Training-Based Schemes are Suboptimal for High Rate Asynchronous Communication

V. Chandar[†] A. Tchamkerten[‡] G. Wornell[†]

[†]MIT [‡]Telecom ParisTech

ITW 2009

Synchronization and error correction

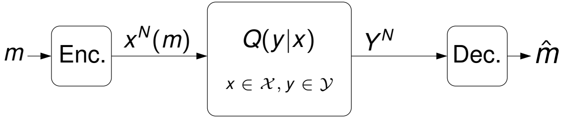
- Information theory has mostly focused on synchronous communication systems.
- Synchronization considered a separate topic from information transmission.
- Synchronization considered as a side engineering problem without much consequence on rate.
- Can asynchronism be systematically ignored in the context of data transmission?

Contribution

- Identify communication regimes where asynchronism can/can't be ignored.
- Identify communication regimes where training based schemes are suboptimal.

Outline

Synchronous DMC



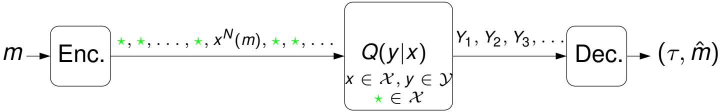
Performance criterion

- Rate: $R = \frac{\log M}{N}$
- error probability: $\mathbb{P}(\hat{m} \neq m)$

R is achievable if, for any $\varepsilon > 0$, there exist length N codewords that

- achieve rate $\geq R - \varepsilon$
- yield error probability $\leq \varepsilon$.

Asynchronous DMC

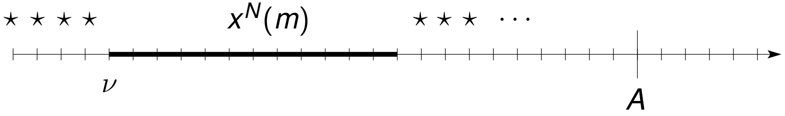


- Transmitter: starts sending $x^N(m)$ at a random time $\nu \in [1, 2, \ldots, A]$. $A =$ "asynchronism level".
- Receiver: decodes on the basis of a *sequential decoder*
  - within $[\nu, \nu+1, \nu+N-1]$ observes a noisy version of $x^N(m)$
  - otherwise observes "noise"

Remark: "$\star \in \mathcal{X}$" is a parameter of the channel; only a *single* message is sent.

Asynchronous DMC (cont.)

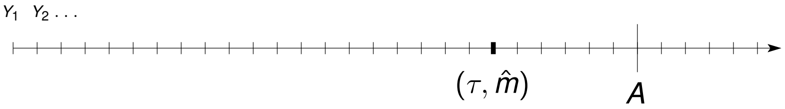
Transmitter:



Channel:

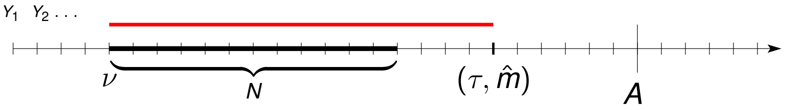
$$Q(y|x),\ (x,y) \in \mathcal{X} \times \mathcal{Y}$$

Sequential decoder:

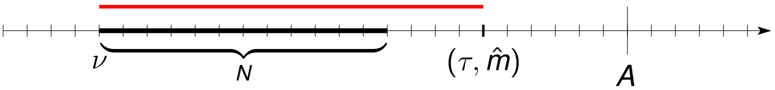


Asynchronous DMC (cont.)

Natural rate definition: $R = \dfrac{\log M}{\mathbb{E}(\tau - \nu)^+}$



Performance criterion



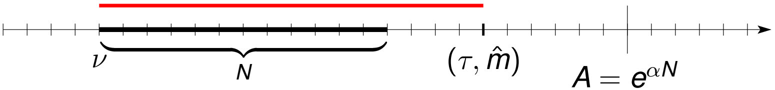
(R, A) is achievable if, for any $\varepsilon > 0$, there exist length N codewords that

- operate under asynchronism level A
- achieve rate $\geq R - \varepsilon$
- yield error probability $\leq \varepsilon$.

Problem is 'trivial' if A doesn't scale with N.

Natural scaling: $\boxed{A = e^{\alpha N} \quad \alpha \geq 0 \quad \text{asynchronism exponent}}$

Performance criterion (cont.)



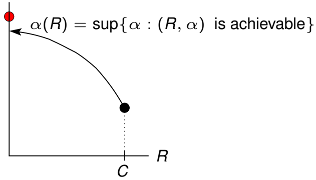
(R, α) is achievable if, for any $\varepsilon > 0$, there exist length N codewords that

- operate under asynchronism level $A = e^{\alpha N}$
- achieve rate $\geq R - \varepsilon$
- yield error probability $\leq \varepsilon$.

Goal: capacity region

Find $R(\alpha) \triangleq \sup\{R : (R, \alpha) \text{ is achievable}\}$
Or, find $\alpha(R) \triangleq \sup\{\alpha : (R, \alpha) \text{ is achievable}\}$

Capacity region: review of results

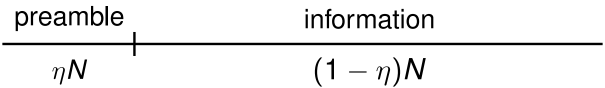


- $\alpha(R)$ non-increasing.
- Lower bound:

$$\alpha(R) \geq \max_{P: I(PQ) \geq R} \min_{V} \max\{D(V||(PQ)_Y), D(V||Q(\cdot|\star))$$

- Upper bound: tight for certain channels.
- Except when $Q(\cdot|\star) =$ capacity achieving output distribution of the synchronized channel, $\alpha(R)$ is always discontinous at $R = C$.
- 'Random coding' capacity.
- $\alpha(R = 0)$ known for all channels, discontinuity at $R = 0$.

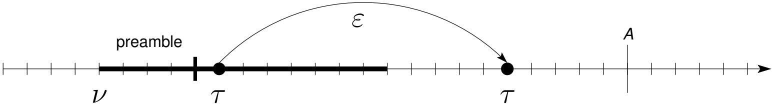
Training based schemes



i. codewords: common preamble followed by information bits
ii. receiver: detects the preamble, decode upcoming bits
iii. information symbols: shouldn't act as a second preamble

$$\Pr(\tau \geq \nu + N + \eta N | \tau \geq \nu + \eta N) \geq \varepsilon$$

for some arbitrary $\varepsilon > 0$ independent of N.



Training based schemes (cont.)

Theorem

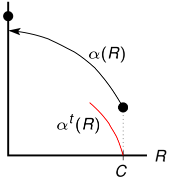
The ‘capacity’ of training based schemes is

$$\alpha^t(R) = \left(1 - \frac{R}{C}\right) \max_{P \in \mathcal{P}^{\mathcal{X}}} \min_{W \in \mathcal{P}^{\mathcal{Y}|\mathcal{X}}} \max\{D(PW||PQ), D(PW||PQ(\cdot|\star))\}\,.$$

Hence, e.g., if $Q(\cdot|\star)$

- can produce all channel outputs
- is not equal to the capacity achieving output distribution

training based schemes are suboptimal at high rates.



Sketch of the proof: achievability

To achieve rate R

1. Choose $\eta = (1 - R/C)$
2. Preamble empirical distribution

$$\arg\max_{P \in \mathcal{P}^{\mathcal{X}}} \min_{W \in \mathcal{P}^{\mathcal{Y}|\mathcal{X}}} \max\{D(PW||PQ), D(PW||PQ(\cdot|\star)\}$$

3. Use capacity achieving code over last $(1 - \eta)$ symbols.
4. Preamble detection: stop at time n whenever

$$D(P\hat{W}_{n-N+1}^{n}||PQ(\cdot|\star)) \geq \alpha$$

5. Decode $(1 - \eta)N$ upcoming symbols.

Sketch of the proof: converse

To prove

$$\alpha(R) \leq \left(1 - \frac{R}{C}\right) \max_{P \in \mathcal{P}^{\mathcal{X}}} \min_{W \in \mathcal{P}^{\mathcal{Y}|\mathcal{X}}} \max\{D(PW||PQ), D(PW||PQ(\cdot|\star)\}$$

for training based schemes.

1. Fix η, P empirical distribution of the preamble
2. If $R > 0$, then for any $W \in \mathcal{P}^{\mathcal{Y}|\mathcal{X}}$ either

$$\eta D(PW||PQ) \geq \alpha \quad \text{or} \quad \eta D(PW||PQ(\cdot|\star)) > \alpha$$

   via a change of measure argument.
3. Hence

$$\alpha \leq \eta \max_{P \in \mathcal{P}^{\mathcal{X}}} \min_{W \in \mathcal{P}^{\mathcal{Y}|\mathcal{X}}} \max\{D(PW||PQ), D(PW||PQ(\cdot|\star)\}$$

4. Highest value of η compatible with rate R communication

$$\eta = (1 - R/C)$$

Summary

- Communication model:
  - for sporadic communication $A \gg N$ (e.g., monitoring sensor networks)
  - rate definition favors ‘very short communication’
- Identified channels and communication regime where training is suboptimal; each sent bit should carry information and help detect the codeword.
- Open problems
  - ‘selfsync’ codes: codes with good isolation *and* detection properties
  - multiple messages, multiple access

Selfsync coding

1. Target rate R
2. Available: family of rate R compatible codes $\{\mathcal{C}^{(i)}\}$, i.e., $I(P^{(i)}Q) \geq R$ where $P^{(i)}$ is the empirical distribution of $\mathcal{C}^{(i)}$
3. Choose best code $\mathcal{C}^{(i^*)}$ according to

$$\arg\max_i \min_W \max\{D(P^{(i)}W||P^{(i)}Q), D(P^{(i)}W||P^{(i)}Q(\cdot|\star))\}$$

4. Detection rule: stop whenever $D(P^{(i^*)}W^n_{n-N+1}||P^{(i^*)}Q(\cdot|\star)) > \alpha$
5. Isolation rule: decode past N symbols using favorite decoding rule.

Gaussian channel

- $Q(\cdot|\star) \sim \mathcal{N}(0, \sigma^2)$
- For rate R communication under power constraint, we can achieve asynchronism exponents up to

$$\max_{\substack{P: I(PQ) \geq R \\ \mathbb{E}_P X^2 \leq \text{power}}} \min_V \max\{D(V||(PQ)_Y), D(V||Q_\star)\},$$

  where P and V in the optimization are distributions over the reals.
- Outer maximization over Gaussian only: best input has mean μ as large as possible given rate and power constraints

$$R = \frac{1}{2} \log \left(1 + \frac{\text{power} - \mu^2}{\sigma^2} \right)$$