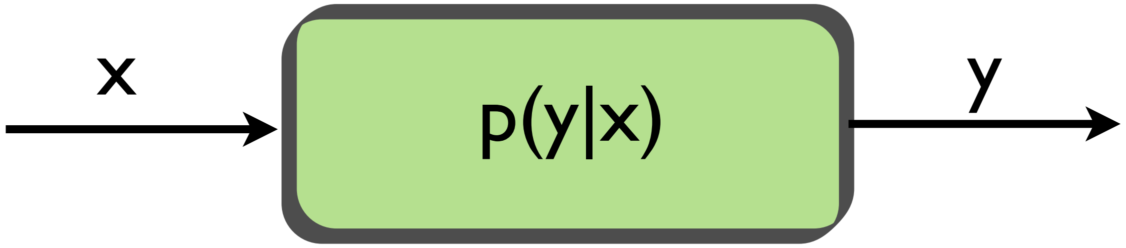
New Vistas in Coding - The Interplay Between Coding and Control

Ravi Teja Sukhavasi and Babak Hassibi
presented by Sormeh Shadbakht

Cyber Physical Systems Meeting, August 1-2, 2011

Single User Information Theory

Single-user information theory (Shannon 1948) deals with the study of the fundamental limits of reliable information transmission between a sender and a receiver over a noisy channel.



$$C = \max_{p_X(.)} \{H(x) + H(y) - H(x, y)\}$$

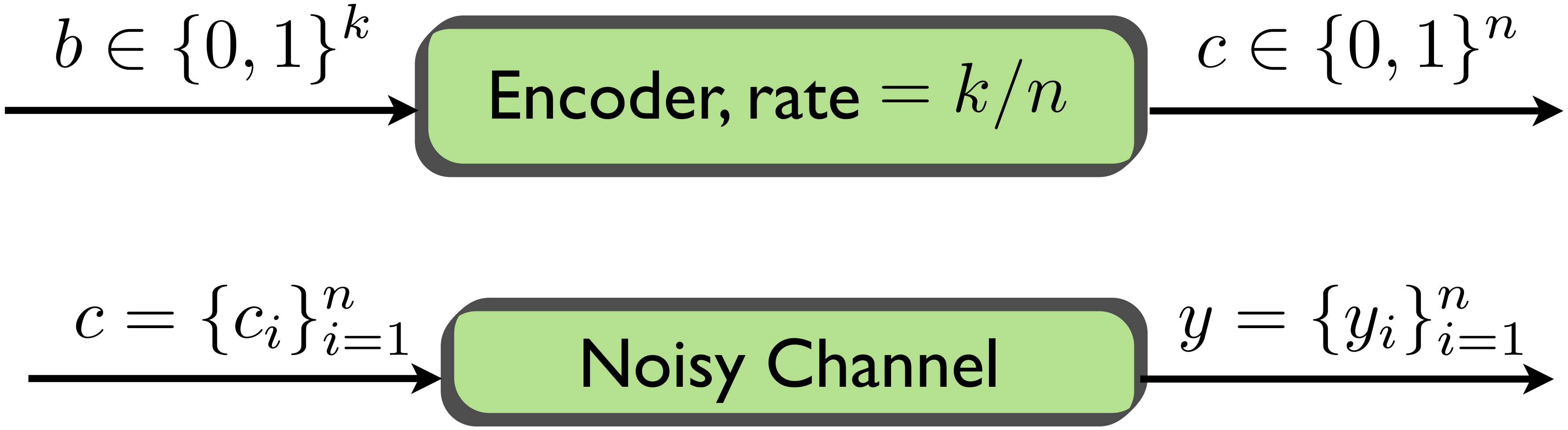
Key Idea: Block Coding

- the behavior of the channel over a single channel use is unpredictable
- but the behavior over many channel uses is:

  if the channel introduces error with probability p, say, over $n \gg 1$ channel uses it introduces $\approx np$ errors.

Coding Theory

The set of all codewords, c, is denoted by $\mathcal{C}, \ (|\mathcal{C}| = 2^k)$



$b \in \{0,1\}^k$

Encoder, rate $= k/n$

$c \in \{0,1\}^n$

$c = \{c_i\}_{i=1}^n$

Noisy Channel

$y = \{y_i\}_{i=1}^n$

Maximum Likelihood Decoder: $\hat{c} = \arg\max_{c \in \mathcal{C}} p(y|c)$

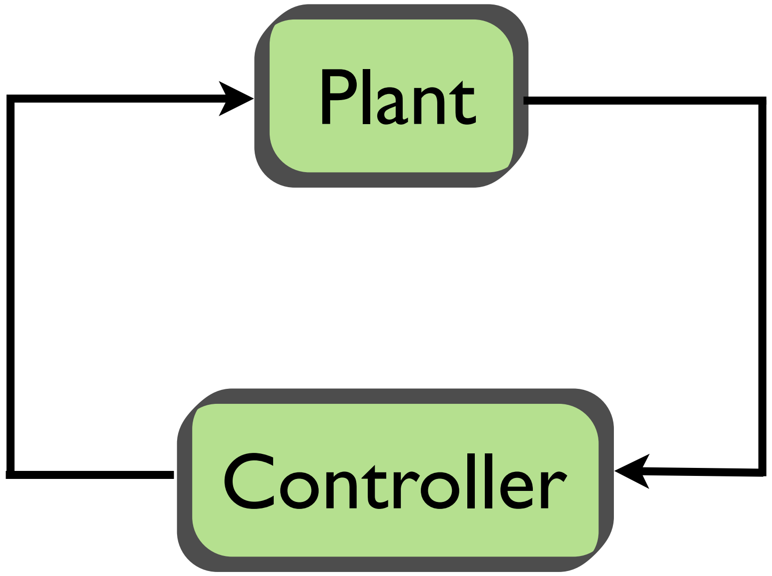
Shannon: For all rates $k/n < C$, *there exists a sequence of codes,* such that

$$\lim_{n \to \infty} P(\hat{c} \neq c) = 0$$

After 60 years of coding theory, there are currently many efficient codes that approach the Shannon limit.

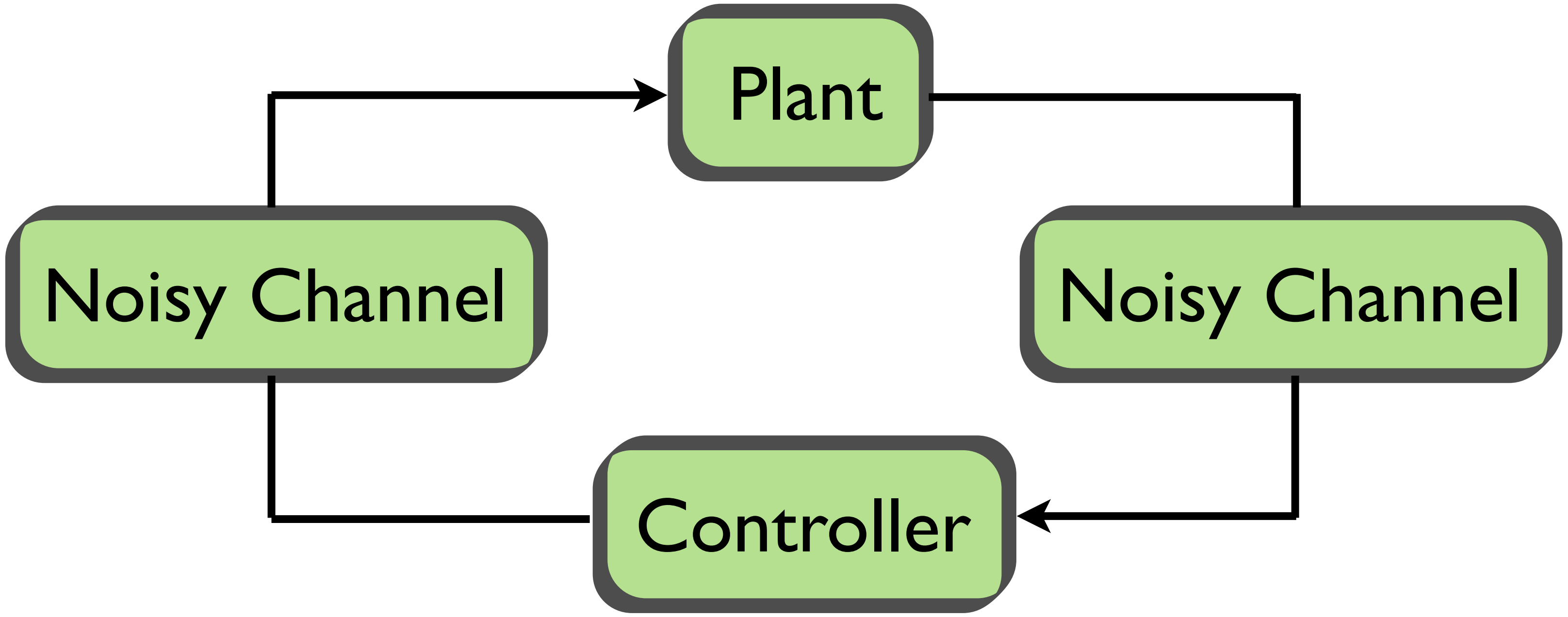
Reliability at the cost of ***delay***

Control Theory



- In control theory, we observe the output of a dynamical system and design a controller to regulate its behavior
- Controller needs to apply control action in *real-time*, delay can result in loss of performance and/or instability
- Very rich theory has been developed (e.g., LQG control, Kalman filtering, H^∞ filtering, separation principle, etc)
- Virtually no interaction with information theory (plant and controller co-located , no measurement loss)

But, what if...?



- increasingly many applications where systems (autonomous agents, sensor/actuator networks, smart grid etc.) are remotely controlled and where measurement and control signals are transmitted across noisy channels.

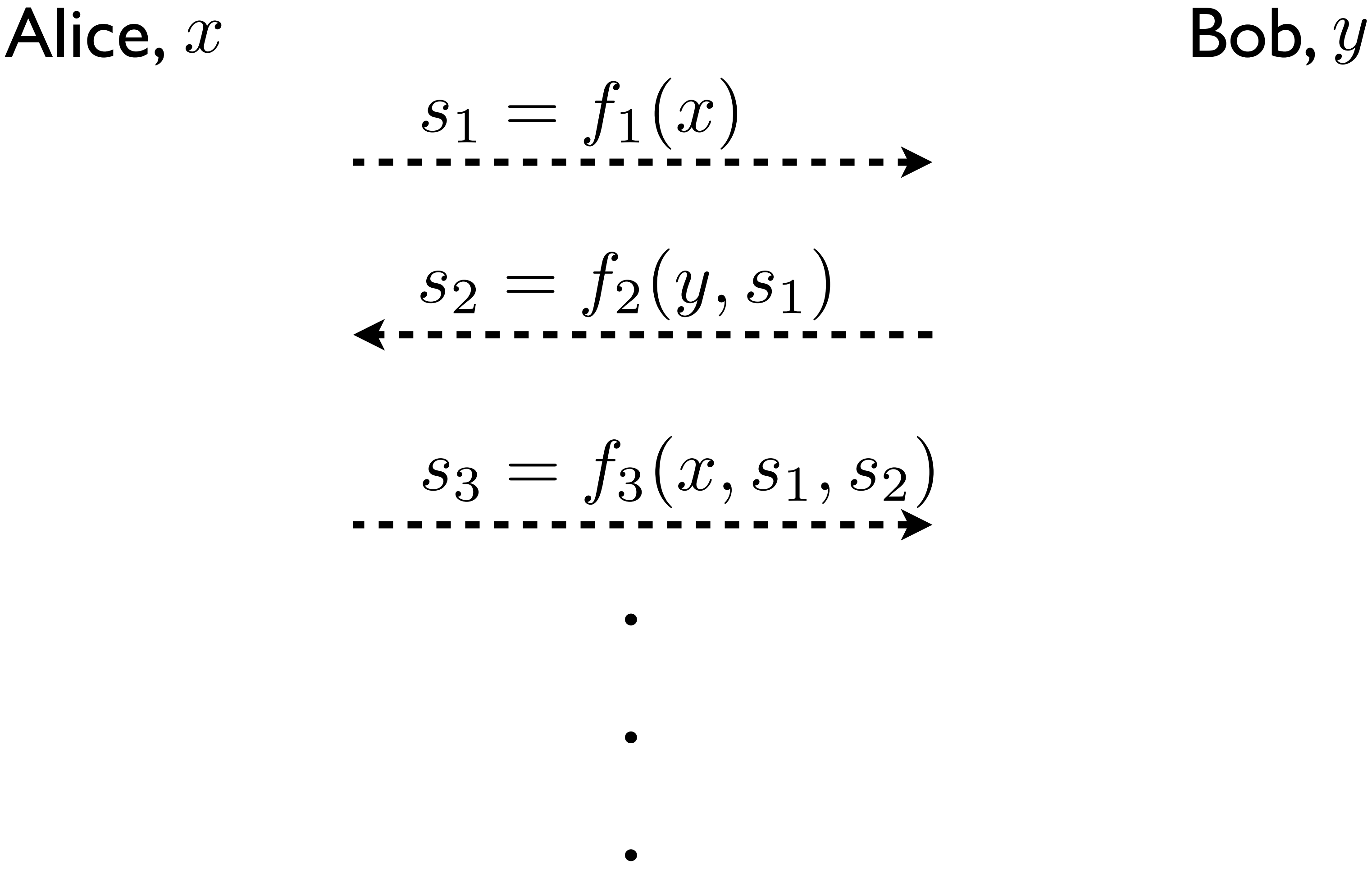
- conventional channel codes do not work as the ensuing delay might lead to instability

- *Can we live with noisy channels?* No! (e.g., if the noisy channels are erasure channels, Sinopoli et al(2005) showed that if the erasure probability is high enough then closed loop system is unstable)

- So, what do we need to guarantee the stability of the closed loop systems?

Coding for Interactive Communication

Consider a two-party communication system



Alice, x

Bob, y

$s_1 = f_1(x)$

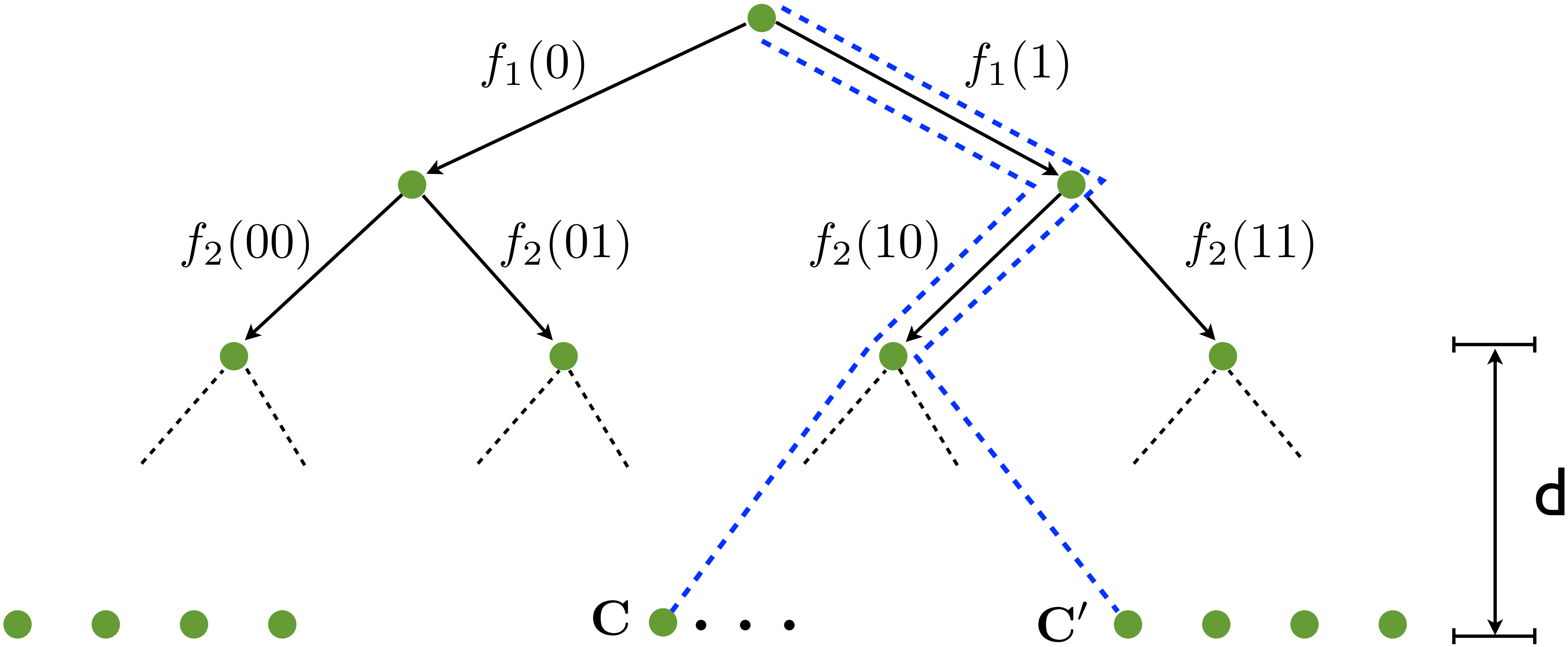
$s_2 = f_2(y, s_1)$

$s_3 = f_3(x, s_1, s_2)$

$\vdots$

Can one do this reliably over noisy links?

Tree Codes



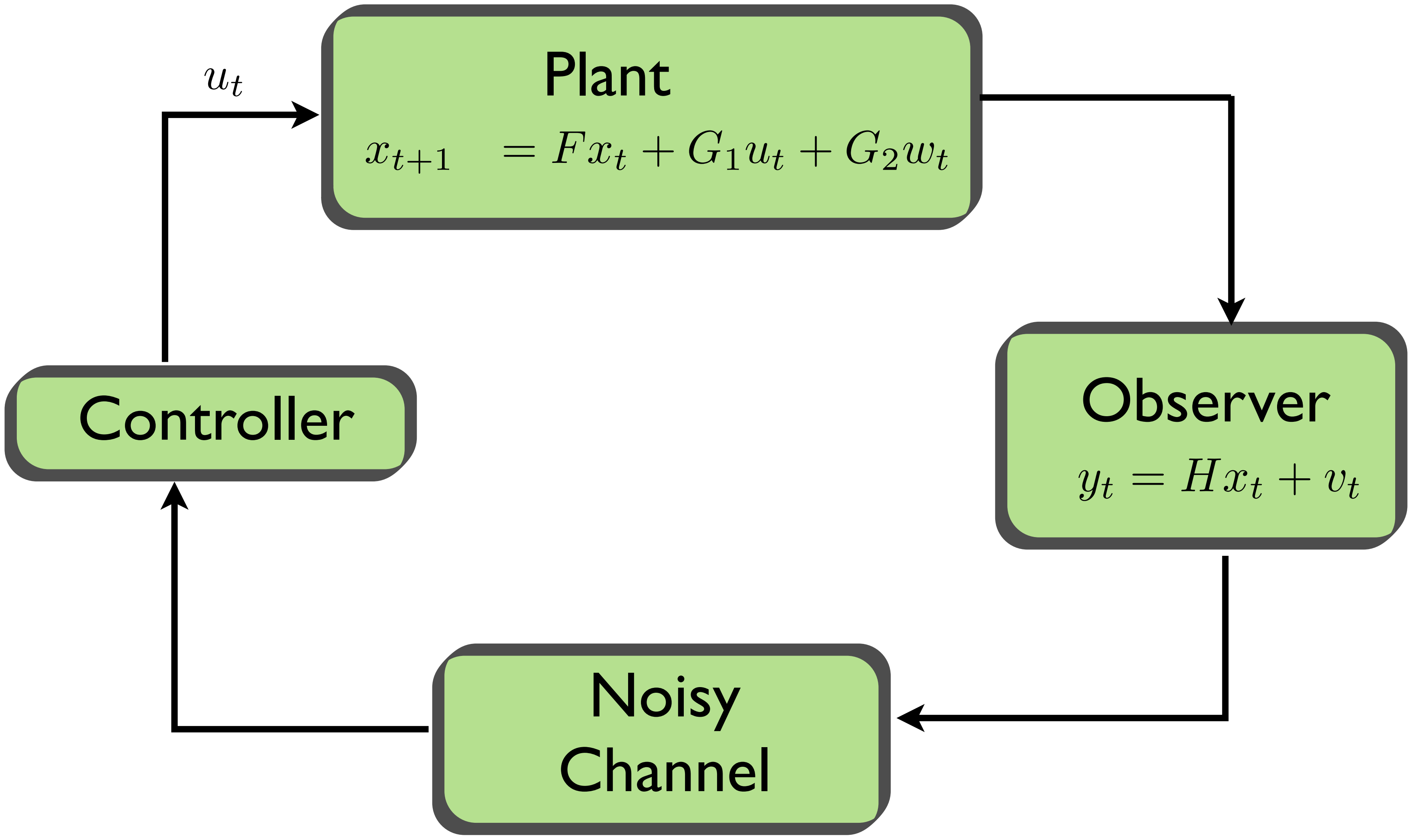
Tree Code: $\|\mathbf{C} - \mathbf{C}'\|_{\mathcal{H}} \propto d$

semi-infinite m-ary tree with each edge labeled by a symbol in an alphabet of size q > m

Tree Codes

- maps a sequence $\{b_i\}_{i>0}$ to a sequence $\{c_i\}_{i>0}$, where $b_i \in \{0,1,\ldots,m-1\}$ and $c_i \in \{0,1,\ldots,q-1\}$
- represents a *causal* code with rate $R = \log(m)/\log(q)$. In the example, $c_i = f_i(b_{\leq i})$.
- for every pair of paths with a common ancestor and length d, say, we require that the "Hamming distance" between the paths be at least a fixed proportion of d
- Schulman proved the existence of tree codes
- *along with ML decoding, allows reliable interactive communication over a noisy channel.*
- ***Problem:*** No explicit constructions, no tractable decoding and existence result is not with high probability

Control over Noisy Channels



- Automatic control over a noisy channel provides an interactive context in which the traditional notion of Shannon Capacity is no longer the right figure of merit [Sahai 2001].
- What sort of communication reliability do we need to guarantee closed loop stability?

Anytime Reliability - An Example

Track the following process over a noisy channel

$$x_{t+1} = \lambda x_t + w_t, \quad x_0 = 0, \quad |\lambda| > 1, \quad w_t = \pm 1$$

Encoder - Encode $\{w_t\}$ causally, Decoder - Generate estimates $\{\hat{w}_{\tau|t}\}_{\tau \leq t}$

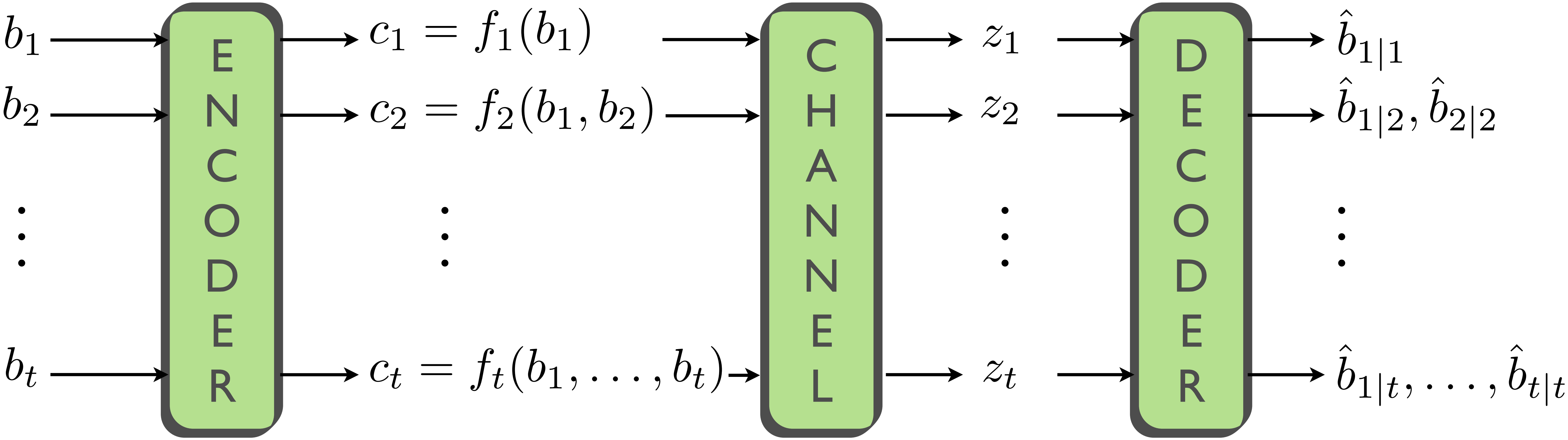
$$P^e_{t,d} \triangleq \left\{ \begin{array}{c} \hat{w}_{t-d+1} \neq w_{t-d+1}, \ \ \hat{w}_{\tau|t} = w_\tau, \ \ \tau \leq t - d \\ \implies |x_{t+1} - \hat{x}_{t+1|t}|^2 \leq K|\lambda|^{2d} \end{array} \right.$$

$P^e_{t,d}$ - Earliest error happens at time $t - d + 1$

$$\mathbb{E}|x_{t+1} - \hat{x}_{t+1|t}|^2 \leq K \sum_{d \geq 0} P^e_{t,d} |\lambda|^{2d}$$

Anytime Reliability

$$P^e_{t,d} \leq |\lambda|^{-(2+\epsilon)d}, \ \ \forall \ t, d \geq d_o$$



$$b_i \in \{0,1\}^{nR}, \;\; c_i \in \{0,1\}^{n}$$

An encoder, decoder pair is said to be (R, β, d_o) anytime reliable if

$$P\left(\hat{b}_{t-d+1|t} \neq b_{t-d+1}\right) \leq K 2^{-n\beta d}, \;\; \forall \;\; t, d \geq d_o$$

Let $\{\mu_i\}$ be the eigen values of F, then

$$R > \sum_{i:|\mu_i|>1} \log|\mu_i|, \;\; \beta > \eta \log \max_i |\mu_i| \implies \limsup_{t\to\infty} \mathbb{E}|x_t|^\eta < \infty$$

Tree codes under maximum likelihood decoding are anytime reliable

Linear Tree Codes Exist

$$\begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \vdots \\ \mathbf{c}_t \end{bmatrix} = \begin{bmatrix} G_{11} & 0 & \dots & \dots \\ G_{21} & G_{22} & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots \\ G_{t1} & G_{t,t-1} & \dots & G_{tt} \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_t \end{bmatrix}, \quad G_{ij} \in \{0,1\}^{n \times k}$$

or equivalently,

$$\begin{bmatrix} H_{11} & 0 & \dots & \dots \\ H_{21} & H_{22} & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots \\ H_{t1} & H_{t,t-1} & \dots & H_{tt} \end{bmatrix} \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \vdots \\ \mathbf{c}_t \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \quad H_{ij} \in \{0,1\}^{(n-k) \times n}$$

Without additional structure, existence is not with high probability (because one needs to union bound both over decoding instants and delay).

The Toeplitz Ensemble

$$\begin{bmatrix} G_1 & 0 & \dots & \dots \\ G_2 & G_1 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots \\ G_t & G_{t-1} & \dots & G_1 \end{bmatrix}$$

OR

$$\begin{bmatrix} H_1 & 0 & \dots & \dots \\ H_2 & H_1 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots \\ H_t & H_{t-1} & \dots & H_1 \end{bmatrix}$$

- the entries of $\{G_i\}$ (or $\{H_i\}$) are chosen i.i.d Bernoulli(p)
- this is a time-invariant infinite constraint length convolutional code
- a code drawn from this ensemble is anytime reliable with a high probability
- because of time-invariance, only a union bound over delay is required which gives a high probability result

Toeplitz Ensemble, $\mathbb{TZ}_p$: draw an infinite sequence $\{G_1, G_2, \ldots, G_t, \ldots\}$ such that each entry of G_i is Bernoulli(p)

Bhattacharya Parameter, $\zeta = \sum_{z \in \mathcal{Z}} \sqrt{p(z|0)p(z|1)}$

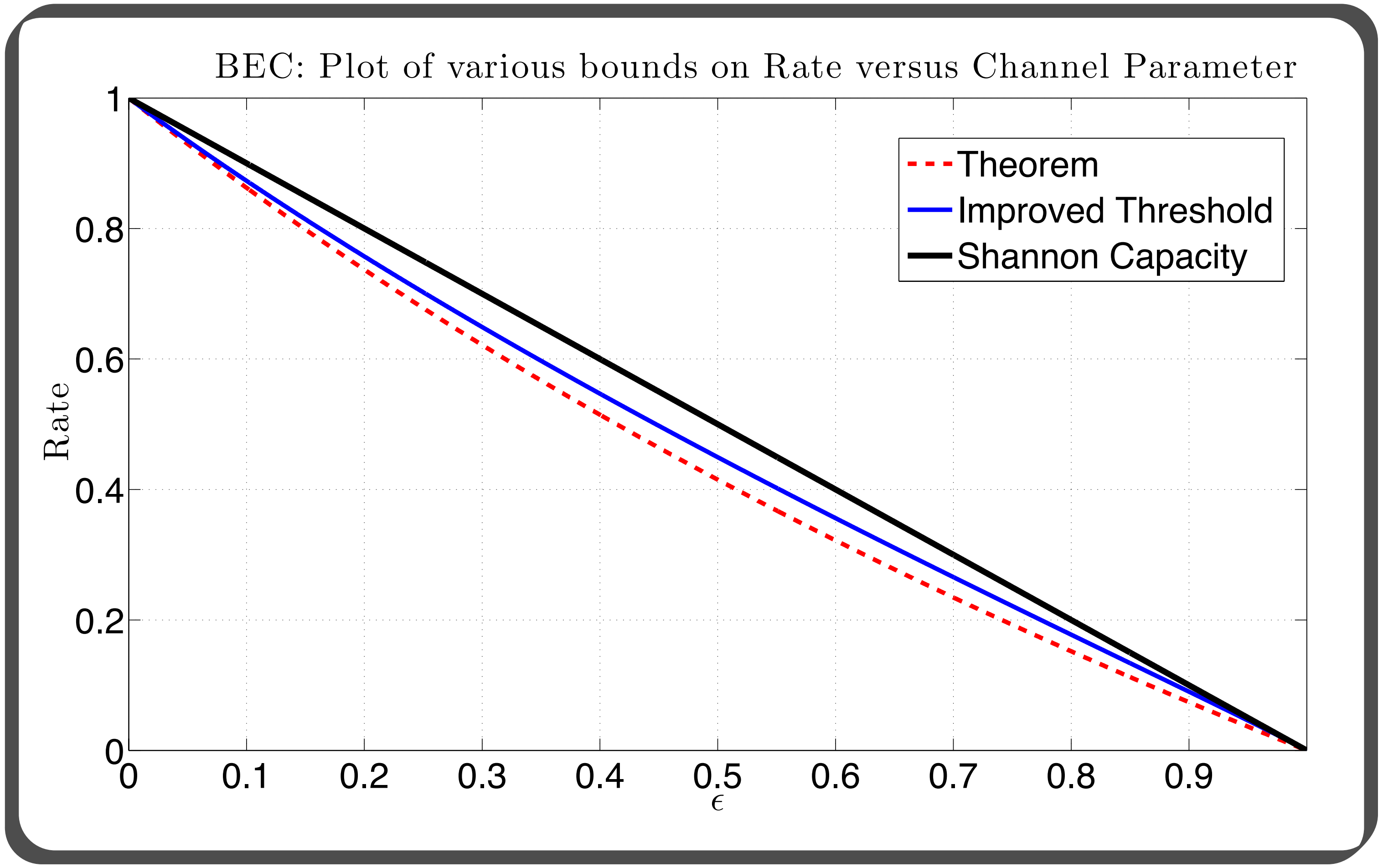
Theorem: For each rate, $R > 0$ and exponent, $\beta > 0$ such that

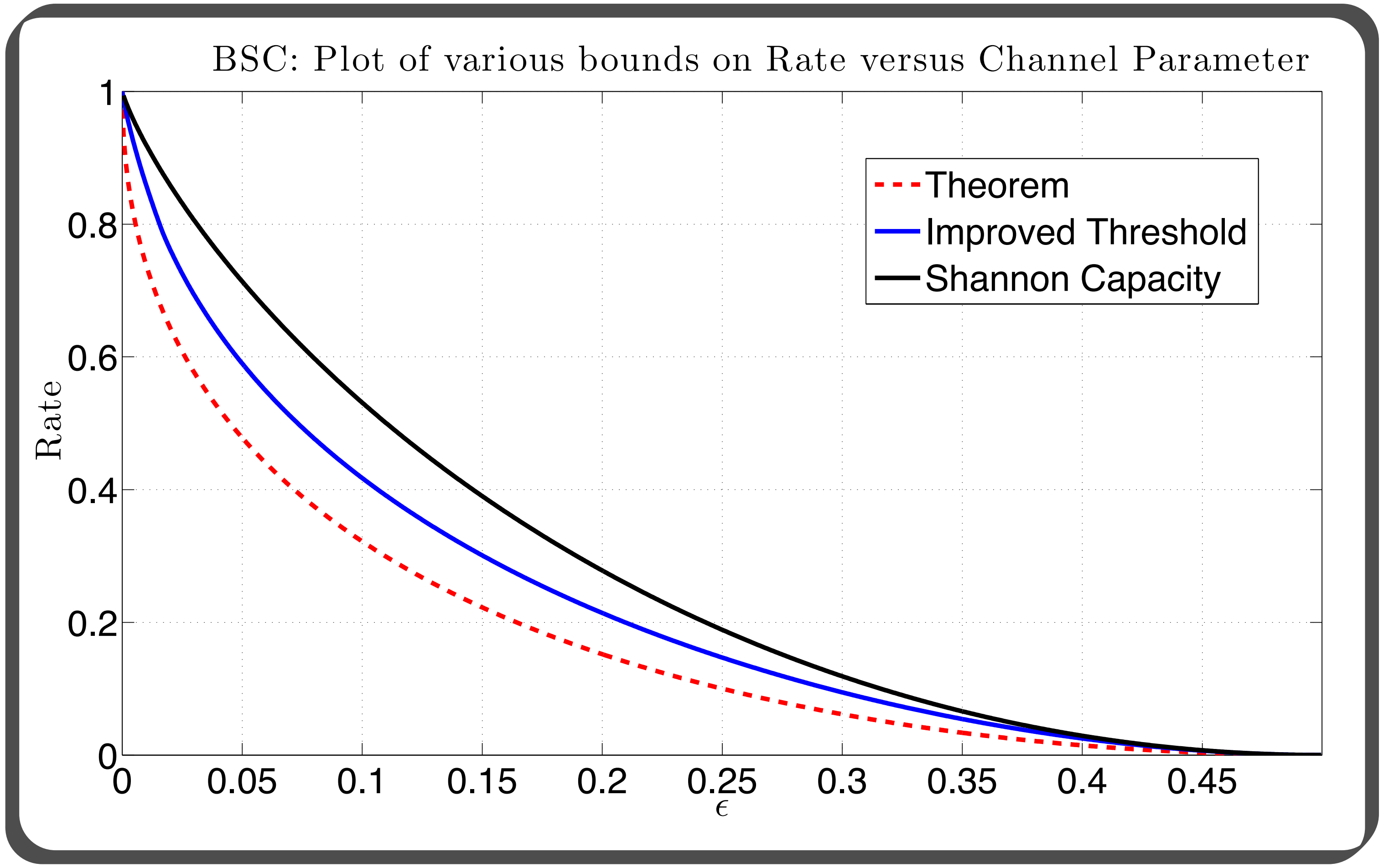
$$R < 1 - \log_2(1+\zeta), \;\; \beta < H^{-1}(1-R)\left[\log_2\left(\frac{1}{\zeta}\right) + \log_2\left(2^{1-R} - 1\right)\right]$$

the probability that a randomly chosen code from $\mathbb{TZ}_{\frac{1}{2}}$ is (R, β, d_o) - anytime reliable is at least $1 - 2^{-\Omega(nd_o)}$

BEC(ϵ): $\zeta = \epsilon$, BSC(ϵ): $\zeta = 2\sqrt{\epsilon(1-\epsilon)}$

e.g., for BEC(ϵ), the Theorem guarantees anytime reliability for rates up to 1-$\log_2$(1+ ϵ). This is the so called *computational cut-off rate* and is a consequence of using a *union bound* in the proof. These thresholds can be *significantly improved* using tighter techniques.


BEC: Plot of various bounds on Rate versus Channel Parameter


BSC: Plot of various bounds on Rate versus Channel Parameter

Decoder for BEC

- Divide the transmitted codeword c into observed entries c_o and erased entries c_e
- Then the parity check condition gives

$$Hc = \left[\begin{array}{cc} H_e & H_o \end{array} \right] \left[\begin{array}{c} c_e \\ c_o \end{array} \right] = 0 \implies H_e c_e = H_o c_o \equiv s$$

- So, maximum likelihood (ML) decoding over the BEC is just matrix inversion. If R < 1- ϵ, the above equation has a unique solution with high probability

The Triangular Case...

- H_e and H_o are lower triangular. So, even though H_e is a tall matrix, it will *most likely* not have a full column rank.
- But can we recover the erased bits with a delay d?

Efficient Decoder for BEC

$$H_e c_e = s \equiv \begin{bmatrix} H_{e,11} & 0 \\ H_{e,21} & H_{e,22} \end{bmatrix} \begin{bmatrix} c_{e,1} \\ c_{e,2} \end{bmatrix} = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$

1. For $d' = 1, 2, \ldots, d$, partition $c_e = \begin{bmatrix} c_{e,1}^T & c_{e,2}^T \end{bmatrix}^T$ and $H_e = \begin{bmatrix} H_{e,11} & 0 \\ H_{e,21} & H_{e,22} \end{bmatrix}$, where $c_{e,1}$ and $c_{e,2}$ denote the erased entries in the intervals $[t-d+1, t-d']$ and $[t-d'+1, t]$ respectively.
2. Check whether $\begin{bmatrix} H_{e,11} \\ H_{e,22}^{\perp} H_{e,21} \end{bmatrix}$ has full column rank.
3. If so, solve for $c_{e,1}$ from $\begin{bmatrix} H_{e,11} \\ H_{e,22}^{\perp} H_{e,21} \end{bmatrix} c_{e,1} = \begin{bmatrix} s_1 \\ H_{e,22}^{\perp} s_2 \end{bmatrix}$
4. Increment $t = t + 1$ and continue

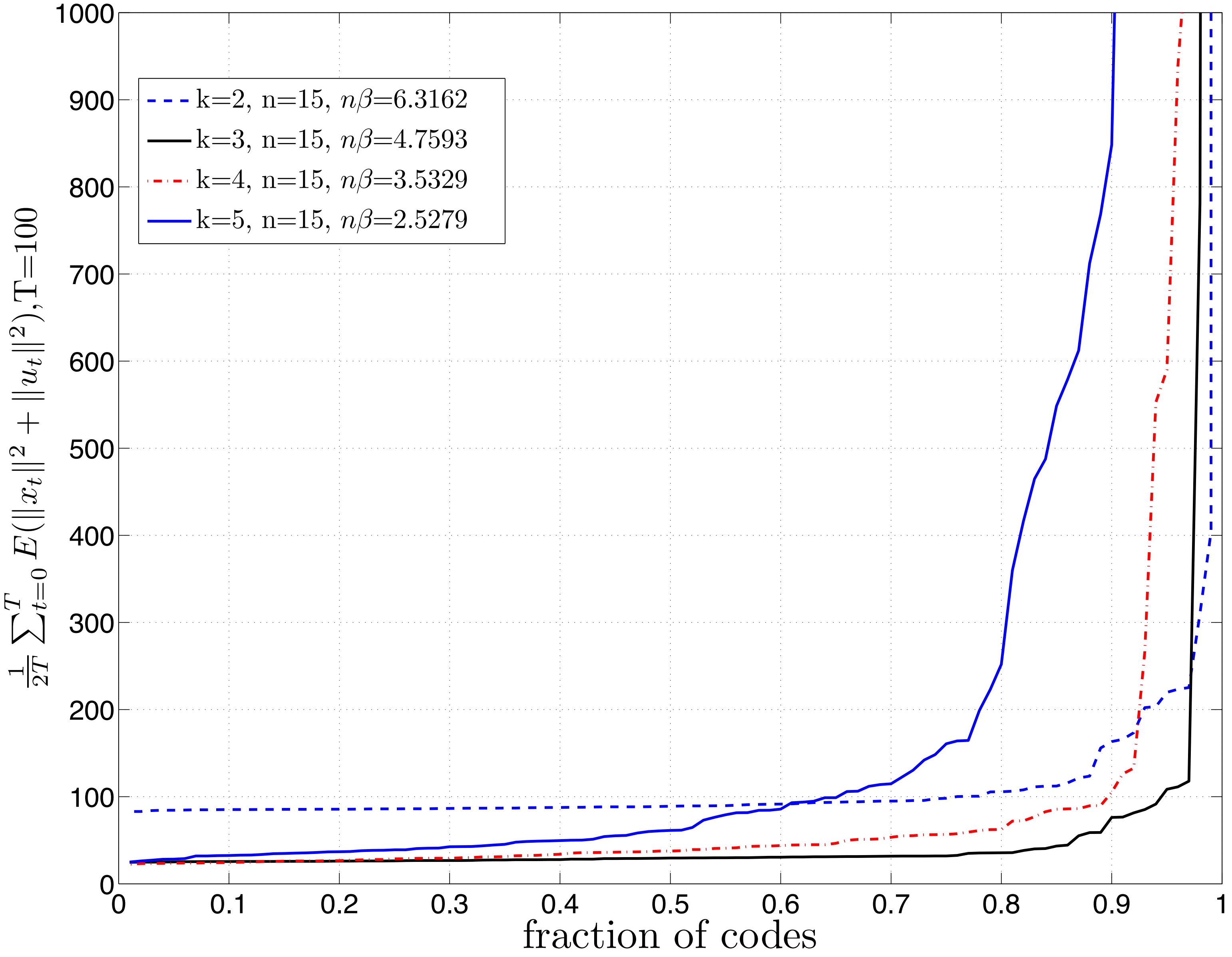
Decoding complexity

- If the position of the earliest uncorrected erasure is at time t-d, then the algorithm takes $O(d^3)$ time.
- But the probability that there is an uncorrected erasure at time t-d is at most $K2^{-n\beta d}$
- So, the expected complexity per time instant is constant:

$$\sum_{d\geq 1} Kd^3 2^{-n\beta d}$$

- Furthermore, the probability that the complexity at any given time instant exceeds $O(d^3)$ decays as $O(2^{-n\beta d})$
- ***Remark:*** With feedback, encoding can also be done with constant expected complexity

A Simulation



$$F = \begin{bmatrix} 2 & 1 & 0 \\ 0.25 & 0 & 1 \\ -0.5 & 0 & 0 \end{bmatrix}$$

$$\lambda(F) = \{2, -0.5, 0.5\}$$

$$H = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$$

$$G_1 = \mathcal{I}_3$$

$$u_t = -\hat{x}_{t+1|t}$$

Higher rates provide a finer resolution of the measurements while larger exponent ensure that the controller's estimate of the state does not drift away; however, we cannot have both.

Future Directions

1. Construct efficiently decodable tree codes for other classes of channels, e.g., BSC, AWGNC.
2. Study the tradeoff between rate and reliability to optimize system performance (e.g., LQR cost).

References

1. LJ Schulman, "Coding for Interactive Communication", IEEE Transactions on Information Theory, vol. 42, no. 6, 1996.
2. A Sahai and S Mitter, "The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link - part i: Scalar systems", IEEE Transactions on Information Theory, vol. 52, no. 8, 2006.
3. RT Sukhavasi and B Hassibi, "Linear Error Correcting Codes With Anytime Reliability", IEEE International Symposium on Information Theory, 2011.
4. RT Sukhavasi and B Hassibi, "Anytime Reliable Codes for Stabilizing Plants over Erasure Channels", IEEE Conference on Decision and Control, 2011