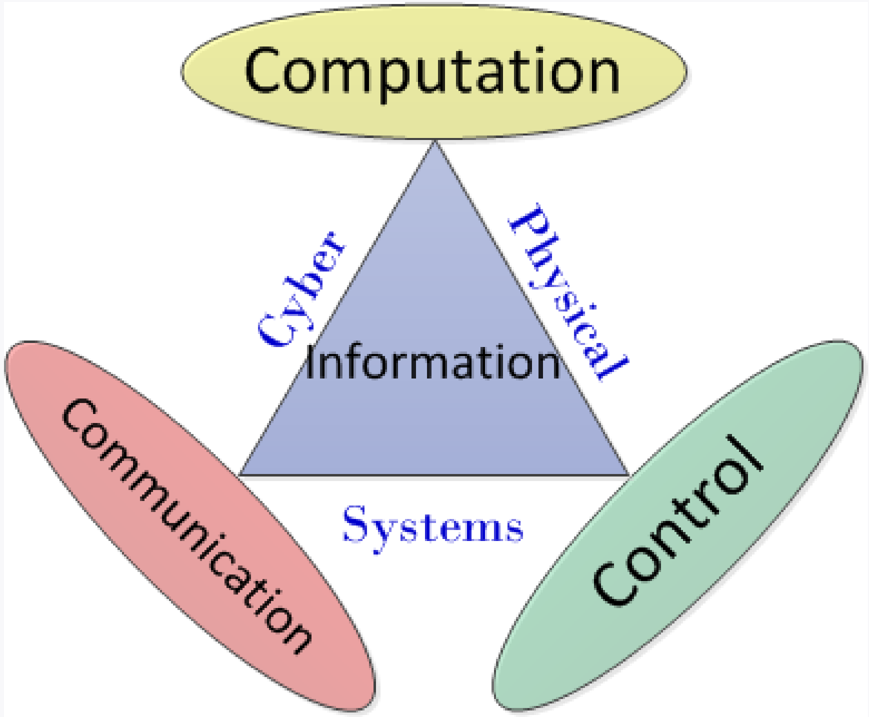
Elements of information theory for networked control systems

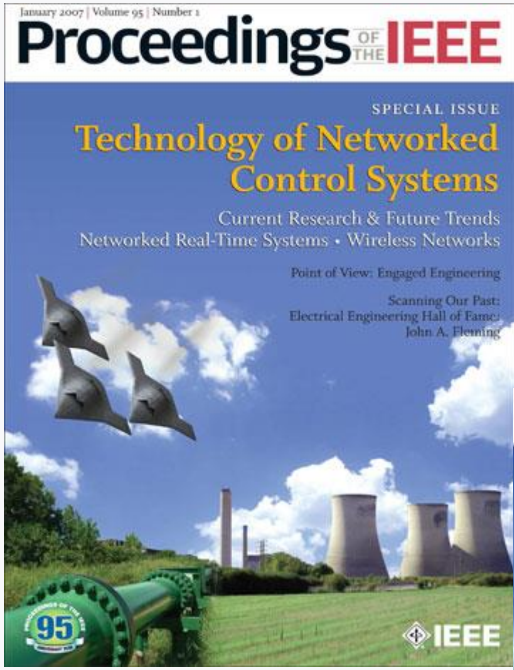
Massimo Franceschetti

PhD School on Control of Networked and Large-Scale Systems, Lucca, Italy, July 2013

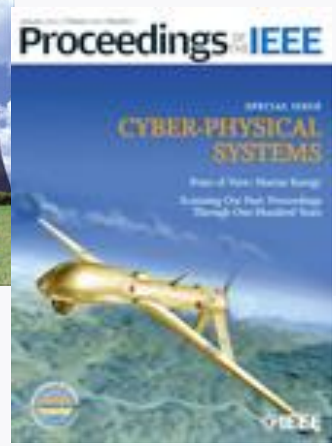
Motivation


Computation
Cyber
Physical
Information
Communication
Control
Systems

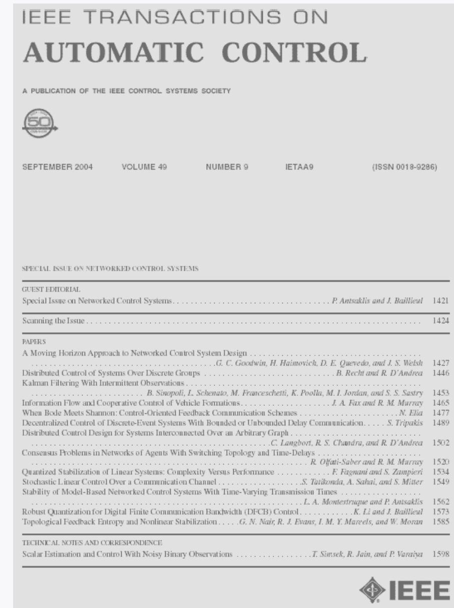
Motivation



January 2007



January 2012



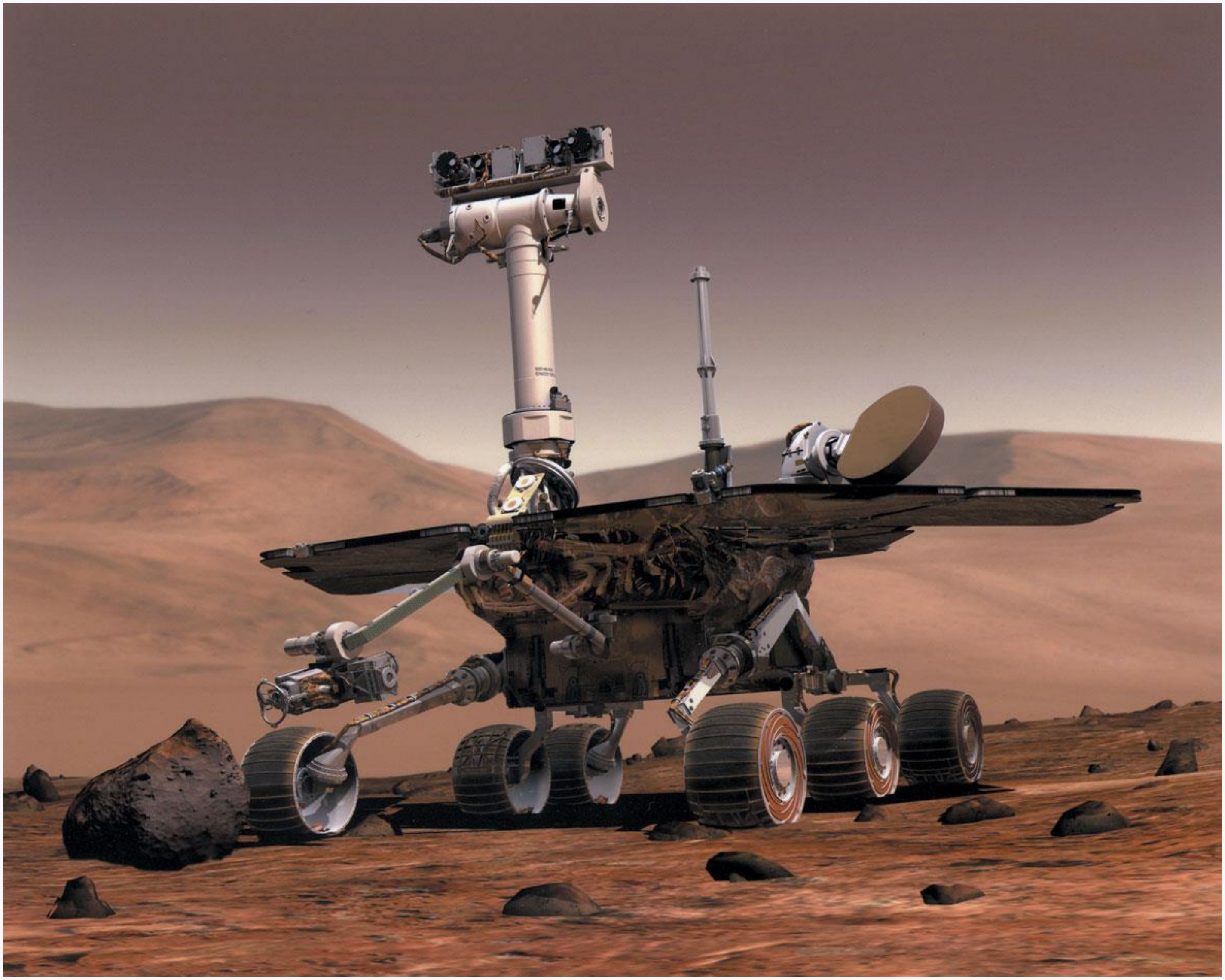
IEEE JOURNAL ON
SELECTED AREAS IN COMMUNICATIONS

IEEE **Transactions on Control of Network Systems**

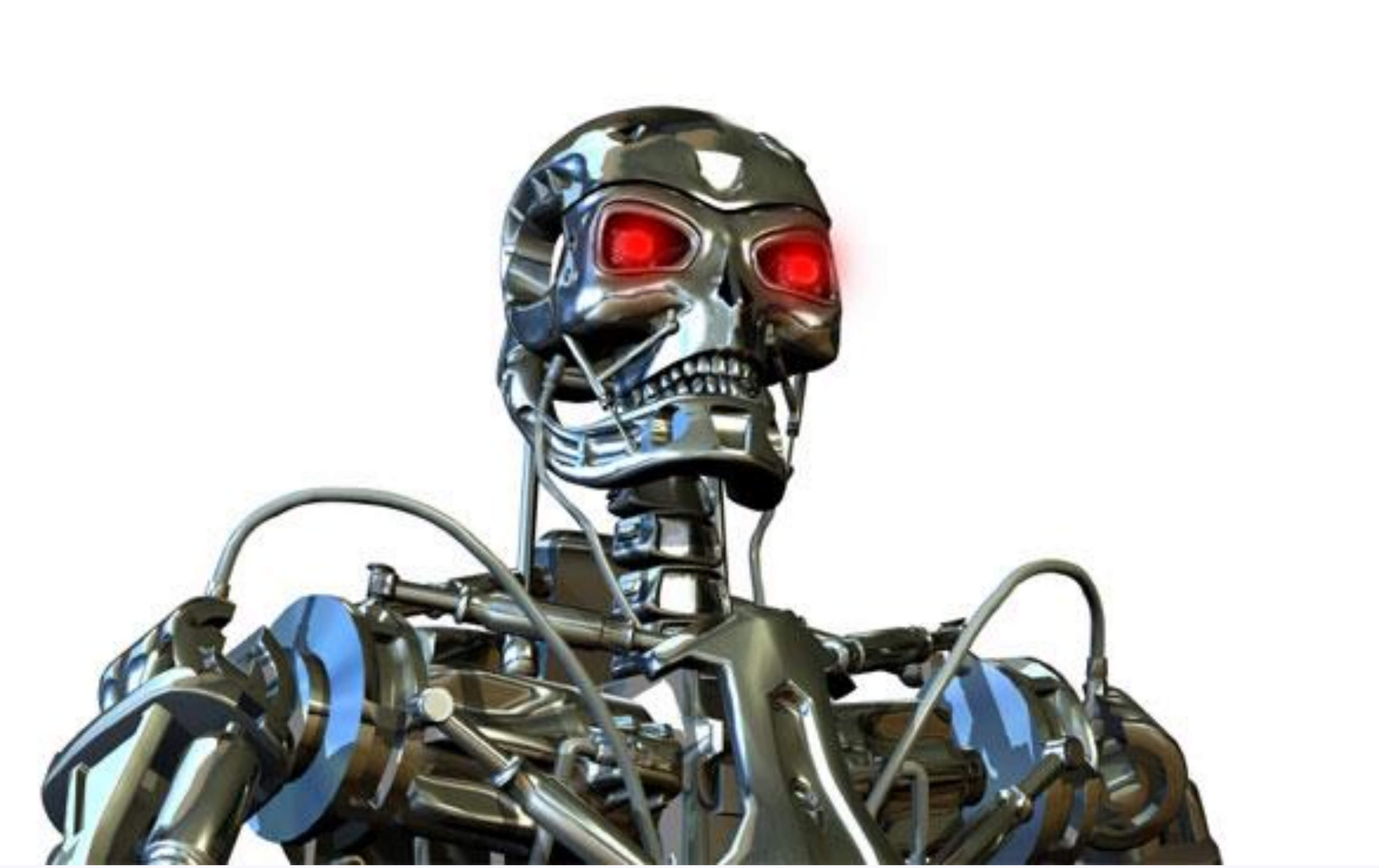
January 2014

Motivation

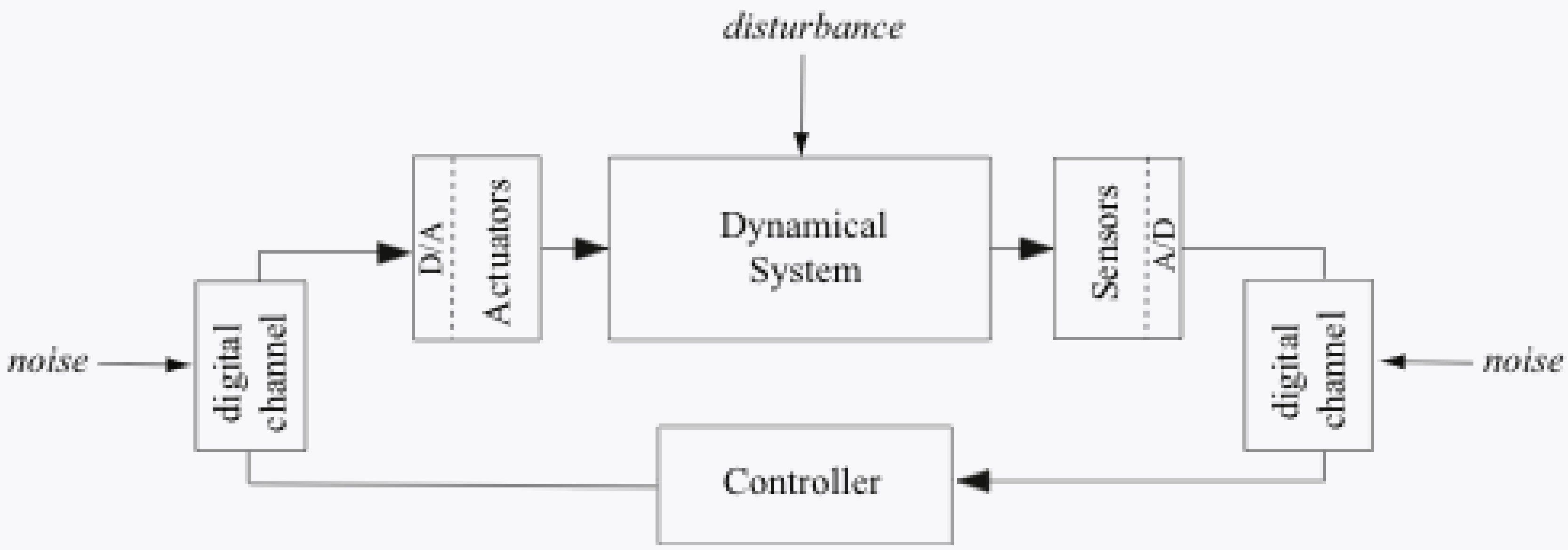
Motivation

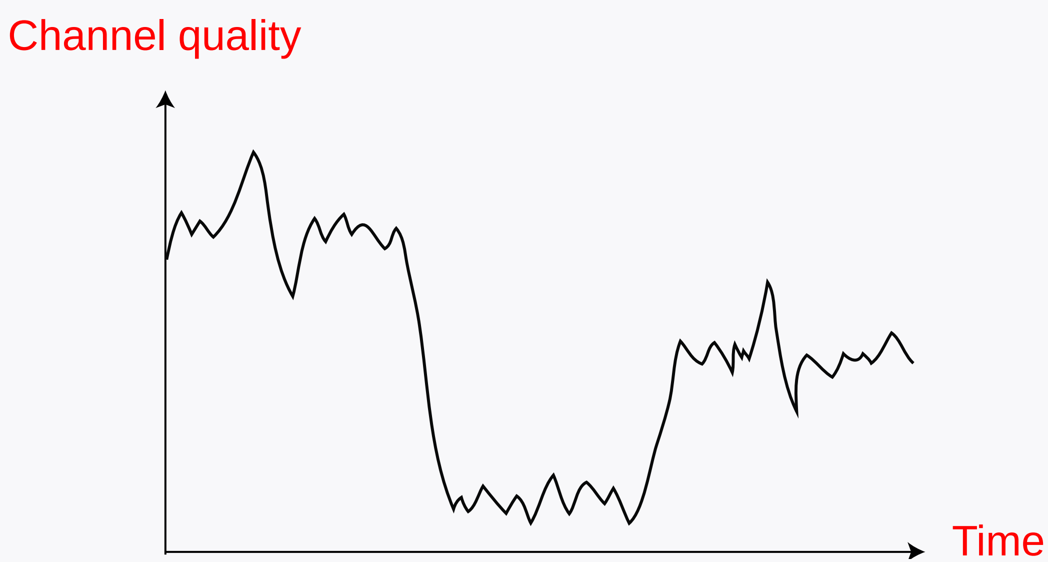


Motivation



Abstraction





Problem formulation

- Linear dynamical system

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + v_k, \\ y_k &= Cx_k + w_k \end{aligned}$$

A composed of unstable modes $|\lambda_1| \geq 1, \cdots, |\lambda_n| \geq 1$

- Time-varying channel

Slotted time-varying channel evolving at the same time scale of the system

Problem formulation

Objective: identify the trade-off between system's unstable modes and channel's rate to guarantee stability:

$$\sup_k ||x_k|| < \infty$$

or

$$\sup_k \mathbb{E}[||x_k||^2] < \infty$$

In the literature…

- Two main approaches to channel model

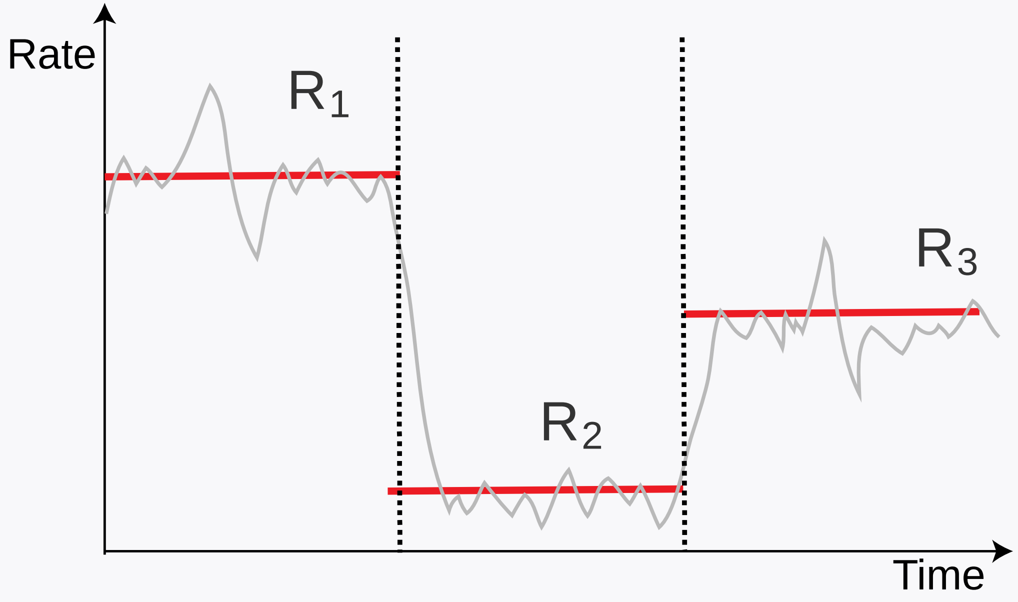
Information-theoretic approach (bit-rate)

Network-theoretic approach (packets)

Information-theoretic approach

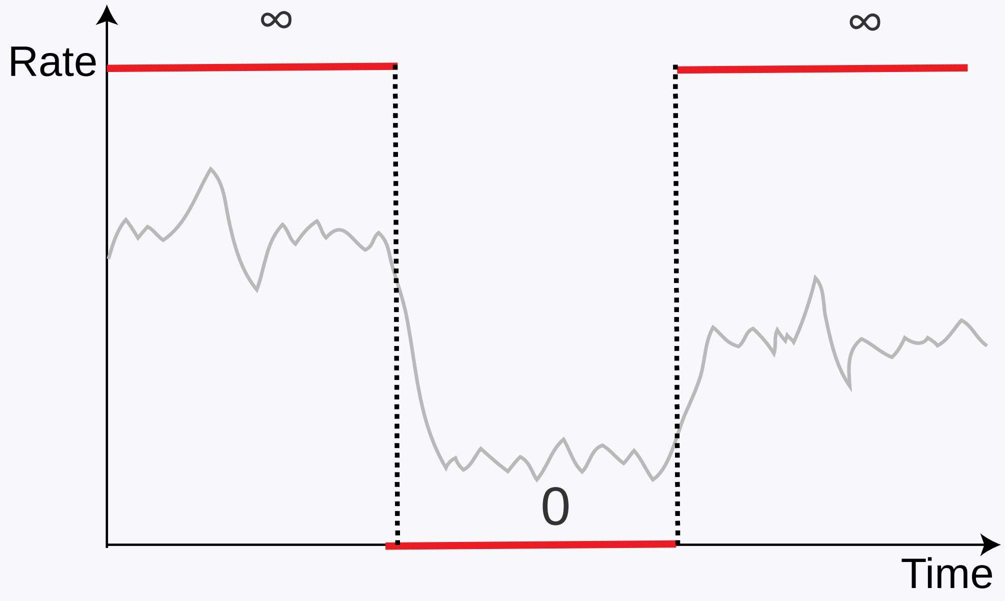
- A **rate-based** approach, transmit R_k bits/time

- Derive **data-rate theorems** quantifying how much rate is needed to construct a stabilizing quantizer/controller pair



Network-theoretic approach

- A **packet-based** approach (a packet models a real number)
- Determine the **critical packet loss probability** above which the system cannot be stabilized by any control scheme



Information-theoretic approach

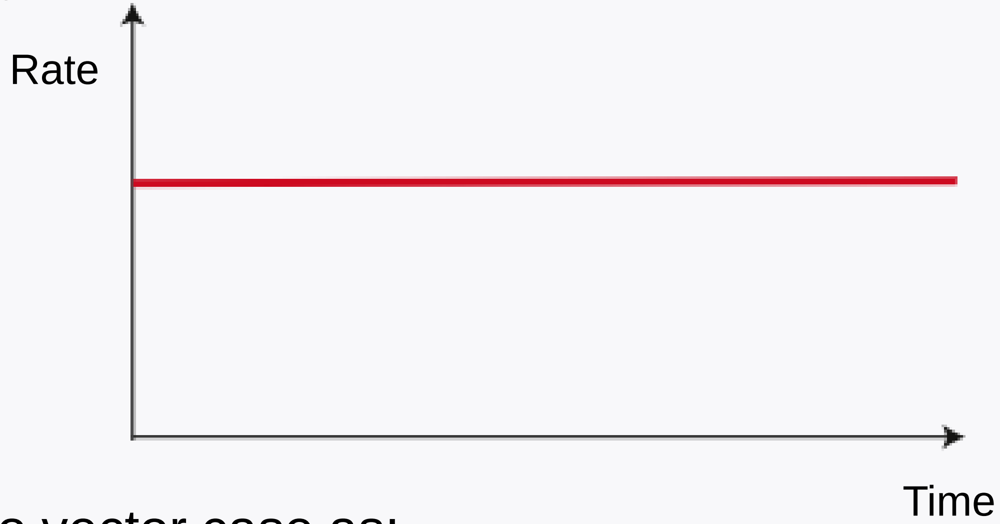
- Tatikonda-Mitter (IEEE-TAC 2002)

  Rate process: $\forall k \;\; R_k = R$ known at the transmitter

  Disturbances and initial state support: bounded

  Data rate theorem: $R > R_c = \log|\lambda|$

  a.s. stability



- Generalizes to vector case as:

$$R > \sum_u m_u \log|\lambda_u|$$

Information-theoretic approach

- Nair-Evans (SIAM-JCO 2004, best paper award)

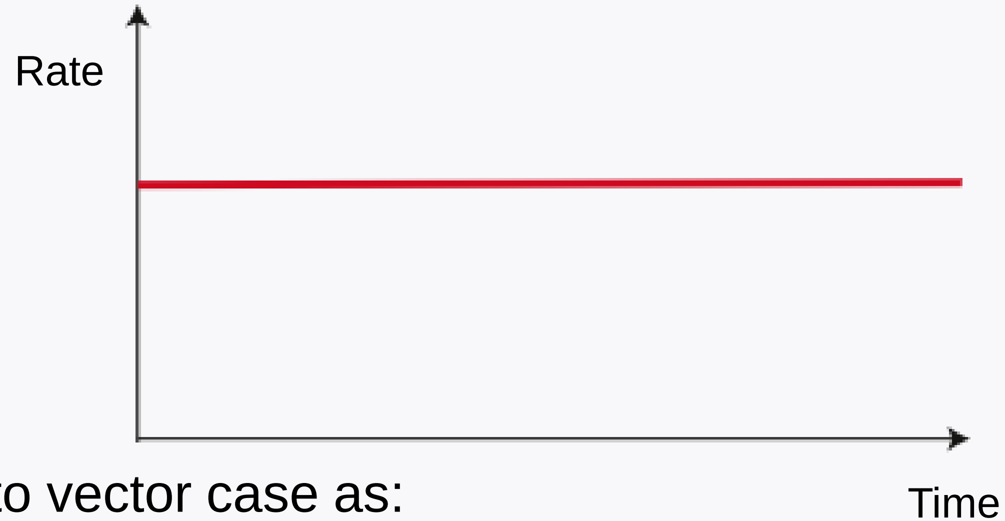
  Rate process: $\forall k \;\; R_k = R$ known at the transmitter

  Disturbances and initial state support: unbounded

  Bounded higher moment (e.g. Gaussian distribution)

  Data rate theorem: $R > R_c = \log|\lambda|$

  Second moment stability



- Generalizes to vector case as:

$$R > \sum_u m_u \log|\lambda_u|$$

Intuition

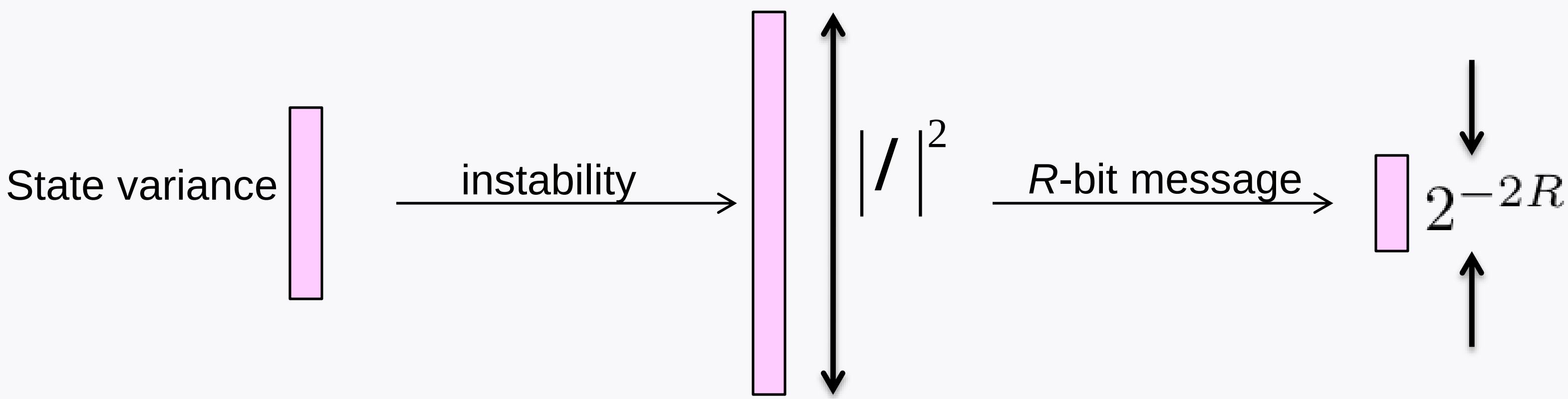
- Want to compensate for the expansion of the state during the communication process

- At each time step, the uncertainty volume of the state

$$\uparrow |\lambda|^2 \qquad \downarrow 2^{-2R}$$

- Keep the product less than one for second moment stability

$$R > \log|\lambda|$$



Information-theoretic approach

- Martins-Dahleh-Elia (IEEE-TAC 2006)

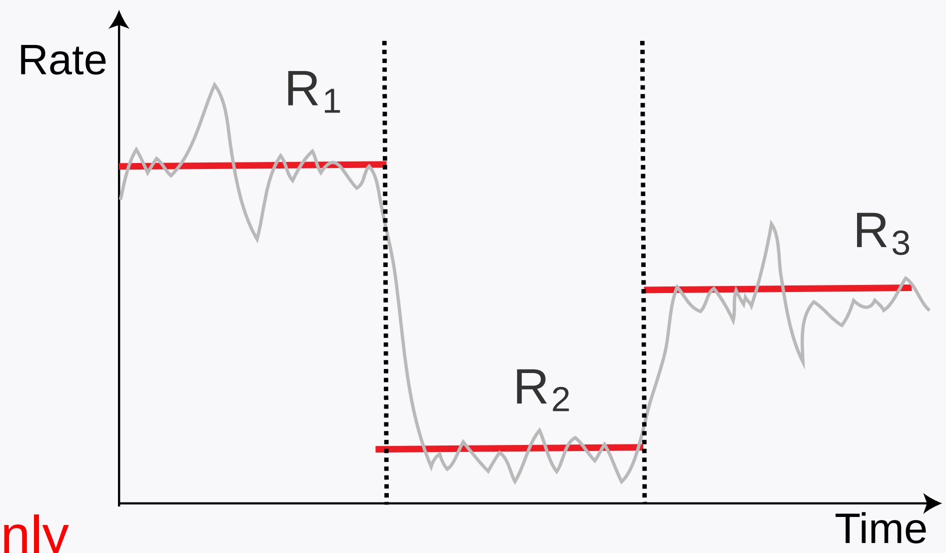
Rate process: $\{R_k\}$ i.i.d process distributed as *R*

Disturbances and initial state support: bounded

Causal knowledge channel: coder and decoder have knowledge of $\{R_i\}_{i=0}^{k}$

Data rate theorem: $|\lambda|^2 \mathbb{E}\left[2^{-2R}\right] < 1$

Second moment stability



- Scalar case only

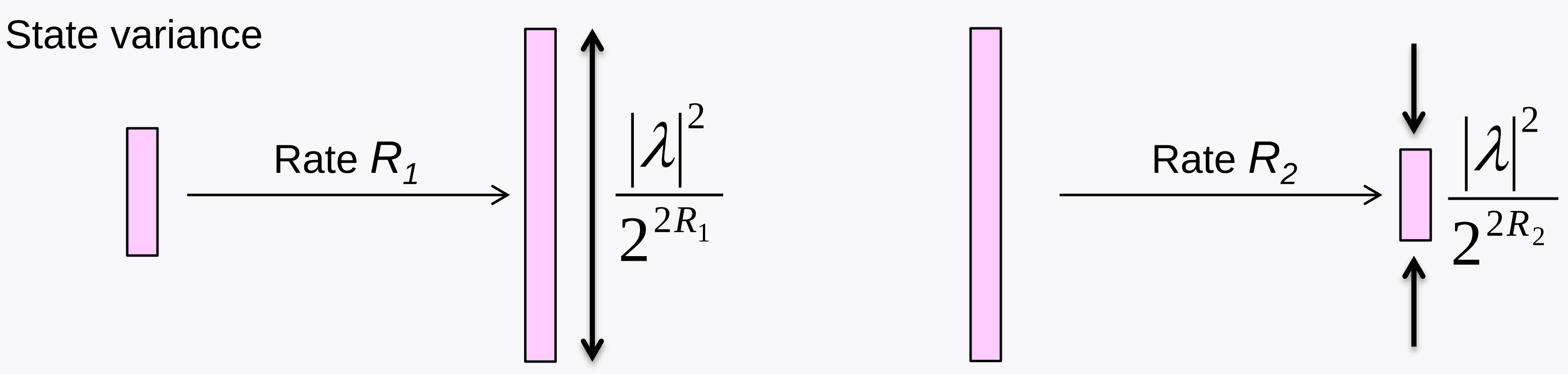
Intuition

- At each time step, the uncertainty volume of the state

$$\uparrow |\lambda|^2 \quad \downarrow 2^{-2R_i}$$

- Keep the *average* of the product less than one for second moment stability

$$|\lambda|^2 \mathbb{E}\left[2^{-2R}\right] < 1$$



Information-theoretic approach

- Minero-F-Dey-Nair (IEEE-TAC 2009)

Rate process: $\{R_k\}$ i.i.d process distributed as *R*

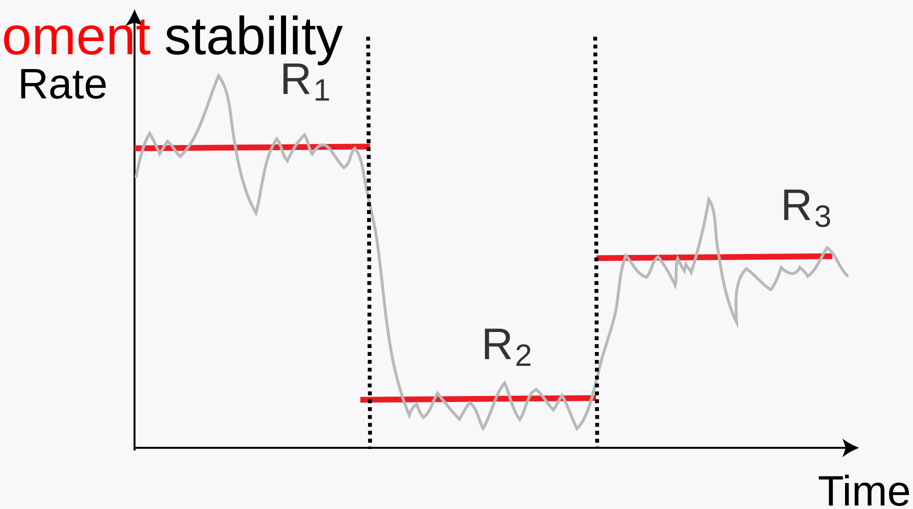
Disturbances and initial state support: unbounded

Bounded higher moment (e.g. Gaussian distribution)

Causal knowledge channel: coder and decoder have knowledge of $\{R_i\}_{i=0}^{k}$

Data rate theorem: $|\lambda|^2 \mathbb{E}\left[2^{-2R}\right] < 1$

Second moment stability



- Vector case, necessary and sufficient conditions almost tight

Proof sketches

- **Necessity:** Using the entropy power inequality find a recursion

$$\mathbb{E}[x_k^2] \geq |\lambda|^2\, \mathbb{E}[2^{-2R}]\, \mathbb{E}[x_{k-1}^2] + \text{const}$$

Thus,

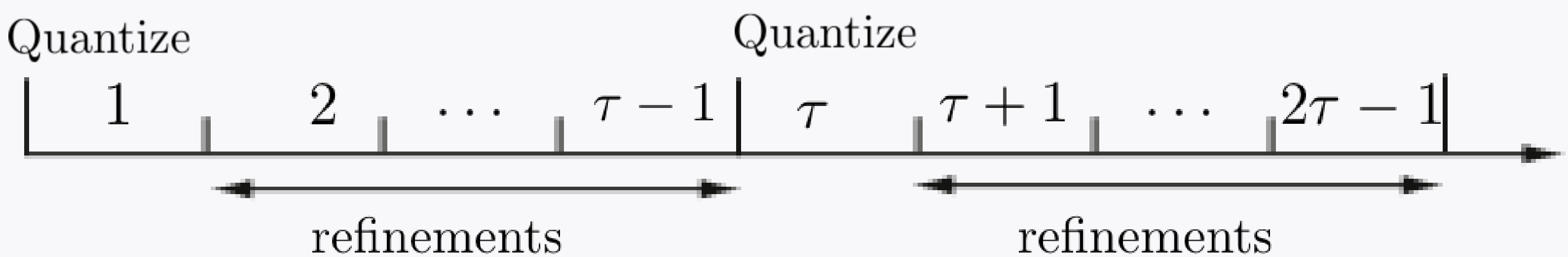
$$\sup_k [x_k^2] < \infty \implies |\lambda|^2\, \mathbb{E}[2^{-2R}] < 1$$

- **Sufficiency:** Difficulty is in the unbounded support, uncertainty about the state cannot be confined in any bounded interval, design an adaptive quantizer, avoid saturation, achieve high resolution through successive refinements.

Proof of sufficiency

- Divide time into cycles of fixed length τ (of our choice)

- Observe the system at the beginning of each cycle and send an initial estimate of the state

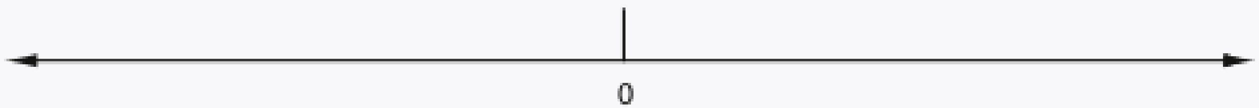


- During the remaining part of the cycle "refine" the initial estimate
- Number of bits per cycle is a random variable dependent of the rate process
- Use refined state at the end of cycle for control

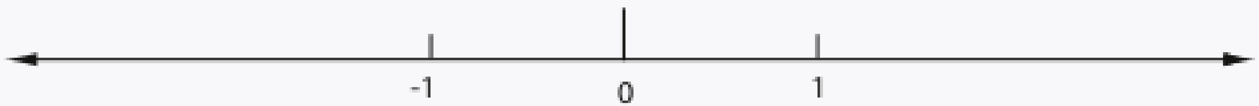
Adaptive quantizer

- Constructed recursively

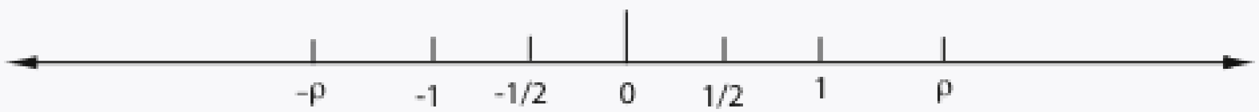
$R_k = $ 1-bit



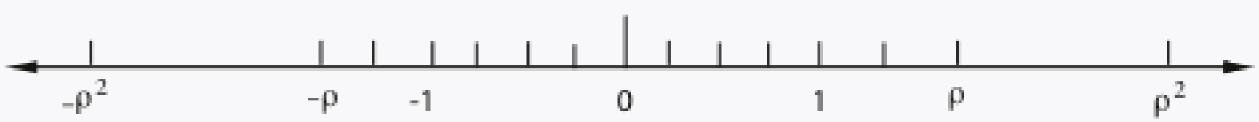
$R_k = $ 2-bit



$R_k = $ 3-bit

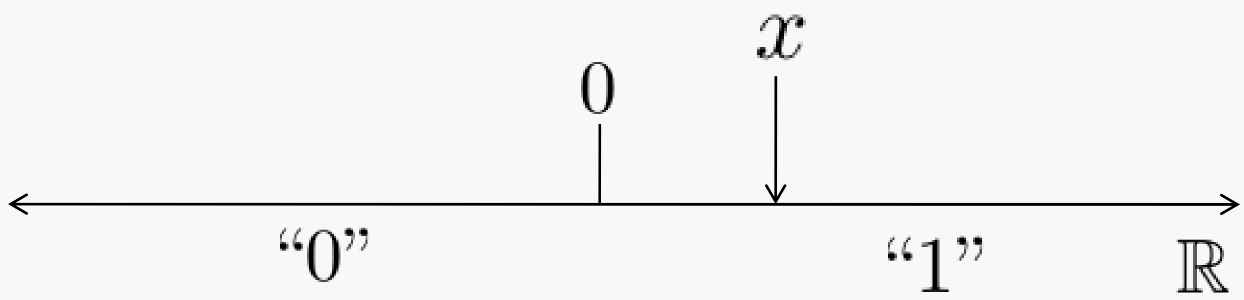


$R_k = $ 4-bit



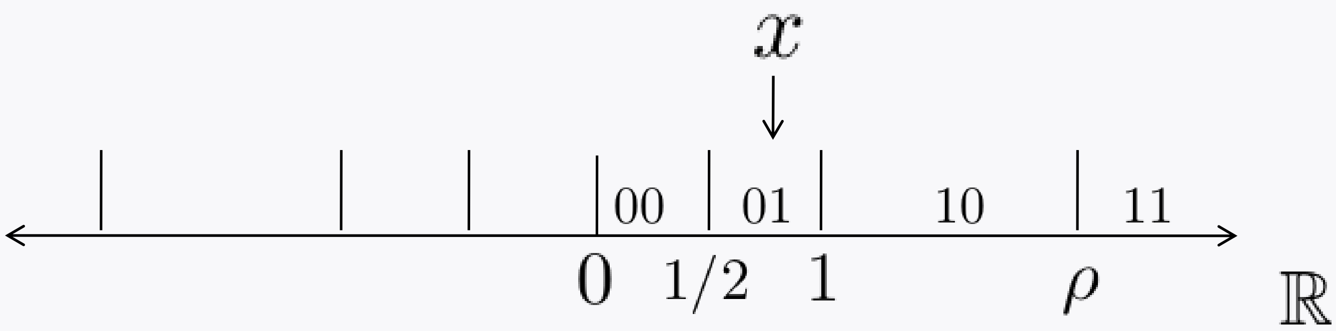
Successive refinements: example

- Suppose we need to quantize a positive real value x
- At time $k = 1$ suppose $R_1 = 1$
- With one bit of information the decoder knows that $x > 0$



Successive refinements: example

- At time $k = 2$ suppose $R_2 = 2$
- Partition the real axis according to the adaptive 3-bit quantizer
- Label only the partitions on the positive real line (2 bits suffice)
- After receiving 01 the decoder knows that $x \in [1/2, 1)$, thus the initial estimate $x \in [0, \infty)$ has been refined
- The scheme works as if we knew ahead of time that $R_1 + R_2 = 3$



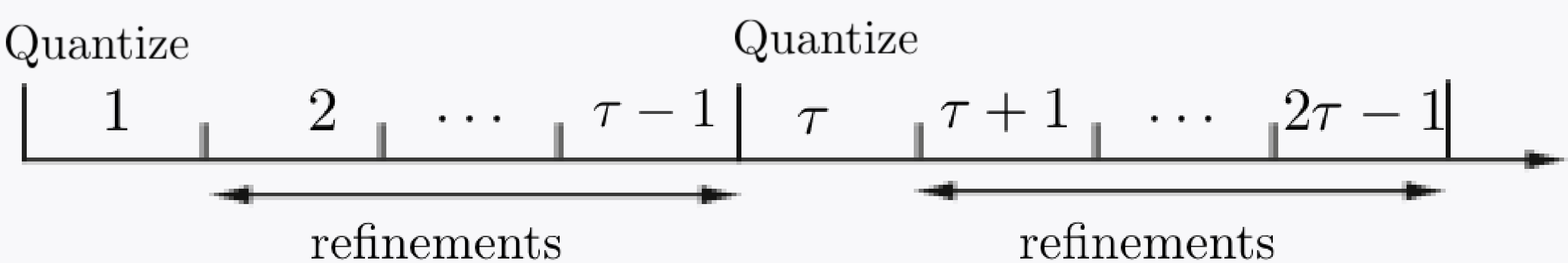
Proof of sufficiency

- Find a recursion for $\mathbb{E}[x_{k\tau}^2]$

$$\mathbb{E}[x_{k\tau}^2] \leq \text{const} \left(\mathbb{E}\left[\frac{|\lambda|^2}{2^{2R}}\right]\right)^{\tau} \mathbb{E}[x_{(k-1)\tau}^2] + \text{const}$$

- Thus, for τ large enough

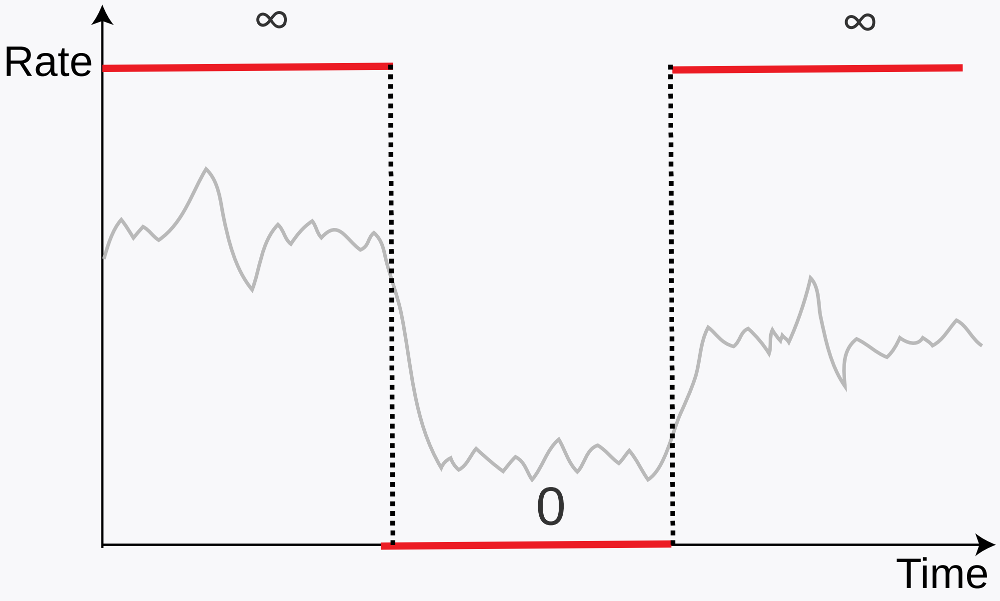
$$\mathbb{E}\left[\frac{|\lambda|^2}{2^{2R}}\right] < 1 \implies \text{const} \left(\mathbb{E}\left[\frac{|\lambda|^2}{2^{2R}}\right]\right)^{\tau} < 1$$



Network-theoretic approach

- A **packet-based** approach (a packet models a real number)

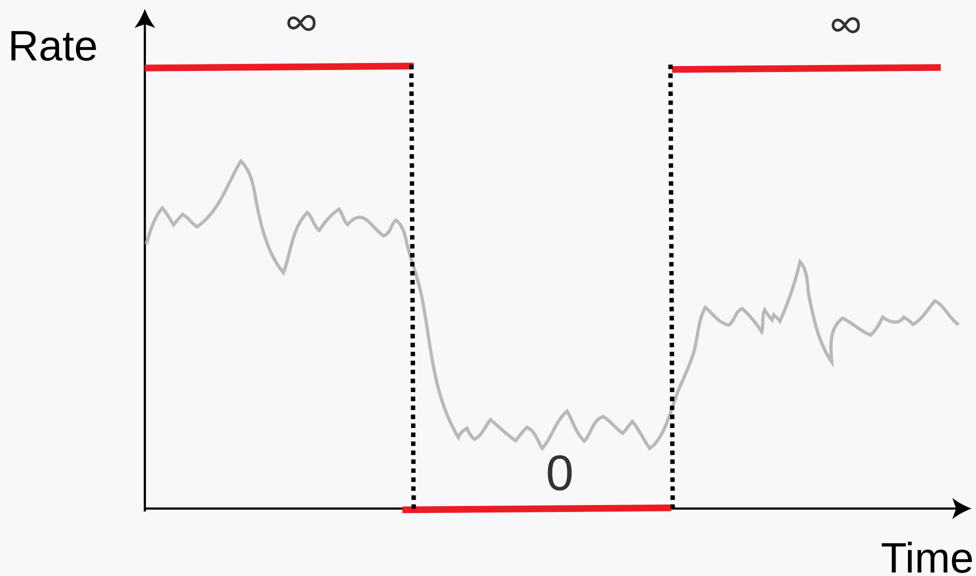
$$R_k = \begin{cases} \infty & \text{w.p. } 1-p \\ 0 & \text{w.p. } p \end{cases}$$



Critical dropout probability

- Sinopoli-Schenato-F-Sastry-Poolla-Jordan (IEEE-TAC 2004)
- Gupta-Murray-Hassibi (System-Control-Letters 2007)

$$p < p_c = \frac{1}{|\lambda|^2}$$



- Generalizes to vector case as:

$$p < p_c = \max_i \frac{1}{|\lambda_i|^2}$$

Critical dropout probability

- Can be viewed as a special case of the information-theoretic approach
- Gaussian disturbance requires unbounded support data rate theorem of Minero, F, Dey, Nair, (2009) to recover the result

$$\mathbb{E}\left[\frac{|\lambda|^2}{2^{2R}}\right] = p\frac{|\lambda|^2}{2^0} + (1-p)\frac{|\lambda|^2}{2^{2r}} < 1$$

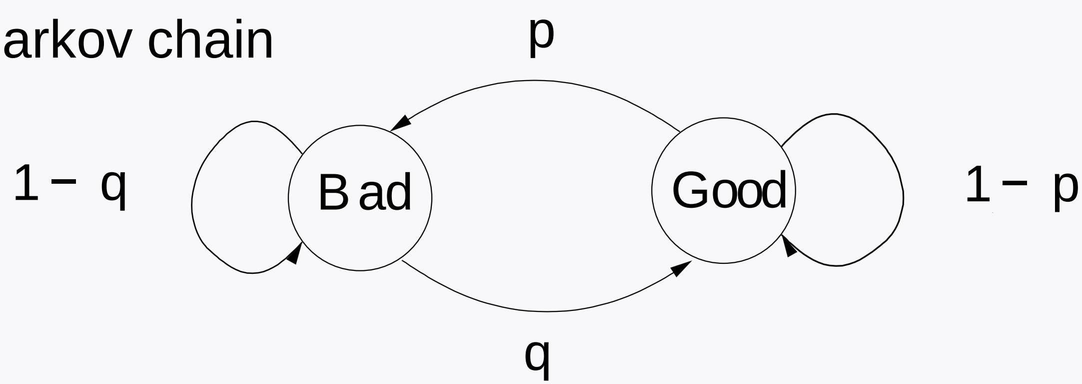
$$\implies p < \frac{1}{|\lambda|^2}, \text{ as } r \to \infty$$

Stabilization over channels with memory

- Gupta-Martins-Baras (IEEE-TAC 2009)

Network theoretic approach

Two-state Markov chain



- Critical “recovery probability”

$$q > q_c = 1 - \frac{1}{|\lambda^2|}$$

Stabilization over channels with memory

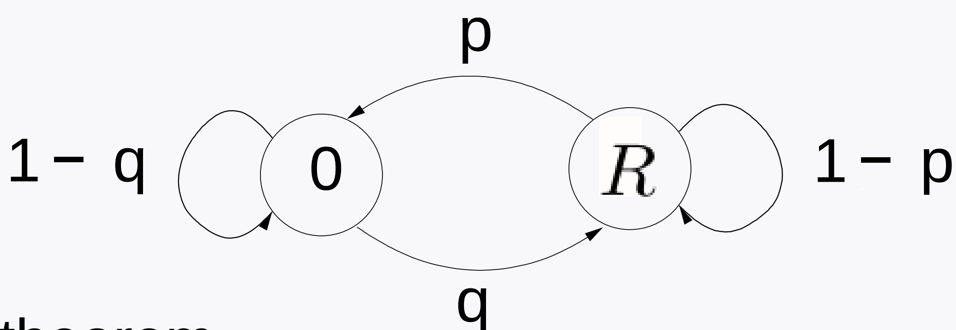
- You-Xie (IEEE-TAC 2010)

  Information-theoretic approach

  Two-state Markov chain, fixed *R* or zero rate

  Disturbances and initial state support: unbounded

  Let T be the excursion time of state *R*



- Data-rate theorem

  $$R > R_c = \frac{1}{2} \log \mathbb{E}[|\lambda|^{2T}]$$

- For $R \to \infty$ recover the critical probability $q > q_c = 1 - \frac{1}{|\lambda^2|}$

Intuition

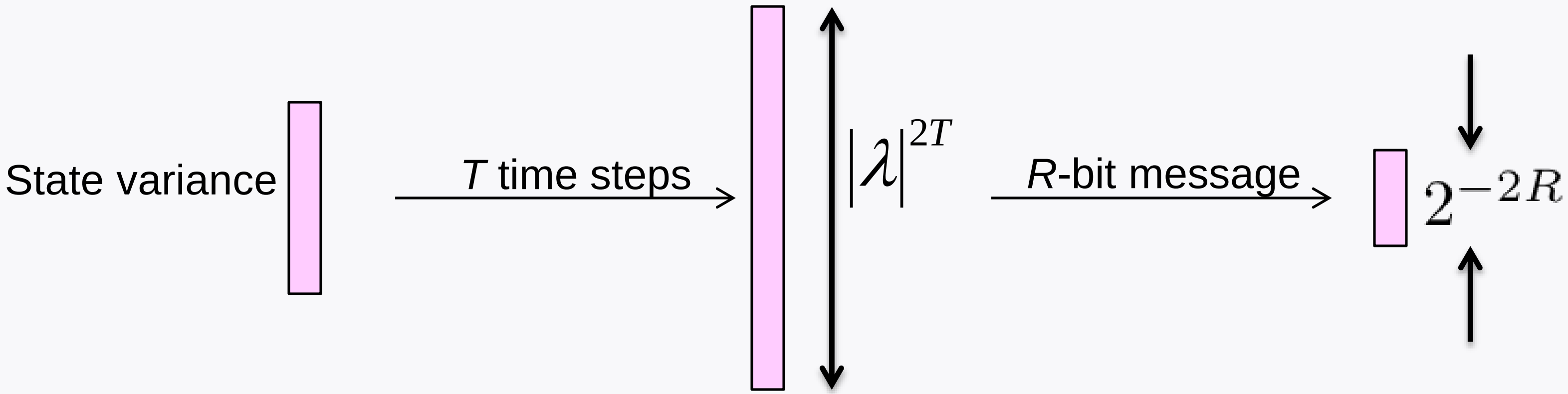
- Send *R* bits after T time steps

- In T time steps the uncertainty volume of the state

$$\uparrow |\lambda|^{2T} \qquad \downarrow 2^{-2R}$$

- Keep the *average* of the product less than one for second moment stability

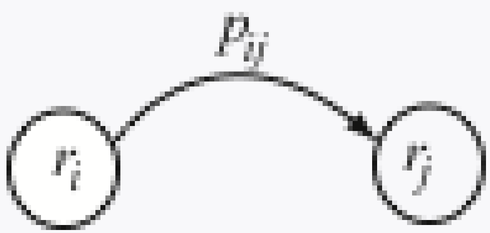
$$R > \frac{1}{2}\log \mathbb{E}[|\lambda|^{2T}]$$



Stabilization over channels with memory

- Coviello-Minero-F (IEEE-TAC 2013)

  Information-theoretic approach
  Disturbances and initial state support: unbounded
  Time-varying rate $R_k \in \{r_1, \cdots, r_n\}$
  Arbitrary positively recurrent time-invariant Markov chain of *n* states $p_{ij} = P\{R_{k+1} = r_j | R_k = r_i\}$



- Obtain a general data rate theorem that recovers all previous results using the theory of Jump Linear Systems

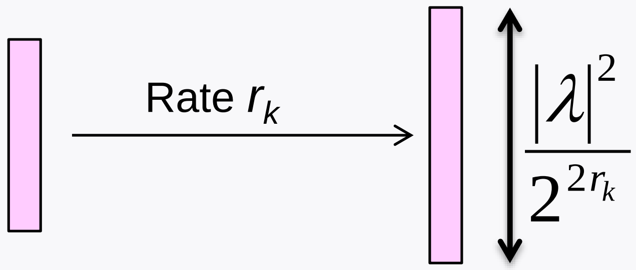
Markov Jump Linear System

- Define an auxiliary dynamical system (MJLS)

$$z_{k+1} = \frac{|\lambda|}{2^{R_k}} z_k + c$$

$$z_0 < \infty, c \geq 0$$

State variance



Markov Jump Linear System

- Let H be the $n \times n$ matrix defined by the transition probabilities and the rates

$$h_{ij} = \frac{1}{2^{2r_j}} p_{ji}$$

- Let $r(H)$ be the spectral radius of *H*
- The MJLS is mean square stable iff $|\cdot|^2 \, r(H) < 1$
- Relate the stability of MJLS to the stabilizability of our system

Data rate theorem

- Stabilization in mean square sense over Markov time-varying channels is possible if and only if the corresponding MJLS is mean square stable, that is:

$$|\lambda|^2 \rho(H) < 1$$

Proof sketch

The second moment of the system state is lower bounded and upper bounded by two MJLS with the same dynamics

Lower bound: using the entropy power inequality

$$\mathbb{E}[|x_k|^2] \geq \frac{1}{2\pi e}\mathbb{E}[z_k^2], \quad \forall k \geq 0$$

$$z_{k+1} = \frac{|\lambda|}{2^{R_k}} z_k + c, \quad z_0 = e^{h(x_0)}, c = e^{h(w)}$$

$\implies |\lambda|^2 \rho(H) < 1$ is a necessary condition

Proof sketch

Upper bound: using an adaptive quantizer at the beginning of each cycle the estimation error is upper bounded as

$$\mathbb{E}[|x_{j\tau} - \hat{x}_{j\tau}|^2] \leq \mathbb{E}[z_{j\tau}^2] + \text{const}$$

where

$$z_{j\tau} = \phi \frac{|\lambda|^\tau}{2^{R_{(j-1)\tau} + \cdots + R_{j\tau-1}}} z_{(j-1)\tau} \qquad \phi > 1$$

This represents the evolution at times $\tau, 2\tau, 3\tau \ldots$ of a MJLS

$$z_k = \phi^{1/\tau} \frac{|\lambda|}{2^{R_k}} z_{k-1}$$

A sufficient condition for stability of $\{z_k\}$ is

$$\phi^{2/\tau} |\lambda|^2 \rho(H) < 1$$

Proof sketch

Assuming $|\lambda|^2 \rho(H) < 1$

Can choose τ large enough so that

$\phi^{2/\tau} |\lambda|^2 \rho(H) < 1$

MJLS $\{z_k\}$ is stable

Second moment of estimation error at the beginning of each cycle is bounded

The state remains second moment bounded.

Previous results as special cases

Coviello, Minero, F (2011)

Tatikonda, Mitter (2002)
Nair Evans (2004)
Martins Dahleh Elia (2006)
Minero, F, Dey, Nair (2009)

Gupta Murray Hassibi (2007)
Gupta Martins Baras (2009)

You Xie (2010)

Previous results as special cases

- **iid** bit-rate

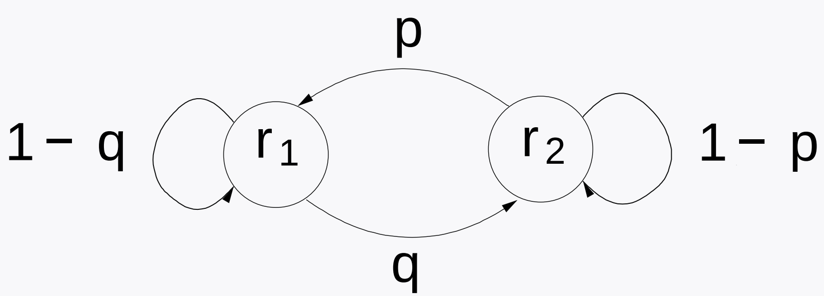
- Data rate theorem reduces to

$$|\lambda|^2 \rho(H) = |\lambda|^2 (2^{-2r_1}, \cdots, 2^{-2r_n})(p_1, \cdots, p_n)^T$$
$$= |\lambda|^2 \mathbb{E}[2^{-2R}] < 1$$

Recover Minero, F, Dey, Nair (2009)

Previous results as special cases

- **Two-state Markov** channel



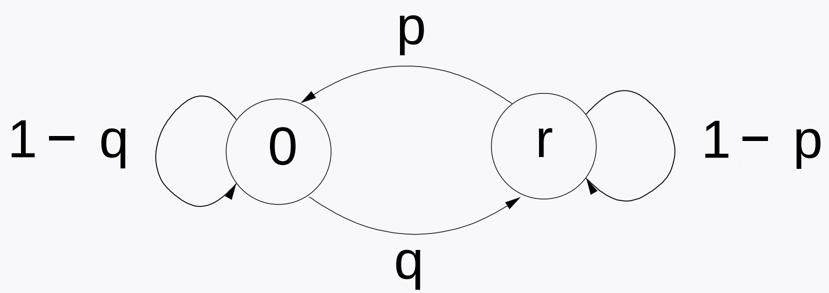
- Data rate theorem reduces to

$$|\lambda|^2\rho(H) = \frac{|\lambda|^2}{2}\mathrm{Tr}\ (H)^2 + \frac{|\lambda|^2}{2}\sqrt{\mathrm{Tr}\ (H)^2 - 4\ \det(H)} < 1$$

Previous results as special cases

- **Two-state Markov** channel

$$r_1 = 0, \; r_2 = r$$



- Data rate theorem further reduces to

$$|\lambda|^2 < \begin{cases} \frac{1}{\mathrm{tr}(H)} & \text{if } \det(H) = 0 \\ \frac{\mathrm{tr}(H) - \sqrt{\mathrm{tr}(H) - 4\det(H)}}{2\det(H)} & \text{otherwise} \end{cases}$$

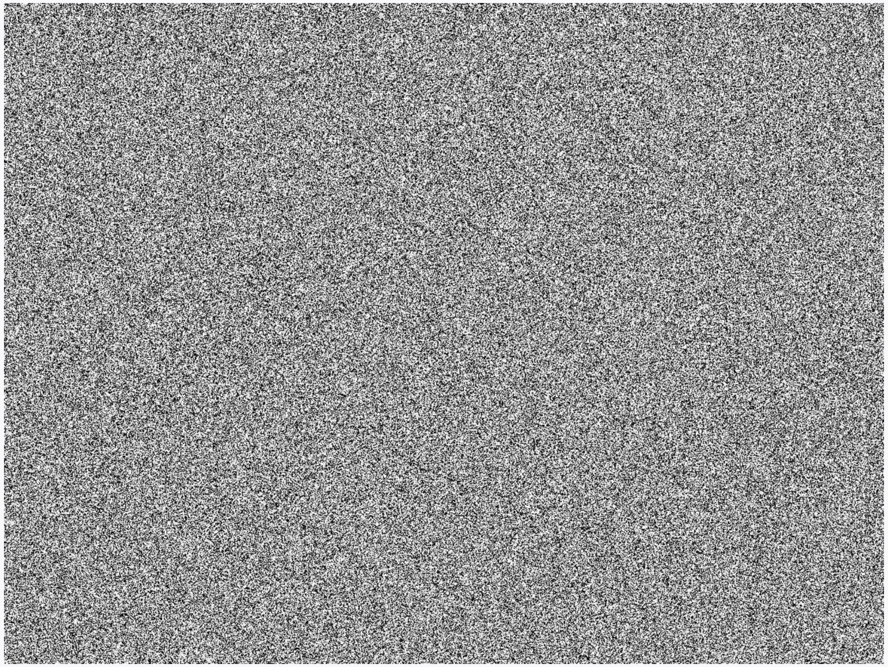
- From which it follows

$$r > \frac{1}{2} \log \mathbb{E}[|\lambda|^{2T}]$$

recovering You, Xie (2010)

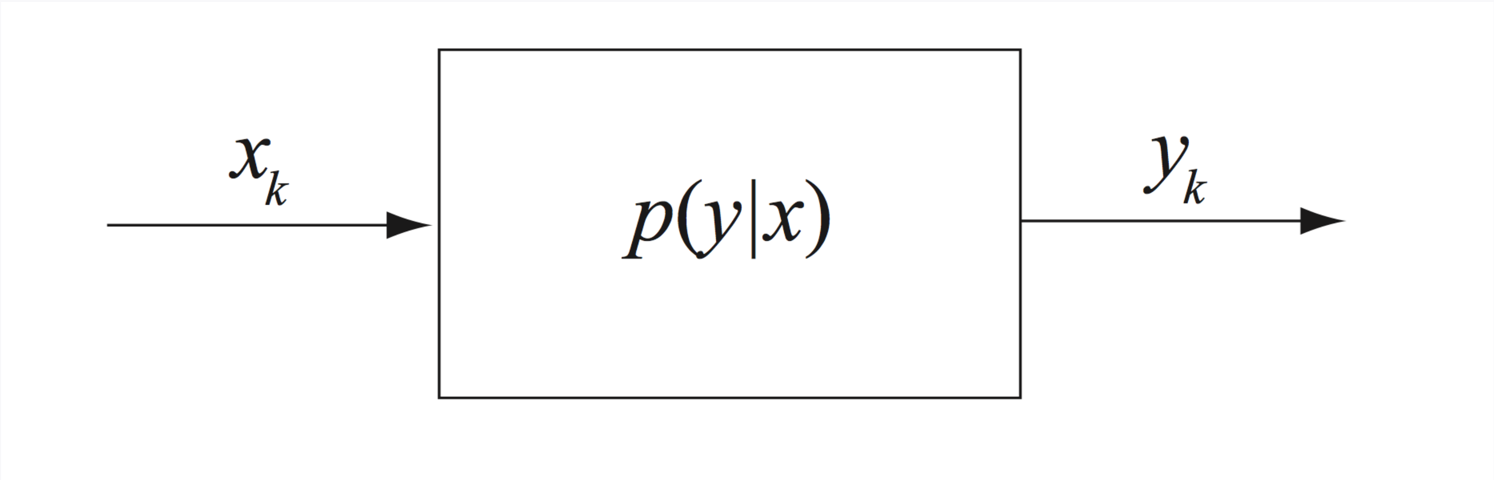
What next

- Is this the end of the journey?



- No! journey is still wide open
- … noisy channels, beyond erasures

Discrete memory-less channel (DMC)



- The communication channel is a stochastic system described by the conditional probability distribution of the channel output given the channel input
- Need to keep track of the state in the presence of decoding errors

Insufficiency of Shannon capacity

- Example: i.i.d. erasure channel

$$R_k \sim R = \begin{cases} r & \text{w.p. } 1-p \\ 0 & \text{w.p. } p \end{cases}$$

- Data rate theorem:

$$|\lambda|^2 \mathbb{E}(2^{-2R}) < 1 \quad \Longrightarrow \quad |\lambda|^2 (2^{-2r}(1-p) + p) < 1$$

$$\text{as } r \to \infty \quad p < \frac{1}{|\lambda|^2}$$

- Shannon capacity:

$$C = (1-p)r \to \infty$$

Capacity with stronger reliability constraints

Sahai-Mitter (IEEE-IT 2006)

- Shannon capacity C soft reliability constraint $P_{err} \to 0$
- Zero-error capacity C_0 hard reliability constraint $P_{err} = 0$
- Anytime capacity C_A medium reliability constraint

$\mathbb{P}((\hat{M}_{0|k}, \ldots, \hat{M}_{d|k}) \neq (M_0, \ldots, M_d)) = O(2^{-\alpha d}) \quad \text{for all } d \leq k$

$$C_A(\alpha) = \sup_r (r, \alpha)$$

$$C_0 \leq C_A \leq C$$

$$C_A(0) = C, \quad C_A(\infty) = C_0$$

Alternative formulations

- Undisturbed systems
- Tatikonda-Mitter (IEEE-AC 2004)
- Matveev-Savkin (SIAM-JCO 2007)

$C > \log|\lambda|$ a.s. stability

- Disturbed systems (bounded)
- Matveev-Savkin (IJC 2007)

$C_0 > \log|\lambda|$ a.s. stability

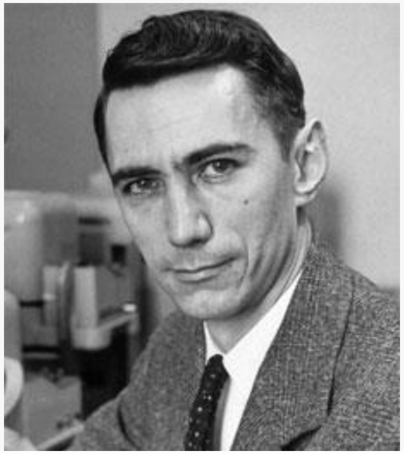
- Sahai-Mitter (IEEE-IT 2006)

$C_A(\eta \log|\lambda|) > \log|\lambda|$ moment stability

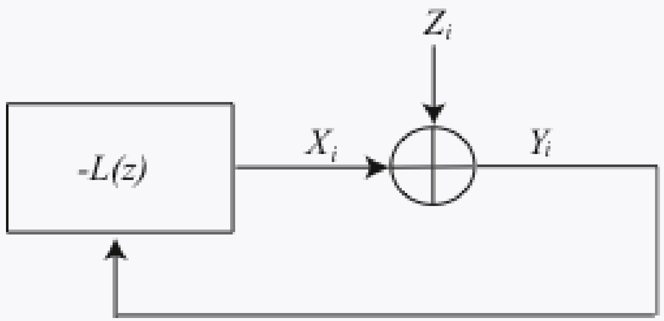
Anytime reliable codes: Shulman (1996), Ostrovsky, Rabani, Schulman (2009), Como, Fagnani, Zampieri (2010), Sukhavasi, Hassibi (2011)

The Bode-Shannon connection

- Connection with the capacity of channels with feedback
- Elia (IEEE-TAC 2004)
- Ardestanizadeh-F (IEEE-TAC 2012)
- Ardestanizadeh-Minero-F (IEEE-IT 2012)



Control over a Gaussian channel Ardestanizadeh, F (2012)



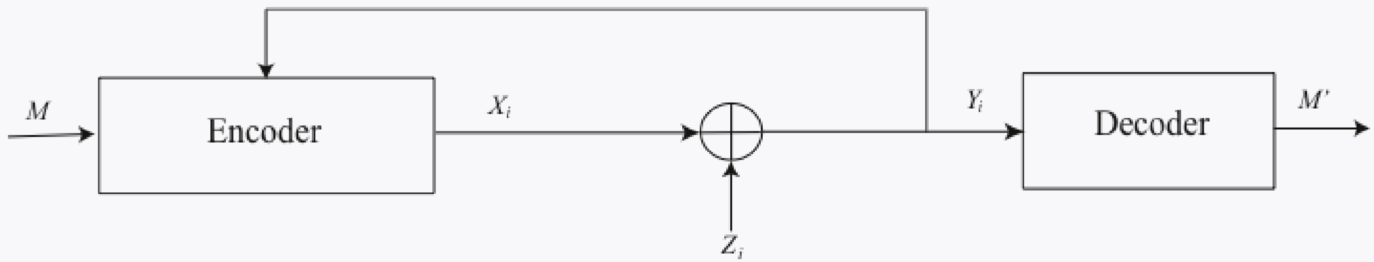
$$U = \sum_{i \in \mathcal{U}} \log|\lambda_i|$$ Instability

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} |T(e^{j\omega})|^2 S_Z(\omega)d\omega \leq P$$ Power constraint

$$T(z) = \frac{L(z)}{1+L(z)}$$ Complementary sensitivity function

Z_i Stationary (colored) Gaussian noise

Control over a Gaussian channel Ardestanizadeh, F (2012)



$$\mathbb{E}(X_i^2) \leq P \quad \forall i$$ Power constraint

C_F Feedback capacity

$$\sup_{\mathcal{L}} U = C_F$$

- The largest instability U over all LTI systems that can be stabilized by unit feedback over the stationary Gaussian channel, with power constraint P corresponds to the Shannon capacity C_F of the stationary Gaussian channel with feedback [Kim(2010)] with the same power constraint P.

Communication using control

- This duality between control and feedback communication for Gaussian channels can be exploited to design communication schemes using control tools
- MAC, broadcast channels with feedback

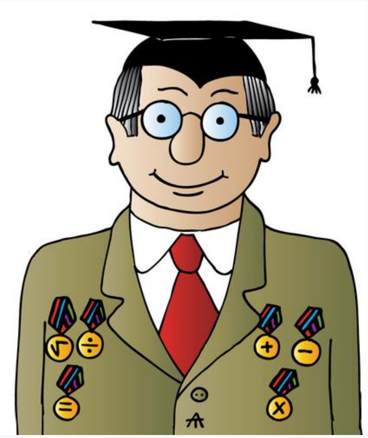
- Elia (IEEE-TAC 2004)
- Ardestanizadeh-Minero-F (IEEE-IT 2012)

Summary of results

Condition	Channel	Stabilization	Disturbance
$C \gtrsim U$	DMC	a.s.	0
$C_0 \gtrsim U$	DMC	a.s	bounded
$C_A(\eta \log \lvert\lambda\rvert) \gtrsim \eta \log \lvert\lambda\rvert$	DMC	η-moment	bounded
$\lvert\lambda\rvert^2(2^{-2r}(1-p)+p) < 1$	Erasure	2^{nd} moment	unbounded
$C > U$	AWGN	η-moment	unbounded
$C_F = \sup U$	ACGN	2^{nd} moment	0

Conclusion

- Data-rate theorems for stabilization over time-varying rate channels, after a beautiful journey of about a decade, are by now fairly well understood
- The journey (quest) for noisy channels is still going on
- The terrible thing about the quest for truth is that you may find it



- For papers: www.circuit.ucsd.edu/~massimo/papers.html