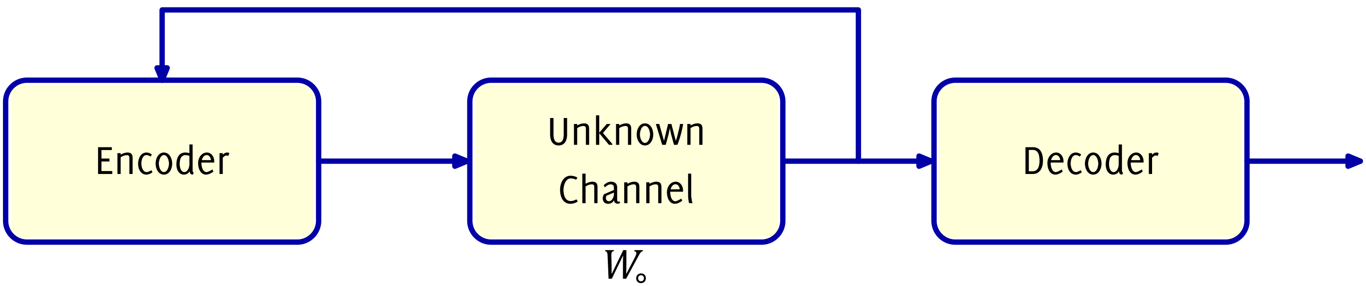
A training based scheme for communicating over unknown channels with feedback

Aditya Mahajan and Sekhar Tatikonda
Yale University

Allerton, Oct 2, 2009

The Setup



- Nature chooses a DMC $W_\circ$ from a family $\mathcal{W}$.
- The family $\mathcal{W}$ is known to the encoder and decoder; The choice of $W_\circ$ is not.
- The choice of $W_\circ$ does not change with time.
- What is the capacity and error exponent of this setup?

Capacity

Training based schemes can achieve any rate $R < C(W_\circ)$.

A feasible scheme: For block length t, train for length $\log t$. Use a standard code for the estimated channel.

Proof based on uniform continuity of entropy (and hence of mutual information) on the input distribution and the channel transition matrix

Not knowing the channel does not affect feedback capacity.

Error Exponents

Error Exponents: Known Channel

Fixed Length Communication

For output symmetric channels sphere packing bound is an upper bound

$$E_{sp}(R,W_\circ) = \sup_{\rho \geq 0} \left(E_0(\rho,W_\circ) - \rho R \right)$$

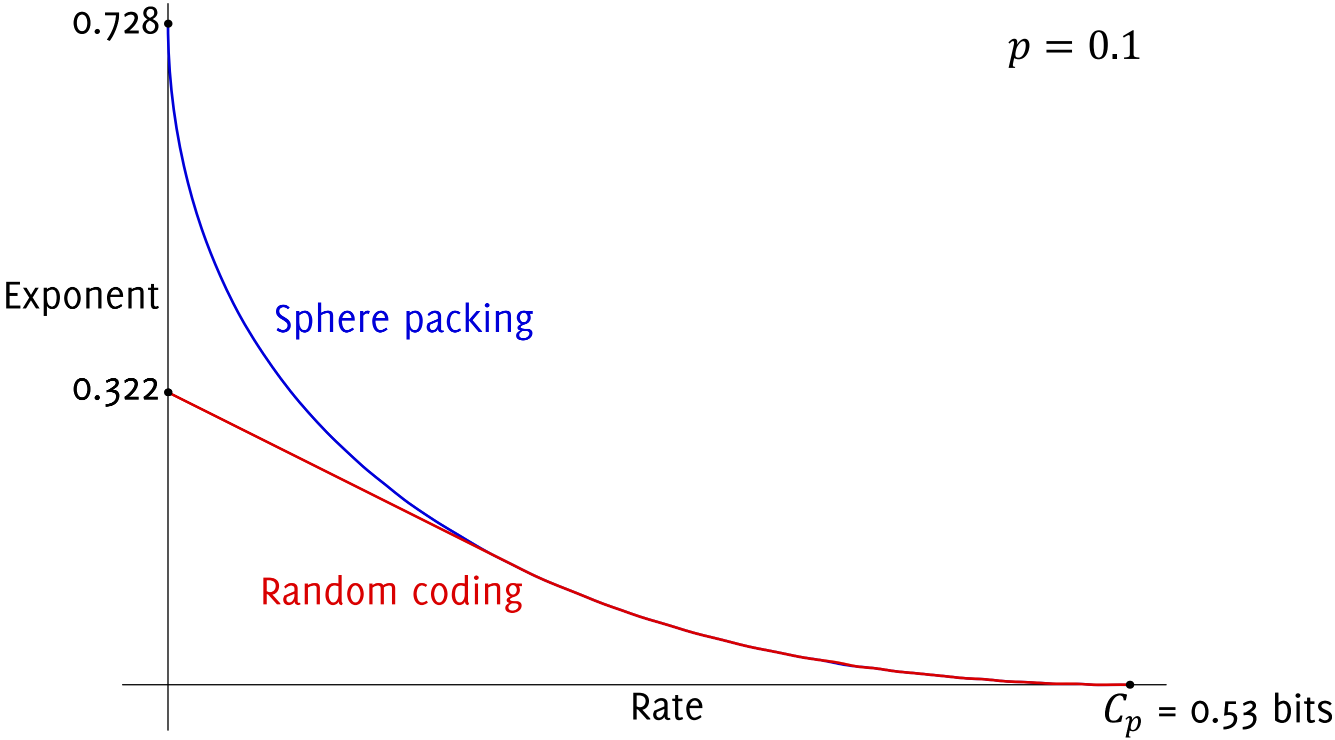
Various lower bounds for different rate regions

Variable Length Communication

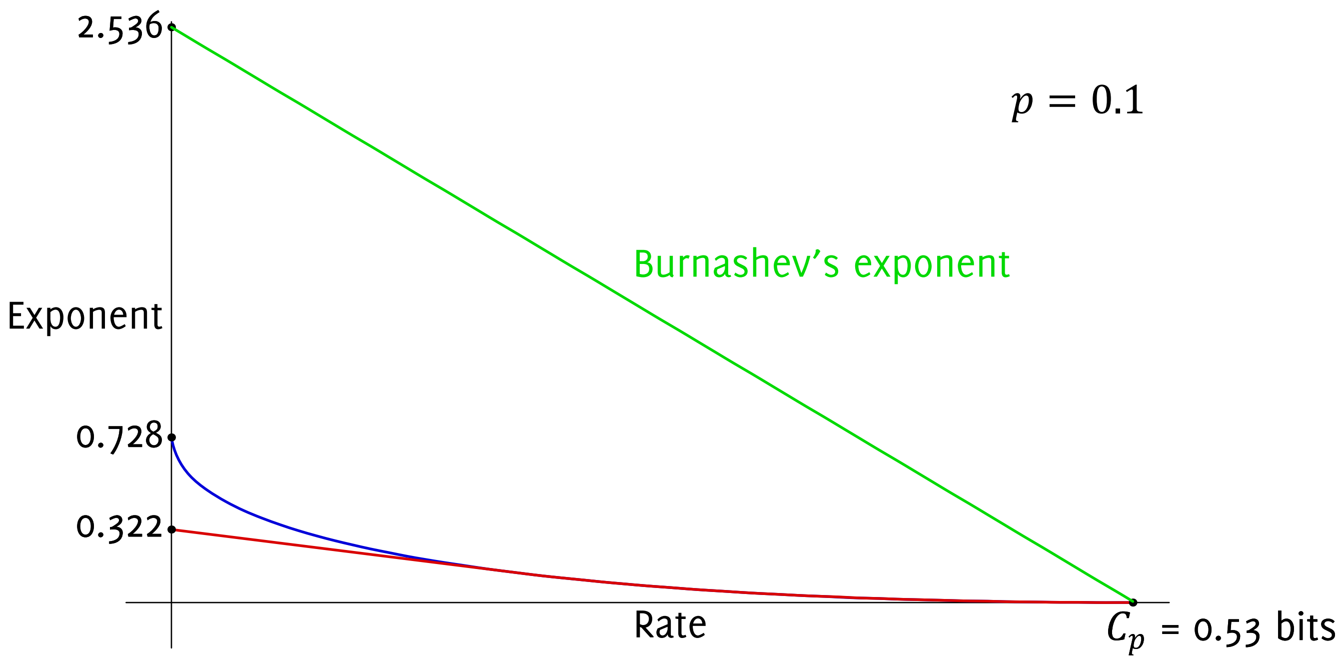
Characterized completely. Burnashev's Exponent

$$E_B(R,W_\circ) = \left(\max_{(x,x') \in \mathcal{X} \times \mathcal{X}} D(W_\circ(\cdot|x) \| W_\circ(\cdot|x')) \right) \left(1 - \frac{R}{C(W_\circ)} \right)$$

Error Exponent: Fixed Block length

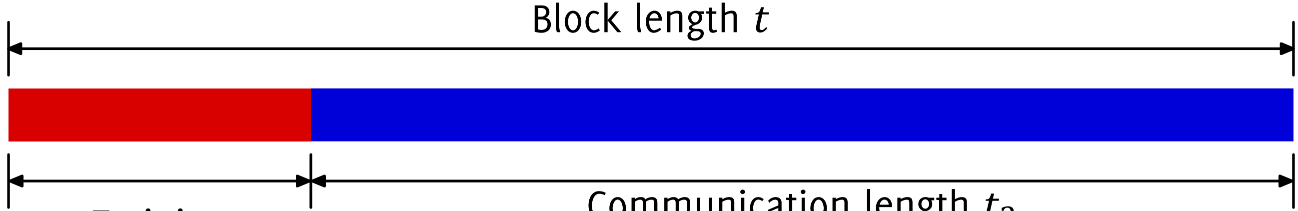


Error Exponent: Variable Block Length



Error Exponent: Unknown channel
Training based schemes for Fixed length communication

M. Feder and A. Lapidoth, Universal decoding for channels with memory, IT-98



$$P_e \approx e^{-t_1\left(\textcolor{red}{\blacksquare}\right)} + e^{-t_2\left(\textcolor{blue}{\blacksquare}\right)}$$

- $t_1 \approx t_2 \left(\textcolor{blue}{\blacksquare}\right) \Big/ \left(\textcolor{red}{\blacksquare}\right) \implies$ loss in rate
- $t_1 = o(t) \implies$ loss in exponent

Error Exponent: Unknown channel Bounds for Variable length communication

A. Tchamkerten and E. Telatar,

Variable length coding over an unknown channel, IT-06

For some families of channels adaptive schemes can achieve Burnashev's exponent

$$\mathcal{W}_{BSC} = \{\text{BSC}(p) : 0 \le p \le 1/2\} \quad \text{or} \quad \mathcal{W}_Z = \{\text{Z}(p) : 0 \le p \le 1\}$$

For some families of channels no scheme can achieve Burnashev's exponent

$$\mathcal{W}_p = \{\text{BSC}(p), \text{BSC}(1-p)\}, \quad 0 \le p \le \frac{1}{2}, \quad p \text{ known}$$

Error Exponent: Unknown channel
Training based schemes for variable length communication

A. Tchamkerten and E. Telatar,
On the use of training sequences for channel estimation, IT-06

The error exponent of training based schemes where the training length is fixed does not have positive slope at capacity.

Seems to suggest that training based schemes loose the biggest advantage of feedback — positive slope of the error exponent at capacity

Are training based schemes really bad?

Tchamkerten and Telatar assume training length is fixed

Feedback boosts error exponents because the transmitter can adapt to channel variations. Fixed length training takes away that advantage.

To boost error exponents, training must adapt to channel variations while communicating (not channel variations while training)

Must train multiple times

How do we achieve error exponents when the channel is known?

Burnashev's adaptive coding scheme

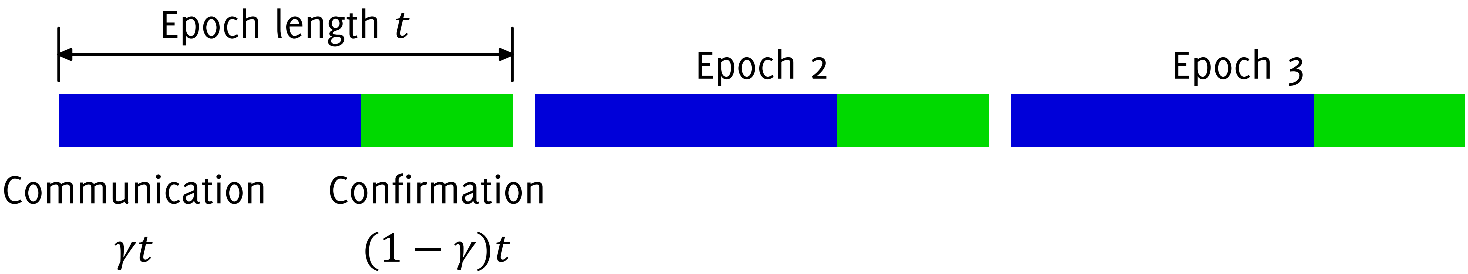
Track the evolution of EMI (empirical mutual information) and stop when EMI is large

Yamamoto-Itoh's iterative scheme

Transmit in multiple epochs consisting of communication phase and confirmation phase.

Yamamoto-Itoh's iterative scheme

H. Yamamoto and K. Itoh, Asymptotic performance of a modified Schalkwijk-Barron scheme with noiseless feedback, IT-79



Communication phase: Fixed length code of rate R/γ and length γt.

Confirmation phase: Confirm whether the decoding was correct or not.

$$P_e \leq e^{-\gamma t(\blacksquare)} \cdot e^{-(1-\gamma)t(\blacksquare)} \cdot \mathbb{E}[\text{number of epochs}]$$

Yamamoto-Itoh's iterative scheme

$$P_e \le e^{-\gamma t\left({\color{blue}\blacksquare}\right)} \cdot e^{(1-\gamma)t\left({\color{green}\blacksquare}\right)} \cdot \mathbb{E}[\text{number of epochs}]$$

Take $\gamma < 1 - \dfrac{R}{C}$

$$R/\gamma < C \implies \left({\color{blue}\blacksquare}\right) > 0\,, \qquad \left({\color{green}\blacksquare}\right) \approx D(W_\circ(\cdot|x = \text{NACK}) \| D(W_\circ(\cdot|x = \text{ACK}))$$

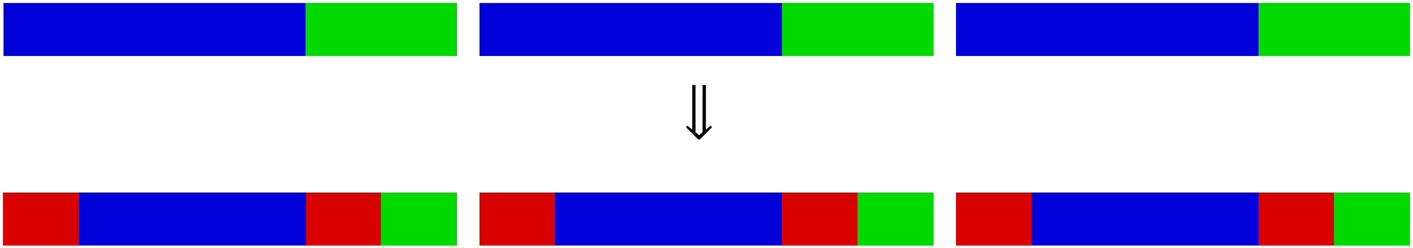
$$\mathbb{E}[\text{number of epochs}] \approx 1$$

$$E_B(R, W_\circ) \ge \left({\color{green}\blacksquare}\right)(1-\gamma) = {\color{green}D}\left(1 - \frac{R}{C}\right)$$

Main idea:

Use Yamamoto-Itoh's scheme with training

Proposed scheme



Train independently in each epoch. Ensures that $\mathbb{E}[\text{number of epochs}] \approx 1$

Within an epoch, train independently in the communication and the confirmation phase. Ensures that

$$P_e \leq \left(e^{-\beta_1 t\left({\color{red}\blacksquare}\right)} + e^{-\beta_2 t\left({\color{blue}\blacksquare}\right)} \right) \cdot \left(e^{-\beta_3\left({\color{red}\blacksquare}\right)} + e^{-\beta_4 t\left({\color{green}\blacksquare}\right)} \right) \cdot \mathbb{E}[\text{number of epochs}]$$

An example

$$\mathcal{W}_p = \{\mathsf{BSC}(p), \mathsf{BSC}(1-p)\}, \quad 0 \le p \le \frac{1}{2}, \quad p \text{ known}$$

Communicate at rate $R < C_p = 1 - h(p)$.

An example

$$\mathcal{W}_p = \{\mathsf{BSC}(p), \mathsf{BSC}(1-p)\}, \quad 0 \le p \le \frac{1}{2}, \quad p \text{ known}$$

Communicate at rate $R < C_p = 1 - h(p)$.

Communicate across multiple epochs of length t using an iterative coding scheme. Each epoch consists of four phases

- Training phase of length $\beta_1 t$: Send $\beta_1 t$ zeros.
- Communication phase of length $\beta_2 t$: Send $\lfloor 2^{tR} \rfloor$ messages at rate R/β_2.
- Re-training phase of length $\beta_3 t$: Send $\beta_3 t$ zeros.
- Confirmation phase of length $\beta_4 t$: Send $\beta_4 t$ zeros (ACK) or $\beta_4 t$ ones (NACK).

Rate and Probability of Error

Average Rate: $$\lim_{t\to\infty} \frac{\log(\text{number of messages})}{\mathbb{E}[\text{number of epochs}]\cdot t} = R.$$

Probability of error:

$$P_e \leq \left(e^{-\beta_1 t\left(\textcolor{red}{\blacksquare}\right)} + e^{-\beta_2 t\left(\textcolor{blue}{\blacksquare}\right)}\right) \cdot \left(e^{-\beta_3\left(\textcolor{red}{\blacksquare}\right)} + e^{-\beta_4 t\left(\textcolor{green}{\blacksquare}\right)}\right) \cdot \mathbb{E}[\text{number of epochs}]$$

Error Exponent

Error Exponent:

$$E_s(R,W_\circ) = -\lim_{t\to\infty} \frac{\log P_e(t)}{\mathbb{E}[\text{number of epochs}]\cdot t}$$

$$\geq -\lim_{t\to\infty}\frac{1}{t}\log\left(e^{-\beta_1 t\left(\textcolor{red}{\blacksquare}\right)} + e^{-\beta_2 t\left(\textcolor{blue}{\blacksquare}\right)}\right)$$

$$-\lim_{t\to\infty}\frac{1}{t}\log\left(e^{-\beta_3\left(\textcolor{red}{\blacksquare}\right)} + e^{-\beta_4 t\left(\textcolor{green}{\blacksquare}\right)}\right)$$

$$-\lim_{t\to\infty}\frac{1}{t}\log\mathbb{E}[\text{number of epochs}]$$

Error Exponent: 1st term

In the training + communication phase, train for $o(t)$ time. This gives a poor decoding error exponents, but for Yamamoto Itoh's scheme all we just want the exponent to be positive. So,

$$-\lim_{t\to\infty}\frac{1}{t}\log\left(e^{-\beta_1 t\left({\color{red}\blacksquare}\right)}+e^{-\beta_2 t\left({\color{blue}\blacksquare}\right)}\right)\approx\lim_{t\to\infty}\beta_1\left({\color{red}\blacksquare}\right)>0$$

Error Exponent: 2nd term

In the re-training + confirmation phase, train such that $\beta_3 = \beta_4 \left({\color{green}\blacksquare}\right) / \left({\color{red}\blacksquare}\right)$. This results in a loss in rate, but the rate of the confirmation phase is zero anyways. So,

$$-\lim_{t\to\infty}\frac{1}{t}\log\left(e^{-\beta_3\left({\color{red}\blacksquare}\right)} + e^{-\beta_4 t\left({\color{green}\blacksquare}\right)}\right) \geq \frac{\left({\color{red}\blacksquare}\right)\cdot\left({\color{green}\blacksquare}\right)}{\left({\color{red}\blacksquare}\right)+\left({\color{green}\blacksquare}\right)}\cdot(\beta_3+\beta_4)$$

Error Exponent: 3rd term

When $P_e \approx 0$, then the number of transmission epochs ≈ 1. So,

$$-\lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}[\text{number of epochs}] \approx 0$$

Error Exponent

Choose $\beta_1 + \beta_2 = \frac{R}{C_p}$ and $\beta_3 + \beta_4 = \left(1 - \frac{R}{C_p}\right)$. Then,

$$E_s(R,W_\circ) \geq \frac{\left({\color{red}\blacksquare}\right)}{\left({\color{red}\blacksquare}\right) + \left({\color{green}\blacksquare}\right)} \cdot \underbrace{\left({\color{green}\blacksquare}\right) \cdot \left(1 - \frac{R}{C}\right)}_{\text{Burnashev's Exp}}$$

Error Exponent

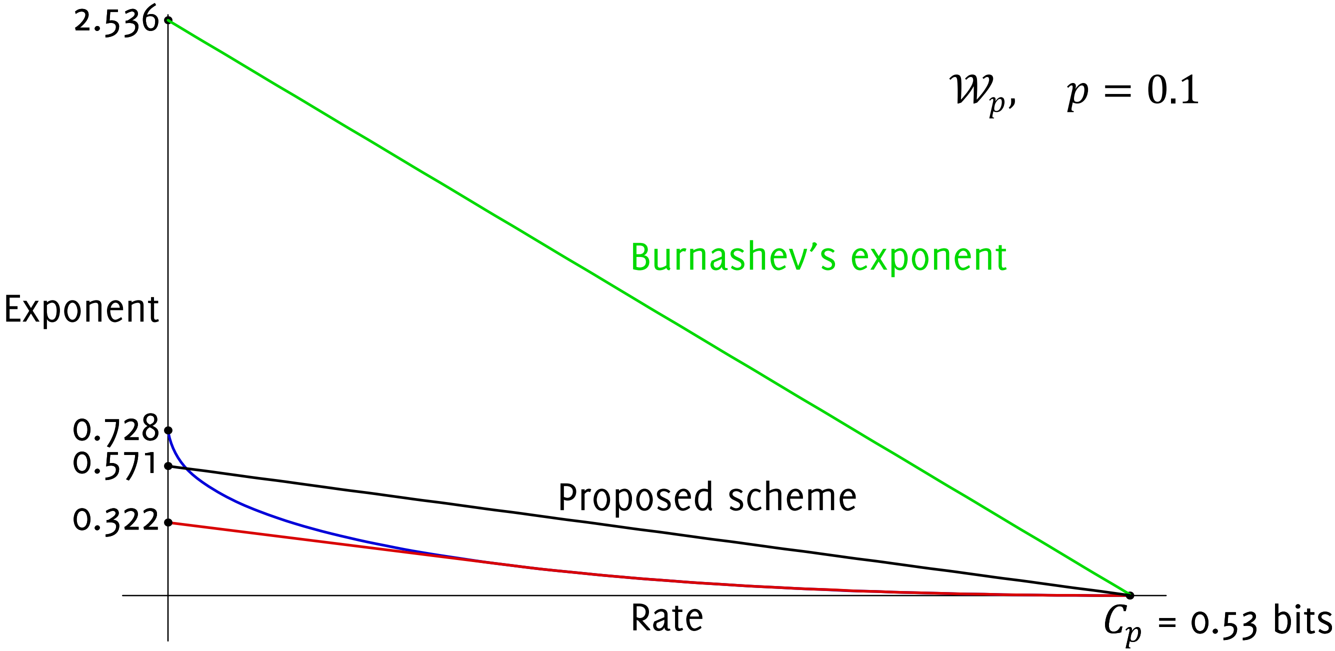
Choose $\beta_1 + \beta_2 = \frac{R}{C_p}$ and $\beta_3 + \beta_4 = \left(1 - \frac{R}{C_p}\right)$. Then,

$$E_s(R, W_\circ) \geq \frac{\left(\color{red}\blacksquare\right)}{\left(\color{red}\blacksquare\right) + \left(\color{green}\blacksquare\right)} \cdot \underbrace{\left(\color{green}\blacksquare\right) \cdot \left(1 - \frac{R}{C}\right)}_{\text{Burnashev's Exp}}$$

$$= {\color{red}\alpha}\, \underbrace{D(p \| 1 - p) \left(1 - \frac{R}{C_p}\right)}_{\text{Burnashev's Exp}}$$

where ${\color{red}\alpha} = \frac{D(0.5 \| p)}{D(0.5 \| p) + D(p \| 1 - p)}$

Error Exponent: Performance



Salient features

- The training based scheme is simple, yet it comes within a constant fraction of the Burnashev's exponent.

- Allowing variable length training ensures that the error exponent has positive slope at capacity.

Proposed scheme will not work when we have to communicate at different rates for different channels in the family.

Conclusion

Training based schemes do not necessarily have poor error exponents. Schemes with variable training length need further investigation.

Future Directions

- Error exponents of best training based schemes.
- Error exponents of best universal schemes.