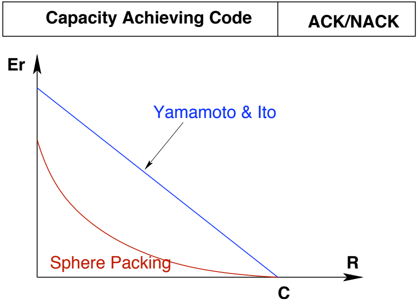
Tilted Matching for Feedback Channels

Barış Nakiboğlu   Lizhong Zheng

ITMANET, Sep. 14, 2009

Feedback for Error Control

- Feedback does not increase the capacity of DMC.
- Feedback is often used to correct errors via *architectural* changes:
  Example: Yamamoto & Ito ' 79



Burnashev'76: optimal error exponent for variable-length codes

Why Variable Length Codes are Nice?

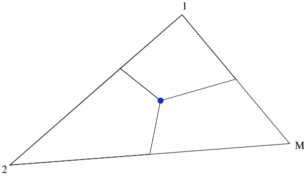
- Ack/NAck/Retransmission is much more efficient than Forward Error Correction without feedback;
- What is hidden from the variable code length?
  - Retransmission costs ignored in average transmission time;
  - Forward transmission remains the same as the no-feedback case;
  - Do not try to recover from a partial error;
- Recent extensions:
  - Fixed length block codes with erasures: Nakiboglu & Z'09
  - Non-block codes with fixed or variable delays: Draper & Sahai'08

Feedback for Forward Error Correction

- Feedback communication schemes that employ "incrementally" tuned encoding
  - *Noise variance reduction in AWGN:* [Elias, 56], [Schalkwijk & Kailath, 66],[Schalkwijk, 66, 68]
  - *Posterior matching:* [Horstein, 63], [Shayevitz & Feder, 07,08]
  - *Dumped down encoding:* [Zigangirov, 70], [D' yachkov, 75]
- Why are these important? Dynamic information exchange.

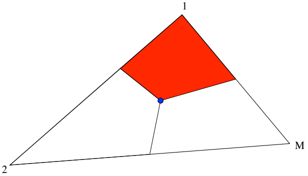
Dynamic View of Coding

- Communication as "driving the belief" at the receiver (Coleman)



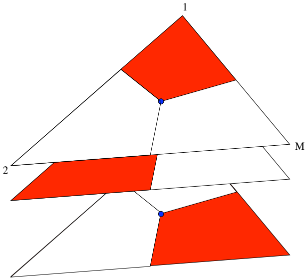
Dynamic View of Coding

- Communication as "driving the belief" at the receiver (Coleman)
- Decision region and reward at the end of the block



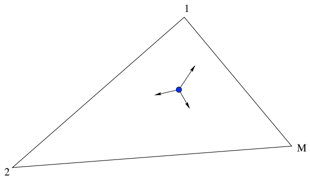
Dynamic View of Coding

- Communication as "driving the belief" at the receiver (Coleman)
- Decision region and reward at the end of the block
- Encoding function does not depend on the correct message, solve parallel problems



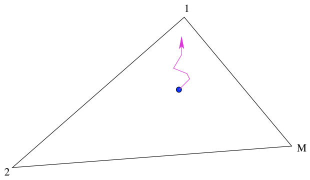
Dynamic View of Coding

- Communication as "driving the belief" at the receiver (Coleman)
- Decision region and reward at the end of the block
- Encoding function does not depend on the correct message, solve parallel problems
- Randomness from the channel



Dynamic View of Coding

- Communication as "driving the belief" at the receiver (Coleman)
- Decision region and reward at the end of the block
- Encoding function does not depend on the correct message, solve parallel problems
- Randomness from the channel
- With feedback (or noisy feedback), encoding can depend on the current belief



Solutions of the Dynamic Coding Problem

- Variety of approximate solutions, depending on how to approximate the reward function, and how to average.
- Example: Posterior matching, Shayevitz & Feder'08

  $$X_{t+1} = F_X^{-1}(F(M|y^t)) \quad \text{is capacity achieving}$$

  Coleman'09: This is the optimal solution when using average K-L divergence as reward;
- Our question: What is the dynamic problem that leads to the optimal error exponent? What are we ignoring when doing block codes?

Error Exponents for K-ary Symmetric Channels with Feedback

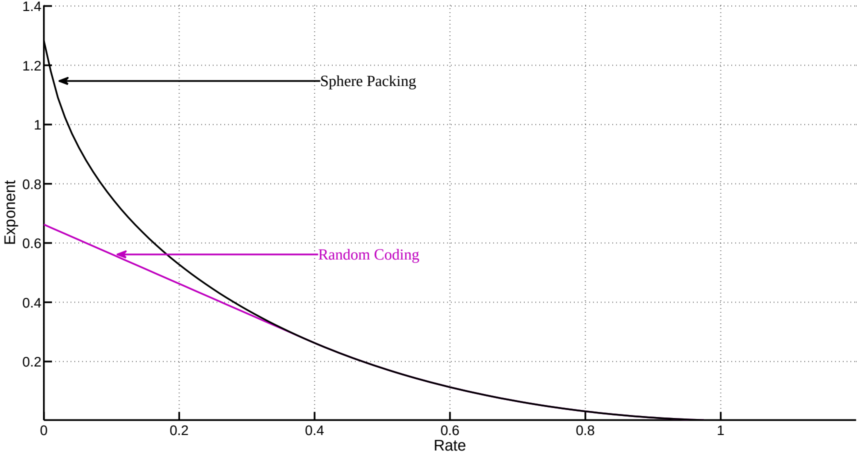


Figure: k-ary symmetric channel with $k = 4$ and $\epsilon = .3/4$.

Error Exponents for K-ary Symmetric Channels with Feedback

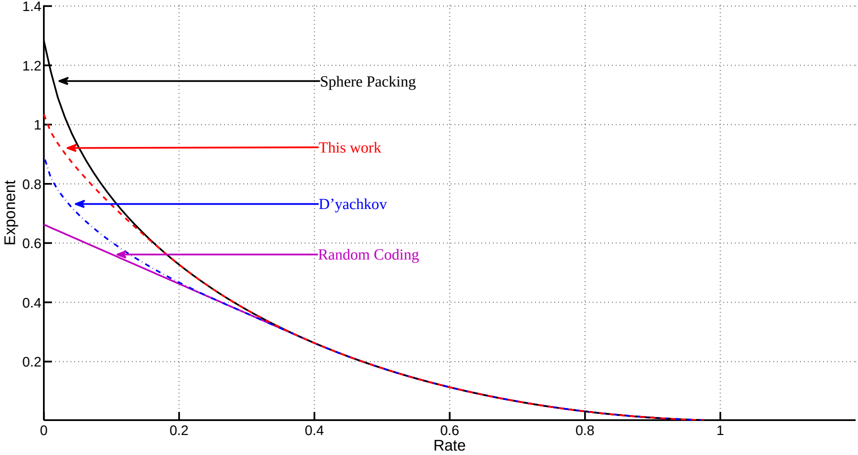
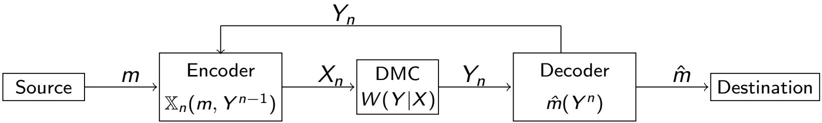


Figure: k-ary symmetric channel with $k = 4$ and $\epsilon = .3/4$.

Notations



- Encoder:

$$\mathbb{X}_t(m, Y^{t-1}) : \mathcal{M} \times \mathcal{Y}^{t-1} \to \mathcal{X} \qquad t \leq n$$

- Decoder:

$$\hat{m}(Y^n) : \qquad \mathcal{Y}^n \to \mathcal{M}$$

- Knowledge at time t, $\varphi_t(\cdot) = \mathbf{P}\left[m = \cdot \mid y^t\right]$

Upper Bound of Error Probability

$$\begin{aligned} P_{\mathbf{e}} &= \sum_{y^n} \mathbf{P}\left[y^n\right] \sum_k \varphi_n(k) \mathbb{I}\{\hat{m}(y^n) \neq k\} \\ &= \mathbf{E}\left[\sum_k \varphi_n(k) \mathbb{I}\{\hat{m}(Y^n) \neq k\}\right] \end{aligned}$$

Upper Bound of Error Probability

$$
\begin{aligned}
P_{\mathbf{e}} &= \sum_{y^n} \mathbf{P}[y^n] \sum_k \varphi_n(k) \mathbb{I}\{\hat{m}(y^n) \neq k\} \\
&= \mathbf{E}\left[\sum_k \varphi_n(k) \mathbb{I}\{\hat{m}(Y^n) \neq k\} \right]
\end{aligned}
$$

- Gallager's trick: for a ML decoder:

$$
\mathbb{I}\{\hat{m}(y^n) \neq k\} \leq \frac{\max_{l \neq k} \varphi_n(l)}{\varphi_n(k)} ,
$$

Upper Bound of Error Probability

$$
\begin{aligned}
P_{\mathbf{e}} &= \sum_{y^n} \mathbf{P}\left[y^n\right] \sum_k \varphi_n(k) \mathbb{I}\{\hat{m}(y^n) \neq k\} \\
&= \mathbf{E}\left[\sum_k \varphi_n(k) \mathbb{I}\{\hat{m}(Y^n) \neq k\}\right]
\end{aligned}
$$

- Gallager's trick: for a ML decoder:

$$
\mathbb{I}\{\hat{m}(y^n) \neq k\} \leq \frac{\max_{l \neq k} \varphi_n(l)^\eta}{\varphi_n(k)^\eta}, \quad \forall \eta > 0,
$$

Upper Bound of Error Probability

$$
\begin{aligned}
P_{\mathbf{e}} &= \sum_{y^n} \mathbf{P}[y^n] \sum_k \varphi_n(k)\mathbb{I}\{\hat{m}(y^n) \neq k\} \\
&= \mathbf{E}\left[\sum_k \varphi_n(k)\mathbb{I}\{\hat{m}(Y^n) \neq k\}\right]
\end{aligned}
$$

- Gallager's trick: for a ML decoder:

$$
\mathbb{I}\{\hat{m}(y^n) \neq k\} \leq \left(\frac{\max_{l \neq k} \varphi_n(l)^\eta}{\varphi_n(k)^\eta}\right)^\rho, \quad \forall \eta > 0, \rho > 0
$$

Upper Bound of Error Probability

$$
\begin{aligned}
P_{\mathbf{e}} &= \sum_{y^n} \mathbf{P}\left[y^n\right] \sum_k \varphi_n(k) \mathbb{I}\{\hat{m}(y^n) \neq k\} \\
&= \mathbf{E}\left[\sum_k \varphi_n(k) \mathbb{I}\{\hat{m}(Y^n) \neq k\}\right]
\end{aligned}
$$

- Gallager's trick: for a ML decoder:

$$
\begin{aligned}
\mathbb{I}\{\hat{m}(y^n) \neq k\} &\leq \left(\frac{\max_{l \neq k} \varphi_n(l)^\eta}{\varphi_n(k)^\eta}\right)^\rho, \quad \forall \eta > 0, \rho > 0 \\
&\leq \left(\frac{\sum_{l \neq k} \varphi_n(l)^\eta}{\varphi_n(k)^\eta}\right)^\rho
\end{aligned}
$$

Upper Bound of Error Probability

$$
\begin{aligned}
P_{\mathbf{e}} &= \sum_{y^n} \mathbf{P}\left[y^n\right] \sum_k \varphi_n(k) \mathbb{I}\{\hat{m}(y^n) \neq k\} \\
&= \mathbf{E}\left[\sum_k \varphi_n(k) \mathbb{I}\{\hat{m}(Y^n) \neq k\}\right]
\end{aligned}
$$

- Gallager's trick: for a ML decoder:

$$
\begin{aligned}
\mathbb{I}\{\hat{m}(y^n) \neq k\} &\leq \left(\frac{\max_{l \neq k} \varphi_n(l)^\eta}{\varphi_n(k)^\eta}\right)^\rho, \quad \forall \eta > 0, \rho > 0 \\
&\leq \left(\frac{\sum_{l \neq k} \varphi_n(l)^\eta}{\varphi_n(k)^\eta}\right)^\rho
\end{aligned}
$$

- Rearrange:

$$
P_{\mathbf{e}} \leq \mathbf{E}\left[\sum_k \varphi_n(k)^{1-\rho\eta} \Big(\sum_{l \neq k} \varphi_n(l)^\eta\Big)^\rho\right] = \mathbf{E}\left[\zeta\left(\varphi_n(\cdot), \eta, \rho\right)\right]
$$

Upper Bound of Error Probability

$$
\begin{aligned}
P_{\mathbf{e}} &= \sum_{y^n} \mathbf{P}[y^n] \sum_k \varphi_n(k) \mathbb{I}\{\hat{m}(y^n) \neq k\} \\
&= \mathbf{E}\left[\sum_k \varphi_n(k) \mathbb{I}\{\hat{m}(Y^n) \neq k\}\right]
\end{aligned}
$$

- Gallager's trick: for a ML decoder:

$$
\begin{aligned}
\mathbb{I}\{\hat{m}(y^n) \neq k\} &\leq \left(\frac{\max_{l \neq k} \varphi_n(l)^\eta}{\varphi_n(k)^\eta}\right)^\rho, \quad \forall \eta > 0, \rho > 0 \\
&\leq \left(\frac{\sum_{l \neq k} \varphi_n(l)^\eta}{\varphi_n(k)^\eta}\right)^\rho
\end{aligned}
$$

- Rearrange:

$$
P_{\mathbf{e}} \leq \mathbf{E}\left[\sum_k \varphi_n(k)^{1-\rho\eta} \Big(\sum_{l \neq k} \varphi_n(l)^\eta\Big)^\rho\right] = \mathbf{E}[\zeta_n]
$$

Smoothed Problem

- $\varphi_t(\cdot)$ summarizes the knowledge up to time t;
- $\zeta_n \triangleq \sum_k \varphi_n(k)^{1-\rho\eta}(\sum_{l\neq k} \varphi_n(l)^\eta)^\rho$;
- ζ_t similarly defined;
- $\zeta_0 = e^{nR\rho}$;
- At the end of the block: $P_{\mathbf{e}} \leq \mathbf{E}[\zeta_n]$;
- Design goal: at time t, with knowledge $\varphi_t(\cdot)$ and ζ_t, make ζ_{t+1} as small as possible, on average.
- Approximations exponentially tight.

Uniform Progress Assumptions

- Suppose there exists an encoding scheme such that

$$\mathbf{E}\left[\zeta_{t+1}|y^t\right] \leq \alpha(\rho,\eta)\zeta_t, \qquad \forall t, y^t$$

Uniform Progress Assumptions

- Suppose there exists an encoding scheme such that

$$\mathbf{E}\left[\zeta_{t+1} | y^t\right] \le \alpha(\rho,\eta)\zeta_t, \qquad \forall t, y^t$$

Then

$$\begin{aligned} P_{\mathbf{e}} &\le \mathbf{E}\left[\zeta_n\right] \\ &\le \mathbf{E}\left[\mathbf{E}\left[\zeta_n | Y^{n-1}\right]\right] \\ &\le \alpha(\rho,\eta)\mathbf{E}\left[\zeta_{n-1}\right] \\ &\le \alpha(\rho,\eta)^n \mathbf{E}\left[\zeta_0\right] \\ &\le \alpha(\rho,\eta)^n e^{nR\rho} \end{aligned}$$

Random Coding Without Feedback

- Recall
$$\mathbf{E}\left[\zeta_{t+1}|y^t\right] \leq \alpha(\rho,\eta)\zeta_t, \qquad \forall t, y^t$$
where ζ_{t+1} depends on Y_{t+1}, and X_{t+1};
- For random coding, and any $\rho \in [0,1]$, take $\eta = \frac{1}{1+\rho}$,
$$\mathbf{E}\left[\zeta_{t+1}|y^t\right] \leq e^{-E_0(P,\rho)}\zeta_t$$
where $E_0(P,\rho) = -\log \sum_y \left(\sum_x W(y|x)^{\frac{1}{1+\rho}} P(x)\right)^{1+\rho}$.
- Uniform progression is natural when there is no feedback

Tilted Matching for Channel with Feedback

- Potentially with feedback, we can design coding map at each time t to depend on y^t, and to minimize $\mathbf{E}[\zeta_n]$.
- We focus on coding schemes that only depend on $\varphi_t(\cdot), \rho$, and η, and follow the uniform progress assumption.
- The optimal choice, under assumption (*), is a posterior matching of $\varphi_t(\cdot)^\eta$, instead of $\varphi_t(\cdot)$ itself.
- (*): every message has a posterior probability exponentially small, valid at the early stage
- This achieves the best known error exponent
  - achieves sphere packing exponent for high rates;
  - optimizing over η necessary;

Discussions

- Interpretations of the parameters ρ and η.
  - ρ is similar to Gallager's, to control the union bound, set up threshold between the correct and incorrect messages;
  - η is a slowing down factor, making $\varphi_t(\cdot)$ "flatter".
    - How much do we trust the history at time t;
    - Avoid a single message dominate $\varphi_t(\cdot)$ too early;
- Beyond the uniform progress assumption, dynamic assignment of η.