Control, feedback, and the nature of information

Anant Sahai
University of California at Berkeley
sahai@eecs.berkeley.edu
August 24, 2004

Outline

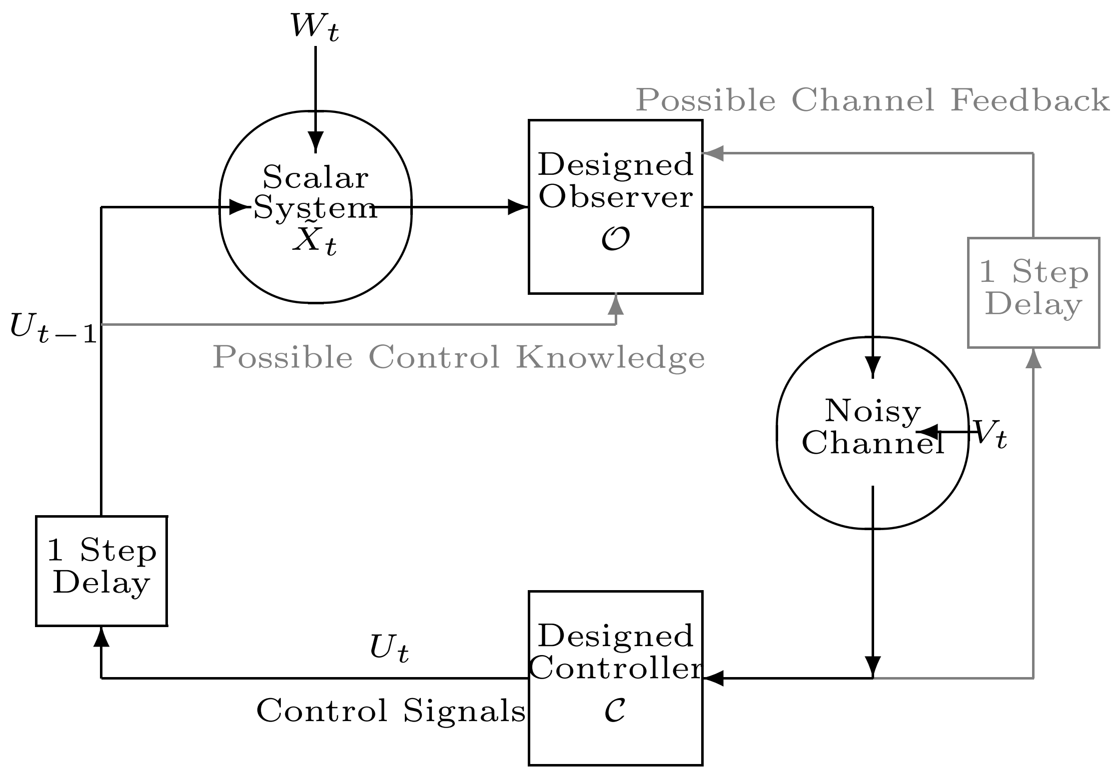
1. Introduction
2. The necessity of feedback anytime reliability for control
3. The relationship of feedback anytime reliability to classical error exponents
4. The nature of information in unstable processes
5. Conclusions and open problems

Big Questions

- Is all information alike?
- How is information produced, stored, and consumed in systems?
- Can we find examples that let us explore the above questions in a concrete setting?

"... can be pursued further and is related to a duality between past and future and the notions of control and knowledge. Thus we may have knowledge of the past and cannot control it; we may control the future but have no knowledge of it." — Claude Shannon 1959

A simple scalar distributed control problem

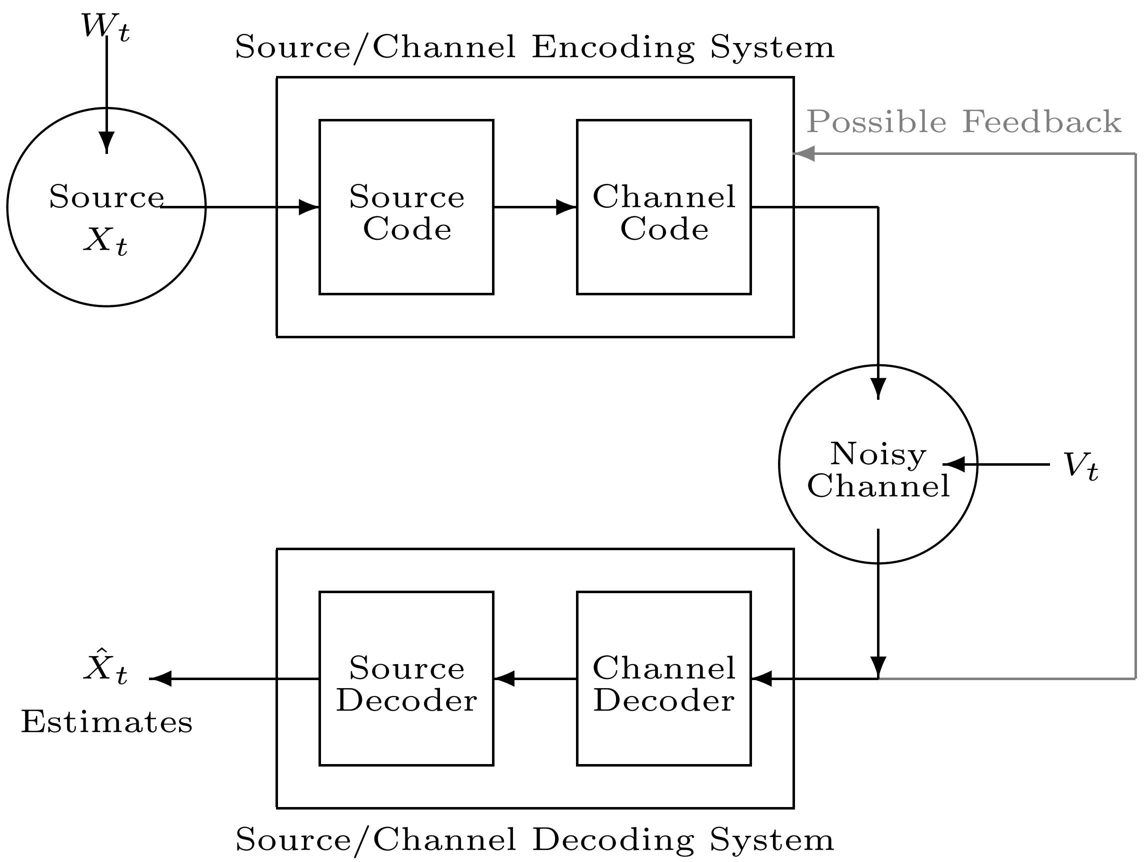


Stabilize the scalar system with $a > 1$

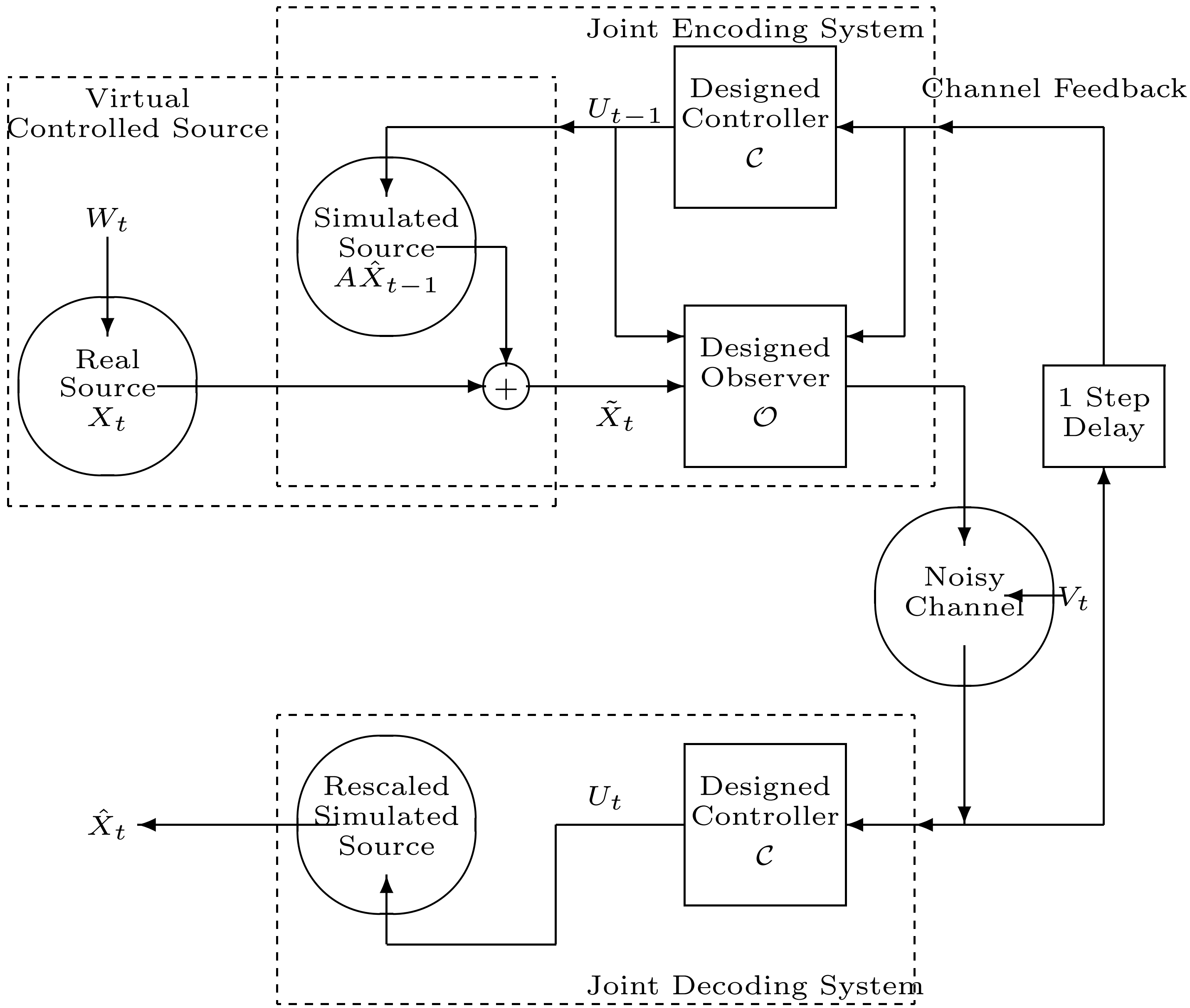
$$\tilde{X}_{t+1} = a\tilde{X}_t + U_t + W_t$$

$\tilde{X}_0 = 0$ and bounded disturbance $\{W_t\}$ with $|W_t| \leq \frac{\Omega}{2}$

The core estimation problem



- Scalar Markov Source $X_{t+1} = aX_t + W_t$ with $a > 1$, $X_0 = 0$ and i.i.d. W_t having bounded support $|W_t| \leq \frac{\Omega}{2}$
- Goal: Keep $\sup_{t>0} E[(X_t - \hat{X}_t)^\eta] \leq D < \infty$


Joint Encoding System
Virtual
Controlled Source
U_{t-1}
Designed
Controller
$\mathcal{C}$
Channel Feedback
W_t
Simulated
Source
$A\hat{X}_{t-1}$
Real
Source
X_t
Designed
Observer
$\mathcal{O}$
$\tilde{X}_t$
1 Step
Delay
Noisy
Channel
V_t
Rescaled
Simulated
Source
U_t
Designed
Controller
$\mathcal{C}$
$\hat{X}_t$
Joint Decoding System

Outline

1. Introduction
2. The necessity of feedback anytime reliability for control
3. The relationship of feedback anytime reliability to classical error exponents
4. The nature of information in unstable processes
5. Conclusions

Want a Shannon's eye view

- Look at the total information flow problem
- See if we can factor it through "bits"
  - Source code into a bitstream
  - Channel code the bitstream across the channel
  - *Communication constraint = limit on how many bits we can transport through the channel per unit time, and how reliably we can do it.*
- Make sure the pieces fit together

Is capacity all we need?

- Consider a system with
  - $a = 2$ for the dynamics
  - noisy channel that sometimes drops packets but is otherwise noiseless (Real erasure channel)

$$Z_t = \begin{cases} Y_t & \text{with Probability } \frac{1}{2} \\ 0 & \text{with Probability } \frac{1}{2} \end{cases}$$

- No other constraints, so design is obvious: $Y_t = X_t$ and $U_t = -aZ_t$
- Resulting closed loop dynamics:

$$\tilde{X}_{t+1} = \begin{cases} W_t & \text{with Probability } \frac{1}{2} \\ 2\tilde{X}_t + W_t & \text{with Probability } \frac{1}{2} \end{cases}$$

- Is it stable?

What should stability mean in such a setting?

- Dealing with uncertainty
  - Noisy channel is stochastic.
  - Disturbance W_t is either worst-case or stochastic.
- Definition 1: state never leaves a box.
- Definition 2: state has a bounded second moment no matter what the disturbance does.
- General definition: $P(|\tilde{X}_t| > x) \leq f(x)$ for all t and W. Reasonable to expect $\lim_{x \to \infty} f(x) = 0$.

$$E[|X|^\eta] \leq K \quad \text{implies} \quad P(|X| > x) \leq K x^{-\eta}$$
$$P(|X| > x) \leq K x^{-\eta(1+\delta)} \quad \text{implies} \quad E[|X|^\eta] \leq K'$$

Is the closed-loop system stable?

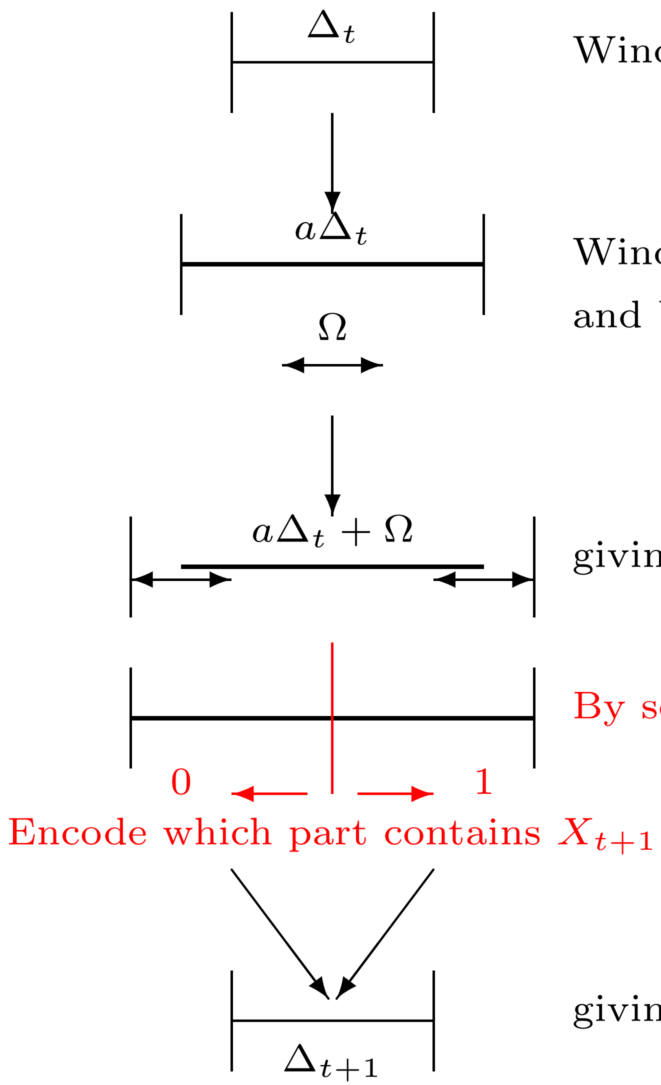
$$\tilde{X}_{t+1} = \begin{cases} W_t & \text{with Probability } \frac{1}{2} \\ 2\tilde{X}_t + W_t & \text{with Probability } \frac{1}{2} \end{cases}$$

- i.i.d. erasures mean arbitrarily long stretches of erasures are possible, though unlikely.
  - System is not guaranteed to stay inside any box.
  - Under even stochastic disturbances, the variance of the state is asymptotically infinite.
- For worst case disturbances $W_t = 1$, the tail probability is dying off as $P(|X| > x) \approx \frac{K}{x}$.
- Meanwhile, $C = \infty$!

Noiseless, but rate-constrained feedback channels

- The feedback link has perfectly reliable deliveries but of only R bits at a time.
- To see that we can keep the system inside a box, look at estimation problem.

Causal Code: Keep X_t In A Box



Window around $\hat{X}_t$ known to contain X_t

Window grows by factor of $a > 1$ due to the dynamics and by a constant due to the bounded driving noise W_t

giving rise to larger window of uncertainty for X_{t+1}

By sending n bits, we cut the window by a factor of 2^{-n}

giving a new window around the updated estimate $\hat{X}_{t+1}$

- If $R > \log_2 a$, we can keep the Δ_t bounded

Classical senses of reliability and capacity

- At the fundamental level, there are senses of reliability
  - Zero-Error: with some fixed delay, we can decode every bit perfectly.
  - ϵ-Error: with some known delay, we can decode every bit with probability of error at most ϵ.
- A sense of reliability has an operational capacity associated with it: the supremal bit-rate for which encoders and decoders exist which satisfy the particular sense of reliability.
- Shannon's capacity is operationally thought of as:

$$C = \lim_{\epsilon \longrightarrow 0} C_\epsilon$$

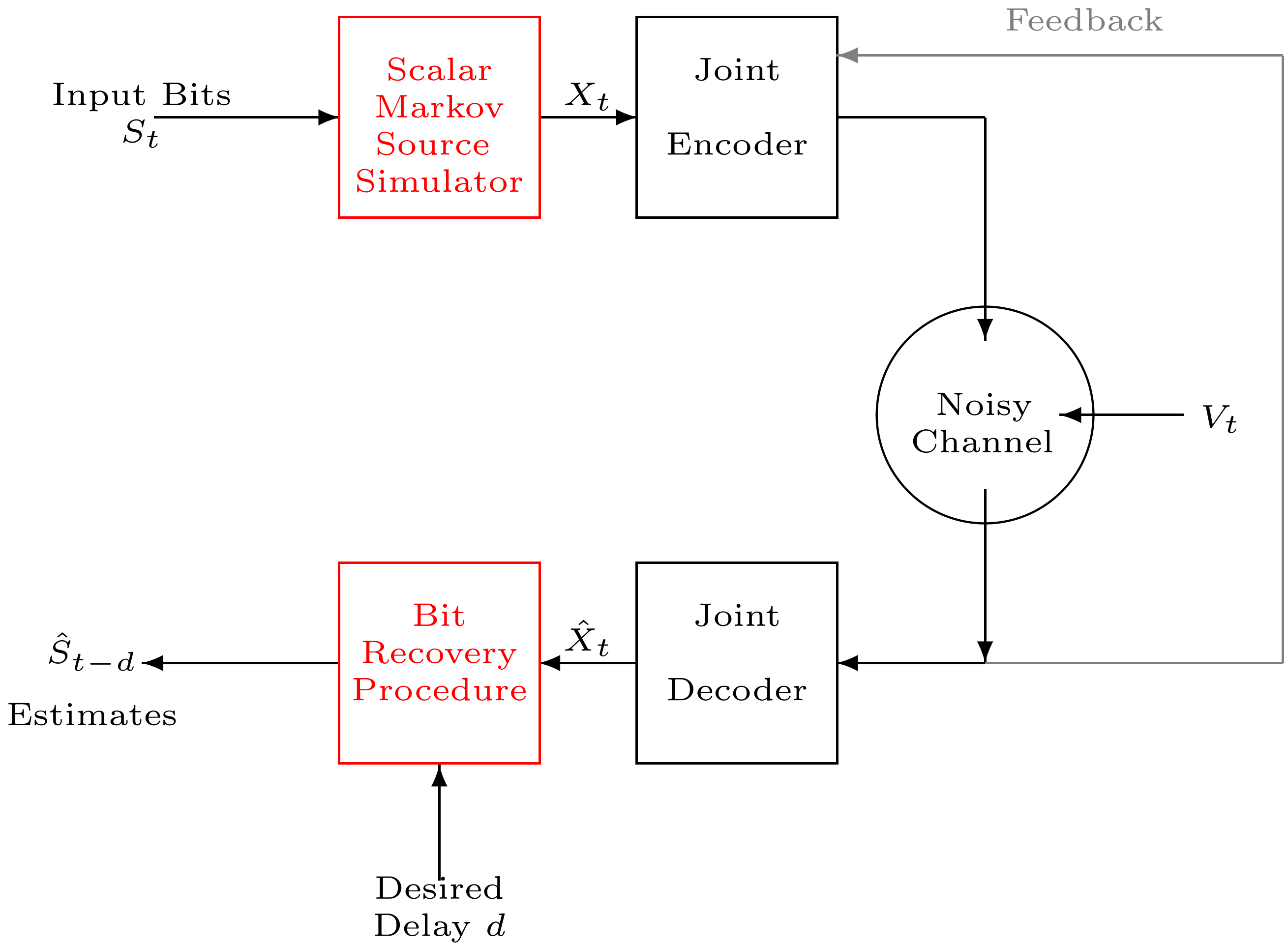
- Sometimes, there is a way of calculating capacity directly. e.g. $C = \sup_{P(A)} I(A; B)$ lets us calculate Shannon capacity for memoryless channels from A to B.

Inadequacy of classical notions

- The real erasure channel has $C_0 = 0$ and $C = \infty$, while the noiseless channel had $C_0 = C > 1$.
- Shows that classical Shannon capacity is not adequate to determine stability in the moment sense.
- Exercise: show that zero-error capacity is sufficient for "stability-in-a-box."
- But is zero-error reliability necessary if we are willing to settle for stability in the η-moment?

Key intuition

- Suppose $a = 2$ and so $X_t = \sum_{i=0}^{t} 2^i W_{t-1}$
- Assume W_j either 0 or 1
- In binary notation: $X_t = W_0 W_1 W_2 \ldots W_{t-1}.00000\ldots$
- If $\hat{X}_t$ is close to X_t, their binary representations likely agree in all the high-order bits.
  - High-order bits correspond to the impact of earlier disturbances.
  - Typically, to get a difference at the W_{t-d} level, we have to be off by about 2^d.
- Estimating the state well is equivalent to communicating bits reliably.

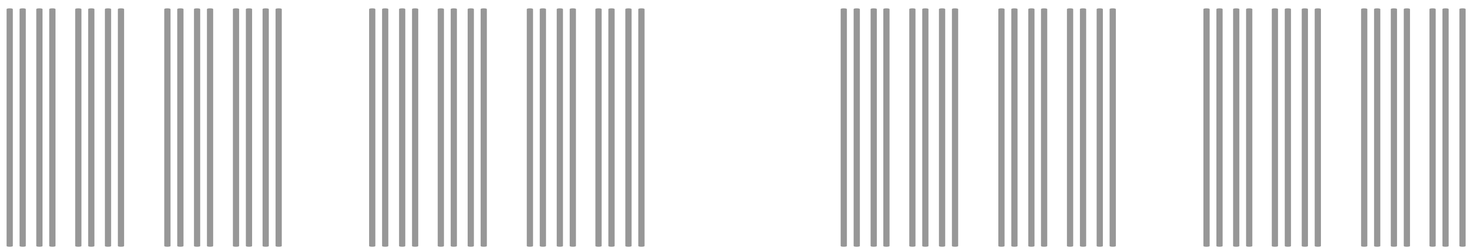


- Use the input bits to drive a source simulator whose output looks like the unstable Markov source ($X_{t+1} = aX_t + W_t$)
- Exploit the source non-ergodicity to get our bits back out.

Binary strings and Cantor sets

- Map the input bitstream bijectively into a Cantor set

$$\check{X} = \sum_{i=0}^{\infty} S_i (2 + \epsilon_1)^{-i}$$

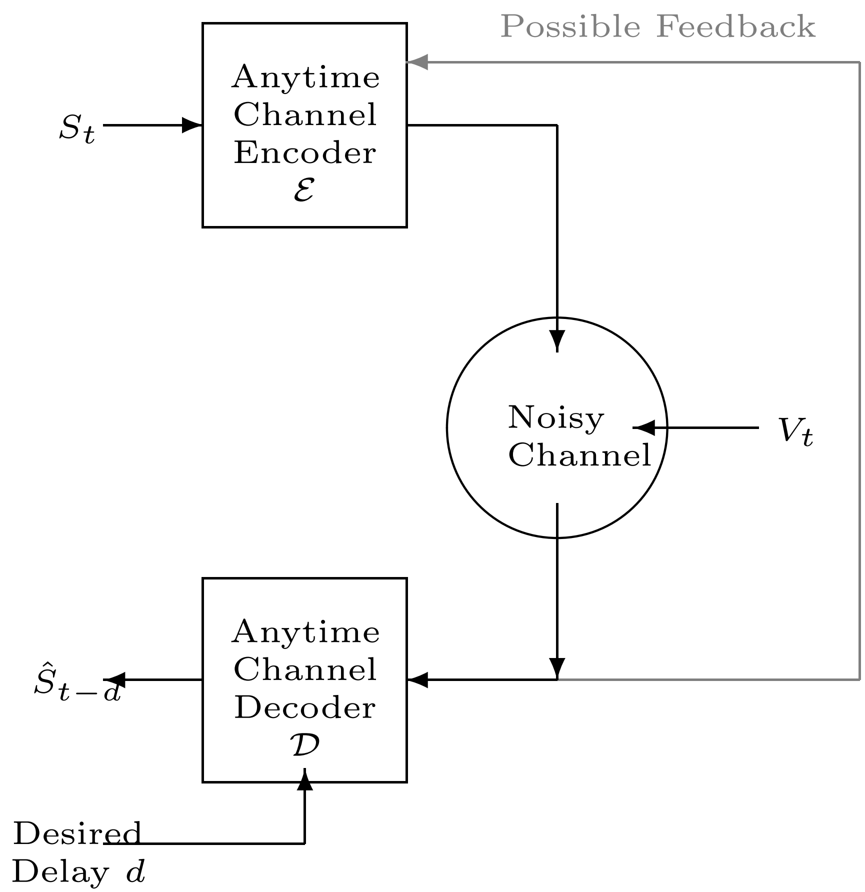


- Embed a suitably scaled, but growing, Cantor set in the unstable $\{X\}$ process
  - Every value for X_t corresponds to a specific neighborhood of the Cantor set
  - Use comparisons to recover the original bits from $\hat{X}_t$. The gaps in the Cantor set give us the ability to distinguish reliably!

What sense of reliability is achieved?

- The gaps in the Cantor set assure us that all $\hat{X}_t$ that differ in their estimates of S_{t-d} are at a distance of at least γa^d from X_t.
  - $|X_t - \hat{X}_t| < \gamma a^d$ implies all bits recovered from $\hat{X}_t$ are correct up through d time steps ago.
  - So $P(S_{t-d} \neq \hat{S}_{t-d}(t)) \leq f(\gamma a^d)$
- Fresh estimates of all bits sent so far.
  - If $E[|\hat{X} - X|^\eta]$ is finite, the probability of error on a bit d time-steps ago is at most $K' a^{-\eta d} = K' 2^{-(\eta \log_2 a) d}$.
  - The reliability of every bit gets better the more we are willing to wait!

Anytime reliable transmission



- Have a fixed encoder, but let the decoder be parametrized by the delay. Want a good estimate "anytime" we ask for one.

- “Reliable Transmission” means every bit is eventually correctly received. Parametrize by the rate at which the probability of bit error $P(S_{t-d} \neq \hat{S}_{t-d}(t))$ goes to zero as delay d increases.

$$C_{\text{anytime}}(\alpha) = \sup \left\{ R \;\middle|\; \begin{array}{l} \exists(\mathcal{E}, \mathcal{D}, K)\ \forall d > 0 \\ \text{Rate} = R, \text{Delay} = d, \\ P_{\text{error}}(\mathcal{E}, \mathcal{D}, d) \leq K 2^{-\alpha d} \end{array} \right\}$$

- Can interpret as the decoder emits not just an estimate for the last bit, but corrections to estimates of previous bits as well.

Separation theorem for control (and estimation)

Necessity: If a scalar system with parameter $a > 1$ can be stabilized with finite η-moment across a noisy channel, then the channel with feedback must have

$$C_{\text{anytime}}(\eta \log_2 a) \geq \log_2 a$$

Sufficiency: If there is an $\delta > 0$ for which the channel with feedback has

$$C_{\text{anytime}}(\eta \log_2 a + \delta) > \log_2 a$$

then the scalar Markov system with parameter $a \geq 1$ with a bounded disturbance can be stabilized across the noisy channel with finite expected η-moment by using observers that have noise-free access to the control signals and channel outputs.

Sufficiency

- Simple sufficiency is left as an exercise.
  - Use classical separation of control and estimation using access to the controls at the observer.
  - Glue together the noise-free encoder with the anytime channel code.
  - Have the controller correct for the impact of any known prior decoding errors.
- Sufficiency can be extended to certain non-perfect information patterns. (CDC 2004)
  - Communicate from controller to encoder through the plant.
  - Have the controller keep these communicating wiggles bounded.

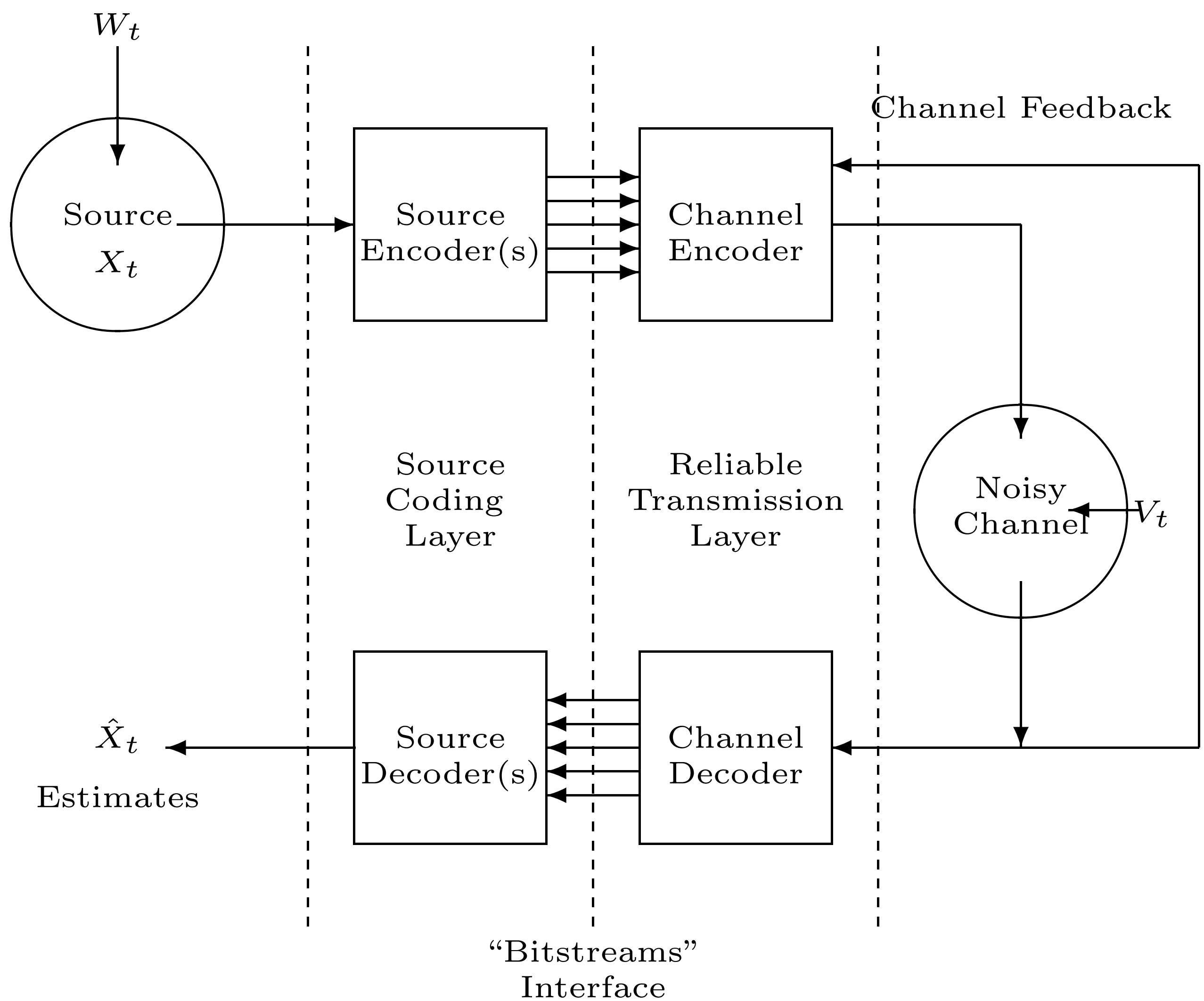
Comments and consequences

$$P(S_{t-d} \neq \hat{S}_{t-d}(t)) \leq f(\gamma a^d)$$

- Applied to an average power-constrained AWGN channel, the construction immediately gives a feedback anytime reliability that is doubly exponential! (*a la* Schalkwijk and Kailath)
- For general DMCs, feedback anytime reliability is usually no better than exponential. This implies that f can have no better than a power-law tail. *Some sufficiently large moment will become infinite.*
- If we demand that X_t stays in a box, then $f(x) = 0$ for large, but finite x. So zero-error reliability is required to achieve "stay-in-the-box" stability or to have stability with an upper bound on actuator effort.

The vector case: differentiated service

- Possibly many unstable eigenvalues
- All unstable eigenspaces need to be estimated with eventually zero error
- Some bits are more important than others.
- Instead of a single α and a single rate R, we get a vector $\vec{\alpha}$ and a rate vector $\vec{R}$.
- Direct and converse both hold on an eigenvalue by eigenvalue basis.


W_t
Source
X_t
Source
Encoder(s)
Channel
Encoder
Channel Feedback
Source
Coding
Layer
Reliable
Transmission
Layer
Noisy
Channel
V_t
$\hat{X}_t$
Estimates
Source
Decoder(s)
Channel
Decoder
"Bitstreams"
Interface

Outline

1. Introduction
2. The necessity of feedback anytime reliability for control
3. The relationship of feedback anytime reliability to classical error exponents
4. The nature of information in unstable processes
5. Conclusions

Reliability functions: conceptual difference

- Classical error exponents
  - Study the rate at which probability of error goes to zero *as we change the encoder block-length.*
  - Considered a way to study encoder/decoder complexity.
  - Internal to channel-coding, do not appear at interface to source-coding.
- Anytime reliability α
  - Requires the probability of error to go to zero at this rate *for a particular encoder.* (ie. The encoder has to be universal over the decoder's acceptable delay and achieve good performance for every delay simultaneously.)
  - Indexes a sense of reliable transmission that cares about delay.
  - Exists at the Source/Channel interface.

Classical reliability functions and feedback

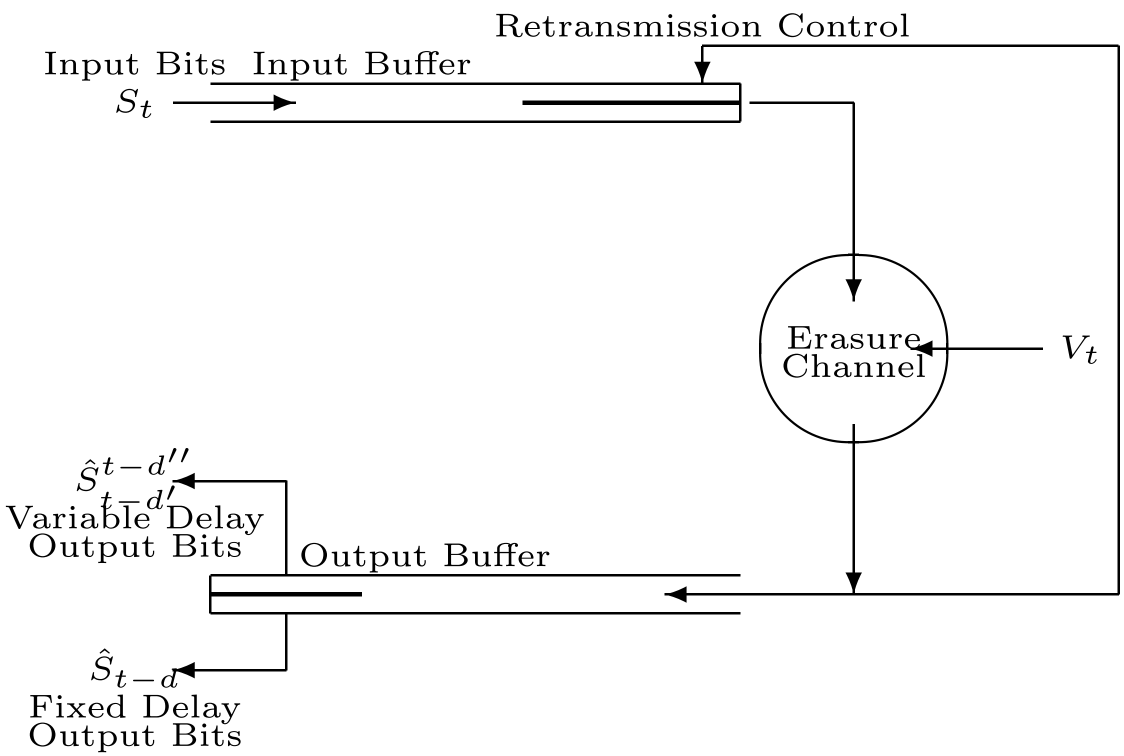
$$E(R) = \lim_{N\to\infty} \sup_{\mathcal{E},\mathcal{D}} \frac{-\log P(\text{block error})}{N}$$

- Average power constrained AWGN
  - Without feedback: Finite, tends to zero as $R \to C$.
  - With noiseless feedback: Infinite, doubly-exponential at all $R < C$
- Binary erasure channel
  - Without feedback: Finite, tends to zero as $R \to C$.
  - With feedback: same as without feedback, at least when R close to C.

Why is the BEC classical reliability unchanged with feedback?

- At rate $R < 1$, have RN bits to transmit in N channel uses.
- Typically $(1-\epsilon)N$ code bits will be received.
- Block errors caused by atypical channel behavior.
  - Doomed if fewer than RN bits arrive intact.
  - *Feedback can not save us.*
  - For rates close to capacity, random coding achieves this bound.
- Dobrushin showed that this type of behavior is common.
  - For sufficiently symmetric channels, the sphere-packing bound $E_{sp}(R)$ is unchanged with feedback.
  - Block-structure is too rigid.

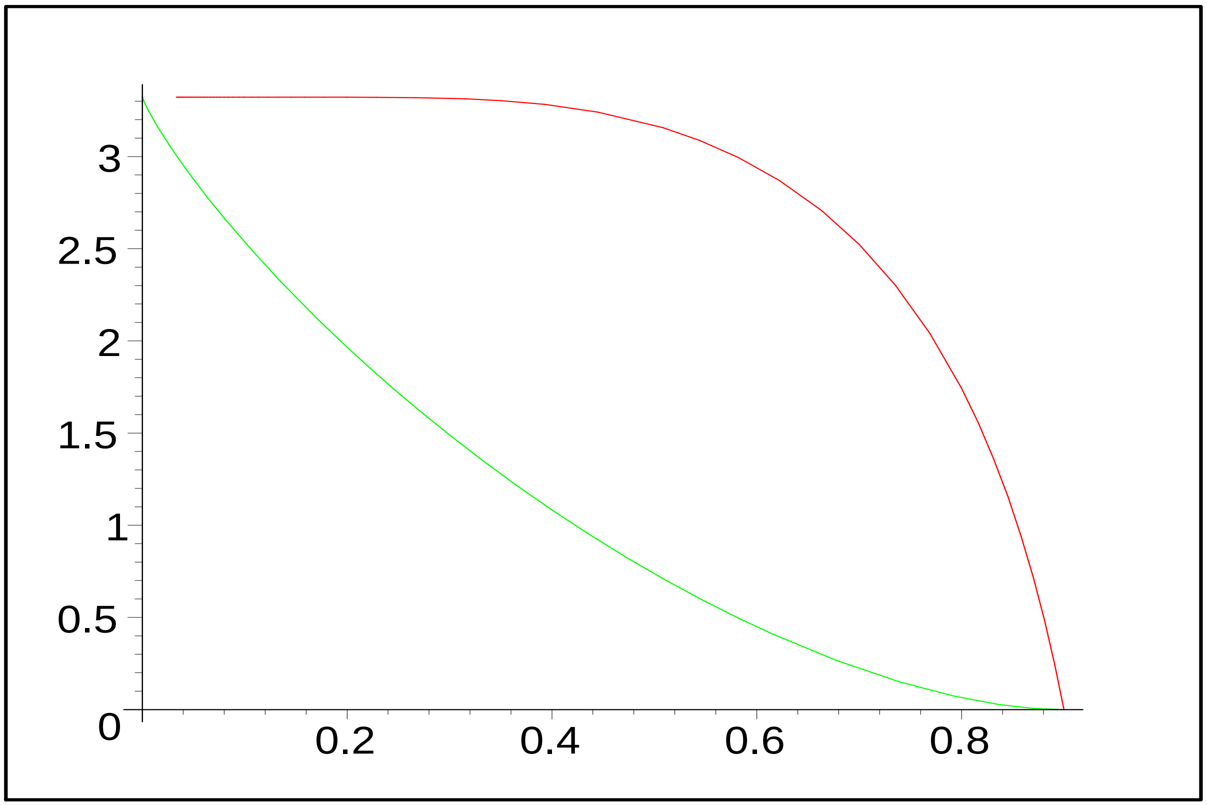
Erasure with feedback



- No block structure to the natural code.
- The probability of error with delay d drops exponentially with d. The further we are from capacity, the faster it drops.
- If $R < 1 - \epsilon$, every bit gets through eventually

Compare BEC feedback anytime reliability to classical E_{sp}

$$C_{\text{anytime}}(\alpha) = \frac{\alpha}{\alpha + \log_2\left(\frac{1-\epsilon}{1-\epsilon 2^{\alpha}}\right)}$$



Using E_{sp} to bound α^*

- Use a rate R anytime-code to make a block code of rate $R' = (1-\lambda)R$ where $\lambda \in [0,1]$.
  - Take $R'N$ bits of data and consider them the first bits to arrive at the anytime encoder.
  - For the rest of the data bits (taking time λN), just choose 0.
- The block error probability is bounded by $K2^{-\alpha\lambda N}$ which has to meet the sphere-packing bound $2^{-E_{sp}((1-\lambda)R)N}$

$$\alpha^*(R) \leq \frac{E_{sp}((1-\lambda)R)}{\lambda}$$

Bound for symmetric DMCs

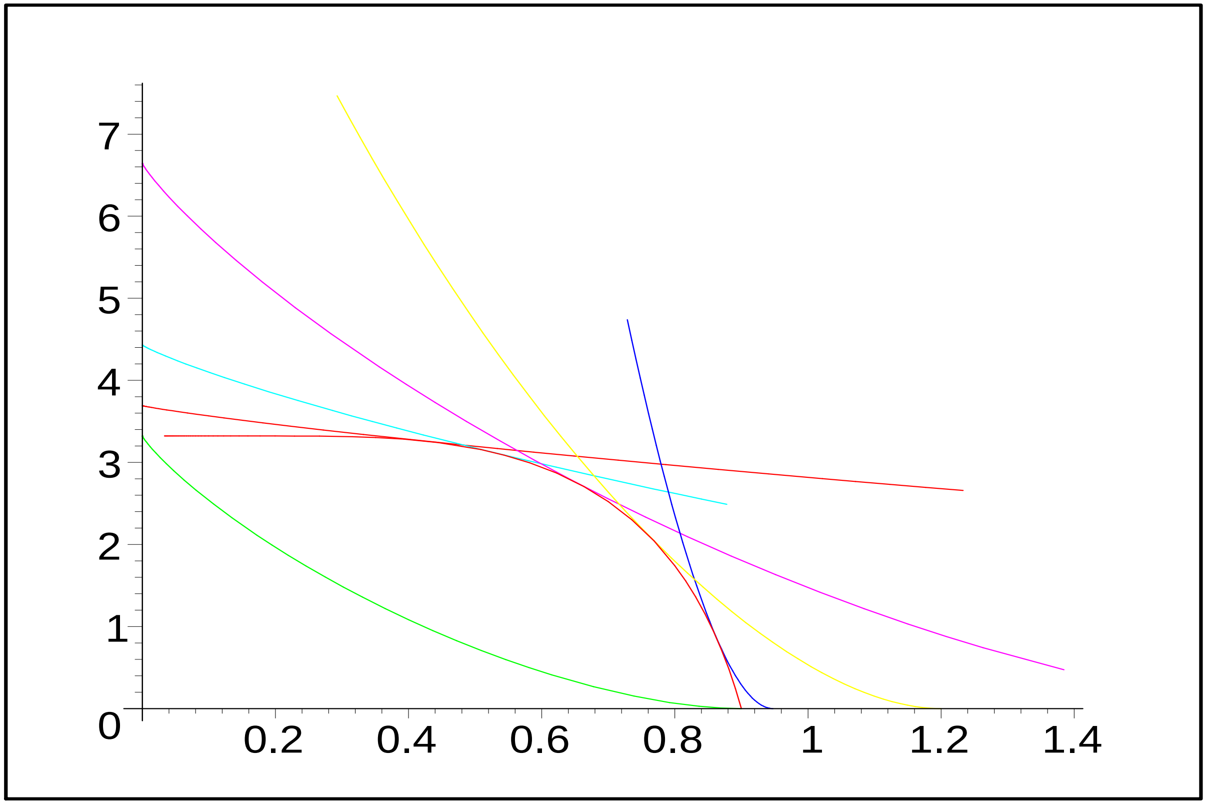
Minimize over λ for symmetric DMCs to sweep out frontier by varying $\rho > 0$:

$$R(\rho) = \frac{E_a^+(\rho)}{\rho}$$

$$E_a^+(\rho) = -\max_q \log_2 \sum_j \left(\sum_i q_i p_{ij}^{\frac{1}{1+\rho}} \right)^{1+\rho}$$

The same basic form as the sphere-packing bound, but convex $\cap$ instead of convex $\cup$.

This upper bound is tight for the BEC with feedback



Known anytime reliabilities with feedback

Characterizing the boundary of possible (α, R) pairs:

- L-bit packet erasure channel

$$C_{\text{anytime}}(\alpha) = \frac{\alpha L}{\alpha + \log_2(\frac{1-\epsilon}{1-\epsilon 2^{\alpha}})}$$

- Variable-sized packet erasure channel with expected packet-size constrained to be $\bar{L}$ and maximum packet-size L_{max} (Allerton 2004)

$$C_{\text{anytime}}(\alpha) = \min\left((1-\epsilon)\bar{L}, \frac{\alpha L_{max}}{\alpha + \log_2(\frac{1-\epsilon}{1-\epsilon 2^{\alpha}})}\right)$$

- Power-constrained AWGN

$$C_{\text{anytime}}(\alpha) = \frac{1}{2}\log_2(1 + \frac{P}{\sigma^2})$$

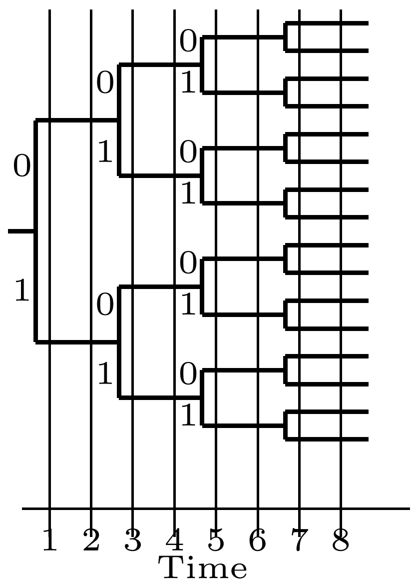
- Power-constrained AWGN+erasure (Allerton 2004)

$$C_{\text{anytime}}(\alpha) = \begin{cases} \frac{(1-\epsilon)}{2}\log_2\left(1 + \frac{P}{\sigma^2}\right) & \text{if } 0 \leq \alpha < -\log_2 \epsilon \\ 0 & \text{otherwise} \end{cases}$$

$$\alpha^*(R) = \begin{cases} -\log_2 \epsilon & \text{if } R < \frac{(1-\epsilon)}{2}\log_2\left(1 + \frac{P}{\sigma^2}\right) \\ 0 & \text{otherwise} \end{cases}$$

It is the tradeoff between α and R that is important, which way to write it is a matter of convenience.

$C_{\textbf{anytime}}(\alpha) > 0$ even without feedback!



- Consider an infinite binary tree, with random labels
  - Choose a path through the tree based on the bits to be sent
  - Transmit the labels along the path through the channel
  - Let the decoder do maximum likelihood decoding
- Achieves the random coding block error exponent $E_r(R)$ for every d and hence has $\alpha(R) \geq E_r(R)$ for every $R < C$

Comments and consequences

- Since feedback can only help, we now know that

$$E_r(R) \leq \alpha^*(R) \leq E_a^+(R)$$

- Positive at all rates below C_{Shannon}
- If $C_{\text{Shannon}} > \log_2 a$, it is possible to stabilize the scalar system for some moment η.

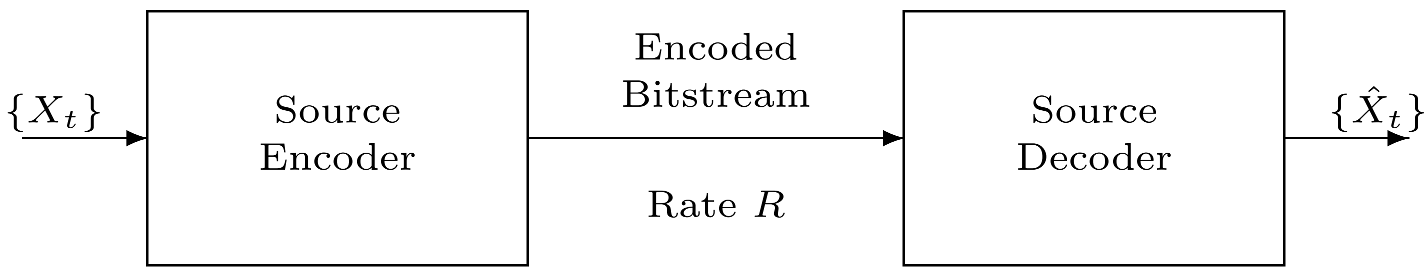
How does feedback help reliability?

- We conjecture that the classical sphere-packing bound continues to bound the anytime-reliability without feedback.
- With feedback, the anytime code has the opportunity to focus its efforts on the earlier bits by postponing the encoding of later ones.
- This phenomenon is also what makes possible the higher feedback reliability in the expected block-length context of Burnashev and Yamamoto-Ito.
- Initial investigations suggest that anytime codes without feedback can be used as part of a total system that involves noisy feedback. (ITW 2004)

Outline

1. Introduction
2. The necessity of feedback anytime reliability for control
3. The relationship of feedback anytime reliability to classical error exponents
4. The nature of information in unstable processes
5. Conclusions

The source coding problem



- Translate ongoing signal ($X_{t+1} = aX_t + W_t$, $a \geq 1$) to bits
- For a given rate R, we would like to minimize average $E[(X_t - \hat{X}_t)^2]$ even as $t \to \infty$
- Encoder and decoder need to work "online," but we are willing to tolerate some large, but finite, delay between X_t and $\hat{X}_t$

Performance bounds

- Distortion-Rate function as limit of finite horizon problems:

$$D(R) = \liminf_{N \to \infty} \inf_{\substack{p(\hat{X}_1^N | X_1^N) \\ \frac{1}{N} I(X_1^N; \hat{X}_1^N) \leq R}} E\left[\frac{1}{N} \sum_{t=1}^{N} |X_t - \hat{X}_t|^\eta\right]$$

- Sequential Rate Distortion when no delay tolerated: {Tatikonda00}

$$D_{\text{seq}}(R) = \liminf_{N \to \infty} \inf_{\substack{p(\hat{X}_i | X_1^i, \hat{X}_1^{i-1}) \\ \frac{1}{N} I(X_1^N \to \hat{X}_1^N) \leq R}} E\left[\frac{1}{N} \sum_{t=1}^{N} |X_t - \hat{X}_t|^\eta\right]$$

This bounds the performance of feedback control systems.

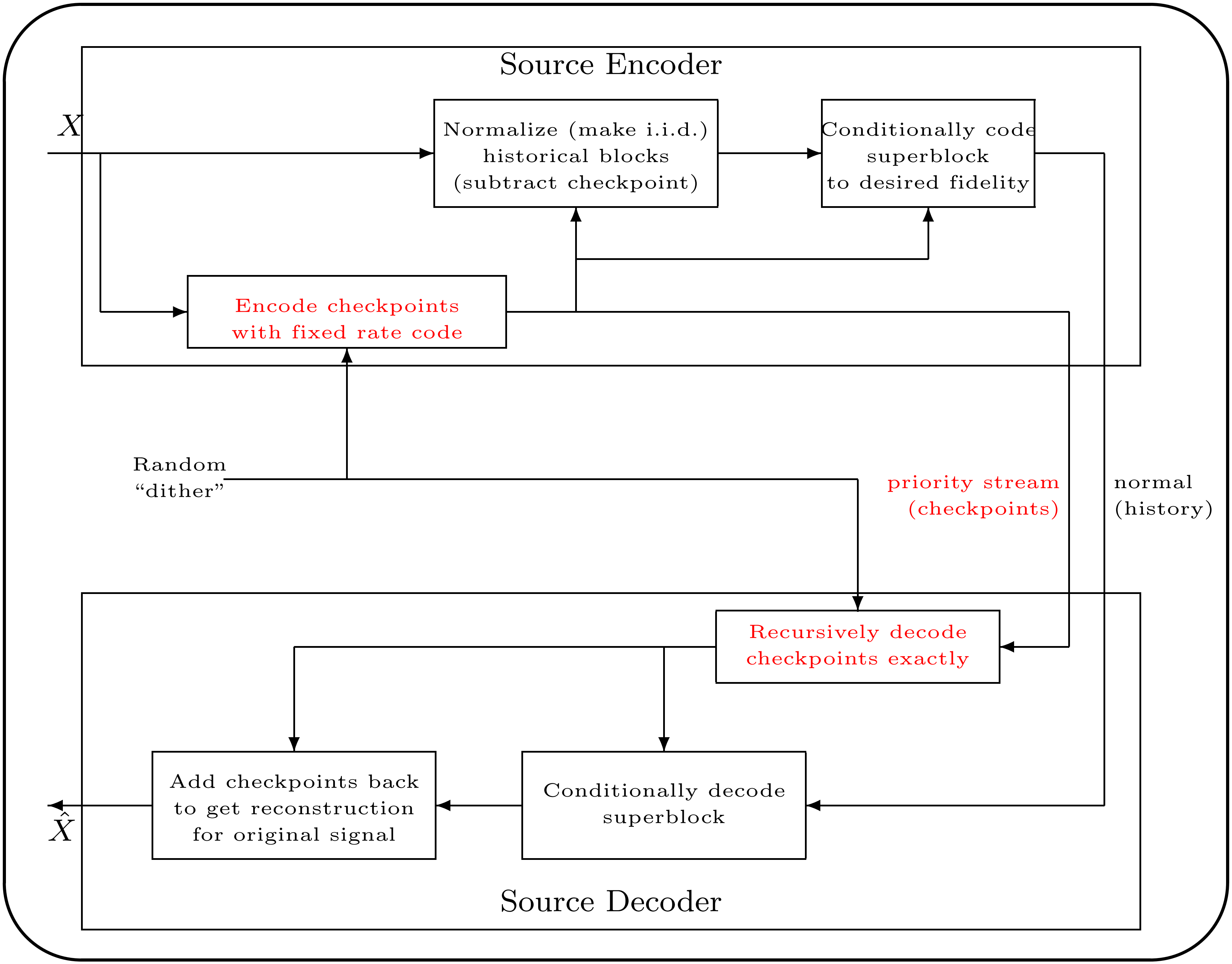
Prior results

- {Gray70} solved finite horizon problem for $a \geq 1$: so we need at least R bits/symbol carried across the channel to achieve $D(R)$ distortion
- {Berger70} solved infinite horizon for $a = 1$
- {Hashimoto80} gave parametric form for $D(R)$ for unstable Gaussian AR processes and squared-error distortion.
- Infinite horizon problem for $a > 1$ remained open till our work.

Two stream encoding

The scalar unstable Markov process can be encoded to distortion arbitrarily close to D using two fixed-rate bitstreams.

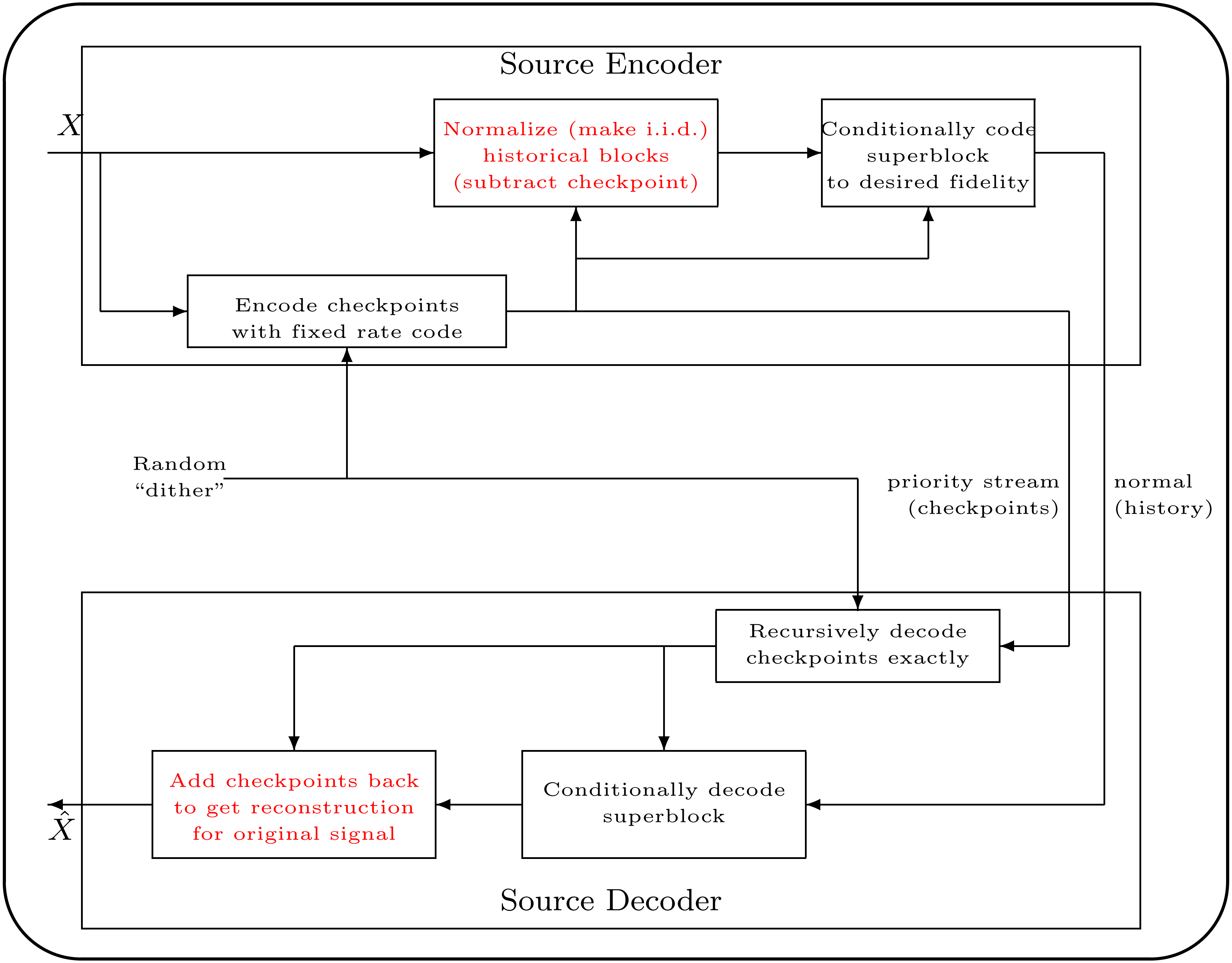
- The sum of the two rates can approach $R(D)$
- The priority bitstream: $R_1 \approx \log_2 a$, and requires *anytime reliability* with $\alpha > \log_2 a$ for noisy channel transport.
- The remaining bitstream only needs reliability in the traditional Shannon sense.



August 24, 2004 at UIUC

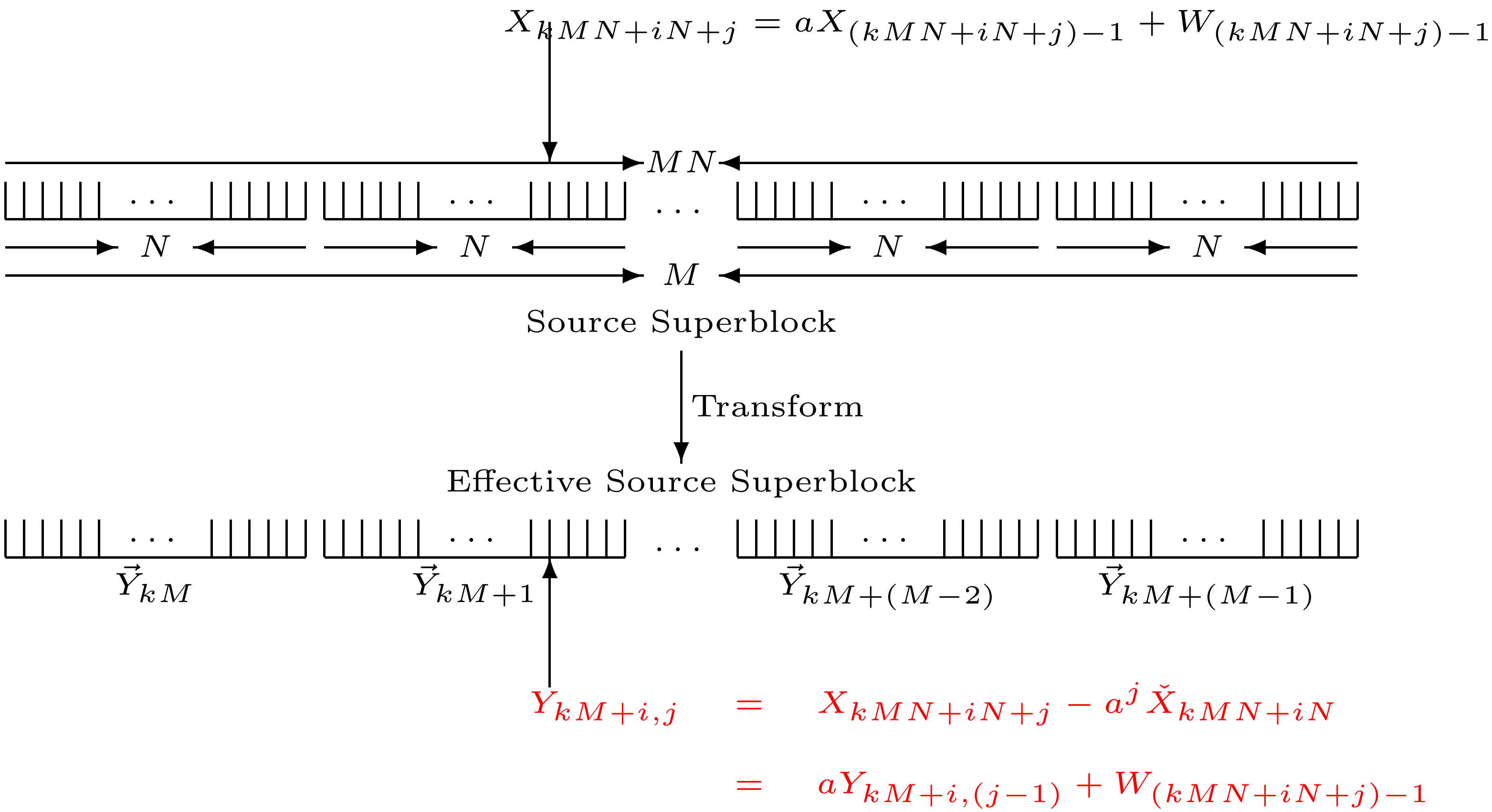
Encoding checkpoints: two-phases

- Phase 1: Build a coarse skeleton of the process using causal code with $R \approx \log_2 a$ — resulting Δ will be large.
- Phase 2: Give higher-fidelity to checkpoints every N samples.
  - Use $N\epsilon_1$ extra bits to cut uncertainty to $\frac{\Delta}{2^{N\epsilon_1}}$.
  - Subtractive dither makes $X_{kN} - \check{X}_{kN}$ small iid uniform rv.
  - By choosing large N, can make the checkpoint $\check{X}_{kN}$ as close to X_{kN} as we would like!

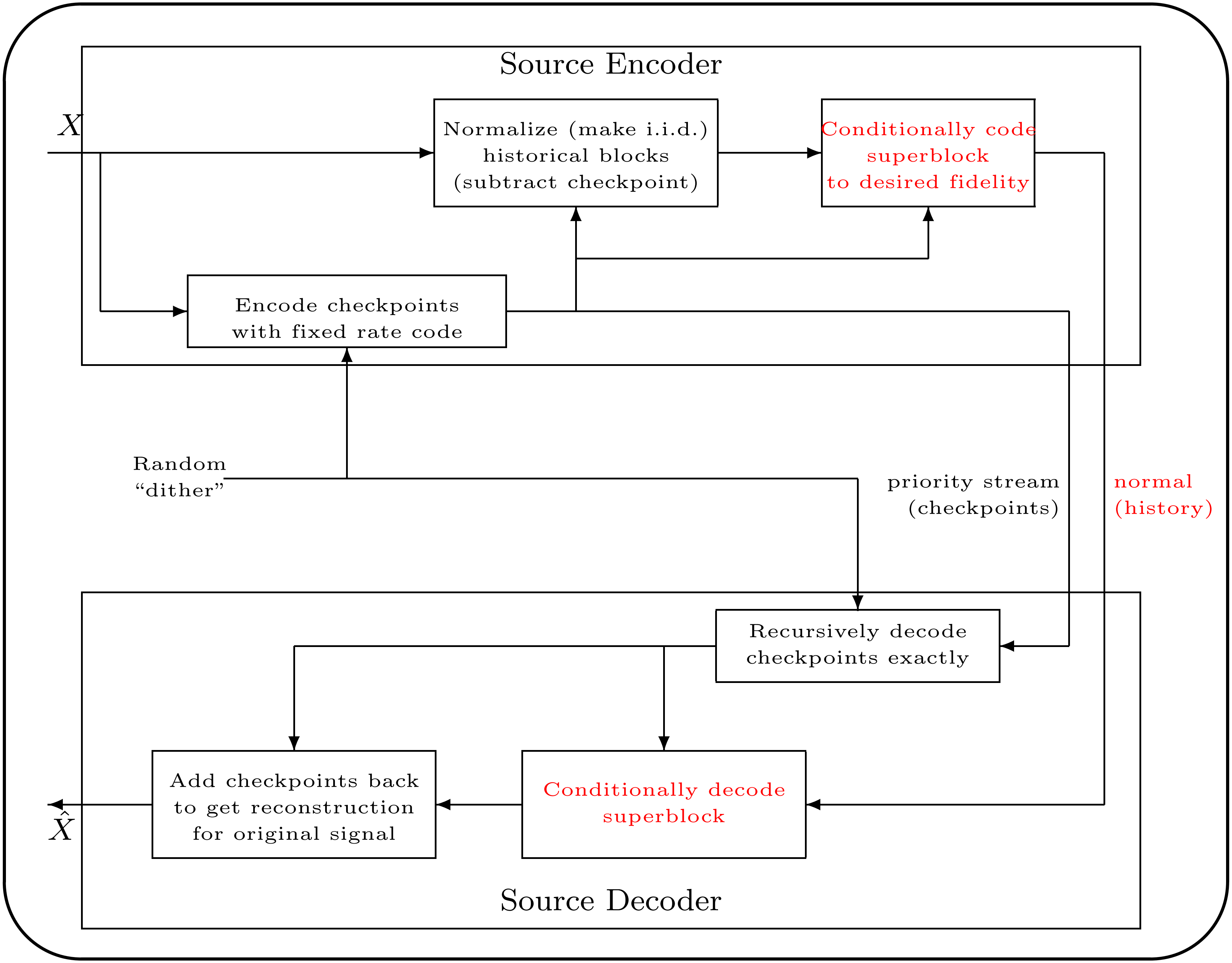


August 24, 2004 at UIUC

How to normalize a superblock



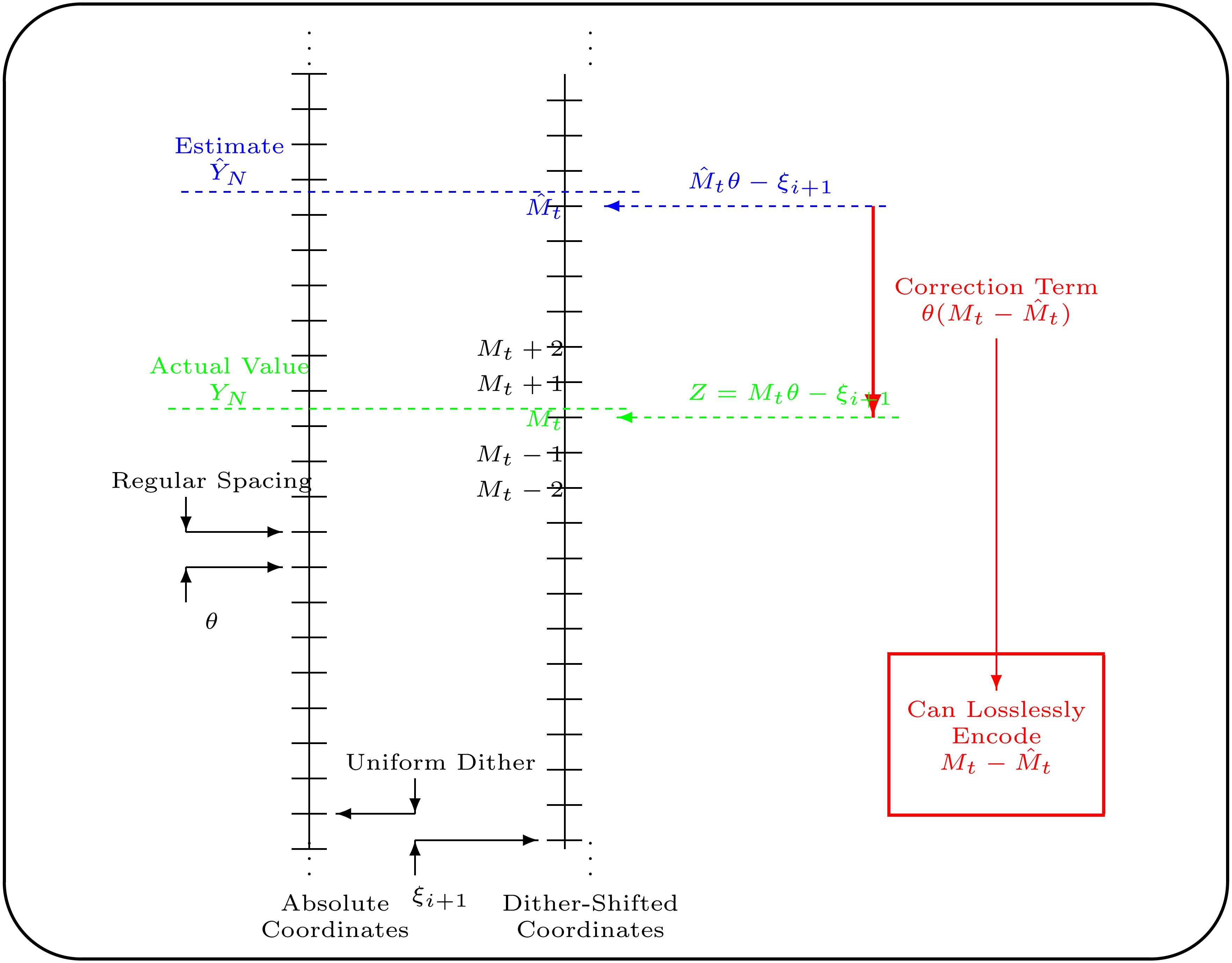
- Use time invariance to make sub-blocks look alike
- Subtract the block's starting checkpoint $\check{X}_t$ (dithered to initialize each to a small uniform initial condition.)



August 24, 2004 at UIUC

Conditional encoding of superblock

- Both encoder and decoder have access to the checkpoints
  - The start of the block checkpoint was used to make the blocks look iid.
  - Use the end of the block checkpoint Z as side information available at both the encoder and decoder.
- Let $(Y_0^N, \hat{Y}_0^N)$ have joint distribution so that $\frac{1}{N}\sum_{i=1}^{N} |Y_i - \hat{Y}_i|^\eta = D + \delta$ and $I(Y_0^N; \hat{Y}_0^N) = N(R + \epsilon_2)$
  - $I(Y_0^N; \hat{Y}_0^N | Z) = I(Y_0^N; \hat{Y}_0^N, Z) - I(Y_0^N; Z)$
  - We know: $I(Y_0^N; Z) = H(Z) - H(Z|Y_0^N) = H(Z) > N \log_2 A$.
  - Need to show: $I(Y_0^N; \hat{Y}_0^N, Z) \leq N(R + \epsilon_2) + o(N)$
  - That would give us: $\frac{1}{N} I(Y_0^N; \hat{Y}_0^N | Z) \leq (R - \log_2 A + \epsilon_2) + \frac{o(N)}{N}$



Estimate
$\hat{Y}_N$
$\hat{M}_t\theta - \xi_{i+1}$
$\hat{M}_t$
Correction Term
$\theta(M_t - \hat{M}_t)$
$M_t + 2$
$M_t + 1$
Actual Value
Y_N
$Z = M_t\theta - \xi_{i+1}$
M_t
$M_t - 1$
$M_t - 2$
Regular Spacing
θ
Can Losslessly Encode $M_t - \hat{M}_t$
Uniform Dither
ξ_{i+1}
Absolute Coordinates
Dither-Shifted Coordinates

Why $I(Y_0^N; \hat{Y}_0^N, Z) \leq N(R + \epsilon_2) + o(N)$

$$
\begin{aligned}
I(Y_0^N; \hat{Y}_0^N, Z) &= I(Y_0^N; \hat{Y}_0^N) + I(Y_0^N; Z|\hat{Y}_0^N) \\
&\leq N(R + \epsilon_2) + H(Z|\hat{Y}_0^N) \\
&\leq N(R + \epsilon_2) + H(Z|\hat{Y}_N)
\end{aligned}
$$

- Since the distortion is finite, the estimates $\hat{Y}$ are close to the actual values Y and $E[|Y_N - \hat{Y}_N|^\eta] \leq K$
- Markov Inequality: the integer correction term $(M_t - \hat{M}_t)$ has at most a power-law tail
- Encoding integers takes logarithmic bits and so the code-length distribution has at most an exponential tail.
- So $H(Z|\hat{Y}_N) \leq E[\text{code-length}] = o(N)$.

The data streams and rate per symbol

- Coarse tracking of X_t: $R_1 = \log A + \epsilon_a$
  - Very sensitive to bit errors.
  - Requires *anytime-reliability* $\alpha > \eta \log_2 A$ to get across the noisy channel.
- Refining of checkpoints $\check{X}_{kN}$: $R_1' = \epsilon_1$
- Conditional history $\hat{Y}_t$: $R_2 = R - \log_2 A + \epsilon_2 + \frac{o(N)}{N}$
  - For R_1', R_2 streams, bit errors can not propagate beyond a single superblock.
  - Standard Shannon bit reliability is sufficient.

Total Rate: $R + \epsilon_a + \epsilon_1 + \epsilon_2 + \frac{o(N)}{N}$

Total Distortion: $D + \delta$

Time-reversal interpretation

- Key Insight: Study history backwards.
  - For $0 \leq i \leq N-1$, let $P_{k,i} = X_{(k+1)N-i} - A^i \check{X}_{(k+1)N}$.
  - $P_{k,i+1} = A^{-1} P_{k,i} - A^{-1} W_{(k+1)N-i-1}$
  - The $P_{k,i}$ process is stable, and is initialized with a tiny uniform random variable for $P_{k,0}$.
- We get an immediate bound for the unstable $D(R)$ for $R > \log_2 A$

$$D(R) \leq D_P(R - \log_2 A)$$

“Phase Transition”

- P_i has finite expected distortion even if $R = 0$.

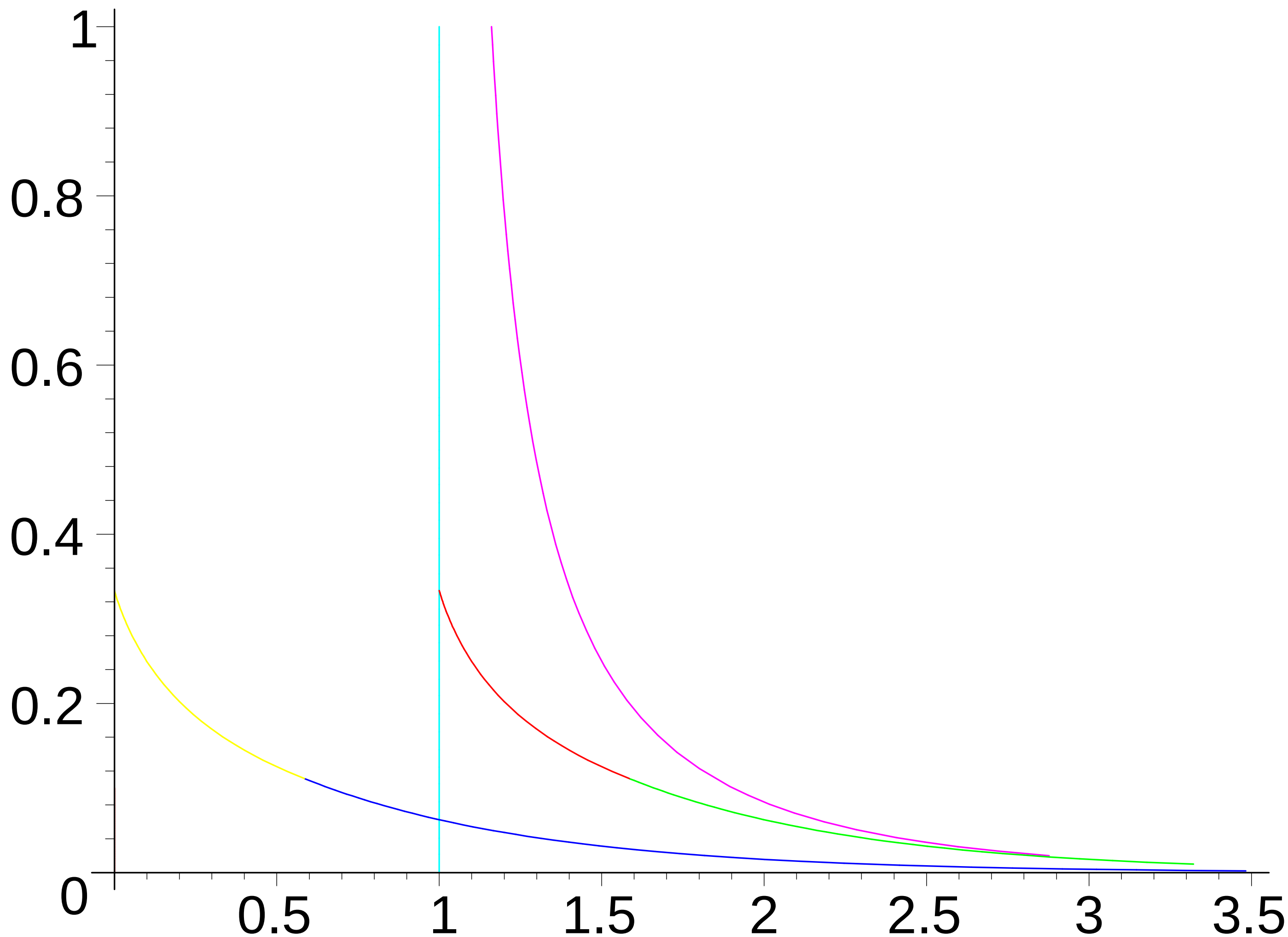
$$\begin{aligned} \lim_{R \to \log_2 A^+} D(R) &\leq \lim_{R \to \log_2 A^+} D_P(R - \log_2 A) \\ &= \lim_{R \to 0^+} D_P(R) \\ &= D_P(0) \end{aligned}$$

- $D(R)$ jumps from infinite to finite distortions at a finite rate!

- In contrast:

$$\lim_{R \to \log_2 A^+} D_{\text{seq}}(R) = \infty$$

- Hashimoto and Arimoto can be reinterpreted to show that in the quadratic Gaussian case, $D(R) = D_P(R - \log_2 A)$.



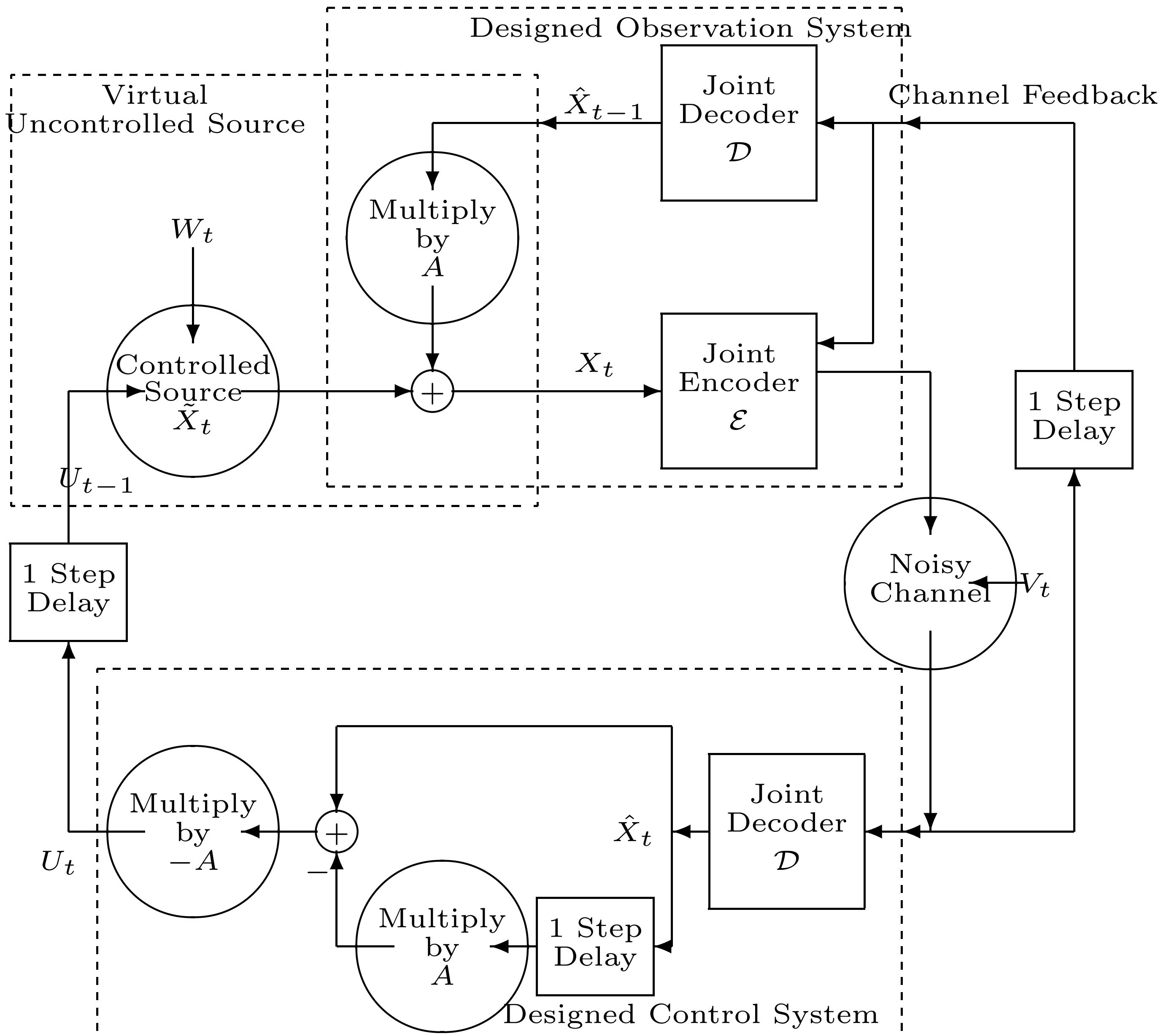
1
0.8
0.6
0.4
0.2
0
0.5
1
1.5
2
2.5
3
3.5

Comments

- Even scalar unstable processes are generating two qualitatively different kinds of information.
  - The high priority bitstream encodes the outline of the process and captures the essential information *accumulating* in the unstable system.
  - The lower priority bitstream encodes the detailed history of the process and captures the desired information that is *dissipating* within the system.
- In vector cases, we get quantitatively different accumulations for the different unstable eigenvalues and then the leftover detailed dissipation.

Conclusions and open problems

- Information theory can *and should* be extended to deal with communication questions relevant to control.
  - Reliability must rise to the source-channel interface.
  - Applications can generate different kinds of information.
- Open problems
  - Need more tools and results on bounding achievable $(\vec{\alpha}, \vec{R})$ regions.
  - Bit-pipes fit together into networks naturally, need analogous understanding for networked reliable bit-pipes
  - Better understand the performance-loss caused by the communication constraint — starting with the gap between $D_{seq}(R)$ and $D(R)$.



Designed Observation System
Virtual
Uncontrolled Source
$\hat{X}_{t-1}$
Joint
Decoder
$\mathcal{D}$
Channel Feedback
W_t
Multiply
by
A
Controlled
Source
$\tilde{X}_t$
X_t
Joint
Encoder
$\mathcal{E}$
1 Step
Delay
U_{t-1}
1 Step
Delay
Noisy
Channel
V_t
Multiply
by
$-A$
$\hat{X}_t$
Joint
Decoder
$\mathcal{D}$
U_t
$-$
Multiply
by
A
1 Step
Delay
Designed Control System