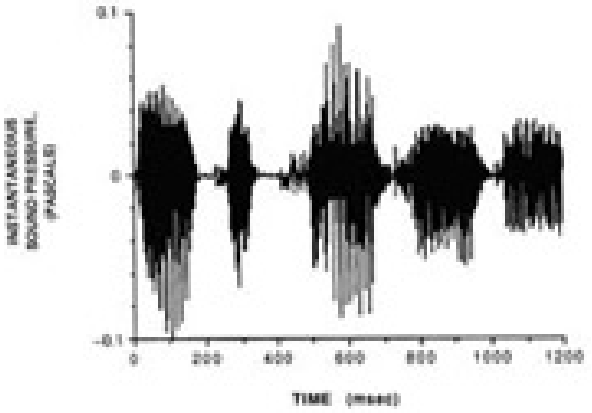
How well can you do it?

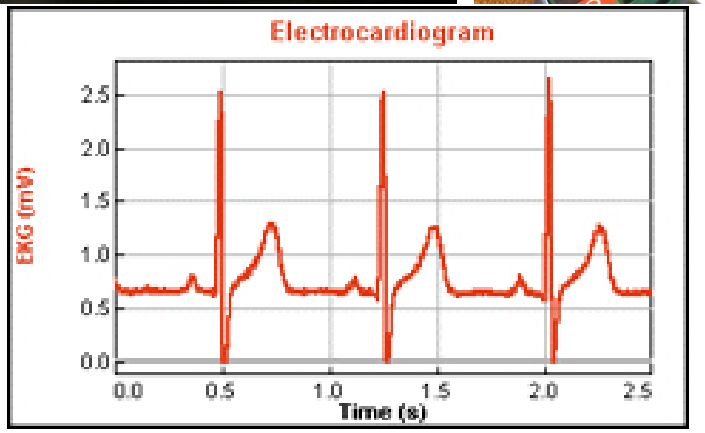
What do you mean by “well”?

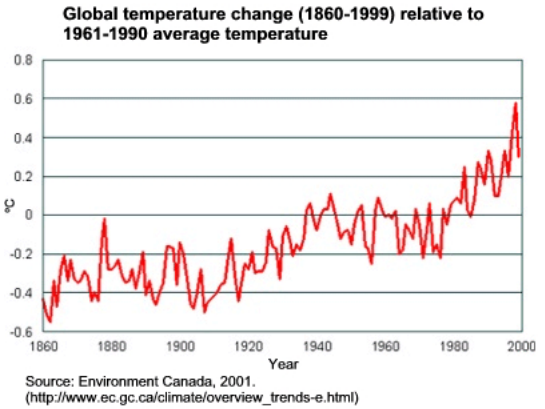
How do you do it?

Consider the pieces

Signals and Sources=Random Processes







Random Processes

Sequence of random numbers

$$\cdots, X_{-2}, X_{-1}, X_0, X_1, X_2, \ldots, X_n, \cdots$$

$$\Uparrow$$

value at time 0

Coin flips, dice rolls, roulette wheels *have no memory*

☞ theory and understanding is simple.

Real world processes are far more complicated

so begin with the simple stuff

Coin Flips

Simple fair coin flips play a fundamentally important role in theory, practice, interpretation, and teaching of random processes
and, of course, in communication

Give it a name: Z represents a sequence

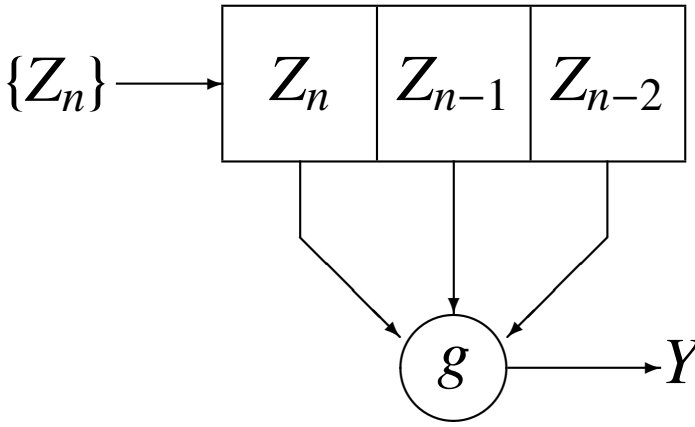
$$\cdots, Z_{-2}, Z_{-1}, Z_0, Z_1, Z_2, \ldots, Z_n, \cdots$$

of independent tosses of a fair coin (the same coin!)

Each Z_n is equally
likely to be 0 or 1, independent of the future and past

Processes with memory

One way: construct from coin flips by *coding/filtering*: simple example

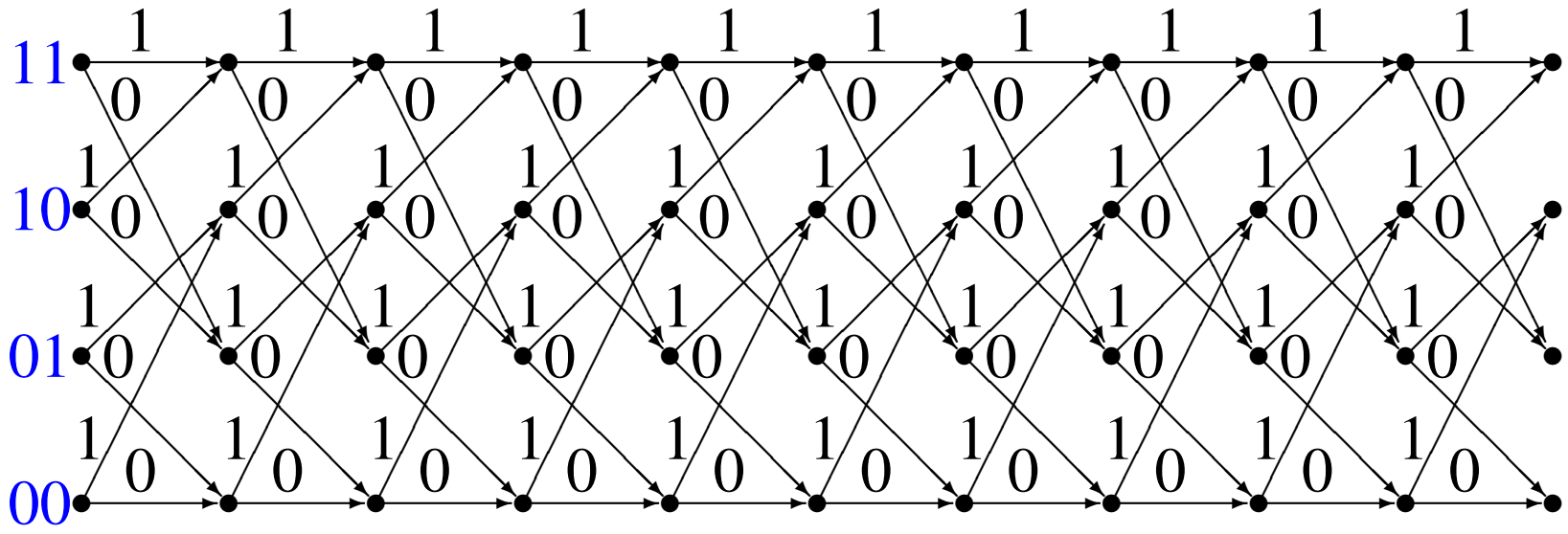


shift register

function, table

$Z_nZ_{n-1}Z_{n-2}$	Y_n
000	0
001	0
010	0
011	0
100	0
101	0
110	0
111	1

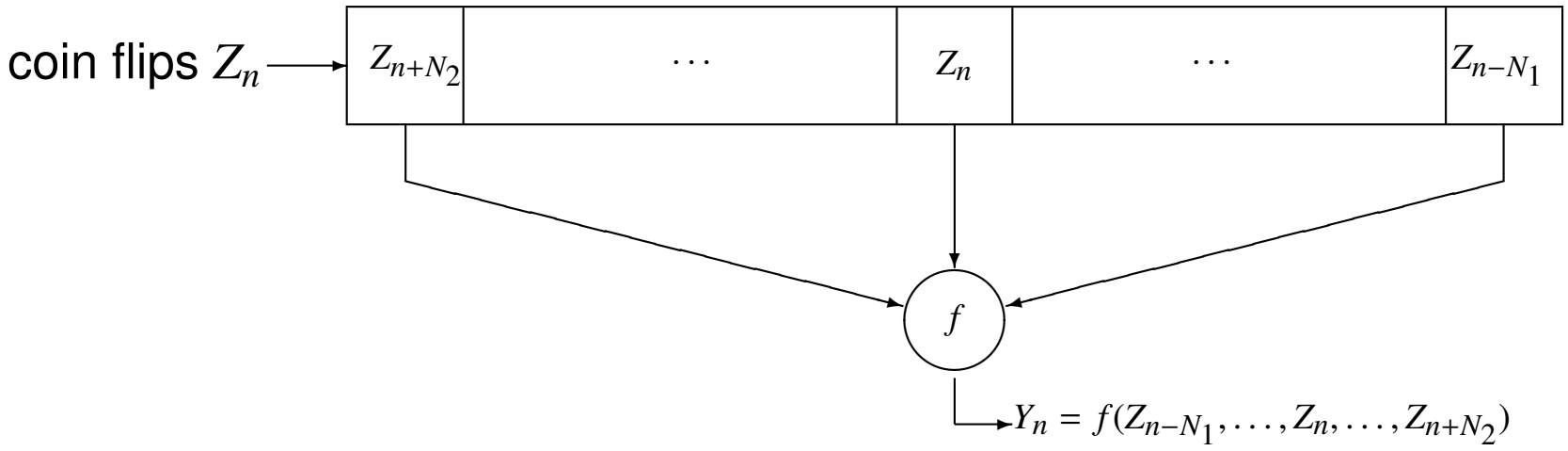
n	-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Z_n	1	1	0	0	0	1	1	0	1	1	1	1	0	0	1
Y_n			0	0	0	0	0	0	0	0	1	1	0	0	0



Trellis diagram of outputs of g

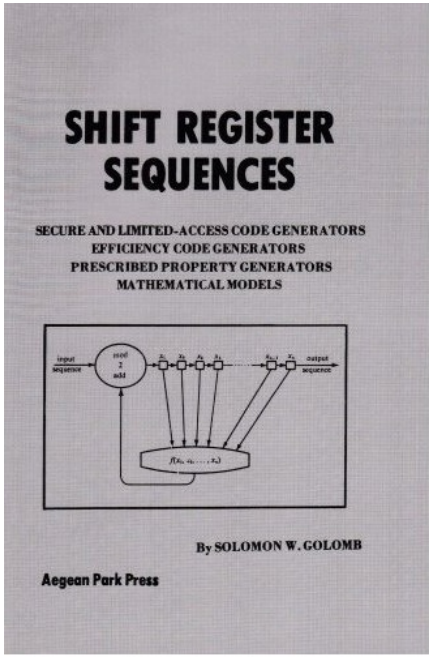
General case: **Sliding-block code**

- Many ways to fill table
- Shift-register can be arbitrarily long
- Can view “future” inputs as well as past (delay)



Sliding-block code (sliding-window code, stationary code)

shift register



+ function of contents

Can you model or simulate *all* processes of interest in this way??

If only use coin flips Z and want one output per flip, then **No**
(Shannon, Komogorov, Sinai)

not random enough.

¿

Question I Can the idea be modified to model most real-world processes?

Question II What does all this have to do with communications?

Answer to Question I

If allow any memoryless input process (not just Z) and big code windows, can generate arbitrarily good  models of *all* well-behaved real-world processes [1]

modeling, simulation

Class of all processes formed by sliding-block coding of memoryless sources is called B-processes (Ornstein (1970s), B for Bernoulli)

[1] well, *almost* all

Answer to Question II

If *forced* to use one coin flip per symbol (limited storage, bandwidth, time), there is a *best*  way to do it. *modeling/simulation*

And this also "solves" the communication problem[2] of communicating optimally  through a bit pipe (like the Internet) *data compression*

If allow R coin flips (bits) for each output sample/pixel, again there is a best model and compression system. As R grows, models and communications improve to arbitrary accuracy 

$\Rightarrow$ *Coin flips lie at the heart of (almost) all interesting random processes (modeling, simulating, teaching).*

[2] in theory

Distortion

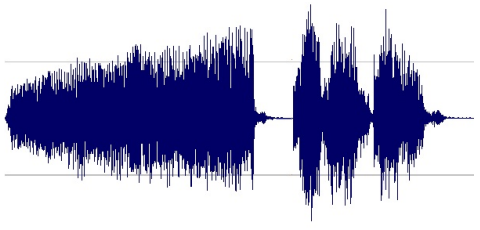
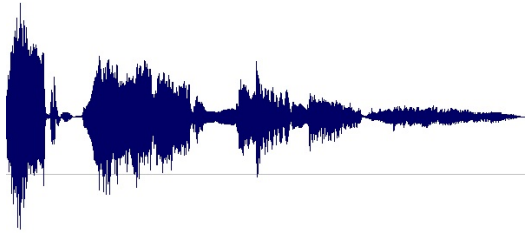
"Good"? "Best"? "Optimally"? "Accuracy"?

These notions require quantifying quality or fidelity,

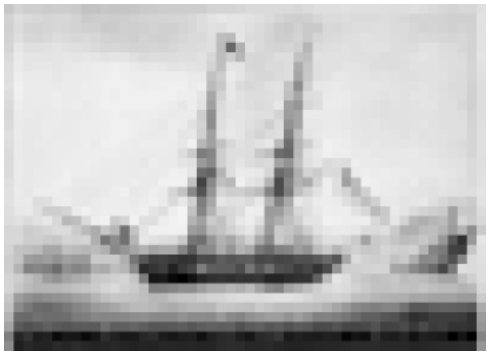
or their opposite — *distortion* $d(a, b) =$

distortion (cost, penalty) resulting if an original a is reproduced by b

$$d(0, 1), d(1, 1), d(3.14, \pi)$$

$d($  ,  $)$

$$d(01100101110100111000 0, 0110010011010111000 0)$$

$d($  ,  $)$

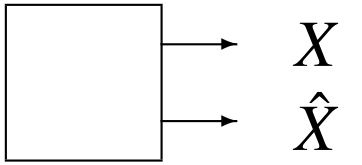
$d($  ,  $)$

$d(a, b) \geq 0$ and $d(a, a) = 0$

$d(a, b)$ Small (large) distortion $\Leftrightarrow$ looks/sounds good (bad)

Average Distortion

Suppose have single box that produces both processes X and $\hat{X}$:



For example, X is fair coin flips, $\hat{X}$ is coin flips with bias q. Independently flipped or connected by a rubber band.

Average distortion Statistical/probabilistic average or expected distortion $E[d(X, \hat{X})]$ (calculus)

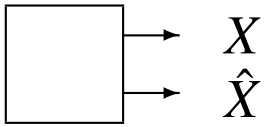
Theory: Law of large numbers $\Rightarrow$ Run process, average distortion over time as n grows

$$\frac{d(X_0, \hat{X}_0) + d(X_1, \hat{X}_1) + d(X_2, \hat{X}_2) + \cdots + d(X_{n-1}, \hat{X}_{n-1})}{n} \longrightarrow E[d(X, \hat{X})]$$

Process Metric

What is the *best* box/rubberband/coupling of two random processes X and $\hat{X}$? Make precise:

$$\bar{d}(X, \hat{X}) = \text{minimum } E[d(X, \hat{X})]$$



Monge (1789)/Kantorovich (1940)/Ornstein (1970)/etc.[3] distance

Fascinating history, rediscovered many times in many fields.

[3] etc. = Salvemini (1943), Dell'Agio (1956) Fréchet (1957), Rubinstein (1958), Vasershtein/Wasserstein (1969), Dobrushin (1970), Mallows (1972), Vallender (1973), Gray, Neuhoff, and Shields (1975), Rubner, Tomasi, and Guibas (1998), Levina and Bickel (2001), *ad nauseum*

Monge/Kantorovich "Optimal transportation cost"
(Kantorovich shared 1974 Nobel prize in Economics for related work)

Ornstein's $\bar{d}$ (d-bar) distance (1974 Bôcher Prize)

Mathematics: How optimally move one pile of sand to another so as to minimize "transportation" (of sand)?

Economics: How optimally map production profile into consumer profile in minimum cost manner? Resource allocation, linear programming.

Random processes: *How much do you have to change one process to make it look like another?* minimum distortion match

For example: two coins, bias p and q: distance is $|p - q|$

In general — linear programming problem

Put the pieces together (home stretch)

$\bar{d}$ distance provides a *geometry* of random processes (like Euclidean distance/geometry)

Back to **Question I** What is the *best possible* simulation of a process X from Z= fair coin flips?

$$\Delta_X = \underset{g}{\text{minimum}}\, \bar{d}(X, g(Z))$$

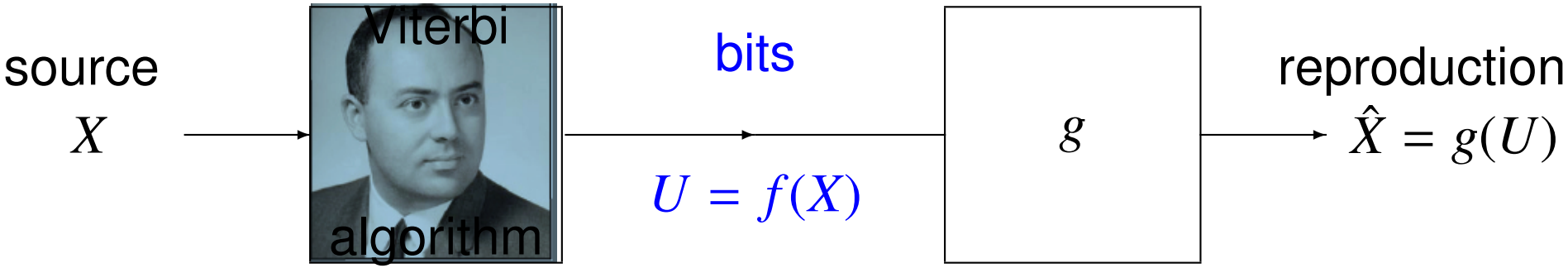
If X is a B-process $\Delta_X = D_X(1)$, Shannon distortion-rate function (much studied in Information Theory)

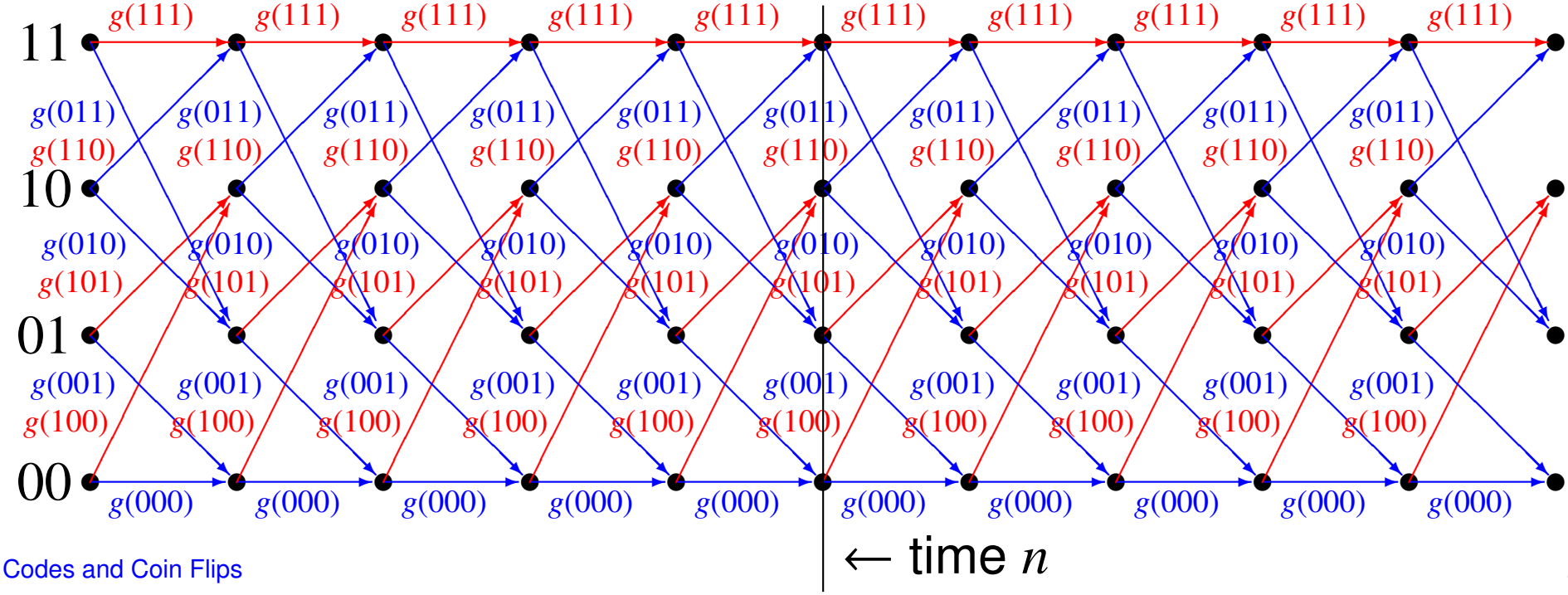
But finding best g is an open problem.

Back to **Question II**

Suppose solve **Question I**, have SBC g such that $\bar{d}(X, g(Z)) \approx \Delta_X$.

Communications? Use minimum distortion search matched to g





If f=Viterbi algorithm encoder (sliding-block approximation) and X is a B-process, then

$$E[d(X, g(f(X)))] \approx \Delta_X = D_X(1)$$

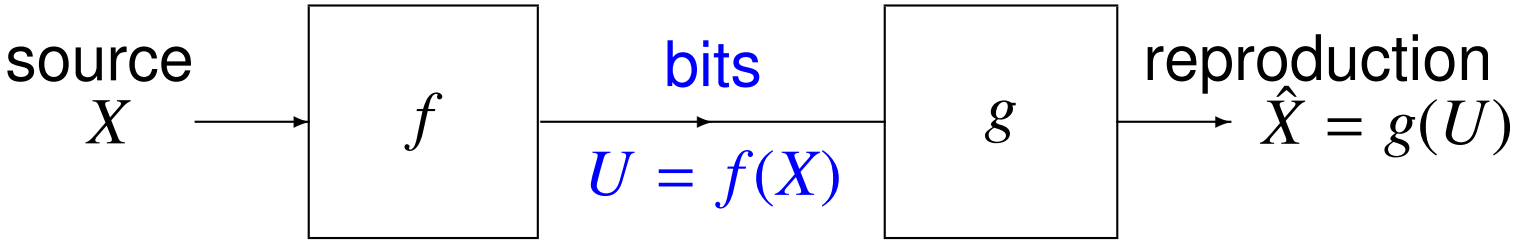
Best match of typical sequences of X and $g(U) \Rightarrow \bar{d}(X, g(U))$

In other words:

optimal simulator as decoder + Viterbi algorithm encoder yields approximately optimal communication through a bit pipe.

Recently shown that the reverse statement is true.

If a communications system



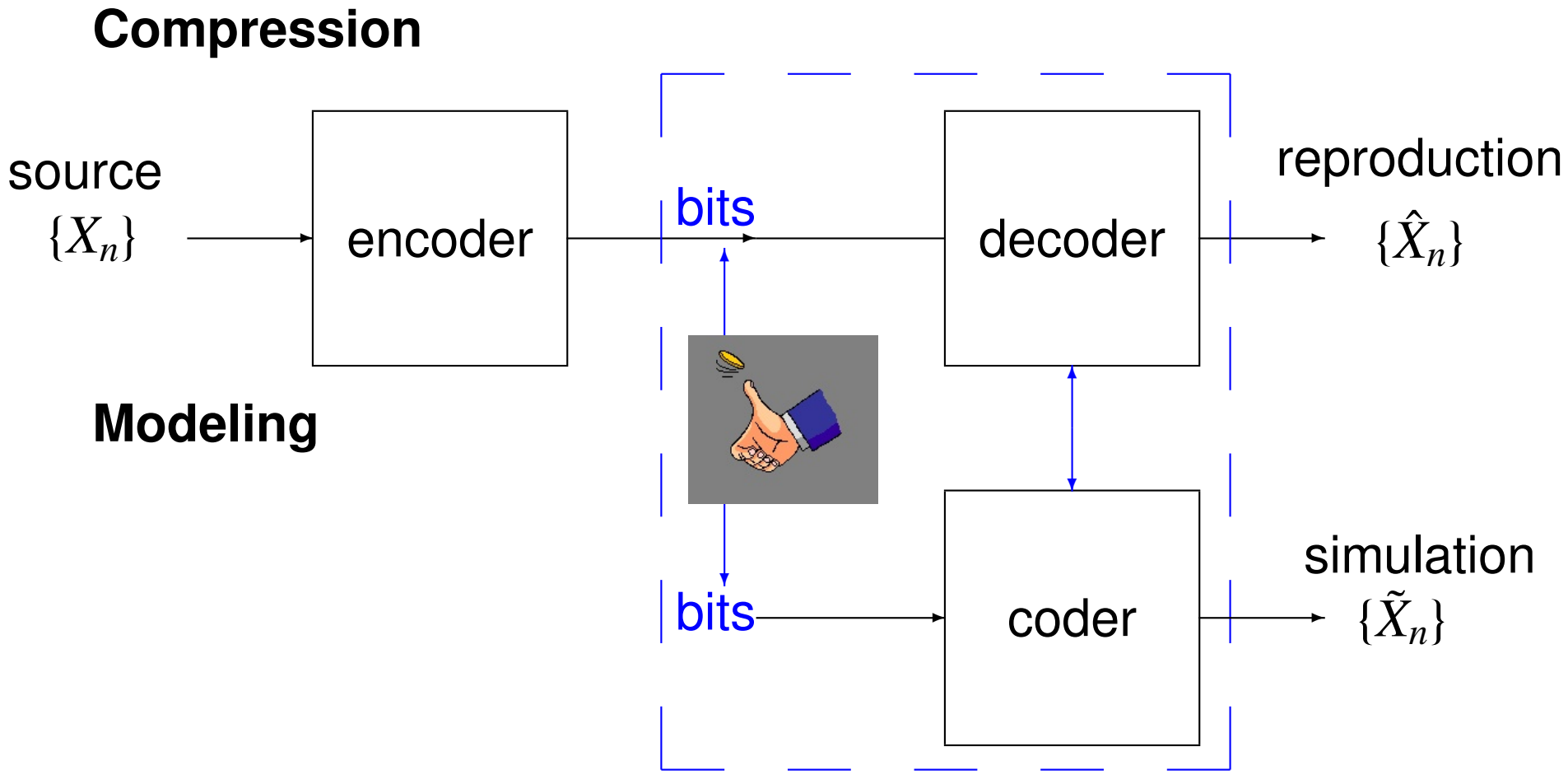
has a nearly optimal encoder f and decoder g for X, i.e., average distortion is near the Shannon limit

$E[d(X, \hat{X})] \approx D_X(1) = \min_{f,g} Ed[X, g(f(X))]$

Then it must also be true that $\bar{d}(Z, U) \approx 0 \quad \Rightarrow$

when working near Shannon limit of communication at 1 bit per symbol, encoded bits look (almost) like fair coin flips.

Summarize with a picture



Good compression ⇔ good modeling/simulation!

and fair coin flips are at the ♡ of it all!

Thanks for your attention!

. . . and the memories