Where do stabilization problems sit in the communication problem hierarchy?

Anant Sahai
University of California at Berkeley
sahai@eecs.berkeley.edu

April 1, 2005

Joint work with: S. Mitter, T. Simsek, Q. Xu, C. Cheng, P. Minero, and A. Avestimehr

Outline

1. Communication problem hierarchy
2. Stabilization problems and anytime communication
3. Core reason: feedback channel coding: block-length is *not* a good proxy for delay

Big questions about communication.

- Are all communication problems asymptotically alike?

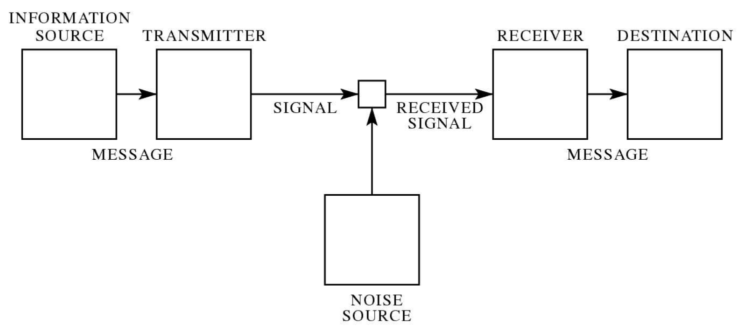
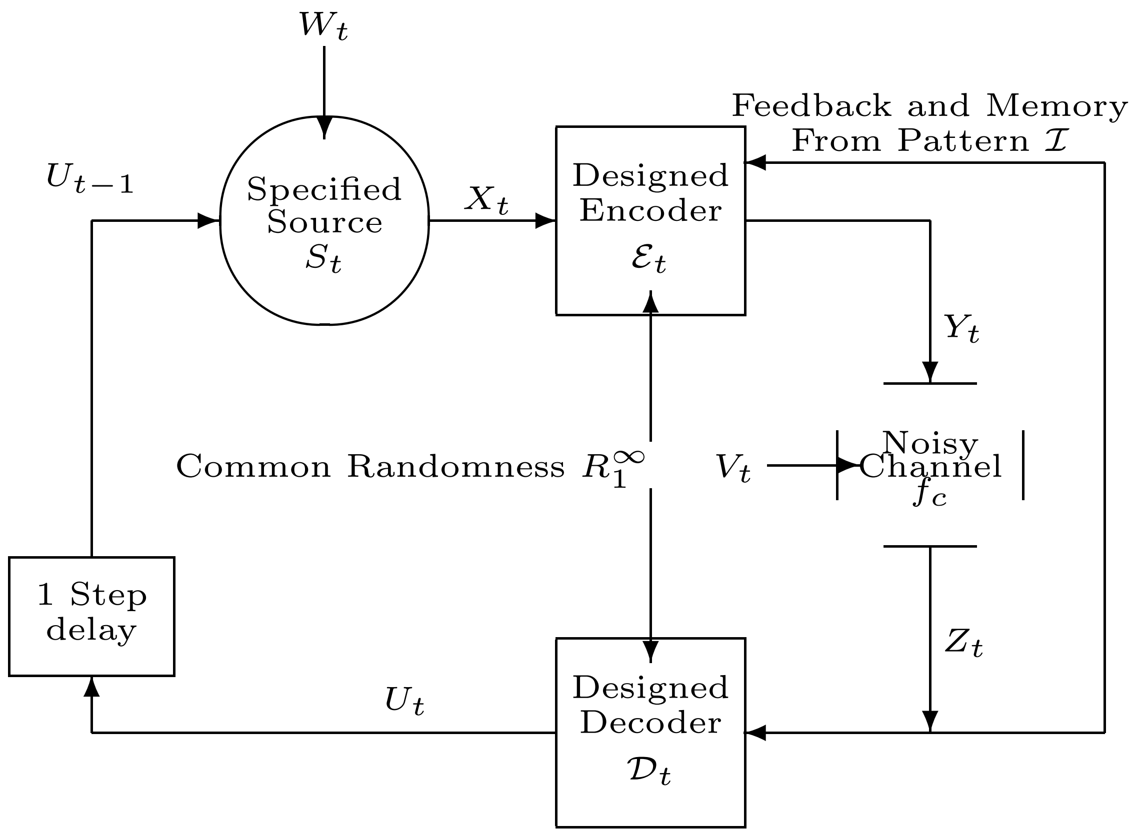


Fig. 1 — Schematic diagram of a general communication system.

- How does delay interact with capacity issues?
- Can we find examples that let us explore these questions in an asymptotic 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

An abstract model of single channel problems



- Problem: Source S, Information pattern $\mathcal{I}$, and Objective $\mathcal{V}$.
- Constrained resource: Noisy channel f_c
- Designed solution: "Encoder" $\mathcal{E}$, "Decoder" $\mathcal{D}$

Focus: what channels are "good enough" for the problem

- f_c *solves* the problem if $\exists \mathcal{E}, \mathcal{D}$ so system satisfies $\mathcal{V}$
- Problem A is *harder* than problem B if any f_c that solves A solves B.
- Information theory is an *asymptotic* theory
  - Pick $\mathcal{V}$ family with an appropriate "slack" parameter
  - Consider the set of channels that solve the problem.
  - Take union over slack parameter choices.

The Shannon problems $A_{R,\epsilon,d}$

- Source: *noninteractive* X_i (R bits): fair coin tosses
- Information pattern: $\mathcal{D}_i$ has access to Z_1^i
  - A^f With feedback: $\mathcal{E}_i$ gets X_1^i and Z_1^{i-1}
  - A^{nf} Without feedback: $\mathcal{E}_i$ gets only X_1^i
- Performance objective: $\mathcal{V}(\epsilon, d)$ is satisfied if $\mathcal{P}(X_i \neq U_{i+d}) \leq \epsilon$ for every $i \geq 0$.
  - Slack parameter: permitted delay d
  - Natural orderings: larger ϵ, d is easier but larger R is harder.
- Classical capacity

$$\mathcal{C}_R^f = \bigcap_{\epsilon>0} \bigcap_{R'<R} \bigcup_{d>0} \{f_c | f_c \text{ solves } A^f_{R',\epsilon,d}\}$$

$$C_{\text{Shannon}}(f_c) = \sup\{R > 0 | f_c \in \mathcal{C}_R\}$$

Classical relationships

- Feedback doesn't change capacity for memoryless channels $\mathcal{C}^m$

$$\mathcal{C}_R^{nf} \cap \mathcal{C}^m = \mathcal{C}_R^f \cap \mathcal{C}^m$$

- Zero-error capacity

$$\mathcal{C}_{0,R}^f = \bigcap_{R' < R} \bigcup_{d > 0} \{f_c | f_c \text{ solves } A_{R',0,d}^f\}$$

$$C_0(f_c) = \sup\{R > 0 | f_c \in \mathcal{C}_{0,R}\}$$

  – Can change with feedback even for memoryless channels

$$(\mathcal{C}_{0,R}^f \cap \mathcal{C}^m) \subset (\mathcal{C}_{0,R}^f \cap \mathcal{C}^m)$$

  – Zero-error problem is fundamentally harder

$$(\mathcal{C}_{0,R}^{nf} \cap \mathcal{C}^m) \subset (\mathcal{C}_{0,R}^f \cap \mathcal{C}^m) \subset (\mathcal{C}_R \cap \mathcal{C}^m)$$

Estimation with distortion: $A_{(F_X,\rho,D,d)}$

- Source: *noninteractive* X_i drawn iid from F_X
- Same information patterns: with/without feedback.
- Performance objective: $\mathcal{V}(\rho, D, d)$ is satisfied if $\lim_{n\to\infty} \frac{1}{n} E[\sum_{i=1}^{n} \rho(X_i, U_{i+d})] \leq D$.
  - Slack parameter: permitted delay d
  - Natural orderings: larger D, d is easier
- Channels that are good enough

$$\mathcal{C}^f_{e,(F_X,\rho,D)} = \bigcap_{D'>D} \bigcup_{d>0} \{f_c | f_c \text{ solves } A^f_{(F_X,\rho,D',d)}\}$$

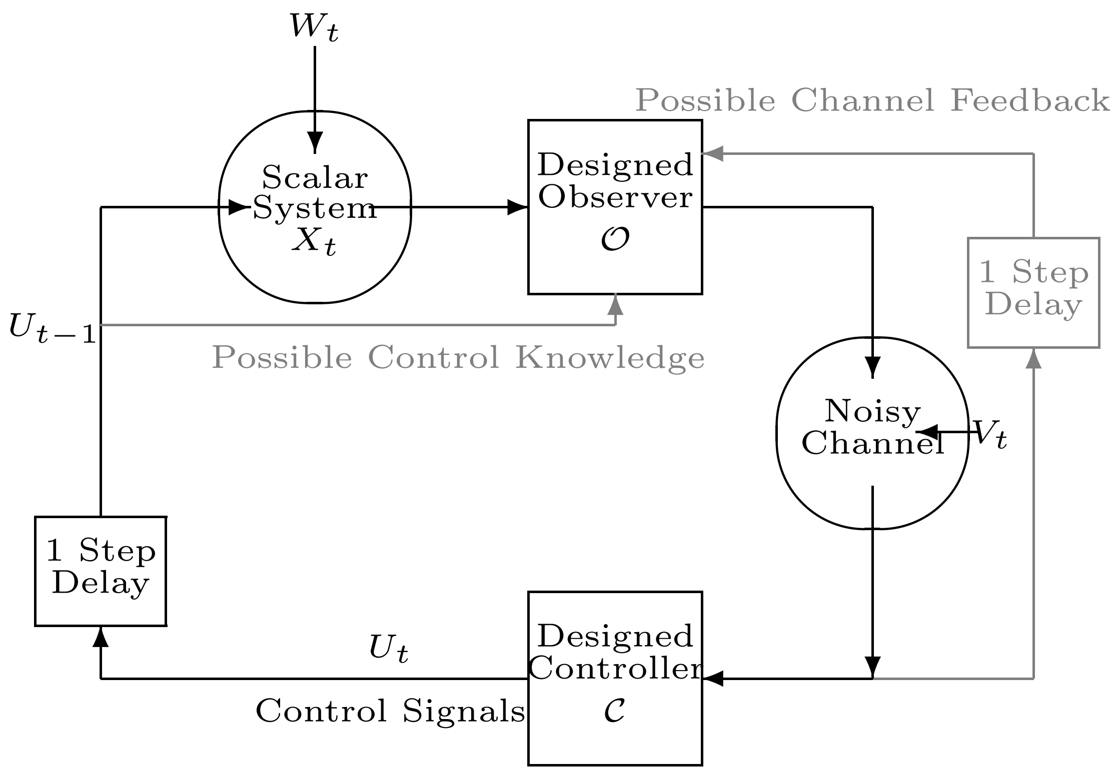
- "Separation Theorem" if ρ is finite.

$$(\mathcal{C}_{R(D)} \cap \mathcal{C}^m) = (\mathcal{C}^{nf}_{e,(F_X,\rho,D)} \cap \mathcal{C}^m) = (\mathcal{C}^f_{e,(F_X,\rho,D)} \cap \mathcal{C}^m)$$

Stabilization problems and anytime communication

- Simple control problem
- Why classical capacity.is not enough.
- Why anytime (delay-universality) is needed
- Implications (power laws, etc.)
- Imperfect information patterns and implicit communication

Our simple scalar distributed control problem



$$X_{t+1} = \lambda X_t + U_t + W_t$$

- Unstable $\lambda > 1$, bounded initial condition and disturbance W.
- Goal: Stability $= \sup_{t>0} E[|X_t|^\eta] \leq K$ for some $K < \infty$.

Is Shannon capacity all we need?

- Consider a system with
  - $\lambda = 2$ 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 $U_t = -\lambda Z_t$
- Resulting closed loop dynamics:

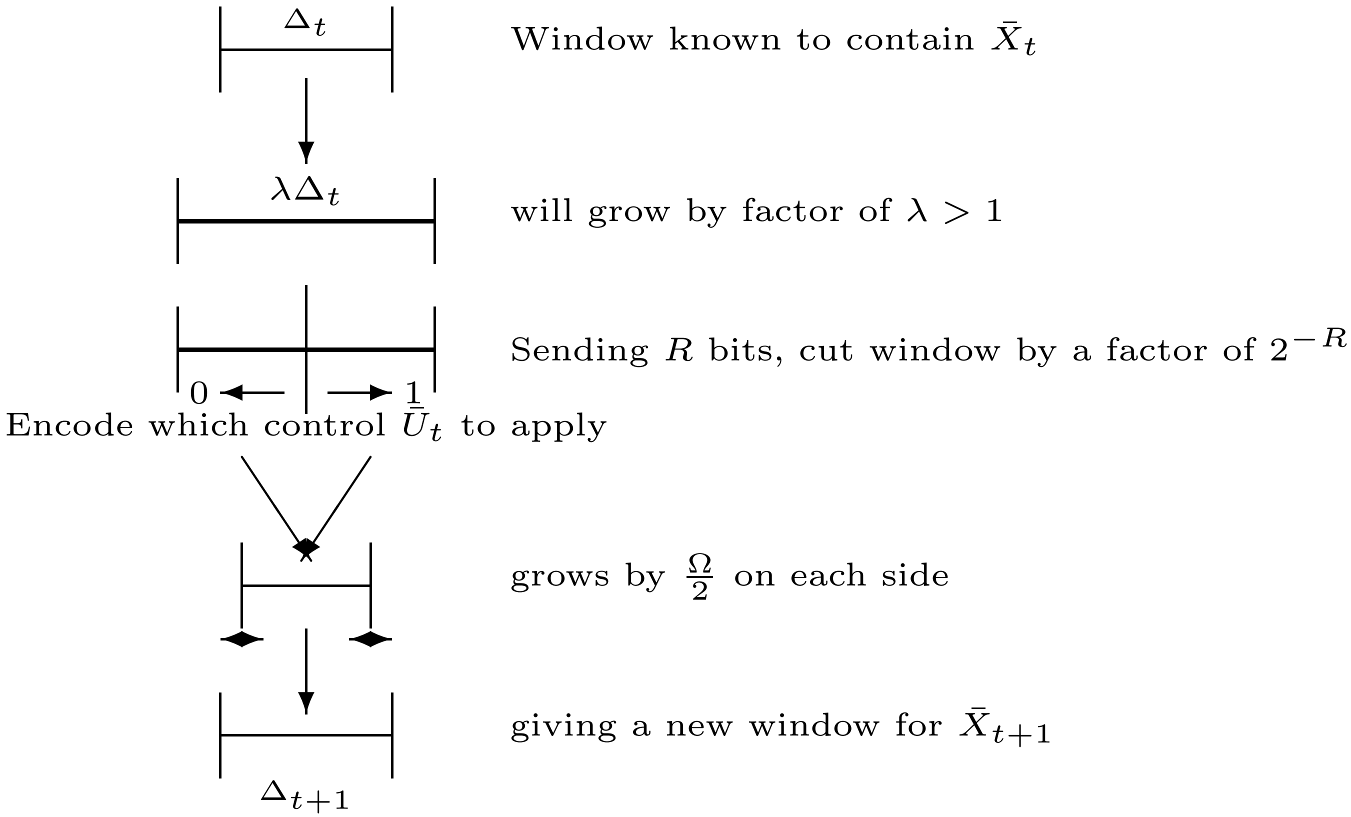
$$X_{t+1} = \begin{cases} W_t & \text{with Probability } \frac{1}{2} \\ 2X_t + W_t & \text{with Probability } \frac{1}{2} \end{cases}$$

Is the closed-loop system stable?

$$X_{t+1} = \begin{cases} W_t & \text{with Probability } \frac{1}{2} \\ 2X_t + W_t & \text{with Probability } \frac{1}{2} \end{cases}$$

- i.i.d. erasures mean arbitrarily long stretches of erasures are possible, though unlikely.
  - System is not guaranteed to stay inside any box.
  - Under 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, $C = \infty$!

Run same plant $\bar{X}$ over noiseless channel



As long as $R > \log_2 \lambda$, we can have Δ stay bounded forever.

What is needed: key intuition

- Break state X into sum of $\check{X}$ (response to disturbance) and $\tilde{X}$ (response to control)
- Suppose $\lambda = 2$ and so $\check{X}_t = \sum_{i=0}^{t} 2^i W_{t-1}$
- Assume W_j either 0 or 1
- In binary notation: $\check{X}_t = W_0 W_1 W_2 \ldots W_{t-1}.00000\ldots$
- If $-\tilde{X}_t$ is close to $\check{X}_t$, their binary representations likely agree in all the high-order bits.
  - High-order bits represent earlier disturbances.
  - Typically, to get a difference at the W_{t-d} level, we have to be off by about 2^d.

Stabilization implies communicating bits reliably in an anytime fashion.

Anytime communication problems: $A_{R,\alpha,K}$

- Same as Shannon problem in source and information pattern.
- Performance objective different:
  - Reinterpret $U_t = 0.\hat{X}_0(t), \hat{X}_1(t), \hat{X}_2(t), \ldots$ in binary
  - $\mathcal{V}_{(K,\alpha)}$ is satisfied if $\mathcal{P}(X_i \neq \hat{X}_i(i+d)) \leq K 2^{-\alpha d}$ for every $i \geq 0, d \geq 0$.
  - Slack parameter: constant factor K
  - Natural orderings: larger K is easier, but larger R, α are harder.
- Capacity

$$\mathcal{C}^f_{a,(R,\alpha)} = \bigcap_{R'<R} \bigcap_{\alpha'<\alpha} \bigcup_{K>0} \{f_c | f_c \text{ solves } A^f_{(R',\alpha',K)}\}$$

$$C_{\text{any}}(f_c, \alpha) = \sup\{R > 0 | f_c \in \mathcal{C}^f_{a,(R,\alpha)}\}$$

Separation theorem for control

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

$$C_{\text{any}}(\eta \log_2 \lambda) \geq \log_2 \lambda$$

In general: If $P(|X| > m) < f(m)$, then $\exists K : P_{\text{error}}(d) < f(K\lambda^d)$

Sufficiency: If there is an $\alpha > \eta \log_2 \lambda$ for which the **channel with noiseless feedback** has

$$C_{\text{any}}(\alpha) > \log_2 \lambda$$

then the scalar system with parameter $\lambda \geq 1$ with a bounded disturbance can be stabilized across the noisy channel with finite η-moment assuming nested information.

What does all this imply?

- If we want $P(|X_t| > m) \leq f(m) = 0$ for some finite m, we require zero-error reliability across the channel.
- For generic DMCs, anytime reliability with feedback is upper-bounded:

$$\begin{aligned} f(K\lambda^d) &\geq \zeta^d \\ f(m) &\geq \zeta^{\frac{\log_2(\frac{m}{K})}{\log_2 \lambda}} \\ f(m) &\geq K' m^{-\frac{\log_2 \frac{1}{\zeta}}{\log_2 \lambda}} \end{aligned}$$

  A controlled state can have at best a power-law tail.

- If we just want $\lim_{m\to\infty} f(m) = 0$, then just Shannon capacity $> \log_2 \lambda$ is required for DMCs.
- Almost-sure stabilization for $W_t = 0$ follows by time-varying transformation.

Known feedback anytime capacities

Characterizing the boundary of possible (R, α) pairs. For generic DMCs, have both upper and lower bounds.

- L-bit packet erasure channel

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

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

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

- Average Power-constrained AWGN or Gilbert-Elliott with CSI (ISIT 05)

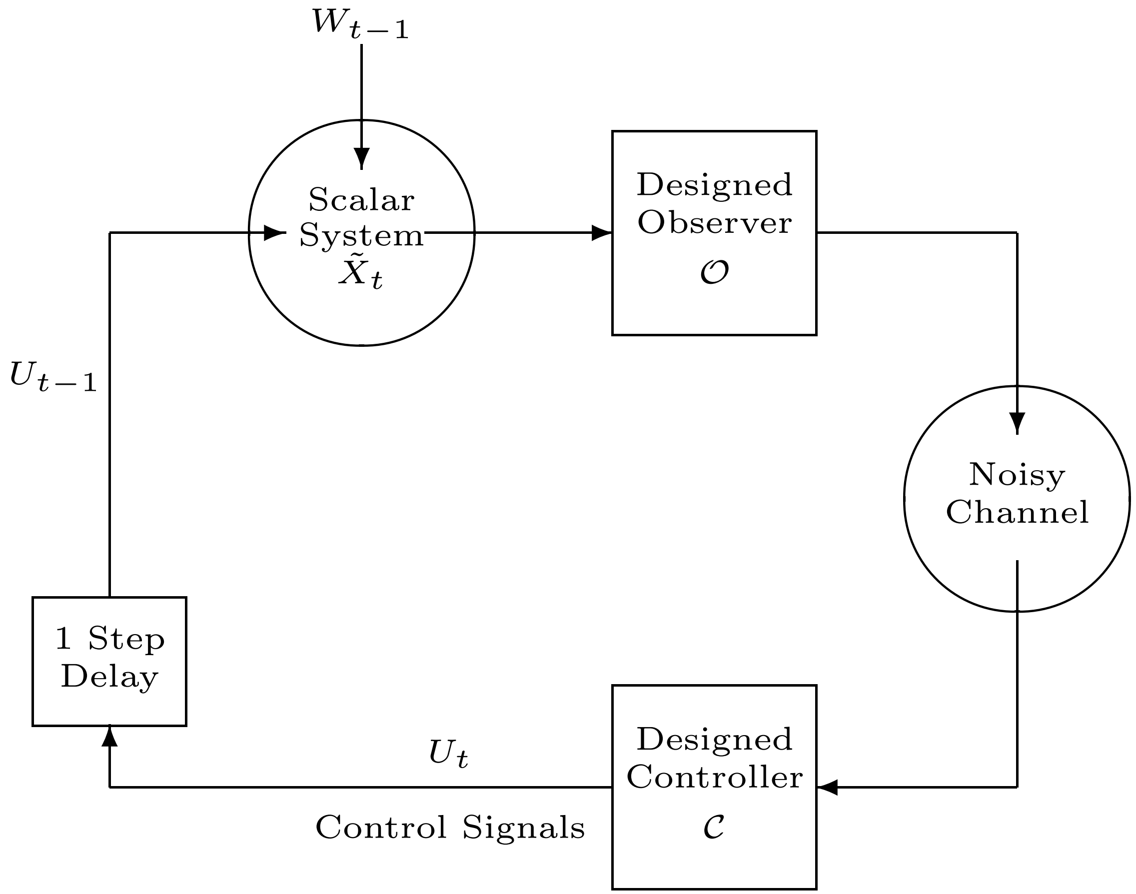
$$C_{\text{any}}(\alpha) = \frac{1}{2}\log_2(1 + \frac{P}{\sigma^2}) = C_{\text{Shannon}}$$

- Power-constrained AWGN+erasure (Allerton 2004)

$$\alpha^*(R) \quad = \quad \begin{cases} -\log_2 \delta & \text{if } R < C_{\text{Shannon}} \\ 0 & \text{otherwise} \end{cases}$$

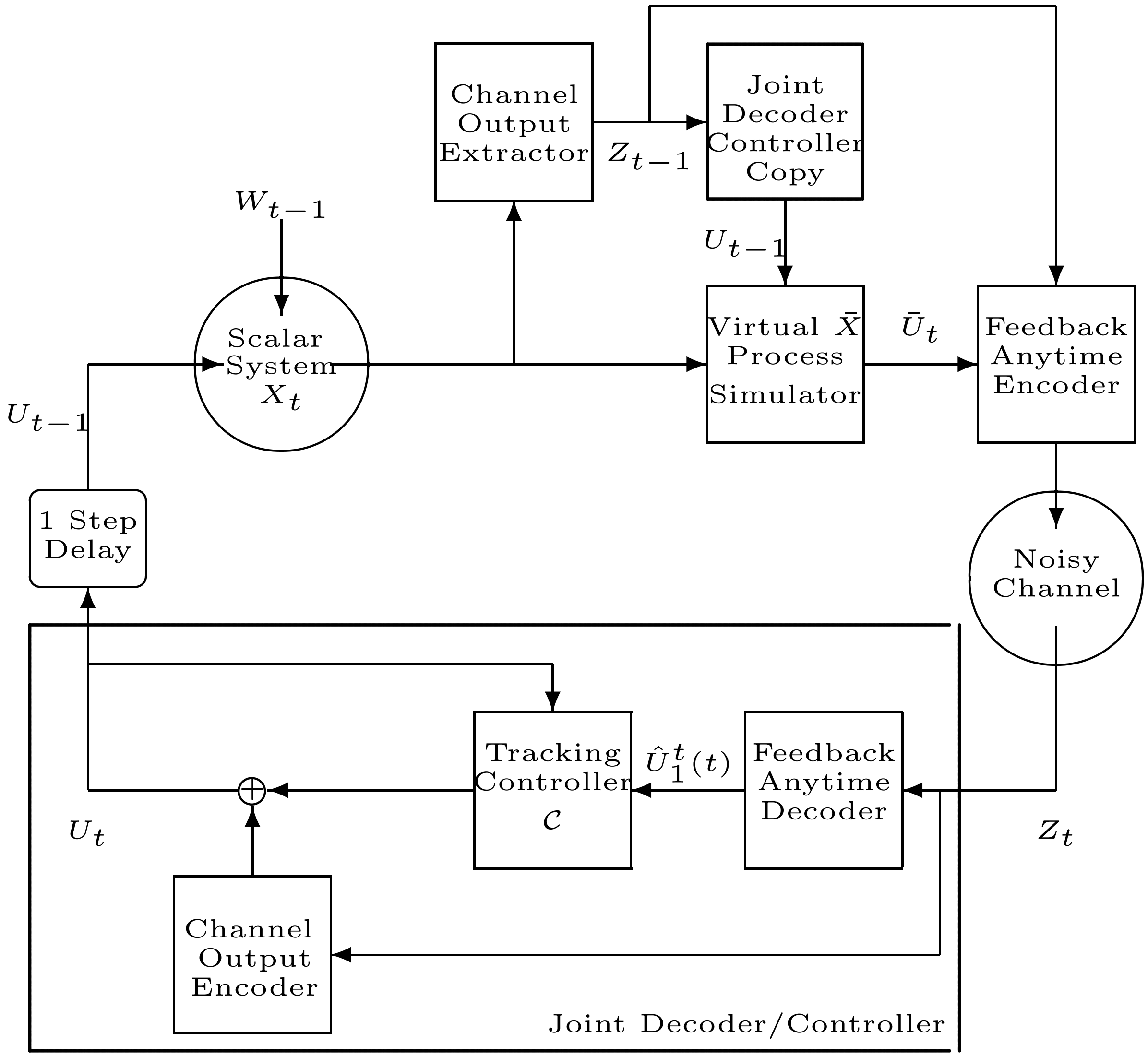
- Similar bound for Markov channels with CSI, but replace δ with the largest eigenvalue of a censored transition matrix.

What about imperfect information patterns?



- Do we now need a higher quality channel?
- The only path from the controller to the observer is through the plant.

Make the plant "dance" in a stable way!



Stabilization and Anytime Equivalence

- With nested information: $A^{f}_{\lambda,\eta,K}$. Without: A^{nf}
  - Slack parameter: K (Performance)
  - Natural ordering: larger η, λ are harder, but larger K is easier.

$$\mathcal{C}^{f}_{s,(\lambda,\eta)} = \bigcap_{\lambda' < \lambda} \bigcap_{\eta' < \eta} \bigcup_{K > 0} \{f_c | f_c \text{ solves } A^{f}_{(\lambda',\eta',K)}\}$$

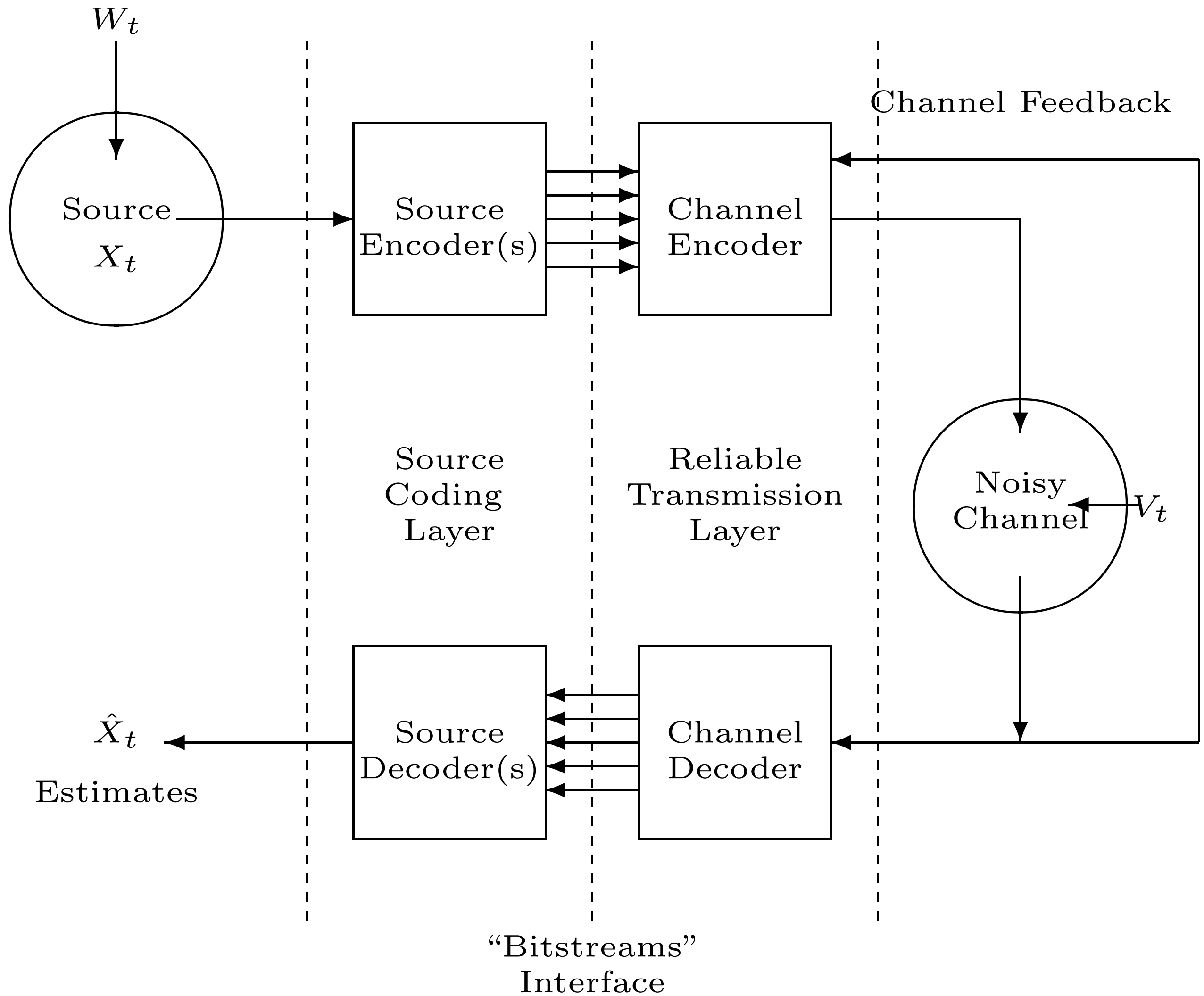
- Equivalences

$$\mathcal{C}^{nf}_{s,(\lambda,\eta)} \subseteq \mathcal{C}^{f}_{s,(\lambda,\eta)} = \mathcal{C}^{f}_{a,(\log_2 \lambda, \eta \log_2 \lambda)}$$

$$(\mathcal{C}^{nf}_{s,(\lambda,\eta)} \cap \mathcal{C}^{\text{finite}}) = (\mathcal{C}^{f}_{s,(\lambda,\eta)} \cap \mathcal{C}^{\text{finite}}) = (\mathcal{C}^{f}_{a,(\log_2 \lambda, \eta \log_2 \lambda)} \cap \mathcal{C}^{\text{finite}})$$

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.
- Imperfect information patterns: need to deal with *intrinsic delays*.


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

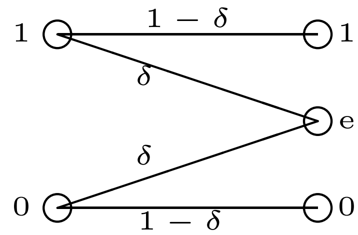
Asymptotic communication problem hierarchy

- The easiest: Shannon communication
  - Asymptotically: a single figure of merit C
  - Equivalent to most estimation problems of stationary ergodic processes with bounded distortion measures.
  - Feedback does not matter.
- Intermediate families: Anytime communication
  - Multiple figures of merit: $(\vec{R}, \vec{\alpha})$
  - Feedback case equivalent to stabilization problems
  - Related nonstationary estimation problems fall here also
  - Feedback matters.
- Hardest level: Zero-error communication
  - Single figure of merit C_0
  - Feedback matters.

Outline

1. Communication problem hierarchy
2. Stabilization problems and anytime communication
3. Feedback channel coding: block-length is *not* a good proxy for delay

Our favorite example: The BEC



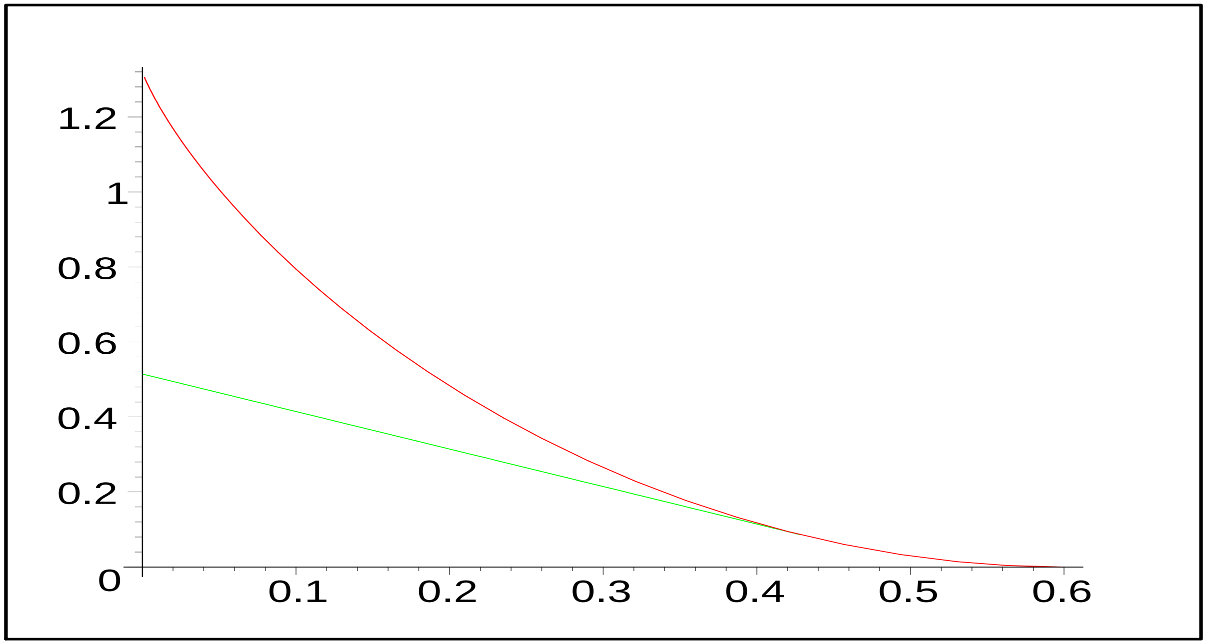
- Simple capacity $1 - \delta$
- With perfect feedback, simple to achieve: retransmit until it gets through
  - Time till success: Geometric$(1 - \delta)$
  - Expected time to get through: $\frac{1}{1-\delta}$
- One size fits all!
  - Strategy works regardless of δ (Universality)
  - All bits eventually get through correctly
  - Gets bits through as soon as possible

Is block-length a good proxy for delay?

- Study erasure case with and without feedback
- Block-codes vs general codes
- Behavior of probability of error with delay

Fixed block length coding

$$E(R) = \lim_{n\to\infty} -\frac{\log_2 P_e(n)}{n}$$



- Classical bounds
  - Random coding bound $\max_{\rho\in[0,1]} E_0(\rho) - \rho R$
  - Sphere-packing bound $D(1-R||\delta)$
- What happens with feedback?

BEC reliability with feedback

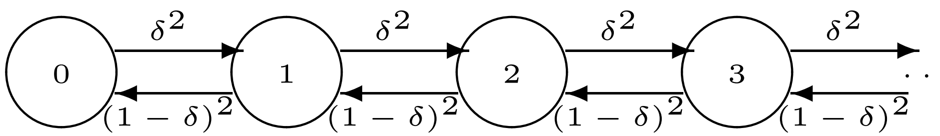
- At rate $R < 1$, have Rn bits to transmit in n channel uses.
- Typically $(1-\delta)n$ code bits will be received.
- Block errors caused by atypical channel behavior.
  - Doomed if fewer than Rn bits arrive intact.
  - *Feedback can not save us.*
  - $D(1-R||\delta)$
- Dobrushin showed that this type of behavior is common.
  - For sufficiently symmetric channels, the sphere-packing bound $E_{sp}(R)$ is unchanged with feedback.
  - In general, can get better but not by much — same convex $\cup$ shape.

Bit error vs Block error

- Block-code with feedback is unfair to later bits.
  - First bit: $-\log_2 \delta$
  - Last bit: $D(1 - R||\delta)$
- Symmetrize by randomly shuffling bits first.
- Makes no difference in exponential order!

BEC with feedback

- $R = \frac{1}{2}$ example:



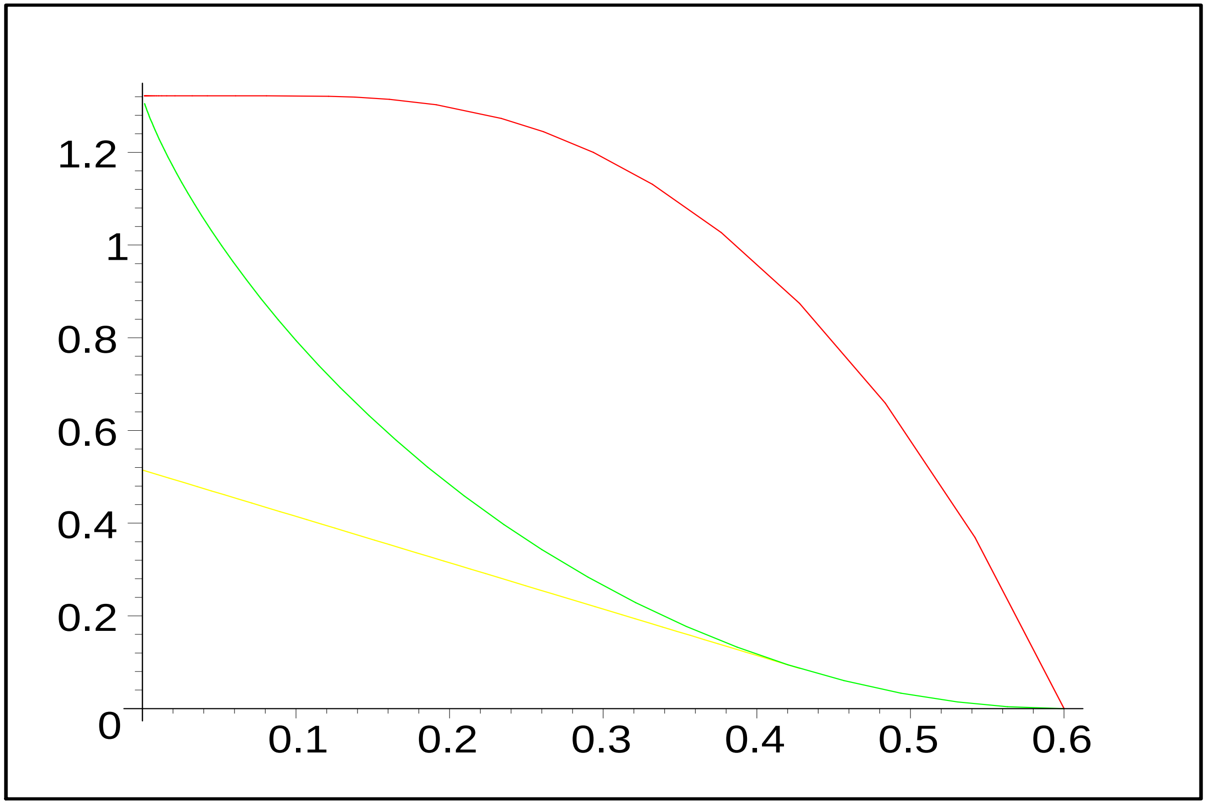
- Birth-death chain: positive recurrent if $\delta < \frac{1}{2}$
- Delay exponent easy to see:

$$P(D \geq d) = P(L > \frac{d}{2}) = K(\frac{\delta}{1-\delta})^d$$

- ≈ 0.584 vs 0.0294 for block-coding!

Compare BEC feedback-delay reliability α to classical E_{sp}

$$R(\alpha) = \frac{\alpha}{\alpha + \log_2(\frac{1-\delta}{1-\delta 2^{\alpha}})}$$

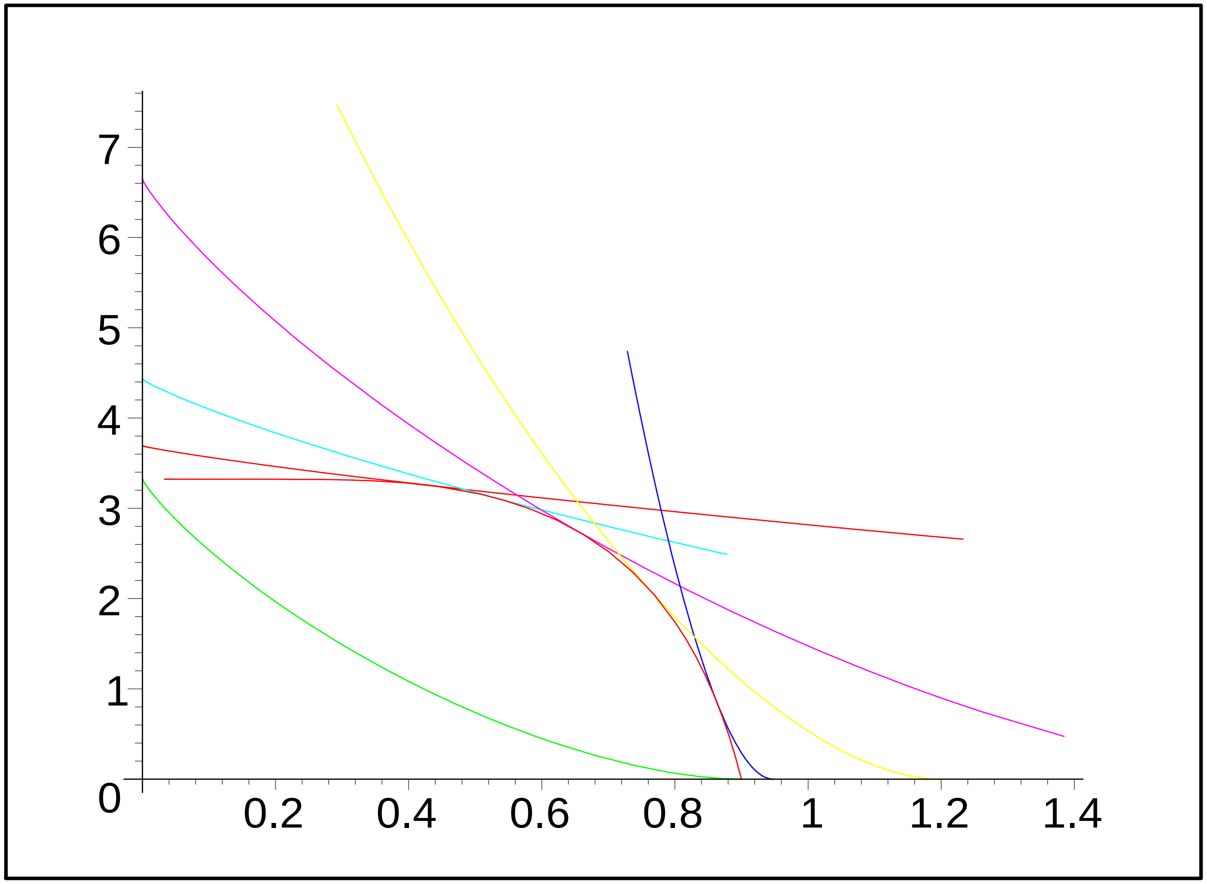


Using E_{sp} to bound α^* in general

- 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 can not exceed the sphere-packing bound $2^{-E_{sp}((1-\lambda)R)n}$

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

Upper bound tight for the BEC with feedback



Bound for symmetric DMCs

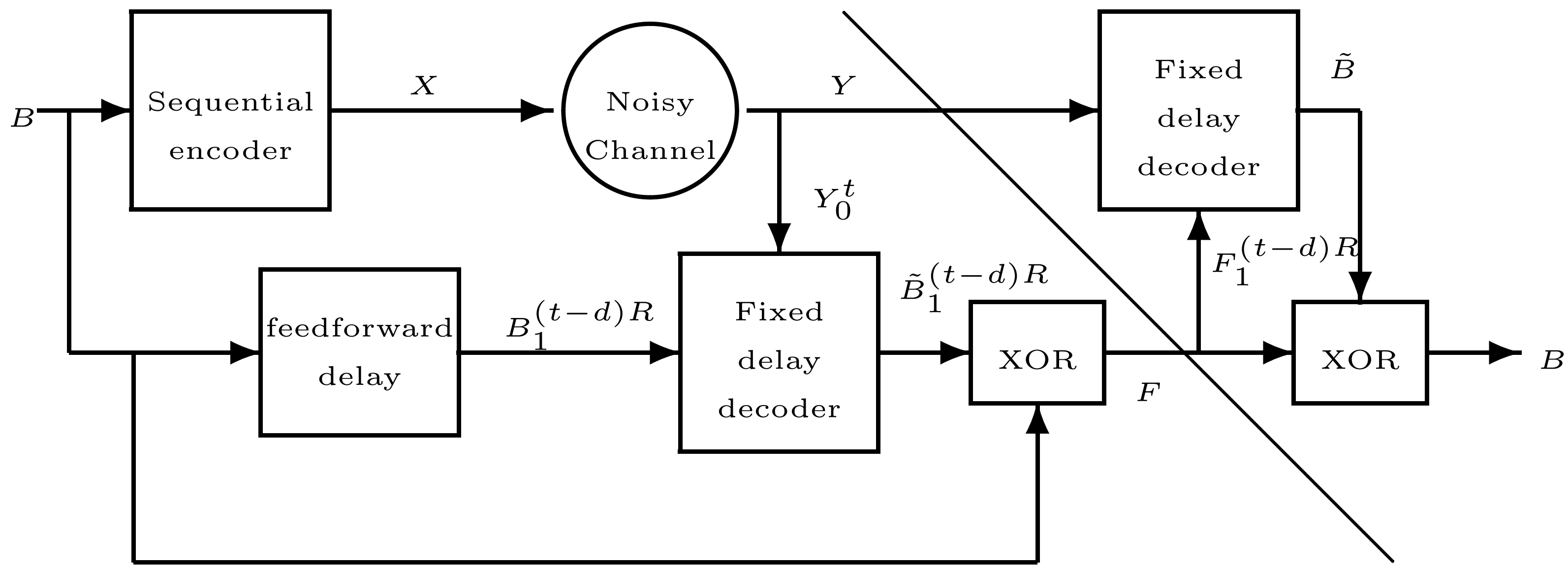
Minimize over λ for symmetric DMCs to sweep out frontier by varying $\rho > 0$:

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

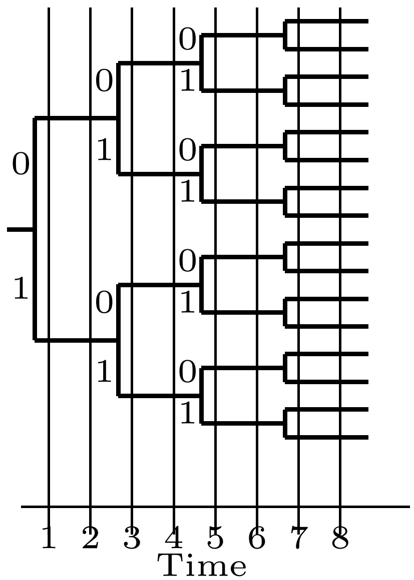
The same basic form as the sphere-packing bound, but concave $\cap$ instead of convex $\cup$.

Reliability: from general coding or from feedback?

- Without feedback: $E_{sp}(R)$ continues to be a bound. (Pinsker)
- Consider a code with target delay d
  - Use it to construct a block-code with blocksize $n >> d$
  - Genie-aided decoder: has the truth of all bits before i
  - Error events for genie-aided system depend only on last d
  - Apply a change of measure argument



Anytime codes without feedback?



Infinite binary tree, with iid random labels:

- Choose a path through the tree based on data bits
- Transmit the path labels through the channel

- Can implement with time-varying convolutional code.
- Decoder does ML decoding
  - Disjoint paths are pairwise indep to true path.
  - $E_r(R)$ analysis applies.
- **Achieves** $\alpha = E_r(R)$ **for every** d **for all** $R < C$

Open questions

- Value of delayed or noisy feedback?
- Bounds for $(\vec{R}, \vec{\alpha})$ regions with/without feedback?
- Distributed problems — many sensors and/or many controllers.
  - Anytime MAC, Broadcast, and Slepian-Wolf exist.
  - Should yield new distributed sufficient conditions.
- Interaction with unmodeled dynamics
- **Performance**
  - Tatikonda's $D_{seq}(R)$ bounds are still the best we have.
  - Anytime tells us more about the ∞ vs finite boundary, but not about performance within the "finite" region.
  - For non-causal estimation of unstable processes, performance is due to residual capacity left over after anytime requirement is met.