Variable-length codes for channels with memory and feedback: fundamental limits and practical transmission schemes

Achilleas Anastasopoulos
anastas@umich.edu

EECS Department
University of Michigan

April 6, 2017
Joint work with Jui Wu
Problems in the intersection of Communications and Control

A) Viewing point-to-point communications as a Control problem

The act of transmitting a signal (partially) controls the overall communication system, with the hope of bringing it to a “desirable” state.
B) Viewing multi-agent communications as a Control problem

Multiple agents (partially) control a communication network to bring it to a state beneficial for all (cooperatively/competitively)
C) More subtle: Viewing off-line optimization problems relevant to Information theory as control problems, e.g., Shannon capacity

\[ C = \sup \{ P_{X_t|X_{t-1},Y_{t-1}(\cdot|\cdot,\cdot)} \}_{t} \frac{1}{T} \sum_{t=1}^{T} I(X_t \land Y_t|Y^{t-1}) \] 

No clear connection to Control: Where is the controller? where is the plant? what is the observation/control action?
Overview

1. Discrete memoryless channels (DMCs)
   - DMC without feedback
   - DMC with feedback and fixed-length (FL) codes
   - DMC with feedback and variable-length (VL) codes

2. Channels with memory and feedback
   - Known capacity results
   - Recent results for error exponents of VL codes
Overview

1. Discrete memoryless channels (DMCs)
   - DMC without feedback
   - DMC with feedback and fixed-length (FL) codes
   - DMC with feedback and variable-length (VL) codes

2. Channels with memory and feedback
   - Known capacity results
   - Recent results for error exponents of VL codes
Discrete memoryless channels without feedback

- Discrete memoryless channel (DMC) \((\mathcal{X}, \mathcal{Y}, Q)\) without feedback

\[ P(Y_t|X^t, Y^{t-1}) = Q(Y_t|X_t) \]
Discrete memoryless channels without feedback

- Discrete memoryless channel (DMC) \((\mathcal{X}, \mathcal{Y}, Q)\) without feedback
  
  \[
P(Y_t|X^t, Y^{t-1}) = Q(Y_t|X_t)
  \]

- Fixed-length (FL) code \(C\) with length \(n\) (channel uses) and size \(M = 2^k\) (messages)

  encoder: \(e: \{1, 2, \ldots, M\} \to \mathcal{X}^n\) with \(e(W) = X^n \overset{\Delta}{=} (X_1, \ldots, X_n)\)

  decoder: \(d: \mathcal{Y}^n \to \{1, 2, \ldots, M\}\) with \(d(Y^n) = \hat{W}\)
Discrete memoryless channels without feedback

- Discrete memoryless channel (DMC) \((\mathcal{X}, \mathcal{Y}, Q)\) without feedback
  \[ P(Y_t|X^t, Y^{t-1}) = Q(Y_t|X_t) \]

- Fixed-length (FL) code \(C\) with length \(n\) (channel uses) and size \(M = 2^k\) (messages)
  - encoder: \(e : \{1, 2, \ldots, M\} \rightarrow \mathcal{X}^n\) with \(e(W) = X^n \triangleq (X_1, \ldots, X_n)\)
  - decoder: \(d : \mathcal{Y}^n \rightarrow \{1, 2, \ldots, M\}\) with \(d(Y^n) = \hat{W}\)

- Rate \(R \triangleq \frac{\log M}{n} = \frac{k}{n}\) (info bits/channel use).
  Error probability \(P_e \triangleq P(W \neq \hat{W})\)
DMCs without feedback: basic results

- **Capacity [Shannon, 1948]:** The maximum transmission rate with arbitrarily low error probability is

\[
C \triangleq \max_{P_X} I(X; Y) = \max_{P_X} \sum_{x,y} Q(y|x)P_X(x) \log \frac{Q(y|x)}{P_Y(y)}
\]
Discrete memoryless channels (DMCs)

DMC without feedback

DMCs without feedback: basic results

- **Capacity [Shannon, 1948]**: The maximum transmission rate with arbitrarily low error probability is

\[ C \triangleq \max_{P_X} I(X; Y) = \max_{P_X} \sum_{x,y} Q(y|x) P_X(x) \log \frac{Q(y|x)}{P_Y(y)} \]

- **Error exponent [Fano, 1961, Gallager, 1965, Shannon et al., 1967]**: The error probability of the optimal codes decays exponentially with code length, \( n \)

\[ P_e \approx 2^{-nE^*(R)} \]

where \( E^*(R) \) is the (rate dependent) error exponent (a.k.a., channel reliability function).
What do we know about $E^*(R)$ for DMCs without feedback (after $\sim 50$ years of research)?

**Upper Bounds:**
- $E_{sp}(R)$
- $E_{st}(R)$
- $E_{md}(R)$

**Lower Bounds:**
- $E_{r}(R)$
- $E_{ex}(R)$
- $E_{T}(R)$

Above $R_{crit}$ the channel reliability function is known (matching bounds). Below $R_{crit}$ we have bounds (not matching).
DMC with feedback

\[ X_t \in \mathcal{X} \xrightarrow{\text{DMC}} Y_t \in \mathcal{Y} \]

Fixed-length (FL) code

C with length \( n \) (channel uses) and size \( M = 2^k \) (messages)

encoder: \( e_t = (e_t)_{t=1}^n \) with \( X_t = e_t(W, Y_{t-1}) \)

decoder: \( d: Y^n \rightarrow \{1, 2, \ldots, M\} \) with \( \hat{W} = d(Y_n) \)

Can also consider randomized encoders, e.g.,

\[ X_t \sim e_t(W, Y_{t-1}, V_t) \]

(with \( V_t \) some RV that induces the required randomness (possibly common information between Tx/Rx))
DMC with feedback

Fixed-length (FL) code $C$ with length $n$ (channel uses) and size $M = 2^k$ (messages)

encoder: $e = (e_t)_{t=1,...,n}$

$e_t : \{1, 2, \ldots, M\} \times \mathcal{Y}^{t-1} \rightarrow \mathcal{X}$ with $X_t = e_t(W, Y^{t-1})$

decoder: $d : \mathcal{Y}^n \rightarrow \{1, 2, \ldots, M\}$ with $\hat{W} = d(Y^n)$
Discrete memoryless channels (DMCs)

DMC with feedback and fixed-length (FL) codes

DMC with feedback

- Fixed-length (FL) code $C$ with length $n$ (channel uses) and size $M = 2^k$ (messages)
  
  encoder: $e = (e_t)_{t=1,...,n}$
  
  $e_t : \{1, 2, \ldots, M\} \times Y^{t-1} \rightarrow X$ with $X_t = e_t(W, Y^{t-1})$
  
  decoder: $d : Y^n \rightarrow \{1, 2, \ldots, M\}$ with $\hat{W} = d(Y^n)$
  
- Can also consider randomized encoders, e.g.,
  
  $X_t \sim e_t(\cdot|W, Y^{t-1}) \Leftrightarrow X_t = e_t(W, Y^{t-1}, V_t)$
  
  (with $V_t$ some RV that induces the required randomness (possibly common information between Tx/Rx))
DMC with feedback: basic results

- **Capacity:** Capacity cannot be improved by feedback for DMCs!

\[
C^{\text{Feedback}} = C^{\text{NoFeedback}}
\]
Discrete memoryless channels (DMCs)

DMC with feedback and fixed-length (FL) codes

**DMC with feedback: basic results**

- **Capacity**: Capacity cannot be improved by feedback for DMCs!

\[
C^{\text{Feedback}} = C^{\text{NoFeedback}}
\]

- **Error exponent for FL codes**: The error exponent with FL codes cannot be improved by feedback (at least for symmetric DMCs) above the critical rate! [Haroutunian, 1977]

\[
E^*_{\text{Feedback}}(R) \leq E_{\text{Haroutunian}}^\text{Feedback}(R)_{\text{symmetric DMCs}} = E_{sp}^{\text{NoFeedback}}(R)
\]

We can only hope for possible improvements in:

1. non-symmetric DMCs (e.g., Z-channel)
2. continuous-alphabet memoryless channels (e.g., Gaussian channels [Schalkwijk and Kailath, 1966])
3. “third-order” performance improvements and/or simpler encoding/decoding schemes
4. variable-length codes ?!
5. channels with memory
DMC with feedback: basic results

- **Capacity:** Capacity cannot be improved by feedback for DMCs!

  \[ C^{\text{Feedback}} = C^{\text{NoFeedback}} \]

- **Error exponent for FL codes:** The error exponent with FL codes cannot be improved by feedback (at least for symmetric DMCs) above the critical rate! [Haroutunian, 1977]

  \[ E^{*, \text{Feedback}}(R) \leq E^{\text{Feedback}}_{\text{Haroutunian}}(R) \mid_{\text{symmetric DMCs}} = E^{\text{NoFeedback}}_{\text{sp}}(R) \]

- We can only hope for possible improvements in:
  1. non-symmetric DMCs (e.g., Z-channel)
  2. continuous-alphabet memoryless channels (e.g., Gaussian channels [Schalkwijk and Kailath, 1966])
  3. “third-order” performance improvements and/or simpler encoding/decoding schemes
  4. variable-length codes ?!?
  5. channels with memory
DMC with feedback and variable-length codes

- Variable-length (VL) code $C$ with size $M = 2^k$ (messages)

  encoder: $e = (e_t)_{t=1,2,...}$
  $e_t : \{1,2,\ldots,M\} \times \mathcal{Y}^{t-1} \rightarrow \mathcal{X}$ with $X_t = e_t(W, Y^{t-1})$

  decoder: $d = (d_t)_{t=1,2,...}$
  $d_t : \mathcal{Y}^t \rightarrow \{1,2,\ldots,M\}$ with $\hat{W}_t = d_t(Y^t)$

  stopping time: $T$ with $\hat{W} = \hat{W}_T = d_T(Y^T)$
Variable-length (VL) code $C$ with size $M = 2^k$ (messages)

encoder: $e = (e_t)_{t=1,2,...}$

$e_t : \{1, 2, \ldots, M\} \times Y^{t-1} \to X$ with $X_t = e_t(W, Y^{t-1})$

decoder: $d = (d_t)_{t=1,2,...}$

$d_t : Y^t \to \{1, 2, \ldots, M\}$ with $\hat{W}_t = d_t(Y^t)$

stopping time: $T$ with $\hat{W} = \hat{W}_T = d_T(Y^T)$

Transmission time (code length), $T$, is a RV \(\Longrightarrow\) “Variable-length codes”.

**Average rate** $\bar{R} \triangleq \frac{\log M}{E[T]} = \frac{k}{E[T]}$

**Error probability** $P_e \triangleq P(W \neq \hat{W}_T \cup T = \infty)$
The reliability function is known exactly [Burnashev, 1976]

\[ E^{*,VL}(\bar{R}) = C_1(1 - \frac{\bar{R}}{C}) \]

where \( C_1 \) is a channel-dependent constant (max divergence)

\[ C_1 \triangleq \max_{x \neq x'} \sum_{y \in Y} Q(y|x) \log \frac{Q(y|x)}{Q(y|x')} = D(Q(·|x_0)||Q(·|x_1)) \]

Transmission schemes achieving this bound are known:

1. The Burnashev scheme [Burnashev, 1976]
2. The Yamamoto-Itoh scheme [Yamamoto and Itoh, 1979]
Error exponent for VL codes: BSC

**BSC(p)**

\[ BSC(p) \]

\[ \begin{array}{cc}
0 & 1-p \\
\hline
p & 0 \\
\hline
1-p & 1
\end{array} \]

\[ X \quad Y \]

\[ p \quad 1-p \]

\[ p \quad 1-p \]

\[ 0 \quad 1 \]

\[ 1 \quad 0 \]

\[ BSC(0.1) \]

C = 0.531

C₁ = 2.54

Burnashev's bound

Sphere-Packing bound

Example

n = 38 E[T]
Upper bound derivation: basic concepts

Define the entropy of the posterior message distribution $\Pi_t(i) \triangleq P(W = i | Y^t)$

$$H_t \triangleq H(\Pi_t) = - \sum_{i=1}^{M} \Pi_t(i) \log \Pi_t(i)$$

- Fano’s inequality: connection between $Pe = P(W \neq \hat{W}_t)$ and $H_t$
- Study the rate of decay of $H_t$ (drift analysis)
  $$E [H_{t+1} - H_t | Y^t] \geq -C$$ (from converse)
- When $H_t$ becomes very small above result is useless. Instead study exponential bounds
  $$E [\log(H_{t+1}) - \log(H_t) | Y^t] \geq -C_1, \quad H_t \approx \text{small}$$
Define a submartingale $\Xi_t$ based on $H_t$ for the two different regimes.

Technical difficulty: stitch together the two regimes in one random process, $\Xi_t$, and perform drift analysis.

Intuition: total transmission time

$$E[T] \geq t_1 + t_2 = \frac{k}{C} + \frac{-\log P_e}{C_1}$$

$$\Rightarrow \frac{-\log P_e}{E[T]} \leq C_1(1 - \frac{k}{E[T]} \frac{1}{C})$$

$$\Rightarrow E^{*,VL}(\bar{R}) \leq C_1(1 - \frac{\bar{R}}{C})$$
The constant $C_1$ is intimately related to **binary hypothesis testing**

When $H_t \approx \text{small}$, one of the messages (say, $W = i$) has very high ($\approx 1$) posterior probability

\[
E \left[ \log(H_{t+1}) - \log(H_t) | y^t \right] \\
\geq - \sum_{y_{t+1}} P(y_{t+1} | W = i, y^t) \log \frac{P(y_{t+1} | W = i, y^t)}{P(y_{t+1} | W \neq i, y^t)}
\]

\[
\geq - \max_{y^t,i} \sum_{y_{t+1}} Q(y_{t+1} | e_{t+1}(i, y^t)) \log \frac{Q(y_{t+1} | e_{t+1}(i, y^t))}{\sum_{j \neq i} Q(y_{t+1} | e_{t+1}(j, y^t)) \frac{\pi_t(j)}{1 - \pi_t(i)}}
\]

\[
\geq - \max_{x \neq x'} \sum_y Q(y|x) \log \frac{Q(y|x)}{Q(y|x')} \\
= - \max_{x \neq x'} D(Q(\cdot|x) || Q(\cdot|x'))
\]
Yamamoto-Itoh transmission scheme

- Transmit in blocks of length $n$
- Two stages in each packet: (1) data transmission; (2) confirmation
- Data transmission: use a capacity-achieving non-feedback code (not specified).
- At end of data transmission both Tx/Rx know if message error occurred
- Confirmation: Send 1 bit information whether decoded message was correct or not ($x_0$ if correct; $x_1$ if error)
- Repeat until correct confirmation is received (may take several blocks)
- Optimize $\gamma \in (0, 1)$ for largest error exponent of $P_{e^{data}} \times P(e \rightarrow c)$.

$$E(R) = \max_{\gamma \in (0,1)} \gamma E^{NF}(R/\gamma) + (1 - \gamma) C_1 \quad \frac{R}{\gamma} = C = 0 + (1 - \frac{R}{C}) C_1$$
Burnashev transmission scheme

Example: BSC(p). Capacity-achieving input distribution $P_X(0) = P_X(1) = 0.5$

\[
\begin{array}{c}
0 \quad 1-p \quad 0 \\
p \quad p \quad p \\
1-p \quad 1-p \quad 1
\end{array}
\]

Keep track of the posterior distribution (pmf) of the message

$$\Pi_t(i) \triangleq P(W = i | Y^t) \quad i = 1, \ldots, M$$

Randomized encoding ($Y^{t-1}$ is summarized in $\Pi_{t-1}$)

$$X_t = e_t(W, \Pi_{t-1}, V_t)$$

with $V_t$ common randomness between Tx/Rx

Two distinct transmission stages...
Burnashev scheme: stage 1

![Graph showing posterior message pmf P_{t-1}() for M=16.](image)
Burnashev scheme: stage 1

Find index \( m \), such that:

\[
\sum_{i=1}^{m} \Pi_{i-1}(i) > P_X(0)
\]

\[
\sum_{i=1}^{m-1} \Pi_{i-1}(i) \leq P_X(0)
\]
Burnashev scheme: stage 1

If \( W \) one of these, then \( X_t = 0 \)

If \( W \) one of these, then \( X_t = 1 \)
Discrete memoryless channels (DMCs)
DMC with feedback and variable-length (VL) codes

Burnashev scheme: stage 1

If $W=m$, then $X_t = 0, 1$
(randomization using common randomness)

If $W$ one of these, then $X_t = 0$
If $W$ one of these, then $X_t = 1$

Posterior message pmf, $\Pi_{1-t}^t(\cdot)$
Message index
Burnashev scheme: stage 1

- Let’s call this scheme TNGTNE (The noisy “guess the number” encoding)!
Burnashev scheme: stage 1

- Let’s call this scheme TNGTNE (The noisy “guess the number” encoding)! well...maybe not...
- **Discrete randomized posterior matching (DRPM)**

\[
X_t = DRPM(\Pi_{t-1}(\cdot), P_X(\cdot), W, V_t)
\]

where
\[
\Pi_{t-1}(\cdot) \in \mathcal{P}\{1, 2, \ldots, M\}
\]
\[
P_X(\cdot) \in \mathcal{P}(\mathcal{X})
\]
\[
W \in \{1, 2, \ldots, M\}
\]
\[
V_t \sim u([0, 1]) \text{ (independent of all } W, X^{t-1}, Y^{t-1}, V^{t-1})
\]
Burnashev scheme: stage 1

We expect that after a number of steps doing DRPM we get something like

- Continue in stage 1 with DRPM until $\max_i \Pi_{t_0}(i) > q$
- Provisional message estimate $\hat{W}_{t_0} = \arg \max_i \Pi_{t_0}(i)$
Burnashev scheme: stage 2

Stage 2:

- If $W = \hat{W}_t$ (hypothesis $h_0$) keep sending the predefined symbol $X_t = x_0$
- If $W \neq \hat{W}_t$ (hypothesis $h_1$) keep sending the predefined symbol $X_t = x_1$
- Continue in stage 2 until either
  \[ \max_i \Pi_t(i) > 1 - Pe \] (say at time $T$) and declare $\hat{W}_T = \arg \max_i \Pi_T(i)$
  or
  \[ \max_i \Pi_t(i) \] drops below threshold $q$, and go back to stage 1
Burnashev scheme: stage 2

Stage 2:

- If $W = \hat{W}_t$ (hypothesis $h_0$) keep sending the predefined symbol $X_t = x_0$
- If $W \neq \hat{W}_t$ (hypothesis $h_1$) keep sending the predefined symbol $X_t = x_1$
- Continue in stage 2 until
  - either $\max_i \Pi_t(i) > 1 - Pe$ (say at time $T$) and declare $\hat{W}_T = \arg \max_i \Pi_T(i)$
  - or $\max_i \Pi_t(i)$ drops below threshold $q$, and go back to stage 1

- No codebook to store at Tx/Rx; simple decoding at Rx
Burnashev scheme: Analysis

[Burnashev, 1976]:

- Would like to analyze how fast \( \log \frac{\max_i \Pi_t(i)}{1-\max_i \Pi_t(i)} \) (log-likelihood ratio of **best** message posterior probability) grows towards the threshold \( \log \frac{1-P_e}{P_e} \)
- Instead, analyze the process \( L_t = \log \frac{\Pi_t(W)}{1-\Pi_t(W)} \) (log-likelihood ratio of **true** message posterior probability)
- \( E [L_{t+1} - L_t | Y^t] \geq C \)
- \( E [L_{t+1} - L_t | Y^t] \geq C_1 > C \) (if \( L_t > \log \frac{q}{1-q} \), for appropriately defined \( q \))
- Create a submartingale \( Z_t \) from \( L_t \) and apply optional stopping theorem
- Intuition: geometric picture
Burnashev scheme: Analysis, Intuition

Intuition: total transmission time

\[
E[T] \leq t_1 + t_2 = \frac{k}{C} + \frac{-\log Pe}{C_1}
\]

\[
\Rightarrow \frac{-\log Pe}{E[T]} \geq C_1 \left(1 - \frac{k}{C} \right)
\]

\[
\Rightarrow E^*, VL(\bar{R}) \geq C_1 \left(1 - \frac{\bar{R}}{C} \right)
\]
Error exponent for VL codes: BSC simulation
Error exponent for VL codes: BSC simulation

- Burnashev's bound
- Sphere-Packing bound
- Example: $E[T]=25.5$ vs. $n=339$

BSC(0.1)
$C=0.531$
$C_1=2.54$

$k=10$
Overview

1. Discrete memoryless channels (DMCs)
   - DMC without feedback
   - DMC with feedback and fixed-length (FL) codes
   - DMC with feedback and variable-length (VL) codes

2. Channels with memory and feedback
   - Known capacity results
   - Recent results for error exponents of VL codes
Unifilar channel with feedback

- Information message $W \in \{1, 2, \ldots, M\}$
- Transmitted symbols $X_t \in \mathcal{X}$, $t = 1, 2, \ldots$
- Channel state $S_t \in \mathcal{S}$, $t = 1, 2, \ldots$
- Received symbols $Y_t \in \mathcal{Y}$, $t = 1, 2, \ldots$
- Input/output conditional distribution $Q(Y_t|X_t, S_t)$
- Deterministic state update $S_{t+1} = g(S_t, X_t, Y_t)$
- Encoding functions $X_t = e_t(W, Y^{t-1}, S_1, V_t)$, $t = 1, 2, \ldots$
- Decoding function $\hat{W}_t = d_t(Y^t)$ (together with a stopping time $T$)
Unifilar channel with feedback

- Information message $W \in \{1, 2, \ldots, M\}$
- Transmitted symbols $X_t \in \mathcal{X}$, $t = 1, 2, \ldots$
- Channel state $S_t \in \mathcal{S}$, $t = 1, 2, \ldots$
- Received symbols $Y_t \in \mathcal{Y}$, $t = 1, 2, \ldots$
- Input/output conditional distribution $Q(Y_t|X_t, S_t)$
- Deterministic state update $S_{t+1} = g(S_t, X_t, Y_t)$
- Encoding functions $X_t = e_t(W, Y_{t-1}, S_1, V_t)$, $t = 1, 2, \ldots$
- Decoding function $\hat{W}_t = d_t(Y^t)$ (together with a stopping time $T$)

- State known to $Tx$ but not to $Rx$ ($Tx$ knows $S_1, X_{t-1}^{t-1}, Y_{t-1}^{t-1} \Rightarrow S^t$)
Unifilar channel with feedback: capacity

Capacity is the result of the **off-line** optimization problem [Permuter et al., 2008] over infinitely many conditional distributions on $\mathcal{X}$

$$C = \lim_{N \to \infty} \sup \{ P(X_t | S_t, Y_{t-1}^t) \}_{t \geq 1} \frac{1}{N} \sum_{i=1}^{N} I(X_t, S_t; Y_t | Y_{t-1}^t).$$

- Observe: $P_{X_t | S_t, Y_{t-1}^t} \in S \times \mathcal{Y}^{t-1} \to \mathcal{P}(\mathcal{X})$, so its domain increases with $t$
Channels with memory and feedback

Unifilar channel with feedback: capacity

Capacity is the result of the **off-line** optimization problem [Permuter et al., 2008] over infinitely many conditional distributions on $\mathcal{X}$

$$C = \lim_{N \to \infty} \sup \left\{ \frac{1}{N} \sum_{i=1}^{N} I(X_t, S_t; Y_t | Y^{t-1}) \right\}$$

- Observe: $P_{X_t|S_t,Y_{t-1}} \in \mathcal{S} \times \mathcal{Y}^{t-1} \to \mathcal{P}(\mathcal{X})$, so its domain increases with $t$

- How can we utilize Control theory to solve this problem?
Unifilar channel with feedback: capacity

Capacity is the result of the **off-line** optimization problem [Permuter et al., 2008] over infinitely many conditional distributions on $\mathcal{X}$

$$ C = \lim_{N \to \infty} \sup_{\{P(X_t|S_t,Y_{t-1})\}_{t \geq 1}} \frac{1}{N} \sum_{i=1}^{N} I(X_t, S_t; Y_t | Y^{t-1}) $$

- Observe: $P_{X_t|S_t,Y_{t-1}} \in S \times Y^{t-1} \to \mathcal{P}(\mathcal{X})$, so its domain increases with $t$

- Define posterior belief of the state ($Tx/Rx$ can evaluate it)

$$ B_t(s) \triangleq P(S_{t+1} = s|Y^t) $$

- $\{B_t\}_t$ forms a (controlled) Markov process, which can be (partially) controlled by $P(X_t|S_t, Y^{t-1})$

- Utilize theory of Markov Decision Processes (MDPs) to derive a single-letter expression [Permuter et al., 2008]

$$ C = \sup_{P(X_t|S_t,B_{t-1})} I(X_t, S_t; Y_t | B_{t-1}) $$
Error exponents for VL coding: upper bound

- How can we generalize Burnashev’s analysis to channels with memory?
Error exponents for VL coding: upper bound

How can we generalize Burnashev’s analysis to channels with memory?

Basic idea #1: analyze multi-step drift (to capture memory effects)
For any $\epsilon > 0$ there exists a large enough step $N$, s.t.

$$
\frac{1}{N} E[H_{t+N} - H_t | Y^t = y^t, S_1 = s_1] \\
\geq - \frac{1}{N} \sum_{k=t}^{t+N-1} I(X_k+1, S_k+1; Y_{k+1}^k | Y_{t+1}^t, Y^t = y^t, S_1 = s_1) \\
\geq -(C + \epsilon), \quad \text{(from ergodicity of } \{B_t\}_t) 
$$

and similarly (for the case of small $H_t$)

$$
\frac{1}{N} E[\log(H_{t+N}) - \log(H_t) | Y^t, S_1] \geq -(C_1 + \epsilon) \quad \text{a.s.}
$$
What is $C_1$ in this case?

\[
C_1 = \max_{s_1, y^t, i} \lim_{N \to \infty} \sup_{\tau = t+1} \max_{e_{\tau}} \left\{ \frac{1}{N} \sum_{Y_{t+1}^{t+N}} P(Y_{t+1}^{t+N} | W = i, y^t, s_1) \log \frac{P(Y_{t+1}^{t+N} | W = i, y^t, s_1)}{P(Y_{t+1}^{t+N} | W \neq i, y^t, s_1)} \right\}.
\]

It relates to a binary hypothesis testing problem with

- h0: $W = i$
- h1: $W \neq i$
Error exponents for VL coding: the $C_1$ constant

- **Basic idea #2**: Define $X_t^i = e_t(i, Y^{t-1}, S_1)$ and $S_t^i = g_t(i, Y^{t-1}, S_1)$ which are the input and the state at time $t$, conditioned on $W = i$. Then,

$$P(Y_{t+1}^{t+N} | W = i, y^t, s_1) = \prod_{\tau = t+1}^{t+N} Q(Y_{\tau} | S_{\tau}^i, X_{\tau}^i)$$

- Define $X_t^i(x|s)$, as the induced input distribution at time $t$, conditioned on $S_t = s$ and $W \neq i$
- Define $B_{t-1}^1(s) \triangleq P(S_t = s | W \neq i, Y^{t-1}, S_1)$ as the posterior state belief at time $t$, conditioned on $W \neq i$. Then,

$$P(Y_{t+1}^{t+N} | y^t, s_1, W \neq i) = \prod_{\tau = t+1}^{t+N} \left[ \sum_{x,s} Q(Y_{\tau} | x, s) X_{\tau}^i(x|s) B_{\tau-1}^1(s) \right]$$
Error exponents for VL coding: the $C_1$ constant

- $C_1$ relates to the average reward per unit time of an MDP with:
  - state: $(S^0_t, B^1_{t-1}) \in S \times \mathcal{P}(S)$,
  - action: $(X^0_t, X^1_t) \in \mathcal{X} \times (S \rightarrow \mathcal{P}(\mathcal{X}))$,
  - instantaneous reward: $R(S^0_t, B^1_{t-1}; X^0_t, X^1_t)$,
  - transition kernel:

$$P(S^0_{t+1}, B^1_t | S^0_t, B^1_{t-1}, X^0_t, X^1_t) = \sum_y \delta_g(S^0_t, X^0_t, y)(S^0_{t+1}) \delta_\phi(B^1_{t-1}, X^1_t, y)(B^1_t) Q(y | X^0_t, S^0_t).$$
Error exponents for VL coding: the $C_1$ constant

- $C_1$ relates to the average reward per unit time of an MDP with:
  - state: $(S^0_t, B^1_{t-1}) \in S \times \mathcal{P}(S)$,
  - action: $(X^0_t, X^1_t) \in \mathcal{X} \times (S \rightarrow \mathcal{P}(\mathcal{X}))$,
  - instantaneous reward: $R(S^0_t, B^1_{t-1}; X^0_t, X^1_t)$,
  - transition kernel:
    \[
    P(S^0_{t+1}, B^1_t | S^0_t, B^1_{t-1}, X^0_t, X^1_t) \\
    = \sum_y \delta_{g(S^0_t, X^0_t, y)}(S^0_{t+1}) \delta_{\phi(B^1_{t-1}, X^1_t, y)}(B^1_t) Q(y | X^0_t, S^0_t).
    \]

- From MDP theory: optimal action only function of current state: $X^0[S^0_t, B^1_{t-1}]$ and $X^1[S^0_t, B^1_{t-1}](\cdot | \cdot)$
Error exponents for VL coding: the $C_1$ constant

- Intuition gained: $C_1$ relates to a binary hypothesis test over a channel with memory, with a weird twist!
  - Under $h_0$, input is a deterministic symbol (function of current state $S_t^0$ and belief $B_{t-1}^1$)
    \[ X_t = X^0[S_t^0, B_{t-1}^1] \]
  - Under $h_1$, input is a random symbol (function of hypothesized state under $h_0$, $S_t^0$, belief $B_{t-1}^1$, and state $S_t \sim B_{t-1}^1(\cdot)$)
    \[ X_t \sim X^1[S_t^0, B_{t-1}^1](\cdot|S_t) \]
Error exponents for VL coding: the $C_1$ constant

- Intuition gained: $C_1$ relates to a binary hypothesis test over a channel with memory, with a weird twist!
  - Under $h0$, input is a deterministic symbol (function of current state $S_t^0$ and belief $B_{t-1}^1$)
    $$X_t = X^0[S_t^0, B_{t-1}^1]$$
  - Under $h1$, input is a random symbol (function of hypothesized state under $h0$, $S_t^0$, belief $B_{t-1}^1$, and state $S_t \sim B_{t-1}^1(\cdot)$)
    $$X_t \sim X^1[S_t^0, B_{t-1}^1](\cdot | S_t)$$

- Why so complicated? dual objective
  1. resolving the hypothesis by transmitting the most distinguishable symbols and
  2. partially controlling the channel state evolution
Error exponents for VL coding: the $C_1$ constant

- Intuition gained: $C_1$ relates to a binary hypothesis test over a channel with memory, with a weird twist!
  - Under $h_0$, input is a deterministic symbol (function of current state $S_t^0$ and belief $B_{t-1}^1$)
    \[ X_t = X^0[S_t^0, B_{t-1}^1] \]
  - Under $h_1$, input is a random symbol (function of hypothesized state under $h_0$, $S_t^0$, belief $B_{t-1}^1$, and state $S_t \sim B_{t-1}^1(\cdot)$)
    \[ X_t \sim X^1[S_t^0, B_{t-1}^1](\cdot | S_t) \]

- Why so complicated? dual objective
  (1) resolving the hypothesis by transmitting the most distinguishable symbols and
  (2) partially controlling the channel state evolution

- Challenge: turn this into an actual transmission scheme!
  Last part of this talk...
A Burnashev-like VL coding scheme

- Keep track of the posterior distribution (pmf) of the message

$$\Pi_t(i) \triangleq P(W = i|\mathcal{F}_t) \quad i = 1, \ldots, M$$

and the vector of states

$$S_t = (S^1_t, S^2_t, \ldots, S^M_t),$$

where $S^i_t$ is the hypothesized state at time $t$ conditioned on $W = i$. 
A Burnashev-like VL coding scheme

- Keep track of the posterior distribution (pmf) of the message

\[ \Pi_t(i) \triangleq P(W = i | F_t) \quad i = 1, \ldots, M \]

and the vector of states

\[ S_t = (S_t^1, S_t^2, \ldots, S_t^M), \]

where \( S_t^i \) is the hypothesized state at time \( t \) conditioned on \( W = i \).

- Calculate posterior beliefs

\[
\hat{B}_{t-1}(s) = \sum_{i=1}^{M} \Pi_{t-1}(i) 1_{\{S_t^i = s\}} = P(S_t = s | F_{t-1}) \\
\Pi_{t-1}^s(i) = \frac{\Pi_{t-1}(i) 1_{\{S_t^i = s\}}}{\hat{B}_{t-1}(s)} = P(W = i | S_t = s, F_{t-1}),
\]

Two distinct transmission stages...
A Burnashev-like VL coding scheme: stage 1

Stage 1: \[ X_t = DRPM(\prod_{t-1}^{S_t}(\cdot), P_{X|SB}(\cdot|S_t, \hat{B}_{t-1}), W, V_t) \]

Continue in stage 1 until \( \max_i \prod_{t_0}^{i}(i) > q \)
Provisional message estimate \( \hat{W} = \arg \max_i \prod_{t_0}^{i}(i) \)
A Burnashev-like VL coding scheme: stage 2

Stage 2:

- Calculate new posteriors (conditioned on h1: $W \neq \hat{W}_0$)

$$
\hat{B}_{t-1}^1(s) = \frac{\sum_{i \neq \hat{W}} \Pi_{t-1}(i) 1\{S_t = s\}}{1 - \Pi_{t-1}(\hat{W})} = P(S_t = s | F_{t-1}, h1)
$$

$$
\Pi_{t-1}^{1,s}(i) = \frac{\Pi_{t-1}(i) 1\{i \neq \hat{W}\} 1\{S_t = s\}}{\hat{B}_{t-1}^1(s)(1 - \Pi_{t-1}(\hat{W}))} = P(W = i | S_t = s, F_{t-1}, h1)
$$

If $W = \hat{W}_0$ (hypothesis $h_0$) transmit $X_t = X_0 [S_{\hat{W}t}, \hat{B}_{t-1}^1]$

If $W \neq \hat{W}_0$ (hypothesis $h_1$) transmit $X_t = DRPM(\Pi_1, S_W t_{t-1}(\cdot), X_1 [S_{\hat{W}t}, \hat{B}_{t-1}^1](\cdot | S_t, W, V_t])$

Continue in stage 2 until either

- $\max_i \Pi_t(i) > 1 - P_e$ (say at time $T$) and declare $\hat{W}_T = \arg \max_i \Pi_T(i)$
- $\max_i \Pi_t(i)$ drops below threshold $q$, and go back to stage 1
A Burnashev-like VL coding scheme: stage 2

Stage 2:

- Calculate new posteriors (conditioned on h1: $W \neq \hat{W}_0$)

$$\hat{B}_{t-1}^1(s) = \frac{\sum_{i \neq \hat{W}} \Pi_{t-1}(i) 1\{s^i = s\}}{1 - \Pi_{t-1}(\hat{W})} = P(S_t = s | \mathcal{F}_{t-1}, h1)$$

$$\Pi_{t-1}^{1,s}(i) = \frac{\Pi_{t-1}(i) 1\{i \neq \hat{W}\} 1\{s^i = s\}}{\hat{B}_{t-1}^1(s)(1 - \Pi_{t-1}(\hat{W}))} = P(W = i|S_t = s, \mathcal{F}_{t-1}, h1)$$

- If $W = \hat{W}$ (hypothesis h0) transmit $X_t = X^0[S_t^{\hat{W}}, \hat{B}_{t-1}^1]$
A Burnashev-like VL coding scheme: stage 2

Stage 2:

- Calculate new posteriors (conditioned on h1: $W \neq \hat{W}_{t_0}$)

$$\hat{B}_{t-1}^1(s) = \frac{\sum_{i \neq \hat{W}} \Pi_{t-1}(i) 1\{s^i = s\}}{1 - \Pi_{t-1}(\hat{W})} = P(S_t = s | \mathcal{F}_{t-1}, h1)$$

$$\Pi_{t-1}^{1,s}(i) = \frac{\Pi_{t-1}(i) 1\{i \neq \hat{W}\} 1\{s^i = s\}}{\hat{B}_{t-1}^1(s)(1 - \Pi_{t-1}(\hat{W}))} = P(W = i | S_t = s, \mathcal{F}_{t-1}, h1)$$

- If $W = \hat{W}$ (hypothesis h0) transmit $X_t = X^0[S_t^{\hat{W}}, \hat{B}_{t-1}^1]$

- If $W \neq \hat{W}$ (hypothesis h1)

$$X_t = DRPM(\Pi_{t-1}^{1,S_t^W}(\cdot), X^1[S_t^{\hat{W}}, \hat{B}_{t-1}^1](\cdot | S_t), W, V_t)$$

- Continue in stage 2 until either

  $\max_i \Pi_t(i) > 1 - Pe$ (say at time $T$) and declare $\hat{W}_T = \arg \max_i \Pi_T(i)$

  or

  $\max_i \Pi_t(i)$ drops below threshold $q$, and go back to stage 1
A Burnashev-like VL coding scheme: Analysis

- Analyze the one-step drift of the process $L_t \triangleq \log \frac{\Pi_t(W)}{1-\Pi_t(W)}$ and use ergodicity to get multi-step results.
A Burnashev-like VL coding scheme: Analysis

- Analyze the one-step drift of the process \( L_t \triangleq \log \frac{\Pi_t(W)}{1 - \Pi_t(W)} \) and use ergodicity to get multi-step results.

**Unresolved issue:** The defined process \( \{\hat{B}_t\}_t \) does not have the same statistics as \( \{B_t\}_t \) (related to capacity expression).

\[
B_{t-1}(s) = P(S_t = s | Y^{t-1}, S_1) \quad \text{vs} \quad \hat{B}_{t-1}(s) = P(S_t = s | Y^{t-1}, V^{t-1}, S_1)
\]

This is because of the introduction of common randomness (RVs \( V_t \))!

In fact \( \{\hat{B}_t\}_t \) is not a Markov chain (but \( \hat{B}_{t-1} \) is measurable wrt a “bigger” Markov chain \( \{(S_t, \Pi_{t-1})\}_t \) )
A Burnashev-like VL coding scheme: Analysis

- Analyze the one-step drift of the process $L_t \triangleq \log \frac{\Pi_t(W)}{1-\Pi_t(W)}$ and use ergodicity to get multi-step results

- **Unresolved issue**: The defined process $\{\hat{B}_t\}_t$ does not have the same statistics as $\{B_t\}_t$ (related to capacity expression)

  - $B_{t-1}(s) = P(S_t = s | Y^{t-1}, S_1)$ vs $\hat{B}_{t-1}(s) = P(S_t = s | Y^{t-1}, V^{t-1}, S_1)$

  - This is because of the introduction of common randomness (RVs $V_t$)!

  - In fact $\{\hat{B}_t\}_t$ is not a Markov chain (but $\hat{B}_{t-1}$ is measurable wrt a “bigger” Markov chain $\{(S_t, \Pi_{t-1})\}_t$)

  - Some kind of “concentration” result is at play here. **Any ideas?**
Numerical/Simulation results

Upper Bound

\[ S_{t+1} = S_t \oplus X_t \oplus Y_t \]

\[ S_t = 0 \]

\[ S_t = 1 \]

\( k = 10 \)

\( k = 20 \)

\( k = 30 \)

\( C = 0.369 \)

\( C_1 = 1.637 \)

\( C_1^* = 1.533 \)
Numerical/Simulation results

Trapdoor channel

\[ S_{t+1} = S_t \oplus X_t \oplus Y_t \]

- \( S_t = 0 \)
- \( S_t = 1 \)

- \( k = 10 \)
- \( k = 20 \)
- \( k = 30 \)

- \( C = 0.697 \)
- \( C_1 = \infty \)
- \( C_1^* = \infty \)
Thank you!


A coding scheme for additive noise channels with feedback–I: No bandwidth constraint.


A mathematical theory of communication.

Lower bounds to error probability for coding on discrete memoryless channels.

Asymptotic performance of a modified schalkwijk-barron scheme for channels with noiseless feedback (corresp.).