(Almost) Practical Tree Codes

Anatoly Khina

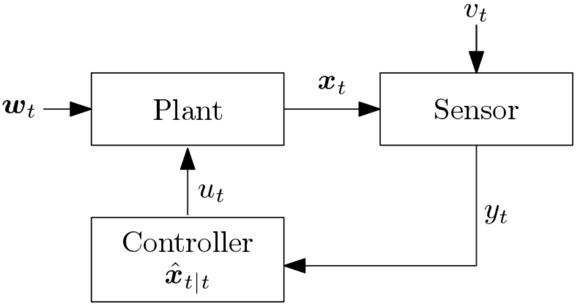
Joint work with Wael Halbawi and Babak Hassibi

Caltech, Pasadena, CA, USA

ITA 2016
San Diego
February 2, 2016

Networked Control vs. Traditional Control

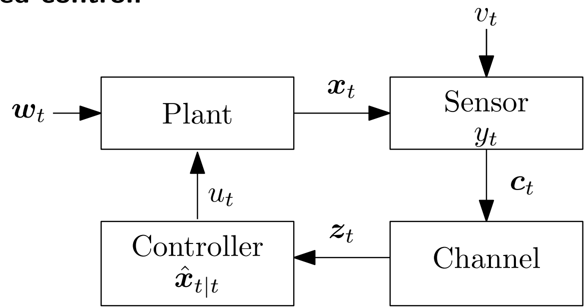
Traditional control:



- Observer and controller are co-located.
- Classical systems are hardwired and well crafted

Networked Control vs. Traditional Control

Networked control:

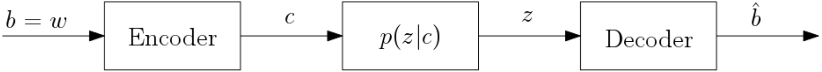


- Observer and controller are not co-located: connected through noisy link
- Suitable for new remote applications (e.g., remote surgery, self-driving cars)

Motivating Example: Tracking a Random Walk [Sahai PhD'01]

$$x_{t+1} = \alpha x_t + w_t$$

- $\alpha > 1 \implies$ not stable!
- $w_t \in \{\pm 1\}$
- We wish to track x_t with bounded expected distortion
- If tracking is possible, stability usually follows
- Allows to distill the coding problem (no quantization)



Distortion requirement

$$\mathbb{E}\left[(x_t - \hat{x}_t)^2\right] < \infty, \qquad \forall t$$

Motivating Example: Tracking a Random Walk [Sahai PhD'01]

- $\hat{b}_{t-d|t}$ – Estimate of b_{t-d} at time t
- Probability of first error event at time $t-d$:
  $P_e(t,d) \triangleq \Pr\left(b_{t-d} \neq \hat{b}_{t-d|t}, \forall \delta > d, b_{t-\delta} = \hat{b}_{t-\delta|t}\right)$

$$\mathbb{E}\left[\left(x_t - \hat{x}_{t|t}\right)^2\right] \propto \sum_{d=1}^{t} P_e(t,d)\alpha^{2d} = \sum_{d=1}^{t} P_e(t,d) 2^{2\log\alpha\cdot d} < \infty$$

Error probability profile: Anytime-reliable code

$$P_e(t,d) < 2^{-(2\log\alpha+\epsilon)d}, \qquad \forall t, d_0 < d < t$$

Larger moments

Higher exponent $\Longrightarrow$ Cannot stabilize all moments!

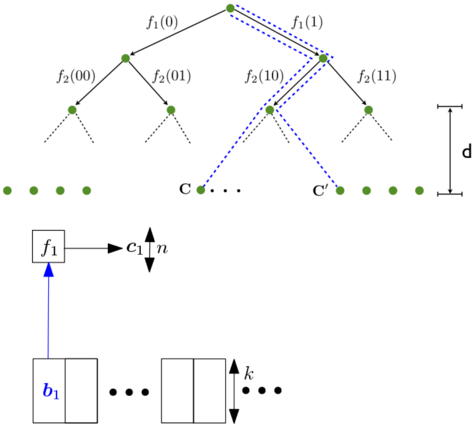
Anytime-Reliable Codes: Basics

Error probability profile

$$P_e(t,d) < 2^{-(2\log\alpha+\epsilon)d}\,, \qquad \forall t, d_0 < d < t$$

How to generate such a code?

$$\begin{aligned} \boldsymbol{c}_1 &= f_1(\boldsymbol{b}_1) \\ \boldsymbol{c}_2 &= f_2(\boldsymbol{b}_1, \boldsymbol{b}_2) \\ &\vdots \\ \boldsymbol{c}_t &= f_t(\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_t) \\ &\vdots \end{aligned}$$



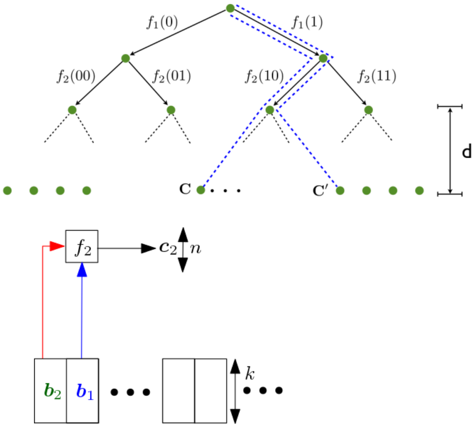
Anytime-Reliable Codes: Basics

Error probability profile

$$P_e(t,d) < 2^{-(2\log\alpha+\epsilon)d}\,, \qquad \forall t, d_0 < d < t$$

How to generate such a code?

$$\mathbf{c}_1 = f_1(\mathbf{b}_1)$$
$$\mathbf{c}_2 = f_2(\mathbf{b}_1, \mathbf{b}_2)$$
$$\vdots$$
$$\mathbf{c}_t = f_t(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_t)$$
$$\vdots$$



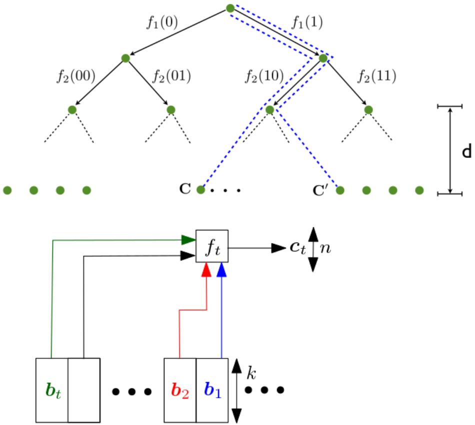
Anytime-Reliable Codes: Basics

Error probability profile

$$P_e(t,d) < 2^{-(2\log\alpha+\epsilon)d}\,, \qquad \forall t, d_0 < d < t$$

How to generate such a code?

$$\begin{aligned} \boldsymbol{c}_1 &= f_1(\boldsymbol{b}_1) \\ \boldsymbol{c}_2 &= f_2(\boldsymbol{b}_1, \boldsymbol{b}_2) \\ &\vdots \\ \boldsymbol{c}_t &= f_t(\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_t) \\ &\vdots \end{aligned}$$



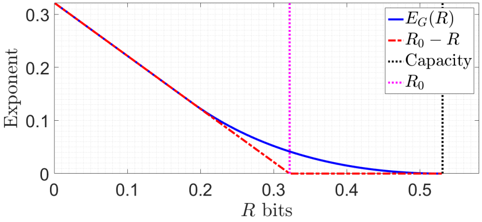
Anytime-Reliable Codes as Convolutional Codes

Random time-varying convolutional-code ensemble [Viterbi, Yudkin, Zigangirov, Shulman–Feder, ...]

- Most results assume infinite stream ($\gg$ delay-line length)
- We wish to recover a bit using subsequent *nd* output symbols
- The random time-varying CC ensemble achieves:

$$\mathbb{E}[P_e(t,d)] \leq 2^{-E_G(R)nd}$$

- $E_G(R) > 0$ for $R < C$ – Gallager's error exponent



Anytime-Reliable Codes as Convolutional Codes

Good ensemble performance $\Rightarrow$ Good specific code performance?

- Block codes: Yes, with high probability!
- Anytime reliable-code?
- Such a code exists [Schulman IT'96], but **not w.h.p.** ☹
  (Proof requires min-distance $\propto$ delay)
- **How to construct a good anytime-reliable code?**

Ensemble Performance ⇒ Specific Code Performance?

Ensemble performance

$$\mathbb{E}[P_e(t,d)] \le 2^{-E_G(R)nd}$$

Specific d and t

Using Markov's inequality:

$$\Pr\left(P_e(t,d) \ge 2^{-[E_G(R)-\epsilon]nd}\right) \le \frac{\mathbb{E}[P_e(t,d)]}{2^{-[E_G(R)-\epsilon]nd}} = 2^{-\epsilon nd}$$

Ensemble Performance $\Rightarrow$ Specific Code Performance?

Ensemble performance

$$\mathbb{E}[P_e(t,d)] \le 2^{-E_G(R)nd}$$

Specific d and t

Using Markov's inequality:

$$\Pr\left(P_e(t,d) \ge 2^{-[E_G(R)-\epsilon]nd}\right) \le \frac{\mathbb{E}[P_e(t,d)]}{2^{-[E_G(R)-\epsilon]nd}} = 2^{-\epsilon nd}$$

All t and $d_0 < d \le t$

Using the union bound:

$$\Pr\left(\bigcup_{t=1}^{\infty}\bigcup_{d=d_0}^{t} P_e(t,d) \ge 2^{-[E_G(R)-\epsilon]nd}\right)$$

$$\le \sum_{t=1}^{\infty}\sum_{d=d_0}^{t} \underbrace{\Pr\left(P_e(t,d) \ge 2^{-[E_G(R)-\epsilon]nd}\right)}_{\le 2^{-\epsilon nd}} \le \sum_{t=1}^{\infty} \text{const} \to \infty$$

Linear Time-Invariant Codes [Sukhavasi–Hassibi ISIT'11]

Linear time-variant code

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{1,1} & \mathbf{0} & \cdots & \cdots & \cdots \\ \mathbf{G}_{2,1} & \mathbf{G}_{2,2} & \mathbf{0} & \cdots & \cdots \\ \vdots & \vdots & \ddots & \ddots & \cdots \\ \mathbf{G}_{t,1} & \mathbf{G}_{t,2} & \cdots & \mathbf{G}_{t,t} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Linear Time-Invariant Codes [Sukhavasi–Hassibi ISIT'11]

Linear time-variant code

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{1,1} & \mathbf{0} & \cdots & \cdots & \cdots \\ \mathbf{G}_{2,1} & \mathbf{G}_{2,2} & \mathbf{0} & \cdots & \cdots \\ \vdots & \vdots & \ddots & \ddots & \cdots \\ \mathbf{G}_{t,1} & \mathbf{G}_{t,2} & \cdots & \mathbf{G}_{t,t} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Linear time-invariant code

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{1} & \mathbf{0} & \cdots & \cdots & \cdots \\ \mathbf{G}_{2} & \mathbf{G}_{1} & \mathbf{0} & \cdots & \cdots \\ \vdots & \vdots & \ddots & \ddots & \cdots \\ \mathbf{G}_{t} & \mathbf{G}_{t-1} & \cdots & \mathbf{G}_{1} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Linear Time-Invariant Codes [Sukhavasi–Hassibi ISIT'11]

Linear time-variant code

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{1,1} & \mathbf{0} & \cdots & \cdots & \cdots \\ \mathbf{G}_{2,1} & \mathbf{G}_{2,2} & \mathbf{0} & \cdots & \cdots \\ \vdots & \vdots & \ddots & \ddots & \cdots \\ \mathbf{G}_{t,1} & \mathbf{G}_{t,2} & \cdots & \mathbf{G}_{t,t} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Linear time-invariant code

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{1} & \mathbf{0} & \cdots & \cdots & \cdots \\ \mathbf{G}_{2} & \mathbf{G}_{1} & \mathbf{0} & \cdots & \cdots \\ \vdots & \vdots & \ddots & \ddots & \cdots \\ \mathbf{G}_{t} & \mathbf{G}_{t-1} & \cdots & \mathbf{G}_{1} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Linear Time-Invariant Codes [Sukhavasi–Hassibi ISIT'11]

Linear time-variant code

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{1,1} & \mathbf{0} & \cdots & \cdots & \cdots \\ \mathbf{G}_{2,1} & \mathbf{G}_{2,2} & \mathbf{0} & \cdots & \cdots \\ \vdots & \vdots & \ddots & \ddots & \cdots \\ \mathbf{G}_{t,1} & \mathbf{G}_{t,2} & \cdots & \mathbf{G}_{t,t} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Linear time-invariant code

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{1} & \mathbf{0} & \cdots & \cdots & \cdots \\ \mathbf{G}_{2} & \mathbf{G}_{1} & \mathbf{0} & \cdots & \cdots \\ \vdots & \vdots & \ddots & \ddots & \cdots \\ \mathbf{G}_{t} & \mathbf{G}_{t-1} & \cdots & \mathbf{G}_{1} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

All t and $d_0 < d \leq t$

Using the union bound:

$$\Pr\left(\bigcup_{t=1}^{\infty}\bigcup_{d=d_0}^{t} P_e(t,d) \geq 2^{-[E_G(R)-\epsilon]nd}\right)$$

$$\leq \sum_{t=1}^{\infty}\sum_{d=d_0}^{t} \underbrace{\Pr\left(P_e(t,d) \geq 2^{-[E_G(R)-\epsilon]nd}\right)}_{\leq 2^{-\epsilon nd}} \leq \sum_{t=1}^{\infty} \text{const} \to \infty$$

Linear Time-Invariant Codes [Sukhavasi–Hassibi ISIT'11]

Linear time-variant code

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{1,1} & \mathbf{0} & \cdots & \cdots & \cdots \\ \mathbf{G}_{2,1} & \mathbf{G}_{2,2} & \mathbf{0} & \cdots & \cdots \\ \vdots & \vdots & \ddots & \ddots & \cdots \\ \mathbf{G}_{t,1} & \mathbf{G}_{t,2} & \cdots & \mathbf{G}_{t,t} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Linear time-invariant code

$$\mathbf{G} = \left[\begin{array}{c|cccc} \mathbf{G}_{1} & \mathbf{0} & \cdots & \cdots & \cdots \\ \hline \mathbf{G}_{2} & \mathbf{G}_{1} & \mathbf{0} & \cdots & \cdots \\ \vdots & \vdots & \ddots & \ddots & \cdots \\ \mathbf{G}_{t} & \mathbf{G}_{t-1} & \cdots & \mathbf{G}_{1} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{array}\right]$$

All t and $d_0 < d \leq t$

Using the union bound:

$$\Pr\left(\bcancel{\bigcup_{t=1}^{\infty}} \bigcup_{d=d_0}^{t \to \infty} P_e(\cancel{t}, d) \geq 2^{-[E_G(R)-\epsilon]nd} \right)$$

$$\leq \bcancel{\sum_{t=1}^{\infty}} \sum_{d=d_0}^{t \to \infty} \underbrace{\Pr\left(P_e(\cancel{t}, d) \geq 2^{-[E_G(R)-\epsilon]nd}\right)}_{\leq 2^{-\epsilon nd}} \leq \bcancel{\sum_{t=1}^{\infty}} \text{const} \to \cancel{\infty}$$

Linear Time-Invariant Codes [Sukhavasi–Hassibi ISIT'11]

Linear time-variant code

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{1,1} & \mathbf{0} & \cdots & \cdots & \cdots \\ \mathbf{G}_{2,1} & \mathbf{G}_{2,2} & \mathbf{0} & \cdots & \cdots \\ \vdots & \vdots & \ddots & \ddots & \cdots \\ \mathbf{G}_{t,1} & \mathbf{G}_{t,2} & \cdots & \mathbf{G}_{t,t} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Linear time-invariant code

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{1} & \mathbf{0} & \cdots & \cdots & \cdots \\ \mathbf{G}_{2} & \mathbf{G}_{1} & \mathbf{0} & \cdots & \cdots \\ \vdots & \vdots & \ddots & \ddots & \cdots \\ \mathbf{G}_{t} & \mathbf{G}_{t-1} & \cdots & \mathbf{G}_{1} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

All t and $d_0 < d \leq t$

Using the union bound:

$$\Pr\left(\bigcup_{d=d_0}^{\infty} P_e(d) \geq 2^{-[E_G(R)-\epsilon]nd}\right) \leq \sum_{d=d_0}^{\infty} 2^{-\epsilon nd}$$

$$= \frac{2^{-\epsilon n d_0}}{1 - 2^{-\epsilon n}}$$

Linear Time-Invariant Codes

- Random LTI convolutional codes are anytime-reliable w.h.p. ✓
- But the exponent result was valid for time-variant codes
- Valid also for LTI codes [S~~c~~hulman–Feder IT'00] ✓
  - (Proved independently in [Sukhavasi–Hassibi ISIT'11])
- **No gain** for general codes over LTI codes in this regime!
  - (Common setting of infinite decoding window: huge gain!)

Linear Time-Invariant Codes

- Random LTI convolutional codes are anytime-reliable w.h.p. ✓
- But the exponent result was valid for time-variant codes
- Valid also for LTI codes [Schulman–Feder IT'00] ✓
  - (Proved independently in [Sukhavasi–Hassibi ISIT'11])
- **No gain** for general codes over LTI codes in this regime!
  - (Common setting of infinite decoding window: huge gain!)

What about decoding?

Decoding of Linear Time-Invariant Codes

- All results assumed maximum-likelihood (ML) decoding
- ML complexity rises exponentially with t

Binary Erasure Channel (BEC)

- For LTI codes: ML $=$ Solving linear equations

- What about other channels?

Sequential Decoding

- Before Viterbi algo.: Sequential decoding *de facto* standard
- Sequential decoding = class of algorithms
- Introduced originally in [Wozencraft '57] for **tree codes**
- Common to all: Explore only subset of (likely) codewords
- Most prominent variants: Stack and Fano's algorithms

Sequential Decoding

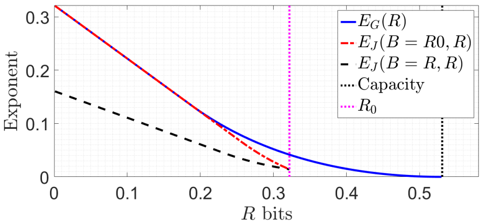
- Fano's metric: $M(\boldsymbol{c}_1, \cdots, \boldsymbol{c}_t) = \sum_{i=1}^{nt} \left[\log \frac{p(z_t|c_t)}{p(z_t)} - \overbrace{B}^{\text{bias}} \right]$
- For ML decoding: $\arg\max_{\{\boldsymbol{c}_i\}} p(z_t|c_t) = \arg\max_{\{\boldsymbol{c}_i\}} \left[\log \frac{p(z_t|c_t)}{p(z_t)} - B \right]$
- For partial tree exploration: Fano's metric penalizes longer incorrect paths via bias B

Sequential Decoding: Error Probability

Error probability of **general** conv. ensemble [Jelinek's Book '68]

$$\mathbb{E}[P_e(t,d)] \le A\,2^{-E_J(B,R)nd}$$

- A is finite for $B < R_0$
- $E_J(B,R)$ properties:
  - $\frac{1}{2}E_G(R) \le E_J(B,R) < E_G(R)$
  - $E_J(B,R) \xrightarrow{B \to R_0} E_G(R)$, for $R < R_{\text{crit}}$
- Does not guarantee a good **specific code** w.h.p.



Sequential Decoding: Error Probability

Proof for general codes requires:

- **Pairwise independence:** Every two paths are independent (from divergence point)
- **Individual codeword distribution:** Entries within each codeword are i.i.d.

Sequential Decoding: Error Probability

Use the following affine time-invariant ensemble:

$$\begin{bmatrix} \boldsymbol{c}_1 \\ \boldsymbol{c}_2 \\ \vdots \\ \boldsymbol{c}_t \\ \vdots \end{bmatrix} = \begin{bmatrix} \mathbf{G}_1 & \mathbf{0} & \cdots & \cdots & \cdots \\ \mathbf{G}_2 & \mathbf{G}_1 & \mathbf{0} & \cdots & \cdots \\ \vdots & \vdots & \ddots & \ddots & \cdots \\ \mathbf{G}_t & \mathbf{G}_{t-1} & \cdots & \mathbf{G}_1 & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \\ \vdots \\ \boldsymbol{b}_t \\ \vdots \end{bmatrix} + \begin{bmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \\ \vdots \\ \boldsymbol{v}_t \\ \vdots \end{bmatrix}$$

- Entries of $\{\mathbf{G}_t\}$, $\{\boldsymbol{b}_t\}$ and $\{\boldsymbol{v}_t\}$ are i.i.d. uniform
- $\{\boldsymbol{v}_t\}$ random translation vectors

Sequential Decoding: Error Probability

Use the following affine time-invariant ensemble:

$$\begin{bmatrix} \boldsymbol{c}_1 \\ \boldsymbol{c}_2 \\ \vdots \\ \boldsymbol{c}_t \\ \vdots \end{bmatrix} = \begin{bmatrix} \mathbf{G}_1 & \mathbf{0} & \cdots & \cdots & \cdots \\ \mathbf{G}_2 & \mathbf{G}_1 & \mathbf{0} & \cdots & \cdots \\ \vdots & \vdots & \ddots & \ddots & \cdots \\ \mathbf{G}_t & \mathbf{G}_{t-1} & \cdots & \mathbf{G}_1 & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \\ \vdots \\ \boldsymbol{b}_t \\ \vdots \end{bmatrix} + \begin{bmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \\ \vdots \\ \boldsymbol{v}_t \\ \vdots \end{bmatrix}$$

- Entries of $\{\mathbf{G}_t\}$, $\{\boldsymbol{b}_t\}$ and $\{\boldsymbol{v}_t\}$ are i.i.d. uniform
- $\{\boldsymbol{v}_t\}$ random translation vectors
- i.i.d. uniformity of $\{\mathbf{G}_t\}$ guarantees pairwise independence ✓

Sequential Decoding: Error Probability

Use the following affine time-invariant ensemble:

$$\begin{bmatrix} \boldsymbol{c}_1 \\ \boldsymbol{c}_2 \\ \vdots \\ \boldsymbol{c}_t \\ \vdots \end{bmatrix} = \begin{bmatrix} \mathbf{G}_1 & \mathbf{0} & \cdots & \cdots & \cdots \\ \mathbf{G}_2 & \mathbf{G}_1 & \mathbf{0} & \cdots & \cdots \\ \vdots & \vdots & \ddots & \ddots & \cdots \\ \mathbf{G}_t & \mathbf{G}_{t-1} & \cdots & \mathbf{G}_1 & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \\ \vdots \\ \boldsymbol{b}_t \\ \vdots \end{bmatrix} + \begin{bmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \\ \vdots \\ \boldsymbol{v}_t \\ \vdots \end{bmatrix}$$

- Entries of $\{\mathbf{G}_t\}$, $\{\boldsymbol{b}_t\}$ and $\{\boldsymbol{v}_t\}$ are i.i.d. uniform
- $\{\boldsymbol{v}_t\}$ random translation vectors
- i.i.d. uniformity of $\{\mathbf{G}_t\}$ guarantees pairwise independence ✓
- i.i.d. uniformity of $\{\boldsymbol{v}_t\}$ guarantees pairwise independence ✓

Sequential Decoding: Complexity

- W_t – Number of branch computations of note t
- W_t is a random variable

Cutoff rate [Arıkan IT'88]

For any "good" code (general or LTI), $\mathbb{E}[W_t]$ is unbounded for $R > R_0$.

Pareto distribution of W_t [Gallager, Zigangirov, Viterbi–Omura, ...]

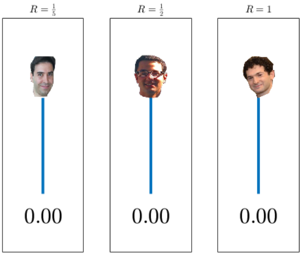
$$\Pr\left(W_t \geq m\right) \leq Am^{-\rho}$$

- For $B, R < R_0$ and $R < \frac{B+\mathrm{R}_0}{2\rho}$, $\rho \in (0, 1]$:
  Tight for **general** and **LTI** codes $\Rightarrow \mathbb{E}[W_t] < \infty$ for $R < R_0$
- For $\rho > 1$, $R = E_0(\rho)/\rho$:
  - Tight for **general** codes
  - Widely conjectured to be true for **LTI** codes
- Heavy tailed even if expectation is finite!

Simulation: Cart–Stick over BSC(0.01)

$$n = 20$$

$k =$	4	10	20
$R =$	$\frac{1}{5}$	$\frac{1}{2}$	1
$E =$	0.5382	0.2382	0



- Cart–stick system model [Franklin–Powell–Emami-Naeini Book]
- BSC(0.01)
- For this setting [Sukhavasi–Hassibi ISIT'11]: $k_{\min} = 3$, $E_{\min} = 0.2052$