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

Asynchronous DMC: review of results

Suboptimality of training based schemes

Summary
Asynchronous DMC: review of results

Summary

Synchronous DMC

\[
m \xrightarrow{\text{Enc.}} x^N(m) \xrightarrow{Q(y|x)} Y^N \xrightarrow{\text{Dec.}} \hat{m}
\]

\(x \in \mathcal{X}, y \in \mathcal{Y}\)

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:

\[ x^N(m) \]

Channel:

\[ Q(y|x), (x, y) \in X \times Y \]

Sequential decoder:

\[ (\tau, \hat{m}) \]
Asynchronous DMC (cont.)

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

\[ \mathcal{E} \]

Diagram showing sequences of events and parameters related to the natural rate definition.
(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: $A = e^{\alpha N}$, $\alpha \geq 0$ asynchronism exponent
Performance criterion (cont.)

\[ \nu \leq N \leq (\tau, \hat{m}) \]

\[ A = e^{\alpha N} \]

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

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

\[ \alpha(R) = \sup \{ \alpha : (R, \alpha) \text{ is achievable} \} \]
Training based schemes

- **Preamble**: common preamble followed by information bits
- **Receiver**: detects the preamble, decodes upcoming bits
- **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

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

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

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

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

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

5. Decode $(1 - \eta)N$ upcoming symbols.
Sketch of the proof: converse

To prove

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

for training based schemes.

1. Fix $\eta$, $P$ empirical distribution of the preamble
2. If $R > 0$, then for any $W \in \mathcal{P}^Y|X$ 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
   $$\alpha \leq \eta \max_{P \in \mathcal{P}^X} \min_{W \in \mathcal{P}^Y|X} \max \{D(PW || PQ), D(PW || PQ(\cdot | \star))\}$$
4. Highest value of $\eta$ 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 $\{C^{(i)}\}$, i.e., $I(P^{(i)}Q) \geq R$ where $P^{(i)}$ is the empirical distribution of $C^{(i)}$
3. Choose best code $C^{(i^*)}$ according to
   \[
   \arg \max_i \min_W \max \{ D(P^{(i)}W \mid P^{(i)}Q), D(P^{(i)}W \mid P^{(i)}Q(\cdot \mid \star)) \}
   \]
4. Detection rule: stop whenever
   \[
   D(P^{(i^*)}W_{n-N+1}^n \mid P^{(i^*)}Q(\cdot \mid \star)) > \alpha
   \]
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

$$\max_{P : I(PQ) \geq R} \min_{V} \max \{ D(V \| (PQ)_V), D(V \| Q_\star) \},$$

where $P$ and $V$ in the optimization are distributions over the reals.

- Outer maximization over Gaussian only: best input has mean $\mu$ as large as possible given rate and power constraints

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