# Anytime Capacity
# Week 8 Presentation

Constantine Caramanis, MIT

Based on the Ph.D. Thesis of Anant Sahai. Also includes work by Sekhar Tatikonda and Sanjoy Mitter.

# Introduction

## Communication and Control

- Control Issues:
  a. Causality
  b. Delay Sensitivity
  c. End–to–End Performance
- Source Issues:
  a. Non-Stationarity
  b. Streaming nature of source.

# Outline

- The Problem Motivation: Unstable Markov Source,
- Current Methods Inadequate,
- New Notion of Reliable Comm: Anytime Capacity

  a. The Binary Erasure Channel,

  b. The Additive White Gaussian Noise Channel,

  c. The General Case.

# Outline

- Solution to the Problem:
    a. The Direct Case.
    b. The Converse Case Constructively.
- Quality of Service: A New Resource
- Conclusion.

# Results

Main Result: We completely characterize the capacity of a larger class of sources, in terms of a useful parameter.

Additional Results:

a. We obtain a new source/channel separation theorem.

b. This parameter has a quality of service interpretation.

# The Problem

**The Problem:** *Given the system:*

- *Parameter:* $A > 1$,
- *Noise:* $\{W_t\}$,
- *Dynamics:*

$$X_{t+1} = AX_t + W_t,$$

*Can we transmit information across a noisy channel to obtain* $\hat{X}_t$ *that tracks* $X_t$ *with finite distortion? Do we need feedback?*

# The Problem

- Classical Shannon capacity is not sufficient.
- A sufficient zero–error capacity is clearly sufficient. Is it necessary? Hopefully not!
- Is there a new notion of “reliable transmission” between these two?

# The Problem

Different From Ordinary Information Transmission:

- Message not explicitly specified at time “-1,”
- Application–specific importance of delay,
- Dependence on past values.

# Current Methods Inadequate

Two attempts via Block Coding:

Idea: Encode the State of $\boldsymbol{X}_t$ itself.

Problem: $\boldsymbol{X}_t$ becomes unbounded.

$\Rightarrow$ Cannot track with finite rate.

# Current Methods Inadequate

Idea: Encode only update, $\boldsymbol{W_t}$, using block coding.

Properties:

a. Dependence on all past errors,

b. Exponential error sensitivity.

Result:

a. Works with noiseless feedback,

b. Does not work without feedback!

# Current Methods Inadequate

**Proposition 1** *Traditional block coding inadequate to track unstable Markov Source!*

In Fact: Much larger class inadequate. Turns out that we **need** sensitivity to past bits!

# Anytime Capacity

**Motivating Ideas**:

- Need codes with dependence on past errors,
- Need to use dependence to **correct** past errors.

Note: Block encoders/decoders do not have these properties.

We define anytime encoders and decoders that do have these properties.

# Anytime Capacity

Let $\theta$ be the offset. Let $\mathcal{B}^i$ be the set of possible outputs by time $i$.

**Definition 1** *A Rate* $R > 0$ *Anytime Channel Decoder is a Sequence of Maps* $\mathcal{D}^a = \{\mathcal{D}^a_i\}$,

$$\mathcal{D}^a_i : \mathcal{B}^i \longrightarrow \{0, 1\}^{\lfloor(\theta+i)R\rfloor}.$$

We will consider max likelihood anytime decoders.

# Anytime Capacity

We now give the key definition:

**Definition 2** *For $f(d) > 0$ a decreasing function of delay, we define the $f$–Anytime Capacity as:*

$$C_a(f) := \sup\{R \mid \exists\, K, \text{ s.t.} P_e(\mathcal{E}, \mathcal{D}^a, d) \leq K f(d)\},$$

*the sup over achievable rates with $f$–decay of error.*

# Anytime Capacity

We care about exponential decay of error:

**Definition 3** *The* $\alpha$*–Anytime Capacity* $C_a(\alpha)$, *is given by:*

$$C_a(\alpha) := C_a(2^{-\alpha d}),$$

*using the above definition.*

<u>Note</u>:

$$C_{\text{zero–error}} \leq C_a(\alpha) \leq C_{\text{Shannon}}.$$

# Anytime Capacity

A stronger notion of reliable communication:

**Proposition 2** *If* $R < C_a(\alpha)$*, for* $\alpha > 0$*, every bit can be eventually decoded correctly.*

proof: Apply the Borel–Cantelli lemma.

# The Encoder

## The Encoder

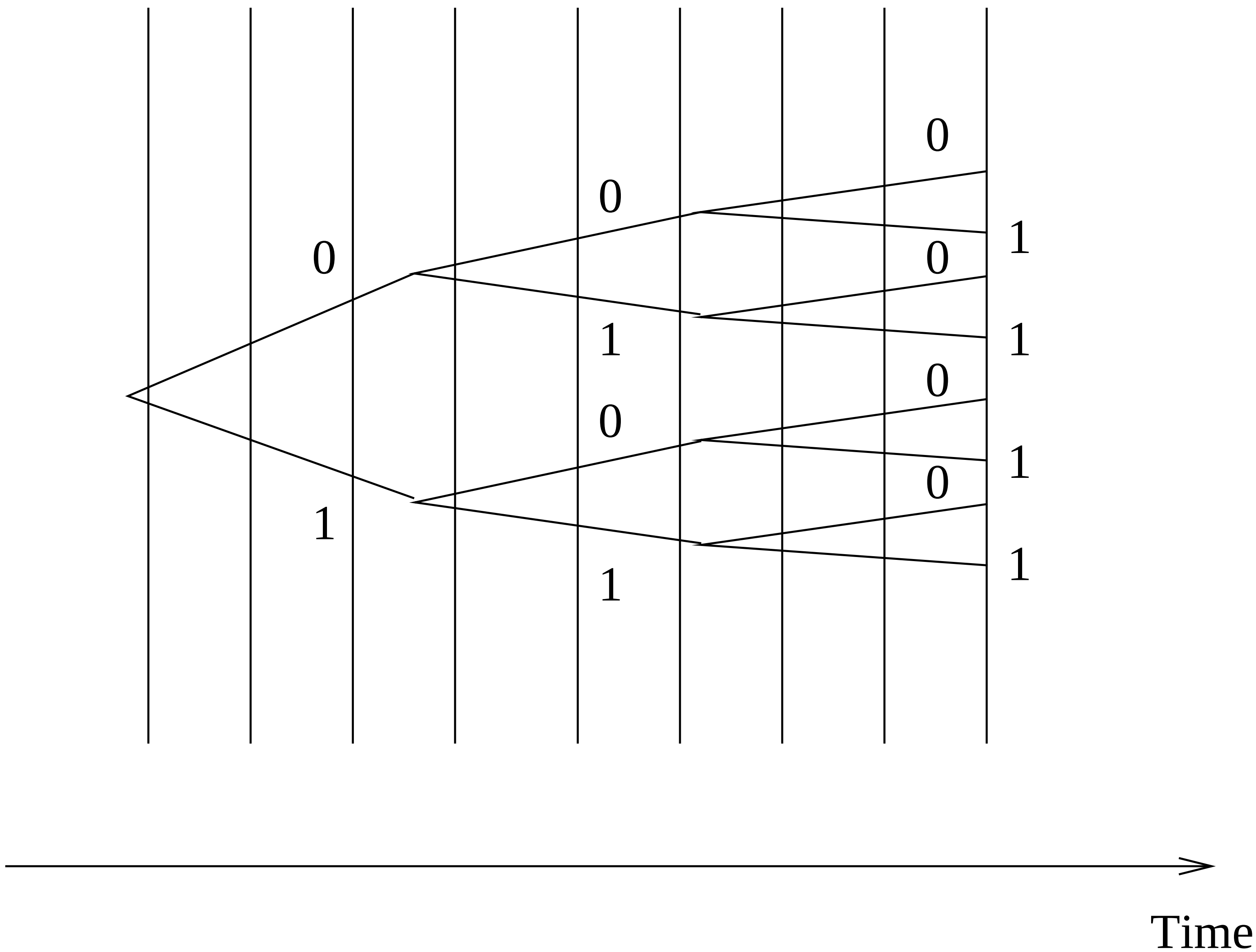




Figure 1: Random Encoder, $R = \frac{1}{3}$

# The Encoder

## The Encoder

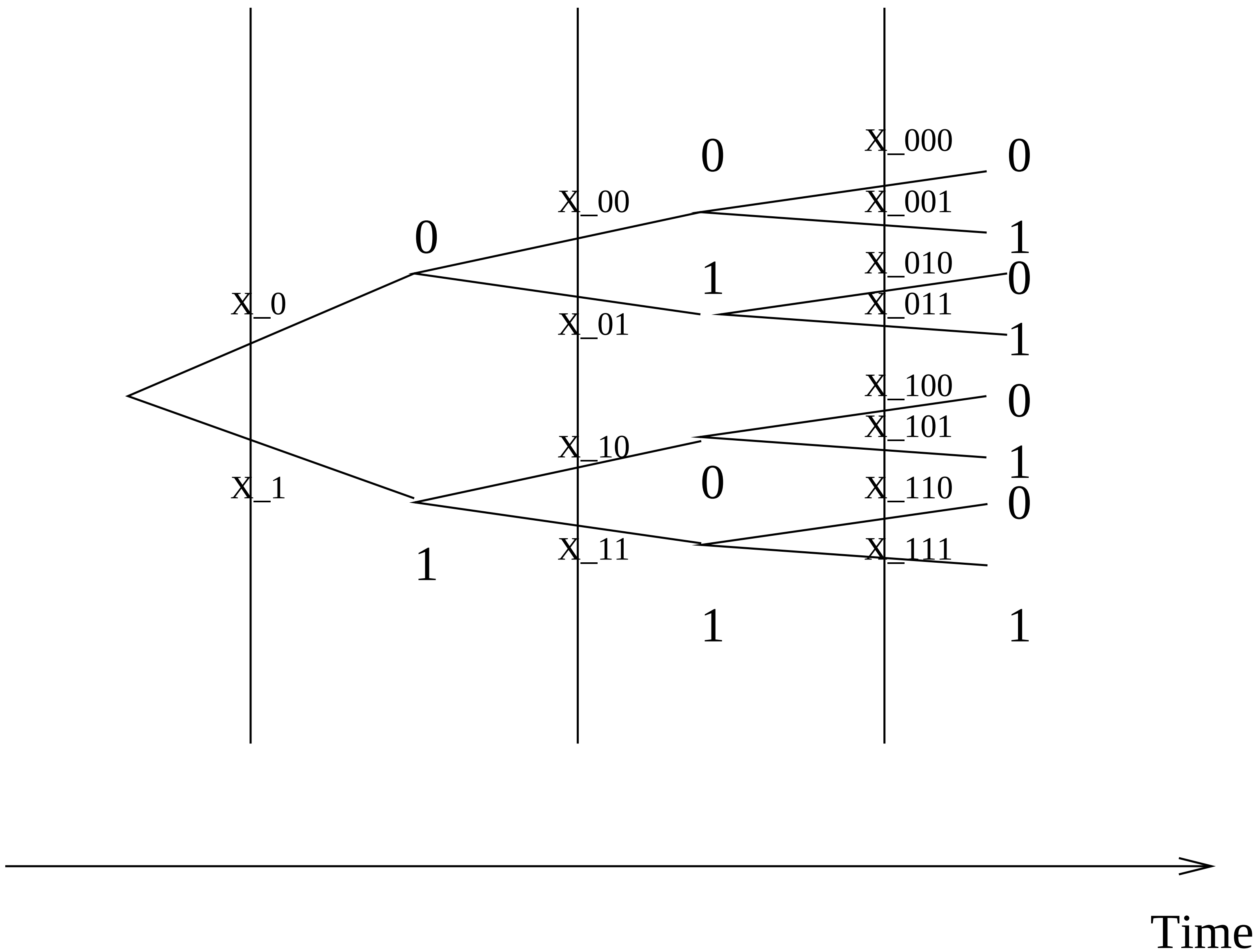




Figure 2: Random Encoder, $R = 1$

## Anytime Capacity



We have the following results:

- BEC($p$): $C_{\text{anytime}}(\alpha) \geq 1 - \log_2(1+p) - \alpha.$
- $P$–constrained AWGN:
  $$C_{\text{anytime}}(\alpha) \geq \frac{1}{2}\log_2(1+\frac{P}{2}) - \alpha.$$
- Memoryless Channel:
  $$C_{\text{anytime}}(E_r(R)\log_2 e) \geq R\log_2 e.$$

# Solution of the Problem

We need to make precise the notion of bit sensitivity.

$$
\begin{aligned}
\mathbf{\Delta}^{+}(j,t) &= \text{the largest distortion at } t \text{ caused by} \\
&\quad \text{bit mistakes from time } j \text{ and on.} \\
\mathbf{\Delta}^{-}(j,t) &= \text{the largest distortino at } t \text{ caused by} \\
&\quad \text{bit mistakes up to time } j.
\end{aligned}
$$

# Solution of the Problem

And also,

$$\mathbf{\Delta}^{+}(d) = \sup_{t>0} \mathbf{\Delta}^{+}(\lfloor R(t-d) \rfloor, t)$$

$$\mathbf{\Delta}^{-}(d) = \sup_{t>0} \mathbf{\Delta}^{-}(\lfloor R(t-d) \rfloor, t).$$

- $\mathbf{\Delta}^{+}$ is the largest possible distortion caused by mistakes made $d$ units ago and on,
- $\mathbf{\Delta}^{-}$ is the largest possible distortion caused by mistakes made up to time $d$ units ago.

# Solution of the Problem

Some observations:

- For a rate $R = \frac{R_{in}}{R_{out}}$ block source code,

  a. $\boldsymbol{\Delta}^{+}(j, t) = \boldsymbol{\Delta}^{-}(k, t) = \mathbf{0}$ for $k$ small enough, and $j$ large enough.

  b. $\lim_{d \to +\infty} \boldsymbol{\Delta}^{-}(d) = \mathbf{0} = \lim_{d \to -\infty} \boldsymbol{\Delta}^{+}(d)$.

## Solution of the Problem

Now we can make precise the inadequacy of block codes.

**Theorem 1** *For an unstable Markov source as above, and any finite expected squared–error–distortion finite rate code,* $\Delta^{+}(d)$ *grows exponentially at least as fast as* $A^{2d}$*, for* $A$ *the source parameter.*

Block codes fall into this category of inappropriate codes.

# Solution of the Problem

Intuition: There are two main points.

- We need the probability of error to decay quickly enough to counter the growth of $\mathbf{\Delta}^+$.
- This will make more explicit the delay/accuracy tradeoff for the decoder.

**Theorem 2** *Consider a source code with finite distortion $D_R$ that emits a rate $R$ stream of bits. Suppose our noisy channel has $C_a(f) > R$ for some function $f(d)$ such that*

$$\lim_{\delta' \to \infty} \sum_{n=0}^{\infty} f(\delta' + \frac{n}{R}) \Delta^{+}(\frac{n}{R}) = 0,$$

*Then we can transmit source code across noisy channel with finite expected distortion that approaches $D_R$ as $\delta' \to \infty$.*

## Converse Direction

The converse is more difficult.

- Want to show that sufficient anytime capacity for $\alpha$ sufficiently large is necessary to transmit an unstable Markov source.
- We show that if we can transmit an unstable source, then we have appropriate anytime capacity.
- We have only an operational definition of anytime capacity. Need to work with this: Constructive Proof.

# Solution of the Problem

## Converse Direction

**Theorem 3** *Suppose we can track any scalar unstable Markov process $X_t$ with parameter $A > 1$, driven by a bounded noise signal $W_t$, with $||W_t|| \leq \frac{\Omega}{2}$, across a given channel. Specifically, suppose that for this channel we have*

$$P(|\hat{X}_t - X_t| \geq \Delta) \leq f(\Delta).$$

*Then, $\forall \varepsilon > 0$, this channel has*

$$C_{\text{anytime}}(\tilde{f}) > \log_2 A - \varepsilon,$$

*where $\tilde{f} := f(\delta' 2^{d \log_2 A})$, for some $\delta' > 0$.*

# Solution of the Problem

## Converse Direction

Idea: We construct an anytime encoder/decoder pair:

- Use input stream $S_1^\infty$ to simulate unstable Markov source $X_t$.
- Transmit $X_t$ across the channel. Obtain $\hat{X}_t$, and recover bits $\hat{S}_1^\infty$.

Some Issues:

- Need to recover bits from tracking process $\hat{X}_t$.
- Moreover: Need to recover earlier bits **more reliably**.

- Take $W_t$ to be binary–valued "noise," $W_t \in \{-\delta/2, \delta/2\}$. The input stream gives the "realization" of the noise.
- We construct:

$$X_t = \sum_{i=1}^{t} A^{t-i} W_i.$$

- The earlier noise gets exponentially amplified by $A$.

Representations of the Reals: Recall some properties of binary and ternary expansions.

- Binary Expansion:

a. $\forall \beta \in [0, 1]$ we can write, for $a_n \in \{0, 1\}$,

$$\beta = \sum_{n=1}^{\infty} 2^{-n} a_n,$$

b. If $\beta_1, \beta_2$ have identical expansions up to (and including) bit $n$,

$$|\beta_1 - \beta_2| \leq 2^{-n}.$$

- Ternary Expansion:

a. The set of all binary expansions, interpreted as ternary expansions, form a scaled Cantor set.

b. If $\beta_1, \beta_2$ have identical expansions up to (and including) bit $n$, we have upper **and** lower bounds:

$$\frac{1}{2}3^{-n} \leq |\beta_1 - \beta_2| \leq 3^{-n+1}.$$

Note that this does not hold for binary case.

# Solution of the Problem

## Converse Direction

### ...Main Idea...

- Similarities: Given some $x \in [0, 1]$, or in the Cantor set, we obtain the expansion by a series of **threshold** comparisons, and **translations**.

- Dissimilarities: Closeness in the usual metric, $|\beta_1 - \beta_2|$, assures a lexicographic closeness of the sequences, and conversely for the ternary case. This is <u>not</u> true for the binary case.

# Solution of the Problem

## Converse Direction

### ...Main Idea...

To construct $X_t$, take $A = (2 + \varepsilon_1) = 2^{1+\varepsilon_2}$. Then

$$X_t = \sum_{i=1}^{t} A^{t-i} W_i = \frac{\delta}{2} \sum_{i=1}^{t} (2 + \varepsilon_1)^{t-i} (2S_i - 1).$$

By translations and normalizations, we obtain the new mapping

$$X'_t = \sum_{i=1}^{t} (2 + \varepsilon_1)^{-i} S_i.$$

# Solution of the Problem

## Converse Direction

### ...Main Idea...

Now consider the mapping

$$x'_\infty(s_1^\infty) := \sum_{i=1}^{\infty} (2+\varepsilon_1)^{-i} s_i.$$

The range is a Cantor set. When $\varepsilon_1 > 0$, we have the very important lower bound property.

Using thresholds and translations we can reverse the map to retrieve bits perfectly.

## Solution of the Problem

### Converse Direction

#### ...Main Idea...

We do not receive $X_t$ perfectly, due to channel distortion.

Question: What happens if we reverse the map on $X_t + \zeta$, rather than $X_t$?

# Solution of the Problem

## Converse Direction

### ...Main Idea...

Answer: As long as $|\zeta|$ is not very big, we are guaranteed to retrieve first $n$ bits perfectly.

- Because $\varepsilon_1 > 0$ we have the lower bound property:
- If two sequences differ for the first time in digit $n$,

$$|x'_\infty(s_1^\infty) - x'_\infty(\hat{s}_1^\infty)| \geq (2 + \varepsilon_1)^{-n} \frac{\varepsilon_1}{1 + \varepsilon_1} > 0.$$

# Solution of the Problem

## ...Main Idea...

- Our mapping of translations and normalizations is:

$$x' = \frac{1}{\delta A^t}\hat{X}_t + \frac{1 + (2+\varepsilon_1)^{-t}}{2+2\varepsilon_1}.$$

- Note that there is a scaling by $A^{-t}$.

We now bound the probability of error in first $k$ bits if we accept delay $d$.

Then we have:

$$
\begin{aligned}
P_e(d,k) &\leq P\left(|\zeta| > \frac{\varepsilon_1}{2+2\varepsilon_1}(2+\varepsilon_1)^{-k}\right) \\
&= P\left(|\delta A^{k+d}\zeta| > \frac{\varepsilon_1}{2+2\varepsilon_1}(2+\varepsilon_1)^{-k}\delta A^{k+d}\right) \\
&= P\left(|X_t - \hat{X}_t| > \frac{\varepsilon_1}{2+2\varepsilon_1}(2+\varepsilon_1)^{-k}\delta A^{k+d}\right) \\
&\leq P\left(|X_t - \hat{X}_t| > \frac{\delta\varepsilon_1}{2+2\varepsilon}(2+\varepsilon_1)^{d}\right),
\end{aligned}
$$

And therefore,

$$
\begin{aligned}
P_e(d,k) &\le f\left(\frac{\delta\varepsilon_1}{2+2\varepsilon}(2+\varepsilon_1)^d\right)\\
&= f\left(\frac{\delta\varepsilon_1}{2+2\varepsilon}2^{d\log_2 A}\right).
\end{aligned}
$$

Thus taking $\delta' = \frac{\delta\varepsilon_1}{2+2\varepsilon_1} > 0$, we have the desired result of the theorem. We have proved the theorem for the case $A = 2+\varepsilon_1$.

# Quality of Service

Consider the vector process $\boldsymbol{X}_t \in \mathbf{R}^5$,

$$\boldsymbol{X}_{t+1} = \boldsymbol{A}\boldsymbol{X}_t + \boldsymbol{W}_t,$$

where

$$\boldsymbol{A} = \begin{pmatrix} \mathbf{1.258} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1.058} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1.058} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1.058} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1.058} \end{pmatrix}.$$

# Quality of Service

We must have

$$C_a(2\log_2(1.258)) > \log_2(1.258),$$

$$C_a(2\log_2(1.058)) > \log_2(1.058).$$

Also, we must have (Tatikonda)

$$C > \log_2(1.258) + 4\log_2(1.058).$$

# Quality of Service

Question: Can we do this for the BEC, $e = \mathbf{0.27}$?

Answer:

1. No, if we insist on treating all bits alike in transmission layer,

2. Yes, if we offer differentiated service, i.e. we provide effectively different $\alpha$.

# Conclusion

- Interaction between Control and Information Theory calls for new paradigm of reliable communication.
- New Notion of Reliable Transmission: Anytime Capacity.
- Necessary and Sufficient for tracking an unstable source across a noisy channel.
- Quality of Service Interpretation.