Achieving the Empirical Capacity for Any Noise Sequence Using Feedback

Ofer Shayevitz and Meir Feder

Tel Aviv University
Problem Setting

A binary modulo-additive channel with noiseless feedback:

- Individual noise sequence $z$ of length $n$
- May be generated by an adversary that knows the message
- Goal - Reliable communications with maximal rate $R$
Berlekamp’s Classical Result

Capacity Upper Bound

- $1 - h_b(p)$
- Tangential Line
- $p$ - Known in Advance
- No Errors Allowed

$R$ vs. $P$ - Allowed Fraction of Errors

$\frac{1}{3}$ to $\frac{1}{2}$
Berlekamp’s Classical Result

Capacity Upper Bound
Our Result...

$1 - h_b(p)$

Tangential Line

$\frac{1}{3}$ $\frac{1}{2}$ $P$ - Precise Fraction of Errors

• $p$ – Unknown in Advance!!
Arbitrarily Varying Channel (AVC) setting with 2 states (clean/inverting) and feedback

This AVC is *symmetrizable* [CsiszárNarayan‘88]  \( \Rightarrow \)
• Arbitrarily Varying Channel (AVC) setting with 2 states (clean/inverting) and feedback

• This AVC is *symmetrizable* [CsiszárNarayan‘88] \( \Rightarrow \)

\[ \textbf{The Capacity is Zero !} \]
For an individual noise sequence \( z \) with any empirical probability \( p \) unknown in advance:

\[
R \approx 1 - h_b(p)
\]
For an individual noise sequence $z$ with any empirical probability $p$ unknown in advance:

$$R \cong 1 - h_b(p)$$

Specifically: We introduce a randomized sequential transmission scheme, attaining a rate $R = R(z)$ so that for some $\delta(n) \rightarrow 0$

$$P_r\left(R \geq 1 - h_b(p) - \delta(n)\right)_{n \rightarrow \infty} 1$$

with a vanishing error probability w.r.t. the randomization
Background - The Horstein Scheme

[Horstein‘63]

- Capacity achieving scheme for the $BSC_p$ with feedback
- The message is represented by a point $\omega \in [0, 1)$
- The transmitter steers the receiver in the direction of $\omega$
- The receiver gradually zooms in on $\omega$
\( f_0(\theta) \) - Start with a uniform density for the message point

\( a_0 \) - Median point w.r.t. \( f_0(\theta) \)
\( f_0(\theta) \) - Start with a uniform density for the message point

\( a_0 \) - Median point w.r.t. \( f_0(\theta) \)

**Background - The Horstein Scheme**
$f_0(\theta)$ - Start with a uniform density for the message point

$a_0$ - Median point w.r.t. $f_0(\theta)$

Background - The Horstein Scheme
$f_1(\theta)$ - a-posteriori density given the received bit

$a_1$ - Median point w.r.t. $f_1(\theta)$
$f_1(\theta)$ - a-posteriori density given the received bit

$a_1$ - Median point w.r.t. $f_1(\theta)$
$f_1(\theta)$ - a-posteriori density given the received bit

$a_1$ - Median point w.r.t. $f_1(\theta)$

Background - The Horstein Scheme
$f_2(\theta)$ - a-posteriori density given two received bits

$a_2$ - Median point w.r.t. $f_2(\theta)$
\[ f_2(\theta) \] - a-posteriori density given two received bits

\[ a_2 \] - Median point w.r.t. \( f_2(\theta) \)

Background - The Horstein Scheme
$f_2(\theta)$ - a-posteriori density given two received bits

$a_2$ - Median point w.r.t. $f_2(\theta)$
$f_3(\theta)$ - a-posteriori density given three received bits

$a_3$ - Median point w.r.t. $f_3(\theta)$
Hopefully after many channel uses...

\[ f_k(\theta) \]

\( \omega \quad 1 \quad \theta \)
Note...

- After $n$ Horstein iterations there are $n + 1$ intervals within each $f_n(\theta)$ is constant.

- Defining $n_1 \triangleq \sum z_k$ we have

\[
    f_n(\theta = \omega) = 2^n p^{n_1} (1 - p)^{n-n_1} = 2^n P_r(z)
\]
Background - The KT Probability Estimator

[Krichevsky Trofimov '81]

- **Goal** - Sequential probability assignment to a sequence $z$

\[
\hat{p}_{k+1} = \frac{\sum_{j=1}^{k} z_j + \frac{1}{2}}{k + 1} \quad \Rightarrow \quad \hat{P}(z) = \prod_{k=1}^{n} \hat{p}_k^{z_k} (1 - \hat{p}_k)^{1-z_k}
\]

- **Known result** - If $z$ has $np$ ones then

\[
-\frac{1}{n} \log \hat{P}(z) \leq h_b(p) + \frac{\log n}{2n}
\]
Our Transmission Scheme

Idea - Perform Horstein iterations. What $p$ to use for an individual noise sequence? Plug in a KT estimate instead!
Our Transmission Scheme

**Idea** - Perform Horstein iterations. What $p$ to use for an individual noise sequence? Plug in a KT estimate instead!

**Problem** - How to make the estimate available to the receiver?
Our Transmission Scheme

Idea - Perform Horstein iterations. What $p$ to use for an individual noise sequence? Plug in a KT estimate instead!

Problem - How to make the estimate available to the receiver?

Solution - Slow estimate update, randomization (Later...)

Assume (for now) the KT estimate is available at the receiver
• Horstein iterations - Replace \((p, 1 - p)\) with \((\hat{p}_k, 1 - \hat{p}_k)\)

• At the end of the block

\[
f_n(\theta = \omega) = 2^n \prod_{k=1}^{n} \hat{p}_k^{z_k} (1 - \hat{p}_k)^{1-z_k} = 2^n \hat{P}(z)
\]

The probability assigned to the noise sequence by the KT estimator

• Again, \(n + 1\) intervals within each \(f_n(\theta)\) is constant
• Horstein iterations - Replace \((p, 1 - p)\) with \((\hat{p}_k, 1 - \hat{p}_k)\)

• At the end of the block

\[
\begin{align*}
 f_n(\theta = \omega) &= 2^n \prod_{k=1}^{n} \hat{p}_k^{z_k} (1 - \hat{p}_k)^{1-z_k} = 2^n \hat{P}(z)
\end{align*}
\]

The probability assigned to the noise sequence by the KT estimator

• Again, \(n + 1\) intervals within each \(f_n(\theta)\) is constant

Assume (for now) the receiver knows the interval containing \(\omega\)
Let $2^{-\ell}$ be the size of the interval containing $\omega$:

$$2^{-\ell} \cdot 2^n \hat{P}(z) \leq 1 \quad \Rightarrow \quad \ell \geq n + \log \hat{P}(z)$$
Let $2^{-\ell}$ be the size of the interval containing $\omega$:

$$2^{-\ell} \cdot 2^n \hat{P}(z) \leq 1 \implies \ell \geq n + \log \hat{P}(z)$$

$$R \triangleq \frac{\ell}{n} \geq 1 + \frac{1}{n} \log \hat{P}(z) \geq 1 - h_b(p) - \frac{\log n}{2n}$$
Let $2^{-\ell}$ be the size of the interval containing $\omega$:

$$2^{-\ell} \cdot 2^n \hat{P}(z) \leq 1 \implies \ell \geq n + \log \hat{P}(z)$$

$$R \triangleq \frac{\ell}{n} \geq 1 + \frac{1}{n} \log \hat{P}(z) \geq 1 - h_b(p) - \frac{\log n}{2n}$$

**Problem** - Must tell the receiver which is the message interval
Let $2^{-\ell}$ be the size of the interval containing $\omega$:

$$2^{-\ell} \cdot 2^n \hat{P}(z) \leq 1 \implies \ell \geq n + \log \hat{P}(z)$$

\[
R \triangleq \frac{\ell}{n} \geq 1 + \frac{1}{n} \log \hat{P}(z) \geq 1 - h_b(p) - \frac{\log n}{2n}
\]

**Problem** - Must tell the receiver which is the message interval

**Solution** - Communicate this together with the KT estimate...

**Estimate + Interval index = Update Information (UI)**
How Much UI is Needed?

• Divide into blocks of size $\sqrt{n}$, send UI once per block

• Estimator redundancy increased by $\sim \frac{\log n}{\sqrt{n}}$ (negligible)

• Estimate updates require $\sim \frac{1}{2} \log n$ bits per block

• Message interval index require $\sim \log n$ bits per block
How Much UI is Needed?

- Divide into blocks of size $\sqrt{n}$, send UI once per block
- Estimator redundancy increased by $\sim \frac{\log n}{\sqrt{n}}$ (negligible)
- Estimate updates require $\sim \frac{1}{2} \log n$ bits per block
- Message interval index require $\sim \log n$ bits per block

Total UI Rate $\sim \frac{\log n}{\sqrt{n}}$ Negligible!
Problem - Adversary may corrupt the UI
Problem - Adversary may corrupt the UI

Solution - Transmit UI over random positions in each block
Problem - Adversary may corrupt the UI

Solution - Transmit UI over random positions in each block

• Positions agreed upon via feedback or common randomness

• Position selection ~ \log^2 n \text{ bits per block (negligible)}

• Use, e.g., a repetition code for UI (slightly increases UI rate)

• Effective channel for UI is ~ “BSC”
Problems - Polarity of the “BSC”, “Bad” blocks
Problems - Polarity of the “BSC”, “Bad” blocks

Solution - Add a training sequence to UI, discard ”bad” blocks
**Problems** - Polarity of the “BSC”, “Bad” blocks

**Solution** - Add a training sequence to UI, discard ”bad” blocks

Mismatch of training estimate and actual channel $\Rightarrow$ Error!

Fortunately...

*Error probability w.r.t. randomization tends to zero as $n \to \infty$*
Final Scheme Outline

Divide into blocks of size $\sqrt{n}$. In each block
Divide into blocks of size $\sqrt{n}$. In each block
Divide into blocks of size $\sqrt{n}$. In each block
Divide into blocks of size $\sqrt{n}$. In each block
Divide into blocks of size $\sqrt{n}$. In each block

- Use training for UI decoding and block discarding
• Decode the bits representing the last message interval

• UI is correctly decoded with probability $\rightarrow 1$

• If UI is correct then the bit decoding rate satisfies

$$R \geq 1 - h_b(p) - \delta(n)$$

with probability $\rightarrow 1$
Summary

- Transmission scheme over a binary modulo-additive channel with an individual noise sequence and feedback

- Rate approaching $1 - h_b(p)$ for any noise sequence, classical capacity is zero

- Beating the Berlekamp bound without knowing $p$ by using randomization, at a cost of a negligible error probability
Future Research

• By using a universal predictor at the transmitter, may approach $1 - h_b(\pi(z))$

• Modify scheme to approach $1 - \rho(z)$?

• Applications in multi-user communications