The necessity and sufficiency of anytime capacity for control over a noisy communication link

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

2004 CDC
December 15, 2004

Outline

1. Introduction and problem setup
2. Why Shannon capacity and mutual information are not enough
3. Necessity of anytime reliability
4. Consequences: power-laws vs zero-error requirements
5. Sufficiency of anytime reliability
6. Conclusions and extensions

Shannon's Formulation of Communication

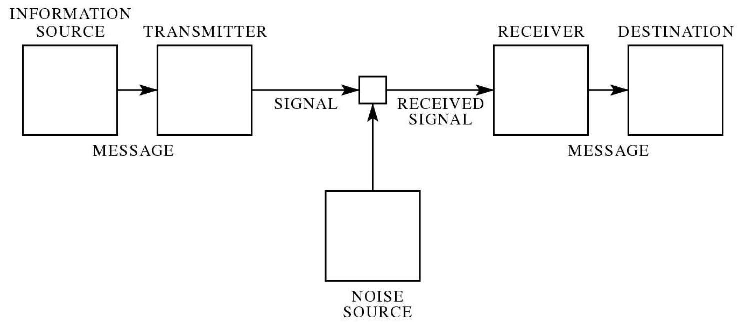


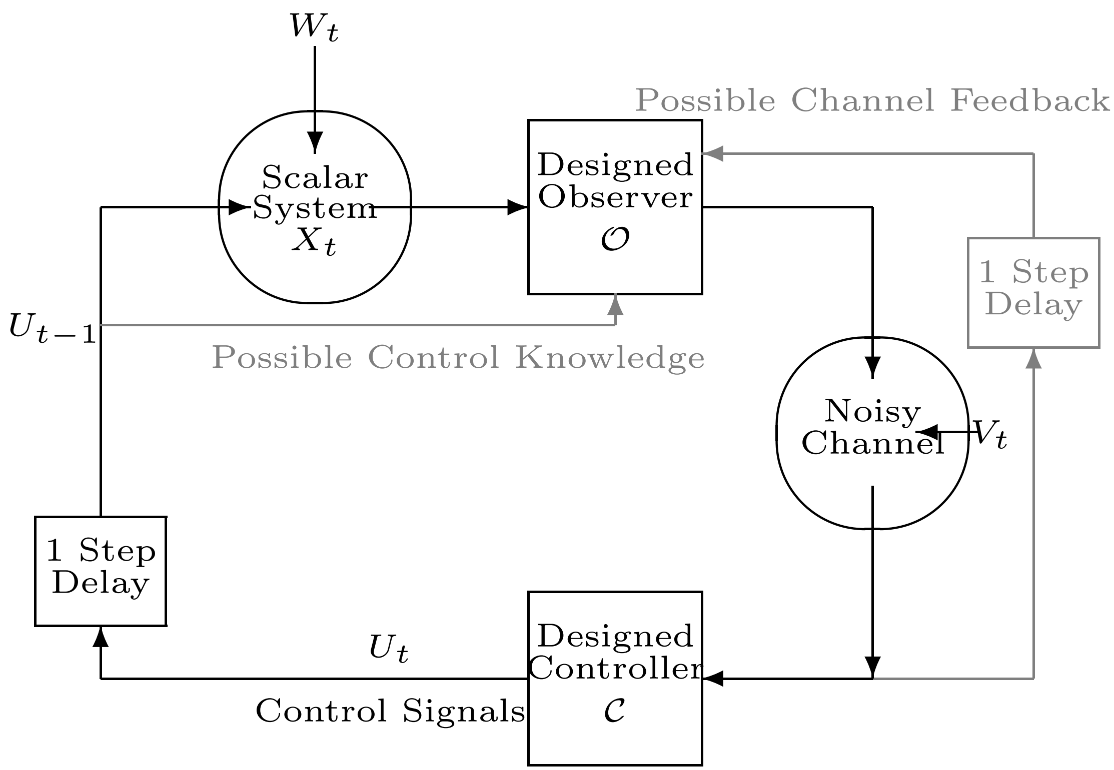
Fig. 1—Schematic diagram of a general communication system.

- Large end-to-end delay is permitted.
- "Meaning" introduced through end-to-end distortion measure.

> "... 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

Many people have taken up the challenge. Prior talk gave some pointers. See Sep. 2004 *Trans. on Auto. Control* for more.

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 $K < \infty$.

Outline

1. Introduction and problem setup
2. Why Shannon capacity and mutual information are not enough
3. Necessity of anytime reliability
4. Consequences: power-laws vs zero-error requirements
5. Sufficiency of anytime reliability
6. Conclusions and extensions

Is Shannon capacity all we need? (Review)

- 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:

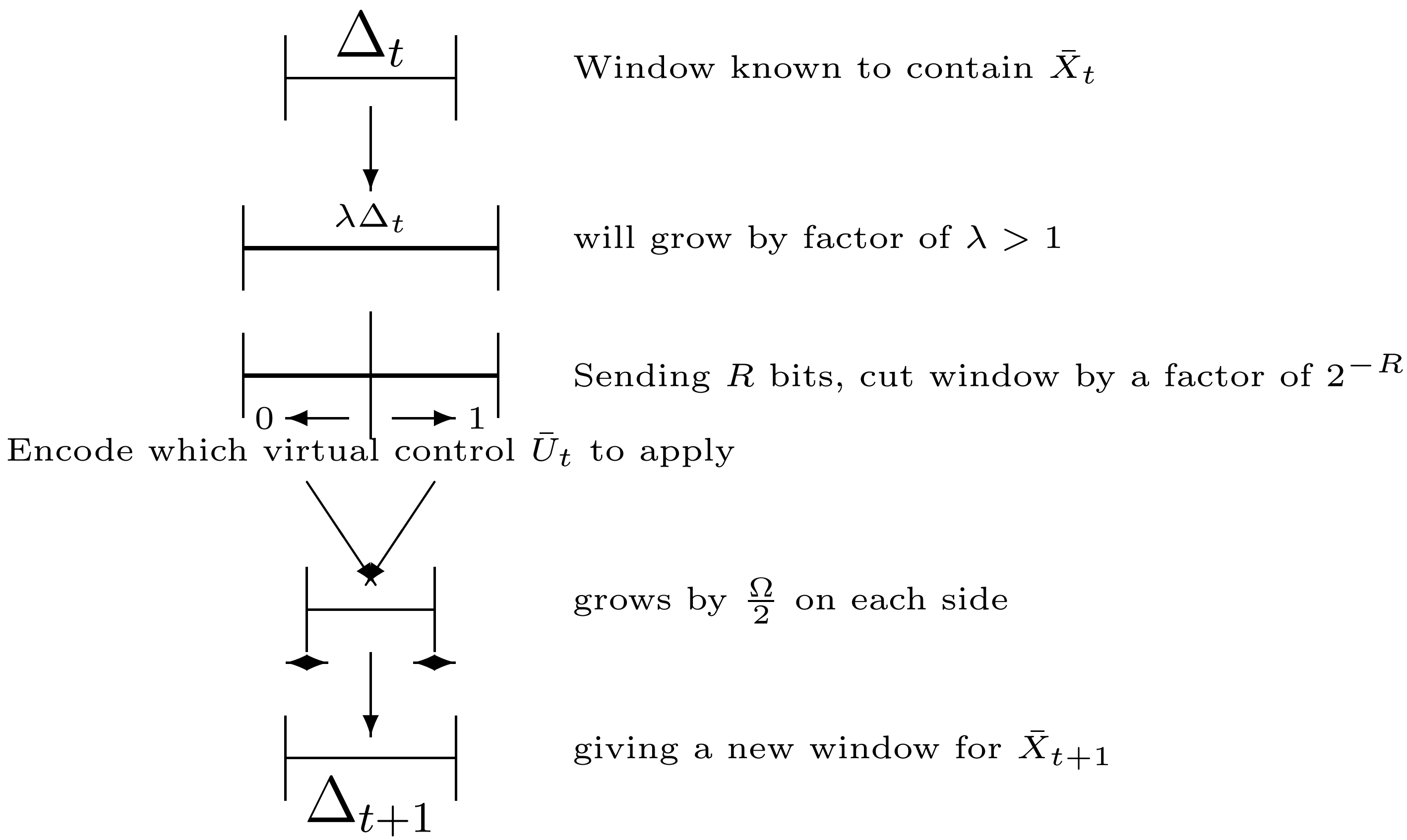
$$X_{t+1} = \begin{cases} W_t & \text{with Probability } \frac{1}{2} \\ 2X_t + W_t & \text{with Probability } \frac{1}{2} \end{cases}$$

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 finite forever.

Outline

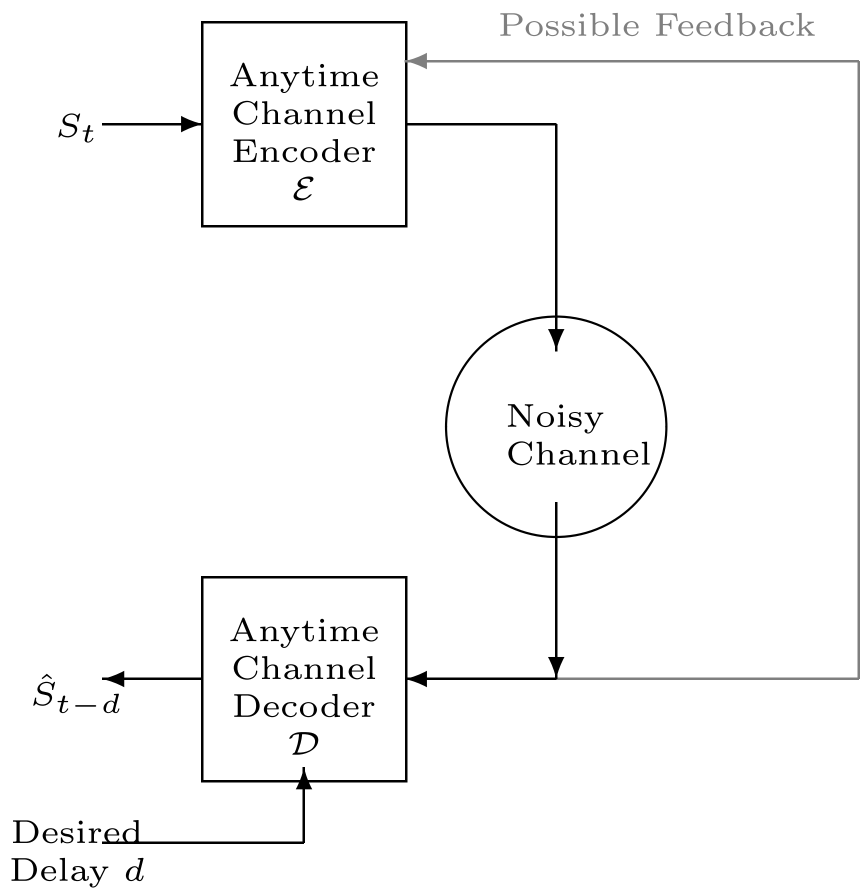
1. Introduction and problem setup
2. Why Shannon capacity and mutual information are not enough
3. Necessity of anytime reliability
4. Consequences: power-laws vs zero-error requirements
5. Sufficiency of anytime reliability
6. Conclusions and extensions

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 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.

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.

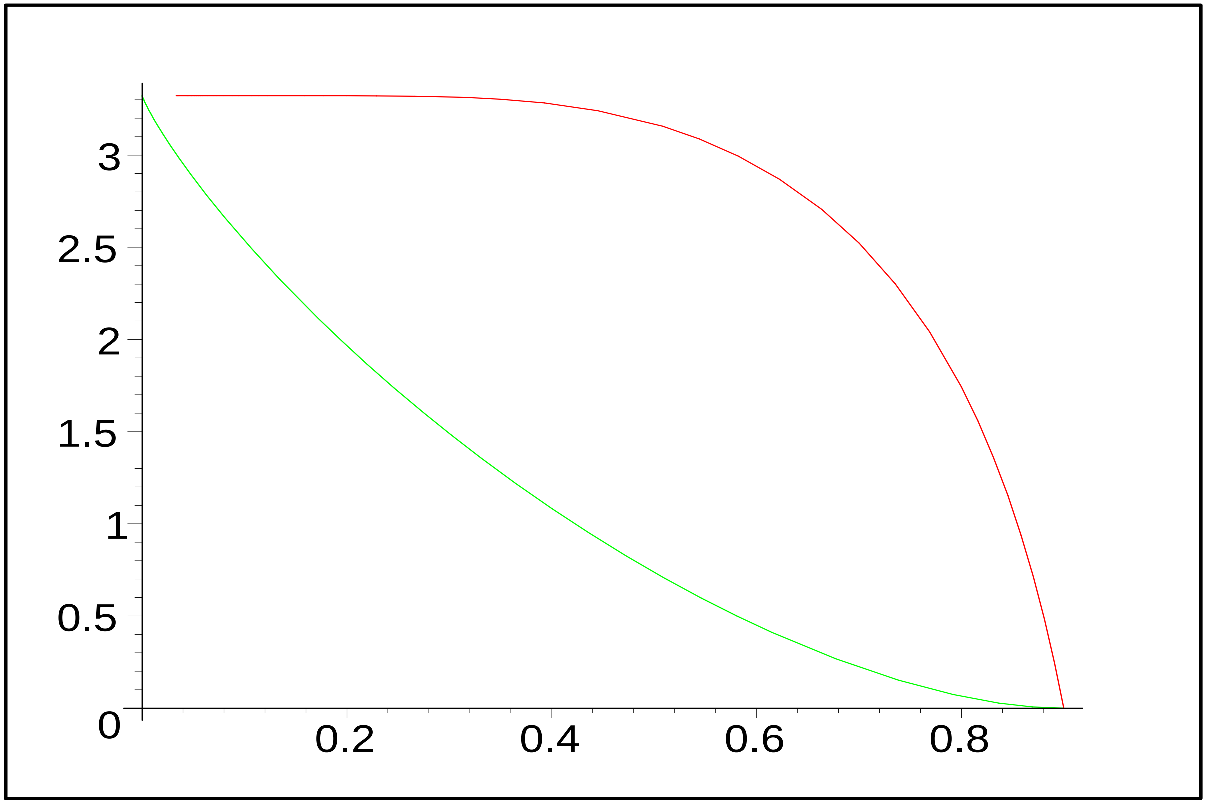
- "Reliable Transmission" means every bit is eventually correctly received. Parametrize by the rate at which the probability of bit error $P(S_{t-d} \neq \hat{S}_{t-d}(t))$ goes to zero as delay d increases.

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

- Related to information-theoretic error exponents, but represent a "delay universal" variation.

BEC feedback anytime reliability vs E_{sp}

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



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

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

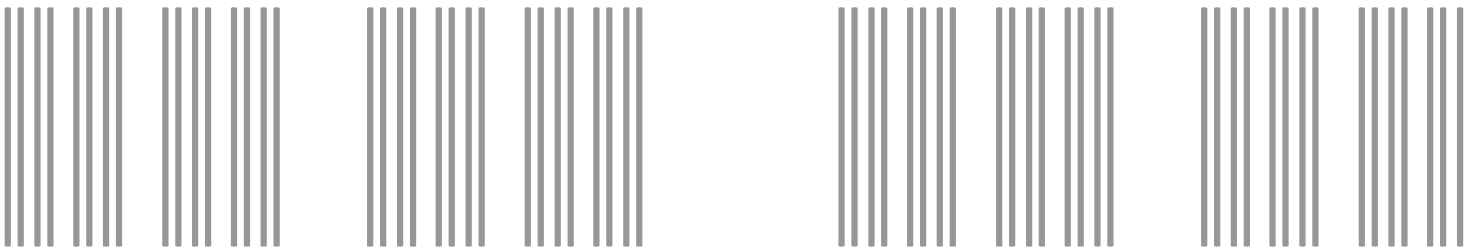
Sufficiency: If there is an $\alpha > \eta \log_2 \lambda$ for which the **channel with noiseless feedback** has

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

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

Proof Idea: Necessity (details in paper)

- Follow the key intuition and embed data to be communicated into a bounded disturbance.
- Technical tool: Instead of binary, use a Cantor set based mapping.



- Recover data from $-\tilde{X}_t$.
- If $P(|X| > m) < f(m)$, then $\exists K$:

$$P_{\text{error}}(d) < f(K\lambda^d)$$

Outline

1. Introduction and problem setup
2. Why Shannon capacity and mutual information are not enough
3. Necessity of anytime reliability
4. Consequences: power-laws vs zero-error requirements
5. Sufficiency of anytime reliability
6. Conclusions and extensions

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, the anytime reliability with feedback can be upper-bounded and it is at most exponential in nature.

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

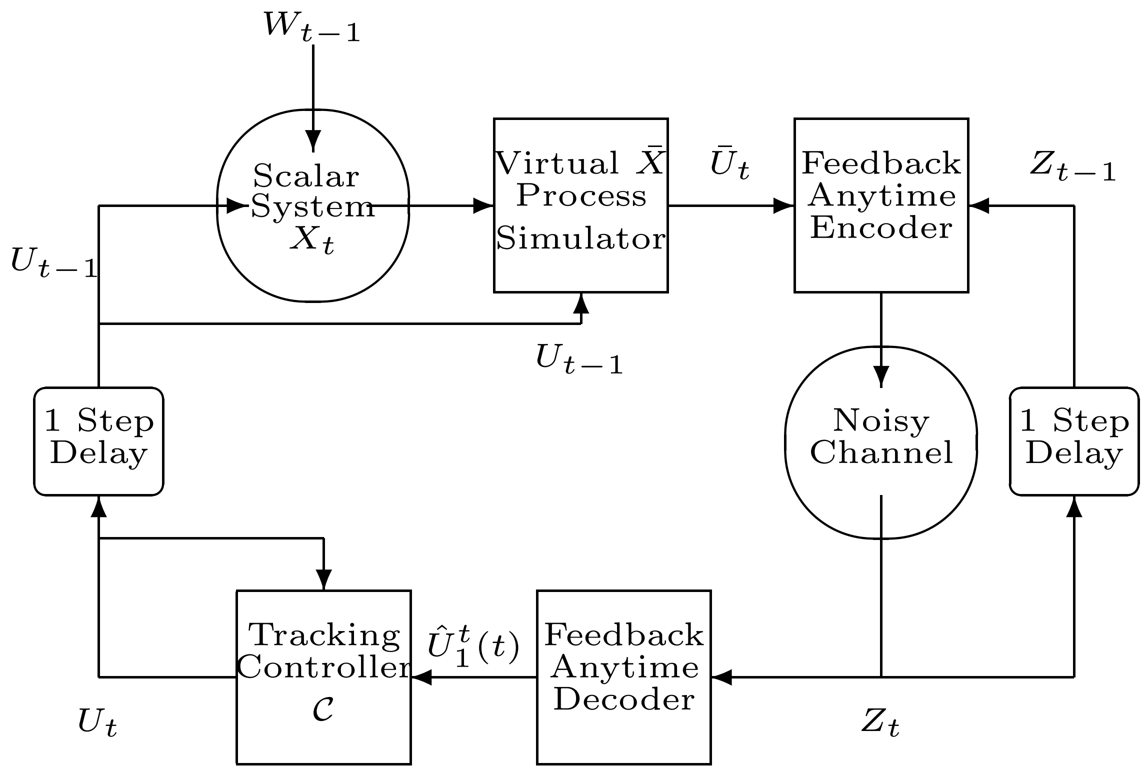
A controlled state can have at best a power-law tail!

- If we just want $\lim_{m \to \infty} f(m) = 0$, then just Shannon capacity is required for DMCs.

Outline

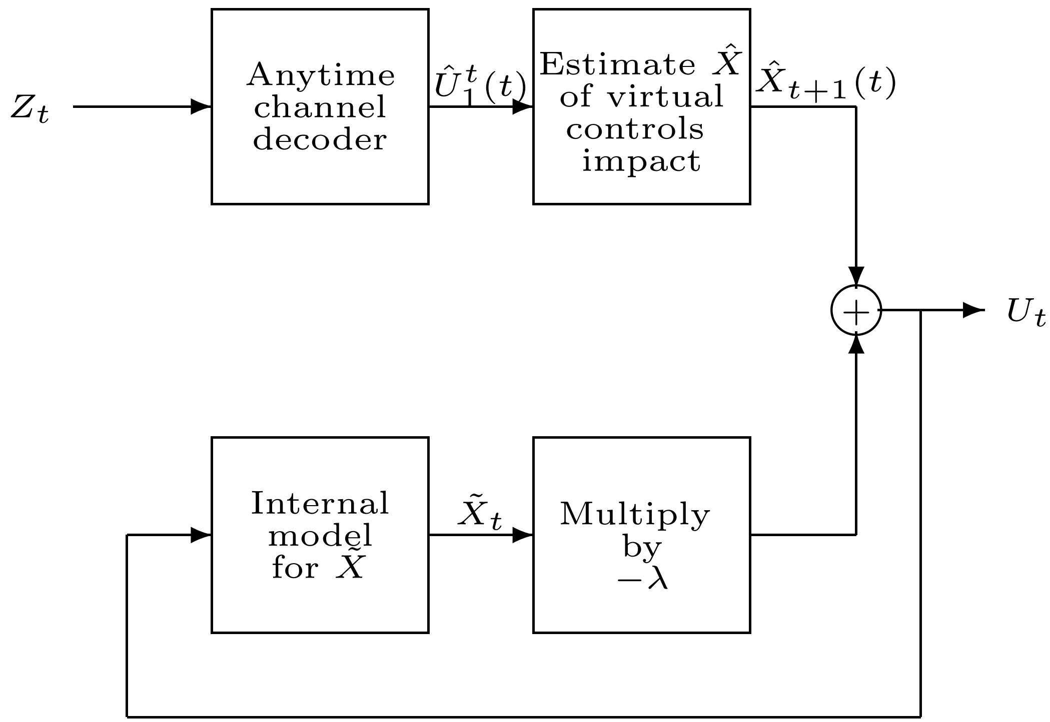
1. Introduction and problem setup
2. Why Shannon capacity and mutual information are not enough
3. Necessity of anytime reliability
4. Consequences: power-laws vs zero-error requirements
5. Sufficiency of anytime reliability
6. Conclusions and extensions

Sufficiency construction: nested information case



- Observer pretends it is controlling $\bar{X}$ with the same disturbance but a noiseless channel.
- Anytime code used to communicate $\bar{U}_t$

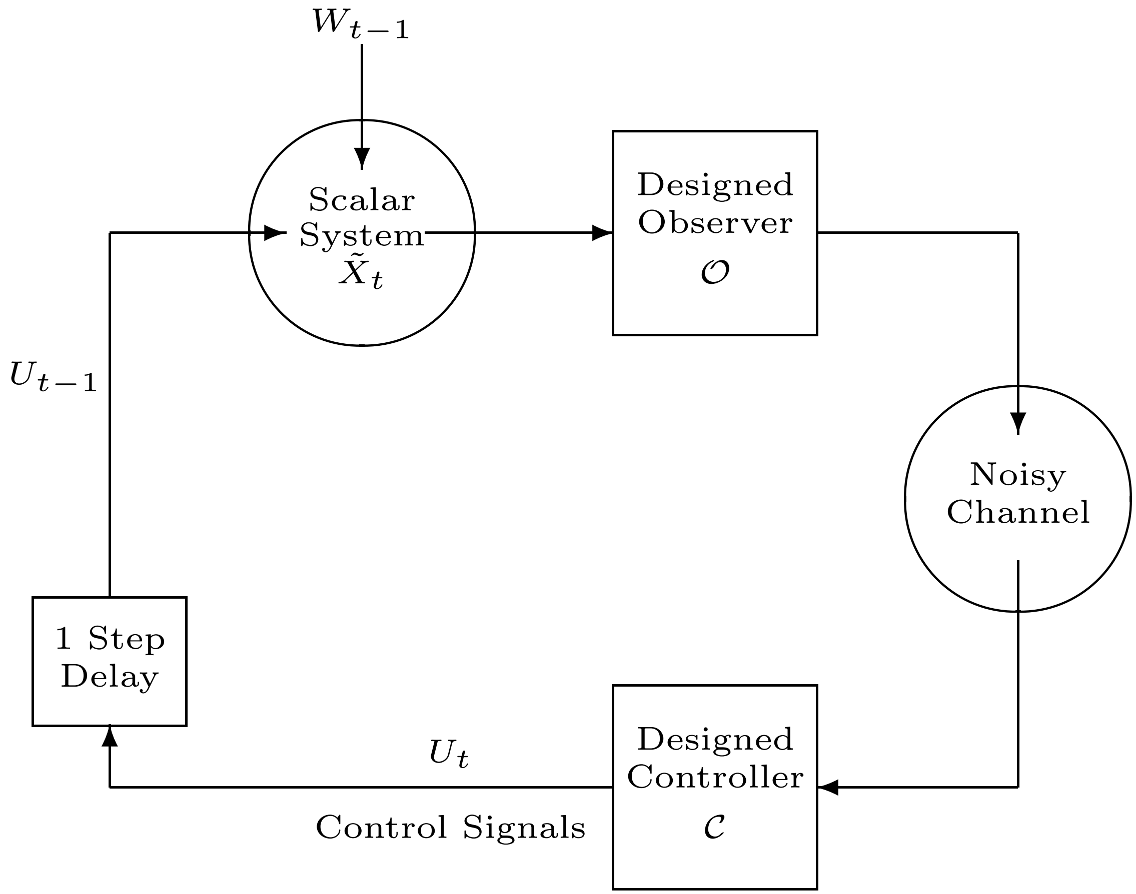
Controller action: try to track $\bar{X}_t$



Uncorrected errors d time steps ago cause at most λ^d of an impact in controlled state.

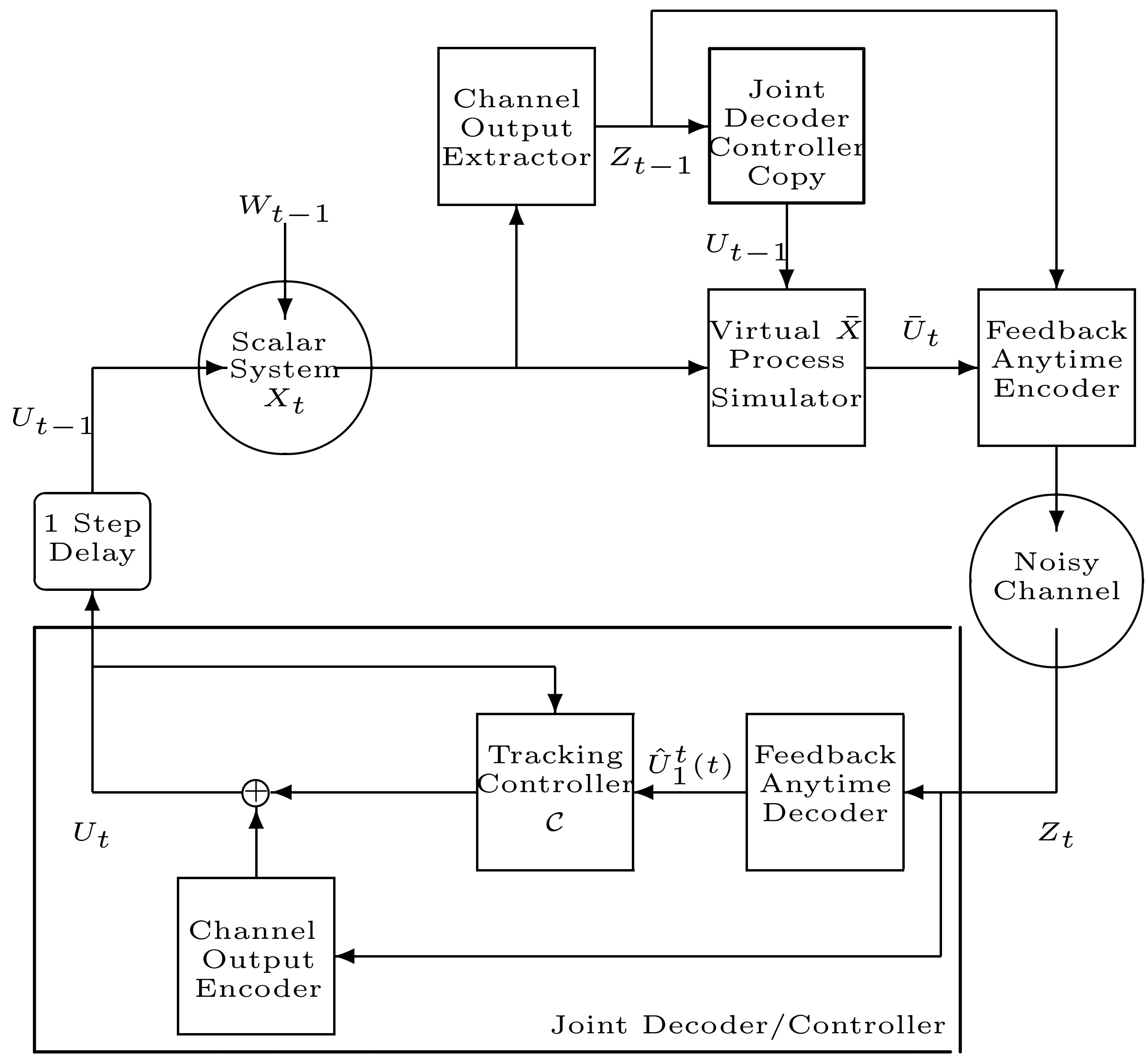
$$P(|X_t| > m) \leq K''' m^{-\frac{\alpha}{\log_2 \lambda}}$$

What about imperfect information patterns?

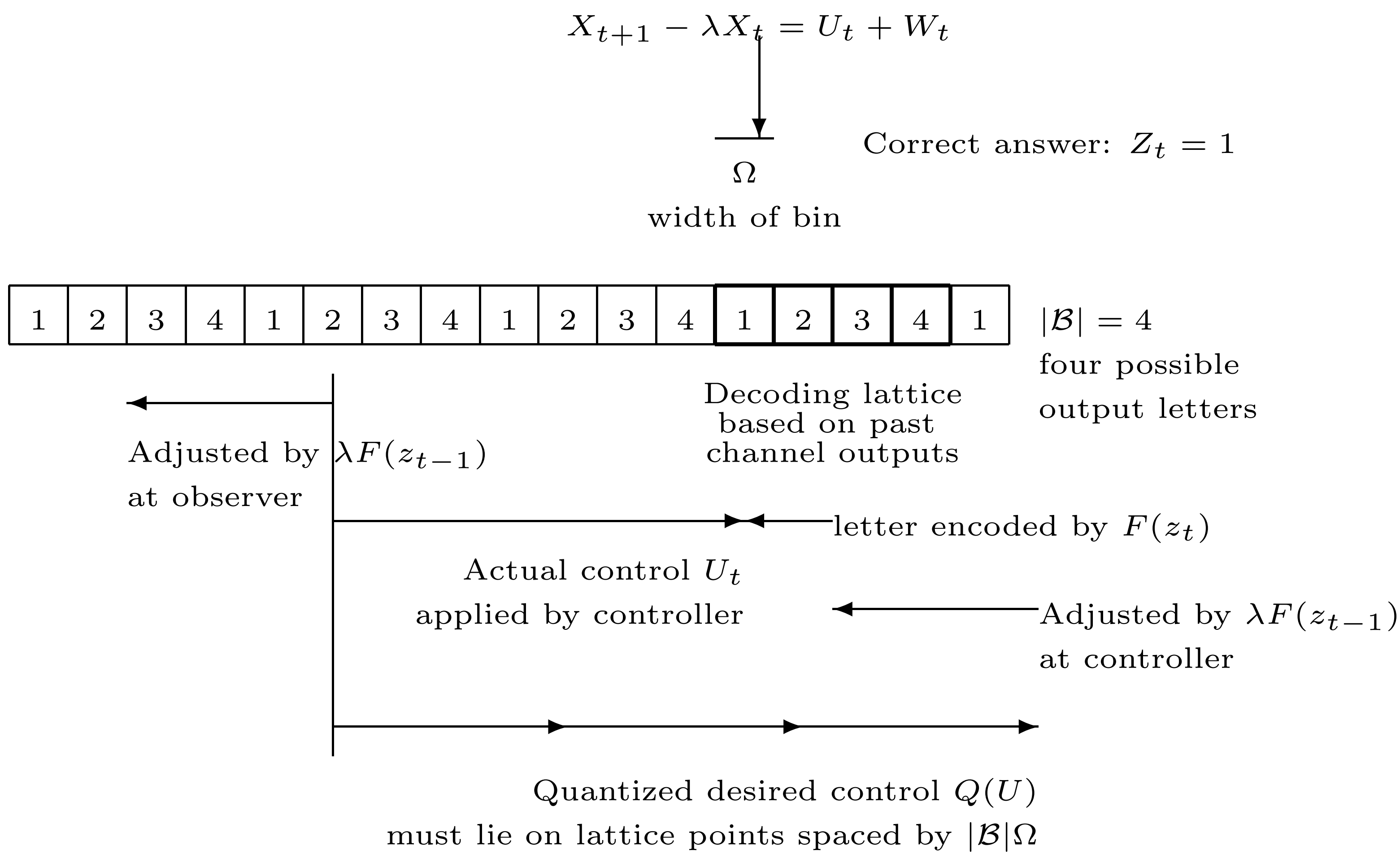


- Do we now need a higher quality channel?
- The only path from the controller to the observer is through the plant.

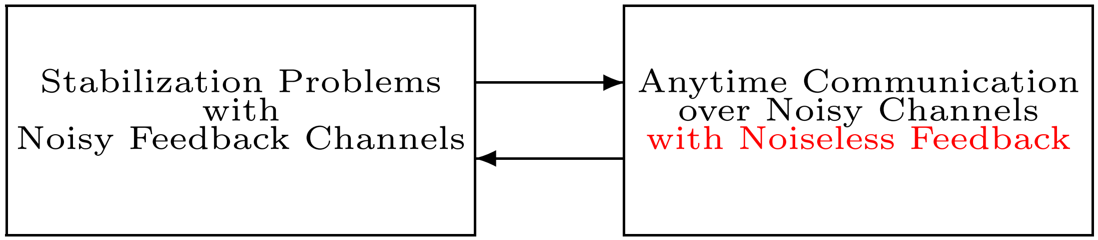
Make the plant "dance" in a stable way!



How is the output information conveyed



Conclusions and extensions



- Anytime reliability is *the* right concept for stabilization.
- Extensions:
  - Boundedly noisy observations
  - Memoryless observers for high quality DMCs
  - Continuous time (use nats)
  - Vector-case
    - Each unstable eigenvalue might need a different reliability: channel code must prioritize
    - Intrinsic delays within the plant impose stricter requirements for non-nested information patterns.

Some backup slides

Open Problems

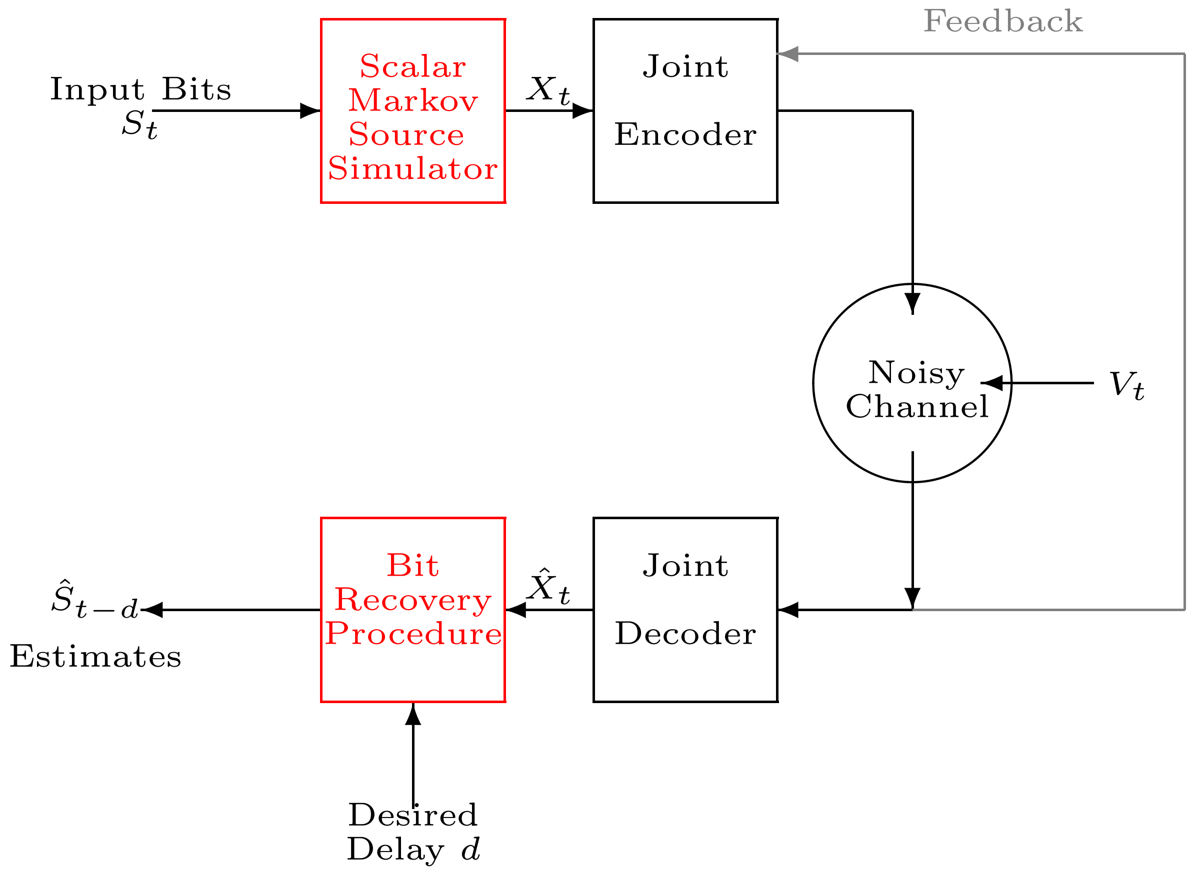
- Need more tools and results on bounding achievable $(\vec{\alpha}, \vec{R})$ regions.
- Bit-pipes fit together into networks naturally, need analogous understanding for networked reliable bit-pipes.
- Anytime (delay-universal) versions of network information theory: Slepian-Wolf, MAC, Broadcast, Relay, etc.
- Better understand the performance-loss caused by the communication constraint — starting with the gap between $D_{seq}(R)$ and $D(R)$.

Connecting with control and estimation

- Classical LQG theory: Everything linear and optimal in quadratic sense. No need for information theory, but nothing generalizes.
- Witsenhausen's 1968 "counterexample:" Nonclassical information patterns cause trouble even in LQG — need to both signal and control simultaneously.

  "[The weak results we have are] in sharp contrast with the elaborate results of information theory. The latter deals with an essentially simpler problem, because the *transmission* of the information is considered independently of its *use*. ...

  Efforts to establish a new theory of information, taking optimal cost into account, have not as yet been convincing." — Hans Witsenhausen 1971

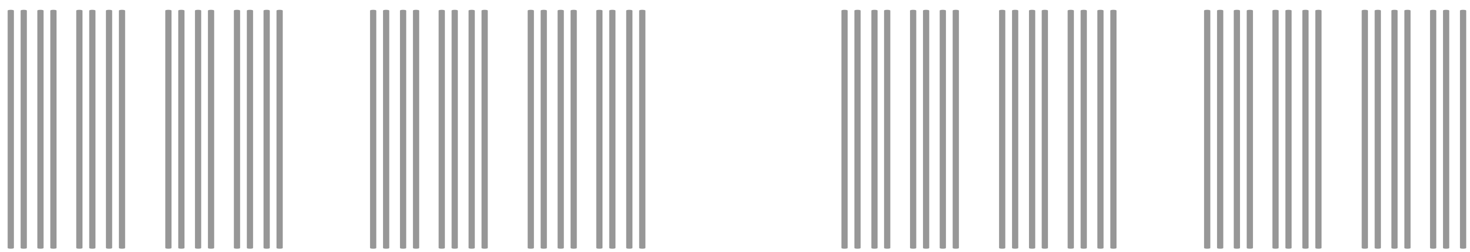


- Use the input bits to drive a source simulator whose output looks like the unstable Markov source $(X_{t+1} = aX_t + W_t)$
- Controlled system state is like the estimation error since the decoder has access to the controls.

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!